POSITIVE DEFINITENESS AND THE STOLARSKY INVARIANCE PRINCIPLE

DMITRIY BILYK, RYAN W. MATZKE, AND OLEKSANDR VLASIUK

ABSTRACT. In this paper we elaborate on the interplay between energy optimization, positive definiteness, and discrepancy. In particular, assuming the existence of a K-invariant measure μ with full support, we show that conditional positive definiteness of a kernel K is equivalent to a long list of other properties: including, among others, convexity of the energy functional, inequalities for mixed energies, and the fact that μ minimizes the energy integral in various senses. In addition, we prove a very general form of the Stolarsky Invariance Principle on compact spaces, which connects energy minimization and discrepancy and extends several previously known versions.

CONTENTS

1. INTRODUCTION

1.1. **Energy Minimization.** Let Ω be a compact metric space and let the kernel $K : \Omega \times \Omega \to \mathbb{R}$ be continuous and symmetric, i.e. $K(x, y) = K(y, x)$ for all $x, y \in \Omega$. We denote by $\mathcal{M}(\Omega)$ the set

of finite regular signed Borel measures on Ω, and by $\mathbb{P}(\Omega)$ the set of Borel probability measures on Ω. Given $\mu, \nu \in \mathcal{M}(\Omega)$, we define their mixed K-energy as

$$(1.1) \qquad I_K(\mu, \nu) = \int\limits_{\Omega} \int\limits_{\Omega} K(x, y) \, d\mu(x) d\nu(y),$$

and the K-energy (also referred to as energy integral or energy functional) of μ to be

$$(1.2) \qquad I_K(\mu) := I_K(\mu, \mu) = \int\limits_{\Omega} \int\limits_{\Omega} K(x, y) \, d\mu(x) d\mu(y).$$

We are interested in finding the optimal (maximal or minimal, depending on K) values of $I_K(\mu)$ over all $\mu \in \mathbb{P}(\Omega)$, as well as extremal measures for which these values are achieved, i.e. equilibrium measures with respect to K.

For a measure $\mu \in \mathbb{P}(\Omega)$, the *potential* U_K^μ of μ with respect to K, defined as

$$(1.3) \qquad U_K^\mu(x) := \int\limits_{\Omega} K(x, y) \, d\mu(y), \quad x \in \Omega,$$

plays an important role in the study of energy minimization, see, e.g., Theorem 2.14.

It is well known that energy minimization is closely connected to the positive definiteness of the kernel. In Section 2 we further explore this property and its variants, such as conditional positive definiteness and positive definiteness up to an additive constant. We discuss various relations between these properties and energy minimization, as well as inequalities for mixed energies, convexity of the energy functional, Hilbert–Schmidt operators, potential theory, etc. A synopsis of the main results presented in Section 2 may be found right before the beginning of §2.1. This section contains mostly background material (although several results do seem to be new) and is completely self-contained.

In Section 3, we restrict our attention to the case when there exists a reasonable candidate for an energy minimizer - namely, a *K-invariant measure* (i.e., a measure whose potential U_K^μ with respect to K is constant on Ω). This is a very natural class of measures which includes, for example, the uniform (Lebesgue) surface measure on the sphere $\mathbb{S}^{d-1}$ when the kernel K is rotationally invariant. Assuming the existence of an invariant measure, much more can be said about the topics of Section 2. In particular, conditional positive definiteness of K is equivalent to the fact that μ minimizes I_K over all *signed measures of mass one* (Theorem 3.5), conditional positive definiteness and positive definiteness up to an additive constant are equivalent, which is not true in general (Lemma 3.7), and local minimizers are necessarily global (see §3.5). Many of the proofs in this section rely on a simple yet crucial identity (3.2) of Lemma 3.3.

Finally, in Section 4, we further focus on the situation when there exists a K-invariant measure *with full support*. These assumptions really tie the discussion together: Theorem 4.1 states that in this case, conditional positive definiteness is equivalent to nine other natural properties, such as various versions of (local or global) convexity, mixed energy bounds, and minimization of I_K by μ. Similar statements are provided for positive definiteness and conditional strict positive definiteness (Theorems 4.2 and 4.3).

Admittedly, some portion of the results discussed above are well known in the field, see, for example, an excellent discussion in [BHS]. However, we undertook an extensive study of the literature, and it appears that many of the implications are in fact new (we carefully point those out in the text), while some others are scattered in the literature. This paper presents a unified, comprehensive, and self-contained discussion of both known and new connections, which results in an impressively long list of equivalent characterizations of conditional positive definiteness provided in Theorem 4.1.

Section 5 introduces an application of these results to discrepancy theory (see §1.2 below), while Section 6 specializes the prior discussion to the sphere, i.e. the case when $\Omega = \mathbb{S}^{d-1}$ and $K(x, y) = F(\langle x, y \rangle)$ is a rotationally invariant kernel. In this setting, our results recover and generalize various well known results.

1.2. **Energy and Discrepancy: the Stolarsky Invariance Principle.** Discrepancy is a classical way to assess the quality of a finite point distribution $\omega_N = \{z_1, \ldots, z_N\} \subset \Omega$ by comparing its empirical measure to some chosen (usually uniform) measure μ on some test sets. Vast literature exists on discrepancy theory [BC, Ma, KN].

Discrepancy is closely connected to energy minimization. One of the first and most famous examples of this connection is given by the *Stolarsky Invariance Principle* [St], see (5.4), which connects the classical L^2 discrepancy with respect to spherical caps and the sum of Euclidean distances, i.e. the discrete energy with the kernel $K(x, y) = \|x - y\|$, showing that maximizing the latter is equivalent to minimizing the former. There has been increased interest in this principle in the recent years, including several new proofs and extensions to various settings and kernels, see, e.g., [BrD, BDM, Sk1, Ba]. A more detailed discussion can be found in the beginning of Section 5.

In Theorem 5.1, we prove a very general form of the Stolarsky Invariance Principle on an arbitrary compact space Ω which connects the (continuous or discrete) energy with a positive definite kernel K to a notion of L^2 discrepancy based on the *convolution square root* of K, whose existence is equivalent to positive definiteness of K (Proposition 2.20). This generalizes several prior versions of this principle.

1.3. **Notation and conventions.** We shall always assume that Ω is a compact metric space, although our discussion up to and including Theorem 2.14 applies also to general topological measure spaces. As defined earlier, the class of signed finite regular Borel measures on Ω will be denoted by $\mathcal{M}(\Omega)$ and, in addition to the class $\mathbb{P}(\Omega)$ of Borel probability measures, we shall consider the following subclasses of $\mathcal{M}(\Omega)$: the class $\widetilde{\mathbb{P}}(\Omega)$ of all *signed* measures of total mass one and $\mathcal{Z}(\Omega)$ – the class of all signed measures $\nu \in \mathcal{M}(\Omega)$ with mean zero, i.e. $\int_\Omega d\nu = \nu(\Omega) = 0$. We shall say that two measures are equal if they coincide on all Borel subsets of Ω; likewise, inequalities between measures will be understood to hold on the Borel subsets of Ω.

While a substantial portion of the theory applies to more general kernels, in the present text we restrict our attention just to continuous functions. Therefore, we shall say that K is a *kernel* on $\Omega \times \Omega$ if $K : \Omega \times \Omega \to \mathbb{R}$ is continuous and symmetric. This ensures that the energy $I_K(\mu)$ in (1.2) is well defined for any measure $\mu \in \mathcal{M}(\Omega)$.

While we will be interested in local and global minimizers over different sets of measures, whenever we say a measure μ is a *minimizer* of I_K without any additional information, we mean that μ is a global minimizer of I_K over $\mathbb{P}(\Omega)$, i.e. for all $\nu \in \mathbb{P}(\Omega)$, $I_K(\mu) \leq I_K(\nu)$. Unless explicitly stated otherwise, local minimizers will be understood in the directional sense, see Definition 2.15. The shorthand $\mu \llcorner_B$ will denote the restriction of μ to a set B, i.e. a measure defined by $\mu \llcorner_B(A) = \mu(A \cap B)$.

The Euclidean sphere in $\mathbb{R}^d$ will be denoted by $\mathbb{S}^{d-1}$ and σ will denote the uniform (Lebesgue) measure on $\mathbb{S}^{d-1}$ normalized so that $\sigma(\mathbb{S}^{d-1}) = 1$.

2. Positive Definite Kernels and Energy Minimization

Positive definite kernels play an extremely important role in various areas of mathematics, such as partial differential equations, machine learning, and probability theory. In this discussion, we will focus on their relation to energy minimization problems, but an exposition on their role in other areas can be found, e.g., in [A, F, Me].

We now state the relevant definition in the form, which is most convenient for applications to energy optimization over Borel measures.

Definition 2.1. *A kernel $K : \Omega^2 \to \mathbb{R}$ is called **conditionally positive definite** if for every $\nu \in \mathcal{Z}(\Omega)$ (i.e. finite signed Borel measures satisfying $\nu(\Omega) = 0$), $I_K(\nu) \geq 0$.*

*If, moreover, $I_K(\nu) \geq 0$ for every finite signed Borel measure, i.e. $\nu \in \mathcal{M}(\Omega)$, then we call K **positive definite**.*

*We call a kernel **strictly positive definite** or **conditionally strictly positive definite** if it is positive definite or conditionally positive definite, respectively, and $I_K(\nu) = 0$ only if $\nu = 0$.*

*If there exists some $C \in \mathbb{R}$ such that $K + C$ is a (strictly) positive definite kernel, we call K **(strictly) positive definite modulo an additive constant** (or **up to an additive constant**).*

A more standard way of defining positive definiteness of a kernel $K : \Omega^2 \to \mathbb{R}$ is by requiring that, for every $N \in \mathbb{N}$ and $\{z_i\}_{i=1}^N \subset \Omega$, the matrix $\left[K(z_i, z_j)\right]_{i,j=1}^N$ is positive semidefinite, i.e. for any sequence $\{c_i\}_{i=1}^N \subset \mathbb{R}$, the kernel K satisfies the inequality

$$(2.1) \qquad \sum_{i,j=1}^N c_i c_j K(z_i, z_j) \geq 0.$$

Since the kernel K is continuous, this is clearly equivalent to Definition 2.1 due to the weak* density of discrete measures in $\mathcal{M}(\Omega)$. Similarly, conditional positive definiteness is equivalent to (2.1) with the additional condition $\sum c_i = 0$. We finally remark that such an equivalence does not hold for the strict version of these properties.

A constant positive kernel, i.e. $K(x, y) = c > 0$ for all $x, y \in \Omega$, is necessarily positive definite, hence such kernels always exist. Moreover, the class of positive definite kernels is easily seen to be closed under addition, multiplication (a result known as the Schur product theorem), and limits of uniformly convergent sequences.

Lemma 2.2. *If K and L are positive definite kernels on Ω, then so are $K + L$ and KL. If $K_1, K_2, ...,$ are positive definite and $\lim_{n \to \infty} K_n = K$ uniformly, then K is positive definite. The statements regarding the sum and limit (but not the product) hold if we replace positive definiteness with conditional positive definiteness.*

For a continuous function $\phi : \Omega \to \mathbb{R}$, the kernel $\phi(x)\phi(y)$ is obviously positive definite, since

$$\int_\Omega \int_\Omega \phi(x)\phi(y) \, d\mu(x)d\mu(y) = \left(\int_\Omega \phi(x)d\mu(x)\right)^2 \geq 0.$$

Therefore, Lemma 2.2 provides a rich class of examples of positive definite kernels.

Lemma 2.3. *For $j \in \mathbb{N}_0$, let $\lambda_j \geq 0$ and $\phi_j : \Omega \to \mathbb{R}$ be continuous. Then if the series converges absolutely and uniformly, the kernel*

$$(2.2) \qquad K(x, y) = \sum_{j=0}^\infty \lambda_j \phi_j(x)\phi_j(y)$$

is positive definite.

In fact, the well-known Mercer's Theorem (see Theorem 2.18 in Section 2.5) demonstrates that the representation (2.2) actually provides a characterization of positive definite kernels, see Corollary 2.19.

In what follows, we provide various results which connect properties of the kernel K, the energy functional I_K, and the minimizers of this energy integral. While some of these results have previously appeared in the literature, a number of them seem to be new (we shall specifically point them out in the exposition below). In addition, it seems that even the known results have not all appeared simultaneously in a single text (perhaps the most complete prior exposition of this kind is the discussion of lower semi-continuous kernels on compact sets in [BHS, Chapter 4]). Moreover, the general results of this section form a basis for the long list of equivalences provided in Theorems 4.1, 4.2, and 4.3 under some additional assumptions. Therefore, for the sake of making our presentation self-contained and coherent, we have decided to include all the relevant background information (not just the new results) in this section.

In Section 2.1, we explore the relation between conditional positive definiteness and positive definiteness up to an additive constant. We show that the latter implies the former (Lemma 2.4), but not vice versa (Example 2.5). This relation will be revisited in Section 3.4, where the converse implication is established under additional assumptions (Lemma 3.7).

Section 2.2 discusses the equivalence of conditional positive definiteness and the arithmetic-mean inequality for mixed energies (Lemma 2.6), as well as the similar equivalence between positive definiteness and the geometric-mean inequality, Lemma 2.7. While the fact that (conditional) positive definiteness implies such inequalities is well-known, we did not find the converse implication in the literature.

In Section 2.3, we concentrate on the interplay between conditional positive definiteness of the kernel K and the convexity of the corresponding energy functional I_K. In particular, Proposition 2.11 demonstrates that the two notions are equivalent. Again, only one direction seems to have appeared in the literature before.

Section 2.4 reviews some basic facts about the potential of the global and local minimizers of the energy integral I_K (Theorem 2.14 and Corollary 2.16, respectively): if μ is a (local) minimizer of I_K, then U_K^μ is constant on $\mathrm{supp}(\mu)$.

In Section 2.5, we recall the connection between the positive definiteness of the kernel K and the properties of the generated Hilbert–Schmidt operator $T_{K,\mu}$. Lemma 2.17 demonstrates the equivalence between positive definiteness of K and positivity of $T_{K,\mu}$, while Mercer's Theorem (Theorem 2.18) provides the absolutely and uniformly convergent expansion of a positive definite kernel in term of the eigenfunctions of the associated Hilbert–Schmidt operator.

In Section 2.6, we demonstrate the existence of the "convolution square root" of a positive definite kernel (see (2.16) in Proposition 2.20). In the case $\Omega = \mathbb{S}^{d-1}$, this fact has been observed in [BD, BDM], but the general case presented here is new.

Finally, Section 2.7 explores the relation between energy minimizers and Hilbert–Schmidt operators. In particular, Lemma 2.21 shows that if μ minimizes I_K, then the associated operator $T_{K,\mu}$ is positive, which leads to an important fact (Lemma 2.22): if μ is a (local) minimizer of I_K, then K is positive definite (up to a constant) on the support of μ. Similar results have appeared in various papers on energy minimization, e.g. [CFP, FS].

2.1. Conditional Positive Definiteness and Positive Definiteness up to a Constant.
Since adding a constant to a kernel obviously does not affect the minimizers, it is natural to consider kernels that are (strictly) positive definite modulo a constant. However, adding a constant also never changes conditional (strict) positive definiteness, as for all $C \in \mathbb{R}$ and $\nu \in \mathcal{Z}(\Omega)$,

$$I_{K+C}(\nu) = I_K(\nu) + (\nu(\Omega))^2 C = I_K(\nu).$$

Since (strict) positive definiteness implies conditional (strict) positive definiteness, we arrive at the following lemma.

Lemma 2.4. *If K is (strictly) positive definite modulo a constant, then K is conditionally (strictly) positive definite.*

In Section 3 (Lemma 3.7), we will demonstrate that the converse of Lemma 2.4 can hold under certain conditions. However, it does not hold in general:

Example 2.5. *Consider $K : \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \to \mathbb{R}$ defined by $K(x,y) = x_1 + y_1$, where $x_1 = \langle x, e_1 \rangle$. Then K is conditionally positive definite, but not positive definite modulo a constant.*

Proof. For all $\nu \in \mathcal{Z}(\mathbb{S}^{d-1})$,

$$I_K(\nu) = 2 \int\limits_{\mathbb{S}^{d-1}} \int\limits_{\mathbb{S}^{d-1}} x_1 \, d\nu(x) d\nu(y) = 0,$$

so K is conditionally positive definite.

Now we show there is no constant C such that $K + C$ is positive definite. If $C < 0$, then

$$I_{K+C}(\sigma) = 2 \int\limits_{\mathbb{S}^{d-1}} x_1 d\sigma(x) + C = C < 0.$$

Suppose that $C \geq 0$ and let $\mu = (C+1)\delta_{-e_1} - C\delta_{e_1} \in \mathcal{M}(\mathbb{S}^{d-1})$. Then

$$I_{K+C}(\mu) = 2\mu(\mathbb{S}^{d-1}) \int\limits_{\mathbb{S}^{d-1}} x_1 \, d\mu(x) + C(\mu(\mathbb{S}^{d-1}))^2$$

$$= 2(-2C - 1) + C = -3C - 2 < 0.$$

The proof is now complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

2.2. **Positive Definiteness and Inequalities for Mixed Energies.** We first make the observation that the (conditional) positive definiteness of the kernel can be characterized by the inequalities for mixed energies in terms of arithmetic or geometric means. While the validity of such inequalities for positive definite kernels is well known, see, e.g., [BHS, Chapter 4], their sufficiency doesn't seem to have appeared in previous literature. We summarize these facts in the following two lemmas. The first one connects conditional positive definiteness to the arithmetic mean inequality.

Lemma 2.6. *Suppose K is a kernel on $\Omega \times \Omega$. Then the following conditions are equivalent:*

  (1) K is conditionally positive definite.
  (2) For every pair of Borel probability measures μ_1 and μ_2 on Ω, the mutual energy $I_K(\mu_1, \mu_2)$ satisfies

(2.3) $$I_K(\mu_1, \mu_2) \leq \frac{1}{2}\big(I_K(\mu_1) + I_K(\mu_2)\big).$$

  (3) Inequality (2.3) is satisfied for any pair of signed Borel measures of total mass one.

In addition, conditional strict positive definiteness of K is equivalent to the fact that the inequality in (2.3) is strict unless $\mu_1 = \mu_2$.

Proof. Suppose that K is conditionally positive definite. Then for any $\mu_1, \mu_2 \in \widetilde{\mathbb{P}}(\Omega)$, $\mu_1 - \mu_2 \in \mathcal{Z}(\Omega)$, so

$$0 \leq I_K(\mu_1 - \mu_2) = I_K(\mu_1) - 2I_K(\mu_1, \mu_2) + I_K(\mu_2),$$

which proves (2.3). Thus, (1) implies (3), which in its turn obviously implies (2).

Now assume condition (2), i.e. that (2.3) holds for all $\mu_1, \mu_2 \in \mathbb{P}(\Omega)$. For any $\mu \in \mathcal{Z}(\Omega)$, there exists $c \geq 0$ and probability measures $\mu_1, \mu_2 \in \mathbb{P}(\Omega)$ such that $\mu = c(\mu_1 - \mu_2)$. We then have that

$$(2.4) \qquad I_K(\mu) = I_K\big(c(\mu_1 - \mu_2)\big) = c^2\Big(I_K(\mu_1) - 2I_K(\mu_1, \mu_2) + I_K(\mu_2)\Big) \geq 0,$$

so K must be conditionally positive definite.

If K is conditionally strictly positive definite and the measure $\mu_1 - \mu_2$ is nonzero, then $I_K(\mu_1 - \mu_2) > 0$, making inequality (2.3) strict. Conversely, if (2.3) is strict whenever $\mu_1 - \mu_2 \neq 0$, then equality in (2.4) can hold only if $\mu_1 = \mu_2$ or $c = 0$, i.e. K is conditionally strictly positive definite. $\qquad\square$

The second lemma is very similar: it shows that positive definiteness is equivalent to the geometric mean inequality for the mixed energy.

Lemma 2.7. *Suppose K is a kernel on $\Omega \times \Omega$. Then K is positive definite if and only if $I_K(\mathbb{P}(\Omega)) \subset [0, \infty)$ and for all $\mu_1, \mu_2 \in \mathbb{P}(\Omega)$, the mutual energy $I_K(\mu_1, \mu_2)$ satisfies*

$$(2.5) \qquad I_K(\mu_1, \mu_2) \leq \sqrt{I_K(\mu_1) I_K(\mu_2)},$$

K is strictly positive definite if and only if the inequality in (2.5) is strict unless $\mu_1 = \mu_2$.

Observe that the proof below shows that, just like in Lemma 2.6, we could replace $\mathbb{P}(\Omega)$ with $\widetilde{\mathbb{P}}(\Omega)$. Heuristically, this statement says that the bilinear form $I_K(\mu_1, \mu_2)$ defines an inner product on measures if and only if K is positive definite, and (2.5) is just the Cauchy–Schwarz inequality.

Proof. Suppose that K is positive definite. For any $t \in \mathbb{R}$ and $\mu_1, \mu_2 \in \mathbb{P}(\Omega)$, we define

$$g(t) := t^2 I_K(\mu_1) - 2t I_K(\mu_1, \mu_2) + I_K(\mu_2) = I_K(t\mu_1 - \mu_2) \geq 0.$$

Thus the discriminant $4I_K(\mu_1, \mu_2)^2 - 4I_K(\mu_1) I_K(\mu_2)$ of the quadratic polynomial $g(t)$ is nonpositive, which yields (2.5). If K is strictly positive definite, $g(t) = 0$ has a root only if $t\mu_1 - \mu_2 = 0$, which implies $t = 1$ and $\mu_1 = \mu_2$.

Suppose instead that (2.5) holds for all probability measures. For any $\mu \in \mathcal{M}(\Omega)$, there exists $a, b \geq 0$ and $\mu_1, \mu_2 \in \mathbb{P}(\Omega)$ such that $\mu = a\mu_1 - b\mu_2$. We then have that

$$I_K(\mu) = a^2 I_K(\mu_1) - 2ab I_K(\mu_1, \mu_2) + b^2 I_K(\mu_2) \geq (a\sqrt{I_K(\mu_1)} - b\sqrt{I_K(\mu_2)})^2 \geq 0,$$

implying that K is positive definite. If (2.5) is strict unless $\mu_1 = \mu_2$, then the inequality above is strict unless $\mu = 0$, i.e. K is strictly positive definite. $\qquad\square$

2.3. **Positive Definiteness and Convexity of the Energy Functional.** Convexity plays an important role in optimization problems, and so do various versions of positive definiteness. Therefore, it is not surprising that the two notions are related. In fact, as we shall see, in many settings, they are equivalent.

Such equivalences, in different forms, have previously appeared in the literature [BFGMPV, CSh, DPZ, Mec, P, PZ, ZDP]. This connections appears to be common knowledge in some references, but is largely overlooked in many other. In this subsection, we take a deeper look at this phenomenon.

Definition 2.8. *Let $K : \Omega \times \Omega \to \mathbb{R}$ be a kernel. We say that I_K is **convex at** $\mu \in \mathbb{P}(\Omega)$ if for every $\nu \in \mathbb{P}(\Omega)$ there exists some $t_\nu \in (0, 1]$ such that for all $t \in [0, t_\nu)$*

$$(2.6) \qquad I_K((1-t)\mu + t\nu) \leq (1-t)I_K(\mu) + tI_K(\nu).$$

*We say I_K is **convex** on $\mathbb{P}(\Omega)$ if inequality (2.6) holds for every $\mu, \nu \in \mathbb{P}(\Omega)$ and all $t \in [0, 1]$; it is said to be **strictly convex**, if the inequality is strict for all $t \in (0, 1)$ unless $\mu = \nu$.*

Similarly to the above definition, we consider (strict) convexity on $\widetilde{\mathbb{P}}(\Omega)$ and $\mathcal{M}(\Omega)$. Unless noted otherwise, convexity is understood on $\mathbb{P}(\Omega)$.

We observe that convexity of I_K on $\mathbb{P}(\Omega)$ is equivalent to the fact that I_K is convex at all $\mu \in \mathbb{P}(\Omega)$. Indeed, if (2.6) fails for some $\mu, \nu \in \mathbb{P}(\Omega)$ and some $t \in (0, 1)$, then the quadratic polynomial $f(t) = I_K((1-t)\mu + t\nu)$ is not convex on the interval $[0, 1]$, i.e. $f''(t) < 0$ for all $t \in [0, 1]$ and I_K fails to be convex at μ.

We first show that convexity is equivalent to the arithmetic mean inequality (2.3) for mixed energies.

Lemma 2.9. *The energy functional I_K is convex at $\mu \in \mathbb{P}(\Omega)$ if and only if for all $\nu \in \mathbb{P}(\Omega)$,*

$$(2.7) \qquad I_K(\mu, \nu) \leq \frac{1}{2}\big(I_K(\mu) + I_K(\nu)\big).$$

Consequently, I_K is convex on $\mathbb{P}(\Omega)$ if and only if inequality (2.7) holds for all $\mu, \nu \in \mathbb{P}(\Omega)$.

In addition, strict convexity of I_K is equivalent to the above inequality being strict unless $\mu = \nu$.

Proof. Let $\nu \in \mathbb{P}(\Omega)$ and assume that the arithmetic-mean inequality (2.7) holds. Then for all $t \in [0, 1]$,

$$(2.8) \ \ I_K((1-t)\mu + t\nu) = (1-t)^2 I_K(\mu) + 2(1-t)t I_K(\mu, \nu) + t^2 I_K(\nu) \leq (1-t)I_K(\mu) + t I_K(\nu).$$

So I_K is indeed convex at μ.

For the converse direction, assume that I_K is convex at μ. Then for any $\nu \in \mathbb{P}(\Omega)$ and $t > 0$ sufficiently small, inequality (2.8) holds, so

$$2(1-t)t I_K(\mu, \nu) \leq t(1-t)(I_K(\mu) + I_K(\nu)).$$

Dividing by $t(1-t)$, we obtain the arithmetic mean inequality (2.7). Lastly, for strictly convex I_K, the proof is the same, with all inequalities strict unless $\mu = \nu$. $\square$

Observe that one can easily replace $\mathbb{P}(\Omega)$ with $\widetilde{\mathbb{P}}(\Omega)$ in Lemma 2.9. At the same time, according to parts (2)-(3) of Lemma 2.6, the validity of (2.7) on $\mathbb{P}(\Omega)$ is equivalent to its validity on $\widetilde{\mathbb{P}}(\Omega)$. Hence we obtain the following corollary.

Corollary 2.10. *I_K is (strictly) convex on $\mathbb{P}(\Omega)$ if and only if it is (strictly) convex on $\widetilde{\mathbb{P}}(\Omega)$.*

Lemmas 2.6 and 2.9 together clearly imply the desired equivalence between convexity and conditional positive definiteness:

Proposition 2.11. *Let $K : \Omega \times \Omega \to \mathbb{R}$ be a kernel. The kernel K is conditionally (strictly) positive definite if and only if the energy functional I_K is (strictly) convex on $\mathbb{P}(\Omega)$ (or, equivalently, on $\widetilde{\mathbb{P}}(\Omega)$).*

This equivalence between convexity of I_K on $\mathbb{P}(\Omega)$ and conditional positive definiteness of K has been observed, e.g. in [BFGMPV, CSh, P, PZ]. In fact, it admits a short direct proof (see Proposition 2.13 and the discussion thereafter). We chose to prove Lemma 2.9 first (an approach taken in a recent paper of the authors, joint with A. Glazyrin, D. Ferizović, and J. Park [BFGMPV]), since it also allows us to establish equivalences between local versions of properties, which will be important later, see e.g. parts (7) and (9) of Theorem 4.1.

Corollary 2.12. *Suppose that K is conditionally strictly positive definite. Then I_K has a unique minimizer (either in $\mathbb{P}(\Omega)$ or in $\widetilde{\mathbb{P}}(\Omega)$).*

Proof. According to Lemmas 2.6 and 2.9, strict conditional positive definiteness of K implies strict convexity of I_K on both $\mathbb{P}(\Omega)$ and $\widetilde{\mathbb{P}}(\Omega)$. To conclude, observe that these measure spaces are themselves convex. $\qquad\square$

Finally, we note that convexity of I_K on the entire space $\mathcal{M}(\Omega)$ is equivalent to positive definiteness of the kernel. More precisely, we have the following.

Proposition 2.13 ([P, Lemma 4]). *The functional I_K is (strictly) convex on $\mathcal{M}(\Omega)$ if and only if K is (strictly) positive definite on Ω.*

We include a simple proof of this proposition for the sake of completeness.

Proof. Simply observe that for any $t \in (0,1)$, the inequality

$$I_K((1-t)\mu + t\nu) = (1-t)^2 I_K(\mu) + 2t(1-t)I_K(\mu,\nu) + t^2 I_K(\nu) \le (1-t)I_K(\mu) + tI_K(\nu)$$

is equivalent to

$$t(1-t)\big(I_K(\mu) - 2I_K(\mu,\nu) + I_K(\nu)\big) = t(1-t)I_K(\mu - \nu) \ge 0,$$

from which the statement easily follows. $\qquad\square$

We note that this argument also gives an alternative proof of Proposition 2.11, since any measure with mass zero can be represented as a multiple of the difference of two measures with mass one.

2.4. **Minimizing Measures: Basic Potential Theory.** It is well known that the behavior of the minimizing measures is closely connected to the behavior of the potential of the minimizing measure with respect to the kernel. For a detailed account of the topic, we refer the reader to Chapter 4 of [BHS]. The following simple statement is classical, see, e.g., [Bj]. We provide its proof for completeness.

Theorem 2.14. *Suppose that μ is a minimizer of I_K over $\mathbb{P}(\Omega)$. Then $U_K^\mu(x) = I_K(\mu)$ on $\mathrm{supp}(\mu)$ and $U_K^\mu(x) \ge I_K(\mu)$ on Ω.*

Proof. Let $\nu \in \mathcal{Z}(\Omega)$ be such that $\mu + \varepsilon\nu \ge 0$ for all $0 < \varepsilon \le \varepsilon_0$ with some positive ε_0. This clearly means that $\mu + \varepsilon\nu \in \mathbb{P}(\Omega)$, so

$$(2.9) \qquad I_K(\mu) \le I_K(\mu + \varepsilon\nu) = I_K(\mu) + 2\varepsilon I_K(\mu,\nu) + \varepsilon^2 I_K(\nu).$$

Thus, for $0 \le \varepsilon \le \varepsilon_0$,

$$0 \le \varepsilon\left(2I_K(\mu,\nu) + \varepsilon I_K(\nu)\right),$$

implying $I_K(\mu,\nu) \ge 0$.

Suppose that there exist $c_1, c_2 \in \mathbb{R}$, $z \in \mathrm{supp}(\mu)$ and $y \in \Omega$ such that

$$c_1 = U_K^\mu(y) < U_K^\mu(z) = c_2.$$

Let B_z be a ball centered at z, so small that $\min_{x \in B_z} U_K^\mu(x) > (c_1 + c_2)/2$. Define

$$(2.10) \qquad \nu := \mu(B_z) \cdot \delta_y - \mu \llcorner B_z.$$

Then $\nu \in \mathcal{Z}(\Omega)$ and satisfies $\mu + \varepsilon\nu \ge 0$ for any $0 \le \varepsilon \le 1$. On the other hand,

$$I_K(\mu,\nu) = \mu(B_z) \cdot U_K^\mu(y) - \int_{B_z} U_K^\mu(x)\,d\mu(x) \le \mu(B_z) \cdot c_1 - \mu(B_z) \cdot \frac{c_1 + c_2}{2} < 0,$$

which is a contradiction. Thus, necessarily $U_K^\mu(y) \ge U_K^\mu(z)$ for y, z as above. Since y can belong to $\mathrm{supp}(\mu)$ and $\int_\Omega U_K^\mu(x)\,d\mu(x) = I_K(\mu)$, both claims of the theorem follow. $\qquad\square$

Notice that if μ has full support, i.e. $\text{supp}(\mu) = \Omega$, the conclusion of Theorem 2.14 states that the potential $U_K^\mu(x)$ is constant on Ω. Measures with constant potentials will be further explored in Section 3.

Definition 2.15. *We shall say that μ is a **local minimizer** of I_K in $\mathbb{P}(\Omega)$ **with respect to a given metric** $d(\cdot, \cdot)$ on $\mathcal{M}(\Omega)$ if it is a local minimizer in the topology induced by this metric, in other words, if there exists $\varepsilon_\mu > 0$, such that for all $\nu \in \mathbb{P}(\Omega)$ satisfying $d(\mu, \nu) \leq \varepsilon_\mu$, we have*

$$I_K(\mu + \nu) \geq I_K(\mu).$$

*We shall say that μ is a **directional local minimizer** of I_K in $\mathbb{P}(\Omega)$ if it is a local minimizer in every direction, i.e., if for each $\nu \in \mathbb{P}(\Omega)$, there exists $\tau_\nu \in (0, 1]$ such that for all $t \in [0, \tau_\nu]$ we have*

$$I_K\big((1 - t)\mu + t\nu\big) \geq I_K(\mu).$$

The difference between the two definitions above is similar to that between the Gateaux and Fréchet derivatives.

Observe that, if Ω is a compact metric space, convergence in total variation implies convergence in Wasserstein W_p metric for $1 \leq p < \infty$. Thus, a local minimizer in one of W_p is also a local minimizer with respect to the total variation distance.

In turn, for $\mu, \nu \in \mathbb{P}(\Omega)$, we have $\big((1 - t)\mu + t\nu\big) - \mu = t(\nu - \mu)$, and the total variation norm satisfies $\|t(\nu - \mu)\|_{TV} \leq 2t < \varepsilon_\mu$ for t small enough. Therefore, local minimizers in total variation are directional local minimizers (but not vice versa). This is summarized below:

$$(2.11) \qquad \begin{pmatrix} \text{local minimizer in} \\ W_p, \ 1 \leq p < \infty \end{pmatrix} \implies \begin{pmatrix} \text{local minimizer in} \\ \text{total variation} \end{pmatrix} \implies \begin{pmatrix} \text{directional} \\ \text{local minimizer} \end{pmatrix}.$$

In this text, unless explicitly specified otherwise, the words "local minimizer" in the assumption of a statement will mean the directional minimizer, as this is the weakest assumption, and thus corresponding results will also hold for the other types of local minimizers mentioned in (2.11).

Analyzing the proof of Theorem 2.14, we find that for ν defined in (2.10), we can write $\mu + \varepsilon\nu = (1 - \varepsilon)\mu + \varepsilon\widetilde{\nu}$ with $\widetilde{\nu} = \mu + \nu \in \mathbb{P}(\Omega)$. Hence, inequality (2.9) holds with sufficiently small ε even if μ is just a directional local minimizer (and therefore, also if it is a local minimizer in total variation or in the Wasserstein distance W_p, $1 \leq p < \infty$), and one arrives at a contradiction in the same way.

Corollary 2.16. *The statement of Theorem 2.14 remains true if we only assume that μ is a local (not global) minimizer of I_K (in either of three senses: directional, in total variation, or in Wasserstein W_p metric, $1 \leq p < \infty$).*

As we shall see in Theorem 3.11, under some additional conditions, in particular, if K is conditionally positive definite, the statement of Theorem 2.14 can be reversed.

2.5. **Positive Definiteness and Hilbert–Schmidt Operators.** Let μ be a Borel probability measure on Ω and let K be a continuous function on $\Omega \times \Omega$. We shall consider the operator $T_{K,\mu}$ associated to K on the space of real-valued functions on Ω that are square-integrable with respect to μ, $L^2(\Omega, \mu)$. This is a linear integral operator with kernel K defined by

$$(2.12) \qquad T_{K,\mu}\psi(x) = \int_\Omega K(x, y)\psi(y)\, d\mu(y).$$

Lemma 2.17. *Let $\widetilde{\Omega} = \text{supp}(\mu)$. The operator $T_{K,\mu}$ is self-adjoint and Hilbert–Schmidt, and the eigenfunctions of $T_{K,\mu}$ corresponding to non-zero eigenvalues are continuous on $\widetilde{\Omega}$. The kernel K is positive definite on $\widetilde{\Omega}$ if and only if $T_{K,\mu}$ is a positive operator on $L^2(\Omega, \mu)$.*

Proof. Self-adjointedness immediately follows from the fact that $K(x, y)$ is symmetric. Since Ω is compact (and hence $\widetilde{\Omega}$ is also) and K continuous, we know that

$$\int\limits_{\Omega} \int\limits_{\Omega} |K(x, y)|^2 \, d\mu(x) d\mu(y) < \infty,$$

which implies that $T_{K,\mu}$ is Hilbert–Schmidt.

Now, suppose that $T_{K,\mu}\phi = \lambda\phi$ for $\lambda \neq 0$. Then the representation

$$\phi(x) = \frac{1}{\lambda} \int\limits_{\Omega} K(x, y)\phi(y) \, d\mu(y)$$

implies that ϕ is continuous on $\widetilde{\Omega}$.

We now show that the positive definiteness of K and the positivity of $T_{K,\mu}$ are equivalent. If K is positive definite on $\widetilde{\Omega}$, then for any $\psi \in L^2(\Omega, \mu) \subset L^1(\Omega, \mu)$,

$$\langle \psi, T_{K,\mu}\psi \rangle_{L^2(\Omega,\mu)} = \int\limits_{\Omega} \int\limits_{\Omega} K(x, y)\psi(x)\psi(y) \, d\mu(x)d\mu(y) = I_K(\psi(\cdot)\mu) \geq 0,$$

so $T_{K,\mu}$ is indeed positive.

Assume instead that $T_{K,\mu}$ is positive. Observe that measures, which are absolutely continuous with respect to μ and have bounded density, i.e. measures of the form $d\nu = f \, d\mu$, where f is a bounded Borel measurable function on $\widetilde{\Omega}$, are weak* dense in $\mathcal{M}(\widetilde{\Omega})$. To show this, notice that for each ball $B(z, r)$ of radius $r > 0$ centered at the point $z \in \widetilde{\Omega}$, we have $\mu(B(z, r)) \neq 0$, and therefore, the functions $f_r(x) = \frac{1}{\mu(B(z,r))} \mathbb{1}_{B(z,r)}(x)$ are well-defined and bounded. Obviously, the measures ν_r defined by $d\nu_r = f_r \, d\mu$ converge weak* to δ_z as $r \to 0$, which suffices due to weak* density of discrete measures.

Then for all such measures of the form $d\nu = f \, d\mu$, since bounded functions are in $L^2(\Omega, \mu)$, we have

$$I_K(\nu) = \langle T_{K,\mu}f, f \rangle \geq 0,$$

and by weak* density, it follows that K is positive definite on $\widetilde{\Omega}$. $\qquad\square$

Moreover, since $T_{K,\mu}$ is a Hilbert-Schmidt operator, it is in fact a compact operator. Hence, we may apply the Spectral Theorem to establish that there exists an orthonormal basis $\{\phi_j\}_{j=1}^{\dim(L^2(\Omega,\mu))}$ of $L^2(\Omega, \mu)$ consisting of eigenfunctions of $T_{K,\mu}$, i.e. $T_{K,\mu}\phi_j = \lambda_j\phi_j$, where the sequence of eigenvalues satisfies $|\lambda_j| \geq |\lambda_{j+1}|$ and $\lim_{j \to \infty} \lambda_j = 0$ if $\dim(L^2(\Omega, \mu)) = \infty$. Moreover, the Spectral Theorem tells us that, for any continuous function K, in the L^2 sense,

$$(2.13) \qquad K(x, y) = \sum_{j=1}^{\dim(L^2(\Omega,\mu))} \lambda_j\phi_j(x)\phi_j(y).$$

When K is positive definite, $T_{K,\mu}$ is positive, i.e. $\lambda_j \geq 0$ for all $j \geq 1$. Mercer's Theorem then strengthens the information about convergence in (2.13). We include its proof for completeness.

Theorem 2.18 (Mercer's Theorem)**.** *Assume that the kernel K on $\Omega \times \Omega$ is positive definite. Fix a measure $\mu \in \mathbb{P}(\Omega)$ with $\operatorname{supp}(\mu) = \Omega$. Let $\lambda_j \geq 0$ be the eigenvalues and ϕ_j be the eigenfunctions of the associated Hilbert–Schmidt operator $T_{K,\mu}$. Then*

$$(2.14) \qquad K(x, y) = \sum_{j=1}^{\dim(L^2(\Omega,\mu))} \lambda_j\phi_j(x)\phi_j(y),$$

where the series converges absolutely and uniformly.

Remark: Here, as well as in Proposition 2.20, we assume that $\operatorname{supp}(\mu) = \Omega$. If $\operatorname{supp}(\mu) \neq \Omega$, then the expansion (2.14) holds for $x, y \in \widetilde{\Omega} = \operatorname{supp}(\mu)$.

Proof. As mentioned above, the fact that (2.14) holds in the L^2 sense follows from the Spectral Theorem for compact operators, hence, only uniform and absolute convergence in the case $\dim(L^2(\Omega, \mu)) = \infty$ need to be proven.

Consider the remainder of the series, i.e. the continuous function

$$R_N(x, y) = K(x, y) - \sum_{j=1}^{N} \lambda_j \phi_j(x) \phi_j(y).$$

The corresponding Hilbert-Schmidt operator defined by $T_{R_N, \mu} \psi(x) = \int_{\Omega} R_N(x, y) \, \psi(y) \, d\mu(y)$ is clearly bounded and positive: if $\psi = \sum_{j=1}^{\infty} \widehat{\psi}_j \phi_j$, then

$$\langle T_{R_N, \mu} \psi, \psi \rangle = \sum_{j=N+1}^{\infty} \lambda_j |\widehat{\psi}_j|^2 \geq 0.$$

For a positive Hilbert–Schmidt operator, the kernel is non-negative on the diagonal, i.e. $R_N(x, x) \geq 0$ for each $x \in \Omega$. Indeed, assume that for some $x \in \Omega$ we have $R_N(x, x) < 0$. Then we can choose a neighborhood U of x so that R_N is negative on $U \times U$, so

$$\langle T_{R_N, \mu} \mathbb{1}_U, \mathbb{1}_U \rangle = \int_U \int_U R_N(x, y) \, d\mu(x) d\mu(y) < 0,$$

which is a contradiction.

Thus, $R_N(x, x) \geq 0$ for each $x \in \Omega$, i.e. for any $N \geq 1$ we have $\sum_{j=1}^{N} \lambda_j \phi_j^2(x) \leq K(x, x)$, and thus

$$\sum_{j=1}^{\infty} \lambda_j \phi_j^2(x) \leq K(x, x).$$

Invoking the Cauchy–Schwarz inequality, we find that

$$\left| \sum_{j=1}^{\infty} \lambda_j \phi_j(x) \phi_j(y) \right| \leq \sum_{j=1}^{\infty} \lambda_j |\phi_j(x) \phi_j(y)| \leq \left(\sum_{j=1}^{\infty} \lambda_j \phi_j^2(x) \right)^{1/2} \left(\sum_{j=1}^{\infty} \lambda_j \phi_j^2(y) \right)^{1/2}$$

$$\leq K^{1/2}(x, x) K^{1/2}(y, y) \leq \sup_{x \in \Omega} K(x, x) < \infty.$$

Therefore, the series in (2.14) converges absolutely and uniformly, by Dini's theorem. □

As a simple corollary of Mercer's Theorem, we find that

$$\int_{\Omega} K(x, x) \, d\mu(x) = \sum_{j \geq 1} \lambda_j \int \phi_j^2 \, d\mu = \sum_{j \geq 1} \lambda_j,$$

and thus

(2.15)
$$\sum_{j=1}^{\dim(L^2(\Omega, \mu))} \lambda_j < \infty.$$

By combining Mercer's Theorem with Lemma 2.3 we arrive at a characterization of all positive definite kernels.

Corollary 2.19. *The kernel K on $\Omega \times \Omega$ is positive definite if and only if for some sequence of functions $\phi_j : \Omega \to \mathbb{R}$ and real numbers $\lambda_j \geq 0$, the kernel K*

$$K(x, y) = \sum_{j \geq 1} \lambda_j \phi_j(x) \phi_j(y)$$

where the sum converges absolutely and uniformly.

2.6. **The Existence of Convolution Square Root.** We now supply another way to characterize positive definite functions, which is similar to the existence of the "square root" for positive semidefinite matrices. A similar statement for the rotation-invariant kernels on the sphere has been obtained in [BD, BDM]. Here we prove a general version of this representation.

Proposition 2.20. *Fix any $\mu \in \mathbb{P}(\Omega)$ with $\operatorname{supp}(\mu) = \Omega$. A kernel K on $\Omega \times \Omega$ is positive definite if and only if there exists some $k \in L^2(\Omega \times \Omega, \mu \times \mu)$ such that for all $x, y \in \Omega$,*

$$(2.16) \qquad K(x, y) = \int_\Omega k(x, z) k(z, y) \, d\mu(z).$$

We again remind the reader that if $\widetilde{\Omega} = \operatorname{supp}(\mu) \subsetneq \Omega$, then the conclusion of Proposition 2.20 holds on $\widetilde{\Omega}$.

Proof. Assume that K is positive definite and define the function $k : \Omega \times \Omega \to \mathbb{R}$ by setting

$$(2.17) \qquad k(x, y) = \sum_{j \geq 1} \sqrt{\lambda_j} \, \phi_j(x) \phi_j(y),$$

where $\lambda_j \geq 0$ are eigenvalues and ϕ_j are eigenfunctions of $T_{K, \mu}$. The sequence $\{\phi_j(x)\phi_j(y)\}$ is orthonormal in $L^2(\Omega \times \Omega, \mu \times \mu)$, hence $k \in L^2(\Omega \times \Omega, \mu \times \mu)$, since

$$\|k\|^2_{L^2(\Omega \times \Omega, \mu \times \mu)} = \sum_{j \geq 1} \lambda_j < \infty$$

according to (2.15).

Moreover, for each $x \in \Omega$, we have that $k(x, y) \in L^2(\Omega, \mu)$ as a function of y. Indeed, using Mercer's theorem and applying Plancherel's theorem with respect to $d\mu(y)$ to (2.17), we can compute the L^2 norm

$$\|k(x, \cdot)\|^2_{L^2(\Omega, \mu)} = \sum_{j \geq 1} \left| \sqrt{\lambda_j} \phi_j(x) \right|^2 = \sum_{j \geq 1} \lambda_j \phi_j^2(x) = K(x, x) < \infty.$$

Similarly, since $k \in L^2(\Omega, \mu)$ in each variable, we have, for any $x, y \in \Omega$,

$$\int_\Omega k(x, z) k(z, y) \, d\mu(z) = \langle k(x, \cdot), k(y, \cdot) \rangle_{L^2(\Omega, \mu)} = \sum_{j \geq 1} \sqrt{\lambda_j} \phi_j(x) \sqrt{\lambda_j} \phi_j(y) = K(x, y),$$

due to Mercer's Theorem, which proves (2.16).

Alternatively, if we assume that (2.16) holds, then for any finite point configuration $\omega_N = \{z_1, ..., z_N\}$ in Ω and $c_1, ..., c_N \in \mathbb{R}$, we have

$$\sum_{j=1}^{N} \sum_{i=1}^{N} K(z_j, z_i) c_i c_j = \sum_{j=1}^{N} \sum_{i=1}^{N} c_j c_i \int_{\Omega} k(z_j, z) k(z, z_i) \, d\mu(z)$$

$$= \int_{\Omega} \left(\sum_{j=1}^{N} c_j k(z_j, z) \right)^2 d\mu(z) \geq 0.$$

Thus, K is clearly positive definite. $\qquad\square$

Remark: Observe that the choice of the "square root" k is not unique. Indeed, instead of (2.17) one could take any

$$k(x, y) = \sum_{j \geq 1} \kappa_j \phi_j(x) \phi_j(y),$$

with the property that $\kappa_j^2 = \lambda_j$ for all $j \geq 1$, i.e. $\kappa_j = \pm\sqrt{\lambda_j}$ for an arbitrary choice of signs, yielding uncountably many functions k satisfying (2.16).

2.7. **Energy Minimizers and Hilbert–Schmidt Operators.** There is a close relation between energy minimizers and the properties of the associated Hilbert–Schmidt operator $T_{K,\mu}$ on $L^2(\Omega, \mu)$. We have the following statement.

Lemma 2.21. *Let K be a kernel on $\Omega \times \Omega$ and assume that $\mu \in \mathbb{P}(\Omega)$ is a global or local minimizer of I_K with $I_K(\mu) \geq 0$. Then the Hilbert–Schmidt operator $T_{K,\mu}$ is positive.*

Proof. We start by observing that if μ is a (global or local) minimizer of I_K, then the constant function $\mathbb{1}_\Omega$ is an eigenfunction of $T_{K,\mu}$ in $L^2(\Omega, \mu)$. Indeed, according to Theorem 2.14 or Corollary 2.16, for each $x \in \text{supp}(\mu)$,

$$(2.18) \qquad T_{K,\mu} \mathbb{1}_\Omega(x) = \int_{\Omega} K(x, y) \, d\mu(y) = U_K^\mu(x) = I_K(\mu) \mathbb{1}_\Omega(x).$$

Assume, indirectly, that $T_{K,\mu}$ is not positive. By Lemma 2.17, $T_{K,\mu}$ is compact and self-adjoint, so there exists an eigenfunction ϕ such that $T_{K,\mu}\phi = \lambda\phi$ with $\lambda < 0$. Since ϕ is continuous, and therefore bounded, on $\text{supp}(\mu)$, we have that for sufficiently small $t > 0$, the measure

$$\mu_t = (1 + t\phi)\mu$$

is positive. As we noted above, $\mathbb{1}_\Omega$ is an eigenfunction of $T_{K,\mu}$ corresponding to the eigenvalue $I_K(\mu) \geq 0$. Clearly, then, $\mathbb{1}_\Omega$ and ϕ are orthogonal, so

$$\mu_t(\Omega) = \int_{\Omega} (1 + t\phi(x)) \, d\mu(x) = \mu(\Omega) = 1,$$

and

$$I_K(\mu_t) = \int_{\Omega} \int_{\Omega} K(x, y)(1 + t\phi(x))(1 + t\phi(y)) \, d\mu(x)d\mu(y)$$

$$= I_K(\mu) + \lambda t^2 \int_{\Omega} |\phi(x)|^2 d\mu(x) < I_K(\mu),$$

which contradicts the (local) minimality of μ over probability measures. $\qquad\square$

Recall that according to Lemma 2.17, the operator $T_{K,\mu}$ is positive if and only if K is positive definite on the support of μ. If the condition $I_K(\mu) \geq 0$ in Lemma 2.21 is not satisfied, we can replace K by $K'(x,y) = K(x,y) - I_K(\mu)$, which does not affect energy minimizers. Then $I_{K'}(\mu) = 0$, and hence, Lemma 2.21 applies. Therefore $T_{K',\mu}$ is positive, i.e. K' is positive definite on $\widetilde{\Omega} = \mathrm{supp}(\mu)$. In other words, K is positive definite up to an additive constant, as a kernel on $\widetilde{\Omega} \times \widetilde{\Omega}$. We arrive at the following important fact.

Lemma 2.22. *Let K be a kernel on $\Omega \times \Omega$. Suppose that μ is a local minimizer of I_K over $\mathbb{P}(\Omega)$. Then the kernel K must be positive definite modulo a constant on $\mathrm{supp}(\mu)$, i.e. as a kernel on $\mathrm{supp}(\mu) \times \mathrm{supp}(\mu)$. If $I_K(\mu) \geq 0$, then K is positive definite on $\mathrm{supp}(\mu)$.*

Various statements of this type are known in the literature [CFP, FS]. Lemma 2.22 clearly implies the following localization statement:

Corollary 2.23. *Assume the the kernel K on $\Omega \times \Omega$ is not positive definite up to an additive constant. Then any (local or global) minimizer μ of I_K must be supported on a proper subset of Ω, i.e. $\mathrm{supp}(\mu) \subsetneq \Omega$.*

3. Invariant Measures

As suggested in Section 2.4, measures with constant potentials are particularly interesting from the point of view of energy minimization. They also naturally arise in metric geometry, in connection with the so-called "rendezvous numbers" [CMY], and we draw the term "invariant" from this literature. These applications and various interesting properties warrant a separate discussion of such measures.

3.1. **Definition, Examples, and Comments.** We start with the following definition.

Definition 3.1. *Let K be a kernel on $\Omega \times \Omega$. We say that a measure $\mu \in \mathbb{P}(\Omega)$ is K-**invariant** on Ω if the potential of this measure with respect to K is constant on Ω, i.e.*

$$(3.1) \qquad U_K^\mu(x) = I_K(\mu) \quad \text{for every} \ \ x \in \Omega.$$

We shall see shortly that these measures have an array of remarkable properties. Notice that the definition does not require that μ has full support: the majority of statements in this section hold even in the absence of this assumption.

Before proceeding to these properties we shall provide some examples, showing that K-invariance is a rather rich notion. Observe that most of the examples below have full support. All statements that we shall prove for K-invariant measures will apply, in particular, to these natural examples.

- If μ (locally) minimizes I_K and has full support, then according to Theorem 2.14 and Corollary 2.16 , the measure μ is K-invariant.
- Let $\Omega = \mathbb{S}^{d-1}$ and assume that K is rotationally invariant, i.e. $K(x,y) = F(\langle x,y\rangle)$. Then the normalized uniform surface measure is K-invariant, since the potential $U_F^\sigma(x) = \int_{\mathbb{S}^{d-1}} F(\langle x,y\rangle)\,d\sigma(y)$ is obviously independent of $x \in \mathbb{S}^{d-1}$.
- If, moreover, the function F from the previous example is a polynomial of degree M, and $\omega_N = \{z_1, ..., z_N\} \subset \mathbb{S}^{d-1}$ is a spherical M-design, then $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{z_i}$ is also K-invariant.
- Similarly, assume that Ω is a compact two-point homogeneous space, K is invariant with respect to the group of isometries, and η is the normalized uniform measure on Ω. Then η is K-invariant. This equally applies to connected (e.g., projective spaces) and discrete (e.g., Hamming cube) two-point homogeneous spaces.

- If Ω is a compact topological group, μ is its normalized Haar measure, and K is invariant with respect to the group operation, i.e. $K(x,y) = F(y^{-1}x)$, then μ is K-invariant.
- A pair K, μ as in the above example can be constructed as follows. Suppose G is a compact group acting on Ω transitively, ν is the Haar measure on G, and μ – a G-invariant measure on Ω. Then for any function $f \in L^2(\Omega, \mu)$, the kernel

$$K(x,y) = \int_G f(\tau x) f(\tau y) \, d\nu(\tau)$$

  is positive definite and μ is K-invariant. Indeed, K is an average of positive definite functions $f(\tau x) f(\tau y)$; also, U_K^μ is a G-invariant function on Ω and, since G acts transitively, must be constant there (see e.g. [CSFSV] for further properties of such K).
- Let $\Omega = [-1,1]$ and $K(x,y) = |x - y|$. Then the measure $\mu = \frac{1}{2}(\delta_0 + \delta_1)$ is K-invariant. Notice that this example (as well as spherical designs) provides an invariant measure which does not have full support.
- More generally, if $\Omega = \mathbb{B}^d$ is the closed unit ball in $\mathbb{R}^d$ and $K(x,y) = \|x - y\|^{-(d-2)}$ (or $K(x,y) = -\log\|x - y\|$ for $d = 2$) is the Newtonian potential, then the uniform surface measure σ on $\mathbb{S}^{d-1}$ is the equilibrium measure for I_K and is K-invariant on $\mathbb{B}^d$ [BHS, L], which is known as the Faraday cage effect. For $d = 1$, one gets exactly the previous example, while for $d \geq 2$, the kernels are discontinuous (and thus are partially beyond the scope of our discussion).
- Still more general form of the Faraday's effect applies to $\Omega \subset \mathbb{R}^3$ that is given by a finite union of disjoint closed domains with smooth boundaries, for $K(x,y) = \|x - y\|^{-s}$, $1 \leq s < 3$: there exists a probability measure with potential constant on Ω [Fr, Section 17].

Despite an abundance of examples, the existence of a K-invariant measure is possible only under certain restrictions on the geometry of the domain Ω and the structure of the kernel K. For example, the following statement is true [CMY].

Lemma 3.2. *Assume that $\mathbb{R}^d$ is endowed with a strictly convex norm $|\cdot|$, i.e. $|x+y| < |x| + |y|$ unless x and $y \in \mathbb{R}^d$ have the same direction. Let $\Omega \subset \mathbb{R}^d$ be compact and set $K(x,y) = |x - y|$. If there exists a K-invariant measure on Ω, then either Ω is a line segment or no three points of Ω are collinear.*

For the case when $K(x,y) = \|x - y\|$ is the Euclidean distance, this lemma shows, for example, that an invariant measure doesn't exist for the unit ball, while, as we know, it does exist for the sphere.

Finally, we make the remark that if a measure μ is K-invariant, it implies that a constant function is an eigenfunction of the Hilbert–Schmidt operator $T_{K,\mu}$ in $L^2(\Omega, \mu)$ with eigenvalue $\lambda = I_K(\mu)$, which is implied by (2.18).

3.2. A Crucial Identity.
The following simple relation provides a powerful direct link between energy minimization and (conditional) positive definiteness and will play a decisive role in many results of this section. It is also an important first step in the proof of the Generalized Stolarsky principle (Theorem 5.1). In a nutshell, this lemma states that, while I_K is a quadratic functional, it behaves linearly around a K-invariant measure.

Lemma 3.3. *Let K be a kernel on $\Omega \times \Omega$ and let μ be a K-invariant measure, i.e. $U_K^\mu(x) = I_K(\mu)$ for all $x \in \Omega$. Then for any $\nu \in \widetilde{\mathbb{P}}(\Omega)$,*

$$(3.2) \qquad\qquad I_K(\nu - \mu) = I_K(\nu) - I_K(\mu).$$

More generally, if $\mu \in \mathbb{P}(\Omega)$ *satisfies* $U_K^\mu(x) \geq I_K(\mu)$, *with equality on* $\text{supp}(\mu)$, *then for any* $\nu \in \mathbb{P}(\Omega)$

$$(3.3) \qquad\qquad I_K(\nu - \mu) \leq I_K(\nu) - I_K(\mu),$$

and equality (3.2) *holds for any measure* $\nu \in \widetilde{\mathbb{P}}(\Omega)$ *with* $\text{supp}(\nu) \subseteq \text{supp}(\mu)$.

Proof. If μ is K-invariant, then for any $\nu \in \widetilde{\mathbb{P}}(\Omega)$,

$$(3.4) \qquad\qquad I_K(\mu, \nu) = \int_\Omega U_K^\mu(x)\, d\nu(x) = \int_\Omega I_K(\mu)\, d\nu(x) = I_K(\mu).$$

Therefore

$$I_K(\nu - \mu) = I_K(\nu) - 2I_K(\mu, \nu) + I_K(\mu) = I_K(\nu) - I_K(\mu).$$

For the second part of our claim, observe that for any $\nu \in \mathbb{P}(\Omega)$, instead of equality (3.4), one has the inequality $I_K(\mu, \nu) \geq I_K(\mu)$, and thus,

$$I_K(\nu - \mu) = I_K(\nu) - 2I_K(\mu, \nu) + I_K(\mu) \leq I_K(\nu) - I_K(\mu).$$

Finally, the last statement follows from the first one by replacing Ω with $\text{supp}(\mu)$. $\qquad\square$

Theorem 2.14 and Corollary 2.16 show that if μ is a global (or at least local) minimizer of I_K, it satisfies the the conditions of the second statement in Lemma 3.3, and if in addition μ has full support, it also satisfies the first condition, i.e. μ is K-invariant. Thus Lemma 3.3 applies to (local) energy minimizers, which results in the following corollary:

Corollary 3.4. *Let K be a kernel on $\Omega \times \Omega$ and μ be a (local) minimizer of I_K. Then for any* $\nu \in \mathbb{P}(\Omega)$

$$(3.5) \qquad\qquad I_K(\nu - \mu) \leq I_K(\nu) - I_K(\mu).$$

For any $\nu \in \widetilde{\mathbb{P}}(\Omega)$ *such that* $\text{supp}(\nu) \subseteq \text{supp}(\mu)$, *then*

$$(3.6) \qquad\qquad I_K(\nu - \mu) = I_K(\nu) - I_K(\mu).$$

3.3. **Conditional Positive Definiteness and Energy Minimization.** Identity (3.2) of Lemma 3.3 provides a link between energy minimization and conditional positive definiteness. We would like to emphasize that relation (3.2) holds not just for probability measures ν, but for arbitrary signed measures of total mass one, i.e. $\nu \in \widetilde{\mathbb{P}}(\Omega)$. Therefore, one can immediately deduce the following equivalence.

Theorem 3.5. *Let K be a kernel on $\Omega \times \Omega$ and assume that μ is K-invariant. Then μ minimizes I_K over the set $\widetilde{\mathbb{P}}(\Omega)$ of normalized signed Borel measures if and only if K is conditionally positive definite.*

Moreover, μ uniquely minimizes I_K over $\widetilde{\mathbb{P}}(\Omega)$ if and only if K is conditionally strictly positive definite.

Proof. Suppose that K is conditionally positive definite. Then for any $\nu \in \widetilde{\mathbb{P}}(\Omega)$, equality (3.2) holds and, since $(\nu - \mu)(\Omega) = 0$, we have

$$I_K(\nu) - I_K(\mu) = I_K(\nu - \mu) \geq 0,$$

which shows that μ minimizes I_K over $\widetilde{\mathbb{P}}(\Omega)$. If K is conditionally strictly positive definite, then $I_K(\nu) = I_K(\mu)$ only if $\nu - \mu = 0$, i.e. μ is the unique minimizer.

Assume conversely that $I_K(\mu) \leq I_K(\nu)$ for each $\nu \in \widetilde{\mathbb{P}}(\Omega)$. Consider an arbitrary signed measure $\gamma \in \mathcal{Z}(\Omega)$. Define $\nu = \mu + \gamma$, then $\nu(\Omega) = 1$, i.e. $\nu \in \widetilde{\mathbb{P}}(\Omega)$. Thus, applying (3.2) once again, we find that

$$I_K(\gamma) = I_K(\nu - \mu) = I_K(\nu) - I_K(\mu) \geq 0,$$

hence K is conditionally positive definite. If μ is the unique minimizer, then the expression above equals zero only for $\gamma = 0$, i.e. K is conditionally strictly positive definite.   □

Obviously, one of the implications holds for minimizers over probability measures $\mathbb{P}(\Omega)$.

Corollary 3.6. *Let K be a kernel on $\Omega \times \Omega$ and assume that μ is K-invariant. If K is conditionally (strictly) positive definite, then μ (uniquely) minimizes I_K over the set $\mathbb{P}(\Omega)$ of Borel probability measures.*

This corollary is well known, see, e.g., Theorem 4.2.11 in [BHS], but the equivalence in Theorem 3.5 appears to be new.

Some remarks are in order. We would like to remind the reader that in some specific cases, such as the sphere with the uniform surface measure (or more generally, two-point homogeneous spaces with the corresponding uniform measures), the relation between energy minimization and some form of positive definiteness of the kernel is well known [Sch, BDM]. However, it is usually demonstrated using the representation theory of the underlying space and the associated orthogonal polynomial (Gegenbauer, Jacobi, Krawtchouk) expansions. In fact, Theorem 3.5 can be viewed as a generalization of the so-called *mean inequality* due to Kabatianskii and Levenshtein [CS, Theorem 5, Chapter 9.3], in which K is an invariant positive definite kernel on a 2-point homogeneous space. The Theorem is thus a blanket statement that covers all of these examples and beyond. Moreover, it relies only on the completely elementary identity (3.2), thus simplifying the known proofs in all of the specific cases. In the spherical case, a similar approach has been recently employed in [BDM].

3.4. **Conditional Positive Definiteness and Positive Definiteness up to a Constant (Revisited).** As we have observed in the previous discussions, two properties, which are somewhat weaker than positive definiteness, play an important role in energy minimization: namely, conditional positive definiteness and positive definiteness up to an additive constant. We have already demonstrated in Lemma 2.4 that the latter always implies the former, and the converse implication is not true in general. We shall now show that the converse implication also holds, i.e. conditional positive definiteness implies positive definiteness up to an additive constant, if we additionally assume the existence of a K-invariant measure. Moreover, the statement also holds for the "strict" version of these properties.

Lemma 3.7. *Let K be a kernel on $\Omega \times \Omega$ and assume that K is conditionally (strictly) positive definite. Suppose also that there exists a K-invariant measure $\mu \in \mathbb{P}(\Omega)$. Then K is (strictly) positive definite up to an additive constant.*

Proof. Let K be conditionally (strictly) positive definite. Set $C = -I_K(\mu) + 1$. Then $K + C$ is still conditionally (strictly) positive definite, μ is $(K+C)$-invariant, and $I_{K+C}(\mu) = (\mu(\Omega))^2 = 1$. For any signed measure ν with $\nu(\Omega) = 1$, identity (3.2) implies that

$$I_{K+C}(\nu) - I_{K+C}(\mu) = I_{K+C}(\nu - \mu) = I_K(\nu - \mu) \geq 0.$$

Therefore, $I_{K+C}(\nu) \geq 1 > 0$.

Now consider an arbitrary measure $\gamma \in \mathcal{M}(\Omega)$. If $\gamma(\Omega) = 0$, then $I_{K+C}(\gamma) \geq 0$ by conditional positive definiteness (and $I_{K+C}(\gamma) > 0$ for $\gamma \neq 0$ for the "strict" version if $\gamma \neq 0$).

If $\gamma(\Omega) = c \neq 0$, we can write $\gamma = c\nu$ for some $\nu \in \widetilde{\mathbb{P}}(\Omega)$. Therefore, $I_{K+C}(\gamma) = c^2 I_{K+C}(\nu) \geq c^2 > 0$. Hence $K + C$ is (strictly) positive definite.   □

Lemmas 2.4 and 3.7 together show that in the presence of a K-invariant measure, conditional positive definiteness and positive definiteness modulo an additive constant are equivalent notions. This is the case, for example, for rotationally invariant kernels K on the sphere, since the uniform surface measure σ is K-invariant for all such kernels.

3.5. **Local and Global Minimizers.** Under some additional assumptions, local minimizers of I_K are necessarily global minimizers. Some facts of this type have been observed in the papers of the authors with D. Ferizović, A. Glazyrin, J. Park [BFGMPV, BGMPV]. Here we prove a variety of more general statements with the same flavor.

In this subsection, the words "local minimizer" in the assumptions mean "directional local minimizer" in the sense of Definition 2.15. Thus these results are also valid for local minimizers in total variation or in the Wasserstein distance W_p with $1 \le p < \infty$, according to (2.11).

Proposition 3.8. *Suppose that μ is a local minimizer of I_K and any of the following two conditions holds:*

   (1) the measure μ is K-invariant, i.e. $U_K^\mu(x) = I_K(\mu)$ for all $x \in \Omega$;
   (2) K is conditionally positive definite.
Then μ is a global minimizer of I_K.

Recall that, according to Corollary 2.16, condition (1) is automatically satisfied if μ has full support (and so is condition (2), due to Lemmas 2.4 and 2.22). Notice also that, unlike Corollary 3.6, part (2) does not require K-invariance of μ, but assumes instead that μ is a local minimizer.

Proof. Assume that (1) holds. Then for any $\nu \in \mathbb{P}(\Omega)$, since μ is a local minimizer, applying equation (3.2) of Lemma 3.3 for small $t > 0$ gives

$$0 \le I_K((1-t)\mu + t\nu) - I_K(\mu) = I_K\big(t(\nu - \mu)\big) = t^2 I_K(\nu - \mu) = t^2\Big(I_K(\nu) - I_K(\mu)\Big).$$

Thus $I_K(\mu) \le I_K(\nu)$, i.e. μ is a global minimizer of I_K in $\mathbb{P}(\Omega)$.

Now assume that (2) holds. Then for each $\nu \in \mathbb{P}(\Omega)$, the measure $\nu - \mu$ has total mass zero. According to conditional positive definiteness of K and inequality (3.5) of Corollary 3.4,

$$(3.7) \qquad\qquad I_K(\nu) - I_K(\mu) \ge I_K(\nu - \mu) \ge 0.$$

Therefore, in this case, μ also minimizes I_K. $\qquad\square$

Observe that if μ is K-invariant, then, according to part (1) of Proposition 3.8, the implications of (2.11) may be reversed, i.e. three definitions of local minimizers (directional, in total variation, and in the Wasserstein distance W_p, $1 \le p < \infty$) are equivalent, since in either of the cases the minimizer is necessarily global.

Corollary 3.9. *Assume that $\mu \in \mathbb{P}(\Omega)$ is K-invariant. Then the following conditions are equivalent:*

   - *μ is a directional local minimizer of I_K;*
   - *μ is local minimizer of I_K with respect to the total variation norm;*
   - *μ is local minimizer of I_K with respect to the Wasserstein distance W_p, $1 \le p < \infty$ (if Ω is a metric space);*
   - *μ is a global minimizer of I_K on $\mathbb{P}(\Omega)$.*

When μ is a local minimizer with full support, both conditions (1) and (2) of Proposition 3.8 hold simultaneously, and an even stronger conclusion can be drawn.

Proposition 3.10. *Let μ be a local minimizer of I_K with $\mathrm{supp}(\mu) = \Omega$. Then μ is a global minimizer of I_K over $\widetilde{\mathbb{P}}(\Omega)$, the set of all signed Borel measures with total mass one.*

Proof. As discussed above, if μ is a local minimizer with full support, then both K-invariance of μ and conditional positive definiteness of K immediately follow, from Corollary 2.16 and Lemmas 2.4 and 2.22, respectively. Theorem 3.5 then shows that μ minimizes I_K over $\widetilde{\mathbb{P}}(\Omega)$. $\square$

Finally, another version of the local-to-global minimization principle may be proved under the assumption that I_K is convex at μ, which according to Lemma 2.9, is equivalent to the fact that the arithmetic mean inequality (2.7) holds for any measure $\nu \in \mathbb{P}(\Omega)$. We have the following statement.

Theorem 3.11. *Suppose that K is a kernel on $\Omega \times \Omega$ and that for some $\mu \in \mathbb{P}(\Omega)$ there exists a constant $M \in \mathbb{R}$ such that $U_K^\mu(x) \geq M$, with equality on $\mathrm{supp}(\mu)$. If I_K is convex at μ, then μ is a global minimizer of I_K over $\mathbb{P}(\Omega)$.*

Before proving this statement we observe that its first assumption is satisfied in either of the following two cases: (a) if μ is a local minimizer, according to Corollary 2.16; (b) if μ is K-invariant. Thus we have two immediate corollaries. The first one is the aforementioned local-to-global principle.

Corollary 3.12. *Let K be a kernel on $\Omega \times \Omega$ and let $\mu \in \mathbb{P}(\Omega)$ be a local minimizer of I_K. If I_K is convex at μ, then μ is a global minimizer of I_K over $\mathbb{P}(\Omega)$.*

Observe that, if convexity at μ were replaced with convexity of I_K on $\mathbb{P}(\Omega)$, then in view of Proposition 2.11, this would be equivalent to the conditional positive definiteness of K. Thus, Corollary 3.12 recovers and strengthens part (2) of Proposition 3.8.

The second corollary will provide a crucial implication in Theorem 4.1.

Corollary 3.13. *Let K be a kernel on $\Omega \times \Omega$ and let $\mu \in \mathbb{P}(\Omega)$ be a K-invariant measure. If I_K is convex at μ, then μ is a global minimizer of I_K over $\mathbb{P}(\Omega)$.*

Proof of Theorem 3.11. Observe first that the constant M is necessarily equal to $I_K(\mu)$:

$$I_K(\mu) = \int_\Omega U_K^\mu(x) \, d\mu(x) = \int_{\mathrm{supp}(\mu)} M d\mu(x) = M.$$

For any $\nu \in \mathbb{P}(\Omega)$, we have that

$$I_K(\mu, \nu) = \int_\Omega U_K^\mu(x) \, d\nu(x) \geq I_K(\mu).$$

Convexity of I_K at μ, according to Lemma 2.9, is equivalent to the arithmetic mean inequality (2.7). Thus,

$$I_K(\mu) \leq I_K(\mu, \nu) \leq \frac{1}{2} I_K(\mu) + \frac{1}{2} I_K(\nu),$$

so $I_K(\nu) \geq I_K(\mu)$. $\square$

4. Invariant Measures and Minimizers with Full Support

It is now time to summarize the results of the previous sections. It may not be yet obvious, but we have proven (sometimes quite surprising) equivalences between many different notions. We shall restrict our attention to the case when the measure μ is K-invariant (i.e. has constant potential) and has full support. As we have discussed before, these conditions are satisfied by many natural candidates (the uniform measure on the sphere or other two-point homogeneous spaces, the Haar measure on a compact topological group, etc). Though a majority of the implications are valid even just for K-invariant measures without the full support assumption,

assuming that μ has full support truly ties the picture together. We shall carefully trace which of the conclusions require this condition.

We start with the following long list of equivalences.

Theorem 4.1. *Let K be a kernel on $\Omega \times \Omega$. Assume that there exists a measure $\mu \in \mathbb{P}(\Omega)$, which is K-invariant and has full support, i.e. $U_K^\mu(x) = I_K(\mu)$ for all $x \in \Omega$ and $\mathrm{supp}(\mu) = \Omega$. Then the following conditions are equivalent:*

(1) K positive definite modulo a constant.

(2) K is conditionally positive definite.

(3) μ is a local minimizer of I_K.

(4) μ is a global minimizer of I_K over $\mathbb{P}(\Omega)$.

(5) μ is a global minimizer of I_K over $\widetilde{\mathbb{P}}(\Omega)$.

(6) I_K is convex on $\mathbb{P}(\Omega)$ (or, equivalently, on $\widetilde{\mathbb{P}}(\Omega)$).

(7) I_K is convex at μ.

(8) The arithmetic mean inequality (2.3) holds for all $\mu_1, \mu_2 \in \mathbb{P}(\Omega)$ (or, equivalently for all $\mu_1, \mu_2 \in \widetilde{\mathbb{P}}(\Omega)$).

(9) The arithmetic mean inequality (2.3) holds when $\mu_1 = \mu$.

(10) The kernel K can be represented as

$$K(x,y) = \sum_{j=1}^{\dim(L^2(\Omega,\mu))} \lambda_j \phi_j(x) \phi_j(y)$$

where the series converges uniformly and absolutely, the function ϕ_1 is constant, and $\lambda_j \geq 0$ for $j \geq 2$.
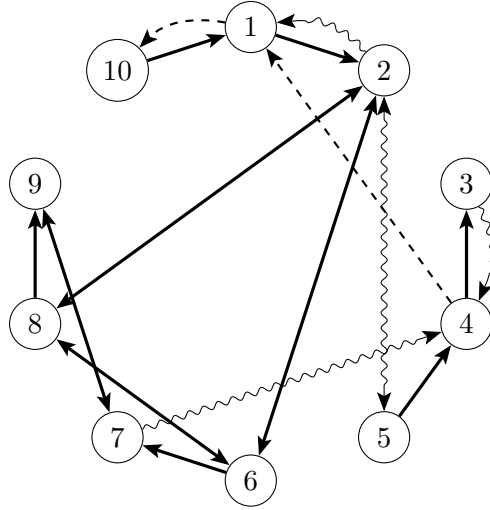

FIGURE 1. Implications in the proof of Theorem 4.1: solid arrows are implications that hold without additional assumptions; wavy arrows require K-invariance, but not full support; dashed arrows represent the implications which do require the assumption of full support.

Proof. For the reader's convenience the implications proving this theorem are summarized in Figure 1.

We open with a list of the implications that do not require any assumptions on μ. It is obvious that (5) implies (4), which in turn implies (3). Also, (6) implies (7), and similarly, (8) implies (9).

The equivalence between (2), (6), and (8) is proved in Lemmas 2.6 and 2.9 together with Proposition 2.11. Lemma 2.9 also establishes the equivalence between (7) and (9). Lemma 2.4 shows that (1) implies (2).

The following implications rely on the fact that μ is K-invariant, but do not require μ to have full support. Lemma 3.7 demonstrates that (2) implies (1). Theorem 3.5 yields the equivalence of (2) and (5). Corollary 3.13 shows that (4) follows from (7). Finally, part (1) of Proposition 3.8 guarantees that (3) implies (4).

The equivalence between (1) and (10) is discussed in Lemma 2.3 and Mercer's Theorem (see Theorem 2.18), or more specifically Corollary 2.19. Condition (1) implies (10) due to the fact that the constant function $\mathbb{1}_\Omega$ is an eigenfunction of the Hilbert–Schmidt operator $T_{K,\mu}$ in $L^2(\Omega, \mu)$ for any K-invariant measure μ, and the expansion in Mercer's Theorem is valid on all of Ω, since $\mathrm{supp}(\mu) = \Omega$. The implication $(10) \Rightarrow (1)$ holds without any additional assumptions, according to Lemma 2.3.

In conclusion, we observe that Lemma 2.22 demonstrates that (4) implies (1), which closes the loop of implications – and, except for the standalone equivalence between (1) and (10), this is the only implication in our proof where the fact that $\mathrm{supp}(\mu) = \Omega$ is used. Indeed, Lemma 2.22 only guarantees that the kernel K is positive definite (up to constant) on the support of the minimizer. Observe also that due to Theorem 2.14, if (4) holds and μ has full support, then μ is automatically K-invariant. $\qquad\square$

To reiterate, this theorem reveals several interesting novel properties of a K-invariant measure (with full support):

- Equivalence between minimization over the set $\mathbb{P}(\Omega)$ of probability measures and the set $\widetilde{\mathbb{P}}(\Omega)$ of all signed measures of mass one. This effect has been observed for rotationally invariant kernels on the sphere and the surface measure σ by two of the authors and F. Dai [BDM]. This is not necessarily the case in other settings. In particular, for the integral over the unit ball $\mathbb{B}^d$

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

  according to [Bj], the unique maximizer over both $\mathbb{P}(\mathbb{B}^d)$ and $\mathbb{P}(\mathbb{S}^{d-1})$ is σ. According to the aforementioned equivalence, σ is also a maximizer over $\widetilde{\mathbb{P}}(\mathbb{S}^{d-1})$, while in the case of signed measures on the ball, the maximizer does not exist [HNW]. Observe, that one cannot expect such minimizers to exist in general, since $\widetilde{\mathbb{P}}(\Omega)$ is not weak* compact.
- Equivalence between being a local and global energy minimizer. This effect, in a slightly less general form, has been observed by the authors and their collaborators in [BFGMPV, BGMPV].
- Equivalence between other local and global properties: e.g., the energy functional I_K is convex on the whole set of probability measures if and only if it is convex just at the special measure μ.
- Equivalence between conditional positive definiteness of the kernel and positive definiteness up to constant, which is not true in general.
- We also note that, due to Corollary 3.9, any of the three notions of local minimizers (directional, in total variation, and in the Wasserstein distance W_p, $1 \le p < \infty$, for

metric spaces Ω) may be assumed in part (3) of Theorem 4.1, as they are equivalent for K-invariant measures.

We now formulate similar theorems for kernels which have the "strict" version of the properties and for positive definite kernels. We start with the latter.

Theorem 4.2. *Suppose that K is a kernel on $\Omega \times \Omega$ and that there exists a measure $\mu \in \mathbb{P}(\Omega)$, which is K-invariant and has full support, i.e. $U_K^\mu(x) = I_K(\mu)$ for all $x \in \Omega$ and $\operatorname{supp}(\mu) = \Omega$. Then the following conditions are equivalent:*

(1) The kernel K is positive definite.
(2) The geometric mean inequality (2.5) and $I_K(\mu_1) \geq 0$ hold for all $\mu_1, \mu_2 \in \mathbb{P}(\Omega)$.
(3) The measure μ is a global minimizer of I_K and satisfies $I_K(\mu) \geq 0$.
(4) I_K is convex on $\mathcal{M}(\Omega)$.
(5) The kernel K can be represented as

$$K(x,y) = \sum_{j=1}^{\dim(L^2(\Omega,\mu))} \lambda_j \phi_j(x) \phi_j(y)$$

where the series converges uniformly and absolutely, and $\lambda_j \geq 0$ for $j \geq 1$.
(6) There exists some symmetric $k \in L^2(\Omega \times \Omega, \mu \times \mu)$ such that for all $x, y \in \Omega$,

$$K(x,y) = \int_\Omega k(x,z) k(z,y) \, d\mu(z).$$

Proof. Lemma 2.7, Proposition 2.13, Corollary 2.19, and Proposition 2.20 show the equivalence of (1), (2), (4), (5), and (6). Positive definiteness, i.e. condition (1), guarantees that $I_K(\mu) \geq 0$, and that μ is a minimizer, due to Theorem 4.1. Conversely, Lemma 2.22 shows that (3) implies (1), finishing our proof. Observe also that according to Theorem 4.1 it does not matter whether we mean global minimization over $\mathbb{P}(\Omega)$ or $\widetilde{\mathbb{P}}(\Omega)$ in condition (3). $\square$

Theorem 4.3. *Suppose that K is a kernel on $\Omega \times \Omega$ and that there exists a measure $\mu \in \mathbb{P}(\Omega)$ which is K-invariant, i.e. $U_K^\mu(x) = I_K(\mu)$ for all $x \in \Omega$. Then the following conditions are equivalent:*

(1) K is conditionally strictly positive definite.
(2) K is strictly positive definite modulo a constant.
(3) μ is the unique minimizer of I_K over $\widetilde{\mathbb{P}}(\Omega)$.
(4) I_K is strictly convex on $\mathbb{P}(\Omega)$ (or, equivalently, on $\widetilde{\mathbb{P}}(\Omega)$).

If in addition $\operatorname{supp}(\mu) = \Omega$, i.e. μ has full support, then each of the conditions (1)–(3) implies the following

(5) The kernel K can be represented as

$$K(x,y) = \sum_{j=1}^{\dim(L^2(\Omega,\mu))} \lambda_j \phi_j(x) \phi_j(y)$$

where $\{\phi_j\}$ is the orthonormal basis consisting of eigenfunctions of the Hilbert–Schmidt operator $T_{K,\mu}$ in $L^2(\Omega, \mu)$, the function ϕ_1 is a constant, the series converges uniformly and absolutely, and $\lambda_j > 0$ for $j \geq 2$.

Moreover, if the span of $\{\phi_j\}$ is dense in $C(\Omega)$, then (5) also implies conditions (1)–(3).

Proof. Lemma 2.4 shows that (2) implies (1), while Lemma 3.7 provides the converse implication. The equivalence of (3) and (1) is proved in Theorem 3.5. For equivalence of (1) and (4), see the discussion after Proposition 2.11.

Before we turn to dealing with condition (5), recall that K-invariance of μ implies that a constant is an eigenfunction of the operator $T_{K,\mu}$, so we shall assume that $\phi_1 = \mathbb{1}_\Omega$.

Now we show that (3) implies (5). Since μ minimizes I_K and has full support, by Theorem 4.1, K is positive definite up to a constant, and since $\phi_1 = 1$, by Mercer's Theorem, the expansion in part (5) holds with some $\lambda_j \in \mathbb{R}$, such that $\lambda_j \geq 0$ for $j \geq 2$, and with the series converging uniformly and absolutely. Suppose, indirectly, that μ is the unique minimizer of I_K over $\widetilde{\mathbb{P}}(\Omega)$, but there exists some $l \geq 2$ such that $\lambda_l = 0$. Let $d\nu(x) = (1 + \phi_l(x)) \, d\mu(x)$. Orthogonality implies that $\int_\Omega \phi_l(x) \, d\mu(x) = 0$, therefore $\nu \in \widetilde{\mathbb{P}}(\Omega)$. Then we obtain

$$
\begin{aligned}
I_K(\nu) &= \int_\Omega \int_\Omega K(x,y)(1 + \phi_l(x))(1 + \phi_l(y)) \, d\mu(x) d\mu(y) \\
&= I_K(\mu) + 2\langle T_{K,\mu}\phi_l, \mathbb{1}_\Omega \rangle_{L^2(\Omega,\mu)} + \langle T_{K,\mu}\phi_l, \phi_l \rangle_{L^2(\Omega,\mu)} \\
&= I_K(\mu) + 2\lambda_l \langle \phi_l, \mathbb{1}_\Omega \rangle_{L^2(\Omega,\mu)} + \lambda_l \|\phi_l\|^2_{L^2(\Omega,\mu)} = I_K(\mu),
\end{aligned}
$$

which contradicts the fact that μ is the unique minimizer over $\widetilde{\mathbb{P}}(\Omega)$.

Finally, we show that (5) implies (2) under the aforementioned additional assumption. Let $K'(x,y) = K(x,y) - \lambda_1 + 1$ and $\nu \in \mathcal{M}(\Omega)$. Then

$$
\begin{aligned}
I_{K'}(\nu) &= \int_\Omega \int_\Omega K'(x,y) \, d\nu(x) d\nu(y) \\
&= (\nu(\Omega))^2 + \sum_{j=2}^{\dim(L^2(\Omega,\mu))} \int_\Omega \int_\Omega \lambda_j \phi_j(x)\phi_j(y) \, d\nu(x) d\nu(y) \\
&= (\nu(\Omega))^2 + \sum_{j=2}^{\dim(L^2(\Omega,\mu))} \lambda_j \Big(\int_\Omega \phi_j(x) \, d\nu(x) \Big)^2 \geq 0.
\end{aligned}
$$

Clearly, the only way that $I_{K'}(\nu) = 0$ is if $\int_\Omega \phi_j(x) \, d\nu(x) = 0$ for all $j \geq 1$. By the density of $\text{span}\{\phi_j\}_{j \geq 1}$ in $C(\Omega)$, we can conclude that this implies $\nu = 0$, so K' must be strictly positive definite, which completes the proof. $\square$

We conclude with the remark that the additional condition imposed for the sufficiency of condition (5) is not very restrictive in practice. For example in the case of rotationally invariant kernels on the sphere and $\mu = \sigma$, the eigenfunctions ϕ_j are simply spherical harmonics, which span all polynomials on the sphere and thus their span is dense in the space continuous functions.

5. The Generalized Stolarsky Principle on Compact Spaces

Discrepancy theory analyzes discrete point configurations by comparing their distribution to some fixed (usually uniform) measure μ on a class of test sets. More precisely, the local discrepancy of a finite configuration $\omega_N = \{z_1, \ldots, z_N\} \subset \Omega$ with respect to a set $A \subseteq \Omega$ is defined as

$$
(5.1) \qquad D(\omega_N, A) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_A(z_i) - \mu(A).
$$

One then takes the supremum over some class $\mathcal{A}$ of test sets A (extremal discrepancy) or a certain average, if the class $\mathcal{A}$ admits a natural measure (quadratic averages lead to the so-called L^2 discrepancy). These quantities provide important information about the distribution of the discrete set ω_N [BC, Ma, KN].

Discrepancy is closely related to discrete energy

$$(5.2) \qquad E_K(\omega_N) := \frac{1}{N^2} \sum_{x,y \in \omega_N} K(x,y).$$

The definitions of discrete (5.2) and continuous (1.2) energies are compatible in the sense that

$$(5.3) \qquad E_K(\omega_N) = I_K(\mu_{\omega_N}), \quad \text{where } \mu_{\omega_N} = \frac{1}{N} \sum_{x \in \omega_N} \delta_x.$$

Similarly, the definition of discrepancy may be extended to the continuous setting by replacing the empirical measure $\mu_{\omega_N}(A) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_A(z_i)$ by an arbitrary measure $\nu \in \widetilde{\mathbb{P}}(\Omega)$.

The foundational example of the connection between energy and discrepancy is given by the classical Stolarsky Invariance Principle [St] which states that

$$(5.4) \qquad c_d D^2_{L^2,\mathrm{cap}}(\omega_N) = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} ||x - y|| \, d\sigma(x) \, d\sigma(y) \; - \; \frac{1}{N^2} \sum_{i,j=1}^{N} ||z_i - z_j||,$$

where $D_{L^2,\mathrm{cap}}(\omega_N)$ is the L^2 spherical cap discrepancy

$$(5.5) \qquad D^2_{L^2,\mathrm{cap}}(\omega_N) = \int_{-1}^{1} \int_{\mathbb{S}^{d-1}} \left| \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{C(x,h)}(z_j) - \sigma\big(C(x,h)\big) \right|^2 d\sigma(x) \, dh$$

and $C(x,h)$ denotes a spherical cap of height $h \in [-1,1]$ centered at $x \in \mathbb{S}^{d-1}$, i.e. $C(x,h) = \{z \in \mathbb{S}^{d-1} : \langle z, x \rangle > h\}$. This principle shows that minimizing the L^2 spherical cap discrepancy is equivalent to an energy optimization problem – maximizing the pairwise sum of Euclidean distances. Observe that by setting $K(x,y) = 1 - c_d^{-1} ||x - y||$, one can rewrite (5.4) as

$$(5.6) \qquad D^2_{L^2,\mathrm{cap}}(\omega_N) = I_K(\nu) - I_K(\sigma)$$

with $\nu = \mu_{\omega_N}$, and the right-hand side of this relation is familiar to the reader from identity (3.2) of Lemma 3.3.

In the recent years, numerous authors (including the first two authors of this paper) revisited this fascinating fact, extended it, and applied it to various problems of discrete geometry and optimization: new proofs of the original Stolarsky Invariance Principle have been given in [BrD, BDM, HBZO], it has been extended to geodesic distances and other rotationally invariant kernels on the sphere [BD, BDM], to compact, connected, two-point homogeneous spaces [Sk1, Sk2] and to the Hamming cube [Ba, BS], and applied to two problems of Fejes Tóth on sums of various distances on the sphere and in projective spaces [BDM, BM].

In this paper, we prove a general version of this principle on arbitrary compact spaces which extends some of the versions mentioned above and does not use any structural information about the underlying domain. Let Ω be a compact topological space and let us fix a measure $\mu \in \mathbb{P}(\Omega)$ – this will usually be an energy minimizing (equilibrium) measure or an invariant measure of full support (its role is similar to that of σ in the spherical case). We now define the L^2 *discrepancy*

of an arbitrary probability measure $\nu \in \mathbb{P}(\Omega)$ (or even a signed measure $\nu \in \widetilde{\mathbb{P}}(\Omega)$) relative to the measure μ with respect to the function $k : \Omega \times \Omega \to \mathbb{R}$ by the identity

$$
\begin{aligned}
D^2_{L^2,k,\mu}(\nu) &= \int\limits_\Omega \left| \int\limits_\Omega k(x,y)\, d\nu(y) - \int\limits_\Omega k(x,y)\, d\mu(y) \right|^2 d\mu(x) \\
&= \int\limits_\Omega \left| \int\limits_\Omega k(x,y)\, d(\nu-\mu)(y) \right|^2 d\mu(x).
\end{aligned}
\tag{5.7}
$$

When ν is the equal-weight discrete measure associated to the N-point set $\omega_N = \{z_1, ..., z_N\} \subset \Omega$, i.e. $\nu = \dfrac{1}{N} \sum\limits_{i=1}^N \delta_{z_i}$, this becomes the discrepancy of the set ω_N with respect to k:

$$
D^2_{L^2,k,\mu}(\omega_N) = D^2_{L^2,k,\mu}\Big(\frac{1}{N}\sum_{i=1}^N \delta_{z_i}\Big) = \int\limits_\Omega \left| \frac{1}{N}\sum_{i=1}^N k(x,z_i) - \int\limits_\Omega k(x,y)\, d\mu(y) \right|^2 d\mu(x)
\tag{5.8}
$$

Notice that, in the spherical case, setting $k(x,y) = \mathbb{1}_{\{\langle x,y \rangle > h\}} = \mathbb{1}_{C(x,h)}(y)$, one finds that (5.8) is equal to the inner integral in (5.5) with the integrand of the form (5.1). Therefore, this definition is indeed an extension of the classical notion of discrepancy with arbitrary functions k in place of indicators of test sets.

We can now obtain the following general version of the Stolarsky Invariance Principle:

Theorem 5.1 (Generalized Stolarsky Principle)**.** *Let K be a positive definite (modulo an additive constant C) kernel on $\Omega \times \Omega$. Let us assume that $\mu \in \mathbb{P}(\Omega)$ is a K-invariant measure with full support, i.e. $U^\mu_K(x) = I_K(\mu)$ for all $x \in \Omega$ and $\mathrm{supp}(\mu) = \Omega$. Then for every measure $\nu \in \widetilde{\mathbb{P}}(\Omega)$, we have the following identity.*

$$
I_K(\nu) - I_K(\mu) = D^2_{L^2,k,\mu}(\nu),
\tag{5.9}
$$

where the function $k \in L^2(\Omega \times \Omega, \mu \times \mu)$ is as in part (6) of Theorem 4.2 applied to the positive definite kernel $K + C$.

In particular, for a discrete set $\omega_N = \{z_1, ..., z_N\} \subset \Omega$,

$$
E_K(\omega_N) - I_K(\mu) = D^2_{L^2,k,\mu}(\omega_N).
\tag{5.10}
$$

This theorem has the following immediate corollary:

Corollary 5.2. *Let K be a kernel on $\Omega \times \Omega$. Assume that $\mu \in \mathbb{P}(\Omega)$ is a global minimizer of the energy functional I_K over $\mathbb{P}(\Omega)$ with $\mathrm{supp}(\mu) = \widetilde{\Omega} \subseteq \Omega$. Then identity (5.9) holds for any signed measure ν with total mass one, whose support is contained in the support of μ, i.e. $\nu \in \widetilde{\mathbb{P}}(\widetilde{\Omega})$. Similarly, relation (5.10) holds for any point set $\omega_N = \{z_1, ..., z_N\} \subset \widetilde{\Omega}$.*

Proof. If μ is a global minimizer of I_K, by Theorem 2.14, the potential U^μ_K is constant on $\widetilde{\Omega}$, i.e. μ is K-invariant and has full support if viewed as an element of $\mathbb{P}(\widetilde{\Omega})$. Moreover, according to Lemma 2.22, the kernel K is positive definite (modulo a constant) on $\widetilde{\Omega}$. Therefore, the statement follows directly from Theorem 5.1 applied to $\widetilde{\Omega}$ in place of Ω. $\qquad\square$

We now turn to the proof of the generalized Stolarsky principle:

Proof of Theorem 5.1. Without loss of generality, we can assume that K is positive definite, since adding a constant to K affects neither the invariance of μ nor the difference $I_K(\nu) - I_K(\mu)$. We

can now use the crucial identity (3.2) of Lemma 3.3, as well as part (6) of Theorem 4.2, to obtain

$$I_K(\nu) - I_K(\mu) = I_K(\nu - \mu) = \int_\Omega \int_\Omega K(x,y) d(\nu - \mu)(x) d(\nu - \mu)(y)$$

(5.11)
$$= \int_\Omega \int_\Omega \int_\Omega k(x,z) k(z,y)\, d\mu(z) d(\nu - \mu)(x) d(\nu - \mu)(y)$$

$$= \int_\Omega \left| \int_\Omega k(x,z) d(\nu - \mu)(x) \right|^2 d\mu(z) = D^2_{L^2, k, \mu}(\nu).$$

$\square$

Remark: Observe that, for $k \in L^2(\Omega \times \Omega, \mu \times \mu)$, it is not technically obvious that the definition of $D^2_{L^2, k, \mu}(\nu)$ in (5.7) is properly justified: we do not know a priori that k is integrable with respect to ν, only with respect to μ. (This problem does not occur in the discrete case since we know that $k(\cdot, z_i) \in L^2(\Omega, \mu)$ for each $i = 1, ..., N$.) However, the proof of Stolarsky principle (5.9) demonstrates that the L^2 discrepancy $D^2_{L^2, k, \mu}$ is well-defined for any Borel measure ν. Indeed, the inner integral with respect to $d\mu(z)$ in (5.11) is defined according to part (6) of Theorem 4.2, and, moreover, produces the function $K(x, y)$, which is continuous and therefore integrable with respect to the finite Borel measure $(\nu - \mu) \times (\nu - \mu)$ on $\Omega \times \Omega$. Hence Fubini's theorem applies and

$$\int_\Omega \int_\Omega k(x,z) k(y,z) d(\nu - \mu)(x) d(\nu - \mu)(y) = \left| \int_\Omega k(x,z) d(\nu - \mu)(x) \right|^2$$

is finite for μ-a.e. z and is integrable with respect to $d\mu(z)$, i.e. $D^2_{L^2, k, \mu}(\nu)$ is well-defined.

6. The Spherical Case

We end with a brief discussion of the results of prior sections in the particular case when the domain is the sphere in Euclidean space. More specifically, let $\Omega = \mathbb{S}^{d-1}$ and let K be a rotationally invariant kernel on the sphere, i.e. $K(x, y) = F(\langle x, y \rangle)$ for each $x, y \in \mathbb{S}^{d-1}$, where $F \in C[-1, 1]$. By a slight abuse of notation, we shall also call the kernel F and the energy I_F.

We immediately observe that σ, the normalized surface measure on $\mathbb{S}^{d-1}$, is an F-invariant measure with full support. Therefore, all of the results of Sections 2–4 apply. In particular, Theorem 4.1 holds with $\mu = \sigma$. Some of its conclusions are interesting even in this classical case. We do not completely restate Theorems 4.1 and 4.2 for $\Omega = \mathbb{S}^{d-1}$, but simply summarize some interesting facts.

For any function $F \in C[-1, 1]$ which generates a kernel $F(\langle x, y \rangle)$ on $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$,
  (i) conditional positive definiteness of F on the sphere is equivalent to positive definiteness up to an additive constant;
 (ii) the facts that σ minimizes in three different ways I_F (locally, globally over probability measures, globally over signed measures of total mass one) are all equivalent to each other;
(iii) in turn, the fact that σ minimizes I_F is equivalent to conditional positive definiteness of F on the sphere.

Some results of part (ii) have been observed in [BDM, BGMPV] and part (iii) is well known, see, e.g., [BDM] (but heuristically it goes back to [Sch]).

Further considerations are connected to the Gegenbauer expansions of positive definite kernels. The Gegenbauer polynomials $\{C_n^\lambda\}_{n \geq 0}$ form an orthogonal basis of the space $L^2([-1, 1], w_\lambda)$ of

functions on $[-1, 1]$, which are square integrable with respect to the weight $w_\lambda(t) = (1 - t^2)^{\lambda - \frac{1}{2}}$. These functions are closely related to harmonic analysis on the sphere, see e.g. [DX] for details, when $\lambda = \frac{d-2}{2}$ (which we assume for the rest of the section).

The kernel F can be expanded in Gegenbauer series

$$(6.1) \qquad F(t) = \sum_{n=0}^{\infty} \widehat{F}(n, \lambda) \frac{n + \lambda}{\lambda} C_n^\lambda(t)$$

in the sense of $L^2([-1, 1], w_\lambda)$. In addition, the Funk–Hecke formula, which states that

$$(6.2) \qquad \int_{\mathbb{S}^{d-1}} F(\langle x, y \rangle) Y_n(y) \, d\sigma(y) = \widehat{F}(n, \lambda) Y_n(x)$$

whenever Y_n is a spherical harmonic of degree n on $\mathbb{S}^{d-1}$, demonstrates that the spherical harmonics are exactly the eigenfunctions of the Hilbert–Schmidt operator $T_{F,\sigma}$, with the corresponding Gegenbauer coefficients as eigenvalues. Thus, equivalence between positivity of $T_{F,\sigma}$ and positive definiteness of F, recovers the seminal result of Schoenberg [Sch] which asserts that positive definiteness on the sphere is equivalent to non-negativity of Gegenbauer coefficients. Moreover, if $\{Y_{n,k}\}_{k=1}^{\dim(\mathcal{H}_n^d)}$ is an orthonormal basis of the space $\mathcal{H}_n^d$ of spherical harmonics of degree n on $\mathbb{S}^{d-1}$, then the *addition formula*

$$(6.3) \qquad \sum_{j=1}^{\dim(\mathcal{H}_n^d)} Y_{n,j}(x) Y_{n,j}(y) = \frac{n + \lambda}{\lambda} C_n^\lambda(\langle x, y \rangle) \quad \text{for all} \quad x, y \in \mathbb{S}^{d-1},$$

together with Mercer's Theorem, Theorem 2.18, implies that for positive definite F the following series converges absolutely and uniformly

$$F(\langle x, y \rangle) = \sum_{n=0}^{\infty} \widehat{F}(n, \lambda) \sum_{k=1}^{\dim(\mathcal{H}_n^d)} Y_{n,k}(x) Y_{n,k}(y) = \sum_{n=0}^{\infty} \widehat{F}(n, \lambda) \frac{n + \lambda}{\lambda} C_n^\lambda(\langle x, y \rangle),$$

and hence the Gegenbauer expansion of a function $F \in C[-1, 1]$, which is positive definite on the sphere, must be absolutely and uniformly convergent. Absolute convergence of the Gegenbauer expansions of positive definite functions has been observed before [G, BD] using special properties of Gegenbauer polynomials – here we see that this fact is a consequence of general spectral theory. Part (5) of Theorem 4.2 and part (10) of Theorem 4.1 may be restated as follows:

 (iv) The kernel F is (conditionally) positive definite on the sphere if and only if the Gegenbauer coefficients satisfy $\widehat{F}(n, \lambda) \geq 0$ for all $n \geq 0$ ($n \geq 1$). Moreover, in this case, the Gegenbauer expansion (6.1) converges uniformly and absolutely.

Similarly, one obtains the characterization in terms of the "convolution square root", as in part (6) of Theorem 4.2:

 (v) positive definiteness of F on the sphere is equivalent to the existence of the function $f \in L^2([-1, 1], w_\lambda)$ such that

$$(6.4) \qquad F(\langle x, y \rangle) = \int_{\mathbb{S}^{d-1}} f(\langle x, z \rangle) f(\langle z, y \rangle) \, d\sigma(z), \quad x, y \in \mathbb{S}^{d-1}.$$

This equivalence had been stated in [BD] and the construction of the function f from the kernel F is almost identical to that of the function k from the kernel K in Proposition 2.20: one chooses the Gegenbauer coefficients of f so that $(\widehat{f}(n, \lambda))^2 = \widehat{F}(n, \lambda)$, which is mimicked in (2.17). In

addition, (6.4) was used to prove a general Stolarsky Principle on the sphere [BDM, BM]: for each $\nu \in \mathbb{P}(\mathbb{S}^{d-1})$,

$$(6.5) \qquad I_F(\nu) - I_F(\sigma) = D^2_{L^2,f,\sigma}(\nu),$$

which we have generalized to arbitrary compact domains in Theorem 5.1.

We finish by mentioning that our results similarly apply to the setting of two-point homogeneous compact spaces, both connected (e.g., projective spaces) and discrete (e.g., Hamming cube) with the corresponding uniform measure, as well as compact topological groups with the Haar measure, although we do not pursue these directions in this paper.

References

[A] N. Aronszajn. *Theory of Reproducing Kernels*. Trans. of the AMS, **68 (3)** (1950), 337–404. 2

[Ba] A. Barg. *Stolarsky's Invariance Principle for Finite Metric Spaces*. Mathematika, **67(1)** (2020), 1-23. 1.2, 5

[BS] A. Barg and M. Skriganov. *Bounds for Discrepancies in the Hamming Space*. Journal of Complexity (2021), available at https://doi.org/10.1016/j.jco.2021.101552. 5

[BC] J. Beck and W. Chen. *Irregularities of Distribution*. Cambridge Tracts in Mathematics, **89**, Cambridge University Press (1987). 1.2, 5

[BD] D. Bilyk and F. Dai. *Geodesic Distance Riesz Energy on the Sphere*. Transactions of the American Mathematical Society, **372** (2019), 3141-3166. 2, 2.6, 5, 6, 6

[BDM] D. Bilyk, F. Dai, and R. Matzke. *Stolarsky Principle and Energy Optimization on the Sphere*. Constructive Approximation, **48(1)** (2018), 31-60. 1.2, 2, 2.6, 3.3, 4, 5, 6, 6

[BFGMPV] D. Bilyk, D. Ferizović, A. Glazyrin, R. Matzke, J. Park, and O. Vlasiuk. *Potential Theory with Multivariate Kernels*. Available at https://arxiv.org/pdf/2104.03410. 2.3, 2.3, 3.5, 4

[BGMPV] D. Bilyk, A. Glazyrin, R. Matzke, J. Park, and O. Vlasiuk. *Energy on Spheres and Discreteness of Minimizing Measures*. Journal of Functional Analysis, **280 (11)**, (2021), 108995. 3.5, 4, 6

[BM] D. Bilyk and R. Matzke. *On the Fejes Tóth Problem on the Sum of Acute Angles*. Proceedings of American Mathematical Society, **147** (2019), 51-59. 5, 6

[Bj] G. Björck. *Distributions of Positive Mass*. Arkiv För Matematik, **3** (1956), 255-269. 2.4, 4

[BHS] S. Borodachov, D. Hardin, and E. Saff. *Discrete Energy of Rectifiable Sets*. Springer Monographs in Mathematics, Springer-Verlag New York (2019). 1.1, 2, 2.2, 2.4, 3.1, 3.3

[BrD] J.S. Brauchart and J. Dick. *A Simple Proof of Stolarsky's Invariance Principle*. Proceedings of the American Mathematical Society, **141** (2013), 2085-2096. 1.2, 5

[CFP] J. Carrillo, R.J. McCann, and C. Vallani. *Geometry of Minimizers for the Interaction Energy with Mildly Repulsive Potentials*. Annales de l'Institut Henri Poincaré C, Analyse Non Linéaire, **34** (2017), 1299-1308. 2, 2.7

[CSh] J. Carillo, R. Shu. *From radial symmetry to fractal behavior of aggregation equilibria for repulsive-attractive potentials*. Available at https://arxiv.org/pdf/2107.05079.pdf. 2.3, 2.3

[CSFSV] D. Castro-Silva, F. M. de O. Filho, L. Slot, and F. Vallentin (2021). *A recursive Lovász theta number for simplex-avoiding sets*. Available at https://arxiv.org/pdf/2106.09360. 3.1

[CMY] J. Cleary, S. A. Morris, and D. Yost. *Numerical Geometry-Numbers for Shapes*. The American Mathematical Monthly **93(4)** (1986), 260-275. 3, 3.1

[CS] J. H. Conway and N. J. A. Sloane. *Sphere packings, lattices, and groups*. Grundlehren der mathematischen Wissenschaften, Springer-Verlag New York (1999). 3.3

[DX] F. Dai and Y. Xu. *Approximation Theory and Harmonic Analysis on Spheres and Balls*. Springer Monographs in Mathematics, Springer, New York, NY (2013). 6

[DPZ] H. Dette, A. Pepelyshev, and A. Zhigljavsky. *Optimal Design for Linear Models with Correlated Observations*. The Annals of Statistics, **41(1)** (2013), 143-176. 2.3

[F] G.E. Fasshauer. *Positive Definite Kernels: Past, Present and Future*. Springer Monographs in Mathematics, Springer-Verlag New York (2019). 2

[FS] F. Finster and D. Schiefeneder. *On the Support of Minimizers of Causal Variational Principles*. Archive for Rational Mechanics and Analysis, **210(2)** (2013), 321-364. 2, 2.7

[Fr] O. Frostman. *Potentiel d'équilibre et capacité des ensembles*. Lund (1935). 3.1

[G] R. Gangolli. *Positive definite kernels on homogeneous spaces and certain stochastic processes related to Lévy's Brownian motion of several parameters*. Annales de l'Institut Henri Poincaré, Section B **3(2)** (1967), 121-226. 6

[HBZO] H. He, K. Basu, Q. Zhao, and A. Owen. *Permutation p-value Approximation via Generalized Stolarsky Invariance*. Annals of Statistics, **47(1)** (2019), 583-611. 5

[HNW] A. Hinrichs, P. Nickolas, and R. Wolf. *A Note on the Metric Geometry of the Unit Ball*. Mathematische Zeitschrift, **268** (2011), 887-896. 4

[KN] L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences*. Dover Publications (2006). 1.2, 5

[L] N. S. Landkoff. *Potential Theory*. Springer-Verlag (1972). 3.1

[Ma] J. Matoušek. *Geometric Discrepancy: An Illustrated Guide*. Algorithms and Combinatorics, **18**, Springer-Verlag, Berlin Heidelberg (1999). 1.2, 5

[Mec] E. Meckes. *The Random Matrix Theory of the Classical Compact Groups*. Cambridge University Press (2019). 2.3

[Me] J. Mercer. *Functions of Positive and Negative Type and their Connection with the Theory of Integral Equations*. Philosphical Transactions of the Royal Society A, **209(441-458)**, 415-446 (1909). 2

[P] M. Petrache. *Decorrelation as an avatar of convexity*. Available at http://arxiv.org/abs/1507.00782. 2.3, 2.3, 2.13

[PZ] L. Pronzato and A. Zhigljavsky. *Bayesian Quadrature and Energy Minimization for Space-filling Design*. SIAM/ASA Journal on Uncertainty Quantification, **8(3)** (2020), 959-1011. 2.3, 2.3

[Sch] I. Schoenberg. *Positive Definite Functions on Spheres*. Duke Mathematical Journal, **9(1)** (1942), 96–108. 3.3, 6, 6

[Sk1] M. Skriganov. *Point Distributions in Compact Metric Spaces*. Mathematika **63(3)** (2017), 1152–1171. 1.2, 5

[Sk2] M. Skriganov. *Stolarsky's Invariance Principle for Projective Spaces*. Journal of Complexity **56** (2020), 101428. 5

[St] K.B. Stolarsky. *Sums of Distances between Points on a Sphere. II*. Proceedings of American Mathematical Society, **41** (1973), 575–582. 1.2, 5

[ZDP] A. Zhigljavsky, H. Dette, and A. Pepelyshev. *A New Approach to Optimal Design for Linear Models with Correlated Observations*. Journal of the American Statistical Association, **105(491)** (2010), 1093-1103. 2.3

School of Mathematics, University of Minnesota, Minneapolis, MN 55408, USA.
Email address: `dbilyk@math.umn.edu`

School of Mathematics, University of Minnesota, Minneapolis, MN 55408, USA.
Email address: `matzk053@umn.edu`

Department of Mathematics, Florida State University, Tallahassee, FL 32306, USA.
Email address: `ovlasiuk@fsu.edu`