



On distance covariance in metric and Hilbert spaces

Svante Janson

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden
E-mail address: svante.janson@math.uu.se
URL: <http://www2.math.uu.se/~svante/>

Abstract. Distance covariance is a measure of dependence between two random variables that take values in two, in general different, metric spaces, see [Székely et al. \(2007\)](#) and [Lyons \(2013\)](#). It is known that the distance covariance, and its generalization α -distance covariance, can be defined in several different ways that are equivalent under some moment conditions. The present paper considers four such definitions and find minimal moment conditions for each of them, together with some partial results when these conditions are not satisfied.

Another purpose of the present paper is to improve existing results on consistency of distance covariance, estimated using the empirical distribution of a sample.

The paper also studies the special case when the variables are Hilbert space valued, and shows under weak moment conditions that two such variables are independent if and only if their (α -) distance covariance is 0; this extends results by [Lyons \(2013\)](#) and [Dehling et al. \(2020\)](#). The proof uses a new definition of distance covariance in the Hilbert space case, generalizing the definition for Euclidean spaces using characteristic functions by [Székely et al. \(2007\)](#).

1. Introduction

Distance covariance is a measure of dependence between two random variables \mathbf{X} and \mathbf{Y} that take values in two, in general different, spaces \mathcal{X} and \mathcal{Y} . This measure appears in [Feuerverger \(1993\)](#) as a test statistic when $\mathcal{X} = \mathcal{Y} = \mathbb{R}$; it was more generally introduced by [Székely et al. \(2007\)](#) for the case of random variables in Euclidean spaces, possibly of different dimensions. (See [Definition 2.4](#) below.) This was extended to general separable measure spaces by [Lyons \(2013\)](#), see also [Jakobsen \(2017\)](#), and to semimetric spaces (of negative type, see below) by [Sejdinovic et al. \(2013\)](#). (See [Definition 2.3](#).)

Our setting throughout this paper is the following (see also [Remark 2.7](#)): (\mathbf{X}, \mathbf{Y}) is a pair of random variables taking values in $\mathcal{X} \times \mathcal{Y}$, where \mathcal{X} and \mathcal{Y} are separable metric spaces, with metrics $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$; we write just d for both metrics when there is no risk of confusion. (All spaces will be separable throughout, whether said explicitly or not.)

Received by the editors November 25th, 2020; accepted March 24th, 2021.

2010 *Mathematics Subject Classification.* 62H20; 60B11, 62G20.

Key words and phrases. Distance covariance; random variables in metric spaces; consistency.

Supported by the Knut and Alice Wallenberg Foundation.

We denote the distance covariance by $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$, where $\alpha > 0$ is a parameter. The standard choice is $\alpha = 1$; in this case we may drop the subscript and write $\text{dcov}(\mathbf{X}, \mathbf{Y})$.

In the first part of the present paper, we consider general separable metric spaces and general $\alpha > 0$. One purpose is to improve existing result on consistency of dcov_α . Székely et al. (2007) showed that in the Euclidean case and with $\alpha \in (0, 2)$, computing dcov_α for the empirical distribution of a sample gives a strongly consistent estimator of dcov_α , provided α moments are finite. This was extended to general metric spaces, with $\alpha = 1$, by Lyons (2013), who claimed consistency in this sense assuming only finite first moments; however, the proof is incorrect as noted in the Errata to Lyons (2013). (For spaces of negative type, the result holds by another proof in Lyons, 2013.) As also noted in Lyons (2013), there is a simple proof assuming second moments, and Jakobsen (2017) proved the result when $\mathbb{E}(\|\mathbf{X}\|\|\mathbf{Y}\|)^{5/6} < \infty$, and thus in particular when \mathbf{X} and \mathbf{Y} have moments of order $5/3$. We remove this condition and show (Theorem 5.4) consistency assuming only first moments (as stated in Lyons, 2013). Furthermore, this is extended to all $\alpha > 0$, now assuming α^* moments with α^* defined in (2.1).

One interesting feature of distance covariance is that it can be defined in several ways that look very different but are equivalent (at least assuming sufficient moment conditions). We will in Section 2 give several definitions, requiring somewhat different moment conditions. Another purpose of the present paper is to show that the different definitions agree under the (minimal) moment conditions assumed in them (Section 4). We also show that dcov_α depends continuously on the distribution of (\mathbf{X}, \mathbf{Y}) , assuming convergence of appropriate moments (Theorem 5.2 and Remark 5.3).

In the second part of the paper, Sections 6–8, we consider Hilbert spaces. As said above, the original definition of distance covariance by Székely et al. (2007) is for the case of Euclidean spaces; their definition (Definition 2.4 below) uses characteristic functions and is thus tied to \mathbb{R}^p . For general separable Hilbert spaces, we give a related definition of dcov_α (Definition 7.1), which replaces the use of characteristic functions (i.e., Fourier transforms) by certain *characteristic random variables*, which are Gaussian random variables that can be defined also for variables in infinite-dimensional Hilbert spaces. We show that this definition is equivalent to the other ones under suitable moment conditions.

In the Euclidean setting in Feuerverger (1993) and Székely et al. (2007), with $\alpha < 2$, the original definition (our Definition 2.4) implies immediately the fundamental property that $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) \geq 0$ for any \mathbf{X} and \mathbf{Y} , and furthermore

$$\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) = 0 \iff \mathbf{X} \text{ and } \mathbf{Y} \text{ are independent.} \quad (1.1)$$

Hence, $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ can be regarded as a measure of dependency, and distance covariance can be used to test independence. (As noted in Székely et al., 2007, (1.1) does not hold for $\alpha = 2$; see Section 8.) Lyons (2013) showed that (1.1) does not hold for general metric spaces, but it holds for $\alpha = 1$ and metric spaces of *strong negative type* (see Lyons (2013) for the definition). In particular, Lyons (2013) showed that a Hilbert space is of strong negative type, and thus (1.1) holds for Hilbert spaces and $\alpha = 1$. Dehling et al. (2020, Theorem 4.2) extended this to all $\alpha \in (0, 2)$. We use our new definition of distance covariance for Hilbert spaces to give a new proof, assuming only α moments, of this theorem by Dehling et al. (2020) (our Theorem 7.6). Our proof (and Definition 7.1) is based on the ideas in Dehling et al. (2020); however, the proof in Dehling et al. (2020) is formulated for the Hilbert space $L^2[0, 1]$ and uses arguments with Brownian motion. Our proof can be regarded as a more abstract version of their proof, stated for arbitrary (separable) Hilbert spaces and using i.i.d. Gaussian sequences instead of Brownian motion; we believe that this makes the proof clearer since it avoids irrelevant details related to the particular choice $L^2[0, 1]$ of the Hilbert space.

Section 8 studies the case $\alpha = 2$ for Hilbert spaces. This case is rather trivial, and markedly different from $\alpha < 2$. In particular, even in one dimension, (1.1) does *not* hold for $\alpha = 2$, as is well known since Székely et al. (2007, §3.1) and Székely and Rizzo (2009a, §4.1). However, this case

is of special interest because, as said in Remark 2.7, and in more detail in Remark 8.4, arbitrary (semi)metric spaces of negative type can be embedded into it.

In the third part of the paper, we return to general metric spaces and study whether the moment conditions in the definitions and results are optimal. Section 9 shows that the exponents in the conditions cannot be decreased, in general. However, some other weakenings are possible, and in Section 10 we further study and compare the various definitions when the moment conditions above fail. We give some results; in particular, we consider Lorentz spaces. We also state some open problems that we have failed to solve.

The appendices contain some general results on uniform integrability and on integrals in a Hilbert space used in the paper; for completeness full proofs are given although some or all results are known.

2. Several definitions of distance covariance

We consider (except when stated otherwise) the general setting described in Section 1, with (\mathbf{X}, \mathbf{Y}) a pair of random variables taking values in $\mathcal{X} \times \mathcal{Y}$, where \mathcal{X} and \mathcal{Y} are separable metric spaces. We begin with three related definitions of distance covariance that work in this general setting, assuming only some moment conditions on the variables \mathbf{X} and \mathbf{Y} .

Let, throughout the paper, $(\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2), \dots$ be independent copies of (\mathbf{X}, \mathbf{Y}) . Also, let $\mathbf{x}_o \in \mathcal{X}$ and $\mathbf{y}_o \in \mathcal{Y}$ be two fixed points, and write for convenience $\|\mathbf{x}\| := d(\mathbf{x}, \mathbf{x}_o)$ and $\|\mathbf{y}\| := d(\mathbf{y}, \mathbf{y}_o)$ for $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$. (In the case of Euclidean spaces, or Hilbert spaces, we choose $\mathbf{x}_o = \mathbf{y}_o = 0$, and $\|\mathbf{x}\|$ is the usual norm.) We use \mathbf{x}_o and \mathbf{y}_o for moment conditions of the type $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$; note that by the triangle inequality, for this condition the choice of \mathbf{x}_o does not matter, and that this condition is equivalent to $\mathbb{E}d(\mathbf{X}_1, \mathbf{X}_2)^\alpha < \infty$.

Also, define for convenience

$$\alpha^* := \max(\alpha, 2\alpha - 2) = \begin{cases} \alpha, & 0 < \alpha \leq 2, \\ 2\alpha - 2, & \alpha > 2. \end{cases} \tag{2.1}$$

As will be seen below, the case of main interest is $\alpha \in (0, 2]$; in this case thus simply $\alpha^* = \alpha$.

When necessary, we distinguish the versions of distance covariance by different superscripts such as dcov_α^* , $\text{dcov}_\alpha^\wedge$, dcov_α^\sim , but usually this is omitted because the choice of definition does not matter, or is clear from the context.

Definition 2.1. Assume $\mathbb{E}\|\mathbf{X}\|^{2\alpha} < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^{2\alpha} < \infty$. Then

$$\begin{aligned} \text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) &= \text{dcov}_\alpha^*(\mathbf{X}, \mathbf{Y}) \\ &:= \mathbb{E}[d(\mathbf{X}_1, \mathbf{X}_2)^\alpha d(\mathbf{Y}_1, \mathbf{Y}_2)^\alpha] + \mathbb{E}[d(\mathbf{X}_1, \mathbf{X}_2)^\alpha] \mathbb{E}[d(\mathbf{Y}_1, \mathbf{Y}_2)^\alpha] \\ &\quad - 2 \mathbb{E}[d(\mathbf{X}_1, \mathbf{X}_2)^\alpha d(\mathbf{Y}_1, \mathbf{Y}_3)^\alpha]. \end{aligned} \tag{2.2}$$

Definition 2.2. Assume $\mathbb{E}\|\mathbf{X}\|^{\alpha^*} < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^{\alpha^*} < \infty$. Then

$$\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y}) := \frac{1}{4} \mathbb{E}[\widehat{X}_\alpha \widehat{Y}_\alpha], \tag{2.3}$$

where

$$\widehat{X}_\alpha := d(\mathbf{X}_1, \mathbf{X}_2)^\alpha - d(\mathbf{X}_2, \mathbf{X}_3)^\alpha + d(\mathbf{X}_3, \mathbf{X}_4)^\alpha - d(\mathbf{X}_4, \mathbf{X}_1)^\alpha \tag{2.4}$$

and similarly for \widehat{Y}_α .

Definition 2.3. Assume $\mathbb{E}\|\mathbf{X}\|^{\alpha^*} < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^{\alpha^*} < \infty$. Then

$$\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y}) := \mathbb{E}[\widetilde{X}_\alpha \widetilde{Y}_\alpha], \tag{2.5}$$

where

$$\tilde{X}_\alpha := \mathbb{E}(\hat{X}_\alpha \mid \mathbf{X}_1, \mathbf{X}_2) \quad (2.6)$$

$$= d(\mathbf{X}_1, \mathbf{X}_2)^\alpha - \mathbb{E}_{\mathbf{X}} d(\mathbf{X}_1, \mathbf{X})^\alpha - \mathbb{E}_{\mathbf{X}} d(\mathbf{X}_2, \mathbf{X})^\alpha + \mathbb{E} d(\mathbf{X}_1, \mathbf{X}_2)^\alpha \quad (2.7)$$

and similarly for \tilde{Y}_α , where $\mathbb{E}_{\mathbf{X}}$ denotes integrating over \mathbf{X} only, i.e., the conditional expectation given all \mathbf{X}_j (but not \mathbf{X}).

The role of the parameter α is thus to replace the metric d by d^α in the definition of $\text{dcov} = \text{dcov}_1$. See further Remark 2.7 below.

Note that $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ only depends on the joint distribution of \mathbf{X} and \mathbf{Y} ; thus distance covariance can be seen as a functional on distributions in $\mathcal{X} \times \mathcal{Y}$.

The moment condition $\mathbb{E}\|\mathbf{X}\|^{2\alpha} < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^{2\alpha} < \infty$ in Definition 2.1 is equivalent to $\mathbb{E}d(\mathbf{X}_1, \mathbf{X}_2)^{2\alpha} < \infty$ and $\mathbb{E}d(\mathbf{Y}_1, \mathbf{Y}_2)^{2\alpha} < \infty$, which implies that all expectations in (2.2) are finite; it implies also $\hat{X}_\alpha, \hat{Y}_\alpha \in L^2$ and thus $\tilde{X}_\alpha, \tilde{Y}_\alpha \in L^2$, so the expectations in (2.3) and (2.5) are also finite. Moreover, in this case, it is easy to see that Definitions 2.1–2.3 are equivalent: by expanding the products $\hat{X}_\alpha \hat{Y}_\alpha$ and $\tilde{X}_\alpha \tilde{Y}_\alpha$ in (2.3) and (2.5), we obtain (2.2) after simple calculations (Lemma 4.1). It is less obvious that the weaker moment condition in Definitions 2.2 and 2.3 is enough to guarantee that the expectations in (2.3) and (2.5) are finite and equal; we show this, and in particular that $\hat{X}_\alpha, \hat{Y}_\alpha, \tilde{X}_\alpha, \tilde{Y}_\alpha \in L^2$, in Section 4 (Lemma 4.3 and Theorem 4.5). In Section 9 we show that the exponents 2α and α^* in the moment conditions are optimal in general; in Section 10 we discuss extensions when the moment conditions fail.

The original definition of distance covariance by Székely et al. (2007), for random variables \mathbf{X} and \mathbf{Y} in Euclidean spaces \mathbb{R}^p and \mathbb{R}^q , see also Feuerverger (1993), is quite different and is based on characteristic functions. The general version with a $\alpha \in (0, 2)$ Székely et al. (2007, Section 3.1) is as follows.

Let $\varphi_{\mathbf{X}}(\mathbf{t}) := \mathbb{E}e^{i\mathbf{t}\cdot\mathbf{X}}$, $\varphi_{\mathbf{Y}}(\mathbf{u}) := \mathbb{E}e^{i\mathbf{u}\cdot\mathbf{Y}}$ and $\varphi_{\mathbf{X},\mathbf{Y}}(\mathbf{t}, \mathbf{u}) := \mathbb{E}e^{i(\mathbf{t}\cdot\mathbf{X}+\mathbf{u}\cdot\mathbf{Y})}$ be the characteristic functions of \mathbf{X} , \mathbf{Y} and (\mathbf{X}, \mathbf{Y}) . Define also the constants

$$c_{\alpha,k} := \frac{2^\alpha \Gamma((k+\alpha)/2)}{-\pi^{k/2} \Gamma(-\alpha/2)} = \frac{\alpha 2^{\alpha-1} \Gamma((k+\alpha)/2)}{\pi^{k/2} \Gamma(1-\alpha/2)} > 0. \quad (2.8)$$

(The values of these normalization constants are unimportant; they are chosen to make the definition agree with the preceding ones.)

Definition 2.4. Let (\mathbf{X}, \mathbf{Y}) be a pair of random vectors in \mathbb{R}^p and \mathbb{R}^q , respectively, where $p, q \geq 1$, and let $0 < \alpha < 2$. Then

$$\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^{\mathbb{E}}(\mathbf{X}, \mathbf{Y}) := c_{\alpha,p} c_{\alpha,q} \int_{\mathbf{t} \in \mathbb{R}^p} \int_{\mathbf{u} \in \mathbb{R}^q} |\varphi_{\mathbf{X},\mathbf{Y}}(\mathbf{t}, \mathbf{u}) - \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{Y}}(\mathbf{u})|^2 \frac{d\mathbf{t} d\mathbf{u}}{|\mathbf{t}|^{p+\alpha} |\mathbf{u}|^{q+\alpha}}. \quad (2.9)$$

Remark 2.5. No moment condition is needed in Definition 2.4, since the integrand in (2.9) is non-negative; with this definition (for Euclidean spaces and $\alpha < 2$), $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ is always defined, although it may be ∞ . As shown in Székely et al. (2007), $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ is finite at least when $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^\alpha < \infty$; this also follows from the equivalence with Definitions 2.2 and 2.3, see Theorems 7.2 and 7.4.

In contrast, we have in Definitions 2.1–2.3 imposed moment conditions making $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ finite. These definitions can be used somewhat more generally when the expectations in them are finite, and even when the result is $+\infty$; see Sections 9 and 10. However, without moment conditions, there are cases, even with $\mathbf{X} = \mathbf{Y} = \mathbb{R}$, when Definitions 2.1–2.3 yield results of the type $\infty - \infty$ and thus cannot be used at all; see Examples 9.4, 9.7, 9.9 and 9.15.

Remark 2.6. Definition 2.4 requires $\alpha < 2$, since typically the integral in (2.9) diverges for $\alpha \geq 2$. For example, if $p = q$ and $\mathbf{X} = \mathbf{Y} \sim N(0, I)$, then $|\varphi_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}, \mathbf{u}) - \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{Y}}(\mathbf{u})| \sim |\langle \mathbf{t}, \mathbf{u} \rangle|$ as $\mathbf{t}, \mathbf{u} \rightarrow 0$, and (2.9) diverges for $\alpha \geq 2$.

Feuerverger (1993) gave Definition 2.4 with $\alpha = 1$ for $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and the special case when (\mathbf{X}, \mathbf{Y}) have the empirical distribution of a finite sample from an unknown bivariate distribution, thus defining a test statistic for independence. He also showed that it has the equivalent forms (2.3) and (2.2). More generally, for arbitrary random (\mathbf{X}, \mathbf{Y}) in Euclidean spaces and $0 < \alpha < 2$, Székely et al. (2007) gave Definition 2.4; they also showed that it is equivalent to Definition 2.1 when the moment condition in the latter holds Székely et al. (2007, Remark 3 for $\alpha = 1$; implicit in §3.1 for $\alpha \in (0, 2)$); see also Székely and Rizzo (2009a, (3.7), (4.1) and Theorem 8). Furthermore, (2.5) was used for finite samples in Székely et al. (2007, (2.8) and §3.1) and Székely and Rizzo (2009a, (2.8) and §4.1). The name *distance covariance* was introduced by Székely et al. (2007) (for the case $\alpha = 1$, and α -*distance covariance* in general). (Actually, Székely et al., 2007 and Székely and Rizzo, 2009a define the distance covariance as the square root of $\text{dcov}(\mathbf{X}, \mathbf{Y})$; we ignore this difference in terminology.)

Lyons (2013) extended the theory to general (separable) metric spaces, with $\alpha = 1$, using Definition 2.3 as his definition. (This was also suggested in Székely and Rizzo, 2009b, §3.) Lyons (2013) showed also that, although the definition works for arbitrary metrics, dcov is useful as a measure of dependence mainly in the case when \mathcal{X} and \mathcal{Y} are metric spaces of *negative type* (see Lyons (2013) for a definition; see also Sejdinovic et al. (2013), Berg et al. (1984) and Remark 2.7 below), because in this case, but not otherwise, $\text{dcov}(\mathbf{X}, \mathbf{Y}) \geq 0$ for any \mathbf{X} and \mathbf{Y} such that $\text{dcov}(\mathbf{X}, \mathbf{Y})$ is defined; as said in the introduction, if furthermore the spaces are of strong negative type, then also (1.1) holds for $\alpha = 1$. (The implication that $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) = 0$ for independent variables is trivial, for any α , but not the converse.) Hence, for metric spaces of strong negative type, dcov can be regarded as a measure of dependence and for tests of independence just as in the Euclidean case.

Dehling et al. (2020) studied dcov_α in the infinite-dimensional Hilbert space $L^2[0, 1]$ for $\alpha \in (0, 2)$, using Definition 2.1. Since all separable infinite-dimensional Hilbert spaces are isomorphic; this is equivalent to considering arbitrary separable Hilbert spaces.

2.1. Further comments and extensions. We have here, as Lyons (2013), assumed that $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ are metrics. However, we can formally use Definitions 2.1–2.3 for any symmetric measurable functions $d_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$ and $d_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{Y} \rightarrow [0, \infty)$. (For \mathcal{X} and \mathcal{Y} such that the expectations exist, and still assuming \mathcal{X} and \mathcal{Y} to be separable metric spaces, to avoid technical problems.) It seems natural to assume at least that $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ are *semimetrics*; a semimetric on a space \mathcal{X} is a symmetric function $d : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$ such that $d(\mathbf{x}_1, \mathbf{x}_2) = 0 \iff \mathbf{x}_1 = \mathbf{x}_2$. (Thus, the triangle inequality is not assumed. Note that the term semimetric also is used in other context with a different meaning.) This extension was made by Sejdinovic et al. (2013); they considered semimetrics of negative type and showed that much of the theory extends to this case.

Remark 2.7. If $0 < \alpha \leq 1$, then d^α is also a metric for any metric d , and dcov_α is just dcov applied to the spaces \mathcal{X} and \mathcal{Y} equipped with the metrics $d_{\mathcal{X}}^\alpha$ and $d_{\mathcal{Y}}^\alpha$. (From an abstract point of view, the case $\alpha \leq 1$ thus does not add anything new.)

If we allow general semimetrics, there is no such restriction; d^α is a semimetric for every $\alpha > 0$, and dcov_α is just dcov applied to the semimetrics $d_{\mathcal{X}}^\alpha$ and $d_{\mathcal{Y}}^\alpha$ for any $\alpha > 0$.

On the other hand, see Lyons (2013) and Schoenberg (1938), a semimetric d on a space \mathcal{X} is of negative type if and only there exists an embedding $\varphi : \mathcal{X} \rightarrow \mathcal{H}$ into a Hilbert space such that

$$d(\mathbf{x}_1, \mathbf{x}_2) = \|\varphi(\mathbf{x}_1) - \varphi(\mathbf{x}_2)\|^2. \quad (2.10)$$

In particular, (2.10) implies that $d^{1/2}$ is a metric. (We assume that balls for the semimetric define the topology, and thus the metric $d^{1/2}$ defines the topology of \mathcal{X} .) Hence, for semimetrics of negative

type, dcov_α is the same as $\text{dcov}_{2\alpha}$ for the metrics $d_{\mathcal{X}}^{1/2}$ and $d_{\mathcal{Y}}^{1/2}$; in particular, dcov equals dcov_2 for these metrics. Consequently, our setting with metrics but arbitrary α includes also semimetrics of negative type. Furthermore, using the embedding φ , we see that dcov for semimetric spaces of negative type can be reduced to dcov_2 for Hilbert spaces, see Remark 8.4. (This is implicit in Sejdinovic et al., 2013, where this embedding is used to give another interpretation of distance covariance, see Remarks 2.10 and 8.5.)

We will in the sequel assume that $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ are metrics (without assuming negative type), but note that as just said, by changing α , this really includes the case of semimetrics of negative type.

In this context we note that if \mathcal{X} is a Euclidean space \mathbb{R}^q , or more generally a Hilbert space, then the semimetric $\|\mathbf{x}_1 - \mathbf{x}_2\|^\alpha$ is of negative type if and only if $0 < \alpha \leq 2$, see Schoenberg (1938). (It is thus a metric of negative type if and only if $0 < \alpha \leq 1$.) Consequently, for Hilbert spaces, if $0 < \alpha \leq 2$, we can conversely regard dcov_α as dcov for the semimetric of negative type $\|\mathbf{x}_1 - \mathbf{x}_2\|^\alpha$.

Remark 2.8. Another version of the definitions above is obtained if we denote the right-hand side of (2.4) by $\widehat{X}_\alpha(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$ and then define

$$\begin{aligned} \text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) &= \text{dcov}_\alpha^-(\mathbf{X}, \mathbf{Y}) \\ &:= \mathbb{E}(\widehat{X}_\alpha(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)\widehat{Y}_\alpha(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_5, \mathbf{Y}_6)). \end{aligned} \quad (2.11)$$

This version is used in proofs in Lyons (2013) and Jakobsen (2017).

It is obvious that if $\widehat{X}_\alpha, \widehat{Y}_\alpha \in L^2$, then the expectation in (2.11) is finite, and, using Fubini's theorem to integrate first over $\mathbf{X}_3, \mathbf{X}_4, \mathbf{Y}_5, \mathbf{Y}_6$, it equals $\mathbb{E}(\widetilde{X}_\alpha\widetilde{Y}_\alpha)$; thus, at least in this case, (2.11) agrees with (2.5). In particular, by Lemma 4.3 below, this holds when $\mathbb{E}\|\mathbf{X}\|^{\alpha^*} < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^{\alpha^*} < \infty$. We will not consider this definition further, and we leave the case when the moment condition just stated fails to the reader. (We conjecture results similar to those in Sections 9 and 10).

Remark 2.9. We have defined \widetilde{X}_α as a conditional expectation of \widehat{X}_α ; this can be regarded as an orthogonal projection in the Hilbert space $L^2(\mathbb{P})$.

If $\mathbb{E}\|\mathbf{X}\|^{2\alpha} < \infty$, so $d(\mathbf{X}_1, \mathbf{X}_2)^\alpha \in L^2$, then, as noted by Jakobsen (2017), \widetilde{X}_α can also be regarded as a projection in another way, viz. as the orthogonal projection of $d(\mathbf{X}_1, \mathbf{X}_2)^\alpha$ onto the subspace of $L^2(\mathbb{P})$ consisting of functions $g(\mathbf{X}_1, \mathbf{X}_2)$ with $\mathbb{E}(g(\mathbf{X}_1, \mathbf{X}_2) \mid \mathbf{X}_1) = \mathbb{E}(g(\mathbf{X}_1, \mathbf{X}_2) \mid \mathbf{X}_2) = 0$ a.s.

Remark 2.10. For semimetrics of negative type, another interpretation of distance covariance is given by Sejdinovic et al. (2013, Theorem 24), showing that it coincides with the *Hilbert-Schmidt independence criterion*, a distance measure between the distributions $\mathcal{L}(\mathbf{X}, \mathbf{Y})$ and $\mathcal{L}(\mathbf{X}_1, \mathbf{Y}_2) = \mathcal{L}(\mathbf{X}) \times \mathcal{L}(\mathbf{Y})$ that is defined using reproducing Hilbert spaces given by some kernels on the spaces, provided one chooses the kernels to be defined in a specific way by the metrics $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$. See also Remark 8.5.

Remark 2.11. Yet another interpretation (or definition) of distance covariance was given by Székely and Rizzo (2009a) for Euclidean spaces; it was called *Brownian covariance distance*. In the one-dimensional case $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, and with $\alpha = 1$, let W and W' be two two-sided Brownian motions, independent of each other and of \mathbf{X} and \mathbf{Y} ; then

$$\text{dcov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}[\text{Cov}(W(\mathbf{X}), W'(\mathbf{Y}) \mid W, W')^2] \quad (2.12)$$

This was extended, also in Székely and Rizzo (2009a), to arbitrary dimension by using Brownian fields on \mathbb{R}^k , and to $\alpha \in (0, 2)$ by using fractional Brownian fields.

This approach was further generalized to arbitrary spaces with semimetrics of negative type by Kanagawa et al. (2018, Section 6.4), letting W and W' be Gaussian stochastic processes on \mathcal{X} and \mathcal{Y} , with suitable covariance kernels.

Remark 2.12. Definitions 2.2–2.4 show immediately that $\text{dcov}_\alpha(\mathbf{X}, \mathbf{X}) \geq 0$ whenever the definition applies (even in the extended sense discussed in Remark 2.5). Moreover, $\text{dcov}_\alpha(\mathbf{X}, \mathbf{X}) > 0$ unless

\mathbf{X} is degenerate (i.e., is concentrated at a single value); this is immediate for Definition 2.4; it was shown by Lyons (2013) for Definition 2.3 (for $\alpha = 1$), and his proof extends to general α , and to Definition 2.2, for the latter even without any moment assumption (allowing $+\infty$).

Remark 2.13. Distance correlation is defined by Székely et al. (2007) as

$$\frac{\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})}{\text{dcov}_\alpha(\mathbf{X}, \mathbf{X})^{1/2} \text{dcov}_\alpha(\mathbf{Y}, \mathbf{Y})^{1/2}}, \tag{2.13}$$

provided \mathbf{X} and \mathbf{Y} are non-degenerate so the denominator is strictly positive (see Remark 2.12).

Various properties of distance correlation follow from properties of distance covariance; we leave this to the reader.

3. Some notation

As said above, (\mathbf{X}, \mathbf{Y}) is a pair of random variables taking values in separable metric spaces \mathcal{X} and \mathcal{Y} , and $(\mathbf{X}_i, \mathbf{Y}_i), i \geq 1$, are independent copies of (\mathbf{X}, \mathbf{Y}) . α is a fixed parameter, and α^* is given by (2.1). Unless stated otherwise, we assume only $\alpha > 0$. (This condition is sometimes repeated for emphasis.)

$\mathcal{P}(\mathcal{X})$ denotes the set of all Borel probability measures in \mathcal{X} .

Convergence almost surely, in probability, in distribution and in L^p are denoted by $\xrightarrow{\text{a.s.}}$, $\xrightarrow{\text{P}}$, $\xrightarrow{\text{d}}$, $\xrightarrow{L^p}$.

We use the standard definition of covariance

$$\text{Cov}(Z, W) := \mathbb{E}[ZW] - \mathbb{E}Z \mathbb{E}W \tag{3.1}$$

not only for real random variables, but also more generally for any complex random variables Z and W with $\mathbb{E}|Z|^2, \mathbb{E}|W|^2 < \infty$; we further extend this notation to conditional covariance.

For real $x, y, x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$; also $x_+ := x \vee 0$ and $x_- := (-x)_+ = -(x \wedge 0)$, so $x = x_+ - x_-$.

The inner product in a Hilbert space is denoted by $\langle x, y \rangle$; for finite-dimensional \mathbb{R}^q we also use $x \cdot y$. All Hilbert spaces have real scalars, so the inner product is real-valued.

C and c will denote some unimportant positive constants that depend only on α (and may be taken as universal constants for $\alpha \leq 2$). Their value may differ from one occurrence to the next.

4. Existence and continuity

We begin by recording the simple fact that with enough moments, Definitions 2.1–2.3 agree.

Lemma 4.1. *Let $\alpha > 0$. If $\mathbb{E}\|\mathbf{X}\|^{2\alpha} < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^{2\alpha} < \infty$, then all expectations in (2.2), (2.3) and (2.5) are finite, and the three definitions of $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ agree, i.e., $\text{dcov}_\alpha^*(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$.*

Proof: As said in Section 2, this is elementary; we omit the details. □

We will extend this to the weaker moment conditions used in Definitions 2.2 and 2.3. We argue similarly to Lyons (2013), who showed the case $\alpha = 1$ (and thus implicitly $0 < \alpha \leq 1$, see Remark 2.7). We first show some useful estimates of the variable \widehat{X}_α defined in (2.4). Note the symmetry up to sign under cyclic permutations of the indices $1, \dots, 4$.

Although we state the next lemma for the random variables \mathbf{X}_i , it is really a pointwise inequality that could have been stated for four non-random points $\mathbf{x}_1, \dots, \mathbf{x}_4$. In sums such as (4.2) and (4.3), the indices are interpreted modulo 4; moreover, a term containing an index $i \pm 1$ should be interpreted as two terms, with $i + 1$ and $i - 1$; the sum in (4.3) is thus really a sum of 8 terms.

Lemma 4.2. *Let \mathcal{X} be a metric space.*

(i) *If $0 < \alpha \leq 1$, then*

$$|\widehat{X}_\alpha| \leq 2 \sum_{i=1}^4 (\|\mathbf{X}_i\|^\alpha \wedge \|\mathbf{X}_{i+1}\|^\alpha). \tag{4.1}$$

(ii) *If $0 < \alpha \leq 2$, then*

$$|\widehat{X}_\alpha| \leq C \sum_{i=1}^4 \|\mathbf{X}_i\|^{\alpha/2} \|\mathbf{X}_{i+1}\|^{\alpha/2}. \tag{4.2}$$

(iii) *If $\alpha \geq 1$, then*

$$|\widehat{X}_\alpha| \leq C \sum_{i=1}^4 \|\mathbf{X}_i\|^{\alpha-1} \|\mathbf{X}_{i\pm 1}\|. \tag{4.3}$$

Proof: Write $d_{ij} := d(\mathbf{X}_i, \mathbf{X}_j)$. Thus $\widehat{X}_\alpha = d_{12}^\alpha - d_{23}^\alpha + d_{34}^\alpha - d_{41}^\alpha$. Note the triangle inequality

$$d_{ij} \leq \|\mathbf{X}_i\| + \|\mathbf{X}_j\|. \tag{4.4}$$

Case 1: $\alpha \leq 1$. Since d^α is a metric when $\alpha \leq 1$, it suffices to consider the case $\alpha = 1$. The triangle inequality yields

$$|\widehat{X}| \leq |d_{12} - d_{41}| + |d_{23} - d_{34}| \leq d_{24} + d_{24} = 2d_{24}. \tag{4.5}$$

Similarly, by shifting the indices,

$$|\widehat{X}| \leq 2d_{13}. \tag{4.6}$$

Hence, using (4.5)–(4.6) and (4.4),

$$|\widehat{X}| \leq 2 \min(d_{13}, d_{24}) \leq 2 \min(\|\mathbf{X}_1\| + \|\mathbf{X}_3\|, \|\mathbf{X}_2\| + \|\mathbf{X}_4\|). \tag{4.7}$$

We claim that for any real $x_1, \dots, x_4 \geq 0$,

$$(x_1 + x_3) \wedge (x_2 + x_4) \leq \sum_{i=1}^4 (x_i \wedge x_{i+1}). \tag{4.8}$$

In fact, by cyclic symmetry, we may without loss of generality assume that x_1 is the largest of x_1, \dots, x_4 , and in this case

$$x_2 + x_4 = x_1 \wedge x_2 + x_4 \wedge x_1 \leq \sum_{i=1}^4 (x_i \wedge x_{i+1}), \tag{4.9}$$

and (4.8) follows. Hence (4.8) holds, and (4.7) implies (4.1) for $\alpha = 1$. As said above, this shows (4.1) in general.

Furthermore, for $\alpha \leq 1$, (4.2) follows from (4.1) since $x \wedge y \leq x^{1/2}y^{1/2}$ when $x, y \geq 0$.

Case 2: $\alpha > 1$. By the cyclic symmetry we may assume that $\|\mathbf{X}_1\|$ is the largest of $\|\mathbf{X}_1\|, \dots, \|\mathbf{X}_4\|$. Then, (4.4) implies

$$d_{ij} \leq 2\|\mathbf{X}_1\|, \quad i, j = 1, \dots, 4. \tag{4.10}$$

As above, the triangle inequality yields

$$|d_{12} - d_{41}| \leq d_{24} \tag{4.11}$$

and thus, by the mean value theorem, for some $\theta \in [0, 1]$,

$$|d_{12}^\alpha - d_{41}^\alpha| \leq d_{24} \alpha (\theta d_{12} + (1 - \theta) d_{41})^{\alpha-1}. \tag{4.12}$$

Using (4.10), this yields

$$|d_{12}^\alpha - d_{41}^\alpha| \leq d_{24}\alpha 2^{\alpha-1} \|\mathbf{X}_1\|^{\alpha-1}. \tag{4.13}$$

Similarly,

$$|d_{23}^\alpha - d_{34}^\alpha| \leq d_{24}\alpha(\theta'd_{23} + (1 - \theta')d_{34})^{\alpha-1} \leq d_{24}\alpha 2^{\alpha-1} \|\mathbf{X}_1\|^{\alpha-1}. \tag{4.14}$$

Summing (4.13) and (4.14) yields, using again (4.4),

$$\begin{aligned} |\widehat{X}_\alpha| &\leq |d_{12}^\alpha - d_{41}^\alpha| + |d_{23}^\alpha - d_{34}^\alpha| \leq \alpha 2^\alpha \|\mathbf{X}_1\|^{\alpha-1} d_{24} \\ &\leq \alpha 2^\alpha \|\mathbf{X}_1\|^{\alpha-1} (\|\mathbf{X}_2\| + \|\mathbf{X}_4\|). \end{aligned} \tag{4.15}$$

This proves (4.3) for any $\alpha \geq 1$.

If $1 \leq \alpha \leq 2$, we further note that our assumption $\|\mathbf{X}_j\| \leq \|\mathbf{X}_1\|$ implies

$$\|\mathbf{X}_1\|^{\alpha-1} \|\mathbf{X}_j\| \leq \|\mathbf{X}_1\|^{\alpha/2} \|\mathbf{X}_j\|^{\alpha/2}, \quad j = 1, \dots, 4, \tag{4.16}$$

and thus (4.15) also yields (4.2). □

Lemma 4.3. *If $\mathbb{E}\|\mathbf{X}\|^{\alpha^*} < \infty$, then $\mathbb{E}\widehat{X}_\alpha^2 < \infty$ and $\mathbb{E}\widetilde{X}_\alpha^2 < \infty$.*

For $\alpha = 1$, this is shown by Lyons (2013, Errata).

Proof: Case 1: $\alpha \leq 2$. In this case $\alpha^* = \alpha$. Recall that, by definition, \mathbf{X}_i and $\mathbf{X}_{i\pm 1}$ are independent. Hence,

$$\mathbb{E}(\|\mathbf{X}_i\|^{\alpha/2} \|\mathbf{X}_{i+1}\|^{\alpha/2})^2 = \mathbb{E}\|\mathbf{X}_i\|^\alpha \mathbb{E}\|\mathbf{X}_{i+1}\|^\alpha < \infty, \tag{4.17}$$

so each term in the sum in (4.2) belongs to L^2 , and thus (4.2) implies $\widehat{X}_\alpha \in L^2$. Since \widetilde{X}_α is defined by (2.6) as a conditional expectation of \widehat{X}_α , this further implies $\widetilde{X}_\alpha \in L^2$.

Case 2: $\alpha \geq 2$. In this case $\alpha^* = 2(\alpha - 1) \geq 2$, and the result follows in the same way from (4.3). □

In the following lemma, we consider together with \mathbf{X} also a sequences $(\mathbf{X}^{(n)})_{n \geq 1}$ of random variables in \mathcal{X} . We then define $\mathbf{X}_i^{(n)}$ for $i \geq 1$ such that the random variables $(\mathbf{X}_i, (\mathbf{X}_i^{(n)})_n)$ in \mathcal{X}^∞ are independent copies of $(\mathbf{X}, (\mathbf{X}^{(n)})_n)$. This extends in the obvious way when we consider sequences $((\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}))_n$. We use the superscript (n) in the natural way and let e.g. $\widehat{X}_\alpha^{(n)}$ be defined as in (2.4) using $\mathbf{X}_i^{(n)}$.

Lemma 4.4. *Let \mathbf{X} and $\mathbf{X}^{(n)}$, $n \geq 1$, be random variables in \mathcal{X} , and assume that $\mathbb{E}\|\mathbf{X}\|^{\alpha^*} < \infty$ and $\mathbb{E}d(\mathbf{X}^{(n)}, \mathbf{X})^{\alpha^*} \rightarrow 0$ as $n \rightarrow \infty$. Then $\mathbb{E}(\widehat{X}_\alpha^{(n)} - \widehat{X}_\alpha)^2 \rightarrow 0$ and $\mathbb{E}(\widetilde{X}_\alpha^{(n)} - \widetilde{X}_\alpha)^2 \rightarrow 0$.*

Proof: We use without further comments some elementary facts about uniform integrability, see e.g. Gut (2013, Theorems 5.5.4, 5.4.5 and 5.4.6).

Since $\mathbb{E}d(\mathbf{X}^{(n)}, \mathbf{X})^{\alpha^*} \rightarrow 0$, the sequence $d(\mathbf{X}^{(n)}, \mathbf{X})^{\alpha^*}$ of random variables is uniformly integrable. The triangle inequality yields $\|\mathbf{X}^{(n)}\| \leq d(\mathbf{X}^{(n)}, \mathbf{X}) + \|\mathbf{X}\|$, and thus

$$\|\mathbf{X}^{(n)}\|^{\alpha^*} \leq C(d(\mathbf{X}^{(n)}, \mathbf{X})^{\alpha^*} + \|\mathbf{X}\|^{\alpha^*}), \tag{4.18}$$

and it follows that the sequence $\|\mathbf{X}^{(n)}\|^{\alpha^*}$ is uniformly integrable. Lemma 4.2 and the argument in the proof of Lemma 4.3, using Lemma A.1 in the appendix, show that the sequence $(\widehat{X}_\alpha^{(n)})^2$ is uniformly integrable.

Furthermore, we have $d(\mathbf{X}^{(n)}, \mathbf{X}) \xrightarrow{P} 0$, and thus $d(\mathbf{X}_i^{(n)}, \mathbf{X}_i) \xrightarrow{P} 0$ for every i . The triangle inequality then implies $d(\mathbf{X}_i^{(n)}, \mathbf{X}_j^{(n)}) \xrightarrow{P} d(\mathbf{X}_i, \mathbf{X}_j)$ for every i and j , and thus the definition (2.4) implies $\widehat{X}_\alpha^{(n)} \xrightarrow{P} \widehat{X}_\alpha$.

This and the uniform square integrability just established yield $\mathbb{E}(\widehat{X}_\alpha^{(n)} - \widehat{X}_\alpha)^2 \rightarrow 0$.

Furthermore, by (2.6), if \mathcal{F} is the σ -field generated by all \mathbf{X}_j and $\mathbf{X}_j^{(n)}$ with $j \in \{1, 2\}$, then $\tilde{X}_\alpha = \mathbb{E}(\hat{X}_\alpha \mid \mathcal{F})$ and $\tilde{X}_\alpha^{(n)} = \mathbb{E}(\hat{X}_\alpha^{(n)} \mid \mathcal{F})$. Consequently,

$$\mathbb{E}|\tilde{X}_\alpha^{(n)} - \tilde{X}_\alpha|^2 = \mathbb{E}|\mathbb{E}(\hat{X}_\alpha^{(n)} - \hat{X}_\alpha \mid \mathcal{F})|^2 \leq \mathbb{E}|\hat{X}_\alpha^{(n)} - \hat{X}_\alpha|^2 \rightarrow 0. \tag{4.19}$$

□

Theorem 4.5. *Definitions 2.1–2.3 are well-defined; more precisely, for any $\alpha > 0$, assuming the stated moment conditions, the expectations in (2.2), (2.3) and (2.5) are finite. Furthermore, any two of these definitions yield the same result, whenever the moment conditions in both are satisfied.*

Proof: Lemma 4.1 shows that all three definitions are valid and agree under the condition of Definition 2.1, i.e., when $\mathbb{E}\|\mathbf{X}\|^{2\alpha} < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^{2\alpha} < \infty$.

It remains to show that (2.3) and (2.5) are finite and agree under the weaker assumption $\mathbb{E}\|\mathbf{X}\|^{\alpha^*} < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^{\alpha^*} < \infty$. In this case, Lemma 4.3 shows that $\hat{X}_\alpha, \hat{Y}_\alpha, \tilde{X}_\alpha, \tilde{Y}_\alpha \in L^2$, and thus (2.3) and (2.5) are finite.

We do not know a simple direct argument to show the equality of the two expressions, so we use truncations as follows. Let, for $n \geq 1$,

$$\mathbf{X}^{(n)} := \begin{cases} \mathbf{X}, & \|\mathbf{X}\| \leq n, \\ \mathbf{x}_o, & \text{otherwise,} \end{cases} \tag{4.20}$$

and define $\mathbf{Y}^{(n)}$ similarly. Then

$$\mathbb{E}d(\mathbf{X}^{(n)}, \mathbf{X})^{\alpha^*} = \mathbb{E}[\|\mathbf{X}\|^{\alpha^*} \mathbf{1}\{\|\mathbf{X}\| > n\}] \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \rightarrow \infty. \tag{4.21}$$

Thus, Lemma 4.4 yields $\|\hat{X}_\alpha^{(n)} - \hat{X}_\alpha\|_{L^2} \rightarrow 0$ and $\|\tilde{X}_\alpha^{(n)} - \tilde{X}_\alpha\|_{L^2} \rightarrow 0$. Similarly, $\|\hat{Y}_\alpha^{(n)} - \hat{Y}_\alpha\|_{L^2} \rightarrow 0$ and $\|\tilde{Y}_\alpha^{(n)} - \tilde{Y}_\alpha\|_{L^2} \rightarrow 0$.

The L^2 -convergence just shown implies that, as $n \rightarrow \infty$,

$$\text{dcov}_\alpha(\hat{\mathbf{X}}^{(n)}, \hat{\mathbf{Y}}^{(n)}) = \frac{1}{4} \mathbb{E}[\hat{X}_\alpha^{(n)} \hat{Y}_\alpha^{(n)}] \rightarrow \frac{1}{4} \mathbb{E}[\hat{X}_\alpha \hat{Y}_\alpha] = \text{dcov}_\alpha(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) \tag{4.22}$$

and similarly

$$\text{dcov}_\alpha(\tilde{\mathbf{X}}^{(n)}, \tilde{\mathbf{Y}}^{(n)}) = \mathbb{E}[\hat{X}_\alpha^{(n)} \hat{Y}_\alpha^{(n)}] \rightarrow \mathbb{E}[\hat{X}_\alpha \hat{Y}_\alpha] = \text{dcov}_\alpha(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}), \tag{4.23}$$

Furthermore, for each n , $\|\mathbf{X}^{(n)}\|$ and $\|\mathbf{Y}^{(n)}\|$ are bounded, and thus Lemma 4.1 applies and shows $\text{dcov}_\alpha(\hat{\mathbf{X}}^{(n)}, \hat{\mathbf{Y}}^{(n)}) = \text{dcov}_\alpha(\tilde{\mathbf{X}}^{(n)}, \tilde{\mathbf{Y}}^{(n)})$. Consequently, (4.22)–(4.23) imply $\text{dcov}_\alpha(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) = \text{dcov}_\alpha(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$. □

We return in Section 9 to the case when the moment conditions fail.

5. Continuity and consistency

The lemmas in Section 4 yield also continuity results. Unspecified convergence is as $n \rightarrow \infty$.

Theorem 5.1. *Let $\alpha > 0$. Let (\mathbf{X}, \mathbf{Y}) and $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)})$, $n \geq 1$, be pairs of random variables in $\mathcal{X} \times \mathcal{Y}$, and assume that $\mathbb{E}\|\mathbf{X}\|^{\alpha^*} < \infty$, $\mathbb{E}\|\mathbf{Y}\|^{\alpha^*} < \infty$ and, as $n \rightarrow \infty$, $\mathbb{E}d(\mathbf{X}^{(n)}, \mathbf{X})^{\alpha^*} \rightarrow 0$ and $\mathbb{E}d(\mathbf{Y}^{(n)}, \mathbf{Y})^{\alpha^*} \rightarrow 0$. Then,*

$$\text{dcov}_\alpha(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}) \rightarrow \text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}). \tag{5.1}$$

Proof: Lemma 4.4 yields $\hat{X}_\alpha^{(n)} \xrightarrow{L^2} \hat{X}_\alpha$ and $\hat{Y}_\alpha^{(n)} \xrightarrow{L^2} \hat{Y}_\alpha$, and thus

$$\text{dcov}_\alpha(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}) = \frac{1}{4} \mathbb{E}[\hat{X}_\alpha^{(n)} \hat{Y}_\alpha^{(n)}] \rightarrow \frac{1}{4} \mathbb{E}[\hat{X}_\alpha \hat{Y}_\alpha] = \text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}). \tag{5.2}$$

□

We can extend this result and assume only convergence in distribution of $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)})$ together with a moment condition.

Theorem 5.2. *Let $\alpha > 0$. Let (\mathbf{X}, \mathbf{Y}) and $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)})$, $n \geq 1$, be pairs of random variables in $\mathcal{X} \times \mathcal{Y}$, and assume that, as $n \rightarrow \infty$, $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}) \xrightarrow{d} (\mathbf{X}, \mathbf{Y})$. Assume further one of the following two conditions.*

- (i) *The sequences $\|\mathbf{X}^{(n)}\|^{\alpha^*}$ and $\|\mathbf{Y}^{(n)}\|^{\alpha^*}$ are uniformly integrable.*
- (ii) *$\mathbb{E}\|\mathbf{X}^{(n)}\|^{\alpha^*} \rightarrow \mathbb{E}\|\mathbf{X}\|^{\alpha^*} < \infty$ and $\mathbb{E}\|\mathbf{Y}^{(n)}\|^{\alpha^*} \rightarrow \mathbb{E}\|\mathbf{Y}\|^{\alpha^*} < \infty$.*

Then,

$$\text{dcov}_\alpha(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}) \rightarrow \text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}). \tag{5.3}$$

Proof: (i): Since $\mathcal{X} \times \mathcal{Y}$ is a separable metric space, we may by the Skorohod coupling theorem [Kallenberg \(2002, Theorem 4.30\)](#) without loss of generality assume that $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}) \xrightarrow{\text{a.s.}} (\mathbf{X}, \mathbf{Y})$. Furthermore, the assumption in (i) implies that $\sup_n \mathbb{E}\|\mathbf{X}^{(n)}\|^{\alpha^*} < \infty$, and thus $\mathbb{E}\|\mathbf{X}\|^{\alpha^*} < \infty$ by Fatou’s lemma. Since $d(\mathbf{X}^{(n)}, \mathbf{X}) \leq \|\mathbf{X}^{(n)}\| + \|\mathbf{X}\|$, it follows, similarly to (4.18), that the sequence $d(\mathbf{X}^{(n)}, \mathbf{X})^{\alpha^*}$ is uniformly integrable. Since we have assumed $d(\mathbf{X}^{(n)}, \mathbf{X}) \xrightarrow{\text{a.s.}} 0$, this implies $\mathbb{E}d(\mathbf{X}^{(n)}, \mathbf{X})^{\alpha^*} \rightarrow 0$. Similarly, $\mathbb{E}d(\mathbf{Y}^{(n)}, \mathbf{Y})^{\alpha^*} \rightarrow 0$. Thus [Theorem 5.1](#) applies and yields (5.3).

(ii): We have $\mathbf{X}^{(n)} \xrightarrow{d} \mathbf{X}$ and thus $\|\mathbf{X}^{(n)}\| \xrightarrow{d} \|\mathbf{X}\|$. This and our assumption $\mathbb{E}\|\mathbf{X}^{(n)}\|^{\alpha^*} \rightarrow \mathbb{E}\|\mathbf{X}\|^{\alpha^*}$ imply that the sequence $\|\mathbf{X}^{(n)}\|^{\alpha^*}$ is uniformly integrable [Gut \(2013, Theorem 5.5.9\)](#). The same holds for $\mathbf{Y}^{(n)}$, and thus part (i) applies. \square

Remark 5.3. Suppose that the metric spaces \mathcal{X} and \mathcal{Y} are complete. (This ensures that all probability measures are tight; see e.g. [Billingsley, 1968](#).) Give $\mathcal{X} \times \mathcal{Y}$ the metric (for example)

$$d((\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2)) := d_{\mathcal{X}}(\mathbf{x}_1, \mathbf{x}_2) + d_{\mathcal{Y}}(\mathbf{y}_1, \mathbf{y}_2). \tag{5.4}$$

Let $\mathcal{P}^\alpha(\mathcal{X} \times \mathcal{Y})$ be the space of all Borel probability measures μ on $\mathcal{X} \times \mathcal{Y}$ such that $\int_{\mathcal{X} \times \mathcal{Y}} \|(\mathbf{x}, \mathbf{y})\|^\alpha d\mu(\mathbf{x}, \mathbf{y}) < \infty$. In other words, $\mathcal{P}^\alpha(\mathcal{X} \times \mathcal{Y})$ is the space of all distributions of pairs of random variables $(\mathbf{X}, \mathbf{Y}) \in \mathcal{X} \times \mathcal{Y}$ such that $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^\alpha < \infty$.

Define a metric in $\mathcal{P}^\alpha(\mathcal{X} \times \mathcal{Y})$ by

$$d_\alpha(\mu, \mu') := \begin{cases} \inf\{\mathbb{E}[d((\mathbf{X}, \mathbf{Y}), (\mathbf{X}', \mathbf{Y}'))^\alpha]\}, & 0 < \alpha \leq 1, \\ \inf\{\mathbb{E}[d((\mathbf{X}, \mathbf{Y}), (\mathbf{X}', \mathbf{Y}'))^\alpha]^{1/\alpha}\}, & \alpha > 1. \end{cases} \tag{5.5}$$

taking the infimum over all pairs of random variables (\mathbf{X}, \mathbf{Y}) and $(\mathbf{X}', \mathbf{Y}')$ in $\mathcal{X} \times \mathcal{Y}$ such that $(\mathbf{X}, \mathbf{Y}) \sim \mu$ and $(\mathbf{X}', \mathbf{Y}') \sim \mu'$; see e.g. [Bogachev and Kolesnikov \(2012, pp. 796–799\)](#) (in the English translation). (This is known under various names, including *Kantorovich distance*, *Wasserstein distance* and *minimal L^α distance*, see also [Rüschendorf, 2020](#)). Convergence of a sequence $\mathcal{L}(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)})$ of distributions to $\mathcal{L}(\mathbf{X}, \mathbf{Y})$ in this metric is equivalent to convergence in distribution $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}) \xrightarrow{d} (\mathbf{X}, \mathbf{Y})$ (i.e., weak convergence of the distributions) together with uniform integrability of $\|\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}\|^{\alpha^*}$ (or, equivalently, convergence of moments $\mathbb{E}\|(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)})\|^{\alpha^*} \rightarrow \mathbb{E}\|(\mathbf{X}, \mathbf{Y})\|^{\alpha^*}$).

[Theorem 5.1](#) then says that dcov_α is a continuous functional on $\mathcal{P}^{\alpha^*}(\mathcal{X} \times \mathcal{Y})$, for every $\alpha > 0$.

5.1. *Consistency.* Let $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be the distribution of (\mathbf{X}, \mathbf{Y}) . Then, $(\mathbf{X}_1, \mathbf{Y}_1), \dots$ can be regarded as a sequence of independent samples from μ . Let ν_n be the empirical distribution of the first n samples, i.e.,

$$\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{(\mathbf{X}_i, \mathbf{Y}_i)} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}). \tag{5.6}$$

Note that ν_n is a random probability measure. Hence, its distance covariance $\text{dcov}_\alpha(\nu_n)$ is a random variable. The following theorem shows that this random variable converges to $\text{dcov}_\alpha(\mu)$ a.s.; in other words, the distance covariance of the empirical distribution is a consistent estimator of the covariance distance of μ . As said in the introduction, this was proved by Székely et al. (2007) for the Euclidean case with $\alpha \in (0, 2)$; for general metric spaces, with $\alpha = 1$, the result was stated by Lyons (2013), but his proof requires a stronger moment condition. Second moments are enough for $\alpha = 1$, see Székely et al. (2007, Remark 3); Jakobsen (2017) improved this and showed that $5/3$ moments are enough. The proof in Székely et al. (2007, Remark 3) generalizes to arbitrary $\alpha > 0$, assuming 2α moments.

We can now show consistency assuming only α^* moments, as required by our definitions. In particular, this shows that for $\alpha = 1$, first moments suffice, as stated in Lyons (2013). (The proofs in Lyons (2013) and Jakobsen (2017) use results for V -statistics which require extra moment assumptions; our proof uses another method.)

Theorem 5.4. *Let μ be the distribution of $(\mathbf{X}, \mathbf{Y}) \in \mathcal{X} \times \mathcal{Y}$ and assume that $\mathbb{E}\|\mathbf{X}\|^{\alpha^*}, \mathbb{E}\|\mathbf{Y}\|^{\alpha^*} < \infty$. If ν_n is the empirical distribution (5.6), then*

$$\text{dcov}_\alpha(\nu_n) \xrightarrow{\text{a.s.}} \text{dcov}_\alpha(\mu). \quad (5.7)$$

Proof: Conditionally on the sequence $(\nu_n)_n$ of empirical measures, let $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)})$ be a random variable with distribution ν_n . Since $\mathcal{X} \times \mathcal{Y}$ is a separable metric space, the distribution ν_n converges a.s. to μ (in the usual weak topology); see Varadarajan (1958) or Billingsley (1968, Problem 4.4). In other words, a.s., conditionally on $(\nu_n)_n$, $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}) \xrightarrow{\text{d}} (\mathbf{X}, \mathbf{Y})$.

Furthermore, by the definition (5.6) of ν_n , conditioning on the sequence $(\nu_k)_k$,

$$\mathbb{E}(\|\mathbf{X}^{(n)}\|^{\alpha^*} \mid (\nu_k)_k) = \frac{1}{n} \sum_{i=1}^n \|\mathbf{X}_i\|^{\alpha^*}. \quad (5.8)$$

Hence, the strong law of large numbers (in \mathbb{R}) shows that a.s., conditioned on $(\nu_k)_k$, $\mathbb{E}\|\mathbf{X}^{(n)}\|^{\alpha^*} \xrightarrow{\text{a.s.}} \mathbb{E}\|\mathbf{X}\|^{\alpha^*}$, and similarly also $\mathbb{E}\|\mathbf{Y}^{(n)}\|^{\alpha^*} \xrightarrow{\text{a.s.}} \mathbb{E}\|\mathbf{Y}\|^{\alpha^*}$. Consequently, Theorem 5.2(ii) applies a.s. to the sequence $(\nu_n)_n$ and the corresponding random variables $(\mathbf{X}^{(n)}, \mathbf{Y}^{(n)})$; hence $\text{dcov}_\alpha(\nu_n) \xrightarrow{\text{a.s.}} \text{dcov}_\alpha(\mu)$. \square

Our proofs of Theorems 5.2 and 5.4 give no information on the rate of convergence, leading to the following problems.

Problem 5.5. *What is the rate of convergence in (5.3), under suitable hypotheses on $(\mathbf{X}_n, \mathbf{Y}_n)$?*

Problem 5.6. *What is the rate of convergence in (5.7), under suitable hypotheses on (\mathbf{X}, \mathbf{Y}) ?*

6. Hilbert spaces, preliminaries

In this and the next two sections we assume that \mathcal{X} and \mathcal{Y} are separable Hilbert spaces; we therefore change notation and write $\mathcal{X} = \mathcal{H}$ and $\mathcal{Y} = \mathcal{H}'$.

We give our extension of Definition 2.4 of covariance distance in Section 7, but we first need some preliminaries.

6.1. *Characteristic random variables.* Let \mathcal{H} be a separable Hilbert space, of finite or infinite dimension $\dim \mathcal{H}$.

Fix an ON-basis $(\mathbf{e}_i)_1^{\dim \mathcal{H}}$ in \mathcal{H} , and let ξ_i , $i = 1, 2, \dots$, be i.i.d. $N(0, 1)$ random variables. Let $\boldsymbol{\xi} := (\xi_i)_1^{\dim \mathcal{H}}$, a random vector of length $\dim \mathcal{H}$ (finite or infinite). Define for any $\mathbf{x} \in \mathcal{H}$,

$$\boldsymbol{\xi} \cdot \mathbf{x} = \mathbf{x} \cdot \boldsymbol{\xi} := \sum_{i=1}^{\dim \mathcal{H}} \langle \mathbf{x}, \mathbf{e}_i \rangle \xi_i. \quad (6.1)$$

Note that in the finite-dimensional case, $\xi \in \mathcal{H}$ and this is the usual inner product. In the infinite-dimensional case $\xi \notin \mathcal{H}$ a.s., but the sum in (6.1) converges a.s. since $\sum_i |\langle \mathbf{x}, \mathbf{e}_i \rangle|^2 = \|\mathbf{x}\|^2 < \infty$. Hence, $\xi \cdot \mathbf{x}$ is defined a.s. in any case. Note also that $\xi \cdot \mathbf{x}$ is a real-valued random variable, and that

$$\xi \cdot \mathbf{x} \sim N(0, \|\mathbf{x}\|^2). \tag{6.2}$$

Let \mathbf{X} be an \mathcal{H} -valued random variable, and assume that ξ is independent of \mathbf{X} . Then $\xi \cdot \mathbf{X}$ exists a.s.; thus $\xi \cdot \mathbf{X}$ is a well-defined real-valued random variable. Consider the conditional expectation

$$\Phi_{\mathbf{X}}(\xi) := \mathbb{E}(e^{i\xi \cdot \mathbf{X}} \mid \xi). \tag{6.3}$$

This is a complex-valued random variable (determined a.s.), which can be written as a (deterministic) function of ξ .

In the finite-dimensional case $\dim \mathcal{H} < \infty$, we may identify \mathcal{H} with \mathbb{R}^q , with $(\mathbf{e}_j)_1^q$ as the standard basis. Then (6.1) and (6.3) show that

$$\Phi_{\mathbf{X}}(\xi) = \varphi_{\mathbf{X}}(\xi) \quad \text{a.s.}, \tag{6.4}$$

where $\varphi_{\mathbf{X}}(\mathbf{t}) := \mathbb{E} e^{i\mathbf{t} \cdot \mathbf{X}}$ is the usual characteristic function. For this reason, we say, for a general Hilbert space \mathcal{H} , that $\Phi_{\mathbf{X}}(\xi)$ is the *characteristic random variable* of \mathbf{X} .

Note that $\Phi_{\mathbf{X}}(\xi)$ is a complex random variable, with

$$|\Phi_{\mathbf{X}}(\xi)| \leq 1 \quad \text{a.s.} \tag{6.5}$$

$\Phi_{\mathbf{X}}(\xi)$ depends on the choices of $(\mathbf{e}_j)_j$ and $(\xi_j)_j$, but these choices are regarded as fixed. Moreover, the following theorem says that $\Phi_{\mathbf{X}}(\xi)$ has the same fundamental property as the usual characteristic function: it depends on \mathbf{X} only through its distribution, and conversely, it characterizes the distribution.

Theorem 6.1. *Let \mathcal{H} be a separable Hilbert space, and let \mathbf{X} and \mathbf{Y} be \mathcal{H} -valued random variables. Fix as above an ON-basis $(\mathbf{e}_i)_{i=1}^{\dim \mathcal{H}}$ in \mathcal{H} , and a random vector $\xi := (\xi_i)_1$ of i.i.d. standard normal random variables $\xi_i, i = 1, 2, \dots$, and assume further that these are independent of \mathbf{X} and \mathbf{Y} . Then*

$$\mathbf{X} \stackrel{d}{=} \mathbf{Y} \iff \Phi_{\mathbf{X}}(\xi) = \Phi_{\mathbf{Y}}(\xi) \quad \text{a.s.} \tag{6.6}$$

We prove first a lemma that will help to reduce to the finite-dimensional case.

Lemma 6.2. *Let \mathbf{X} be an \mathcal{H} -valued random variable and let $\xi = (\xi_i)_i$ be as above, and in particular independent of \mathbf{X} . Then, for any $\varepsilon > 0$, the event $\{\mathbb{E}(1 \wedge |\xi \cdot \mathbf{X}| \mid \xi) < \varepsilon\}$ has positive probability.*

More generally, for any finite set of random variables $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}$ in \mathcal{H} , all independent of ξ , the events $\{\mathbb{E}(1 \wedge |\xi \cdot \mathbf{X}^{(j)}| \mid \xi) < \varepsilon\}$ hold simultaneously with positive probability.

Proof: For finite $N \leq \dim \mathcal{H}$, let Π_N be the orthogonal projection of \mathcal{H} onto the subspace \mathcal{H}_N spanned by $\mathbf{e}_1, \dots, \mathbf{e}_N$. Let $X_{\leq N} := \Pi_N \mathbf{X}$ and $\mathbf{X}_{>N} := \mathbf{X} - X_{\leq N}$, and define $\xi_N := (\xi_1, \dots, \xi_N)$ and $\xi_{>N} := (\xi_{N+1}, \xi_{N+2}, \dots)$. Then we can write, interpreting the dot products in the obvious way in analogy with (6.1),

$$\xi \cdot \mathbf{X} = \xi_N \cdot X_{\leq N} + \xi_{>N} \cdot \mathbf{X}_{>N}. \tag{6.7}$$

Assume in the remainder of the proof that $\dim \mathcal{H} = \infty$; the case $\dim \mathcal{H} < \infty$ is similar but simpler, taking $N := \dim \mathcal{H}$ below so $\mathbf{X}_{>N} = 0$.

Since the sum in (6.1) converges a.s., and $\xi_{>N} \cdot \mathbf{X}_{>N}$ is the tail of this sum, it follows that $\xi_{>N} \cdot \mathbf{X}_{>N} \xrightarrow{\text{a.s.}} 0$ as $N \rightarrow \infty$. Consequently, by dominated convergence,

$$\mathbb{E}(1 \wedge |\xi_{>N} \cdot \mathbf{X}_{>N}|) \rightarrow 0 \quad \text{as } N \rightarrow \infty. \tag{6.8}$$

Let

$$W_N := \mathbb{E}(1 \wedge |\xi_{>N} \cdot \mathbf{X}_{>N}| \mid \xi). \tag{6.9}$$

Then (6.8) shows $\mathbb{E} W_N \rightarrow 0$; hence we may choose $N < \infty$ such that $\mathbb{E} W_N < \varepsilon/4$. Then Markov's inequality yields

$$\mathbb{P}(W_N < \varepsilon/2) \geq 1 - \frac{\mathbb{E} W_N}{\varepsilon/2} > \frac{1}{2}. \tag{6.10}$$

Moreover, for each $i \leq N$, again by dominated convergence,

$$\mathbb{E}(1 \wedge |s \langle \mathbf{X}, \mathbf{e}_i \rangle|) \rightarrow 0 \quad \text{as } s \rightarrow 0, \tag{6.11}$$

and thus there exists $\delta_i > 0$ such that if $|s| < \delta_i$, then

$$\mathbb{E}(1 \wedge |s \langle \mathbf{X}, \mathbf{e}_i \rangle|) < \frac{\varepsilon}{2N}. \tag{6.12}$$

Recalling (6.7) and (6.1), we see that

$$|\boldsymbol{\xi} \cdot \mathbf{X}| \leq \sum_{i=1}^N |\xi_i \langle \mathbf{X}, \mathbf{e}_i \rangle| + |\boldsymbol{\xi}_{>N} \cdot \mathbf{X}_{>N}| \tag{6.13}$$

and thus

$$1 \wedge |\boldsymbol{\xi} \cdot \mathbf{X}| \leq \sum_{i=1}^N (1 \wedge |\xi_i \langle \mathbf{X}, \mathbf{e}_i \rangle|) + (1 \wedge |\boldsymbol{\xi}_{>N} \cdot \mathbf{X}_{>N}|). \tag{6.14}$$

Hence, recalling (6.9),

$$\mathbb{E}(1 \wedge |\boldsymbol{\xi} \cdot \mathbf{X}| \mid \boldsymbol{\xi}) \leq \sum_{i=1}^N \mathbb{E}(1 \wedge |\xi_i \langle \mathbf{X}, \mathbf{e}_i \rangle| \mid \xi_i) + W_N. \tag{6.15}$$

Consequently, if $\boldsymbol{\xi}$ is such that $W_N < \varepsilon/2$ and $|\xi_i| < \delta_i$ for $i = 1, \dots, N$, then (6.12) implies

$$\mathbb{E}(1 \wedge |\boldsymbol{\xi} \cdot \mathbf{X}| \mid \boldsymbol{\xi}) < \sum_{i=1}^N \frac{\varepsilon}{2N} + \frac{\varepsilon}{2} = \varepsilon. \tag{6.16}$$

Since the events $\{W_N < \varepsilon/2\}$ and $\{|\xi_i| < \delta_i\}$ are independent and each has positive probability, they occur together with positive probability, and thus (6.16) holds with positive probability.

This proves the first part of the lemma. The second is proved in the same way, choosing N so large that (6.10) holds with W_N replaced by $\sum_{j=1}^m W_N^{(j)}$, where $W_N^{(j)}$ is defined by (6.9) but using $\mathbf{X}^{(j)}$ instead of \mathbf{X} , and then choosing δ_i so small that (6.12) holds for each $\mathbf{X}^{(j)}$ \square

Proof of Theorem 6.1: \implies : If $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$, then $(\mathbf{X}, \boldsymbol{\xi}) \stackrel{d}{=} (\mathbf{Y}, \boldsymbol{\xi})$ and (6.1) implies $(\boldsymbol{\xi} \cdot \mathbf{X}, \boldsymbol{\xi}) \stackrel{d}{=} (\boldsymbol{\xi} \cdot \mathbf{Y}, \boldsymbol{\xi})$ which by (6.3) implies $\Phi_{\mathbf{X}}(\boldsymbol{\xi}) = \Phi_{\mathbf{Y}}(\boldsymbol{\xi})$ a.s.

\impliedby : We let $N \leq \dim \mathcal{H}$ be finite and use the notation in the proof of Lemma 6.2. Then (6.7) holds, and thus

$$|e^{i\boldsymbol{\xi} \cdot \mathbf{X}} - e^{i\boldsymbol{\xi}_N \cdot \mathbf{X}_{\leq N}}| = |e^{i\boldsymbol{\xi}_{>N} \cdot \mathbf{X}_{>N}} - 1| \leq 2 \wedge |\boldsymbol{\xi}_{>N} \cdot \mathbf{X}_{>N}|. \tag{6.17}$$

Hence,

$$|\mathbb{E}(e^{i\boldsymbol{\xi} \cdot \mathbf{X}} \mid \boldsymbol{\xi}) - \mathbb{E}(e^{i\boldsymbol{\xi}_N \cdot \mathbf{X}_{\leq N}} \mid \boldsymbol{\xi})| \leq \mathbb{E}(2 \wedge |\boldsymbol{\xi}_{>N} \cdot \mathbf{X}_{>N}| \mid \boldsymbol{\xi}) \quad \text{a.s.} \tag{6.18}$$

Using (6.3), (6.18) can be written, since $\boldsymbol{\xi}_N$ and $\boldsymbol{\xi}_{>N}$ are independent,

$$|\Phi_{\mathbf{X}}(\boldsymbol{\xi}) - \Phi_{X_{\leq N}}(\boldsymbol{\xi}_N)| \leq \mathbb{E}(2 \wedge |\boldsymbol{\xi}_{>N} \cdot \mathbf{X}_{>N}| \mid \boldsymbol{\xi}_{>N}) \quad \text{a.s.} \tag{6.19}$$

Similarly, with analogous notation,

$$|\Phi_{\mathbf{Y}}(\boldsymbol{\xi}) - \Phi_{Y_{\leq N}}(\boldsymbol{\xi}_N)| \leq \mathbb{E}(2 \wedge |\boldsymbol{\xi}_{>N} \cdot \mathbf{Y}_{>N}| \mid \boldsymbol{\xi}_{>N}) \quad \text{a.s.} \tag{6.20}$$

The assumption $\Phi_{\mathbf{X}}(\boldsymbol{\xi}) = \Phi_{\mathbf{Y}}(\boldsymbol{\xi})$ a.s. thus implies

$$\begin{aligned} &|\Phi_{X_{\leq N}}(\boldsymbol{\xi}_N) - \Phi_{Y_{\leq N}}(\boldsymbol{\xi}_N)| \\ &\leq \mathbb{E}(2 \wedge |\boldsymbol{\xi}_{>N} \cdot \mathbf{X}_{>N}| \mid \boldsymbol{\xi}_{>N}) + \mathbb{E}(2 \wedge |\boldsymbol{\xi}_{>N} \cdot \mathbf{Y}_{>N}| \mid \boldsymbol{\xi}_{>N}) \quad \text{a.s.} \end{aligned} \tag{6.21}$$

Lemma 6.2 (applied to $\mathbf{X}_{>N}$ and $\mathbf{Y}_{>N}$) implies that for any $\varepsilon > 0$, the right-hand side of (6.21) is less than 4ε with positive probability. Furthermore, the left-hand side of (6.21) is a function of $\boldsymbol{\xi}_N$, and the right-hand side is a function of $\boldsymbol{\xi}_{>N}$; thus the two sides are independent. Consequently, (6.21) implies

$$|\Phi_{X_{\leq N}}(\boldsymbol{\xi}_N) - \Phi_{Y_{\leq N}}(\boldsymbol{\xi}_N)| < 4\varepsilon \quad \text{a.s.} \tag{6.22}$$

Since ε is arbitrary, this shows

$$\Phi_{X_{\leq N}}(\boldsymbol{\xi}_N) = \Phi_{Y_{\leq N}}(\boldsymbol{\xi}_N) \quad \text{a.s.} \tag{6.23}$$

Since $X_{\leq N}$ and $Y_{\leq N}$ live in the finite-dimensional space \mathcal{H}_N , (6.4) applies and shows

$$\varphi_{X_{\leq N}}(\boldsymbol{\xi}_N) = \Phi_{X_{\leq N}}(\boldsymbol{\xi}_N) = \Phi_{Y_{\leq N}}(\boldsymbol{\xi}_N) = \varphi_{Y_{\leq N}}(\boldsymbol{\xi}_N) \quad \text{a.s.,} \tag{6.24}$$

where $\varphi_{X_{\leq N}}(\mathbf{t})$ and $\varphi_{Y_{\leq N}}(\mathbf{t})$ are the ordinary characteristic functions in \mathbb{R}^N (identified with \mathcal{H}_N). Hence,

$$\varphi_{X_{\leq N}}(\mathbf{t}) = \varphi_{Y_{\leq N}}(\mathbf{t}) \tag{6.25}$$

for a.e. $\mathbf{t} \in \mathbb{R}^N$, and since characteristic functions are continuous, (6.25) holds for all $\mathbf{t} \in \mathbb{R}^N$, and thus

$$X_{\leq N} \stackrel{d}{=} Y_{\leq N}. \tag{6.26}$$

If $\dim \mathcal{H} < \infty$, we may choose $N = \dim \mathcal{H}$ and the result $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$ follows. (Much of the argument above is not needed in this case.)

If $\dim \mathcal{H} = \infty$, then (6.26) holds for every finite N . Furthermore, as $N \rightarrow \infty$, we have $X_{\leq N} \xrightarrow{\text{a.s.}} \mathbf{X}$ and thus $X_{\leq N} \xrightarrow{d} \mathbf{X}$ and similarly $Y_{\leq N} \xrightarrow{d} \mathbf{Y}$. Consequently, $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$, which completes the proof. \square

Remark 6.3. The mapping $\mathbf{x} \mapsto \boldsymbol{\xi} \cdot \mathbf{x}$ is an isometry of \mathcal{H} onto the Gaussian Hilbert space spanned by the random variables ξ_i , and it can be regarded as an abstract stochastic integral, cf. Janson (1997, Chapter VII.2). It replaces the Itô integrals used in Dehling et al. (2020).

Remark 6.4. The arguments above are related to the proof of Lyons (2013, Theorem 3.16). We sketch the connection: That proof uses an embedding ϕ of the Hilbert space into $L^2(\mathbb{R}^\infty \times \mathbb{R})$; if we compose ϕ with the Fourier transform $f \mapsto \int e^{2\pi i t x} f(x) dx$ acting on the last variable (which is an isometry), we obtain an equivalent embedding $\hat{\phi}$, which in our notation equals

$$\hat{\phi} : \mathbf{x} \rightarrow \frac{i}{2\pi t} (e^{i c' t \boldsymbol{\xi} \cdot \mathbf{x}} - 1) \in L^2(\mathbb{P} \times dt) \tag{6.27}$$

for a constant $c' > 0$. Hence, if $\mu = \mathcal{L}(\mathbf{X})$, the distribution of \mathbf{X} , then, combining the notation of Lyons (2013) and ours,

$$\beta_{\hat{\phi}}(\mu) := \mathbb{E}(\phi'(X) \mid \boldsymbol{\xi}) = \frac{i}{2\pi t} (\Phi_{c' t \mathbf{X}}(\boldsymbol{\xi}) - 1). \tag{6.28}$$

Hence, the result in Lyons (2013, Theorem 3.16) that $\beta_{\phi}(\mu)$ characterises μ is closely related to, and follows from, Theorem 6.1. Furthermore, the two proofs are similar; both are based on approximating with the finite-dimensional case which is easy.

6.2. *Independence and characteristic random variables.* Now consider a pair of random variables (\mathbf{X}, \mathbf{Y}) taking values in two, possibly different, separable Hilbert spaces \mathcal{H} and \mathcal{H}' . Fix, as above, an ON-basis $(\mathbf{e}_i)_1^{\dim \mathcal{H}}$ in \mathcal{H} , and i.i.d. $N(0, 1)$ random variables $\xi_i, i = 1, 2, \dots$. Similarly, fix an ON-basis $(\mathbf{e}'_j)_1^{\dim \mathcal{H}'}$ in \mathcal{H}' , and i.i.d. $N(0, 1)$ random variables $\eta_j, j = 1, 2, \dots$. Assume that all ξ_i and η_j are independent of each other and of (\mathbf{X}, \mathbf{Y}) .

Then (\mathbf{X}, \mathbf{Y}) is a random variable in the Hilbert space $\mathcal{H} \oplus \mathcal{H}' = \mathcal{H} \times \mathcal{H}'$, and $\mathbf{e}_1, \mathbf{e}'_1, \mathbf{e}_2, \mathbf{e}'_2, \dots$ is an ON-basis in this space. Let $\boldsymbol{\xi} = (\xi_i)_1^{\dim \mathcal{H}}, \boldsymbol{\eta} := (\eta_j)_1^{\dim \mathcal{H}'}$, and $\boldsymbol{\zeta} := (\xi_1, \eta_1, \xi_2, \eta_2, \dots)$.

Theorem 6.5. *Let (\mathbf{X}, \mathbf{Y}) be a pair of random variables taking values in separable Hilbert spaces \mathcal{H} and \mathcal{H}' . Then, with notation as above, \mathbf{X} and \mathbf{Y} are independent if and only if*

$$\mathbb{E}(e^{i\boldsymbol{\xi} \cdot \mathbf{X} + i\boldsymbol{\eta} \cdot \mathbf{Y}} \mid \boldsymbol{\xi}, \boldsymbol{\eta}) = \mathbb{E}(e^{i\boldsymbol{\xi} \cdot \mathbf{X}} \mid \boldsymbol{\xi}) \mathbb{E}(e^{i\boldsymbol{\eta} \cdot \mathbf{Y}} \mid \boldsymbol{\eta}) \quad \text{a.s.} \tag{6.29}$$

Proof: Let \mathbf{Y}' be a copy of \mathbf{Y} , independent of $\mathbf{X}, \boldsymbol{\xi}, \boldsymbol{\eta}$. Then, \mathbf{X} and \mathbf{Y} are independent if and only if $(\mathbf{X}, \mathbf{Y}) \stackrel{d}{=} (\mathbf{X}, \mathbf{Y}')$, and the result follows from Theorem 6.1, applied to the Hilbert space $\mathcal{H} \times \mathcal{H}'$, noting that with the bases and Gaussian variables above, $\boldsymbol{\zeta} \cdot (\mathbf{X}, \mathbf{Y}) = \boldsymbol{\xi} \cdot \mathbf{X} + \boldsymbol{\eta} \cdot \mathbf{Y}$ a.s., and thus

$$\Phi_{(\mathbf{X}, \mathbf{Y})}(\boldsymbol{\zeta}) = \mathbb{E}(e^{i\boldsymbol{\zeta} \cdot (\mathbf{X}, \mathbf{Y})} \mid \boldsymbol{\xi}, \boldsymbol{\eta}) = \mathbb{E}(e^{i\boldsymbol{\xi} \cdot \mathbf{X} + i\boldsymbol{\eta} \cdot \mathbf{Y}} \mid \boldsymbol{\xi}, \boldsymbol{\eta}), \tag{6.30}$$

while, by independence and $\mathbf{Y} \stackrel{d}{=} \mathbf{Y}'$,

$$\begin{aligned} \Phi_{(\mathbf{X}, \mathbf{Y}') }(\boldsymbol{\zeta}) &= \mathbb{E}(e^{i\boldsymbol{\xi} \cdot \mathbf{X} + i\boldsymbol{\eta} \cdot \mathbf{Y}'} \mid \boldsymbol{\xi}, \boldsymbol{\eta}) = \mathbb{E}(e^{i\boldsymbol{\xi} \cdot \mathbf{X}} \mid \boldsymbol{\xi}) \mathbb{E}(e^{i\boldsymbol{\eta} \cdot \mathbf{Y}'} \mid \boldsymbol{\eta}) \\ &= \mathbb{E}(e^{i\boldsymbol{\xi} \cdot \mathbf{X}} \mid \boldsymbol{\xi}) \mathbb{E}(e^{i\boldsymbol{\eta} \cdot \mathbf{Y}} \mid \boldsymbol{\eta}) \quad \text{a.s.} \end{aligned} \tag{6.31}$$

□

Note that, by (3.1), (6.29) may be written

$$\text{Cov}(e^{i\boldsymbol{\xi} \cdot \mathbf{X}}, e^{i\boldsymbol{\eta} \cdot \mathbf{Y}} \mid \boldsymbol{\xi}, \boldsymbol{\eta}) = 0 \quad \text{a.s.} \tag{6.32}$$

7. Covariance distance in Hilbert space

We give a new definition of covariance distance for Hilbert spaces; it can be seen as a version of Definition 2.4 for Euclidean spaces, where we replace the characteristic functions there by the characteristic random variables defined in Section 6, which makes the extension to infinite-dimensional Hilbert spaces possible. (The definition is inspired by Dehling et al., 2020, Lemma 4.1; see Remark 6.3.)

Define, for $0 < \alpha < 2$,

$$c_\alpha := \frac{2^{1+\alpha/2}}{-\Gamma(-\alpha/2)} = \frac{\alpha 2^{\alpha/2}}{\Gamma(1 - \alpha/2)}. \tag{7.1}$$

Definition 7.1. Let (\mathbf{X}, \mathbf{Y}) be a pair of random vectors in separable Hilbert spaces, and let $0 < \alpha < 2$. Then, with notation as in Section 6,

$$\begin{aligned} \text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) &= \text{dcov}_\alpha^{\text{H}}(\mathbf{X}, \mathbf{Y}) \\ &:= c_\alpha^2 \int_0^\infty \int_0^\infty \mathbb{E} \left| \Phi_{(r\mathbf{X}, s\mathbf{Y})}(\boldsymbol{\xi}, \boldsymbol{\eta}) - \Phi_{r\mathbf{X}}(\boldsymbol{\xi}) \Phi_{s\mathbf{Y}}(\boldsymbol{\eta}) \right|^2 \frac{dr ds}{r^{\alpha+1} s^{\alpha+1}} \end{aligned} \tag{7.2}$$

$$= c_\alpha^2 \int_0^\infty \int_0^\infty \mathbb{E} \left| \mathbb{E}(e^{ir\boldsymbol{\xi} \cdot \mathbf{X} + is\boldsymbol{\eta} \cdot \mathbf{Y}} \mid \boldsymbol{\xi}, \boldsymbol{\eta}) - \mathbb{E}(e^{ir\boldsymbol{\xi} \cdot \mathbf{X}} \mid \boldsymbol{\xi}) \mathbb{E}(e^{is\boldsymbol{\eta} \cdot \mathbf{Y}} \mid \boldsymbol{\eta}) \right|^2 \frac{dr ds}{r^{\alpha+1} s^{\alpha+1}} \tag{7.3}$$

$$= c_\alpha^2 \int_0^\infty \int_0^\infty \mathbb{E} \left| \text{Cov}(e^{ir\boldsymbol{\xi} \cdot \mathbf{X}}, e^{is\boldsymbol{\eta} \cdot \mathbf{Y}} \mid \boldsymbol{\xi}, \boldsymbol{\eta}) \right|^2 \frac{dr ds}{r^{\alpha+1} s^{\alpha+1}}. \tag{7.4}$$

The expressions (7.2)–(7.4) are equal by the definitions (6.3) and (3.1) above, cf. (6.30) and (6.32). Note that no moment assumptions are made; as for Definition 2.4, the definition works for any (\mathbf{X}, \mathbf{Y}) in these spaces, but $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ may be infinite. Furthermore, as shown in the next theorem, for the special case of Euclidean spaces, Definition 7.1 agrees with Definition 2.4, again without moment conditions.

Theorem 7.2. *Let $0 < \alpha < 2$. If (\mathbf{X}, \mathbf{Y}) is a pair of random vectors in Euclidean spaces \mathbb{R}^p and \mathbb{R}^q , then Definitions 2.4 and 7.1 agree, i.e., $\text{dcov}_\alpha^E(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$.*

Proof: Assume that $\mathcal{H} = \mathbb{R}^p$ and $\mathcal{H}' = \mathbb{R}^q$. Then (6.4) implies

$$\Phi_{(r\mathbf{X}, s\mathbf{Y})}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \varphi_{(r\mathbf{X}, s\mathbf{Y})}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \varphi_{(\mathbf{X}, \mathbf{Y})}(r\boldsymbol{\xi}, s\boldsymbol{\eta}) \tag{7.5}$$

and thus, since $r\boldsymbol{\xi} \sim N(0, r^2 I_p)$ and $s\boldsymbol{\eta} \sim N(0, s^2 I_q)$, where I_k is the identity matrix in \mathbb{R}^k ,

$$\begin{aligned} \mathbb{E}|\Phi_{(r\mathbf{X}, s\mathbf{Y})}(\boldsymbol{\xi}, \boldsymbol{\eta}) - \Phi_{r\mathbf{X}}(\boldsymbol{\xi})\Phi_{s\mathbf{Y}}(\boldsymbol{\eta})|^2 &= \mathbb{E}|\varphi_{(\mathbf{X}, \mathbf{Y})}(r\boldsymbol{\xi}, s\boldsymbol{\eta}) - \varphi_{\mathbf{X}}(r\boldsymbol{\xi})\varphi_{\mathbf{Y}}(s\boldsymbol{\eta})|^2 \\ &= \int_{\mathbf{t} \in \mathbb{R}^p} \int_{\mathbf{u} \in \mathbb{R}^q} |\varphi_{(\mathbf{X}, \mathbf{Y})}(\mathbf{t}, \mathbf{u}) - \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{Y}}(\mathbf{u})|^2 \frac{e^{-|\mathbf{t}|^2/2r^2}}{(2\pi r^2)^{p/2}} \frac{e^{-|\mathbf{u}|^2/2s^2}}{(2\pi s^2)^{q/2}} d\mathbf{t} d\mathbf{u}. \end{aligned} \tag{7.6}$$

Substituting this in (7.2), we obtain (2.9) by interchanging the order of integration, because, by elementary calculations,

$$\int_0^\infty \frac{e^{-|\mathbf{t}|^2/2r^2}}{(2\pi r^2)^{p/2}} \frac{dr}{r^{\alpha+1}} = \frac{2^{\alpha/2-1}\Gamma((p+\alpha)/2)}{\pi^{p/2}} |\mathbf{t}|^{-p-\alpha} = \frac{c_{\alpha,p}}{c_\alpha} |\mathbf{t}|^{-p-\alpha}, \tag{7.7}$$

see (2.8) and (7.1), and similarly for the integral over s . □

Remark 7.3. The proof of Theorem 7.2 together with Remark 2.6 shows that the restriction $\alpha < 2$ in Definition 7.1 is necessary; for $\alpha \geq 2$, the integrals diverge typically, for example for $\mathcal{H} = \mathcal{H}' = \mathbb{R}$ and $\mathbf{X} = \mathbf{Y} \sim N(0, 1)$. (We conjecture that for $\alpha \geq 2$, the integrals always diverge except when \mathbf{X} and \mathbf{Y} are independent, but we have not verified that.)

We return to the general Hilbert space case, and show that Definition 7.1 agrees with the earlier ones; this is an abstract version of Dehling et al. (2020, Lemma 4.1), where the Hilbert spaces are $L^2[0, 1]$, see Remark 6.3.

Theorem 7.4. *Let $0 < \alpha < 2$. If (\mathbf{X}, \mathbf{Y}) is a pair of random vectors in Hilbert spaces \mathcal{H} and \mathcal{H}' , and $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^\alpha < \infty$, then Definitions 2.2, 2.3 and 7.1 agree, i.e., $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$, and this value is finite.*

Proof: Let again $(\mathbf{X}_1, \mathbf{Y}_1), \dots$ be i.i.d. copies of (\mathbf{X}, \mathbf{Y}) , and assume that $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are independent of all of them. Then, using (6.30)–(6.31) and (6.2),

$$\begin{aligned} &\mathbb{E}|\Phi_{(r\mathbf{X}, s\mathbf{Y})}(\boldsymbol{\xi}, \boldsymbol{\eta}) - \Phi_{r\mathbf{X}}(\boldsymbol{\xi})\Phi_{s\mathbf{Y}}(\boldsymbol{\eta})|^2 \\ &= \mathbb{E} \mathbb{E} \left[\left(e^{ir\boldsymbol{\xi} \cdot \mathbf{X}_1 + is\boldsymbol{\eta} \cdot \mathbf{Y}_1} - e^{ir\boldsymbol{\xi} \cdot \mathbf{X}_1 + is\boldsymbol{\eta} \cdot \mathbf{Y}_2} \right) \left(e^{-ir\boldsymbol{\xi} \cdot \mathbf{X}_3 - is\boldsymbol{\eta} \cdot \mathbf{Y}_3} - e^{-ir\boldsymbol{\xi} \cdot \mathbf{X}_3 - is\boldsymbol{\eta} \cdot \mathbf{Y}_4} \right) \mid \boldsymbol{\xi}, \boldsymbol{\eta} \right] \\ &= \mathbb{E} \left[\left(e^{ir\boldsymbol{\xi} \cdot \mathbf{X}_1 + is\boldsymbol{\eta} \cdot \mathbf{Y}_1} - e^{ir\boldsymbol{\xi} \cdot \mathbf{X}_1 + is\boldsymbol{\eta} \cdot \mathbf{Y}_2} \right) \left(e^{-ir\boldsymbol{\xi} \cdot \mathbf{X}_3 - is\boldsymbol{\eta} \cdot \mathbf{Y}_3} - e^{-ir\boldsymbol{\xi} \cdot \mathbf{X}_3 - is\boldsymbol{\eta} \cdot \mathbf{Y}_4} \right) \right] \\ &= \mathbb{E} e^{ir\boldsymbol{\xi} \cdot (\mathbf{X}_1 - \mathbf{X}_3) + is\boldsymbol{\eta} \cdot (\mathbf{Y}_1 - \mathbf{Y}_3)} - \mathbb{E} e^{ir\boldsymbol{\xi} \cdot (\mathbf{X}_1 - \mathbf{X}_3) + is\boldsymbol{\eta} \cdot (\mathbf{Y}_1 - \mathbf{Y}_4)} \\ &\quad - \mathbb{E} e^{ir\boldsymbol{\xi} \cdot (\mathbf{X}_1 - \mathbf{X}_3) + is\boldsymbol{\eta} \cdot (\mathbf{Y}_2 - \mathbf{Y}_3)} + \mathbb{E} e^{ir\boldsymbol{\xi} \cdot (\mathbf{X}_1 - \mathbf{X}_3) + is\boldsymbol{\eta} \cdot (\mathbf{Y}_2 - \mathbf{Y}_4)} \\ &= \mathbb{E} e^{-\frac{r^2}{2}\|\mathbf{X}_1 - \mathbf{X}_3\|^2 - \frac{s^2}{2}\|\mathbf{Y}_1 - \mathbf{Y}_3\|^2} - \mathbb{E} e^{-\frac{r^2}{2}\|\mathbf{X}_1 - \mathbf{X}_3\|^2 - \frac{s^2}{2}\|\mathbf{Y}_1 - \mathbf{Y}_4\|^2} \\ &\quad - \mathbb{E} e^{-\frac{r^2}{2}\|\mathbf{X}_1 - \mathbf{X}_3\|^2 - \frac{s^2}{2}\|\mathbf{Y}_2 - \mathbf{Y}_3\|^2} + \mathbb{E} e^{-\frac{r^2}{2}\|\mathbf{X}_1 - \mathbf{X}_3\|^2 - \frac{s^2}{2}\|\mathbf{Y}_2 - \mathbf{Y}_4\|^2}. \end{aligned} \tag{7.8}$$

Define the real-valued random variable

$$\Lambda_X(u) := \mathbb{E} e^{-u\|\mathbf{X}_1-\mathbf{X}_2\|^2} - \mathbb{E} e^{-u\|\mathbf{X}_2-\mathbf{X}_3\|^2} + \mathbb{E} e^{-u\|\mathbf{X}_3-\mathbf{X}_4\|^2} - \mathbb{E} e^{-u\|\mathbf{X}_4-\mathbf{X}_1\|^2} \tag{7.9}$$

and define $\Lambda_Y(u)$ similarly. Then, by expanding the product and using symmetry,

$$\begin{aligned} &\mathbb{E}[\Lambda_X(u)\Lambda_Y(v)] \\ &= 4\left(\mathbb{E} e^{-u\|\mathbf{X}_1-\mathbf{X}_3\|^2-v\|\mathbf{Y}_1-\mathbf{Y}_3\|^2} - \mathbb{E} e^{-u\|\mathbf{X}_1-\mathbf{X}_3\|^2-v\|\mathbf{Y}_1-\mathbf{Y}_4\|^2} \right. \\ &\quad \left. - \mathbb{E} e^{-u\|\mathbf{X}_1-\mathbf{X}_3\|^2-v\|\mathbf{Y}_2-\mathbf{Y}_3\|^2} + \mathbb{E} e^{-u\|\mathbf{X}_1-\mathbf{X}_3\|^2-v\|\mathbf{Y}_2-\mathbf{Y}_4\|^2}\right). \end{aligned} \tag{7.10}$$

Consequently, (7.8) yields

$$\mathbb{E}|\Phi_{(r\mathbf{X},s\mathbf{Y})}(\boldsymbol{\xi}, \boldsymbol{\eta}) - \Phi_{r\mathbf{X}}(\boldsymbol{\xi})\Phi_{s\mathbf{Y}}(\boldsymbol{\eta})|^2 = \frac{1}{4} \mathbb{E}\left[\Lambda_X\left(\frac{r^2}{2}\right)\Lambda_Y\left(\frac{s^2}{2}\right)\right] \tag{7.11}$$

and the definition (7.3) yields, with a change of variables,

$$\begin{aligned} \text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y}) &= \frac{c_\alpha^2}{4} \int_0^\infty \int_0^\infty \mathbb{E}\left[\Lambda_X\left(\frac{r^2}{2}\right)\Lambda_Y\left(\frac{s^2}{2}\right)\right] \frac{dr ds}{r^{\alpha+1}s^{\alpha+1}} \\ &= \frac{c_\alpha^2}{2^{4+\alpha}} \int_0^\infty \int_0^\infty \mathbb{E}[\Lambda_X(u)\Lambda_Y(v)] \frac{du dv}{u^{\alpha/2+1}v^{\alpha/2+1}}. \end{aligned} \tag{7.12}$$

We rewrite (7.9) as, with indices interpreted modulo 4,

$$\Lambda_X(u) = \sum_{i=1}^4 (-1)^{i-1} e^{-u\|\mathbf{X}_i-\mathbf{X}_{i+1}\|^2} = \sum_{i=1}^4 (-1)^i (1 - e^{-u\|\mathbf{X}_i-\mathbf{X}_{i+1}\|^2}). \tag{7.13}$$

Recall that for $0 < \gamma < 1$, see [Olver et al. \(2010, \(5.9.5\)\)](#),

$$\int_0^\infty (1 - e^{-x})x^{-\gamma-1} dx = -\Gamma(-\gamma). \tag{7.14}$$

Hence, (7.13) and a change of variables yield

$$\begin{aligned} \int_0^\infty \Lambda_X(u) \frac{du}{u^{\alpha/2+1}} &= \sum_{i=1}^4 (-1)^i \int_0^\infty (1 - e^{-u\|\mathbf{X}_i-\mathbf{X}_{i+1}\|^2}) \frac{du}{u^{\alpha/2+1}} \\ &= -\Gamma(-\alpha/2) \sum_{i=1}^4 (-1)^i \|\mathbf{X}_i - \mathbf{X}_{i+1}\|^\alpha \\ &= \Gamma(-\alpha/2) \widehat{X}_\alpha. \end{aligned} \tag{7.15}$$

If we naively interchange order of integrations and expectation in (7.12), and use (7.15), we obtain (2.3) and thus $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^{\widehat{}}(\mathbf{X}, \mathbf{Y})$, since c_α is defined in (7.1) so that constant factors cancel. However, this interchange requires justification; indeed it is not always allowed, since the expectation in (2.3) does not always exist, not even as an extended real number, see [Example 9.7](#), while (7.2)–(7.4) always exist in $[0, \infty]$.

Hence, we introduce an integrating factor. Let $M > 0$; we will later let $M \rightarrow \infty$. Similarly to (7.15), we have

$$\begin{aligned} \int_0^\infty e^{-Mu} \Lambda_X(u) \frac{du}{u^{\alpha/2+1}} &= \sum_{i=1}^4 (-1)^i \int_0^\infty (e^{-Mu} - e^{-u(\|\mathbf{X}_i-\mathbf{X}_{i+1}\|^2+M)}) \frac{du}{u^{\alpha/2+1}} \\ &= -\Gamma(-\alpha/2) \sum_{i=1}^4 (-1)^i ((\|\mathbf{X}_i - \mathbf{X}_{i+1}\|^2 + M)^{\alpha/2} - M^{\alpha/2}). \end{aligned} \tag{7.16}$$

Let $\alpha \in (0, 2)$ be given and define, for $x \geq 0$,

$$h_M(x) := x^{\alpha/2} + M^{\alpha/2} - (x + M)^{\alpha/2}. \tag{7.17}$$

Then, (7.15) and (7.16) yield

$$\begin{aligned} \int_0^\infty (1 - e^{-Mu}) \Lambda_X(u) \frac{du}{u^{\alpha/2+1}} &= \Gamma(-\alpha/2) \sum_{i=1}^4 (-1)^{i-1} h_M(\|\mathbf{X}_i - \mathbf{X}_{i+1}\|^2) \\ &=: \Gamma(-\alpha/2) \widehat{X}_{\alpha;M}, \end{aligned} \tag{7.18}$$

where thus we define

$$\widehat{X}_{\alpha;M} := \sum_{i=1}^4 (-1)^{i-1} h_M(\|\mathbf{X}_i - \mathbf{X}_{i+1}\|^2). \tag{7.19}$$

Note also that the integrand in (7.12) is non-negative by (7.11). Hence, (7.12) and monotone convergence yield

$$\begin{aligned} \text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y}) &= \lim_{M \rightarrow \infty} \frac{c_\alpha^2}{2^{4+\alpha}} \int_0^\infty \int_0^\infty (1 - e^{-Mu})(1 - e^{-Mv}) \mathbb{E}[\Lambda_X(u)\Lambda_Y(v)] \frac{du dv}{u^{\alpha/2+1}v^{\alpha/2+1}}. \end{aligned} \tag{7.20}$$

Furthermore, $|\Lambda_X(u)|$ and $|\Lambda_Y(v)|$ are bounded (by 4) by (7.13), and thus Fubini applies so we may interchange expectation and integrations in (7.20), which by (7.18) yields, recalling (7.1),

$$\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y}) = \lim_{M \rightarrow \infty} \frac{1}{4} \mathbb{E}[\widehat{X}_{\alpha;M} \widehat{Y}_{\alpha;M}]. \tag{7.21}$$

Since $\alpha/2 \in (0, 1)$, the function h_M in (7.17) is increasing, with $h_M(0) = 0$ and $h_M(x) \nearrow M^{\alpha/2}$ as $x \rightarrow \infty$. Similarly, $h_M(x) = h_x(M) \nearrow x^{\alpha/2}$ as $M \rightarrow \infty$; hence, the definitions (7.19) and (2.4) yield

$$\widehat{X}_{\alpha;M} \rightarrow \widehat{X}_\alpha \quad \text{as } M \rightarrow \infty. \tag{7.22}$$

Furthermore, if $0 \leq x \leq y$, then

$$0 \leq h_M(y) - h_M(x) \leq y^{\alpha/2} - x^{\alpha/2}, \tag{7.23}$$

and it follows that for any $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathcal{H}$,

$$|h_M(\|\mathbf{Z}_1\|^2) - h_M(\|\mathbf{Z}_2\|^2)| \leq \left| \|\mathbf{Z}_1\|^\alpha - \|\mathbf{Z}_2\|^\alpha \right|. \tag{7.24}$$

We claim that Lemma 4.2 holds for $\widehat{X}_{\alpha;M}$ too, so that, in particular,

$$|\widehat{X}_{\alpha;M}| \leq C \sum_{i=1}^4 \|\mathbf{X}_i\|^{\alpha/2} \|\mathbf{X}_{i+1}\|^{\alpha/2}, \tag{7.25}$$

where the constant C does not depend on M . This is seen by repeating the proof of Lemma 4.2, recalling the definition (7.19) of $\widehat{X}_{\alpha;M}$ and using (7.24); we omit the details.

Let \widehat{X}^* be the right-hand side of (7.25). We now use the assumption $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$, which implies that $\widehat{X}^* \in L^2$. Similarly, $|\widehat{Y}_{\alpha;M}| \leq \widehat{Y}^*$ with $\widehat{Y}^* \in L^2$. Consequently, $|\widehat{X}_{\alpha;M} \widehat{Y}_{\alpha;M}| \leq \widehat{X}^* \widehat{Y}^* \in L^1$, so dominated convergence applies to (7.21) and we obtain, by (7.22),

$$\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y}) = \frac{1}{4} \mathbb{E}[\lim_{M \rightarrow \infty} \widehat{X}_{\alpha;M} \widehat{Y}_{\alpha;M}] = \frac{1}{4} \mathbb{E}[\widehat{X}_\alpha \widehat{Y}_\alpha] = \text{dcov}_\alpha^{\widehat{}}(\mathbf{X}, \mathbf{Y}), \tag{7.26}$$

using (2.3). Hence, Definitions 7.1 and 2.2 agree (under the given moment condition). By Theorem 4.5, they agree with Definition 2.3 too; furthermore, the value is finite. \square

Remark 7.5. Note that the proof shows that (7.21) holds for any random variables in Hilbert spaces, without any moment condition. (With the result possibly $+\infty$.)

7.1. *Independence and distance covariance.* For (separable) Hilbert spaces, as said in the introduction, Lyons (2013, Theorem 3.16) showed that (1.1) holds for $\alpha = 1$, and Dehling et al. (2020, Theorem 4.2) extended this to all $\alpha \in (0, 2)$. That is, they proved (a version of) the following, which we now easily can prove using the results above.

Theorem 7.6 (Dehling et al., 2020, Theorem 4.2). *Let $\mathcal{X} = \mathcal{H}$ and $\mathcal{Y} = \mathcal{H}'$ be separable Hilbert spaces and let $\alpha \in (0, 2)$. Use Definition 2.1, 2.2, 2.3 or 7.1, and assume (for the first three) the moment condition there. Then $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) = 0$ if and only if \mathbf{X} and \mathbf{Y} are independent.*

Proof: For Definitions 2.1–2.3, the moment condition there and Theorems 4.5 and 7.2 show that $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ equals $\text{dcov}_\alpha^{\text{H}}(\mathbf{X}, \mathbf{Y})$ given by Definition 7.1. Hence, we may in all cases use $\text{dcov}_\alpha^{\text{H}}$. It follows from (7.3) that $\text{dcov}_\alpha^{\text{H}}(\mathbf{X}, \mathbf{Y}) = 0$ if and only if (6.29) holds, and the result follows by Theorem 6.5. \square

Remark 7.7. This theorem is stated in Dehling et al. (2020) for the case $\mathcal{X} = \mathcal{Y} = L^2[0, 1]$ (so \mathbf{X} and \mathbf{Y} are stochastic processes on $[0, 1]$), but since all infinite-dimensional separable Hilbert spaces are isomorphic; the result can be stated as above. (Only stochastic processes \mathbf{X}, \mathbf{Y} that satisfy some smoothness conditions are considered in Dehling et al. (2020), but this is for other reasons and is not needed for Theorem 7.6.)

The theorem in Dehling et al. (2020) is stated assuming only finite α moments, as we do above for Definitions 2.2 and 2.3; however, Dehling et al. (2020) uses Definition 2.1 which in general requires somewhat more for existence, see Theorem 9.1 below.

Remark 7.8. Theorem 7.6 includes the case when \mathcal{X} or \mathcal{Y} has finite dimension, i.e., is a Euclidean space.

Furthermore, although the theorem is stated for separable Hilbert spaces, it extends also to non-separable spaces, provided we assume that \mathbf{X} and \mathbf{Y} are Bochner measurable, for the trivial reason that this implies that \mathbf{X} and \mathbf{Y} a.s. take values in some separable subspaces \mathcal{H}_1 and \mathcal{H}'_1 .

Remark 7.9. The proof of Theorem 7.6 would be much simpler if distance covariance was monotone under orthogonal projections, so that we would have $\text{dcov}_\alpha(\Pi_N \mathbf{X}, \Pi_N \mathbf{Y}) \leq \text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$. However, this is not always the case, even in finite dimension, as is seen by the following example.

Example 7.10. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^2$ and let $\mathbf{X} = (X', X'')$ and $\mathbf{Y} = (Y', Y'')$, where $X' = Y'$, but X', X'', Y'' are independent and non-degenerate. (For definiteness, we may take $X', X'', Y'' \sim \text{Be}(1/2)$, or $N(0, 1)$.) Let $\Pi : \mathbb{R}^2 \rightarrow \mathbb{R}$ be the standard projection onto the first coordinate, so $(\Pi \mathbf{X}, \Pi \mathbf{Y}) = (X', Y')$.

For $a \in \mathbb{R}$, let $\mathbf{X}(a) := (X', aX'')$ and $\mathbf{Y}(a) := (Y', aY'')$; thus $(\mathbf{X}(1), \mathbf{Y}(1)) = (\mathbf{X}, \mathbf{Y})$ and $(\mathbf{X}(0), \mathbf{Y}(0)) = (X', Y')$ (regarding \mathbb{R} as a subspace of \mathbb{R}^2). For $\mathbf{t} = (t', t'')$ and $\mathbf{u} = (u', u'')$, we have

$$\begin{aligned} \varphi_{\mathbf{X}(a), \mathbf{Y}(a)}(\mathbf{t}, \mathbf{u}) &= \mathbb{E} e^{i(t'X' + u'X' + t''aX'' + u''aY'')} \\ &= \varphi_{X'}(t' + u') \varphi_{X''}(at'') \varphi_{Y''}(au'') \end{aligned} \quad (7.27)$$

and similarly (or by taking $\mathbf{t} = 0$ or $\mathbf{u} = 0$ in (7.27))

$$\varphi_{\mathbf{X}(a)}(\mathbf{t}) = \varphi_{X'}(t') \varphi_{X''}(at''), \quad \varphi_{\mathbf{Y}(a)}(\mathbf{u}) = \varphi_{X'}(u') \varphi_{Y''}(au''). \quad (7.28)$$

Hence, (2.9) yields

$$\begin{aligned} \text{dcov}_\alpha(\mathbf{X}(a), \mathbf{Y}(a)) \\ = c_{\alpha, 2} c_{\alpha, 2} \int_{\mathbf{t} \in \mathbb{R}^2} \int_{\mathbf{u} \in \mathbb{R}^2} |\varphi_{X'}(t' + u') - \varphi_{X'}(t') \varphi_{X'}(u')|^2 |\varphi_{X''}(at'') \varphi_{Y''}(au'')|^2 \frac{d\mathbf{t} d\mathbf{u}}{|\mathbf{t}|^{2+\alpha} |\mathbf{u}|^{2+\alpha}} \end{aligned} \quad (7.29)$$

and it is obvious that

$$\begin{aligned} \text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) &= \text{dcov}_\alpha(\mathbf{X}(1), \mathbf{Y}(1)) \\ &< \text{dcov}_\alpha(\mathbf{X}(0), \mathbf{Y}(0)) = \text{dcov}_\alpha(X', Y') = \text{dcov}_\alpha(\Pi\mathbf{X}, \Pi\mathbf{Y}). \end{aligned} \tag{7.30}$$

Thus, an orthogonal projection might increase distance covariance.

It can obviously also decrease it; for example the projection onto the second coordinate above yields (X'', Y'') with $\text{dcov}_\alpha(X'', Y'') = 0$.

8. Hilbert spaces and $\alpha = 2$

We continue to assume that \mathcal{X} and \mathcal{Y} are Hilbert spaces; we now consider the case $\alpha = 2$. Note that Definition 7.1 does not apply (it requires $\alpha < 2$, see Remark 7.3), so we return to the general Definitions 2.1–2.3.

In this case, (2.4) yields, by expanding all $\|\mathbf{X}_i - \mathbf{X}_j\|^2$,

$$\begin{aligned} \widehat{X}_2 &= -2\langle \mathbf{X}_1, \mathbf{X}_2 \rangle + 2\langle \mathbf{X}_2, \mathbf{X}_3 \rangle - 2\langle \mathbf{X}_3, \mathbf{X}_4 \rangle + 2\langle \mathbf{X}_4, \mathbf{X}_1 \rangle \\ &= 2\langle \mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_4 - \mathbf{X}_2 \rangle. \end{aligned} \tag{8.1}$$

Assume, as in Definitions 2.2 and 2.3, that $\mathbb{E}\|\mathbf{X}\|^2 < \infty$. Then $\mathbb{E}\mathbf{X}$ exists, in Bochner sense (see Appendix B), and (2.6) together with (8.1) yield

$$\widetilde{X}_2 = \mathbb{E}(\widehat{X}_2 \mid \mathbf{X}_1, \mathbf{X}_2) = -2\langle \mathbf{X}_1 - \mathbb{E}\mathbf{X}, \mathbf{X}_2 - \mathbb{E}\mathbf{X} \rangle. \tag{8.2}$$

We thus see directly that (4.2) and (4.3) hold, and thus $\widehat{X}_2, \widetilde{X}_2 \in L^2$ if $\mathbb{E}\|\mathbf{X}\|^2 < \infty$, as asserted by Lemma 4.3.

In particular, in the 1-dimensional case $\mathcal{X} = \mathbb{R}$,

$$\widehat{X}_2 = 2(\mathbf{X}_1 - \mathbf{X}_3)(\mathbf{X}_4 - \mathbf{X}_2), \quad \widetilde{X}_2 = -2(\mathbf{X}_1 - \mathbb{E}\mathbf{X})(\mathbf{X}_2 - \mathbb{E}\mathbf{X}), \tag{8.3}$$

with the latter assuming $\mathbb{E}|\mathbf{X}|^2 < \infty$. Consequently, if $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and $\mathbb{E}|\mathbf{X}|^2, \mathbb{E}|\mathbf{Y}|^2 < \infty$, then Definition 2.3 yields, using (8.3) and independence,

$$\text{dcov}_2(\mathbf{X}, \mathbf{Y}) = \mathbb{E}[\widetilde{X}_2\widetilde{Y}_2] = 4\text{Cov}(\mathbf{X}, \mathbf{Y})^2, \tag{8.4}$$

as noted by Székely et al. (2007). (Definitions 2.1–2.2 agree by Theorems 4.5.) This extends to higher dimensional Euclidean spaces and, more generally, Hilbert spaces as follows. Let $\mathcal{H} \otimes \mathcal{H}'$ denote the Hilbert space tensor product of \mathcal{H} and \mathcal{H}' , see e.g. Janson (1997, Appendix E); recall that this is a Hilbert space such that there is a bilinear map $\otimes : \mathcal{H} \times \mathcal{H}' \rightarrow \mathcal{H} \otimes \mathcal{H}'$ with

$$\langle \mathbf{x}_1 \otimes \mathbf{y}_1, \mathbf{x}_2 \otimes \mathbf{y}_2 \rangle_{\mathcal{H} \otimes \mathcal{H}'} = \langle \mathbf{x}_1, \mathbf{x}_2 \rangle_{\mathcal{H}} \langle \mathbf{y}_1, \mathbf{y}_2 \rangle_{\mathcal{H}'}; \tag{8.5}$$

furthermore, if $\{\mathbf{e}_i\}_i$ and $\{\mathbf{e}'_j\}_j$ are ON-bases in \mathcal{H} and \mathcal{H}' , then $\{\mathbf{e}_i \otimes \mathbf{e}'_j\}_{i,j}$ is an ON-basis in $\mathcal{H} \otimes \mathcal{H}'$. (Note that the mapping \otimes is neither injective nor surjective, but the set of finite linear combinations $\sum_i \mathbf{x}_i \otimes \mathbf{y}_j$ is dense in $\mathcal{H} \otimes \mathcal{H}'$.) Hence, $\mathbf{X} \otimes \mathbf{Y}$ is a random variable in $\mathcal{H} \otimes \mathcal{H}'$ with $\|\mathbf{X} \otimes \mathbf{Y}\| = \|\mathbf{X}\| \|\mathbf{Y}\|$.

Theorem 8.1. *Let $\mathcal{X} = \mathcal{H}$ and $\mathcal{Y} = \mathcal{H}'$ be separable Hilbert spaces, and assume $\mathbb{E}\|\mathbf{X}\|^2 < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^2 < \infty$. Let $(\mathbf{e}_i)_i$ and $(\mathbf{e}'_j)_j$ be ON-bases in \mathcal{H} and \mathcal{H}' . Then,*

$$\text{dcov}_2(\mathbf{X}, \mathbf{Y}) = 4 \sum_{i,j} \text{Cov}(\langle \mathbf{X}, \mathbf{e}_i \rangle, \langle \mathbf{Y}, \mathbf{e}'_j \rangle)^2 \tag{8.6}$$

$$= 4 \|\mathbb{E}(\mathbf{X} \otimes \mathbf{Y}) - \mathbb{E}\mathbf{X} \otimes \mathbb{E}\mathbf{Y}\|_{\mathcal{H} \otimes \mathcal{H}'}^2 \tag{8.7}$$

Proof: Since $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ and the expressions in (8.6)–(8.7) are invariant under (deterministic) shifts of \mathbf{X} and \mathbf{Y} , we may for convenience assume $\mathbb{E}\mathbf{X} = \mathbb{E}\mathbf{Y} = 0$. Then, by (8.2),

$$\tilde{X}_2 \tilde{Y}_2 = 4 \langle \mathbf{X}_1, \mathbf{X}_2 \rangle \langle \mathbf{Y}_1, \mathbf{Y}_2 \rangle = 4 \sum_{i,j} \langle \mathbf{X}_1, \mathbf{e}_i \rangle \langle \mathbf{X}_2, \mathbf{e}_i \rangle \langle \mathbf{Y}_1, \mathbf{e}'_j \rangle \langle \mathbf{Y}_2, \mathbf{e}'_j \rangle. \quad (8.8)$$

We have, by the Cauchy–Schwarz inequality,

$$\sum_i |\langle \mathbf{X}_1, \mathbf{e}_i \rangle \langle \mathbf{X}_2, \mathbf{e}_i \rangle| \leq \left(\sum_i \langle \mathbf{X}_1, \mathbf{e}_i \rangle^2 \right)^{1/2} \left(\sum_i \langle \mathbf{X}_2, \mathbf{e}_i \rangle^2 \right)^{1/2} = \|\mathbf{X}_1\| \|\mathbf{X}_2\| \quad (8.9)$$

and thus, by independence and the Cauchy–Schwarz inequality again,

$$\begin{aligned} \mathbb{E} \sum_{i,j} |\langle \mathbf{X}_1, \mathbf{e}_i \rangle \langle \mathbf{X}_2, \mathbf{e}_i \rangle \langle \mathbf{Y}_1, \mathbf{e}'_j \rangle \langle \mathbf{Y}_2, \mathbf{e}'_j \rangle| &\leq \mathbb{E} [\|\mathbf{X}_1\| \|\mathbf{X}_2\| \|\mathbf{Y}_1\| \|\mathbf{Y}_2\|] \\ &= (\mathbb{E} [\|\mathbf{X}_1\| \|\mathbf{Y}_1\|])^2 \leq \mathbb{E} \|\mathbf{X}_1\|^2 \mathbb{E} \|\mathbf{Y}_1\|^2 < \infty. \end{aligned} \quad (8.10)$$

Hence, (8.8) yields by Fubini’s theorem, justified by (8.10),

$$\begin{aligned} \mathbb{E} [\tilde{X}_2 \tilde{Y}_2] &= 4 \sum_{i,j} \mathbb{E} [\langle \mathbf{X}_1, \mathbf{e}_i \rangle \langle \mathbf{X}_2, \mathbf{e}_i \rangle \langle \mathbf{Y}_1, \mathbf{e}'_j \rangle \langle \mathbf{Y}_2, \mathbf{e}'_j \rangle] \\ &= 4 \sum_{i,j} (\mathbb{E} [\langle \mathbf{X}, \mathbf{e}_i \rangle \langle \mathbf{Y}, \mathbf{e}'_j \rangle])^2 \end{aligned} \quad (8.11)$$

which yields (8.6).

Moreover, $\{\mathbf{e}_i \otimes \mathbf{e}'_j\}_{i,j}$ is an ON-basis in $\mathcal{H} \otimes \mathcal{H}'$, and thus

$$\begin{aligned} \|\mathbb{E}(\mathbf{X} \otimes \mathbf{Y})\|^2 &= \sum_{i,j} \langle \mathbb{E}(\mathbf{X} \otimes \mathbf{Y}), \mathbf{e}_i \otimes \mathbf{e}'_j \rangle^2 = \sum_{i,j} (\mathbb{E} \langle \mathbf{X} \otimes \mathbf{Y}, \mathbf{e}_i \otimes \mathbf{e}'_j \rangle)^2 \\ &= \sum_{i,j} (\mathbb{E} [\langle \mathbf{X}, \mathbf{e}_i \rangle \langle \mathbf{Y}, \mathbf{e}'_j \rangle])^2 \end{aligned} \quad (8.12)$$

which together with (8.11) yields (8.7). \square

Corollary 8.2. *Let $\mathcal{X} = \mathcal{H}$ and $\mathcal{Y} = \mathcal{H}'$ be separable Hilbert spaces, and assume $\mathbb{E}\|\mathbf{X}\|^2 < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^2 < \infty$. Then, the following are equivalent:*

- (i) $\text{dcov}_2(\mathbf{X}, \mathbf{Y}) = 0$.
- (ii) $\text{Cov}(\langle \mathbf{X}, \mathbf{x} \rangle, \langle \mathbf{Y}, \mathbf{y} \rangle) = 0$ for every $\mathbf{x} \in \mathcal{H}$, $\mathbf{y} \in \mathcal{H}'$.
- (iii) $\mathbb{E}(\mathbf{X} \otimes \mathbf{Y}) - \mathbb{E}\mathbf{X} \otimes \mathbb{E}\mathbf{Y} = 0$.

Proof: For (i) \implies (ii), and $\mathbf{x}, \mathbf{y} \neq 0$, choose ON-bases such that $\mathbf{e}_1 = \mathbf{x}/\|\mathbf{x}\|$ and $\mathbf{e}'_1 = \mathbf{y}/\|\mathbf{y}\|$. The rest is immediate from Theorem 8.1. \square

Székely et al. (2007) observed that for $\alpha = 2$ and real-valued variables, $\text{dcov}_2(\mathbf{X}, \mathbf{Y}) = 0$ does not characterize independence but instead that \mathbf{X} and \mathbf{Y} are uncorrelated; Corollary 8.2 extends this to Hilbert spaces, in the sense (ii) or (iii) above.

Remark 8.3. $\mathbb{E}(\mathbf{X} \otimes \mathbf{Y}) - \mathbb{E}\mathbf{X} \otimes \mathbb{E}\mathbf{Y} \in \mathcal{H} \otimes \mathcal{H}'$ can be regarded as the covariance of the vector-valued variables \mathbf{X} and \mathbf{Y} ; cf. the general theory of higher moments of Banach space valued variables in Janson and Kaijser (2015), where the moment lives in a suitable tensor product. (The general theory in Janson and Kaijser (2015) focusses on a single variable and on the projective and injective tensor products, but see Janson and Kaijser (2015, Remarks 3.24 and 3.25). Since we assume separable spaces and $\mathbb{E}\|\mathbf{X}\|^2, \mathbb{E}\|\mathbf{Y}\|^2 < \infty$, there are no problems with integrability; cf. Janson and Kaijser (2015, Theorem 5.14).)

Remark 8.4. Let \mathcal{X} and \mathcal{Y} both be metric spaces such that d^α is a semimetric of negative type. Then, see Remark 2.7, there are embeddings $\varphi : \mathcal{X} \rightarrow \mathcal{H}$ and $\varphi' : \mathcal{Y} \rightarrow \mathcal{H}'$ into Hilbert spaces such that

$$d_{\mathcal{X}}(\mathbf{x}_1, \mathbf{x}_2)^\alpha = \|\varphi(\mathbf{x}_1) - \varphi(\mathbf{x}_2)\|^2, \quad d_{\mathcal{Y}}(\mathbf{y}_1, \mathbf{y}_2)^\alpha = \|\varphi'(\mathbf{y}_1) - \varphi'(\mathbf{y}_2)\|^2. \tag{8.13}$$

It follows immediately that, for any of Definitions 2.1–2.3,

$$\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y}) = \text{dcov}_2(\varphi(\mathbf{X}), \varphi'(\mathbf{Y})). \tag{8.14}$$

Hence, $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ can be interpreted as in Theorem 8.1 for the embedded variables, as shown (for $\alpha = 1$) in Lyons (2013, Proposition 3.7).

Remark 8.5. The Hilbert space tensor product $\mathcal{H} \otimes \mathcal{H}'$ can be identified with the space of Hilbert–Schmidt operators $\mathcal{H} \rightarrow \mathcal{H}'$ (see (B.18) and the proof of Lemma B.7); then $\mathbb{E}(\mathbf{X} \otimes \mathbf{Y}) - \mathbb{E} \mathbf{X} \otimes \mathbb{E} \mathbf{Y} = \mathbb{E}[(\mathbf{X} - \mathbb{E} \mathbf{X}) \otimes (\mathbf{Y} - \mathbb{E} \mathbf{Y})]$ corresponds to the operator $\mathbf{x} \mapsto \mathbb{E}[\langle \mathbf{x}, \mathbf{X} - \mathbb{E} \mathbf{X} \rangle (\mathbf{Y} - \mathbb{E} \mathbf{Y})]$, known as the *covariance operator* (or *cross-covariance operator*, Baker, 1973). Thus Theorem 8.1 says that $\text{dcov}_2(\mathbf{X}, \mathbf{Y})$ is 4 times the squared Hilbert–Schmidt norm of the covariance operator.

More generally, if \mathcal{X} and \mathcal{Y} both are metric spaces such that d^α is a semimetric of negative type, then (8.14) shows that $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ equals $\text{dcov}_2(\varphi(\mathbf{X}), \varphi'(\mathbf{Y}))$ for some embeddings $\varphi : \mathcal{X} \rightarrow \mathcal{H}$ and $\varphi' : \mathcal{Y} \rightarrow \mathcal{H}'$ into Hilbert spaces. Hence, $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ equals 4 times the squared Hilbert–Schmidt norm of the covariance operator corresponding to the embedded variables, as shown in Sejdinovic et al. (2013, Theorem 24); this Hilbert–Schmidt norm (or its square) is called the *Hilbert–Schmidt independence criterion (HSIC)* Gretton et al. (2005), Sejdinovic et al. (2013, §3.3); cf. Remark 2.10.

Remark 8.6. If α is an even integer larger than 2, we can similarly express dcov_α in moments of \mathbf{X} and \mathbf{Y} , but the resulting formulas are complicated and do not seem to be of any interest. For example, for $\alpha = 4$, for $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, and taking for simplicity $\mathbf{X} = \mathbf{Y}$ with $\mathbb{E} \mathbf{X} = 0$,

$$\begin{aligned} \text{dcov}_4(\mathbf{X}, \mathbf{X}) &= 32 \mathbb{E}[\mathbf{X}^2] \mathbb{E}[\mathbf{X}^6] - 96 \mathbb{E}[\mathbf{X}^3] \mathbb{E}[\mathbf{X}^5] + 68(\mathbb{E}[\mathbf{X}^4])^2 \\ &\quad - 72(\mathbb{E}[\mathbf{X}^2])^2 \mathbb{E}[\mathbf{X}^4] + 64 \mathbb{E}[\mathbf{X}^2] (\mathbb{E}[\mathbf{X}^3])^2 + 36(\mathbb{E}[\mathbf{X}^2])^4. \end{aligned} \tag{8.15}$$

We do not know any application or interesting properties of dcov_α with $\alpha > 2$.

9. Optimality of moment conditions

We have so far assumed the moment conditions stated in Definitions 2.1–2.3; these seem natural and convenient for applications. Nevertheless, it is of interest to study whether they really are required for the definitions, and what happens when we try to extend one of the definitions to cases when the moment condition fails. Definitions 2.4 and 7.1 are stated without moment conditions, but we similarly can ask when the results are finite and whether they agree with the other definitions.

In this section, we will give examples showing that the moment conditions in Definitions 2.1–2.3 are optimal in general, in the sense that if we reduce the exponent in the moment condition, then there exist counterexamples where the definition either yields an infinite value or is meaningless. On the other hand, there are also cases where the moment conditions do not hold but the definitions yield a finite value. We explore these possibilities in the next section, but our results are incomplete, and we leave a number of (explicit or implicit) open problems.

In general, if we try to define $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ by (2.2) or (2.3) for some \mathbf{X} and \mathbf{Y} , there are three possibilities:

- (dc1) The expression yields a finite value; this may then be taken to be $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$. This happens when all expectations in (2.2) or (2.3), respectively, are finite. (For (2.2), it also includes the trivial case when \mathbf{X} or \mathbf{Y} is degenerate, so $d(\mathbf{X}_i, \mathbf{X}_j) = 0$ a.s. or $d(\mathbf{Y}_i, \mathbf{Y}_j) = 0$ a.s.; then all terms in (2.2) are 0, if necessary interpreting $0 \cdot \infty = 0$.)

- (dc2) The expression makes sense as either $+\infty$ or $-\infty$. We may then take it as defining $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$, now with an infinite value in $\{-\infty, \infty\}$. (We do not know whether $-\infty$ can happen, see Problem 10.20.) Thus, at least one expectation is infinite. Furthermore, for (2.2), where all expectations are of non-negative variables and thus defined in $[0, \infty]$, this means that either the two first expectations are finite, or the third expectation is finite; for (2.3) this means that one of $\mathbb{E}[(\widehat{X}_\alpha \widehat{Y}_\alpha)_+]$ and $\mathbb{E}[(\widehat{X}_\alpha \widehat{Y}_\alpha)_-]$ is finite and the other infinite, so the expectation $\mathbb{E}[\widehat{X}_\alpha \widehat{Y}_\alpha]$ is defined as $+\infty$ or $-\infty$.
- (dc3) The expression (2.2) or (2.3) is of the type $\infty - \infty$. Then it is meaningless, and $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ is undefined (by this definition).

For Definition 2.3, we have the same possibilities as for Definition 2.2, but also the complication that \widetilde{X}_α and \widetilde{Y}_α have to be defined, see (2.6)–(2.7). We thus have another bad case:

- (dc4) \widetilde{X}_α or \widetilde{Y}_α is not defined. Then $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$ is undefined.

For Euclidean spaces, we also have Definition 2.4, and for Hilbert spaces we have Definition 7.1. Since (2.9) and (7.2)–(7.4) are integrals of non-negative functions, Definitions 2.4 and 7.1 are always meaningful, but may yield $+\infty$. In other words, we have only the cases (dc1) and (dc2). Again we may ask when the definition yields a finite value, and when it agrees with other definitions; in particular whether the moment conditions in Theorem 7.4 are best possible.

The moment conditions assumed in Definitions 2.1–2.3 guarantee, as seen in Theorem 4.5, that the good case (dc1) occurs. In the following subsections we investigate more generally when the cases (dc1)–(dc4) occur, and whether the different definitions still agree when more than one of them applies.

9.1. *Optimality in Definition 2.1.* We begin with Definition 2.1, where we have a simple necessary and sufficient condition.

Theorem 9.1. (i) *If $\mathbb{E}\|\mathbf{X}\|^\alpha + \mathbb{E}\|\mathbf{Y}\|^\alpha + \mathbb{E}[\|\mathbf{X}\|^\alpha \|\mathbf{Y}\|^\alpha] < \infty$, then all expectations in (2.2) are finite, so (2.2) defines $\text{dcov}_\alpha^*(\mathbf{X}, \mathbf{Y})$ as a finite number.*

Moreover, in this case also the definitions (2.3) and (2.5) yield the same result, i.e., $\text{dcov}_\alpha^(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$.*

(ii) *Conversely, if $\mathbb{E}\|\mathbf{X}\|^\alpha + \mathbb{E}\|\mathbf{Y}\|^\alpha + \mathbb{E}[\|\mathbf{X}\|^\alpha \|\mathbf{Y}\|^\alpha] = \infty$, and \mathbf{X} and \mathbf{Y} are non-degenerate, then (2.2) is of the type $\infty - \infty$ and thus meaningless.*

In particular, Case (dc2), i.e., a well-defined infinite value of dcov_α^* , never occurs for Definition 2.1.

Proof: (i): This follows by minor modifications of the argument used under slightly stronger assumptions in Section 2 and Lemma 4.1. Note that the assumption implies that $\mathbb{E}[\|\mathbf{X}_i\|^\alpha \|\mathbf{Y}_j\|^\alpha] < \infty$ for all i and j , and thus it follows from the triangle inequality (4.4) that all expectations in (2.2) are finite. Moreover, the assumption implies, using (4.4) again, that $\widehat{X}_\alpha, \widehat{Y}_\alpha \in L^1$, and thus \widetilde{X}_α and \widetilde{Y}_α are defined by (2.6)–(2.7), and also that $\widehat{X}_\alpha \widehat{Y}_\alpha \in L^1$ and $\widetilde{X}_\alpha \widetilde{Y}_\alpha \in L^1$. We omit the details.

(ii): If $\mathbb{E}\|\mathbf{X}\|^\alpha = \infty$, then $\mathbb{E}d(\mathbf{x}, \mathbf{X}_2)^\alpha = \infty$ for any \mathbf{x} , and thus, by first conditioning on $(\mathbf{X}_1, \mathbf{Y}_1)$ and $(\mathbf{X}_3, \mathbf{Y}_3)$ and integrating over \mathbf{X}_2 only, both $\mathbb{E}[d(\mathbf{X}_1, \mathbf{X}_2)^\alpha] = \infty$ and $\mathbb{E}[d(\mathbf{X}_1, \mathbf{X}_2)^\alpha d(\mathbf{Y}_1, \mathbf{Y}_3)^\alpha] = \infty$; hence, since $\mathbb{E}[d(\mathbf{Y}_1, \mathbf{Y}_2)^\alpha] > 0$, we see that (2.2) is of the type $\infty - \infty$.

By symmetry, the same holds if $\mathbb{E}\|\mathbf{Y}\|^\alpha = \infty$.

Finally, suppose that $\mathbb{E}[\|\mathbf{X}\|^\alpha \|\mathbf{Y}\|^\alpha] = \infty$. By the cases just treated, we may assume that also $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^\alpha < \infty$. Then, using the triangle inequality and integrating only over the event $\{\|\mathbf{X}_2\|, \|\mathbf{Y}_2\|, \|\mathbf{Y}_3\| \leq M\}$, for an M so large that this event has positive probability, we see that both the first and last expectations in (2.2) are ∞ , and thus (2.2) is $\infty - \infty$. \square

Remark 9.2. If Theorem 9.1(i) applies and $\text{dcov}_\alpha^E(\mathbf{X}, \mathbf{Y})$ or $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$ is defined, i.e., if $\alpha < 2$ and the spaces are Euclidean spaces or Hilbert spaces, respectively, then it too equals $\text{dcov}_\alpha^*(\mathbf{X}, \mathbf{Y})$. This follows by Theorem 9.1 together with Theorems 7.2 and 7.4.

Example 9.3. If \mathbf{X} and \mathbf{Y} are independent with $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$ and $\mathbb{E}\|\mathbf{Y}\|^\alpha < \infty$, then Theorem 9.1(i) applies and (2.2) makes perfect sense; Definitions 2.1–2.3 all can be used, and all yield 0.

Example 9.4. Let \mathbf{X} be arbitrary with $\mathbb{E}\|\mathbf{X}\|^{2\alpha} = \infty$, and let $\mathbf{Y} = \mathbf{X}$. Then, Theorem 9.1(ii) shows that $\text{dcov}_\alpha^*(\mathbf{X}, \mathbf{X})$ is of the type $\infty - \infty$ and does not make sense. Consequently, in general, the moment condition in Definition 2.1 is necessary. (In particular, for every (\mathbf{X}, \mathbf{Y}) with $\mathbf{Y} = \mathbf{X}$.)

9.2. *Optimality in Definition 2.2.* We have already seen in Example 9.4 that the moment condition in Definition 2.1 is necessary, in a strong sense. We next show that the moment conditions in Definitions 2.2 and 2.3 also are optimal, in the sense that if we reduce the exponent, there are counterexamples. However, there are also examples where these definitions yield finite values although the moment condition fails.

Consider first Definition 2.2. \widehat{X}_α and \widehat{Y}_α are always defined by (2.4), so the question is whether $\mathbb{E}[\widehat{X}_\alpha \widehat{Y}_\alpha]$ exists or not, and whether its value is finite or not. Note, in particular, that $\text{dcov}_\alpha(\widehat{\mathbf{X}}, \widehat{\mathbf{X}}) := \frac{1}{4} \mathbb{E} \widehat{X}_\alpha^2$ always is defined, although it may be $+\infty$; we have

$$\text{dcov}_\alpha(\widehat{\mathbf{X}}, \widehat{\mathbf{X}}) < \infty \iff \widehat{X}_\alpha \in L^2. \tag{9.1}$$

Note also that, by rotational symmetry in the indices in (2.4), \widehat{X}_α has a symmetric distribution. Thus $\mathbb{E} \widehat{X}_\alpha = 0$ whenever the expectation exists.

Example 9.5. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, and suppose that $\mathbf{X} \geq 0$ with $\mathbb{P}(\mathbf{X} = 0) > 0$. On the event $\mathbf{X}_3 = \mathbf{X}_4 = 0$, we have

$$-\widehat{X}_\alpha = \mathbf{X}_1^\alpha + \mathbf{X}_2^\alpha - |\mathbf{X}_1 - \mathbf{X}_2|^\alpha \geq \mathbf{X}_1^\alpha \wedge \mathbf{X}_2^\alpha. \tag{9.2}$$

Hence, if $\mathbb{E}|\widehat{X}_\alpha|^2 < \infty$, then

$$\begin{aligned} \infty &> \mathbb{E}[(\mathbf{X}_1^\alpha \wedge \mathbf{X}_2^\alpha)^2] = \mathbb{E}[\mathbf{X}_1^{2\alpha} \wedge \mathbf{X}_2^{2\alpha}] \\ &= \int_0^\infty \mathbb{P}[\mathbf{X}_1^{2\alpha} \wedge \mathbf{X}_2^{2\alpha} > t] dt = \int_0^\infty \mathbb{P}[\mathbf{X}^{2\alpha} > t]^2 dt \\ &= 2\alpha \int_0^\infty \mathbb{P}[\mathbf{X} > x]^2 x^{2\alpha-1} dx. \end{aligned} \tag{9.3}$$

If we choose \mathbf{X} such that, for $x \geq 2$ say,

$$\mathbb{P}(\mathbf{X} > x) = x^{-\alpha}, \tag{9.4}$$

then $\mathbb{E}|\mathbf{X}|^\gamma < \infty$ for every $\gamma < \alpha$, but the integral in (9.3) diverges and thus $\mathbb{E}|\widehat{X}_\alpha|^2 = \infty$; hence (2.3) yields $\text{dcov}_\alpha(\widehat{\mathbf{X}}, \widehat{\mathbf{X}}) = \infty$ by (9.1). (Case (dc2).) Consequently, when $\alpha \leq 2$, the exponent $\alpha^* = \alpha$ is optimal in Definition 2.2 (in order to yield a finite value).

Example 9.6. Let $\alpha > 1$ and $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, and suppose, for simplicity, that $\mathbf{X} \geq 0$ with $\mathbb{P}(\mathbf{X} = 0) = \mathbb{P}(\mathbf{X} = 1) = 1/4$. On the event $\mathbf{X}_2 = \mathbf{X}_3 = 0, \mathbf{X}_4 = 1$, we have for some $c > 0$, assuming that $\mathbf{X}_1 \geq 2$, say,

$$\widehat{X}_\alpha = \mathbf{X}_1^\alpha - |\mathbf{X}_1 - 1|^\alpha + 1 \geq c\mathbf{X}_1^{\alpha-1}. \tag{9.5}$$

Hence, for these values of $\mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$, we have $\widehat{X}_\alpha \geq c\mathbf{X}_1^{\alpha-1} - C$. Consequently, $\mathbb{E}|\widehat{X}_\alpha|^2 < \infty \implies \mathbb{E} \mathbf{X}^{2(\alpha-1)} < \infty$.

We can choose \mathbf{X} as above such that $\mathbb{E} \mathbf{X}^\gamma < \infty$ for every $\gamma < 2\alpha - 2$, but $\mathbb{E} \mathbf{X}^{2\alpha-2} = \infty$ and consequently $\mathbb{E}|\widehat{X}_\alpha|^2 = \infty$; thus, (2.3) yields $\text{dcov}_\alpha(\widehat{\mathbf{X}}, \widehat{\mathbf{X}}) = \infty$. (Case (dc2).) Hence, when $\alpha \geq 2$, the exponent $\alpha^* = 2\alpha - 2$ is optimal in Definition 2.2.

Example 9.7. We have here given examples with $\mathbb{E}|\widehat{X}_\alpha|^2 = \infty$, so that (2.3) gives $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X}) = +\infty$.

Similarly, (9.2) and a calculation as in (9.3) show that if, say, $\mathbb{P}(\mathbf{X} > x) = x^{-\alpha/2}$ for $x > 2$, then $\mathbb{E}|\widehat{X}_\alpha| = \infty$. Since \widehat{X}_α has a symmetric distribution by (2.3), it follows that if \mathbf{Y} is any non-degenerate random variable such that \mathbf{X} and \mathbf{Y} are independent, then the expectation in (2.3) is of the type $\infty - \infty$ and thus undefined (Case (dc3) above); hence Definition 2.2 cannot be applied at all (even allowing $\pm\infty$ as a result).

Example 9.8. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and consider the special (and rather exceptional) case $\alpha = 2$, cf. Section 8. Then \widehat{X}_2 is given by (8.3), and it follows easily that

$$\widehat{X}_2 \in L^2 \iff \mathbb{E}|\mathbf{X}|^2 < \infty, \quad (9.6)$$

and that for $\mathbf{X} = \mathbf{Y}$, (8.4) holds in the form

$$\text{dcov}_2^\wedge(\mathbf{X}, \mathbf{X}) = \frac{1}{4} \mathbb{E}[\widehat{X}_2^2] = 4(\text{Var } \mathbf{X})^2 \quad (9.7)$$

for any \mathbf{X} , where the expressions all are infinite when $\mathbb{E}|\mathbf{X}|^2 = \infty$. This shows again that the condition of finite α^* moment in Definition 2.2 cannot be improved when $\alpha = 2$, if we want $\text{dcov}_2(\mathbf{X}, \mathbf{X})$ to be finite. Furthermore, if we take $\mathbf{Y} = \zeta \mathbf{X}$ where \mathbf{X} and ζ are independent with $\mathbb{E} \mathbf{X} = 0$, $\mathbb{E}[\mathbf{X}^2] = \infty$, $\zeta \in \{\pm 1\}$ and $\mathbb{E} \zeta = 0$, then $\mathbb{E}[\widehat{X}_2 \widehat{Y}_2]$ is of the type $\infty - \infty$; hence, even allowing infinite values, $\text{dcov}_2^\wedge(\mathbf{X}, \mathbf{Y})$ cannot be defined by Definition 2.2 without assuming second moments.

9.3. Optimality in Definition 2.3. We now turn to Definition 2.3. As noted above, \widetilde{X}_α is only defined for some \mathbf{X} . If we use the conditional expectation definition in (2.6), then we have to require $\widehat{X}_\alpha \in L^1$, i.e., $\mathbb{E}|\widehat{X}_\alpha| < \infty$. On the other hand, the explicit formula (2.7) makes sense only if $\mathbb{E}d(\mathbf{X}_1, \mathbf{X}_2)^\alpha < \infty$, or equivalently $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$, since otherwise also the conditional expectations in (2.7) are $+\infty$ a.s., and thus (2.7) is $\infty - \infty$. Moreover, if $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$, then $\widehat{X}_\alpha \in L^1$ by (2.4), and (2.6) agrees with (2.7). Hence we may take (2.6) as the primary definition of \widetilde{X}_α , and say that \widetilde{X}_α is defined when $\widehat{X}_\alpha \in L^1$. This holds in particular when $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$, and then (2.7) holds too, but note that Lemma 4.2 shows that $\mathbb{E}\|\mathbf{X}\|^{\alpha^*/2} < \infty$ suffices for $\widehat{X}_\alpha \in L^1$.

Hence, \widetilde{X}_α is defined if and only if $\widehat{X}_\alpha \in L^1$, and then $\widetilde{X}_\alpha \in L^1$; furthermore

$$\mathbb{E} \widetilde{X}_\alpha = \mathbb{E} \mathbb{E}(\widehat{X}_\alpha \mid \mathbf{X}_1, \mathbf{X}_2) = \mathbb{E} \widehat{X}_\alpha = 0. \quad (9.8)$$

Moreover, in this case, also

$$\mathbb{E}(\widetilde{X}_\alpha \mid \mathbf{X}_1) = \mathbb{E}(\mathbb{E}(\widehat{X}_\alpha \mid \mathbf{X}_1, \mathbf{X}_2) \mid \mathbf{X}_1) = \mathbb{E}(\widehat{X}_\alpha \mid \mathbf{X}_1) = 0, \quad (9.9)$$

since \widehat{X}_α has a symmetric distribution also when conditioned on \mathbf{X}_1 , by symmetry in (2.4).

Example 9.9. Recall that Example 9.7 gives an example where $\widehat{X}_\alpha \notin L^1$; hence, \widetilde{X}_α is not defined and thus $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X})$ is undefined.

We note a general result relating \widetilde{X}_α and \widehat{X}_α . By (2.6), \widetilde{X}_α is (a.s.) a function of \mathbf{X}_1 and \mathbf{X}_2 ; let us (temporarily) write \widetilde{X}_α as $\widetilde{X}_\alpha(\mathbf{X}_1, \mathbf{X}_2)$, so that we can substitute other \mathbf{X}_i as arguments. The following lemma shows that \widehat{X}_α can be recovered from \widetilde{X}_α .

Lemma 9.10. *Suppose that $\widehat{X}_\alpha \in L^1$. Then, a.s.,*

$$\widehat{X}_\alpha = \widetilde{X}_\alpha(\mathbf{X}_1, \mathbf{X}_2) - \widetilde{X}_\alpha(\mathbf{X}_2, \mathbf{X}_3) + \widetilde{X}_\alpha(\mathbf{X}_3, \mathbf{X}_4) - \widetilde{X}_\alpha(\mathbf{X}_4, \mathbf{X}_1). \quad (9.10)$$

Consequently, for any $p \geq 1$,

$$\widetilde{X}_\alpha \text{ exists and } \widetilde{X}_\alpha \in L^p \iff \widehat{X}_\alpha \in L^p. \quad (9.11)$$

Proof: If $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$, this is obvious from (2.7) and cancellations. In general, we use truncations. Let, for $M > 0$, $I^M := \mathbf{1}\{|\mathbf{X}| \leq M\}$, $I_i^M := \mathbf{1}\{|\mathbf{X}_i| \leq M\}$, and let $p_M := \mathbb{E}I^M = P(|\mathbf{X}_i| \leq M)$. Then,

$$\begin{aligned} &\mathbb{E}(I_3^M I_4^M \widehat{X}_\alpha \mid \mathbf{X}_1, \mathbf{X}_2) \\ &= p_M^2 d(\mathbf{X}_1, \mathbf{X}_2)^\alpha - p_M \mathbb{E}_{\mathbf{X}}(I^M d(\mathbf{X}_2, \mathbf{X})^\alpha) \\ &\quad + p_M^2 \mathbb{E}(I_3^M I_4^M d(\mathbf{X}_3, \mathbf{X}_4)^\alpha) - p_M \mathbb{E}_{\mathbf{X}}(I^M d(\mathbf{X}_1, \mathbf{X})^\alpha) \end{aligned} \tag{9.12}$$

and consequently, by rotational symmetry and cancellations, interpreting all indices modulo 4,

$$\sum_{i=1}^4 (-1)^{i-1} \mathbb{E}(I_{i+2}^M I_{i+3}^M \widehat{X}_\alpha \mid \mathbf{X}_i, \mathbf{X}_{i+1}) = p_M^2 \sum_{i=1}^4 (-1)^{i-1} d(\mathbf{X}_i, \mathbf{X}_{i+1})^\alpha = p_M^2 \widehat{X}_\alpha. \tag{9.13}$$

Since we assume $\widehat{X}_\alpha \in L^1$, we have $I_{i+2}^M I_{i+3}^M \widehat{X}_\alpha \xrightarrow{L^1} \widehat{X}_\alpha$ as $M \rightarrow \infty$, and thus

$$\mathbb{E}(I_{i+2}^M I_{i+3}^M \widehat{X}_\alpha \mid \mathbf{X}_i, \mathbf{X}_{i+1}) \xrightarrow{L^1} \mathbb{E}(\widehat{X}_\alpha \mid \mathbf{X}_i, \mathbf{X}_{i+1}) = \widetilde{X}_\alpha(\mathbf{X}_i, \mathbf{X}_{i+1}). \tag{9.14}$$

Hence, as $M \rightarrow \infty$, the left-hand side of (9.13) converges in L^1 to the right-hand side of (9.10), while the right-hand side of (9.13) obviously converges to \widehat{X}_α . Hence, (9.10) follows.

Finally, (9.11) is an immediate consequence of (9.10) and (2.6). □

In particular, this leads to the following for the case $\mathbf{Y} = \mathbf{X}$. Note that $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X}) := \mathbb{E} \widetilde{X}_\alpha^2$ is defined whenever \widetilde{X}_α is, although it may be $+\infty$; cf. $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X})$ discussed above.

Theorem 9.11. *Let \mathbf{X} be a random variable in a metric space. Then the following are equivalent:*

- (i) $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X}) < \infty$.
- (ii) $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X}) < \infty$ (which includes that \widetilde{X}_α is defined).
- (iii) $\widehat{X}_\alpha \in L^2$.
- (iv) \widetilde{X}_α is defined and $\widetilde{X}_\alpha \in L^2$.

Furthermore, if these hold, then $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X}) = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X})$.

Proof: (i) \iff (iii): This follows directly from the definition (2.3), as noted in (9.1).

(ii) \iff (iv): Follows similarly from the definition (2.5).

(iii) \iff (iv): By Lemma 9.10.

Finally, suppose that (i)–(iv) hold. Use (9.10) and expand $(\widehat{X}_\alpha)^2$ as a sum of products. Since $\widetilde{X}_\alpha \in L^2$, each product is in L^1 , so we may take their expectations separately. Furthermore, (9.9) implies that all off-diagonal terms such as $\mathbb{E}[\widetilde{X}_\alpha(\mathbf{X}_1, \mathbf{X}_2)\widetilde{X}_\alpha(\mathbf{X}_2, \mathbf{X}_3)] = 0$, and we obtain

$$\mathbb{E}[\widehat{X}_\alpha^2] = \sum_{i=1}^4 \mathbb{E}[\widetilde{X}_\alpha(\mathbf{X}_i, \mathbf{X}_{i+1})^2] = 4 \mathbb{E}[\widetilde{X}_\alpha^2]. \tag{9.15}$$

Hence, $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X}) = \frac{1}{4} \mathbb{E}[\widehat{X}_\alpha^2] = \mathbb{E}[\widetilde{X}_\alpha^2] = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X})$. □

Corollary 9.12. (i) *If $\widehat{X}_\alpha \in L^1$, so \widetilde{X}_α is defined, then $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X}) = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X})$ (finite or infinite).*

(ii) *If $\widehat{X}_\alpha \notin L^1$, then $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X}) = \infty$ and $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X})$ is undefined.*

Proof: Follows from Theorem 9.11, considering the three cases $\widehat{X}_\alpha \in L^2$, $\widehat{X}_\alpha \in L^1 \setminus L^2$ and $\widehat{X}_\alpha \notin L^1$ separately. □

If we only care about finite values and regard ∞ as 'undefined', we thus see that $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X})$ for all \mathbf{X} .

Example 9.13. Let $\alpha \leq 2$ and let $\mathbf{X} \in \mathbb{R}$ be as in Example 9.5; thus $\mathbf{X} \geq 0$ and (9.4) holds. We have $\mathbb{E}|\mathbf{X}|^\gamma < \infty$ for every $\gamma < \alpha$, and in particular $\mathbb{E}|\mathbf{X}|^{\alpha/2} < \infty$; hence Lemma 4.2(ii) implies that $\widehat{X}_\alpha \in L^1$. Thus, \widetilde{X}_α exists, but $\widehat{X}_\alpha \notin L^2$ by Example 9.5; hence Theorem 9.11 shows that $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X}) = \infty$. Consequently, the exponent $\alpha^* = \alpha$ is optimal in Definition 2.3 when $\alpha \leq 2$.

Example 9.14. Similarly, let $\alpha > 2$ and let $\mathbf{X} \in \mathbb{R}$ be as in Example 9.6. Then, $\mathbb{E}|\mathbf{X}|^\gamma < \infty$ for every $\gamma < 2\alpha - 2$, and in particular $\mathbb{E}|\mathbf{X}|^{\alpha-1} < \infty$; hence Lemma 4.2(iii) implies that $\widehat{X}_\alpha \in L^1$. Thus, \widetilde{X}_α exists, but $\widehat{X}_\alpha \notin L^2$ by Example 9.6; hence Theorem 9.11 shows that $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X}) = \infty$. Consequently, the exponent $\alpha^* = 2\alpha - 2$ is optimal in Definition 2.3 when $\alpha > 2$.

Hence, the exponent α^* is optimal in Definition 2.3 too.

Example 9.15. Let $\alpha = 2$ and $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ as in Example 9.8, and assume that $\mathbb{E}\mathbf{X} = 0$. Then, by (8.3), \widetilde{X}_2 exists and $\widetilde{X}_2 = -2\mathbf{X}_1\mathbf{X}_2$. Hence, we find directly the same conclusions for dcov_α^\sim as found for $\text{dcov}_\alpha^\wedge$ in Example 9.8. In particular, with \mathbf{X} and $\mathbf{Y} = \zeta\mathbf{X}$ as in the final part of Example 9.8, $\mathbb{E}[\widetilde{X}_2\widetilde{Y}_2]$ is of the type $\infty - \infty$ and thus undefined. (Case (dc3).)

9.4. *Optimality for dcov_α^E and dcov_α^H .* Definitions 2.4 and 7.1 do not require any moment conditions; if \mathcal{X} and \mathcal{Y} are Euclidean spaces or Hilbert spaces, respectively, then $\text{dcov}_\alpha^E(\mathbf{X}, \mathbf{Y})$ and $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$ are always defined, but may be $+\infty$. (Recall also that Theorem 7.2 shows that for spaces where both are defined, we always have $\text{dcov}_\alpha^E(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$, finite or not.) Theorem 7.4 shows that the moment condition $\mathbb{E}\|\mathbf{X}\|^\alpha, \mathbb{E}\|\mathbf{Y}\|^\alpha < \infty$ is sufficient to guarantee that $\text{dcov}_\alpha^E(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$ is finite. (Recall that this is the same moment condition as in Definitions 2.2 and 2.3.) The following example shows that the exponent α in this moment condition is optimal, even for random variables in \mathbb{R} .

Example 9.16. Let $0 < \alpha < 2$, and let \mathbf{X} be a symmetric stable random variable in \mathbb{R} with the characteristic function $\varphi_{\mathbf{X}}(t) = e^{-|t|^\alpha}$. Then $\mathbb{E}|\mathbf{X}|^\alpha = \infty$, but $\mathbb{E}|\mathbf{X}|^\gamma < \infty$ for every $\gamma < \alpha$.

Take $\mathbf{Y} = \mathbf{X}$. Then, for $0 \leq t \leq 1$ and $t \leq u \leq 2t$,

$$\begin{aligned} \varphi_{\mathbf{X},\mathbf{X}}(t, -u) - \varphi_{\mathbf{X}}(t)\varphi_{\mathbf{X}}(-u) &= e^{-|t-u|^\alpha} - e^{-|t|^\alpha - |u|^\alpha} \\ &\geq e^{-t^\alpha} - e^{-2t^\alpha} \geq ct^\alpha, \end{aligned} \tag{9.16}$$

for some $c > 0$. Consequently, (2.9) yields, changing the sign of u ,

$$\text{dcov}_\alpha^E(\mathbf{X}, \mathbf{X}) \geq c \int_{t=0}^1 \int_{u=t}^{2t} t^{2\alpha} \frac{dt du}{t^{1+\alpha}u^{1+\alpha}} = c \int_{t=0}^1 \frac{t^{2\alpha}}{t^{1+2\alpha}} dt = \infty. \tag{9.17}$$

Hence, using Theorem 7.2, $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{X}) = \text{dcov}_\alpha^E(\mathbf{X}, \mathbf{X}) = \infty$. The condition in Theorem 7.4 on finite α moments thus cannot be replaced by any lower moments in order to guarantee finite values.

10. Beyond the moment conditions

We continue to investigate cases when the moment condition in Definitions 2.2–2.3 fails; now with the aim of obtaining positive results.

10.1. *A weaker condition.* We begin with $\text{dcov}_\alpha^\wedge$ in Definition 2.2, and show first that the counterexample in Example 9.5 is optimal, at least when $\alpha \leq 1$.

Theorem 10.1. *Let \mathcal{X} be any separable metric space, and let $0 < \alpha \leq 1$. If*

$$\int_0^\infty \mathbb{P}[\|\mathbf{X}\| > x]^2 x^{2\alpha-1} dx < \infty, \tag{10.1}$$

then $\mathbb{E}|\widehat{X}_\alpha|^2 < \infty$ and thus $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X}) < \infty$.

Proof: The calculation in (9.3) shows that (10.1) is equivalent to

$$\mathbb{E}[(\|\mathbf{X}_1\|^\alpha \wedge \|\mathbf{X}_2\|^\alpha)^2] < \infty. \tag{10.2}$$

In other words, $\|\mathbf{X}_1\|^\alpha \wedge \|\mathbf{X}_2\|^\alpha \in L^2$. Hence, Lemma 4.2(i) shows that $\widehat{X}_\alpha \in L^2$. □

Remark 10.2. Let $0 < \alpha \leq 1$. Then, e.g. using Lemma 10.4 below, the argument in Example 9.5 is easily extended to show that if $\mathcal{X} = \mathbb{R}$, then (10.1) is also necessary for $\mathbb{E}|\widehat{X}_\alpha|^2 < \infty$. Thus, at least for $\alpha \leq 1$ and $\mathcal{X} = \mathbb{R}$, (10.1) is both necessary and sufficient for $\text{dcov}_\alpha(\mathbf{X}, \mathbf{X}) < \infty$.

Remark 10.3. It is easy to see directly that the condition (10.1) follows from the condition $\mathbb{E}\|\mathbf{X}\|^\alpha < \infty$ in Lemma 4.3. (We omit the details.) Furthermore, (10.1) is a strictly weaker condition, and thus, for $\alpha \leq 1$, Theorem 10.1 is stronger than Lemma 4.3. For example, if we instead of (9.4) choose, for $x \geq e$,

$$\mathbb{P}(\mathbf{X} > x) = x^{-\alpha} / \log x, \tag{10.3}$$

then $\mathbb{E}\mathbf{X}^\alpha = \infty$, but the integral in (9.3) converges and Theorem 10.1 shows that $\mathbb{E}|\widehat{X}_\alpha|^2 < \infty$ and $\text{dcov}_\alpha(\mathbf{X}, \mathbf{X}) < \infty$.

Hence, although we have seen that the exponent in the moment condition in Definition 2.2 is best possible, Theorem 10.1 shows that for $\alpha \leq 1$, the moment condition can be weakened to the condition (10.1) (together with the same for \mathbf{Y}); we postpone the details to Theorem 10.6, where we also extend it to dcov_α^E and dcov_α^H .

Before proceeding, we note that when $\alpha \leq 1$, we may simplify the condition $\widehat{X}_\alpha \in L^2$ by the following lemma.

Lemma 10.4. *Let $p > 0$. If $0 < \alpha \leq 1$, then*

$$\widehat{X}_\alpha \in L^p \iff \|\mathbf{X}_1\|^\alpha + \|\mathbf{X}_2\|^\alpha - d(\mathbf{X}_1, \mathbf{X}_2)^\alpha \in L^p. \tag{10.4}$$

Note that (for $\alpha \leq 1$) the right-hand side is non-negative by the triangle inequality.

Proof: \implies : Since $\mathbb{E}|\widehat{X}_\alpha|^p < \infty$, the conditional expectation $\mathbb{E}(|\widehat{X}_\alpha|^p \mid X_3, X_4) \in L^1$. Hence, there exist some \mathbf{x}_3 and \mathbf{x}_4 such that $\mathbb{E}(|\widehat{X}_\alpha|^p \mid X_3 = \mathbf{x}_3, X_4 = \mathbf{x}_4) \in L^1$, which by the definition (2.4) means

$$d(\mathbf{X}_1, \mathbf{X}_2)^\alpha - d(\mathbf{X}_1, \mathbf{x}_4)^\alpha - d(\mathbf{X}_2, \mathbf{x}_3)^\alpha \in L^p. \tag{10.5}$$

The triangle inequality yields, for $j = 3, 4$,

$$|d(\mathbf{X}, \mathbf{x}_j) - \|\mathbf{X}\|^\alpha| \leq d(\mathbf{x}_j, \mathbf{x}_o) = O(1), \tag{10.6}$$

and thus, since $\alpha \leq 1$,

$$|d(\mathbf{X}, \mathbf{x}_j)^\alpha - \|\mathbf{X}\|^\alpha| = O(1), \tag{10.7}$$

and the result follows.

\impliedby : Immediate (for any α), since the definition (2.4) can be written

$$\widehat{X}_\alpha = \sum_{i=1}^4 (-1)^i (\|\mathbf{X}_i\|^\alpha + \|\mathbf{X}_{i+1}\|^\alpha - d(\mathbf{X}_i, \mathbf{X}_{i+1})^\alpha). \tag{10.8}$$

□

Remark 10.5. We do not know (even for $\mathcal{X} = \mathbb{R}$) whether Lemma 10.4 holds also for $\alpha > 1$, and leave that as an open problem. (It holds, by a minor modification of the proof above, for $\alpha > 1$ under the additional assumption $\mathbb{E}\|\mathbf{X}\|^{p(\alpha-1)} < \infty$, but that seems less useful.)

We next introduce a class of function spaces.

10.2. *Lorentz spaces.* The condition (10.1) can be expressed as follows using *Lorentz spaces*, a generalization of the Lebesgue spaces L^p ; see e.g. Bennett and Sharpley (1988); Bergh and Löfström (1976). Let X^* be the *decreasing rearrangement* of $\|\mathbf{X}\|$; this is the (weakly) decreasing function $(0, 1) \rightarrow [0, \infty)$ defined by

$$X^*(t) := \inf\{x : \mathbb{P}(\|\mathbf{X}\| > x) \leq t\}. \quad (10.9)$$

In probabilistic terms, X^* is characterized as the decreasing function on $(0, 1)$ that, regarded as a random variable when $(0, 1)$ is equipped with the Lebesgue measure, has the same distribution as $\|\mathbf{X}\|$.

For a given probability space (Ω, \mathcal{F}, P) , and $p, q \in (0, \infty)$, the *Lorentz space* $L^{p,q}(\Omega, \mathcal{F}, P)$ is defined as the linear space of all real-valued random variables \mathbf{X} such that

$$\int_0^1 (t^{1/p} X^*(t))^q \frac{dt}{t} < \infty. \quad (10.10)$$

It is well-known that $L^{p,p} = L^p$, and that if $q_1 < q_2$ then $L^{p,q_1} \subset L^{p,q_2}$, with strict inequality provided the probability space is large enough.

A standard Fubini argument shows that

$$\begin{aligned} \int_0^1 (t^{1/p} X^*(t))^q \frac{dt}{t} &= q \int_0^1 \int_0^\infty \mathbf{1}\{X^*(t) > x\} x^{q-1} t^{q/p-1} dx dt \\ &= q \int_0^1 \int_0^\infty \mathbf{1}\{\mathbb{P}(\|\mathbf{X}\| > x) > t\} x^{q-1} t^{q/p-1} dx dt \\ &= p \int_0^\infty \mathbb{P}[\|\mathbf{X}\| > x]^{q/p} x^{q-1} dx. \end{aligned} \quad (10.11)$$

In particular, taking $p = \alpha$ and $q = 2\alpha$, we see that (10.1) is equivalent to $\|\mathbf{X}\| \in L^{\alpha, 2\alpha}$.

Consequently, for $\alpha \leq 1$, Theorem 10.1 says that if $\|\mathbf{X}\| \in L^{\alpha, 2\alpha}$, then $\widehat{X}_\alpha \in L^2$, which weakens the condition $\|\mathbf{X}\| \in L^\alpha$ in Lemma 4.3 to $L^{\alpha, 2\alpha}$. Hence, we can extend the use of Definition 2.2; moreover, as shown below, also Definitions 2.3, 2.4 and 7.1 yield the same result in this case.

Theorem 10.6. *Let $0 < \alpha \leq 1$, and assume that $\|\mathbf{X}\|, \|\mathbf{Y}\| \in L^{\alpha, 2\alpha}$. Then:*

- (i) *Definition 2.2 yields a finite value $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y})$.*
- (ii) *Definition 2.3 yields a finite value $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$, and $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y})$.*
- (iii) *If \mathcal{X} and \mathcal{Y} are Euclidean spaces, then Definition 2.4 yields a finite value $\text{dcov}_\alpha^E(\mathbf{X}, \mathbf{Y})$, and $\text{dcov}_\alpha^E(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y})$.*
- (iv) *If \mathcal{X} and \mathcal{Y} are Hilbert spaces, then Definition 7.1 yields a finite value $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$, and $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y})$.*

Thus, the values that are defined are all equal (and finite).

The proof is postponed to the following subsections.

Note also that by Remark 10.2, the Lorentz space $L^{\alpha, 2\alpha}$ is optimal in the strong sense that, for $\alpha \leq 1$ and $\mathcal{X} = \mathbb{R}$,

$$\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X}) < \infty \iff \widehat{X}_\alpha \in L^2 \iff \|\mathbf{X}\| \in L^{\alpha, 2\alpha}. \quad (10.12)$$

The proofs of the results above assume $\alpha \leq 1$. We leave the case $\alpha > 1$ as open problems. For example:

Problem 10.7. *For $\alpha > 1$, what is the optimal Lorentz space condition that guarantees $\mathbb{E}|\widehat{X}_\alpha|^2 < \infty$ and thus $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X}) < \infty$?*

By Theorem 9.11, the answer for $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{X})$ is the same.

Remark 10.8. Example 9.8 shows that in the special case $\alpha = 2$, the condition $\|\mathbf{X}\| \in L^2$ in Definition 2.2 cannot be improved; it is actually necessary for $\widehat{X}_\alpha \in L^2$ and $\text{dcov}_2(\mathbf{X}, \mathbf{X}) < \infty$ in the case $\mathcal{X} = \mathbb{R}$. Hence, for $\alpha = 2$, the answer to Problem 10.7 is $L^2 = L^{2,2}$.

A naive interpolation with (10.12) yields the conjecture that for $1 < \alpha < 2$, the answer is $L^{\alpha,2}$.

Remark 10.9. The equivalence (10.12) does not hold for all metric spaces \mathcal{X} , not even for $\alpha = 1$. For a counterexample, let $\mathcal{X} = \ell^1$ with the standard basis $(\mathbf{e}_n)_1^\infty$, let $0 < \gamma \leq 1/2$, and let N be an integer-valued random variable with $\mathbb{P}(N = n) = p_n := cn^{-1-\gamma}$, $n \geq 1$, where c is a normalization constant. Finally, let $\mathbf{X} := N^{1/2}\mathbf{e}_N$. It is easily seen that, with \mathbf{X}_i defined in the same way by N_i ,

$$\widehat{X} \leq 2 \sum_{i=1}^4 N_i^{1/2} \mathbf{1}\{N_i = N_{i+1}\}, \tag{10.13}$$

and thus, using Cauchy–Schwarz’s (or Minkowski’s) inequality,

$$\mathbb{E} \widehat{X}^2 \leq C \mathbb{E}[N_1 \mathbf{1}\{N_1 = N_2\}] = C \sum_{n=1}^\infty np_n^2 = C \sum_{n=1}^\infty n^{-1-2\gamma} < \infty, \tag{10.14}$$

while for $x \geq 1$,

$$\mathbb{P}(\|\mathbf{X}\| > x) = \mathbb{P}(N > x^2) = \sum_{n>x^2} cn^{-1-\gamma} \geq cx^{-2\gamma} \geq cx^{-1}, \tag{10.15}$$

so (10.1) fails, and thus $\|\mathbf{X}\| \notin L^{1,2}$.

Problem 10.10. *Does the equivalence (10.12) hold in Euclidean spaces? In infinite-dimensional Hilbert spaces?*

We have not investigated whether the results on continuity and consistency in Section 5 can be extended (for $\alpha \leq 1$) by replacing the moment conditions with the corresponding Lorentz space condition. In particular:

Problem 10.11. *Let $\alpha \leq 1$. Does Theorem 5.4 hold if the moment condition is replaced by $\mathbf{X}, \mathbf{Y} \in L^{\alpha,2\alpha}$?*

10.3. *More on $\text{dcov}_\alpha^\wedge$ and dcov_α^\sim .* Theorem 9.11 considers only the case $\mathbf{X} = \mathbf{Y}$. We do not know whether it extends to $\text{dcov}_\alpha(\mathbf{X}, \mathbf{Y})$ in general, without further conditions. We give a partial result.

Theorem 10.12. *Suppose that $\widehat{X}_\alpha, \widehat{Y}_\alpha \in L^1$, so \widetilde{X}_α and \widetilde{Y}_α exist. Suppose further that $\widetilde{X}_\alpha \widetilde{Y}_\alpha \in L^1$ and $\widetilde{X}_\alpha(\mathbf{X}_1, \mathbf{X}_2) \widetilde{Y}_\alpha(\mathbf{Y}_1, \mathbf{Y}_3) \in L^1$. Then $\widehat{X}_\alpha \widehat{Y}_\alpha \in L^1$, and $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$; furthermore, this value is finite.*

In particular, this holds if $\widehat{X}_\alpha, \widehat{Y}_\alpha \in L^2$.

Proof: This is similar to the proof of Theorem 9.11. We have $\widetilde{X}_\alpha, \widetilde{Y}_\alpha \in L^1$ by (2.6), and thus $\widetilde{X}_\alpha(\mathbf{X}_1, \mathbf{X}_2) \widetilde{Y}_\alpha(\mathbf{Y}_3, \mathbf{Y}_4) \in L^1$ by independence. Express \widehat{X}_α and \widehat{Y}_α by (9.10) and expand $\widehat{X}_\alpha \widehat{Y}_\alpha$ as a sum of 16 terms. By the assumptions (and symmetry), every term is in L^1 , so we may take their expectations separately. Furthermore, (9.9) implies that e.g. $\mathbb{E}[\widetilde{X}_\alpha(\mathbf{X}_1, \mathbf{X}_2) \widetilde{Y}_\alpha(\mathbf{Y}_1, \mathbf{Y}_3)] = 0$, and we obtain

$$\mathbb{E}[\widehat{X}_\alpha \widehat{Y}_\alpha] = \sum_{i=1}^4 \mathbb{E}[\widetilde{X}_\alpha(\mathbf{X}_i, \mathbf{X}_{i+1}) \widetilde{Y}_\alpha(\mathbf{Y}_i, \mathbf{Y}_{i+1})] = 4 \mathbb{E}[\widetilde{X}_\alpha \widetilde{Y}_\alpha]. \tag{10.16}$$

If $\widehat{X}_\alpha, \widehat{Y}_\alpha \in L^2$, then $\widetilde{X}_\alpha, \widetilde{Y}_\alpha \in L^2$ and the assumptions above follow by the Cauchy–Schwarz inequality. □

Proof of Theorem 10.6(i)(ii): By the comments before Theorem 10.6, the assumptions imply $\widehat{X}_\alpha, \widehat{Y}_\alpha \in L^2$, and thus Theorem 10.12 shows (i) and (ii). □

Problem 10.13. *Let either \mathcal{X} and \mathcal{Y} be arbitrary, or consider only $\mathcal{X} = \mathcal{Y} = \mathbb{R}$.*

- (i) *Is it true for arbitrary random $\mathbf{X} \in \mathcal{X}$ and $\mathbf{Y} \in \mathcal{Y}$ that $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y})$ is defined and finite \iff $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$ is defined and finite?*
- (ii) *If this holds, is furthermore always $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$?*

10.4. *More on dcov_α^E and dcov_α^H .* Consider now the case of Euclidean or, more generally, Hilbert spaces and Definitions 2.4 and 7.1. We complete the proof of Theorem 10.6; recall that this assumes $\alpha \leq 1$.

Proof of Theorem 10.6(iii)(iv): (iv): This follows by essentially the same proof as for Theorem 7.4. As noted in Remark 7.5, (7.21) holds without any moment condition. Moreover, as said in the proof of Theorem 7.4, Lemma 4.2 holds for $\widehat{X}_{\alpha;M}$ defined in (7.19) too, uniformly in M ; we now use Lemma 4.2(i), and denote the right-hand side by \widehat{X}^{**} . Hence, $|\widehat{X}_{\alpha;M}| \leq \widehat{X}^{**}$, and similarly $|\widehat{Y}_{\alpha;M}| \leq \widehat{Y}^{**}$.

As noted above, $X \in L^{\alpha,2\alpha}$ is equivalent to (10.1) and to (10.2). Consequently, $\widehat{X}^{**} \in L^2$ and, similarly, $\widehat{Y}^{**} \in L^2$. Hence, $\widehat{X}^{**}\widehat{Y}^{**} \in L^1$ and dominated convergence applies to (7.21), just as in the proof of Theorem 7.4, yielding $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y}) < \infty$.

(iii): Theorem 7.2 shows the general equality $\text{dcov}_\alpha^E(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$, and thus (iii) follows from (iv).

This completes the proof of Theorem 10.6. □

Problem 10.14. *For $1 < \alpha < 2$, what is the optimal Lorentz space condition that guarantees $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{X}) < \infty$ (for variables in a Hilbert space)? Does this also imply $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{X}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{X})$? Does this condition imply $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y})$ for two variables \mathbf{X} and \mathbf{Y} ?*

Problem 10.15. *Let either \mathcal{X} and \mathcal{Y} be arbitrary Hilbert spaces, or consider only $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. Let $0 < \alpha < 2$.*

- (i) *Is it true for arbitrary random $\mathbf{X} \in \mathcal{X}$ and $\mathbf{Y} \in \mathcal{Y}$ that $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y})$ is defined and finite \iff $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$ is finite?*
- (ii) *If this holds, is furthermore always $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y}) = \text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$?*

10.5. *Hilbert spaces, $\alpha = 2$.* Consider now the case when $\mathcal{X} = \mathcal{H}$ and $\mathcal{Y} = \mathcal{H}'$ are Hilbert spaces, as in the preceding subsection, but take $\alpha = 2$. Then $\text{dcov}_\alpha^H(\mathbf{X}, \mathbf{Y})$ is not defined, so we consider $\text{dcov}_2^\wedge(\mathbf{X}, \mathbf{Y})$ and $\text{dcov}_2^\sim(\mathbf{X}, \mathbf{Y})$. In Section 8, we did this assuming second moments; we now remove that assumption and generalise the results. (This is partly for its own sake, but mainly for the application in the next subsection.)

In this subsection, expectations $\mathbb{E}\mathbf{X}$ of Hilbert space valued random variables are always interpreted in Pettis sense, see Appendix B. (This is sometimes said explicitly for emphasis.) We use some technical results stated and proved in Appendix B.

Recall that $\widehat{X}_2 = 2\langle \mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_4 - \mathbf{X}_2 \rangle$ by (8.1), for any \mathbf{X} . We next show that (8.2) holds under weaker conditions than assumed in Section 8.

Lemma 10.16. *Let $\mathcal{X} = \mathcal{H}$ be a Hilbert space. If \widetilde{X}_2 exists, then $\mathbb{E}\mathbf{X}$ exists in Pettis sense, and*

$$\widetilde{X}_2 = -2\langle \mathbf{X}_1 - \mathbb{E}\mathbf{X}, \mathbf{X}_2 - \mathbb{E}\mathbf{X} \rangle. \tag{10.17}$$

Proof: By (8.1), we have $\widehat{X}_2 = 2\langle \mathbf{Z}, \mathbf{Z}' \rangle$ where $\mathbf{Z} := \mathbf{X}_1 - \mathbf{X}_3$ and $\mathbf{Z}' := \mathbf{X}_4 - \mathbf{X}_2$. Assume that \widetilde{X}_2 exists, which by our definition means that $\mathbb{E}|\widehat{X}_2| < \infty$. Thus $\mathbb{E}|\langle \mathbf{Z}, \mathbf{Z}' \rangle| < \infty$. Lemma B.1(ii) applies and shows that $\mathbf{Z} = \mathbf{X}_1 - \mathbf{X}_3$ is Pettis integrable. Hence, for every $\mathbf{x} \in \mathcal{H}$,

$$\langle \mathbf{X}_1, \mathbf{x} \rangle - \langle \mathbf{X}_3, \mathbf{x} \rangle = \langle \mathbf{X}_1 - \mathbf{X}_3, \mathbf{x} \rangle = \langle \mathbf{Z}, \mathbf{x} \rangle \in L^1. \tag{10.18}$$

Since $\langle \mathbf{X}_1, \mathbf{x} \rangle$ and $\langle \mathbf{X}_3, \mathbf{x} \rangle$ are independent random variables, this implies $\mathbb{E}|\langle \mathbf{X}_1, \mathbf{x} \rangle| < \infty$, for any $\mathbf{x} \in \mathcal{H}$, and thus $\mathbb{E} \mathbf{X}_1$ exists in Pettis sense.

Using (8.1), we may now integrate over first \mathbf{X}_4 and then \mathbf{X}_3 and obtain

$$\mathbb{E}(\widehat{X}_2 \mid \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) = 2\langle \mathbf{X}_1 - \mathbf{X}_3, \mathbb{E} \mathbf{X} - \mathbf{X}_2 \rangle, \tag{10.19}$$

$$\widetilde{X}_2 = \mathbb{E}(\widehat{X}_2 \mid \mathbf{X}_1, \mathbf{X}_2) = 2\langle \mathbf{X}_1 - \mathbb{E} \mathbf{X}, \mathbb{E} \mathbf{X} - \mathbf{X}_2 \rangle. \tag{10.20}$$

showing (10.17). □

Theorem 10.17. *Let $\mathcal{X} = \mathcal{H}$ and $\mathcal{Y} = \mathcal{H}'$ be Hilbert spaces.*

- (i) *If $\text{dcov}_2^{\widehat{}}(\mathbf{X}, \mathbf{Y})$ is defined, i.e., $\mathbb{E}[\widehat{X}_2 \widehat{Y}_2]$ is defined as an extended real number, then $\text{dcov}_2^{\widehat{}}(\mathbf{X}, \mathbf{Y}) \in [0, \infty]$.*
- (ii) *If $\text{dcov}_2^{\widehat{}}(\mathbf{X}, \mathbf{Y}) < \infty$, then*

$$\text{dcov}_2^{\widehat{}}(\mathbf{X}, \mathbf{Y}) = \|\mathbb{E}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes (\mathbf{Y}_1 - \mathbf{Y}_2)]\|_{\mathcal{H} \otimes \mathcal{H}'}^2, \tag{10.21}$$

where the expectation exists in Pettis sense.

- (iii) *If $\text{dcov}_2^{\widetilde{}}(\mathbf{X}, \mathbf{Y})$ is defined, i.e., \widetilde{X}_2 and \widetilde{Y}_2 are defined and $\mathbb{E}[\widetilde{X}_2 \widetilde{Y}_2]$ is defined as an extended real number, then $\text{dcov}_2^{\widetilde{}}(\mathbf{X}, \mathbf{Y}) \in [0, \infty]$.*
- (iv) *If furthermore $\text{dcov}_2^{\widetilde{}}(\mathbf{X}, \mathbf{Y}) < \infty$, then*

$$\text{dcov}_2^{\widetilde{}}(\mathbf{X}, \mathbf{Y}) = 4\|\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}] - \mathbb{E} \mathbf{X} \otimes \mathbb{E} \mathbf{Y}\|_{\mathcal{H} \otimes \mathcal{H}'}^2, \tag{10.22}$$

where the expectations exist in Pettis sense.

- (v) *If $\text{dcov}_2^{\widehat{}}(\mathbf{X}, \mathbf{Y})$ and $\text{dcov}_2^{\widetilde{}}(\mathbf{X}, \mathbf{Y})$ both are defined, as in (i) and (iii), and furthermore $\text{dcov}_2^{\widehat{}}(\mathbf{X}, \mathbf{Y})$ and $\text{dcov}_2^{\widetilde{}}(\mathbf{X}, \mathbf{Y})$ both are finite, then*

$$\text{dcov}_2^{\widehat{}}(\mathbf{X}, \mathbf{Y}) = \text{dcov}_2^{\widetilde{}}(\mathbf{X}, \mathbf{Y}). \tag{10.23}$$

Proof: (i),(ii): By (8.1) and (8.5),

$$\begin{aligned} \widehat{X}_2 \widehat{Y}_2 &= 4\langle \mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_4 - \mathbf{X}_2 \rangle \langle \mathbf{Y}_1 - \mathbf{Y}_3, \mathbf{Y}_4 - \mathbf{Y}_2 \rangle \\ &= 4\langle (\mathbf{X}_1 - \mathbf{X}_3) \otimes (\mathbf{Y}_1 - \mathbf{Y}_3), (\mathbf{X}_4 - \mathbf{X}_2) \otimes (\mathbf{Y}_4 - \mathbf{Y}_2) \rangle_{\mathcal{H} \otimes \mathcal{H}'}. \end{aligned} \tag{10.24}$$

This is an example of $\langle \mathbf{Z}, \mathbf{Z}' \rangle$ as in Lemma B.1, with $\mathbf{Z} := (\mathbf{X}_1 - \mathbf{X}_3) \otimes (\mathbf{Y}_1 - \mathbf{Y}_3) \stackrel{d}{=} (\mathbf{X}_1 - \mathbf{X}_2) \otimes (\mathbf{Y}_1 - \mathbf{Y}_2)$. Thus, (i) follows from Lemma B.1(iii), and (ii) from Lemma B.1(ii).

(iii),(iv): Similarly, if \widetilde{X}_2 and \widetilde{Y}_2 exist, then Lemma 10.16 shows that $\mathbb{E} \mathbf{X}$ and $\mathbb{E} \mathbf{Y}$ exist in Pettis sense, and furthermore, using (8.5),

$$\begin{aligned} \widetilde{X}_2 \widetilde{Y}_2 &= 4\langle \mathbf{X}_1 - \mathbb{E} \mathbf{X}, \mathbf{X}_2 - \mathbb{E} \mathbf{X} \rangle \langle \mathbf{Y}_1 - \mathbb{E} \mathbf{Y}, \mathbf{Y}_2 - \mathbb{E} \mathbf{Y} \rangle \\ &= 4\langle (\mathbf{X}_1 - \mathbb{E} \mathbf{X}) \otimes (\mathbf{Y}_1 - \mathbb{E} \mathbf{Y}), (\mathbf{X}_2 - \mathbb{E} \mathbf{X}) \otimes (\mathbf{Y}_2 - \mathbb{E} \mathbf{Y}) \rangle. \end{aligned} \tag{10.25}$$

This is another example of $\langle \mathbf{Z}, \mathbf{Z}' \rangle$ as in Lemma B.1, now with $\mathbf{Z} := (\mathbf{X}_1 - \mathbb{E} \mathbf{X}) \otimes (\mathbf{Y}_1 - \mathbb{E} \mathbf{Y})$. Thus, (iii) follows from Lemma B.1(iii).

Finally, assume $\text{dcov}_2^{\widetilde{}}(\mathbf{X}, \mathbf{Y}) < \infty$, i.e., $\widetilde{X}_2 \widetilde{Y}_2 \in L^1$. Then (10.25) and Lemma B.1(ii) show that $\mathbb{E} \mathbf{Z}$ exists in Pettis sense, and that

$$\text{dcov}_2^{\widetilde{}}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}[\widetilde{X}_2 \widetilde{Y}_2] = 4\|\mathbb{E} \mathbf{Z}\|^2 = 4\|\mathbb{E}[(\mathbf{X} - \mathbb{E} \mathbf{X}) \otimes (\mathbf{Y} - \mathbb{E} \mathbf{Y})]\|^2. \tag{10.26}$$

We have

$$\mathbf{X}_1 \otimes \mathbf{Y}_1 = \mathbf{Z} + (\mathbb{E} \mathbf{X}) \otimes (\mathbf{Y}_1 - \mathbb{E} \mathbf{Y}) + \mathbf{X}_1 \otimes \mathbb{E} \mathbf{Y}. \tag{10.27}$$

Furthermore, since $\mathbb{E} \mathbf{X}$ and $\mathbb{E} \mathbf{Y}$ are constant vectors, it is easy to see that $\mathbb{E}[\mathbf{X}_1 \otimes \mathbb{E} \mathbf{Y}] = \mathbb{E} \mathbf{X} \otimes \mathbb{E} \mathbf{Y}$ and $\mathbb{E}[(\mathbb{E} \mathbf{X}) \otimes (\mathbf{Y}_1 - \mathbb{E} \mathbf{Y})] = \mathbb{E} \mathbf{X} \otimes \mathbb{E}[\mathbf{Y}_1 - \mathbb{E} \mathbf{Y}] = 0$. (This also follows from the more general Lemma B.7.) Hence, (10.27) shows that $\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}]$ exists, and

$$\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}] = \mathbb{E}[\mathbf{X}_1 \otimes \mathbf{Y}_1] = \mathbb{E} \mathbf{Z} + \mathbb{E} \mathbf{X} \otimes \mathbb{E} \mathbf{Y}. \tag{10.28}$$

Thus (10.22) follows from (10.26).

(v): In this case, (i)–(iv) all hold. By (iv), the expectations $\mathbb{E}\mathbf{X}$, $\mathbb{E}\mathbf{Y}$ and $\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}]$ exist. Hence, $\mathbb{E}[\mathbf{X}_i \otimes \mathbf{Y}_i] = \mathbb{E}[\mathbf{X} \otimes \mathbf{Y}]$ exists for every i . Furthermore, if $i \neq j$ so \mathbf{X}_i and \mathbf{Y}_j are independent, $\mathbb{E}[\mathbf{X}_i \otimes \mathbf{Y}_j]$ exists by Lemma B.7 and equals $\mathbb{E}[\mathbf{X}_i] \otimes \mathbb{E}[\mathbf{Y}_j]$. Hence, $\mathbb{E}[\mathbf{X}_i \otimes \mathbf{Y}_j]$ exists for every i and j , and thus

$$\begin{aligned} \mathbb{E}[(\mathbf{X}_1 - \mathbf{X}_2) \otimes (\mathbf{Y}_1 - \mathbf{Y}_2)] &= \mathbb{E}[\mathbf{X}_1 \otimes \mathbf{Y}_1] + \mathbb{E}[\mathbf{X}_2 \otimes \mathbf{Y}_2] - \mathbb{E}[\mathbf{X}_1 \otimes \mathbf{Y}_2] - \mathbb{E}[\mathbf{X}_2 \otimes \mathbf{Y}_1] \\ &= 2(\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}] - \mathbb{E}[\mathbf{X}] \otimes \mathbb{E}[\mathbf{Y}]). \end{aligned} \tag{10.29}$$

Consequently, (10.23) follows from (10.21) and (10.22). □

10.6. *Metric spaces of negative type.* In this subsection we assume that \mathcal{X} and \mathcal{Y} are metric spaces such that $d_{\mathcal{X}}^{\alpha}$ and $d_{\mathcal{Y}}^{\alpha}$ both are of negative type, see Remark 2.7. We then can embed the spaces into Hilbert spaces as in Remark 8.4 and transfer the results in Section 10.5.

Theorem 10.18. *Let $\alpha > 0$ and let \mathcal{X} and \mathcal{Y} be metric spaces such that $d_{\mathcal{X}}^{\alpha}$ and $d_{\mathcal{Y}}^{\alpha}$ both are of negative type.*

- (i) *If $\text{dcov}_{\alpha}^{\widehat{}}(\mathbf{X}, \mathbf{Y})$ is defined, i.e., $\mathbb{E}[\widehat{X}_{\alpha}\widehat{Y}_{\alpha}]$ is defined as an extended real number, then $\text{dcov}_{\alpha}^{\widehat{}}(\mathbf{X}, \mathbf{Y}) \in [0, \infty]$.*
- (ii) *If $\text{dcov}_{\alpha}^{\sim}(\mathbf{X}, \mathbf{Y})$ is defined, i.e., \widetilde{X}_{α} and \widetilde{Y}_{α} are defined and $\mathbb{E}[\widetilde{X}_{\alpha}\widetilde{Y}_{\alpha}]$ is defined as an extended real number, then $\text{dcov}_{\alpha}^{\sim}(\mathbf{X}, \mathbf{Y}) \in [0, \infty]$.*
- (iii) *If $\text{dcov}_{\alpha}^{\widehat{}}(\mathbf{X}, \mathbf{Y})$ and $\text{dcov}_{\alpha}^{\sim}(\mathbf{X}, \mathbf{Y})$ both are defined, as in (i) and (ii), and furthermore both are finite, then*

$$\text{dcov}_{\alpha}^{\widehat{}}(\mathbf{X}, \mathbf{Y}) = \text{dcov}_{\alpha}^{\sim}(\mathbf{X}, \mathbf{Y}). \tag{10.30}$$

Proof: Immediate by Remark 8.4 and Theorem 10.17(i)(iii)(v). □

This gives a partial (but not complete) answer to Problem 10.13 for spaces with d^{α} of negative type; recall from Remark 2.7 that when $0 < \alpha \leq 2$, this includes Hilbert spaces, in particular \mathbb{R} .

Remark 10.19. If d is a metric of negative type, then so is d^{α} for every $\alpha \leq 1$. Hence, if \mathcal{X} and \mathcal{Y} are metric spaces of negative type, then Theorem 10.18 applies at least with $0 < \alpha \leq 1$.

10.7. *Negative values?* If \mathcal{X} and \mathcal{Y} are metric spaces such that d^{α} is of negative type, then Theorem 10.18 shows that $\text{dcov}_{\alpha}^{\widehat{}}(\mathbf{X}, \mathbf{Y})$ and $\text{dcov}_{\alpha}^{\sim}(\mathbf{X}, \mathbf{Y})$ may not be negative and finite, nor $-\infty$. Theorem 9.1 then shows the same for $\text{dcov}_{\alpha}^{*}(\mathbf{X}, \mathbf{Y})$. The same is also, trivially, true for $\text{dcov}_{\alpha}^{\text{E}}(\mathbf{X}, \mathbf{Y})$ and $\text{dcov}_{\alpha}^{\text{H}}(\mathbf{X}, \mathbf{Y})$ when they are applicable. More precisely, we have the possibilities shown in Table 10.1, by Theorems 4.5, 7.4, 9.1 and 10.18; Examples 9.5, 9.7, 9.9, 9.13, 9.16; (2.9) and (7.2).

	$[0, \infty)$	$+\infty$	$(-\infty, 0)$	$-\infty$	undefined
$\text{dcov}_{\alpha}^{*}(\mathbf{X}, \mathbf{Y})$	+	–	–	–	+
$\text{dcov}_{\alpha}^{\widehat{}}(\mathbf{X}, \mathbf{Y})$	+	+	–	–	+
$\text{dcov}_{\alpha}^{\sim}(\mathbf{X}, \mathbf{Y})$	+	+	–	–	+
$\text{dcov}_{\alpha}^{\text{E}}(\mathbf{X}, \mathbf{Y})$	+	+	–	–	–
$\text{dcov}_{\alpha}^{\text{H}}(\mathbf{X}, \mathbf{Y})$	+	+	–	–	–

TABLE 10.1. Possibilities when d^{α} is of negative type

Conversely, if \mathcal{X} or \mathcal{Y} is a metric space that is not of negative type then $\text{dcov}(\mathbf{X}, \mathbf{Y}) < 0$ is possible (as soon as both spaces have at least two points), see Lyons (2013, Proposition 3.15); by Theorem 4.5, this holds for any of $\text{dcov}_{\alpha}^{*}(\mathbf{X}, \mathbf{Y})$, $\text{dcov}_{\alpha}^{\widehat{}}(\mathbf{X}, \mathbf{Y})$, $\text{dcov}_{\alpha}^{\sim}(\mathbf{X}, \mathbf{Y})$. Theorem 9.1 still rules out $\pm\infty$ for $\text{dcov}_{\alpha}^{*}(\mathbf{X}, \mathbf{Y})$, and we find the possibilities shown in Table 10.2.

	$[0, \infty)$	$+\infty$	$(-\infty, 0)$	$-\infty$	undefined
$\text{dcov}_\alpha^*(\mathbf{X}, \mathbf{Y})$	+	-	+	-	+
$\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y})$	+	+	+	?	+
$\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$	+	+	+	?	+

TABLE 10.2. Possibilities when d^α is not of negative type

For $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y})$ and $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y})$, we do not know whether $-\infty$ is possible (in Case (dc2) in Section 9):

Problem 10.20. *Is $\text{dcov}_\alpha^\wedge(\mathbf{X}, \mathbf{Y}) = -\infty$ or $\text{dcov}_\alpha^\sim(\mathbf{X}, \mathbf{Y}) = -\infty$ possible?*

Acknowledgement

This work was inspired by a lecture by Thomas Mikosch at a mini-mini-workshop in Gothenburg in April, 2019. I thank Thomas Mikosch and Russ Lyons for helpful comments.

Appendix A. A uniform integrability lemma

We use above some well-known standard results on uniform integrability, see e.g. Gut (2013, §5.4 and §5.5). We use also the following simple result which perhaps is less well-known; since we have not found a good reference, we provide a proof for completeness.

In this section, all random variables are real-valued. We state the lemmas below for sequences of random variables (the case that we use), but they are valid (with the same proofs) for families $(X_\iota)_{\iota \in \mathcal{I}}$ and $(Y_\iota)_{\iota \in \mathcal{I}}$ with an arbitrary index set \mathcal{I} .

Lemma A.1. *Let $(X_n)_n$ and $(Y_n)_n$ be uniformly integrable sequences of random variables, and suppose that for each n , X_n and Y_n are independent. Then the sequence $(X_n Y_n)_n$ is also uniformly integrable.*

To prove this, we use another simple result that perhaps is less well-known than it deserves.

Lemma A.2. *Let $(X_n)_n$ be a sequence of random variables. Then $(X_n)_n$ is uniformly integrable if and only if for every $\varepsilon > 0$ there exists $K_\varepsilon < \infty$ and a sequence $(X_n^\varepsilon)_n$ of random variables such that for every n ,*

$$|X_n^\varepsilon| \leq K_\varepsilon \quad \text{a.s.}, \tag{A.1}$$

$$\mathbb{E}|X_n - X_n^\varepsilon| < \varepsilon. \tag{A.2}$$

Proof: This is a simple exercise, using your favourite definition of uniform integrability. (See e.g. Gut, 2013, Definition 5.4.1 and Theorem 5.4.1.) \square

Proof of Lemma A.1: The uniform integrability implies the existence of constants B and B' such that $\mathbb{E}|X_n| \leq B$ and $\mathbb{E}|Y_n| \leq B'$ for all n .

Let $0 < \varepsilon < 1$. Lemma A.2 shows that there exists $K_\varepsilon < \infty$ and random variables X_n^ε and Y_n^ε such that both (A.1)–(A.2) and the corresponding inequalities with Y hold. Then $|X_n^\varepsilon Y_n^\varepsilon| \leq K_\varepsilon^2$ a.s. Since X_n and Y_n are independent, we may also assume that the pairs (X_n, X_n^ε) and (Y_n, Y_n^ε) are independent, and then

$$\begin{aligned} \mathbb{E}|X_n Y_n - X_n^\varepsilon Y_n^\varepsilon| &\leq \mathbb{E}|X_n(Y_n - Y_n^\varepsilon)| + \mathbb{E}|(X_n - X_n^\varepsilon)Y_n| \\ &\quad + \mathbb{E}|(X_n - X_n^\varepsilon)(Y_n - Y_n^\varepsilon)| \\ &\leq B\varepsilon + B'\varepsilon + \varepsilon^2 = (B + B' + 1)\varepsilon. \end{aligned} \tag{A.3}$$

Lemma A.2 in the opposite direction shows that the sequence $(X_n Y_n)_n$ is uniformly integrable. \square

Appendix B. Bochner and Pettis integrals

The expectation $\mathbb{E}\mathbf{X}$ of an \mathcal{H} -valued random variable \mathbf{X} , where \mathcal{H} is a separable Hilbert space, can be defined using either the Bochner integral or the Pettis integral; see e.g. the summary in Janson and Kaijser (2015, §2.4) and the references given there. Both integrals are defined for general Banach spaces, but in this paper we need them only for separable Hilbert spaces. In this case, $\mathbb{E}\mathbf{X}$ exists in Bochner sense if and only if $\mathbb{E}\|\mathbf{X}\| < \infty$, and $\mathbb{E}\mathbf{X}$ exists in Pettis sense if and only if $\mathbb{E}|\langle \mathbf{X}, \mathbf{x} \rangle| < \infty$ for every $\mathbf{x} \in \mathcal{H}$, and then $\mathbb{E}\mathbf{X}$ is the element of \mathcal{H} determined by

$$\langle \mathbb{E}\mathbf{X}, \mathbf{x} \rangle = \mathbb{E}\langle \mathbf{X}, \mathbf{x} \rangle, \quad \mathbf{x} \in \mathcal{H}. \quad (\text{B.1})$$

If $\mathbb{E}\mathbf{X}$ exists in Bochner sense, then it exists in Pettis sense, and the value is the same. (Hence, the reader may choose to always interpret $\mathbb{E}\mathbf{X}$ in Pettis sense. However, the Bochner integral is more convenient when applicable.) The converse is not true; there are \mathbf{X} such that $\mathbb{E}\mathbf{X}$ exists in Pettis sense but not Bochner sense. (See e.g. Example B.3.)

It is well-known, and easy to see, that if $\mathbb{E}\mathbf{X}$ exists in Pettis sense, then there exists $C < \infty$ (depending on \mathbf{X}) such that

$$\mathbb{E}|\langle \mathbf{X}, \mathbf{x} \rangle| \leq C\|\mathbf{x}\|, \quad \mathbf{x} \in \mathcal{H}. \quad (\text{B.2})$$

We use in Section 10.5 some results on Pettis integrals in (separable) Hilbert spaces, stated in the lemmas below. We believe that at least some of these are known, but since we have not found references, we give complete proofs.

Lemma B.1. *Let \mathbf{Z} be random variable in a separable Hilbert space \mathcal{H} , and let \mathbf{Z}' be an independent copy of \mathbf{Z} .*

(i) *If \mathbf{Z} is Bochner integrable, i.e., if $\mathbb{E}\|\mathbf{Z}\| < \infty$, then*

$$\mathbb{E}|\langle \mathbf{Z}, \mathbf{Z}' \rangle| < \infty. \quad (\text{B.3})$$

(ii) *If (B.3) holds, then \mathbf{Z} is Pettis integrable, i.e., $\mathbb{E}\mathbf{Z}$ exists in Pettis sense. Moreover,*

$$\mathbb{E}\langle \mathbf{Z}, \mathbf{Z}' \rangle = \|\mathbb{E}\mathbf{Z}\|^2 \geq 0. \quad (\text{B.4})$$

(iii) *If $\mathbb{E}\langle \mathbf{Z}, \mathbf{Z}' \rangle_+ < \infty$, then (B.3) holds. In other words, $\mathbb{E}\langle \mathbf{Z}, \mathbf{Z}' \rangle$ may be finite (and then ≥ 0 by (B.4)), $+\infty$ or undefined, but never $-\infty$.*

Remark B.2. We show in Examples B.3 and B.4 that the implications in (i) and (ii) are strict, i.e., their converses do not hold.

Furthermore, it is easy find examples, even with $\mathcal{H} = \mathbb{R}$, where $\mathbb{E}\langle \mathbf{Z}, \mathbf{Z}' \rangle$ is $+\infty$ or undefined (i.e., $\infty - \infty$); take any real-valued random \mathbf{Z} with $\mathbf{Z} \geq 0$ or with a symmetric distribution, respectively, and further $\mathbb{E}|\mathbf{Z}| = \infty$.

Proof of Lemma B.1: (i): By the Cauchy–Schwarz inequality, $|\langle \mathbf{Z}, \mathbf{Z}' \rangle| \leq \|\mathbf{Z}\|\|\mathbf{Z}'\|$, and (B.3) follows by the independence of \mathbf{Z} and \mathbf{Z}' .

(ii): Let $A := \mathbb{E}|\langle \mathbf{Z}, \mathbf{Z}' \rangle|$ and let $\mathbf{u} \in \mathcal{H}$ with $\|\mathbf{u}\| = 1$. Furthermore, let $W := \text{sgn}\langle \mathbf{Z}, \mathbf{u} \rangle$ and $W' := \text{sgn}\langle \mathbf{Z}', \mathbf{u} \rangle$, and let for $M > 0$, $I_M := \mathbf{1}\{\|\mathbf{Z}\| \leq M\}$ and $I'_M := \mathbf{1}\{\|\mathbf{Z}'\| \leq M\}$. Since $I_M W \mathbf{Z} \stackrel{d}{=} I'_M W' \mathbf{Z}'$ is measurable and bounded, $\mathbb{E}[I_M W \mathbf{Z}] = \mathbb{E}[I'_M W' \mathbf{Z}']$ exists, even in Bochner sense, and we have, for any finite M ,

$$\begin{aligned} A &\geq \mathbb{E}[I_M W I'_M W' \langle \mathbf{Z}, \mathbf{Z}' \rangle] = \mathbb{E}\langle I_M W \mathbf{Z}, I'_M W' \mathbf{Z}' \rangle \\ &= \mathbb{E}[\mathbb{E}(\langle I_M W \mathbf{Z}, I'_M W' \mathbf{Z}' \rangle \mid \mathbf{Z})] = \mathbb{E}\langle I_M W \mathbf{Z}, \mathbb{E}[I'_M W' \mathbf{Z}'] \rangle \\ &= \langle \mathbb{E}[I_M W \mathbf{Z}], \mathbb{E}[I'_M W' \mathbf{Z}'] \rangle = \|\mathbb{E}[I_M W \mathbf{Z}]\|^2. \end{aligned} \quad (\text{B.5})$$

Hence, by the Cauchy–Schwarz inequality, $\|\mathbf{u}\| = 1$, and (B.5),

$$\begin{aligned} \mathbb{E}[I_M|\langle \mathbf{u}, \mathbf{Z} \rangle|] &= \mathbb{E}[I_M W \langle \mathbf{u}, \mathbf{Z} \rangle] = \mathbb{E}\langle \mathbf{u}, I_M W \mathbf{Z} \rangle = \langle \mathbf{u}, \mathbb{E}[I_M W \mathbf{Z}] \rangle \\ &\leq \|\mathbb{E}[I_M W \mathbf{Z}]\| \leq A^{1/2}. \end{aligned} \tag{B.6}$$

Letting $M \rightarrow \infty$ yields, by monotone convergence,

$$\mathbb{E}|\langle \mathbf{u}, \mathbf{Z} \rangle| \leq A^{1/2} \tag{B.7}$$

for every \mathbf{u} with $\|\mathbf{u}\| = 1$, which (since \mathcal{H} is reflexive) shows that \mathbf{Z} is Pettis integrable.

Finally, the Pettis integrability yields first

$$\mathbb{E}(\langle \mathbf{Z}, \mathbf{Z}' \rangle \mid \mathbf{Z}) = \langle \mathbf{Z}, \mathbb{E} \mathbf{Z}' \rangle \tag{B.8}$$

and then, taking the expectation of (B.8),

$$\mathbb{E} \langle \mathbf{Z}, \mathbf{Z}' \rangle = \mathbb{E} \langle \mathbf{Z}, \mathbb{E} \mathbf{Z}' \rangle = \langle \mathbb{E} \mathbf{Z}, \mathbb{E} \mathbf{Z}' \rangle = \langle \mathbb{E} \mathbf{Z}, \mathbb{E} \mathbf{Z} \rangle, \tag{B.9}$$

which is (B.4).

(iii): We have, similarly to (B.5),

$$\begin{aligned} \mathbb{E}[I_M I'_M \langle \mathbf{Z}, \mathbf{Z}' \rangle] &= \mathbb{E} \langle I_M \mathbf{Z}, I'_M \mathbf{Z}' \rangle = \mathbb{E}[\mathbb{E}(\langle I_M \mathbf{Z}, I'_M \mathbf{Z}' \rangle \mid \mathbf{Z})] \\ &= \mathbb{E} \langle I_M \mathbf{Z}, \mathbb{E}[I'_M \mathbf{Z}'] \rangle = \langle \mathbb{E}[I_M \mathbf{Z}], \mathbb{E}[I'_M \mathbf{Z}'] \rangle \\ &= \|\mathbb{E}[I_M \mathbf{Z}]\|^2 \geq 0. \end{aligned} \tag{B.10}$$

Hence,

$$\mathbb{E}[I_M I'_M \langle \mathbf{Z}, \mathbf{Z}' \rangle_-] \leq \mathbb{E}[I_M I'_M \langle \mathbf{Z}, \mathbf{Z}' \rangle_+] \leq \mathbb{E}[\langle \mathbf{Z}, \mathbf{Z}' \rangle_+] < \infty, \tag{B.11}$$

and letting $M \rightarrow \infty$ yields $\mathbb{E} \langle \mathbf{Z}, \mathbf{Z}' \rangle_- < \infty$ by monotone convergence. Hence, (B.3) holds, and the result follows. \square

We give counterexamples to converses of the statements in Lemma B.1.

Example B.3. Let N be a positive integer-valued random variable and let $p_n := \mathbb{P}(N = n)$, let $(a_n)_{n=1}^\infty$ be a sequence of positive numbers, and let $(\mathbf{e}_i)_i$ be an ON-basis in \mathcal{H} . Define $\mathbf{Z} := a_N \mathbf{e}_N$. Then

$$\mathbb{E} \|\mathbf{Z}\| = \mathbb{E} a_N = \sum_{n=1}^\infty a_n p_n. \tag{B.12}$$

If N' is an independent copy of N , and $\mathbf{Z}' := a_{N'} \mathbf{e}_{N'}$, then $\langle \mathbf{Z}, \mathbf{Z}' \rangle = a_N^2 \mathbf{1}\{N = N'\}$, and thus

$$\mathbb{E}|\langle \mathbf{Z}, \mathbf{Z}' \rangle| = \mathbb{E} \langle \mathbf{Z}, \mathbf{Z}' \rangle = \sum_{n=1}^\infty a_n^2 p_n^2. \tag{B.13}$$

Consequently, choosing $p_n = c/n^2$ and $a_n = n$, $\mathbb{E}|\langle \mathbf{Z}, \mathbf{Z}' \rangle| < \infty$ but $\mathbb{E} \|\mathbf{Z}\| = \infty$, so $\mathbb{E} \mathbf{Z}$ does not exist in Bochner sense. Hence the converse to Lemma B.1(i) does not hold.

In this example, as is easily seen, $\mathbb{E} \mathbf{Z}$ exists in Pettis sense if and only if $\sum_{n=1}^\infty a_n^2 p_n^2 < \infty$, and then $\mathbb{E} \mathbf{Z} = \sum_n a_n p_n \mathbf{e}_n$.

Example B.4. Let $(\mathbf{e}_i)_i$ be an ON-basis in \mathcal{H} , let $\xi_i \sim N(0, 1)$, $i \geq 1$, be independent, and let N be a positive integer-valued random variable, independent of $(\xi_i)_i$. Define $\mathbf{Z} := \sum_{i=1}^N \xi_i \mathbf{e}_i$. Then, for any $\mathbf{x} \in \mathcal{H}$,

$$\langle \mathbf{Z}, \mathbf{x} \rangle = \sum_{i=1}^N \langle \mathbf{e}_i, \mathbf{x} \rangle \xi_i. \tag{B.14}$$

Conditioned on N , this has a normal distribution with variance $\sum_1^N \langle \mathbf{e}_i, \mathbf{x} \rangle^2 \leq \|\mathbf{x}\|^2$. Hence,

$$\mathbb{E}(|\langle \mathbf{Z}, \mathbf{x} \rangle| \mid N) = \sqrt{\frac{2}{\pi}} \left(\sum_{i=1}^N \langle \mathbf{e}_i, \mathbf{x} \rangle^2 \right)^{1/2} \leq \|\mathbf{x}\| \tag{B.15}$$

and thus $\mathbb{E}|\langle \mathbf{Z}, \mathbf{x} \rangle| \leq \|\mathbf{x}\| < \infty$. Consequently, $\mathbb{E}\mathbf{Z}$ exists in Pettis sense. (With $\mathbb{E}\mathbf{Z} = 0$, by symmetry.)

On the other hand, if $N' \stackrel{d}{=} N$ and $\xi'_i \sim N(0, 1)$ are independent of each other and of N and $(\xi_i)_i$, so $\mathbf{Z}' := \sum_{i=1}^{N'} \xi'_i \mathbf{e}_i$ is an independent copy of \mathbf{Z} , then $\langle \mathbf{Z}, \mathbf{Z}' \rangle = \sum_1^{N \wedge N'} \xi_i \xi'_i$. The sequence $(\xi_i \xi'_i)_i$ is i.i.d. with mean 0 and variance $\mathbb{E}[(\xi_i \xi'_i)^2] = \mathbb{E}[\xi_i^2] \mathbb{E}[(\xi'_i)^2] = 1$, and thus by the central limit theorem, for some $c > 0$ and every $n \geq 0$,

$$\mathbb{E}(|\langle \mathbf{Z}, \mathbf{Z}' \rangle| \mid N \wedge N' = n) = \mathbb{E} \left| \sum_1^n \xi_i \xi'_i \right| \geq c\sqrt{n}. \tag{B.16}$$

Hence,

$$\begin{aligned} \mathbb{E}|\langle \mathbf{Z}, \mathbf{Z}' \rangle| &\geq c \mathbb{E} \sqrt{N \wedge N'} = c \int_0^\infty \mathbb{P}(\sqrt{N \wedge N'} > t) dt \\ &= c \int_0^\infty \mathbb{P}(N > t^2, N' > t^2) dt = c \int_0^\infty \mathbb{P}(N > t^2)^2 dt. \end{aligned} \tag{B.17}$$

Choose N with $\mathbb{P}(N > n) = n^{-\gamma}$ for $n \geq 1$, where $0 < \gamma \leq \frac{1}{4}$. Then $\mathbb{P}(N > t) \geq t^{-\gamma}$ for $t \geq 1$, and (B.17) yields $\mathbb{E}|\langle \mathbf{Z}, \mathbf{Z}' \rangle| \geq c \int_1^\infty t^{-4\gamma} dt = \infty$. Consequently, $\mathbb{E}\mathbf{Z}$ exists in Pettis sense, but (B.3) does not hold. Hence, the converse to Lemma B.1(ii) does not hold.

Note also that (B.16) and (B.17) hold in the opposite direction with another c ; hence, in this example, (B.3) holds if we take $\gamma > \frac{1}{4}$. Moreover, $\|\mathbf{Z}\| = (\sum_1^N \xi_i^2)^{1/2}$, and it follows from the law of large numbers that $\mathbb{E}(\|\mathbf{Z}\| \mid N = n) \sim \sqrt{n}$ as $n \rightarrow \infty$, and thus, if $\gamma \leq \frac{1}{2}$, we have $\mathbb{E}\|\mathbf{Z}\| \geq c \mathbb{E} N^{1/2} = \infty$. Consequently, taking $\gamma \in (\frac{1}{4}, \frac{1}{2}]$ gives another example showing that the converse to (i) does not hold.

Recall that a *Hilbert–Schmidt operator* $T : \mathcal{H} \rightarrow \mathcal{H}'$, where \mathcal{H} and \mathcal{H}' are Hilbert spaces, is a linear operator such that if $(\mathbf{e}_i)_i$ is an ON-basis in \mathcal{H} , then

$$\|T\|_{\text{HS}}^2 := \sum_i \|T\mathbf{e}_i\|^2 < \infty. \tag{B.18}$$

(This is independent of the choice of basis $(\mathbf{e}_i)_i$.) See e.g. Lax (2002, §30.8) or Conway (1990, Exercise IX.2.19). The following lemma is a version of the fact that a Hilbert–Schmidt operator is absolutely 1-summing Pietsch (1965, Theorem 2.5.5).

Lemma B.5. *Let \mathcal{H} and \mathcal{H}' be separable Hilbert spaces, let \mathbf{X} be random variable in \mathcal{H} such that $\mathbb{E}\mathbf{X}$ exists in Pettis sense, and let $T : \mathcal{H} \rightarrow \mathcal{H}'$ be a Hilbert–Schmidt operator. Then $\mathbb{E}\|T\mathbf{X}\| < \infty$.*

Proof: Since T is a Hilbert–Schmidt operator, T^*T is a positive self-adjoint trace class operator in \mathcal{H} , and thus there exists an ON-basis $(\mathbf{e}_i)_i$ in \mathcal{H} consisting of eigenvectors, so $T^*T\mathbf{e}_i = \lambda_i \mathbf{e}_i$, where $\lambda_i \geq 0$ and

$$\sum_i \lambda_i = \|T\|_{\text{HS}}^2 < \infty. \tag{B.19}$$

(See again e.g. [Lax, 2002](#), §30 and [Conway, 1990](#), Exercise IX.2.19.) Let $s_i := \lambda^{1/2}$. (These are known as the *singular values* of T .) Then, for any $\mathbf{x} \in \mathcal{H}$,

$$\begin{aligned} \|T\mathbf{x}\|^2 &= \langle T^*T\mathbf{x}, \mathbf{x} \rangle = \sum_i \langle T^*T\mathbf{x}, \mathbf{e}_i \rangle \langle \mathbf{x}, \mathbf{e}_i \rangle = \sum_i \langle \mathbf{x}, T^*T\mathbf{e}_i \rangle \langle \mathbf{x}, \mathbf{e}_i \rangle \\ &= \sum_i \lambda_i \langle \mathbf{x}, \mathbf{e}_i \rangle \langle \mathbf{x}, \mathbf{e}_i \rangle = \sum_i s_i^2 \langle \mathbf{x}, \mathbf{e}_i \rangle^2. \end{aligned} \tag{B.20}$$

Let $(\varepsilon_i)_i$ be i.i.d. random variables with $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = \frac{1}{2}$, and let them also be independent of \mathbf{X} . Let

$$\mathbf{Z} := \sum_i s_i \varepsilon_i \mathbf{e}_i, \tag{B.21}$$

where the sum converges in \mathcal{H} (surely) since $\sum_i s_i^2 < \infty$ by [\(B.19\)](#). Let $\mathbf{x} \in \mathcal{H}$ and note that

$$\langle \mathbf{x}, \mathbf{Z} \rangle = \sum_i s_i \langle \mathbf{x}, \mathbf{e}_i \rangle \varepsilon_i. \tag{B.22}$$

Hence, using [\(B.20\)](#),

$$\mathbb{E}|\langle \mathbf{x}, \mathbf{Z} \rangle|^2 = \mathbb{E} \left| \sum_i s_i \langle \mathbf{x}, \mathbf{e}_i \rangle \varepsilon_i \right|^2 = \sum_i s_i^2 \langle \mathbf{x}, \mathbf{e}_i \rangle^2 = \|T\mathbf{x}\|^2. \tag{B.23}$$

Moreover, Khintchine’s inequality [Gut \(2013, Lemma 3.8.1\)](#) applies to [\(B.22\)](#) and yields

$$(\mathbb{E}|\langle \mathbf{x}, \mathbf{Z} \rangle|^2)^{1/2} \leq C \mathbb{E}|\langle \mathbf{x}, \mathbf{Z} \rangle|. \tag{B.24}$$

Combining [\(B.23\)](#) and [\(B.24\)](#) we find

$$\|T\mathbf{x}\| \leq C \mathbb{E}|\langle \mathbf{x}, \mathbf{Z} \rangle|. \tag{B.25}$$

Let $\mathbb{E}_{\mathbf{X}}$ and \mathbb{E}_ε denote integration over \mathbf{X} and (ε_i) , respectively. Then [\(B.25\)](#) yields $\|T\mathbf{X}\| \leq C \mathbb{E}_\varepsilon |\langle \mathbf{X}, \mathbf{Z} \rangle|$ and thus

$$\mathbb{E}\|T\mathbf{X}\| \leq C \mathbb{E}_{\mathbf{X}} \mathbb{E}_\varepsilon |\langle \mathbf{X}, \mathbf{Z} \rangle| = C \mathbb{E}|\langle \mathbf{X}, \mathbf{Z} \rangle|. \tag{B.26}$$

On the other hand, [\(B.2\)](#) yields, using also the definition [\(B.21\)](#) and [\(B.19\)](#),

$$\mathbb{E}_{\mathbf{X}} |\langle \mathbf{X}, \mathbf{Z} \rangle| \leq C \|\mathbf{Z}\| = C \left(\sum_i s_i^2 \right)^{1/2} = C \|T\|_{\text{HS}}. \tag{B.27}$$

Thus,

$$\mathbb{E}|\langle \mathbf{X}, \mathbf{Z} \rangle| = \mathbb{E} \mathbb{E}_{\mathbf{X}} |\langle \mathbf{X}, \mathbf{Z} \rangle| \leq C \|T\|_{\text{HS}} < \infty. \tag{B.28}$$

The result follows by [\(B.26\)](#) and [\(B.28\)](#). □

Remark B.6. Example [B.3](#) shows that the result in Lemma [B.5](#) does not hold for $T = I$, the identity operator (if $\dim \mathcal{H} = \infty$). In fact, the result holds if and only if T is Hilbert–Schmidt: if T is a bounded operator that is not Hilbert–Schmidt, then there exists \mathbf{X} such that $\mathbb{E}\mathbf{X}$ exists but $\mathbb{E}\|T\mathbf{X}\| = \infty$; this can be seen by a modification of Example [B.3](#). (We omit the details.)

Lemma B.7. *Let \mathbf{X} and \mathbf{Y} be independent random variables with values in separable Hilbert spaces \mathcal{H} and \mathcal{H}' . If $\mathbb{E}\mathbf{X}$ and $\mathbb{E}\mathbf{Y}$ exist in Pettis sense, then $\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}]$ exists in Pettis sense, in $\mathcal{H} \otimes \mathcal{H}'$, and $\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}] = (\mathbb{E}\mathbf{X}) \otimes (\mathbb{E}\mathbf{Y})$.*

Proof: Let $\mathbf{z} \in \mathcal{H} \otimes \mathcal{H}'$, and define a linear operator $T_{\mathbf{z}} : \mathcal{H} \rightarrow \mathcal{H}'$ by

$$\langle T_{\mathbf{z}}\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x} \otimes \mathbf{y}, \mathbf{z} \rangle. \tag{B.29}$$

Let $(\mathbf{e}_i)_i$ and $(\mathbf{e}'_j)_j$ be ON-bases in \mathcal{H} and \mathcal{H}' . Then $(\mathbf{e}_i \otimes \mathbf{e}'_j)_{i,j}$ is an ON-basis in $\mathcal{H} \otimes \mathcal{H}'$, and thus, using (B.18) and (B.29),

$$\begin{aligned} \|T_{\mathbf{z}}\|_{\text{HS}}^2 &= \sum_i \|T_{\mathbf{z}}\mathbf{e}_i\|^2 = \sum_i \sum_j \langle T_{\mathbf{z}}\mathbf{e}_i, \mathbf{e}'_j \rangle^2 = \sum_i \sum_j \langle \mathbf{e}_i \otimes \mathbf{e}'_j, \mathbf{z} \rangle^2 \\ &= \|\mathbf{z}\|^2 < \infty, \end{aligned} \tag{B.30}$$

and thus $T_{\mathbf{z}}$ is a Hilbert–Schmidt operator. (In fact, as is well-known, it is easy to see that $\mathbf{z} \mapsto T_{\mathbf{z}}$ yields an isometry between $\mathcal{H} \otimes \mathcal{H}'$ and the space of Hilbert–Schmidt operators $\mathcal{H} \rightarrow \mathcal{H}'$.) Hence, Lemma B.5 applies and shows $\mathbb{E}\|T_{\mathbf{z}}\mathbf{X}\| < \infty$.

Furthermore, since \mathbf{Y} is Pettis integrable, (B.29) and (B.2) show that for every $\mathbf{x} \in \mathcal{H}$,

$$\mathbb{E}|\langle \mathbf{x} \otimes \mathbf{Y}, \mathbf{z} \rangle| = \mathbb{E}|\langle T_{\mathbf{z}}\mathbf{x}, \mathbf{Y} \rangle| \leq C\|T_{\mathbf{z}}\mathbf{x}\|. \tag{B.31}$$

Consequently, with $\mathbb{E}_{\mathbf{Y}}$ denoting the integral over \mathbf{Y} ,

$$\mathbb{E}|\langle \mathbf{X} \otimes \mathbf{Y}, \mathbf{z} \rangle| = \mathbb{E}\mathbb{E}_{\mathbf{Y}}|\langle \mathbf{X} \otimes \mathbf{Y}, \mathbf{z} \rangle| \leq C\mathbb{E}\|T_{\mathbf{z}}\mathbf{X}\| < \infty. \tag{B.32}$$

Since $\mathbf{z} \in \mathcal{H} \otimes \mathcal{H}'$ is arbitrary, this shows that $\mathbf{X} \otimes \mathbf{Y}$ is Pettis integrable, i.e., that $\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}]$ exists in Pettis sense.

Finally, by (B.1), (8.5) and independence, for any \mathbf{e}_i and \mathbf{e}'_j in the bases,

$$\begin{aligned} \langle \mathbb{E}[\mathbf{X} \otimes \mathbf{Y}], \mathbf{e}_i \otimes \mathbf{e}'_j \rangle &= \mathbb{E}\langle \mathbf{X} \otimes \mathbf{Y}, \mathbf{e}_i \otimes \mathbf{e}'_j \rangle = \mathbb{E}[\langle \mathbf{X}, \mathbf{e}_i \rangle \langle \mathbf{Y}, \mathbf{e}'_j \rangle] \\ &= \mathbb{E}[\langle \mathbf{X}, \mathbf{e}_i \rangle] \mathbb{E}[\langle \mathbf{Y}, \mathbf{e}'_j \rangle] = \langle \mathbb{E}\mathbf{X}, \mathbf{e}_i \rangle \langle \mathbb{E}\mathbf{Y}, \mathbf{e}'_j \rangle \\ &= \langle (\mathbb{E}\mathbf{X}) \otimes (\mathbb{E}\mathbf{Y}), \mathbf{e}_i \otimes \mathbf{e}'_j \rangle. \end{aligned} \tag{B.33}$$

Since the set of such $\mathbf{e}_i \otimes \mathbf{e}'_j$ is a basis, $\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}] = (\mathbb{E}\mathbf{X}) \otimes (\mathbb{E}\mathbf{Y})$ follows. \square

Remark B.8. In this paper we consider only the Hilbert space tensor product defined in Section 8. Nevertheless, we note that Lemma B.7 *a fortiori* holds also for the injective tensor product $\mathcal{H} \check{\otimes} \mathcal{H}'$, since there is a natural continuous mapping $\mathcal{H} \otimes \mathcal{H}' \rightarrow \mathcal{H} \check{\otimes} \mathcal{H}'$ mapping $\mathbf{x} \otimes \mathbf{y} \mapsto \mathbf{x} \otimes \mathbf{y}$. On the other hand, the result does not hold for the projective tensor product $\mathcal{H} \widehat{\otimes} \mathcal{H}'$, which can be seen as follows: Let $\mathcal{H} = \mathcal{H}'$ and note that then $\mathbf{x} \otimes \mathbf{y} \mapsto \langle \mathbf{x}, \mathbf{y} \rangle$ extends to a continuous linear functional on $\mathcal{H} \widehat{\otimes} \mathcal{H}'$. Hence, if $\mathbb{E}[\mathbf{X} \otimes \mathbf{Y}]$ exists in $\mathcal{H} \widehat{\otimes} \mathcal{H}'$, then $\mathbb{E}\langle \mathbf{X}, \mathbf{Y} \rangle$ exists in \mathbb{R} , so $\mathbb{E}|\langle \mathbf{X}, \mathbf{Y} \rangle| < \infty$, but Example B.4 shows that this does not always hold for independent Pettis integrable \mathbf{X} and \mathbf{Y} .

References

- Baker, C. R. Joint measures and cross-covariance operators. *Trans. Amer. Math. Soc.*, **186**, 273–289 (1973). [MR336795](#).
- Bennett, C. and Sharpley, R. *Interpolation of operators*, volume 129 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA (1988). ISBN 0-12-088730-4. [MR928802](#).
- Berg, C., Christensen, J. P. R., and Ressel, P. *Harmonic analysis on semigroups. Theory of positive definite and related functions*, volume 100 of *Graduate Texts in Mathematics*. Springer-Verlag, New York (1984). ISBN 0-387-90925-7. [MR747302](#).
- Bergh, J. and Löfström, J. *Interpolation spaces. An introduction*. Grundlehren der Mathematischen Wissenschaften, No. 223. Springer-Verlag, Berlin-New York (1976). [MR0482275](#).
- Billingsley, P. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney (1968). [MR0233396](#).
- Bogachev, V. I. and Kolesnikov, A. V. The Monge–Kantorovich problem: achievements, connections, and prospects. *Uspekhi Mat. Nauk*, **67** (5(407)), 3–110 (2012). [MR3058744](#). English translation: *Russian Math. Surveys* 67 (2012), no. 5, 785–890.
- Conway, J. B. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition (1990). ISBN 0-387-97245-5. [MR1070713](#).

- Dehling, H., Matsui, M., Mikosch, T., Samorodnitsky, G., and Tafakori, L. Distance covariance for discretized stochastic processes. *Bernoulli*, **26** (4), 2758–2789 (2020). [MR4140528](#).
- Feuerverger, A. A consistent test for bivariate dependence. *Int. Stat. Rev.*, **61** (2), 419–433 (1993). [MR4205649](#).
- Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic learning theory*, volume 3734 of *Lecture Notes in Comput. Sci.*, pp. 63–77. Springer, Berlin (2005). [MR2255909](#).
- Gut, A. *Probability: a graduate course*. Springer Texts in Statistics. Springer, New York, second edition (2013). ISBN 978-1-4614-4707-8; 978-1-4614-4708-5. [MR2977961](#).
- Jakobsen, M. E. Distance covariance in metric spaces: non-parametric independence testing in metric spaces (2017). Master's thesis. [arXiv: 1706.03490v1](#).
- Janson, S. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge (1997). ISBN 0-521-56128-0. [MR1474726](#).
- Janson, S. and Kaijser, S. Higher moments of Banach space valued random variables. *Mem. Amer. Math. Soc.*, **238** (1127), vii+110 (2015). [MR3402381](#).
- Kallenberg, O. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition (2002). ISBN 0-387-95313-2. [MR1876169](#).
- Kanagawa, M., Hennig, P., Sejdinovic, D., and Sriperumbudur, B. K. Gaussian processes and kernel methods: a review on connections and equivalences. *ArXiv Mathematics e-prints* (2018). [arXiv: 1807.02582v1](#).
- Lax, P. D. *Functional analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York (2002). ISBN 0-471-55604-1. [MR1892228](#).
- Lyons, R. Distance covariance in metric spaces. *Ann. Probab.*, **41** (5), 3284–3305 (2013). [MR3127883](#). Errata at *Ann. Probab.* **46** (2018), no. 4, 2400–2405. [MR3813995](#).
- Olver, F. W. J., Lozier, D. W., Boisvert, R. F., and Clark, C. W., editors. *NIST handbook of mathematical functions*. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge University Press, Cambridge (2010). ISBN 978-0-521-14063-8. [MR2723248](#). URL: <http://dlmf.nist.gov/>.
- Pietsch, A. *Nukleare lokalkonvexe Räume*. Schriftenreihe Inst. Math. Deutsch. Akad. Wiss. Berlin, Reihe A, Reine Mathematik, Heft 1. Akademie-Verlag, Berlin, second edition (1965). [MR0181888](#). English translation in *Nuclear locally convex spaces* (1972). [MR0350360](#).
- Rüschendorf, L. Wasserstein metric (2020). Encyclopedia of Mathematics. Available at https://www.encyclopediaofmath.org/index.php?title=Wasserstein_metric.
- Schoenberg, I. J. Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, **44** (3), 522–536 (1938). [MR1501980](#).
- Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.*, **41** (5), 2263–2291 (2013). [MR3127866](#).
- Székely, G. J. and Rizzo, M. L. Brownian distance covariance. *Ann. Appl. Stat.*, **3** (4), 1236–1265 (2009a). [MR2752127](#).
- Székely, G. J. and Rizzo, M. L. Rejoinder: Brownian distance covariance. *Ann. Appl. Stat.*, **3** (4), 1303–1308 (2009b). [MR2752135](#).
- Székely, G. J., Rizzo, M. L., and Bakirov, N. K. Measuring and testing dependence by correlation of distances. *Ann. Statist.*, **35** (6), 2769–2794 (2007). [MR2382665](#).
- Varadarajan, V. S. On the convergence of sample probability distributions. *Sankhyā*, **19**, 23–26 (1958). [MR94839](#).