Generalization of the Energy Distance by Bernstein Functions

Jean Carlo Guella

Received: 31 August 2021 / Revised: 4 February 2022 / Accepted: 7 February 2022 / Published online: 4 March 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
We reprove the well-known fact that the energy distance defines a metric on the space of Borel probability measures on a Hilbert space with finite first moment by a new approach, by analyzing the behavior of the Gaussian kernel on Hilbert spaces and a maximum mean discrepancy analysis. From this new point of view, we are able to generalize the energy distance metric to a family of kernels related to Bernstein functions and conditionally negative definite kernels. We also explain what occurs on the energy distance on the kernel $\|x - y\|^a$ for every $a > 2$, by describing in which circumstances it defines a distance between probabilities. We also generalize this idea to a family of kernels related to completely monotone functions of finite order and conditionally negative definite kernels.

Keywords Energy distance · Metric spaces of strong negative type · Metrics on probabilities · Bernstein functions · Conditionally negative definite kernels

Mathematics Subject Classification (2020) 30L05 · 42A82 · 43A35

1 Introduction

A popular method to compare two probabilities is done by embedding the space (or a subset) of probabilities into a Hilbert space and use the metric provided by the embedding. Currently, there are two main approaches for this task:

(I) Maximum mean discrepancy: Let $K : X \times X \to \mathbb{R}$ be a bounded, continuous, positive definite kernel that is characteristic [1,2], the distance between two
Radon regular probabilities $P$ and $Q$ is defined by

$$MMD_K(P, Q) := \sqrt{\int_X \int_X K(x, y)(P - Q)(x)(P - Q)(y)}.$$

(II) Energy distance: Let $\gamma : X \times X \to \mathbb{R}$ be a continuous conditionally negative definite kernel, such that $\gamma(x, x) = 0$ for every $x \in X$ and must additionally satisfy the equality

$$\int_X \int_X -\gamma(x, y)(P - Q)(x)(P - Q)(y) = 0$$

for two Radon regular probabilities $P$ and $Q$ that integrates the function $x \to \gamma(x, z)$ for every $z \in X$ only when $P = Q$, [3]. It can be proved that the above double integral is always a nonnegative number and when this property occurs

$$E_\gamma(P, Q) := \sqrt{\int_X \int_X -\gamma(x, y)(P - Q)(x)(P - Q)(y)},$$

is a metric on the aforementioned subspace of probabilities on $X$.

In this paper, we focus on the second method. The most popular example of energy distance is when $X = \mathbb{R}^d$, $\gamma(x, y) = \|x - y\|^\theta$, where $0 < \theta < 2$, [4,5]. When $\theta = 2$, the kernel is conditionally negative definite but does not satisfy the additional property of Eq. 1.

A more geometrical approach is when $\gamma$ is a metric on $X$ that satisfies Eq. 1 (the topology is the one from the metric), and in this case, we say that the metric space $(X, \gamma)$ has strong negative type. Examples of such spaces include:

- **Euclidean spaces (finite-dimensional)** Proved in [4] using Fourier analysis.
- **Hilbert spaces (infinite-dimensional)** Proved in [6] as a generalization of [4], using probabilistic arguments.
- **Real hyperbolic spaces (finite-dimensional)** Proved in [7] using geometrical properties of this space.

In some cases, the conditionally negative definite kernel $\gamma$ may define a metric on the set $X$, but $\gamma$ is not of strong type. A metric space where we only know that the distance is a conditionally negative definite kernel is called a metric space of negative type. Examples of such spaces include:

- **Real hyperbolic spaces (finite-dimensional)** Proved in [7] using geometrical properties of this space.

In [6], it is also proved that if $(X, \gamma)$ is a metric space of negative type then $\gamma^\theta$, $0 < \theta < 1$ is a conditionally negative definite kernel that satisfies Eq. 1, with the topology of the metric $\gamma$. Interestingly, the kernel $\gamma^\theta$ is a metric on $X$, with the same topology as $\gamma$, so we can rephrase the result of Lyon as $(X, \gamma^\theta)$ being a metric space.
of strong negative type. We provide more details and generalizations of this property in Corollary 3.6.

The major aim of this paper is to reprove the fact that those metric spaces have strong negative type in a new and unified method and also provide a large amount of examples of conditionally negative definite kernels that satisfy Eq. 1 using Bernstein functions in Theorem 3.3. We also provide a proof that the real hyperbolic spaces (infinite-dimensional) and complex hyperbolic spaces (any dimension) are metric spaces of strong negative type in Theorems 3.4 and 3.5.

In [10], Mattner analyzed the behavior of the kernel $\|x - y\|^\theta$, for $\theta > 2$, defined in $\mathbb{R}^d$. What occurs is that we can still provide a metric structure on the space of probabilities with certain integrability assumptions using this kernel, but we can only compare them if they have the same vector mean ($2 < a < 4$), the same vector mean and the same covariance matrix ($4 < a < 6$), and so on. It also provided the same analysis for other radial kernels. We generalize its results in Theorems 4.4 and 4.5 to a broader setting using conditionally negative definite kernels.

The rest of the paper is separated into 5 sections. Section 2 is entirely focused on definitions and known results in the literature that we use. Section 3 is the main contribution of this paper and is focused on the probabilistic properties of applying a conditionally negative definite kernel to a Bernstein function. Section 4 is a higher-order generalization of the previous result, as its focus is on the probabilistic properties of applying a conditionally negative definite kernel to a completely monotone function of order $\ell$. In Sect. 5, we move to a functional analysis problem, by focusing on the behavior of the space of functions

$$y \in \mathcal{H} \rightarrow \int_\mathcal{H} \psi(\|x - y\|^2) d\mu(x) \in \mathbb{R}, \quad \|x\|^{2\ell} \in L^1(|\mu|)$$

where $\psi$ is a continuous function that is completely monotone of order $\ell$. More precisely, we analyze either when this space of functions is uniquely described by the measure $\mu$ or its relation with different completely monotone functions of order $k$. The proofs and some technical results are presented separately in Sect. 6.

2 Definitions

We recall that a nonnegative measure $\lambda$ on a Hausdorff space $X$ is Radon regular (which we simply refer to as Radon) when it is a Borel measure which is finite on every compact subset of $X$ and

(i) (Inner regular)$\lambda(E) = \sup\{\lambda(K) | K \text{ is compact, } K \subset E\}$ for every Borel set $E$.
(ii) (Outer regular)$\lambda(E) = \inf\{\lambda(U) | U \text{ is open, } E \subset U\}$ for every Borel set $E$.

We then say that a real-valued measure $\lambda$ of bounded variation is Radon if its variation is a Radon measure. The vector space of such measures is denoted by $\mathfrak{M}(X)$. Recall that every Borel measure of finite variation (in particular, probability measures) on a separable complete metric space is necessarily Radon.
A semi-inner product on a real vector space $V$ is a symmetric and bilinear real-valued function $(\cdot, \cdot)_V$ defined on $V \times V$ such that $(u, u)_V \geq 0$ for every $u \in V$. When this inequality is an equality only for $u = 0$, we say that $(\cdot, \cdot)_V$ is an inner product. Similarly, a pseudometric on a set $X$ is a symmetric function $d : X \times X \rightarrow [0, \infty)$, such that $d(x, x) = 0$ that satisfies the triangle inequality. If $d(x, y) = 0$ only when $x = y$, $d$ is a metric on $X$.

A kernel $K : X \times X \rightarrow \mathbb{R}$ is called positive definite (PD) if it is symmetric and for every finite quantity of distinct points $x_1, \ldots, x_n \in X$ and scalars $c_1, \ldots, c_n \in \mathbb{R}$, we have that

$$\int_X \int_X K(x, y)d\lambda(x)d\lambda(y) = \sum_{i, j=1}^n c_i c_j K(x_i, x_j) \geq 0,$$

where $\lambda = \sum_{i=1}^n c_i \delta_{x_i}$. The set of measures on $X$ used before are denoted by the symbol $\mathcal{M}_\delta(X)$.

The reproducing kernel Hilbert space (RKHS) of a positive definite kernel $K : X \times X \rightarrow \mathbb{R}$ is the Hilbert space $\mathcal{H}_K \subset \mathcal{F}(X, \mathbb{R})$, and it satisfies [11]

(i) $x \in X \rightarrow K_y(x) := K(x, y) \in \mathcal{H}_K$;
(ii) $\langle K_x, K_y \rangle = K(x, y)$
(iii) $\text{span}\{K_y, \ y \in X\} = \mathcal{H}_K$.

When $X$ is a Hausdorff space and $K$ is continuous, it holds that $\mathcal{H}_K \subset C(X)$.

The following widely known result, usually denoted as kernel mean embedding, describes how it is possible to define a semi-inner product structure on a subspace of $\mathcal{M}(X)$ using a continuous positive definite kernel.

**Lemma 2.1** If $K : X \times X \rightarrow \mathbb{R}$ is a continuous positive definite kernel and $\mu \in \mathcal{M}(X)$ with $\sqrt{K}(x, x) \in L^1(\mu)$ ($\mu \in \mathcal{M}(\sqrt{K}(X))$, then

$$z \in X \rightarrow K_{\mu}(z) := \int_X K(x, z) d\mu(x) \in \mathbb{R}$$

is an element of $\mathcal{H}_K$, and if $\eta$ is another measure with the same conditions as $\mu$, we have that

$$\langle K_{\eta}, K_{\mu} \rangle_{\mathcal{H}_K} = \int_X \int_X K(x, y) d\eta(x) d\mu(y).$$

In particular, $(\eta, \mu) \in \mathcal{M}(\sqrt{K}(X)) \times \mathcal{M}(\sqrt{K}(X)) \rightarrow \langle K_{\eta}, K_{\mu} \rangle_{\mathcal{H}_K}$ is a semi-inner product.

Usually, the kernel $K$ is bounded, so $\mathcal{M}(\sqrt{K}(X)) = \mathcal{M}(X)$. In this case, if the semi-inner product is in fact an inner product we say that $K$ is integrally strictly positive definite (ISPD), and when it is an inner product on the vector space of measures in $\mathcal{M}(X)$ with $\mu(X) = 0$, we say that $K$ is characteristic.
When the kernel is characteristic, we define the maximum mean discrepancy (MMD) as the metric on the space of probability measures in $\mathfrak{M}(X)$ by

$$\text{MMD}(P, Q)_K := \sqrt{\langle K_P - K_Q, K_P - K_Q \rangle_{\mathcal{H}_K}}$$

$$= \sqrt{\int_X \int_X K(x, y)d[P - Q](x)d[P - Q](y)} \quad (2)$$

As mentioned in Introduction, the focus of this paper is to analyze metrics on the space of probabilities using conditionally negative definite kernels. We present a more general definition which will be useful to the analysis of the energy distance through the kernel $\|x - y\|^a$, $a > 2$, defined on a Hilbert space.

**Definition 2.2** Let $\gamma : X \times X \to \mathbb{R}$ be a symmetric kernel and $\tau : X \times X \to \mathbb{R}$ be a positive definite kernel. We say that $\gamma$ is $\tau$-conditionally positive definite ($\tau$-CPD) if for every finite quantity of points $x_1, \ldots, x_n \in X$ and scalars $c_1, \ldots, c_n \in \mathbb{R}$, under the restriction that

$$\sum_{i, j=1}^n c_i c_j \tau(x_i, x_j) = 0,$$  

we have that

$$\sum_{i, j=1}^n c_i c_j\gamma(x_i, x_j) \geq 0.$$  

This definition generalizes the concepts of positive definite kernels ($\tau$ is the zero kernel) and conditionally positive definite (CPD) kernels ($\tau$ is the constant 1 kernel) and the more general approach of Chapter 10.3 in [12]. Sometimes it might be more convenient to work with the opposite sign in Definition 2.2; in this case, we say that the kernel is $\tau$-conditionally negative definite ($\tau$-CND). In the special case that $\tau$ is a strictly positive definite kernel, we define that the only $\tau$-CPD (or $\tau$-CND) kernel is the constant zero kernel; we do so to avoid unnecessary counterexamples.

The concept of CND kernels is intrinsically related to PD kernels, as a symmetric kernel $\gamma : X \times X \to \mathbb{R}$ is CND if and only if for any (or equivalently, for every) $w \in X$ the kernel

$$K^w_{\gamma}(x, y) := \gamma(x, w) + \gamma(w, y) - \gamma(x, y) - \gamma(w, w) \quad (3)$$

is positive definite. Due to the setting we are working in, sometimes it is convenient to omit the term $w$, and we use the notation $K_{\gamma}$, as the term $w$ does not influence the results we propose. Another famous relation is that a symmetric kernel $\gamma : X \times X \to \mathbb{R}$ is CND if and only if for every $r > 0$ the kernel

$$(x, y) \in X \times X \to e^{-r\gamma(x, y)} \quad (4)$$
is PD. Those two classic characterizations are crucial for the development of the subject and can be found in Chapter 3 at [13]. The characterization of the continuous CND radial kernels in all Euclidean spaces was proved in [14] and is the following:

**Theorem 2.3** Let \( \psi : [0, \infty) \to \mathbb{R} \) be a continuous function. The following conditions are equivalent

(i) The kernel

\[
(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \to \psi(\|x - y\|^2) \in \mathbb{R}
\]

is CND for every \( d \in \mathbb{N} \).

(ii) The function \( \psi \) can be represented as

\[
\psi(t) = \psi(0) + at + \int_{(0, \infty)} (1 - e^{-rt}) \frac{1 + r}{r} d\sigma(r),
\]

for all \( t \geq 0 \), where \( a \) is a nonnegative real number; \( \sigma \) is a nonnegative measure in \( \mathcal{M}((0, \infty)) \). The number \( a \) and the finite measure \( \sigma \) are unique.

(iii) The function \( \psi \in C^\infty((0, \infty)) \) and \( \psi^{(1)} \) is completely monotone, that is, \((-1)^n \psi^{(n+1)}(t) \geq 0\), for every \( n \in \mathbb{Z}_+ \) and \( t > 0 \).

A continuous function \( \psi : [0, \infty) \to \mathbb{R} \) that satisfies the relation (iii) in Theorem 2.3 is called a Bernstein function (we do not need to assume that Bernstein functions are nonnegative), and the same theorem provides a representation for it. For more information on Bernstein functions, see [15].

Another important example of Definition 2.2 is when \( X \) is a finite-dimensional Euclidean space and \( \tau(x, y) = \sum_{j=0}^{\ell-1} \langle x, y \rangle^j \), which is equivalent to the definition of a conditionally positive semidefinite kernel of order \( \ell \) given in [12,16,17].

In [16,18], a generalization of Theorem 2.3 to certain \( \tau \)-CPD kernels is given as follows:

**Theorem 2.4** The following conditions are equivalent for a continuous function \( \psi : [0, \infty) \to \mathbb{R} \)

(i) The kernel

\[
(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \to \psi(\|x - y\|) \in \mathbb{R}
\]

is \( \tau \)-CPD for every \( d \in \mathbb{N} \) and \( \tau(x, y) = \sum_{j=0}^{\ell-1} \langle x, y \rangle^j \).

(ii) The function \( \psi \) can be represented as

\[
\psi(t) = \int_{(0, \infty)} (e^{-tr} - e_\ell(r) e_\ell(rt)) \frac{1 + \ell}{r^\ell} d\sigma(r) + \sum_{k=0}^{\ell} a_k t^k
\]

where \( \sigma \) is a nonnegative measure on \( \mathcal{M}((0, \infty)) \), with

\[
\omega_\ell(s) := \sum_{l=0}^{\ell-1} (-1)^l \frac{s^l}{l!}, \quad e_\ell(s) := e^{-s} \sum_{l=0}^{\ell-1} \frac{s^l}{l!}
\]
and \( a_k \in \mathbb{R}, (-1)^\ell a_\ell \geq 0 \). The representation is unique.

(iii) The function \( \psi \in C^\infty((0, \infty)) \) and the function \((-1)^\ell \psi^{(\ell)}(t)\) are completely monotone, that is, \((-1)^{n+\ell} \psi^{(n+\ell)}(t) \geq 0\), for every \( n \in \mathbb{Z}_+ \) and \( t > 0 \).

A continuous function \( \psi : [0, \infty) \to \mathbb{R} \) that satisfies the relation \((iii)\) in Theorem 2.4 is called a completely monotone function of order \( \ell \), and we use the notation \( \psi \in CM_\ell \). For instance, the functions

- \((-1)^\ell t^a;\)
- \((-1)^\ell t^{\ell-1} \log(t), \) with the restriction that \( \ell \geq 2; \)
- \((-1)^\ell (c + t)^a;\)

are elements of \( CM_\ell \), for \( \ell - 1 < a \leq \ell , c > 0 \).

Those examples are not only in \( CM_\ell \), but they are \( \ell - 1 \) continuously differentiable on \([0, \infty)\). In Theorem 4.2, we prove a simpler expression for those functions compared to the one in Theorem 2.4. This new representation is important for the objectives of Sect. 4, as it implies better integrability conditions compared to the general case proved in Theorem 4.5.

### 3 Energy Distance and Bernstein Functions

In [19], it was proved the following result:

**Theorem 3.1** Let \( \gamma : X \times X \to \mathbb{R} \) be a continuous CND kernel. Then, the kernel

\[
(x, y) \in X \times X \to e^{-\gamma(x, y)}
\]

is ISPD if and only if \( 2\gamma(x, y) = \gamma(x, x) + \gamma(y, y) \) only when \( x = y \).

A CND kernel that satisfies the condition of Theorem 3.1 is called metrizable, as the condition in the theorem is equivalent to demanding that \( D_\gamma(x, y) := \sqrt{2\gamma(x, y) - \gamma(x, x) - \gamma(y, y)} \) is a metric.

We can reobtain the fact that Hilbert spaces are metric spaces of strong negative type proved in [6] using Theorem 3.1. For that, we describe which probabilities in \( \mathcal{M}(X) \) can be analyzed by a continuous CND kernel. The result is proved on a broader setting, as the arguments for this general version are similar to the case \( \theta = 1 \), and we use this version in Sect. 4.

**Lemma 3.2** Let \( \gamma : X \times X \to \mathbb{R} \) be a continuous CND kernel such that \( \gamma(x, x) \) is a bounded function, \( \mu \in \mathcal{M}(X) \) and \( \theta > 0 \). Then, the following assertions are equivalent

- \( \gamma \in L^\theta(|\mu| \times |\mu|); \)
- \( \text{The function } x \in X \to \gamma(x, z) \in L^\theta(|\mu|) \text{ for some } z \in X; \)
- \( \text{The function } x \in X \to \gamma(x, z) \in L^\theta(|\mu|) \text{ for every } z \in X. \)

The set of measures that satisfies these relations is a vector space.
The function \( t \in [0, \infty) \to \psi(t) := \sqrt{t} \in \mathbb{R} \) is an example of a Bernstein function, and the representation of Theorem 2.3 for it is

\[-\sqrt{t} = \frac{1}{2\sqrt{\pi}} \int_{(0,\infty)} (e^{-rt} - 1) \frac{1}{r^{3/2}} dr.
\]

So,

\[(x, y) \in \mathcal{H} \times \mathcal{H} \to -\|x - y\|_\mathcal{H} = \frac{1}{2\sqrt{\pi}} \int_{(0,\infty)} (e^{-r\|x-y\|^2} - 1) \frac{1}{r^{3/2}} dr.
\]

By Fubini–Tonelli theorem, we have that if \( \mu \in \mathcal{M}(\mathcal{H}) \) with \( \mu(\mathcal{H}) = 0 \) and \( \|x\| \in L^1(\mu) \), then

\[
\int_{\mathcal{H}} \int_{\mathcal{H}} (1)\|x - y\|_\mathcal{H} d\mu(x) d\mu(y)
= \frac{1}{2\sqrt{\pi}} \int_{(0,\infty)} \left( \int_{\mathcal{H}} \int_{\mathcal{H}} e^{-r\|x-y\|^2} d\mu(x) d\mu(y) \right) \frac{1}{r^{3/2}} dr \geq 0.
\]

Further, because of Theorem 3.1, the double inner integral is positive whenever \( \mu \) is not the zero measure, implying that the final result is a positive number. Thus, we reobtained the fact that Hilbert spaces are metric spaces of strong negative type by a different argument. More generally, we have the following result.

**Theorem 3.3** Let \( \psi : [0, \infty) \to \mathbb{R} \) be a Bernstein function and \( \gamma : X \times X \to [0, \infty) \) be a continuous CND kernel such that \( x \to \gamma(x, x) \) is a bounded function. Consider the vector space

\[
\mathcal{M}_1(X; \gamma, \psi) := \{ \eta \in \mathcal{M}(X), \quad \psi(\gamma(x, y)) \in L^1(|\eta| \times |\eta|) \text{ and } \eta(X) = 0 \},
\]

then the function

\[(\mu, \nu) \in \mathcal{M}_1(X; \gamma, \psi) \times \mathcal{M}_1(X; \gamma, \psi) \to I(\mu, \nu)_{\gamma, \psi} := -\int_X \int_X \psi(\gamma(x, y)) d\mu(x) d\nu(y)
\]

defines a semi-inner product on \( \mathcal{M}_1(X; \gamma, \psi) \). If \( \psi \) is not a linear function and \( \gamma \) is metrizable, then \( I(\mu, \nu)_{\gamma, \psi} \) defines an inner product on \( \mathcal{M}_1(X; \gamma, \psi) \).

It is relevant to say that usually the inner product in Theorem 3.3 is not complete (hence, \( \mathcal{M}_1(X; \gamma, \psi) \) is not a Hilbert space). For instance, in [20] it is proved that the Gaussian kernel can be used to define an inner product on the space of tempered distributions on Euclidean spaces.

We prefer to state Theorem 3.3 in terms of inner products on vector spaces of measures, because this is often a neglected aspect of energy distance. We obtain the standard definition of energy distance if we take probabilities \( P, Q \in \mathcal{M}(X) \) that
satisfies any of the three relations in Lemma 3.2 for the CND kernel \( \psi(\gamma(x, y)) \), then

\[
E_{\gamma, \psi}(P, Q) := \sqrt{-\int_X \int_X \psi(\gamma(x, y))d[P - Q](x)d[P - Q](y)}
\]

\[
= \sqrt{I(P - Q, P - Q)_{\gamma, \psi}}
\]

is a metric in this space of probabilities. Now we present several consequences of Theorem 3.3.

For instance, if \( X \) is a real Hilbert space \( \mathcal{H} \), \( \gamma(x, y) = \|x - y\|^2 \) and \( \psi(t) = ta/2 \), \( 0 < a < 2 \), then

\[
(\mu, \nu) \in M_1(\mathcal{H}; \frac{t}{a}/2) \times M_1(\mathcal{H}; \frac{t}{a}/2) \rightarrow I(\mu, \nu)_{\frac{a}{2}} := -\int_\mathcal{H} \int_\mathcal{H} \|x - y\|^ad\mu(x)d\nu(y)
\]

defines an inner product on

\[
M_1(\mathcal{H}; \frac{t}{a}/2) := \{\eta \in M(\mathcal{H}), \|x\|^a \in L^1(|\eta|) \text{ and } \eta(X) = 0\}.
\]

Another example occurs on the generalized real hyperbolic space. Let \( \mathcal{H} \) be a real Hilbert space and define \( \mathbb{H} := \{(x, z) \in \mathcal{H} \times (0, \infty), \quad z^2 - \|x\|^2 = 1\} \) as the real hyperbolic space relative to \( \mathcal{H} \) and consider the kernel

\[
((x, z), (y, w)) \in \mathbb{H} \times \mathbb{H} \rightarrow [(x, z), (y, w)] := zw - \langle x, y \rangle \in [1, \infty),
\]

which satisfies the relation

\[
\cosh(d_{\mathbb{H}}((x, z), (y, w))) = [(x, z), (y, w)],
\]

where \( d_{\mathbb{H}} \) is a metric in \( \mathbb{H} \). In [21] or chapter 5 in [13], it is proved that the metric \( d_{\mathbb{H}} \) on \( \mathbb{H} \) is a CND kernel. A proof that when \( \mathbb{H} \) is finite-dimensional the metric space \( (\mathbb{H}, d_{\mathbb{H}}) \) has strong negative type was provided in [7] using geometric properties of hyperbolic spaces. Our proof of the next theorem relies on a Laurent type of approximation for the function \( \text{arccosh} \).

**Theorem 3.4** Let \( \mathbb{H} \) be a real hyperbolic space, and consider the vector space

\[
M_1(\mathbb{H}; t) := \{\eta \in M(\mathbb{H}), \quad x \in \mathbb{H} \rightarrow d_{\mathbb{H}}(x, z) \in L^1(|\eta|)
\]

for some (or every) \( z \in \mathbb{H} \) and \( \eta(\mathbb{H}) = 0 \}.

Then

\[
(\mu, \nu) \in M_1(\mathbb{H}; t) \times M_1(\mathbb{H}; t) \rightarrow H(\mu, \nu)_1 := -\int_{\mathbb{H}} \int_{\mathbb{H}} d_{\mathbb{H}}(x, y)d\mu(x)d\nu(y)
\]

is an inner product.
Interestingly, by a similar method we can obtain these properties for the generalized complex hyperbolic space. Let \( \mathcal{H}_C \) be a complex Hilbert space and define \( \mathbb{H}_C := \{(x, z) \in \mathcal{H}_C \times \mathbb{C}, \quad |z|^2 - \|x\|^2 = 1\} \) as the complex hyperbolic space relative to \( \mathcal{H}_C \) and consider the kernel

\[
((x, z), (y, w)) \in \mathbb{H}_C \times \mathbb{H}_C \rightarrow [(x, z), (y, w)] := z\overline{w} - \langle x, y \rangle,
\]

which satisfies the relation

\[
\cosh(d_{\mathbb{H}_C}((x, z), (y, w))) = |[(x, z), (y, w)]|,
\]

where \( d_{\mathbb{H}_C} \) is a metric in \( \mathbb{H}_C \). In [21], it is proved that the metric \( d_{\mathbb{H}_C} \) is a CND kernel.

**Theorem 3.5** Let \( \mathbb{H}_C \) be a complex hyperbolic space, and consider the vector space

\[
\mathcal{M}_1(\mathbb{H}_C; \iota) := \{\eta \in \mathcal{M}(\mathbb{H}_C), \quad x \in \mathbb{H}_C \rightarrow d_{\mathbb{H}_C}(x, z) \in L^1(|\eta|)
\]

for some (or every) \( z \in \mathbb{H}_C \) and \( \eta(\mathbb{H}_C) = 0 \).

Then

\[
(\mu, \nu) \in \mathcal{M}_1(\mathbb{H}_C; \iota) \times \mathcal{M}_1(\mathbb{H}_C; \iota) \rightarrow H(\mu, \nu)_1 := -\int_{\mathbb{H}_C} \int_{\mathbb{H}_C} d_{\mathbb{H}_C}(x, y)d\mu(x)d\nu(y)
\]

is an inner product.

A different behavior occurs on the generalized real spheres. Let \( \mathcal{H} \) be a Hilbert space and define \( S^\mathcal{H} := \{x \in \mathcal{H}, \quad \|x\| = 1\} \) be the real sphere relative to \( \mathcal{H} \). The kernel \( d_{S^\mathcal{H}} \) defined on \( S^\mathcal{H} \) by the relation

\[
\cos(d_{S^\mathcal{H}}(x, y)) = \langle x, y \rangle^\mathcal{H}, \quad x, y \in \mathcal{H}
\]

is a metric and defines a CND kernel as shown in [8]. However, unlike the Hilbert space and the real hyperbolic space, \( d_{S^\mathcal{H}} \) is not a metric space of strong negative type, [22]. Gangolli also proved in [8] that the metric on the other compact two-point homogeneous spaces (real/complex/quaternionic projective spaces and the Cayley projective plane) does not define a CND kernel.

The following corollary of Theorem 3.3 connects the setting of metric spaces of strong negative type and the kernels in Theorem 3.3.

**Corollary 3.6** Let \( \psi : [0, \infty) \rightarrow \mathbb{R} \) be a nonzero Bernstein function such that \( \psi(0) = 0, \lim_{t \to \infty} \psi(t)/t = 0 \) and \( (X, \gamma) \) is a metric space of negative type. Then,

\[
(x, y) \in X \times X \rightarrow D_{\psi, \gamma}(x, y) := \psi(\gamma(x, y))
\]

is a metric on \( X \) and \( (X, D_{\psi, \gamma}) \) is a metric space of strong negative type homomorphic to \( (X, \gamma) \).
As an example of Corollary 3.6 consider the Bernstein function 
\[ \psi(t) = \log(t + 1), \]
sO if \( \mathcal{H} \) is a Hilbert space \( \log(\| x - y \| + 1) \) is a metric on \( \mathcal{H} \) that is homomorphic with the Hilbertian topology and this metric is of strong negative type. Interestingly we can apply Corollary 3.6 again in order to obtain that the same occurs with the metric \( \log(\log(\| x - y \| + 1) + 1) \). The condition \( \lim_{t \to \infty} \psi(t)/t = 0 \) is equivalent to \( a = 0 \) in the representation of \( \psi \) in Theorem 2.3 by Remark 3.3.(iv) in [15].

The variety of examples of inner products obtained from Theorem 3.3 are not only important for providing metrics in a subspace of the space of probabilities, but can also be used to create independence tests. For instance, due to the results in [23,24], if \( \gamma_1, \gamma_2, \gamma_3 \) are CND kernels obtained from Theorem 3.3 and that defines inner products, then for probabilities \( P, Q \in \mathcal{M}(X_1 \times X_2 \times X_3) \) that satisfies reasonable integrability conditions and have the same marginals

\[
\int_{X_1 \times X_2 \times X_3} \int_{X_1 \times X_2 \times X_3} [\gamma_1 \gamma_2 \gamma_3 - \gamma_1 \gamma_2 - \gamma_1 \gamma_3 - \gamma_2 \gamma_3]d[P - Q]d[P - Q] = 0
\]

if and only if \( P = Q \). This generalizes the concept of distance covariance [6] and can be generalized to any number of variables.

4 Higher-Order Energy Distance and Completely Monotone Functions of Order \( \ell \)

Returning to the kernel \( (x, y) \in \mathcal{H} \times \mathcal{H} \to \| x - y \|^a \), we may ask ourselves what occurs when \( a \geq 2 \). The case \( a = 2 \) is simpler, because

\[
- \int_{\mathcal{H}} \int_{\mathcal{H}} \| x - y \|^2d\mu(x)d\nu(y) = 2 \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle d\mu(x)d\nu(y),
\]

for every \( \mu, \nu \in \mathcal{M}_1(\mathcal{H}; t) := \{ \eta \in \mathcal{M}(\mathcal{H}), \| x \|^2 \in L^1(\| \eta \|) \text{ and } \eta(X) = 0 \} \). This still defines a semi-inner product on \( \mathcal{M}_1(\mathcal{H}; t) \), but the infinite-dimensional vector space

\[
\mathcal{M}_2(\mathcal{H}; t) := \{ \eta \in \mathcal{M}_1(\mathcal{H}; t), \int_{\mathcal{H}} \langle x, y \rangle d\eta(x) = 0, \text{ for every } y \in \mathcal{H} \} \subset \mathcal{M}_1(\mathcal{H}; t)
\]

is equivalent to the zero measure in this semi-inner product space.

For an arbitrary measure \( \eta \in \mathcal{M}(\mathcal{H}) \) such that \( \| x \| \in L^1(\| \eta \|) \), the linear functional

\[
y \in \mathcal{H} \to \int_{\mathcal{H}} \langle x, y \rangle d\eta(x) \in \mathbb{R}
\]

is continuous, so there exists a vector \( v_\eta \), which we define as the vector mean of \( \eta \), that represents the above continuous linear functional.

In the case \( a > 2 \), a different behavior emerges. The double integral kernel does not define a semi-inner product on \( \mathcal{M}_1(\mathcal{H}, t^{a/2}) \); however, if we restrict ourselves to
the vector space

\[ \mathcal{M}_2(\mathcal{H}; t^{a/2}) := \{ \eta \in \mathcal{M}(\mathcal{H}), \| x \|^a \in L^1(\| \eta \|), \eta(\mathcal{H}) = 0, v_\eta = 0 \} \]

for \( 2 < a < 4 \) and use the representation for the \( CM_2 \) function

\[ t^{a/2} = \frac{a(a - 2)}{4\Gamma(2 - a/2)} \int_{(0, \infty)} (e^{-rt} - 1 + rt) \frac{1}{r^{a/2 + 1}} dr, \]

by Fubini–Tonelli we obtain that if \( \mu, \nu \in \mathcal{M}_2(\mathcal{H}; t^{a/2}) \)

\[ \int_{\mathcal{H}} \int_{\mathcal{H}} \| x - y \|^a d\mu(x) d\nu(y) = \frac{a(a - 2)}{4\Gamma(2 - a/2)} \int_{(0, \infty)} \left( \int_{\mathcal{H}} \int_{\mathcal{H}} e^{-r\| x - y \|^2} d\mu(x) d\nu(y) \right) \frac{1}{r^{a + 1}} dr \geq 0. \]

In particular, we can use the kernel \( \| x - y \|^a, 2 < a < 4 \), in order to define a metric on the space of Radon probability measures on \( \mathcal{H} \) with finite \( a \)-moment, but with a fixed vector mean. The representation we used for the \( CM_2 \) function \( t^{a/2} \) is different from the one in Theorem 2.4. This occurs because we additionally have that \( t^{a/2} \in C^1([0, \infty)) \), as described by the next lemma.

**Lemma 4.1** A function \( \psi \in CM_\ell \cap C^{\ell - 1}([0, \infty)), \ell \geq 2, \) if and only if it can be represented as

\[ \psi(t) = \int_{(0, \infty)} (e^{-rt} - \omega_\ell(rt)) \frac{1+r}{r^{\ell}} d\sigma(r) + \sum_{k=0}^{\ell} a_k t^k \]

where \( a_j = \psi^{(j)}(0), (-1)^{\ell} a_\ell \geq 0 \) and \( \sigma \) is a nonnegative measure in \( \mathcal{M}((0, \infty)) \).

The representation is unique.

A generalization of Lemma 3.2 to the functions appearing in Lemma 4.1 is possible.

**Lemma 4.2** Let \( \gamma : X \times X \to [0, \infty) \) be a continuous CND kernel such that \( \gamma(x, x) \) is a bounded function, \( \mu \in \mathcal{M}(X) \) and \( \psi \in CM_\ell \cap C^{\ell - 1}([0, \infty)) \). Then, the following assertions are equivalent

(i) \( \psi(\gamma) \in L^1(\| \mu \| \times \| \mu \|); \)

(ii) The function \( x \in X \to \psi(\gamma(x, z)) \in L^1(\| \mu \|) \) for some \( z \in X; \)

(iii) The function \( x \in X \to \psi(\gamma(x, z)) \in L^1(\| \mu \|) \) for every \( z \in X. \)

The set of measures that satisfy these relations is a vector space.

Taking into account the example at the beginning of this section, in order to obtain a higher-order generalization of Theorem 3.3 using a CND kernel \( \gamma \), we must have a substitute for the vector mean. For that, note that from the kernel mean embedding, \( v_\eta = 0 \) if and only if \( \int_{\mathcal{H}} \int_{\mathcal{H}} (x, y) d\eta(x) d\eta(y) = 0. \) If we choose \( w = 0 \in \mathcal{H} \), the
kernel $2\langle x, y \rangle$ is equal to the PD kernel $K^0_\gamma$ of Eq. (3), for $\gamma(x, y) = \|x - y\|^2$. Although we have fixed $w = 0$, it is not difficult to prove that if $\eta(H) = 0$, then the following equivalence holds

$$\int_H \int_H \langle x, y \rangle d\eta(x) d\eta(y) = 0 \iff \int_H \int_H \langle x - w, y - w \rangle d\eta(x) d\eta(y) = 0$$

for any $w \in H$.

**Lemma 4.3** Let $\gamma : X \times X \to [0, \infty)$ be a continuous CND kernel such that $x \to \gamma(x, x)$ is a constant function. Then for $n \in \mathbb{N}$ and $\mu \in \mathcal{M}(X)$ such that $\gamma^n \in L^1(|\mu| \times |\mu|)$, the kernel $K^w_\gamma$ defined in Eq. (3) satisfies $K^w_\gamma \in L^m(|\mu| \times |\mu|)$, $0 \leq m \leq n$ for every $w \in X$. Also

$$\int_X \int_X [K^w_\gamma(x, y)]^m d\mu(x) d\mu(y) = 0, \quad 0 \leq m \leq n - 1,$$

for some $w \in X$ if and only if this holds for every $w \in X$, and these $n$ equalities imply that

$$(−1)^n \int_X \int_X \gamma(x, y)^n d\mu(x) d\mu(y) \geq 0$$

and

$$\int_X \int_X \gamma(x, y)^m d\mu(x) d\mu(y) = 0, \quad 0 \leq m \leq n - 1.$$

It is not clear yet whether the condition that $x \to \gamma(x, x)$ is constant is a necessary condition for the results in the previous lemma. The result and the proof is based in Lemma 3.1 in [18], but in this reference they were interested in the case $X = \mathbb{R}^d$, $\gamma(x, y) = \|x - y\|^2$ and $\mu \in \mathcal{M}_b(\mathbb{R}^d)$.

Gathering all the previous results, we are able to obtain our first higher-order generalization of Theorem 3.3.

**Theorem 4.4** Let $\ell \in \mathbb{N}$, $\psi : [0, \infty) \to \mathbb{R}$ be a continuous function on $CM_\ell \cap C^{\ell-1}([0, \infty))$ and $\gamma : X \times X \to [0, \infty)$ be a continuous CND kernel such that $x \to \gamma(x, x)$ is a constant function. Consider the vector space

$$\mathcal{M}_\ell(X; \gamma, \psi) := \{ \eta \in \mathcal{M}(X), \psi(\gamma(x, y)) \in L^1(|\eta| \times |\eta|), \eta(X) = 0, \text{ and } \int_X \int_X K_\gamma(x, y)^j d\eta(x) d\eta(y) = 0, \quad 1 \leq j \leq \ell - 1 \}$$

where $K_\gamma$ is the kernel in Eq. (3), then the function

$$(\mu, \nu) \in \mathcal{M}_\ell(X; \gamma, \psi) \times \mathcal{M}_\ell(X; \gamma, \psi) \to I(\mu, \nu)_{\gamma, \psi} := \int_X \int_X \psi(\gamma(x, y)) d\mu(x) d\nu(y)$$
defines a semi-inner product on $\mathcal{M}_\ell(X; \gamma, \psi)$. If $\psi$ is not a polynomial of degree $\ell$ or less and $\gamma$ is metrizable, then $I(\mu, v)_{\gamma, \psi}$ defines an inner product on $\mathcal{M}_\ell(X; \gamma, \psi)$.

As an example of Theorem 4.4, if $X$ is a Hilbert space $\mathcal{H}$, $\gamma(x, y) = \|x - y\|^2$ and $\psi(t) = (-1)^{\ell}t^{a/2}$, $2(\ell - 1) < a < 2\ell$, $\ell \in \mathbb{N}$ then

$$(\mu, v) \in \mathcal{M}_\ell(\mathcal{H}; t^{a/2}) \times \mathcal{M}_\ell(\mathcal{H}; t^{a/2}) \rightarrow I(\mu, v)_{a/2} := \int_{\mathcal{H}} \int_{\mathcal{H}} (-1)^{\ell} \|x - y\|^a d\mu(x) d\nu(y)$$

defines an inner product on the vector space

$$\mathcal{M}_\ell(\mathcal{H}; t^{a/2}) := \{\mu \in \mathcal{M}(\mathcal{H}), \|x\|^a \in L^1(\|\mu\|), \mu(\mathcal{H}) = 0, \text{ and } \int_\mathcal{H} \int_\mathcal{H} (x, y_1) \ldots (x, y_j) d\mu(x) = 0, y_1, \ldots, y_j \in \mathcal{H} \ 1 \leq j \leq \ell - 1\}.$$

Theorems 3.3 and 4.4 in the case where $X$ is an Euclidean space $\mathbb{R}^d$ and $\gamma(x, y) = \|x - y\|^2$ were partially proved in [10].

In the second higher-order version of Theorem 3.3, we do not assume that $\gamma \in C^{\ell-1}([0, \infty))$. Since we do not have a version of Lemma 4.2 in this general setting, we must use a different vector space. By Lemma 2.4 in [16], a function $\gamma \in CM_\ell$ satisfies $|\gamma(t)| \leq C(1 + t^\ell)$ for some $C > 0$ (we reprove this result by a more direct approach in Lemma 6.3). So, we can demand integrability of the kernel $\gamma(x, y)^{\ell}$.

**Theorem 4.5** Let $\ell \in \mathbb{N}$, $\psi : [0, \infty) \rightarrow \mathbb{R}$ be a continuous function on $CM_\ell$ and $\gamma : X \times X \rightarrow [0, \infty)$ be a continuous CND kernel such that $x \rightarrow \gamma(x, x)$ is a constant function. Consider the vector space

$$\mathcal{M}_\ell(X; \gamma, t^\ell) := \{\eta \in \mathcal{M}(X), \gamma(x, y)^{\ell} \in L^1(\|\eta\| \times \|\eta\|), \eta(X) = 0, \text{ and } \int_X \int_X K_\gamma(x, y)^{\ell} d\eta(x) d\eta(y) = 0, 1 \leq j \leq \ell - 1\}$$

where $K_\gamma$ is the kernel in Eq. (3), then the function

$$(\mu, v) \in \mathcal{M}_\ell(X; \gamma, t^\ell) \times \mathcal{M}_\ell(X; \gamma, t^\ell) \rightarrow I(\mu, v)_{\gamma, \psi} := \int_X \int_X \psi(\gamma(x, y)) d\mu(x) d\nu(y)$$

defines a semi-inner product on $\mathcal{M}_\ell(X; \gamma, t^\ell)$. If $\psi$ is not a polynomial of degree $\ell$ or less and $\gamma$ is metrizable, then $I(\mu, v)_{\gamma, \psi}$ defines an inner product on $\mathcal{M}_\ell(X; \gamma, t^\ell)$.

### 5 Space of Functions Defined by Derivatives of Completely Monotone Functions

As mentioned in [6], the fact that the energy distance defines a metric on a separable Hilbert space can be proved using the proposed method, but also follows as a consequence of the fact that if $\mathcal{H}$ is a separable Hilbert space, then a measure $\mu \in \mathcal{M}(\mathcal{H})$
such that $\|x\|^a \in L^1(|\mu|)$, $a \in (0, \infty) \setminus 2\mathbb{N}$, satisfies

$$\int_{\mathcal{H}} \|x - y\|^a d\mu(x) = 0, \quad y \in \mathcal{H}$$

(5)

if and only if $\mu$ is the zero measure, proved in [25,26].

As a consequence of Theorem 3.1 and the kernel mean embedding, if $\psi \in CM_0$ and is not a constant function, then

$$\int_{\mathcal{H}} \psi(\|x - y\|^2) d\mu(x) = 0, \quad y \in \mathcal{H},$$

if and only if $\mu$ is the zero measure. In this section, we prove similar results on a much broader setting, as a consequence of the results presented in Sect. 4.

**Theorem 5.1** Let $\mathcal{H}$ be an infinite-dimensional Hilbert space, $\ell \in \mathbb{Z}_+$ and $\phi, \varphi \in CM_{\ell}$. If a measure $\mu \in \mathfrak{M}(\mathcal{H})$ such that $\|x\|^{2\ell} \in L^1(|\mu|)$ satisfies

$$\int_{\mathcal{H}} \psi(\|x - y\|^2) d\mu(x) = 0, \quad y \in \mathcal{H},$$

where $\psi := \phi - \varphi$ then it must hold that

$$\int_{\mathcal{H}} \psi(\|x - y\|^2 + c) d\mu(x) = 0, \quad y \in \mathcal{H}, c \geq 0.$$

In addition (even if $\mathcal{H}$ is not infinite-dimensional), $\psi$ is not a polynomial if and only if the only measure $\mu$ that satisfies

$$\int_{\mathcal{H}} \psi(\|x - y\|^2 + c) d\mu(x) = 0, \quad y \in \mathcal{H}, c \geq 0$$

is the zero measure.

A complement to the second part of Theorem 5.1 is that if $\gamma : X \times X \to \mathbb{R}$ is a continuous metrizable CND kernel with bounded diagonal, then for a nonpolynomial function $\psi : [0, \infty) \to \mathbb{R}$ that is the difference of two functions in $CM_{\ell}$ it holds that

$$\int_X \psi(\gamma(x, y) + c) d\mu(x) = 0, \quad y \in X, c \geq 0 \Leftrightarrow \mu = 0.$$

The proof is analogous to the one in the Theorem 5.1 and thus omitted. For some functions, we can provide a version of Theorem 5.1 on finite-dimensional spaces.

**Theorem 5.2** Let $\ell \in \mathbb{N}$ and $\mathcal{H}$ be a Hilbert space. A measure $\mu \in \mathfrak{M}(\mathcal{H})$ such that $\|x\|^{2\ell} \in L^1(|\mu|)$ and $\psi(\|x - y\|^2) \in L^1(|\mu| \times |\mu|)$ satisfies

$$\int_{\mathcal{H}} \psi(\|x - y\|^2) d\mu(x) = 0, \quad y \in \mathcal{H} \Leftrightarrow \mu = 0$$
when $\psi : [0, \infty) \to \mathbb{R}$ is one of the following functions:

(i) $\psi(t) = t^{a/2}$, $2(\ell - 1) < a < 2\ell$;

(ii) $\psi(t) = t^{\ell-1} \log(t)$, $\ell > 1$;

(iii) $\ell = 1$ and $\psi \in \text{CM}_\ell$ is not a polynomial, $\psi(0) < 0$;

(iv) $\ell = 2$ and $\psi \in \text{CM}_\ell$ is not a polynomial, $\psi(0) \leq 0$.

We remark that we may replace the assumption $\|x\|^{2\ell} \in L^1(\mu)$ by $\phi(\|x\|^2), \varphi(\|x\|^2) \in L^1(\mu)$, if $\phi, \varphi \in C^{\ell-1}[0, \infty)$ in Theorem 5.1 and its complement. In Theorem 5.2, the same occurs, but not for the case (ii), because on the proof we make use of the kernel $\|x - y\|^{2\ell - 4}$.

6 Proofs

6.1 Section 3

Because of the relation between the CND kernel $\gamma$ and the PD kernel $K_\gamma$, if $\gamma : X \times X \to \mathbb{R}$ is CND then there exists a Hilbert space $\mathcal{H}$ and a function $h : X \to \mathcal{H}$, such that (Proposition 3.2 at page 82 in [13])

$$\gamma(x, y) = \|h(x) - h(y)\|^2 + \gamma(x, x)/2 + \gamma(y, y)/2, \quad x, y \in X.$$  

**Proof of Lemma 3.2** Since $\gamma$ is CND, the kernel $\beta(x, y) := \gamma(x, y) - \gamma(x, x)/2 - \gamma(y, y)/2$ is CND, $\beta(x, x) = 0$ for every $x \in X$ and $\beta^{1/2}$ is a pseudometric on $X$. Since $\gamma(x, x)$ is bounded and $\mu$ is a finite measure, for $\theta \geq 1$ the three equivalences for $\gamma$ are, respectively, equivalent to the three equivalences for the CND kernel $\beta$ by the Minkowski inequality. If $\theta \in (0, 1)$, the same relation occurs, but it follows from the general relation on $L^\theta$

$$\int |f + g|^{\theta} \leq \int |f|^{\theta} + \int |g|^{\theta},$$

which holds because $(a + 1)^p \leq 1 + a^p$, for every $a \geq 0$ and $p \in (0, 1)$. In particular, we may suppose that $\gamma$ is a CND kernel for which $\gamma(x, x) = 0$ for every $x \in X$. $\square$

If $\gamma(x, y)^\theta \in L^1(\mu \times \mu)$, then there exists $z \in X$ for which $\gamma(x, z)^\theta \in L^1(\mu)$ by the Fubini–Tonelli theorem.

If $x \in X \rightarrow \gamma(x, z)^\theta \in L^1(\mu)$ for some $z \in X$, then for every $y \in X$

$$(\gamma(x, y))^\theta = ((\gamma(x, y))^{1/2})^{2\theta} \leq (\gamma(x, z)^{1/2} + \gamma(y, z)^{1/2})^{2\theta} \leq 2^{2\theta}(\gamma(x, z)^\theta + \gamma(y, z)^\theta).$$

This inequality directly implies the integrability of $x \in X \rightarrow \gamma(x, y)^\theta$. The same inequality implies that if $x \in X \rightarrow \gamma(x, z)^\theta \in L^1(\mu)$ for every $z \in X$ then $\gamma^\theta \in L^1(\mu \times \mu)$.  

Springer
Proof of Theorem 3.3 The fact that $I(\mu, \nu)_{\gamma, \psi}$ is a semi-inner product on $M_1(X; \gamma, \psi)$ is a consequence of the kernel mean embedding. Indeed, consider the CND kernel $\gamma' = \psi(\gamma)$ and the related PD kernel $K^{\gamma'}_{\gamma'}$, for an arbitrary $z \in X$. Note that

$$K^{\gamma'}_{\gamma'}(x, x) = 2\gamma'(x, z) - \gamma'(x, x) - \gamma'(z, z),$$

then, if $x \rightarrow \gamma'(x, x)$ is a bounded function, the fact that $x \rightarrow \gamma'(x, z) \in L^1(|\mu|)$ implies that $x \rightarrow \sqrt{K^{\gamma'}_{\gamma'}(x, x)} \in L^1(|\mu|)$ because $\mu$ is finite. Hence, if we add the hypothesis that $\nu(\gamma(x, y)) = 0$, we obtain by the kernel mean embedding that

$$(K^{\gamma'}_{\gamma'}(x, x), K^{\gamma'}_{\gamma'}(y, y))_{K^{\gamma'}} = \int_X \int_X K^{\gamma'}_{\gamma'}(x, y)d\mu(x)d\nu(y)$$

$$= \int_X \int_X [\gamma'(x, z) + \gamma'(z, y) - \gamma'(x, y) - \gamma'(z, z)]d\mu(x)d\nu(y)$$

$$= \int_X \int_X -\gamma'(x, y)d\mu(x)d\nu(y).$$

For the second claim, since the representation in Theorem 2.3 is unique, if $\psi$ is not a linear function then $\sigma((0, \infty)) > 0$. If $\mu \in M_1(X; \gamma, \psi)$, then the 3 functions that describe $\psi(\gamma(x, y))$ are in $L^1(|\mu| \times |\mu|)$, because $e^{-r\gamma(x,y)} - 1 < 0$ for every $r > 0$ and $x, y \in X$. So, we can apply Fubini–Tonelli and obtain that

$$-\int_X \int_X \psi(\gamma(x, y))d\mu(x)d\mu(y) = a \int_X \int_X -\gamma(x, y)d\mu(x)d\mu(y)$$

$$+ \int_{(0, \infty)} \left[ \int_X \int_X e^{-r\gamma(x,y)}d\mu(x)d\mu(y) \right] \frac{1 + r}{r} d\sigma(r).$$

The first double integral is nonnegative by the first part of the proof. Since the kernel $\gamma$ is metrizable, the kernel $e^{-r\gamma(x,y)}$ is ISPD for every $r > 0$ by Theorem 3.1, so if $\mu \neq 0$

$$\int_X \int_X e^{-r\gamma(x,y)}d\mu(x)d\mu(y) > 0, \quad r > 0$$

and the conclusion follows because $\sigma((0, \infty)) > 0$. \(\Box\)

Proof of Theorem 3.4 By equation 4.38.2 in [27], we have that for $t \geq 1$

$$\text{arcosh}(t) = \log(2) + \log(t) - \sum_{k=1}^{\infty} \frac{(2k)!}{2^{2k}(k!)^2} \frac{t^{-2k}}{2k}.$$ 

In [13], it is proved that $\log([x, y])$ is a CND kernel on $\mathcal{H}$ while by [19] the PD kernel $[x, y]^{-2k}$ on $\mathbb{H}$ is ISPD for every $k \in \mathbb{N}$. Since the series appearing on the arcosh formula above only contains nonnegative numbers, we may reverse the order
of summation with integration for any $\eta \in \mathcal{M}_1(\mathbb{H}; t)$. Consequently, if $\mu$ is not the zero measure

$$- \int_{\mathbb{H}} \int_{\mathbb{H}} d_{\mathbb{H}}(x, y) d\mu(x) d\mu(y) = - \int_{\mathbb{H}} \int_{\mathbb{H}} \text{arcosh}([x, y]) d\mu(x) d\mu(y)$$

$$= \int_{\mathbb{H}} \int_{\mathbb{H}} \text{arcosh}([x, y]) \sum_{k=1}^{\infty} \frac{(2k)!}{2^{2k} (k!)^2 2k} [x, y]^{-2k} d\mu(x) d\mu(y)$$

$$\geq \sum_{k=1}^{\infty} \frac{(2k)!}{2^{2k} (k!)^2 2k} \int_{\mathbb{H}} \int_{\mathbb{H}} [x, y]^{-2k} d\mu(x) d\mu(y) > 0. \tag*{\Box}$$

Before proving Theorem 3.5, we prove that $\log[(x, z), (y, w)]$ is a CND kernel. The proof is based in the one given in [13] for the real hyperbolic case. Indeed, the kernel $\langle x, y \rangle / \langle z \rangle$ is PD, as it is the product of two PD kernels, then by Cauchy–Schwarz we have that

$$|\langle x, y \rangle|^2 \leq (x, x) (y, y) = (|z|^2-1)(|w|^2-1) \leq (||z|||w|-1)^2;$$

hence, $[(x, z), (y, w)] = |z \langle w - \langle x, y \rangle| \geq 1$. In particular, $1 - |\langle x, y \rangle / \langle z \rangle| \geq 1/|z \rangle$, and this implies that the real part of the complex number $1 - \langle x, y \rangle / \langle z \rangle$ is nonnegative. Hence, for any $r > 0$

$$e^{-r \log |z \langle w - \langle x, y \rangle| = |z \langle w - \langle x, y \rangle|^{-r} = |z|^{-r} |w|^{-r} |1 - \langle x, y \rangle / \langle z \rangle|^r.$$

The kernel $|1 - \langle x, y \rangle / \langle z \rangle|^r$ is PD because

$$|1 - \langle x, y \rangle / \langle z \rangle|^r = (1 - \langle x, y \rangle / \langle z \rangle)^{-r/2} (1 - \langle y, x \rangle / \langle w \rangle)^{-r/2},$$

and the kernel $(1 - \langle x, y \rangle / \langle z \rangle)^{-r/2}$ is PD as we can rewrite it using the binomial series as

$$(1 - \langle x, y \rangle / \langle z \rangle)^{-r/2} = 1 + \sum_{n=1}^{\infty} \left( \frac{1/2}{n} \right) (\langle x, y \rangle / \langle z \rangle)^n.$$

Similar for $(1 - \langle x, y \rangle / \langle w \rangle)^{-r/2}$. This implies that $\log[(x, z), (y, w)]$ is CND and that $[(x, z), (y, w)]^{-r}$ is PD for any $r > 0$.

**Proof of Theorem 3.5** Using Equation 4.38.2 in [27] once again, we have that

$$d_{\mathbb{H}}(x, y) = \text{arcosh}([(x, z), (y, w)]) = \log(2) + \log(|[x, y]|) - \sum_{k=1}^{\infty} \frac{(2k)!}{2^{2k} (k!)^2 2k} [x, y]^{-2k}.$$
for every \( x, y \in \mathbb{H}^C \), which implies that \( d_{\mathbb{H}^C}(x, y) \) is a CND kernel. Similarly, the PD kernel \( ||x, y||^{-2k} \) on \( \mathbb{H} \) is ISPD, because \( ||x, y|| \) is a metrizable CND kernel and the function \( \psi(t) = t^{-2k} \) can be written as

\[
t^{-2k} = \frac{1}{\Gamma(2k)} \int_{(0, \infty)} r^{2k-1} e^{-rt} dr,
\]

and the conclusion that the kernel is ISPD follows by an integration argument using Theorem 3.1. The remaining arguments are similar to the proof of Theorem 3.4, so we omit them.

\[\Box\]

**Proof of Corollary 3.6** By Remark 3.3-(iv) on [15], if \( \psi \) satisfies these assumptions then we can write the kernel \( D_{\psi, \gamma} \) as

\[
D_{\psi, \gamma}(x, y) = \psi(\gamma(x, y)) = \int_{(0, \infty)} (1 - e^{-r\gamma(x, y)}) \frac{1 + r}{r} d\sigma(r)
\]

where \( \sigma \) is a nonnegative Radon measure in \( M((0, \infty)) \). Because \( \gamma \) is a metric, we have that

\[
1 - e^{-r\gamma(x, y)} \leq [1 - e^{-r\gamma(x, z)}] + [1 - e^{-r\gamma(z, y)}], \quad x, y, z \in X,
\]

which proves that \( D_{\psi, \gamma}(x, y) \leq D_{\psi, \gamma}(x, z) + D_{\psi, \gamma}(z, y) \).

The topologies are equivalent because \( \psi \) is necessarily an increasing function with \( \psi(0) = 0 \), so \( \psi(t_n) \to 0 \) if and only if \( t_n \to 0 \).

The metric space \( (X, D_{\psi, \gamma}) \) has strong negative type because the kernel \( \gamma \) is continuous on the metric topology \( (X, \gamma) \), \( \psi \) is not a linear function and the remaining requirements for Theorem 3.3 are satisfied.

\[\Box\]

6.2 Section 4

In order to prove Lemmas 4.1 and 4.2, we need a few estimates for the functions involved.

**Lemma 6.1** For \( \ell \in \mathbb{Z}_+ \), the function \( E_{\ell}(s) = (-1)^\ell (e^{-s} - \omega_{\ell}(s)) \) is nonnegative, increasing, convex and \( (E_{\ell+1})^{(1)} = E_{\ell} \). Also, there exists constants \( M_i > 0 \) which only depend on \( \ell \) and satisfy

\[
M_1 \min(t^\ell, t^{\ell-1}) \leq (-1)^\ell (e^{-rt} - \omega_{\ell}(rt)) \frac{1 + r}{r^\ell} \leq M_2(1 + t^\ell), \quad t, r \geq 0,
\]

and this inequality implies that

\[
0 \leq \frac{(e^{-st} - \omega_{\ell}(st))}{(e^{-t} - \omega_{\ell}(t))} \leq \frac{M_2}{M_1} s^\ell, \quad t \geq 0, s \geq 1
\]

\[\Box\]
Proof The case \( \ell = 0, 1 \) is immediate. If \( \ell \geq 2 \), note that \( E_2^{(1)} = -(e^{-s} - 1) \), which is a nonnegative and increasing function. Since by induction, \((E_{\ell+1})^{(1)} = E_{\ell}\), we obtain that \( E_{\ell} \) is increasing and nonnegative. It is convex because \( E_{\ell}^{(2)} = E_{\ell-2} \) and this function is nonnegative.

For the second part, as

\[
\lim_{s \to \infty} (-1)^{\ell} (e^{-s} - \omega_\ell(s)) \frac{1+s}{s^\ell} = \frac{1}{(\ell-1)!}, \quad \lim_{s \to 0} (-1)^{\ell} (e^{-s} - \omega_\ell(s)) \frac{1+s}{s^\ell} = \frac{1}{\ell!} > 0,
\]

and the function \( E_{\ell} \) is increasing, we obtain the existence of \( M_1, M_2 > 0 \) for which

\[
M_1 \leq (-1)^{\ell} (e^{-rt} - \omega_\ell(rt)) \frac{1+rt}{(rt)^\ell} \leq M_2, \quad r, t \geq 0.
\]

Since \( \min(1, 1/t) \leq (1 + r)/(1 + rt) \leq \max(1, 1/t) \), we obtain the first inequality. For the second inequality we have

\[
0 \leq \frac{(e^{-st} - \omega_\ell(st))}{(e^{-t} - \omega_\ell(t))} \leq \frac{M_2(st)^{\ell}}{1+st} \frac{1+t}{M_1 t^{\ell}} \leq \frac{M_2 s^{\ell}}{M_1}.
\]

\( \Box \)

Proof of Lemma 4.1 If \( \psi \in CM_\ell \cap C^{\ell-1}([0, \infty)) \), the continuous function \( \phi := (-1)^{\ell} \psi^{(\ell-1)} \) is a Bernstein function, so by Theorem 2.3 there exists a nonnegative measure \( \sigma \in \mathcal{M}((0, \infty)) \) and \( b_\ell \geq 0 \), for which

\[
\phi(t) = \phi(0) + b_\ell t + \int_{(0, \infty)} (1-e^{-rt}) \frac{1+r}{r} d\sigma(r), \quad t \geq 0.
\]

Define the function

\[
\varphi(t) = (-1)^{\ell} \frac{b_\ell}{\ell!} t^{\ell-1} + \int_{(0, \infty)} (e^{-rt} - \omega_\ell(rt)) \frac{1+r}{r^{\ell}} d\sigma(r), \quad t \geq 0
\]

which is well defined because of the previous lemma. The same lemma implies that \( \varphi \) is differentiable in \([0, \infty)\) and

\[
\varphi^{(1)}(t) = (-1)^{\ell} \frac{b_\ell}{(\ell-1)!} t^{\ell-1} + (-1) \int_{(0, \infty)} (e^{-rt} - \omega_{\ell-1}(rt)) \frac{1+r}{r^{\ell-1}} d\sigma(r), \quad t > 0.
\]

By induction, we obtain that

\[
\varphi^{(\ell-1)}(t) = (-1)^{\ell} b_\ell t + (-1)^{\ell-1} \int_{(0, \infty)} (e^{-rt} - 1) \frac{1+r}{r} d\sigma(r), \quad t \geq 0,
\]

so \([\psi - \varphi]^{(\ell-1)}(t) = (-1)^{\ell} \phi(0) = \psi^{(\ell-1)}(0)\); hence, \( \psi - \varphi \) is a polynomial of degree less than or equal to \( \ell - 1 \). Readjusting the terms we obtain the desired representation,
since $\varphi^{(j)}(0) = 0$ for $0 \leq j \leq \ell - 1$. The converse follows by the same arguments of the first part using the previous lemma.

For the proof of Lemma 4.2, we emphasize that if $\gamma : X \times X \to [0, \infty)$ is CND and $x, y, z \in X$, then

$$\gamma(x, y) - 2\gamma(x, z) - 2\gamma(y, z) \leq 2\gamma(x, y) - 4\gamma(x, z)$$

$$-4\gamma(y, z) + \gamma(x, x) + \gamma(y, y) + 4\gamma(z, z) \leq 0$$

**Proof of Lemma 4.2** If $\psi$ is a polynomial, the result is a direct application of Lemma 3.2. If the measure $\sigma$ in the representation of $\psi$ is not the zero measure, then by Lemma 6.1

$$\int_{(0, \infty)} (-1)^\ell (e^{-r^\ell} - \omega_\ell(r^\ell)) \frac{1 + r}{r^\ell} d\sigma(r) \geq M_1 \min(r^\ell, r^{\ell-1}) \sigma((0, \infty));$$

hence, if $\psi(\gamma)$ satisfy any of the three equivalences, then $\gamma^k$ satisfy the respective equivalence for $k \leq \ell - 1$. Lemma 3.2 implies that we may assume in this setting that $a_k = 0$ for $k \leq \ell - 1$ in the representation of Lemma 4.1. As $(-1)^\ell a_\ell$ and $E_\ell$ are nonnegative functions, the proof of the three equivalences can be done separately for the remaining polynomial part and the integral part. The polynomial part is a consequence of Lemma 3.2, so we only focus on the integral part.

If relation (i) holds, then by Fubini–Tonelli there must exist a $z \in X$ for which $x \in X \to \psi(\gamma(x, z)) \in L^1(\mu)$.

Suppose now that relation (ii) holds, then by Lemma 6.1

$$E_\ell(\gamma(x, y)) \leq E_\ell(2\gamma(x, z) + 2\gamma(y, z)) = E_\ell \left( \frac{4(2\gamma(x, z) + 2\gamma(y, z))}{4} \right)$$

$$\leq \frac{M_2}{M_1} 4^\ell \frac{2\gamma(x, z) + 2\gamma(y, z)}{4} \leq \frac{M_2}{M_1} 4^\ell (E_\ell(\gamma(x, z)) + E_\ell(\gamma(y, z)))$$

as $\mu$ is finite, this implies that $x \in X \to \psi(\gamma(x, y)) \in L^1(\mu)$ for every $y \in X$. The same inequality used before implies that relation (iii) implies relation (i).

Lemma 3.1 in [18] is done by using coordinates. The same method could be used in the setting of Lemma 4.3, because of Mercer theorem. However, we think that the arguments are more easy to follow if we obtain this result as a consequence of the case $X = H$ and $\gamma = \|x - y\|^2$.

It is worth mentioning that although we are not assuming that $\mathcal{H}$ is separable, as we are using a measure $\mu \in \mathcal{M}(\mathcal{H})$, so $\text{span}(\text{Supp}(\mu))$ is a separable space; hence, for the objectives of Lemma 4.3, we may suppose that there is an orthonormal basis $(e_\xi)_{\xi \in \mathbb{N}}$ such that $(x, y) = \sum_{\xi \in \mathbb{N}} x_\xi y_\xi$, $|\mu| \times |\mu|$ almost everywhere.

Also, we use the same infinite-dimensional multinomial theorem that was used to prove that the Gaussian kernel is ISPD on Hilbert spaces in [19]. If $\mathcal{H}$ is a real Hilbert
space and \((e_\xi)_{\xi \in \mathbb{N}}\) is an orthonormal basis for it, then for every \(n \in \mathbb{N}\)

\[
\left( \sum_{\xi \in \mathbb{N}} x_\xi y_\xi \right)^n = \sum_{\alpha \in \mathbb{N}, |\alpha| = n} \frac{n!}{\alpha!} x^\alpha y^\alpha
\]

(6)

where \((\mathbb{N}, \mathbb{Z}_+)\) is the space of functions from \(\mathbb{N}\) to \(\mathbb{Z}_+\), the condition \(|\alpha| = n\) means that \(\sum_{\xi \in \mathbb{N}} \alpha(\xi) = n\) (in particular \(\alpha\) must be the zero function except for a finite number of points). Also \(\alpha! = \prod_{\xi \in \mathbb{N}} \alpha(\xi)!\) (which makes sense because \(0! = 1\)) and \(x^\alpha = \prod_{\alpha(\xi) \neq 0} x_\xi^{\alpha(\xi)}\). This result can be proved using approximations of \((\sum_{\xi \in \mathbb{N}} x_\xi y_\xi)^n\) on finite-dimensional spaces and the multinomial theorem in those spaces. The number \([l]\) stands for the smallest integer less then or equal to \(l\).

**Lemma 6.2** Let \(\mathcal{H}\) be a real Hilbert space, \(n \in \mathbb{N}\) and \(\mu \in \mathcal{M}(\mathcal{H})\). Suppose that \(\|x - y\|^{2n} \in L^1(|\mu| \times |\mu|)\), then

\[
\langle x, y \rangle^k \|x\|^{2i} \|y\|^{2j} \in L^1(|\mu| \times |\mu|), \quad k, i, j \in \mathbb{Z}_+, \quad k + i + j \leq n.
\]

Moreover, for some \(w \in \mathcal{H}\) it holds that \(\int_{\mathcal{H}} \int_{\mathcal{H}} \langle x - w, y - w \rangle^k d\mu(x) d\mu(y) = 0\) for every \(0 \leq k \leq n - 1\) if and only if these \(n\) equalities hold for all \(w \in \mathcal{H}\). These \(n\) equalities imply that

\[
(-1)^n \int_{\mathcal{H}} \int_{\mathcal{H}} \|x - y\|^{2n} d\mu(x) d\mu(y) = \sum_{l=0}^{\lceil n/2 \rceil} \binom{n}{2l} \binom{2l}{l} 2^{n-2l} \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle^{n-2l} \|x\|^{2i} \|y\|^{2j} d\mu(x) d\mu(y) \geq 0,
\]

and

\[
\int_{\mathcal{H}} \int_{\mathcal{H}} \|x - y\|^{2m} d\mu(x) d\mu(y) = 0, \quad 0 \leq m \leq n - 1.
\]

**Proof** By Lemma 3.2, the fact that \(\|x - y\|^{2n} \in L^1(|\mu| \times |\mu|)\) is equivalent to \(\|x\|^{2n} \in L^1(|\mu|)\). Since \(|\langle x, y \rangle|^k \|x\|^{2i} \|y\|^{2j}| \leq \|x\|^{2i+k} \|y\|^{2j+k} \leq \max\{1, \|x\|^{2n}\}\), we obtain the desired integrability.

Suppose that the \(n\) equalities hold for \(w = 0\), hence for any \(z \in \mathcal{H}\)

\[
\int_{\mathcal{H}} \langle x - z, y - z \rangle^k d\mu(x) = \sum_{j=0}^{k} \binom{k}{j} \langle -z, y - z \rangle^{n-k} \int_{\mathcal{H}} \langle x, y - z \rangle^k d\mu(x) = 0,
\]

for every \(y \in \mathcal{H}\) and \(0 \leq k \leq n - 1\). The converse follows by a similar argument.
Now we prove the last assertion. Note that

\[ \|x - y\|^{2m} = (\|x\|^2 + \|y\|^2 - 2\langle x, y \rangle)^m \]

\[ = \sum_{k=0}^{m} \sum_{i=0}^{m-k} \binom{m}{k} \binom{m-k}{i} (-2)^k \langle x, y \rangle^k \|x\|^{2i} \|y\|^{2(m-k-i)} \]

If \( k + 2i \leq n - 1 \), then by the hypothesis

\[ 0 = \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle^{k+2i} d\mu(x) d\mu(y) = \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle^k \left( \sum_{\xi \in \mathbb{N}} x_\xi y_\xi \right)^{2i} d\mu(x) d\mu(y) \]

\[ = \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle^k \left( \sum_{\xi \in \mathbb{N}} x_\xi y_\xi \right)^{2i} d\mu(x) d\mu(y) \]

\[ = \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle^k \left( \sum_{|\alpha|=2i} \frac{2!}{\alpha!} x^\alpha y^\alpha \right) d\mu(x) d\mu(y) \]

\[ = \sum_{|\alpha|=2i} \frac{2!}{\alpha!} \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle^k x^\alpha y^\alpha d\mu(x) d\mu(y). \]

But then, \( \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle^k x^\alpha y^\alpha d\mu(x) d\mu(y) = 0 \) for every \( \alpha \in (\mathbb{N}, \mathbb{Z}_+) \) with \( |\alpha| = 2i \), because the kernel inside the double integral is PD, continuous and satisfies the conditions in Lemma 2.1. In particular, we have that

\[ \int_{\mathcal{H}} \langle x, y \rangle^k x^\alpha d\mu(x) = 0, \quad y \in \mathcal{H}, \alpha \in (\mathbb{N}, \mathbb{Z}_+), |\alpha| = 2i. \]

Then

\[ \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle^k \|x\|^{2i} \|y\|^{2(m-k-i)} d\mu(x) d\mu(y) \]

\[ = \sum_{|\beta|=m-k-i} \sum_{|\alpha|=i} \frac{(m-k-i)! i!}{\beta!} \int_{\mathcal{H}} \int_{\mathcal{H}} \langle x, y \rangle^k x^{2\alpha} y^{2\beta} d\mu(x) d\mu(y) = 0. \]

By symmetry, the same double integral is zero when \( k + 2(m-k-i) \leq n - 1 \). Those two relations do not occur only when \( n = m \) and \( n = k + 2i \). The remaining terms on the sum when \( n = m \) are exactly those on the statement on the theorem after a simplification using that \( n = k + 2i \). The conclusion follows because the kernel \( \langle x, y \rangle^k \|x\|^{2i} \|y\|^{2i} \) is continuous, PD and satisfies the conditions in Lemma 2.1. \( \square \)

**Proof of Lemma 4.3** By the hypothesis on \( \gamma \), there exists a Hilbert space \( \mathcal{H} \) and a continuous function \( T : X \to \mathcal{H} \), such that \( \gamma(x, y) = \|T(x) - T(y)\|^2_{\mathcal{H}} + c \), where \( c \geq 0 \) is the value of \( \gamma \) on the diagonal. In particular, \( K^w_\gamma(x, y) = 2(T(x) - T(w), T(y) - T(w)) \)
$T(w)$. If $\mu \in M(X)$ is a measure satisfying the conditions on the lemma, then the image measure $\mu_T \in M(H)$ satisfies the same conditions of Lemma 6.2. The conclusion follows by standard properties of image measures.

Lemma 6.3 There exists an $M > 0$, which only depends on $\ell \in \mathbb{Z}_+$ for which

$$|e^{-rt} - e_\ell(r)\omega_\ell(rt)|(1 + r^\ell) \leq Mr^\ell(1 + t^\ell), \quad r > 0, t \geq 0. \quad (7)$$

Proof Case $r \geq 1$: In this case, the left-hand side of Equation 7 is

$$|e^{-rt} - e_\ell(r)\omega_\ell(rt)|(1 + r^\ell) \leq 2(1 + |e_\ell(r)\omega_\ell(rt)|)r^\ell \leq 2(1 + \sum_{l=0}^{\ell-1}|e_\ell(r)r^l|t^l/l!)r^\ell.\quad \text{ (8)}$$

Since each function $|e_\ell(r)r^l|$ is bounded, the results follows from the fact that $t^l \leq 1 + t^\ell$.

Case $r < 1$: In this case, the left-hand side of Eq. 7 is

$$|e^{-rt} - e_\ell(r)\omega_\ell(rt)|(1 + r^\ell) \leq (|e^{-rt} - \omega_\ell(rt)| + |(e_\ell(r) - 1)\omega_\ell(rt)|)(1 + r^\ell).\quad \text{ (9)}$$

By Lemma 6.1, it holds that

$$|(e^{-rt} - \omega_\ell(rt))(1 + r^\ell) \leq M_2(1 + t^\ell)r^\ell \frac{1 + r^\ell}{1 + r} \leq 2M_2(1 + t^\ell)r^\ell.\quad \text{ (10)}$$

In the other function, we have that

$$|(e_\ell(r) - 1)\omega_\ell(rt)|(1 + r^\ell) \leq 2\sum_{l=0}^{\ell-1}|e_\ell(r) - 1|t^l/l!,\quad \text{ (11)}$$

since $|e_\ell(r) - 1| = e^{-r} \sum_{k=\ell}^{\infty} r^k/k! \leq er^\ell$, we obtain the desired inequality. \hfill \square

Proof of Theorem 4.4 By the representation of $\psi$ in Theorem 2.4, we have that

$$\psi(\gamma(x, y)) = \int_{(0, \infty)} \frac{e^{-r\gamma(x, y)} - e_\ell(r)\omega_\ell(\gamma(x, y)r)}{r^\ell} d\lambda(r) + \sum_{k=0}^{\ell} a_\ell \gamma(x, y)^k. \quad \text{ (12)}$$

Because of Lemma 6.3, the $\ell + 2$ functions above are in $L^1(|\mu| \times |\mu|)$. Lemma 4.3 implies that

$$\int_X \int_X \sum_{k=0}^{\ell} a_\ell \gamma(x, y)^k d\mu(x)d\mu(y) = \int_X \int_X e_\ell \gamma(x, y)^\ell d\mu(x)d\mu(y) \geq 0. \quad \text{ (13)}$$
On the other hand, because of Lemma 6.3 we can apply Fubini–Tonelli, and then

\begin{align*}
\int_X \left[ \int_{(0,\infty)} e^{-\gamma(x,y)r} - e^r(r)\omega(r)(\gamma(x,y)r) \right] d\mu(x)d\mu(y) \\
= \int_{(0,\infty)} \left[ \int_X \int_X e^{-\gamma(x,y)r} d\mu(x)d\mu(y) \right] d\lambda(r) \geq 0,
\end{align*}

because the inner double integral is a nonnegative number for every \( r > 0 \) by Theorem 3.1.

Because the representation for \( \psi \) is unique, if \( \psi \) is not a polynomial of degree \( \ell \) or less then \( \lambda((0,\infty)) > 0 \), also, if \( \gamma \) is metrizable, by Theorem 3.1 the inner double integral is a positive number for every \( r > 0 \) when \( \mu \) is not the zero measure, and then, the triple integral is a positive number as well. \( \Box \)

6.3 Section 5

We use throughout this section that a function \( \phi \in CM_\ell \) satisfies

\[ |\phi(t)| \leq C(1 + t^\ell), \]

for some \( C > 0 \), which is a consequence of Lemma 6.3.

**Proof of Theorem 5.1** Since \( \mathcal{H} \) is infinite-dimensional, take \((e_i)_{i \in \mathbb{N}}\) to be an orthonormal sequence of vectors in \( \mathcal{H} \). By the dominated convergence theorem, we have that

\begin{align*}
0 = \int_{\mathcal{H}} \psi(\|x - y - ce_i\|^2) d\mu(x) \rightarrow \int_{\mathcal{H}} \psi(\|x - y\|^2 + c^2) d\mu(x), \quad y \in \mathcal{H}, \quad c \in [0,\infty)
\end{align*}

because \( \langle x - y, e_i \rangle \to 0 \) as \( t \to \infty \) and \( |\psi(t)| \leq |\phi(t)| + |\phi(t)| \), which proves the first assertion.

Now, if \( \psi \) is a polynomial of degree \( n \), let \( t_1, \ldots, t_N \in \mathbb{R}, \ c_1, \ldots, c_N \in \mathbb{R} \) (not all null) such that \( \sum_{i=1}^N c_i p(t_i) = 0 \) for every \( p \in \pi_{2n}(\mathbb{R}) \). Then if \( \|v\| = 1 \), the measure \( \mu := \sum_{i=1}^N c_i \delta(t_i v) \in \mathcal{M}(\mathcal{H}) \) is nonzero and

\begin{align*}
\int_{\mathcal{H}} \psi(\|x - y\|^2) d\mu(x) = \sum_{i=1}^N c_i \psi(\|y - (y, v)v\|^2 + ((y, v) - t_i)^2) = 0
\end{align*}

because this function is a polynomial of degree \( 2n \) for every fixed \( y \in \mathcal{H} \).

For the converse, first, we show that it is sufficient to prove the case \( \ell = 0 \).

Indeed, the function \( c \in (0,\infty) \to F(c) := \psi(\|x - y\|^2 + c) \in \mathbb{R} \) is differentiable for every \( x, y \in \mathcal{H} \), and

\[ F^{(1)}(c) = \psi^{(1)}(c + \|x - y\|^2). \]
Since $\psi = \varphi - \phi$, and those functions are elements of $CM_\ell$, we have that $(-1)\psi^{(1)}(t + c) \in CM_{\ell-1}$. In particular,

$$\int_{\mathcal{H}} \psi^{(1)}(c + \|x - y\|^2)d\mu(x) = 0, \quad y \in \mathcal{H}, \quad c > 0. \quad (8)$$

Since $\psi^{(1)}(c + \cdot)$ also is the difference between two functions in $CM_{\ell-1}$ for every $c > 0$, by induction, we may assume that $\ell = 0$.

Assume that $\psi$ is not a polynomial and $\mu$ is a nonzero measure that satisfies the equality on the statement of the theorem. The function $\psi(c + \cdot)$ is the difference between two completely monotone functions on $[0, \infty)$, so there exists a measure $\beta_c$ in $\mathcal{M}([0, \infty))$ (not necessarily nonnegative) for which

$$\psi(c + t) = \int_{[0, \infty)} e^{-rt}d\beta_c(r), \quad c > 0, \quad t \geq 0.$$ 

Since the representation for completely monotone functions is unique, we have that $d\beta_{c+s}(r) = e^{-rs}d\beta_c(r)$ for every $c, s > 0$. Integrating the function on the hypotheses with respect to the measure $d\mu(y)$, we obtain that

$$0 = \int_{\mathcal{H}}\int_{\mathcal{H}} \psi(c + \|x - y\|^2)d\mu(x)d\mu(y)$$

$$= \int_{[0, \infty)}\int_{\mathcal{H}}\int_{\mathcal{H}} e^{-r\|x-y\|^2}d\mu(x)d\mu(y)d\beta_c(r), \quad c > 0. \quad (9)$$

The continuous and bounded function $I_\mu(r) := \int_{\mathcal{H}}\int_{\mathcal{H}} e^{-r\|x-y\|^2}d\mu(x)d\mu(y), r \geq 0$, is positive for every $r > 0$ by Theorem 3.1. Equation 9 implies that $(c = s + 1)$

$$0 = \int_{[0, \infty)} e^{-sr}I_\mu(r)d\beta_1(r), \quad s \geq 0.$$ 

By the uniqueness representation of Laplace transform, this can only occur if the finite measure $I_\mu d\beta_1$ is the zero measure on $[0, \infty)$. The behavior of $I_\mu$ implies that this occurs if and only if $I_\mu(0) = 0$ and $\beta_1$ is a multiple of $\delta_0$, the latter implies that $\psi$ is a constant function, which is a contradiction. \[ \square \]

**Proof of Theorem 5.2** By Theorem 5.1, we only need to focus on the finite-dimensional case. We prove $(i)$ and $(ii)$ by showing that it is sufficient to prove the case $\ell = 1, 2$, which will follow from $(iii)$ and $(iv)$. For the induction argument on $(ii)$, we assume a more general setting that $\psi(t) = t^{\ell-1}\log(t) + bt^{\ell-1}$, with $b \geq 0$.

Indeed, suppose that $\ell \geq 3$. Note then that the function $y \in \mathcal{H} \to F(y) := \psi(\|x - y\|^2) \in \mathbb{R}$ is twice differentiable on each direction of an orthonormal basis $(e_i)_{i \in J}$ for $\mathcal{H}$, and

$$\frac{\partial^2 F}{\partial^2 e_i} (y) = 4\psi^{(2)}(\|x - y\|^2)(y_i - x_i)^2 + 2\psi^{(1)}(\|x - y\|^2).$$
The second derivative is a function in $L^1(\mu)$, and summing on the $\iota$ variable, we obtain ($m = \dim(H)$)

$$0 = \int_H 4\psi''(\|x - y\|^2)\|x - y\|^2 + 2m\psi'(\|x - y\|^2)d\mu(x), \quad y \in H. \quad (10)$$

When $\psi$ is a function of type (i) or (ii), the integrand on this equation is equal to a positive multiple of $\|x - y\|^2$ (or $\|x - y\|^{\ell-4}$) plus a nonnegative multiple of $\|x - y\|^{\ell-4}$, which is the induction argument.

Now, let $\psi$ be an arbitrary function on $CM_{\ell}, \ell = 1, 2$, that is not a polynomial. For every $t > 0$, define $\eta_t := t\mu - \tau_t$, where $\tau_t = t\mu(H)\delta_0 - (\delta_{tv\mu} - \delta_{-tv\mu})/2$ and $v\mu$ is the vector mean, that is

$$\int_H \langle x, y \rangle d\mu(x) = \langle v\mu, y \rangle, \quad y \in H.$$

In the case $\ell = 1$, the vector $v\mu$ might not be well defined; in this case, define it as the vector zero. Then $\eta_t(H) = 0$, and if it is well defined $v_{\eta_t} = 0$. By the hypothesis, we obtain that

$$4 \int_H \int_H \psi(\|x - y\|^2)d\eta_t(x)d\eta_t(y) = \int_H \int_H \psi(\|x - y\|^2)d2\tau_t(x)d2\tau_t(y)$$

$$= \psi(0)(4t^2\mu(H))^2 + 2 - 2\psi(4t^2\|v\mu\|^2).$$

By Theorem 4.5, this is a nonnegative number for every $t > 0$. Suppose that $\ell = 2$. Since $(-1)^2\psi(t + 1) \in CM_2 \cap C^1([0, \infty))$, the inequalities in Lemma 6.1 implies that $(-1)^2\psi(t)$ converges to $+\infty$ as $t \to \infty$, so if $\|v\mu\| \neq 0$ or $\psi(0) < 0$ we would reach a contradiction, consequently $v\mu = 0, \psi(0) = 0$. In particular, we obtain that the double integral with respect to $\eta_1$ is zero, which by Theorem 4.5 we must have that $\mu = \mu(H)\delta_0$, because $\psi$ is not a polynomial. From this equality and the initial assumption on $\mu$, we obtain that $\mu(H)\psi(\|y\|^2) = 0$ for every $y \in H$, which can only occur if $\mu$ is the zero measure because $\psi$ is not a polynomial.

The case $\ell = 1$ follows by a similar analysis.

\textbf{Acknowledgements} This research was done while the author was a member of the Functional Analytic Team at Riken Center for Advanced Intelligence Project (AIP), Tokyo, Japan.

\textbf{Funding:} Not applicable

\textbf{Declarations}

\textbf{Conflict of interest} Not applicable

\textbf{References}

1. Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B., Smola, A.: A kernel method for the two-sample-problem. Adv. Neural. Inf. Process. Syst. \textbf{19}, 513–520 (2007)
2. Sriperumbudur, B.K., Gretton, A., Fukumizu, K., Schölkopf, B., Lanckriet, G.R.G.: Hilbert space embeddings and metrics on probability measures. J. Mach. Learn. Res. 11(50), 1517–1561 (2010)
3. Sejdinovic, D., Sriperumbudur, B., Gretton, A., Fukumizu, K.: Equivalence of distance-based and rkhs-based statistics in hypothesis testing. The Ann. Stat. (2013). https://doi.org/10.1214/13-AOS1140
4. Székely, G.J., Rizzo, M.L.: Testing for equal distributions in high dimension. Inter. Stat. 5(16.10), 1249–1272 (2004)
5. Székely, G.J., Rizzo, M.L.: Energy statistics: a class of statistics based on distances. J. Stat. Plan. Inference 143(8), 1249–1272 (2013). https://doi.org/10.1016/j.jspi.2013.03.018
6. Lyons, R.: Distance covariance in metric spaces. Ann. Probab. 41(5), 3284–3305 (2013). https://doi.org/10.1215/12-AOP803
7. Lyons, R.: Hyperbolic space has strong negative type. Illinois J. Math. 58(4), 1009–1013 (2014). https://doi.org/10.1215/144619297
8. Gangolli, R.: Positive definite kernels on homogeneous spaces and certain stochastic processes related to lévy brownian motion of several parameters. Anna. de Institut Henri Poincaré Probabilités et Stat. 3(2), 121–226 (1967)
9. Robertson, G.: Crofton formulae and geodesic distance in hyperbolic spaces. J. Lie Theory 8(1), 163–172 (1998)
10. Mattner, L.: Strict definiteness of integrals via complete monotonicity of derivatives. Trans. Am. Math. Soc. 349(8), 3321–3342 (1997). https://doi.org/10.1090/S0002-9947-97-01966-1
11. Steinwart, I., Christmann, A.: Support Vector Machines. Springer, New York (2008)
12. Wendland, H.: Scattered Data Approximation, vol. 17. Cambridge University Press, Cambridge (2005)
13. Berg, C., Christensen, J., Ressel, P.: Harmonic analysis on semigroups: theory of positive definite and related functions. Graduate Texts in Mathematics, vol. 100. Springer, New York (1984)
14. Schoenberg, J.J.: Metric spaces and completely monotone functions. Ann. Math. 81–841 (1938)
15. Schilling, R.L., Song, R., Vondracek, Z.: Bernstein Functions: Theory and Applications, vol. 37. Walter de Gruyter, Berlin (2012)
16. Guo, K., Hu, S., Sun, X.: Conditionally positive definite functions and Laplace-Stieltjes integrals. J. Approx. Theory 74(3), 249–265 (1993). https://doi.org/10.1006/jath.1993.1065
17. Gelfand, I.M., Vilenkin, N.Y.: Generalized Functions, Vol. 4: Applications of Harmonic Analysis. Academic Press, New York (1964)
18. Micchelli, C.A.: Interpolation of scattered data: distance matrices and conditionally positive definite functions. Constr. Approx. 2(11), 11–22 (1984). https://doi.org/10.1007/978-94-009-6466-2_7
19. Guella, J.C.: On Gaussian kernels on Hilbert spaces and kernels on Hyperbolic spaces. Preprint at https://arxiv.org/abs/2102.00633
20. Simon-Gabriel, C.-J., Schölkopf, B.: Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions. J. Mach. Learn. Res. 19(44), 1–29 (2018)
21. Faraut, J., Harzallah, K.: Distances hilbertiennes invariantes sur un espace homogène. Ann. de l’Institut Fourier 24(3), 171–217 (1974). https://doi.org/10.5802/aif.525
22. Lyons, R.: Strong negative type in spheres. Pac. J. Math. 307(2), 383–390 (2020)
23. Böttcher, B., Keller-Ressel, M., Schilling, R.: Detecting independence of random vectors: generalized distance covariance and gaussian covariance. Mod. Stoch. Theory Appl. 5(3), 353–383 (2018) https://doi.org/10.15559/18-VMSTA116 https://doi.org/10.15559/18-VMSTA116
24. Böttcher, B., Keller-Ressel, M., Schilling, R.L.: Distance multivariance: new dependence measures for random vectors. Ann. Stat. 47(5), 2757–2789 (2019). https://doi.org/10.1214/18-AOS1764
25. Linde, W.: On rudin’s equimeasurability theorem for infinite dimensional hilbert spaces. Indiana Univ. Math. J. 35(2), 235–243 (1986)
26. Koldobskii, A.L.: Isometric operators in vector-valued lp-spaces. J. Sov. Math. 36, 420–423 (1987). https://doi.org/10.1007/BF01839616
27. NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.1.2 of 2021-06-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. http://dlmf.nist.gov/

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.