Characteristic and Universal Tensor Product Kernels

Zoltán Szabó, Bharath K Sriperumbudur

Characteristic and Universal Tensor Product Kernels

Zoltán Szabó* Bharath K. Sriperumbudur†

Abstract

Kernel mean embeddings provide a versatile and powerful nonparametric representation of probability distributions with several fundamental applications in machine learning. Key to the success of the technique is whether the embedding is injective. This characteristic property of the underlying kernel ensures that probability distributions can be discriminated via their representations. In this paper, we consider kernels of tensor product type and various notions of characteristic property (including the one that captures joint independence of random variables) and provide a complete characterization for the corresponding embedding to be injective. This has applications, for example in independence measures such as Hilbert-Schmidt independence criterion (HSIC) to characterize the joint independence of multiple random variables.

Keywords: tensor product kernel, kernel mean embedding, characteristic kernel, $\mathcal{I}$-characteristic kernel, universality, maximum mean discrepancy, Hilbert-Schmidt independence criterion

1 Introduction

Kernel methods [Schölkopf and Smola, 2002] are among the most flexible and influential tools in machine learning, with superior performance demonstrated in a large number of areas and applications. The key idea in these methods is to map the data samples into a possibly infinite-dimensional feature space—precisely, a reproducing kernel Hilbert space (RKHS) [Aronszajn, 1950]—and apply linear methods in the feature space, without the explicit need to compute the map. A generalization of this idea to probability measures, i.e., mapping probability measures into an RKHS (Berlinet and Thomas-Agnan, 2004, Chapter 4; Smola et al., 2007) has found novel applications in nonparametric statistics and machine learning. Formally, given a probability measure $\mathbb{P}$ defined on a measurable space $\mathcal{X}$ and an RKHS $\mathcal{H}_k$ with k as the reproducing kernel (which is symmetric and positive definite), $\mathbb{P}$ is embedded into $\mathcal{H}_k$ as

$$\mathbb{P} \mapsto \int_{\mathcal{X}} k(\cdot, x)\, \mathrm{d}\mathbb{P}(x) =: \mu_k(\mathbb{P}), \tag{1}$$

where $\mu_k(\mathbb{P})$ is called the *mean element* or *kernel mean embedding* of $\mathbb{P}$. The *mean embedding* of $\mathbb{P}$ has lead to a new generation of solutions in two-sample testing [Gretton et al.,

*Center for Applied Mathematics (CMAP), École Polytechnique, Route de Saclay, 91128 Palaiseau, France. E-mail: zoltan.szabo@polytechnique.edu

†Department of Statistics, Pennsylvania State University, 314 Thomas Building, University Park, PA 16802. E-mail: bks18@psu.edu

2012], domain adaptation [Zhang et al., 2013], kernel belief propagation [Song et al., 2011], kernel Bayes' rule [Fukumizu et al., 2013], model criticism [Lloyd et al., 2014], approximate Bayesian computation [Park et al., 2016], probabilistic programming [Schölkopf et al., 2015], distribution classification [Muandet et al., 2011], distribution regression [Szabó et al., 2016] and topological data analysis [Kusano et al., 2016]. For a recent survey on the topic, the reader is referred to [Muandet et al., 2017].

Crucial to the success of the mean embedding based representation is whether it encodes all the information about the distribution, in other words whether the map in (1) is injective in which case the kernel is referred to as *characteristic* [Fukumizu et al., 2008, Sriperumbudur et al., 2010]. Various characterizations for the characteristic property of k is known in the literature (e.g., see Fukumizu et al., 2008, 2009, Gretton et al., 2012, Sriperumbudur et al., 2010) using which the popular kernels on $\mathbb{R}^d$ such as Gaussian, Laplacian, B-spline, inverse multiquadrics, and the Matérn class are shown to be characteristic. The characteristic property is closely related to the notion of *universality* (Steinwart, 2001; Micchelli et al., 2006; Carmeli et al., 2010; Sriperumbudur et al., 2011)—k is said to be universal if the corresponding RKHS $\mathcal{H}_k$ is dense in a certain target function class, for example, the class of continuous functions on compact domains—and the relation between these notions has recently been explored in [Sriperumbudur et al., 2011, Simon-Gabriel and Schölkopf, 2016].

Based on the mean embedding in (1), Smola et al. [2007] and Gretton et al. [2012] defined a semi-metric, called the maximum mean discrepancy (MMD) on the space of probability measures:

$$\mathrm{MMD}_k(\mathbb{P}, \mathbb{Q}) := \|\mu_k(\mathbb{P}) - \mu_k(\mathbb{Q})\|_{\mathcal{H}_k},$$

which is a metric if k is characteristic. A fundamental application of MMD is non-parametric hypothesis testing that includes two-sample [Gretton et al., 2012] and independence tests [Gretton et al., 2008]. Particularly in independence testing, as a measure of independence, MMD measures the distance between the joint distribution $\mathbb{P}_{XY}$ and the product of marginals $\mathbb{P}_X \otimes \mathbb{P}_Y$ of two random variables X and Y which are respectively defined on measurable spaces $\mathcal{X}$ and $\mathcal{Y}$, with the kernel k being defined on $\mathcal{X} \times \mathcal{Y}$. As aforementioned, if k is characteristic, then $\mathrm{MMD}_k(\mathbb{P}_{XY}, \mathbb{P}_X \otimes \mathbb{P}_Y) = 0$ implies $\mathbb{P}_{XY} = \mathbb{P}_X \otimes \mathbb{P}_Y$, i.e., X and Y are independent. A simple way to define a kernel on $\mathcal{X} \times \mathcal{Y}$ is through the tensor product of kernels k_X and k_Y defined on $\mathcal{X}$ and $\mathcal{Y}$ respectively: $k = k_X \otimes k_Y$, i.e., $k\left((x, y), (x', y')\right) = k_X(x, x')k_Y(y, y')$, $x, x' \in \mathcal{X}$, $y, y' \in \mathcal{Y}$, with the corresponding RKHS $\mathcal{H}_k = \mathcal{H}_{k_X} \otimes \mathcal{H}_{k_Y}$ being the tensor product space generated by $\mathcal{H}_{k_X}$ and $\mathcal{H}_{k_Y}$. This means, when $k = k_X \otimes k_Y$,

$$\mathrm{MMD}_k(\mathbb{P}_{XY}, \mathbb{P}_X \otimes \mathbb{P}_Y) = \|\mu_{k_X \otimes k_Y}(\mathbb{P}_{XY}) - \mu_{k_X \otimes k_Y}(\mathbb{P}_X \otimes \mathbb{P}_Y)\|_{\mathcal{H}_{k_X} \otimes \mathcal{H}_{k_Y}}. \tag{2}$$

In addition to the simplicity of defining a joint kernel k on $\mathcal{X} \times \mathcal{Y}$, the tensor product kernel offers a principled way of combining inner products (k_X and k_Y) on domains that correspond to different modalities (say images, texts, audio). By exploiting the isomorphism between tensor product Hilbert spaces and the space of Hilbert-Schmidt operators, it follows from (2) that

$$\mathrm{MMD}_k(\mathbb{P}_{XY}, \mathbb{P}_X \otimes \mathbb{P}_Y) = \|C_{XY}\|_{\mathrm{HS}} =: \mathrm{HSIC}_k(\mathbb{P}_{XY}),$$

which is the Hilbert-Schmidt norm of the cross-covariance operator $C_{XY} := \mu_{k_X \otimes k_Y}(\mathbb{P}_{XY}) - \mu_{k_X}(\mathbb{P}_X) \otimes \mu_{k_Y}(\mathbb{P}_Y)$ and is known as the *Hilbert-Schmidt independence criterion* (HSIC) [Gretton et al., 2005]. HSIC has enjoyed tremendous success in a variety of applications such as

blind source separation [Gretton et al., 2005], feature selection [Song et al., 2012], independence testing [Gretton et al., 2008], post selection inference [Yamada et al., 2016] and causal detection [Mooij et al., 2016, Pfister et al., 2017]. Recently, Pfister et al. [2017] generalized HSIC to test for the joint independence of M random variables as:

$$\mathrm{HSIC}_k\left(\mathbb{P}\right) = \left\| \mu_{\otimes_{m=1}^M k_m}(\mathbb{P}) - \otimes_{m=1}^M \mu_{k_m}\left(\mathbb{P}_m\right) \right\|_{\otimes_{m=1}^M \mathcal{H}_{k_m}},$$

where $\mathbb{P}$ is a joint measure on the product space $\mathcal{X} := \times_{m=1}^M \mathcal{X}_m$ and $(\mathbb{P}_m)_{m=1}^M$ are the marginal measures of $\mathbb{P}$ defined on $(\mathcal{X}_m)_{m=1}^M$ respectively. The key requirement in these applications of HSIC is that $k = \otimes_{m=1}^M k_m$ captures the joint independence of M random variables (with joint distribution $\mathbb{P}$)—we call this property as $\mathcal{I}$-characteristic—, which is guaranteed if k is characteristic. Since k is defined in terms of $(k_m)_{m=1}^M$, it is of fundamental importance to understand the characteristic and $\mathcal{I}$-characteristic properties of k in terms of the characteristic property of $(k_m)_{m=1}^M$, which is the goal of this work.

For $M = 2$, the characterization of independence, i.e., the $\mathcal{I}$-characteristic property of k, is studied by Waegeman et al. [2012] and Gretton [2015] where it has been shown that if k_1 and k_2 are universal, then k is universal[1] and therefore HSIC captures independence. A stronger version of this result can be obtained by combining [Lyons, 2013, Theorem 3.11] and [Sejdinovic et al., 2013, Proposition 29]: if k_1 and k_2 are characteristic, then the HSIC associated with $k = k_1 \otimes k_2$ characterizes independence. Apart from these results, not much is known about the characteristic/$\mathcal{I}$-characteristic/universality properties of k in terms of the individual kernels. In Section 3, we conduct a comprehensive analysis about these properties of k and $(k_m)_{m=1}^M$ for any positive integer M. To this end, we define various notions of characteristic property on the product space $\mathcal{X}$ (see Definition 1 and Figure 2(a) in Section 3) and explore the relation between them. In order to keep our presentation in this section to be non-technical, we relegate the problem formulation to Section 3, with the main results of the paper being presented in Section 4. A summary of the results is captured in Figure 1 while the proofs are provided in Section 5. Various definitions and notation that are used throughout the paper are collected in Section 2.

2 Definitions & Notation

$\mathbb{N} := \{1, 2, \ldots\}$ and $\mathbb{R}$ denotes the set of natural numbers and real numbers respectively. For $M \in \mathbb{N}$, $[M] := \{1, \ldots, M\}$. $\mathbf{1}_d := (1, 1, \ldots, 1) \in \mathbb{R}^d$ and $\mathbf{0}$ denotes the matrix of zeros. For $a := (a_1, \ldots, a_d) \in \mathbb{R}^d$ and $b := (b_1, \ldots, b_d) \in \mathbb{R}^d$, $\langle a, b\rangle = \sum_{i=1}^d a_i b_i$ is the Euclidean inner product. For sets A and B, $A \backslash B = \{a \in A : a \notin B\}$ is their difference, $|A|$ is the cardinality of A and $\times_{m=1}^M A_m = \{(a_1, \ldots, a_M) : a_m \in A_m\, m \in [M]\}$ is the Descartes product of sets $(A_m)_{m=1}^M$. $\mathcal{P}(\mathcal{X})$ denotes the power set of a set $\mathcal{X}$, i.e., all subsets of $\mathcal{X}$ (including the empty set and $\mathcal{X}$). The Kronecker delta is defined as $\delta_{a,b} = 1$ if $a = b$, and zero otherwise. χ_A is the indicator function of set A: $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ otherwise. $\mathbb{R}^{d_1 \times \ldots \times d_M}$ is the set of $d_1 \times \ldots \times d_M$-sized tensors.

For a topological space $(\mathcal{X}, \tau_{\mathcal{X}})$, $\mathcal{B}(\mathcal{X}) := \mathcal{B}(\tau_{\mathcal{X}})$ is the Borel sigma-algebra on $\mathcal{X}$ induced by the topology $\tau_{\mathcal{X}}$. Probability and finite signed measures in the paper are meant w.r.t. the

[1] Waegeman et al. [2012] deals with c-universal kernels while Gretton [2015] deals with c_0-universal kernels. A brief description of these notions are provided in Section 3. We refer the reader to [Carmeli et al., 2010, Sriperumbudur et al., 2010] for more details on these notions of universality.

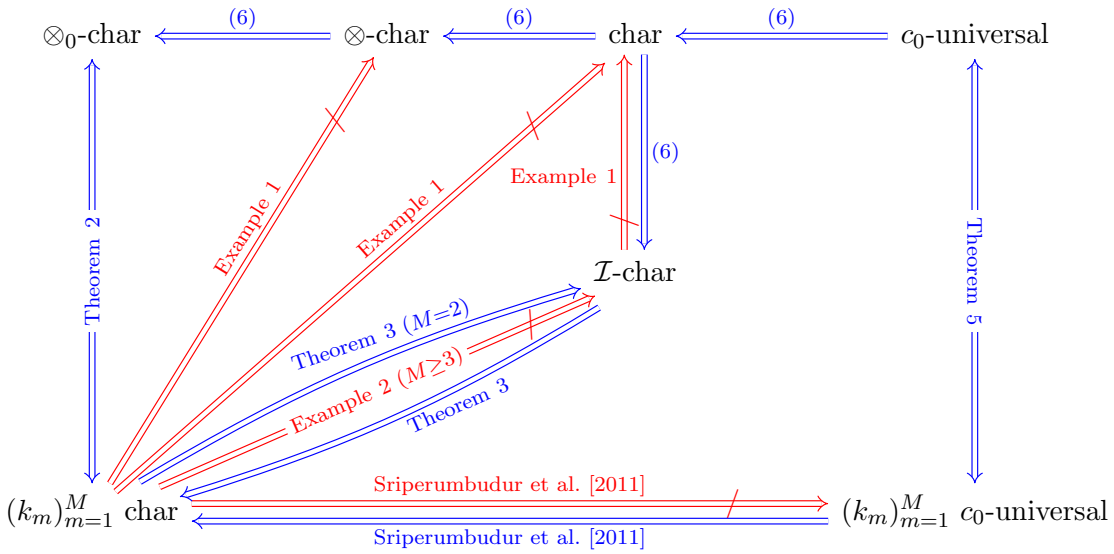


Figure 1: Summary of results: "char" denotes characteristic. In addition to the usual characteristic property, three new notions $\otimes_0$-characteristic, $\otimes$-characteristic and $\mathcal{I}$-characteristic are introduced in Definition 1 which along with c_0-universal (in the top right corner) correspond to the property of the tensor product kernel $\otimes_{m=1}^M k_m$, while the bottom part of the picture corresponds to the individual kernels $(k_m)_{m=1}^M$ being characteristic or c_0-universal. If $(k_m)_{m=1}^M$-s are continuous, bounded and translation invariant kernels on $\mathbb{R}^{d_m}$, $m \in [M]$, all the notions are equivalent (see Theorem 4).

measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Given $\{(\mathcal{X}_i, \tau_i)\}_{i\in I}$ topological spaces, their product $\times_{i\in I}\mathcal{X}_i$ is enriched with the product topology; it is the coarsest topology for which the canonical projections $\pi_i : \times_{i\in I}\mathcal{X}_i \to (\mathcal{X}_i, \tau_i)$ are continuous for all $i \in I$. $C(\mathcal{X})$ denotes the space of continuous functions on $\mathcal{X}$. $C_0(\mathcal{X})$ denotes the class of real-valued functions vanishing at infinity on a locally compact Hausdorff (LCH) space[2] $\mathcal{X}$, i.e., for any $\epsilon > 0$, the set $\{x \in \mathcal{X} : |f(x)| \geq \epsilon\}$ is compact. $C_0(\mathcal{X})$ is endowed with the uniform norm $\|f\|_\infty = \sup_{x\in\mathcal{X}} |f(x)|$. $\mathcal{M}_b(\mathcal{X})$ and $\mathcal{M}_1^+(\mathcal{X})$ are the space of finite signed measures and probability measures on $\mathcal{X}$, respectively. For $\mathbb{P}_m \in \mathcal{M}_1^+(\mathcal{X}_m)$, $\otimes_{m=1}^M \mathbb{P}_m$ denotes the product probability measure on the product space $\times_{m=1}^M \mathcal{X}_m$, i.e., $\otimes_{m=1}^M \mathbb{P}_m \in \mathcal{M}_1^+(\times_{m=1}^M \mathcal{X}_m)$. δ_x denotes the Dirac measure supported on $x \in \mathcal{X}$.

$\mathcal{H}_{k_m}$ is the reproducing kernel Hilbert space (RKHS) associated with the reproducing kernel $k_m : \mathcal{X}_m \times \mathcal{X}_m \to \mathbb{R}$, which in this paper is assumed to be measurable and bounded. The tensor product of $(k_m)_{m=1}^M$ is a kernel, defined as

$$\otimes_{m=1}^M k_m \left((x_1, \ldots, x_M), (x'_1, \ldots, x'_M)\right) = \prod_{m=1}^M k_m \left(x_m, x'_m\right), \quad x_m, x'_m \in \mathcal{X}_m, \tag{3}$$

whose associated RKHS is denoted as $\mathcal{H}_{\otimes_{m=1}^M k_m} = \otimes_{m=1}^M \mathcal{H}_{k_m}$ [Berlinet and Thomas-Agnan, 2004, Theorem 13], where the r.h.s. is the tensor product of RKHSs $(\mathcal{H}_{k_m})_{m=1}^M$. For $h_m \in \mathcal{H}_m$,

[2]LCH spaces include $\mathbb{R}^d$, discrete spaces, and topological manifolds. Open or closed subsets, finite products of LCH spaces are LCH. Infinite-dimensional Hilbert spaces are *not* LCH.

$m \in [M]$, the multi-linear operator $\otimes_{m=1}^{M} h_m \in \otimes_{m=1}^{M} \mathcal{H}_m$ is defined as

$$\left(\otimes_{m=1}^{M} h_m\right)(v_1, \ldots, v_M) = \prod_{m=1}^{M} \langle h_m, v_m \rangle_{\mathcal{H}_m}, \quad v_m \in \mathcal{H}_m.$$

For an LCH space $\mathcal{X}$, $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a c_0-kernel if $k(\cdot, x) \in C_0(\mathcal{X})$ for all $x \in \mathcal{X}$. $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is said to be translation invariant if $k(x, y) = \psi(x - y)$, $x, y \in \mathbb{R}^d$ for a positive definite function $\psi : \mathbb{R}^d \to \mathbb{R}$. $\mu_k(\mathbb{F})$ denotes the kernel mean embedding of $\mathbb{F} \in \mathcal{M}_b(\mathcal{X})$ to $\mathcal{H}_k$ which is defined as $\mu_k(\mathbb{F}) = \int_{\mathcal{X}} k(\cdot, x) \, \mathrm{d}\mathbb{F}(x)$, where the integral is meant in the Bochner sense.

3 Problem Formulation

In this section, we formally introduce the goal of the paper. To this end, we start with a definition. For simplicity, throughout the paper, we assume that all kernels are bounded. The definition is based on the observation [Sriperumbudur et al., 2010, Lemma 8] that a bounded kernel k on a topological space $(\mathcal{X}, \tau_{\mathcal{X}})$ is characteristic if and only if

$$\int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') \, \mathrm{d}\mathbb{F}(x) \, \mathrm{d}\mathbb{F}(x') > 0, \ \forall \, \mathbb{F} \in \mathcal{M}_b(\mathcal{X}) \backslash \{0\} \text{ such that } \mathbb{F}(\mathcal{X}) = 0.$$

In other words, characteristic kernels are integrally strictly positive definite (ispd) (see Sriperumbudur et al., 2010, p. 1523) w.r.t. the class of finite signed measures that assign zero measure to $\mathcal{X}$. The following definition extends this observation to tensor product kernels on product spaces.

Definition 1 ($\mathcal{F}$-ispd tensor product kernel) *Suppose $k_m : \mathcal{X}_m \times \mathcal{X}_m \to \mathbb{R}$ is a bounded kernel on a topological space $(\mathcal{X}_m, \tau_{\mathcal{X}_m})$, $m \in [M]$. Let $\mathcal{F} \subseteq \mathcal{M}_b(\mathcal{X})$ be such that $0 \in \mathcal{F}$ where $\mathcal{X} := \times_{m=1}^{M} \mathcal{X}_m$. $k := \otimes_{m=1}^{M} k_m$ is said to be* $\mathcal{F}$-ispd *if*

$$\mu_k(\mathbb{F}) = 0 \Rightarrow \mathbb{F} = 0 \quad (\mathbb{F} \in \mathcal{F}), \text{ or equivalently}$$

$$\|\mu_k(\mathbb{F})\|_{\mathcal{H}_k}^2 = \int_{\times_{m=1}^{M} \mathcal{X}_m} \int_{\times_{m=1}^{M} \mathcal{X}_m} \left(\otimes_{m=1}^{M} k_m\right)\left(x, x'\right) \, \mathrm{d}\mathbb{F}(x) \, \mathrm{d}\mathbb{F}(x') > 0, \quad \forall \, \mathbb{F} \in \mathcal{F} \backslash \{0\}. \tag{4}$$

Specifically,

- *if k_m-s are c_0-kernels on locally compact Polish*[3] *spaces $\mathcal{X}_m$-s and $\mathcal{F} = \mathcal{M}_b(\mathcal{X})$, then k is called c_0-universal.*
- *if*

$$\begin{aligned}
\mathcal{F} &= [\mathcal{M}_b(\mathcal{X})]^0 && := \{\mathbb{F} \in \mathcal{M}_b(\mathcal{X}) : \mathbb{F}(\mathcal{X}) = 0\}, \\
\mathcal{F} &= \left[\otimes_{m=1}^{M} \mathcal{M}_b(\mathcal{X}_m)\right]^0 && := \left\{\mathbb{F} \in \otimes_{m=1}^{M} \mathcal{M}_b\left(\mathcal{X}_m\right), \mathbb{F}(\mathcal{X}) = 0\right\}, \\
\mathcal{F} &= \mathcal{I} && := \left\{\mathbb{P} - \otimes_{m=1}^{M} \mathbb{P}_m : \mathbb{P} \in \mathcal{M}_1^+\left(\times_{m=1}^{M} \mathcal{X}_m\right)\right\}, \\
\mathcal{F} &= \otimes_{m=1}^{M} \mathcal{M}_b^0(\mathcal{X}_m) && := \left\{\mathbb{F} = \otimes_{m=1}^{M} \mathbb{F}_m \in \otimes_{m=1}^{M} \mathcal{M}_b\left(\mathcal{X}_m\right), \mathbb{F}_m(\mathcal{X}_m) = 0, \ \forall m \in [M]\right\},
\end{aligned}$$

  then k is called characteristic*, $\otimes$-*characteristic*, $\mathcal{I}$-*characteristic *and* $\otimes_0$-characteristic*, respectively.*

[3] A topological space is called Polish if it is complete, separable and metrizable. For example, $\mathbb{R}^d$ and countable discrete spaces are Polish. Open and closed subsets, products and disjoint unions of countably many Polish spaces are Polish. Every second-countable (i.e., its topology has countable basis) LCH space is Polish.

In Definition 1, k being characteristic matches the usual notion of characteristic kernels on a product space, i.e., there are no two distinct probability measures on $\mathcal{X} = \times_{m=1}^{M} \mathcal{X}_m$ such that the MMD between them is zero. The other notions such as $\otimes$-characteristic, $\mathcal{I}$-characteristic and $\otimes_0$-characteristic are typically weaker than the usual characteristic property as

$$\otimes_{m=1}^{M} \mathcal{M}_b^0(\mathcal{X}_m) \subseteq \left[\otimes_{m=1}^{M} \mathcal{M}_b(\mathcal{X}_m)\right]^0 \subseteq \left[\mathcal{M}_b\left(\times_{m=1}^{M} \mathcal{X}_m\right)\right]^0 \subseteq \mathcal{M}_b\left(\times_{m=1}^{M} \mathcal{X}_m\right). \quad (5)$$

$$\cup| \quad \mathcal{I}$$

Remark. *(i)* The family $\mathcal{I}$ is useful to describe the joint *independence* of M random variables—hence the name $\mathcal{I}$-characteristic—defined on kernel-endowed domains $(\mathcal{X}_m)_{m=1}^{M}$: If $\mathbb{P}$ denotes the joint distribution of random variables $(X_m)_{m=1}^{M}$, then by definition $k = \otimes_{m=1}^{M} k_m$ is $\mathcal{I}$-characteristic iff

$$\mathrm{HSIC}_k(\mathbb{P}) = 0 \Leftrightarrow \mathbb{P} = \otimes_{m=1}^{M} \mathbb{P}_m.$$

In other words, HSIC captures joint independence exactly with $\mathcal{I}$-characteristic kernels.

(ii) For the notions of $\otimes$-characteristic and $\otimes_0$-characteristic, the class $\mathcal{F}$ is chosen to be the product of finite signed measures on $\mathcal{X}$ with a slight difference that in the latter case, each marginal measure $\mathbb{F}_m$ is assumed to assign zero measure to the corresponding space $\mathcal{X}_m$ while in the former, the entire joint measure $\mathbb{F}$ is restricted to assign measure zero to $\mathcal{X}$.

(iii) When $\mathcal{F} = \mathcal{M}_b(\mathcal{X})$ with k_m being c_0-kernels on LCH space $\mathcal{X}_m$ for all $m \in [M]$, then k is also a c_0-kernel on LCH space $\mathcal{X}$ implying that if k satisfies (4), then k is c_0-universal (see Sriperumbudur et al., 2010, Proposition 2). It is well known that c_0-universality reduces to c-universality (i.e., the notion of universality proposed by Steinwart, 2001) if $\mathcal{X}$ is compact (see Sriperumbudur et al., 2010 for details) which is guaranteed if and only if each $\mathcal{X}_m$, $m \in [M]$ is compact.

Given the relations in (5), it immediately follows that $k = \otimes_{m=1}^{M} k_m$ satisfies

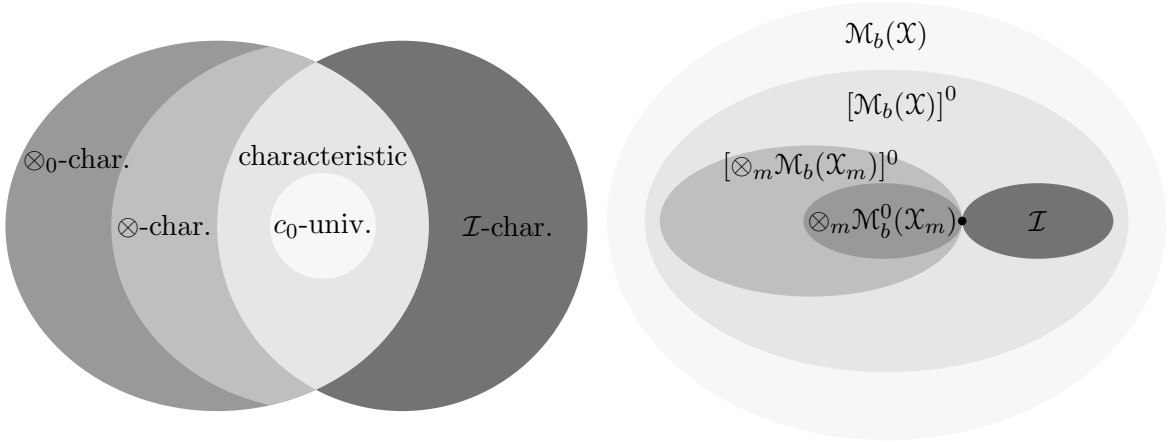
$$\otimes_0\text{-characteristic} \Leftarrow \otimes\text{-characteristic} \Leftarrow \text{characteristic} \Leftarrow c_0\text{-universal} \quad (6)$$
$$\Downarrow$$
$$\mathcal{I}\text{-characteristic}$$

when $\mathcal{X}_m$ for all $m \in [M]$ are LCH. A visual illustration of (5) and (6) is provided in Figure 2. The goal of this paper is to investigate whether the characteristic or c_0-universal property of k_m-s ($m \in [M]$) imply different $\mathcal{F}$-ispd properties of $\otimes_{m=1}^{M} k_m$, and vice versa.

4 Main Results

In this section, we present our main results related to the $\mathcal{F}$-ispd property of tensor product kernels, which are summarized in Figure 1. First, in the following result, we show that the characteristic property of individual kernels $(k_m)_{m=1}^{M}$ need *not* be equivalent to that of $\otimes_{m=1}^{M} k_m$, but is equivalent to $\otimes_0$-characteristic property of $\otimes_{m=1}^{M} k_m$.

Theorem 2 *Let $k_m : \mathcal{X}_m \times \mathcal{X}_m \to \mathbb{R}$ be bounded kernels on topological spaces $\mathcal{X}_m$ for all $m \in [M]$. Then the following holds.*



(a) (b)

Figure 2: (a) $\mathcal{F}$-ispd $\otimes_{m=1}^{M} k_m$ kernels (see (6)); (b) $\mathcal{F} \subseteq \mathcal{M}_b(\mathcal{X})$, $\mathcal{X} = \times_{m=1}^{M} \mathcal{X}_m$.

(i) *If $\otimes_{m=1}^{M} k_m$ is characteristic, then it is $\otimes$-characteristic.*
(ii) *If $\otimes_{m=1}^{M} k_m$ is $\otimes$-characteristic, then it is $\otimes_0$-characteristic.*
(iii) *$\otimes_{m=1}^{M} k_m$ is $\otimes_0$-characteristic iff $k_1, \ldots, k_M$ are characteristic.*

In the above result, the converse to assertion *(ii)* is in general not true as demonstrated by the following example, which therefore demonstrates that the tensor product kernel need not be characteristic even if all k_m-s are characteristic.

Example 1 *Let $\mathcal{X}_1 = \mathcal{X}_2 = \{1, 2\}$, $\tau_{\mathcal{X}_1} = \tau_{\mathcal{X}_2} = \mathcal{P}(\{1, 2\})$, $k_1(x, x') = k_2(x, x') = 2\delta_{x,x'} - 1$. It is easy to verify that k_1 and k_2 are characteristic. However, it can be proved that $k_1 \otimes k_2$ is not $\otimes$-characteristic[4] and therefore not characteristic. On the hand, interestingly, $k_1 \otimes k_2$ is $\mathcal{I}$-characteristic. We refer the reader to Section 5.2 for details.*

In the above example, we showed that the tensor product of k_1 and k_2 (which are characteristic kernels) is $\mathcal{I}$-characteristic. The following result generalizes this behavior for any bounded characteristic kernels. In addition, under a mild extra assumption, it shows the converse to be true for any M.

Theorem 3 *Let $k_m : \mathcal{X}_m \times \mathcal{X}_m \to \mathbb{R}$ be bounded kernels on topological spaces $\mathcal{X}_m$ for all $m \in [M]$. Then the following holds.*

(i) *If k_1 and k_2 are characteristic, then $k_1 \otimes k_2$ is $\mathcal{I}$-characteristic.*
(ii) *Suppose $\mathcal{X}_m$ is Hausdorff and $|\mathcal{X}_m| \geq 2$ for all $m \in [M]$. If $\otimes_{m=1}^{M} k_m$ is $\mathcal{I}$-characteristic, then $k_1, \ldots, k_M$ are characteristic.*

Lyons [2013] has showed an analogous result to Theorem 3*(i)* for distance covariances ($M = 2$) on metric spaces of negative type (see Theorem 3.11 in Lyons, 2013), which via

[4] An interesting and different example is constructed by Sejdinovic et al. [2013, Remark 31], illustrating that the tensor product of characteristic kernels need not be (using our terminology) $\otimes$-characteristic. However, unlike Example 1, the construction in [Sejdinovic et al., 2013] does not hint the $\mathcal{I}$-characteristic property of the tensor product kernel.

Sejdinovic et al. [2013, Proposition 29] holds for HSIC yielding the $\mathcal{I}$-characteristic property of $k_1 \otimes k_2$. Recently, Gretton [2015] presented a direct proof showing that HSIC corresponding to $k_1 \otimes k_2$ captures independence if k_1 and k_2 are translation invariant characteristic kernels on $\mathbb{R}^d$ (which is equivalent to c_0-universality). In contrast, Theorem 3*(i)* establishes the result for bounded kernels on general topological spaces. In fact, the result of [Gretton, 2015] is a special case of Theorems 4 and 5 below. Theorem 3*(i)* raises a pertinent question: whether $\otimes_{m=1}^{M} k_m$ is $\mathcal{I}$-characteristic if k_m-s are characteristic for all $m \in [M]$ where $M > 2$? The following example provides a negative answer to this question. On a positive side, however, we will see in Theorem 5 that the $\mathcal{I}$-characteristic property of $\otimes_{m=1}^{M} k_m$ can be guaranteed for any $M \geq 2$ if a stronger condition is imposed on k_m-s (and $\mathcal{X}_m$-s). Theorem 3*(ii)* generalizes Proposition 3.15 of [Lyons, 2013] for any $M > 2$.

Example 2 *Let $M = 3$ and $\mathcal{X}_m := \{1, 2\}$, $\tau_{\mathcal{X}_m} = \mathcal{P}(\mathcal{X}_m)$, $k_m(x, x') = 2\delta_{x,x'} - 1$ $(m = 1, 2, 3)$. As mentioned in Example 1, $(k_m)_{m=1}^{3}$ are characteristic. However, it can be shown that $\otimes_{m=1}^{3} k_m$ is not $\mathcal{I}$-characteristic. See Section 5.3 for details.*

In Theorem 2 and Example 1, we showed that in general, only the $\otimes_0$-characteristic property of $\otimes_{m=1}^{M} k_m$ is equivalent to the characteristic property of k_m-s. Our next result shows that all the various notions of characteristic property of $\otimes_{m=1}^{M} k_m$ *coincide* if k_m-s are translation-invariant, continuous bounded kernels on $\mathbb{R}^d$.

Theorem 4 *Suppose $k_m : \mathbb{R}^{d_m} \times \mathbb{R}^{d_m} \to \mathbb{R}$ are continuous, bounded and translation-invariant kernels for all $m \in [M]$. Then the following statements are equivalent:*

(i) *k_m-s are characteristic for all $m \in [M]$;*
(ii) *$\otimes_{m=1}^{M} k_m$ is $\otimes_0$-characteristic;*
(iii) *$\otimes_{m=1}^{M} k_m$ is $\otimes$-characteristic;*
(iv) *$\otimes_{m=1}^{M} k_m$ is $\mathcal{I}$-characteristic;*
(v) *$\otimes_{m=1}^{M} k_m$ is characteristic.*

Our final result shows that on LCP spaces, the tensor product of $M \geq 2$ c_0-universal kernels is also c_0-universal, and vice versa.

Theorem 5 *Suppose $k_m : \mathcal{X}_m \times \mathcal{X}_m \to \mathbb{R}$ are c_0-kernels on LCP spaces $\mathcal{X}_m$ $(m \in [M])$. Then $\otimes_{m=1}^{M} k_m$ is c_0-universal iff k_m-s are c_0-universal for all $m \in [M]$.*

A special case of Theorem 5 for $M = 2$ is proved by Lyons [2013, Lemma 3.8] in the context of distance covariance which reduces to Theorem 5 through the equivalence established by Sejdinovic et al. [2013]. Another special case of Theorem 5 is proved[5] by Waegeman et al. [2012, Theorem VII.2] for c-universality with $M = 2$ using the Stone-Weierstrass theorem: if k_1 and k_2 are c-universal then $k_1 \otimes k_2$ is c-universal. Since the notions of c_0-universality and characteristic property are equivalent for translation-invariant c_0-kernels on $\mathbb{R}^d$ (see Carmeli et al., 2010, Proposition 5.16 and Sriperumbudur et al., 2010, Theorem 9), Theorem 4 can be

[5]In fact there is a small technical error in [Waegeman et al., 2012]: $C(\mathcal{X}) \times C(\mathcal{X}) = \{(x_1, x_2) \mapsto f_1(x_1) f_2(x_2), f_m \in C(\mathcal{X}), m = 1, 2\}$ is claimed to be an algebra. This is not true since $C(\mathcal{X}) \times C(\mathcal{X})$ is not closed w.r.t. addition; in other words $f, g \in C(\mathcal{X}) \times C(\mathcal{X}) \not\Rightarrow f + g \in C(\mathcal{X}) \times C(\mathcal{X})$. However, the issue can be resolved by taking the linear span of $C(\mathcal{X}) \times C(\mathcal{X})$.

considered as a special case of Theorem 5. Since the c_0-universality of $\otimes_{m=1}^M k_m$ implies its $\mathcal{I}$-characteristic property (see (6)), Theorem 5 also provides a generalization of Theorem 3*(i)* to $M \geq 2$ under additional assumptions on k_m-s, while constraining $\mathcal{X}_m$-s to LCP-s instead of general topological spaces.

5 Proofs

In this section, we provide the proofs of our results presented in Section 4.

5.1 Proof of Theorem 2

By (6), it is sufficient to prove (*iii*). Note that for any $\mathbb{F} = \otimes_{m=1}^M \mathbb{F}_m \in \otimes_{m=1}^M \mathcal{M}_b(\mathcal{X}_m)$, $\mathbb{F}_m(\mathcal{X}_m) = 0$ $(\forall m \in [M])$,

$$\begin{aligned}
\|\mu_k(\mathbb{F})\|^2_{\mathcal{H}_{\otimes_{m=1}^M k_m}} &= \int_{\times_{m=1}^M \mathcal{X}_m} \int_{\times_{m=1}^M \mathcal{X}_m} \left(\otimes_{m=1}^M k_m\right)(x, x') \, \mathrm{d}\mathbb{F}(x) \, \mathrm{d}\mathbb{F}(x') \\
&= \int_{\times_{m=1}^M \mathcal{X}_m} \int_{\times_{m=1}^M \mathcal{X}_m} \left(\otimes_{m=1}^M k_m\right)(x, x') \, \mathrm{d}\left[\otimes_{m=1}^M \mathbb{F}_m\right](x) \, \mathrm{d}\left[\otimes_{m=1}^M \mathbb{F}_m\right](x') \\
&= \prod_{m=1}^M \int_{\mathcal{X}_m \times \mathcal{X}_m} k_m(x_m, x'_m) \, \mathrm{d}\mathbb{F}_m(x_m) \, \mathrm{d}\mathbb{F}_m(x'_m) \\
&= \prod_{m=1}^M \|\mu_{k_m}(\mathbb{F}_m)\|^2_{\mathcal{H}_{k_m}} . \qquad (7)
\end{aligned}$$

$(\Rightarrow)$ If k_m-s are characteristic, then $\|\mu_{k_m}(\mathbb{F}_m)\|_{\mathcal{H}_{k_m}} > 0$, $\forall\, \mathbb{F}_m \in \mathcal{M}_b(\mathcal{X}_m) \backslash \{0\}$, $\mathbb{F}_m(\mathcal{X}_m) = 0$ and all $m \in [M]$. It therefore follows from (7) that $\|\mu_k(\mathbb{F})\|_{\mathcal{H}_{\otimes_{m=1}^M k_m}} > 0$, $\forall\, \mathbb{F} = \otimes_{m=1}^M \mathbb{F}_m \in \otimes_{m=1}^M \mathcal{M}_b(\mathcal{X}_m)$, $\mathbb{F}_m(\mathcal{X}_m) = 0$ $(\forall m \in [M])$, implying that $\otimes_{m=1}^M k_m$ is $\otimes_0$-characteristic.

$(\Leftarrow)$ If $\otimes_{m=1}^M k_m$ is $\otimes_0$-characteristic, then $\|\mu_k(\mathbb{F})\|_{\mathcal{H}_{\otimes_{m=1}^M k_m}} > 0$ for all $\mathbb{F} = \otimes_{m=1}^M \mathbb{F}_m \in \otimes_{m=1}^M \mathcal{M}_b(\mathcal{X}_m)$, $\mathbb{F}_m(\mathcal{X}_m) = 0$ $(\forall m \in [M])$. Therefore (7) implies $\|\mu_{k_m}(\mathbb{F}_m)\|_{\mathcal{H}_{k_m}} > 0$, $\forall\, \mathbb{F}_m \in \mathcal{M}_b(\mathcal{X}_m) \backslash \{0\}$, $\mathbb{F}_m(\mathcal{X}_m) = 0$ and all $m \in [M]$, i.e., k_m-s are characteristic.

5.2 Proof of Example 1

The proof is structured as follows.

1. First we show that $k := k_1 = k_2$ is a kernel and it is characteristic.
2. Next it is proved that $k_1 \otimes k_2$ is not $\otimes$-characteristic and therefore is not characteristic.
3. Finally, the $\mathcal{I}$-characteristic property of $k_1 \otimes k_2$ is established.

The individual steps are as follows:

k **is a kernel.** Assume w.l.o.g. that $x_1 = \ldots = x_N = 1$, $x_{N+1} = \ldots = x_n = 2$. Then it is easy to verify that the Gram matrix $\mathbf{G} = [k(x_i, x_j)]_{i,j=1}^n = \mathbf{a}\mathbf{a}^\top$ where $\mathbf{a} := \left(\mathbf{1}_N^\top, -\mathbf{1}_{n-N}^\top\right)^\top$ and $\mathbf{a}^\top$ is the transpose of $\mathbf{a}$. Clearly $\mathbf{G}$ is positive semidefinite and so k is a kernel.

k **is characteristic**. We will show that k satisfies (4). On $\mathcal{X} = \{1, 2\}$ a finite signed measure $\mathbb{F}$ takes the form $\mathbb{F} = a_1\delta_1 + a_2\delta_2$ for some $a_1, a_2 \in \mathbb{R}$. Thus,

$$\mathbb{F} \in \mathcal{M}_b(\mathcal{X})\backslash\{0\} \Leftrightarrow (a_1, a_2) \neq \mathbf{0} \qquad \text{and} \qquad \mathbb{F}(\mathcal{X}) = 0 \Leftrightarrow a_1 + a_2 = 0. \tag{8}$$

Consider

$$\begin{aligned}\int_{\mathcal{X}}\int_{\mathcal{X}} k(x, x')\,\mathrm{d}\mathbb{F}(x)\,\mathrm{d}\mathbb{F}(x') &= a_1^2 k(1,1) + a_2^2 k(2,2) + 2a_1a_2k(1,2)\\ &= a_1^2 + a_2^2 - 2a_1a_2 = (a_1 - a_2)^2 = 4a_1^2 > 0,\end{aligned} \tag{9}$$

where we used (8) and the facts that $k(1,1) = k(2,2) = 1$, $k(1,2) = -1$.

$k_1 \otimes k_2$ **is not** $\otimes$**-characteristic**. Our goal is to construct a witness $\mathbb{F} = \mathbb{F}_1 \otimes \mathbb{F}_2 \in \otimes_{m=1}^2 \mathcal{M}_b(\mathcal{X}_m)\backslash\{0\}$ such that

$$\mathbb{F}(\mathcal{X}_1 \times \mathcal{X}_2) = \mathbb{F}_1(\mathcal{X}_1)\mathbb{F}_2(\mathcal{X}_2) = 0, \tag{10}$$

and

$$\begin{aligned}0 &= \int_{\mathcal{X}_1\times\mathcal{X}_2}\int_{\mathcal{X}_1\times\mathcal{X}_2} \underbrace{(k_1 \otimes k_2)((i_1, i_2), (i_1', i_2'))}_{k_1(i_1,i_1')k_2(i_2,i_2')}\,\mathrm{d}\mathbb{F}(i_1, i_2)\,\mathrm{d}\mathbb{F}(i_1', i_2')\\ &= \prod_{m=1}^{2}\int_{\mathcal{X}_m}\int_{\mathcal{X}_m} k_m(i_m, i_m')\,\mathrm{d}\mathbb{F}_m(i_m)\,\mathrm{d}\mathbb{F}_m(i_m').\end{aligned} \tag{11}$$

Finite signed measures on $\{1, 2\}$ take the form $\mathbb{F}_1 = \mathbb{F}_1(\mathbf{a}) = a_1\delta_1 + a_2\delta_2$, $\mathbb{F}_2 = \mathbb{F}_2(\mathbf{b}) = b_1\delta_1 + b_2\delta_2$ form, where $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2, \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$. With these notations, (10) and (11) can be rewritten as

$$\begin{aligned}0 &= (a_1 + a_2)(b_1 + b_2),\\ 0 &= \left[\sum_{i,i'=1}^{2} k_1(i, i')a_ia_{i'}\right]\left[\sum_{j,j'=1}^{2} k_2(j, j')b_jb_{j'}\right] = (a_1 - a_2)^2(b_1 - b_2)^2.\end{aligned}$$

Keeping the solutions where neither $\mathbf{a}$ nor $\mathbf{b}$ is the zero vector, there are 2 (symmetric) possibilities: (i) $a_1 + a_2 = 0$, $b_1 = b_2$ and (ii) $a_1 = a_2$, $b_1 + b_2 = 0$. In other words, for any $a, b \neq 0$, the possibilities are (i) $\mathbf{a} = (a, -a)$, $\mathbf{b} = (b, b)$ and (ii) $\mathbf{a} = (a, a)$, $\mathbf{b} = (b, -b)$. This establishes the non-$\otimes$-characteristic property of $k_1 \otimes k_2$.

$k_1 \otimes k_2$ **is** $\mathcal{I}$**-characteristic**. Our goal is to show that $k_1 \otimes k_2$ is $\mathcal{I}$-characteristic, i.e., for any $\mathbb{P} \in \mathcal{M}_1^+(\mathcal{X}_1 \times \mathcal{X}_2)$, $\mu_{k_1\otimes k_2}(\mathbb{F}) = 0$ implies $\mathbb{F} = 0$, where $\mathbb{F} = \mathbb{P} - \mathbb{P}_1 \otimes \mathbb{P}_2$. We divide the proof into two parts:

1. First we derive the equations of

$$\mathbb{F}(\mathcal{X}_1 \times \mathcal{X}_2) = 0 \quad \text{and} \quad \int\int_{(\mathcal{X}_1\times\mathcal{X}_2)^2} (k_1 \otimes k_2)\left((i, j), (r, s)\right)\,\mathrm{d}\mathbb{F}(i, j)\,\mathrm{d}\mathbb{F}(r, s) = 0 \tag{12}$$

for general finite signed measures $\mathbb{F} = \sum_{i,j=1}^{2} a_{ij}\delta_{(i,j)}$ on $\mathcal{X}_1 \times \mathcal{X}_2$.

2. Then, we apply the $\mathbb{F} = \mathbb{P} - \mathbb{P}_1 \otimes \mathbb{P}_2$ parameterization and solve for $\mathbb{P}$ that satisfies (12) to conclude that $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$, i.e., $\mathbb{F} = 0$. Note that in the chosen parametrization for $\mathbb{F}$, $\mathbb{F}(\mathcal{X}_1 \times \mathcal{X}_2) = 0$ holds automatically.

The details are as follows.

Step 1.

$$0 = \mathbb{F}(\mathcal{X}_1 \times \mathcal{X}_2) \Leftrightarrow 0 = a_{11} + a_{12} + a_{21} + a_{22}, \tag{13}$$

$$\begin{aligned}
0 &= \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\mathcal{X}_1 \times \mathcal{X}_2} \underbrace{(k_1 \otimes k_2)\left((i,j),(r,s)\right)}_{k_1(i,r)k_2(j,s)} \mathrm{d}\mathbb{F}(i,j)\, \mathrm{d}\mathbb{F}\,(r,s) \\
&= \sum_{i,j=1}^{2} \sum_{r,s=1}^{2} k_1(i,r)k_2(j,s)a_{ij}a_{rs} = \sum_{i,r=1}^{2} k_1(i,r) \sum_{j,s=1}^{2} k_2(j,s)a_{ij}a_{rs} \\
&= k_1(1,1)\left[k_2(1,1)a_{11}a_{11} + k_2(1,2)a_{11}a_{12} + k_2(2,1)a_{12}a_{11} + k_2(2,2)a_{12}a_{12}\right] \\
&\quad + k_1(1,2)\left[k_2(1,1)a_{11}a_{21} + k_2(1,2)a_{11}a_{22} + k_2(2,1)a_{12}a_{21} + k_2(2,2)a_{12}a_{22}\right] \\
&\quad + k_1(2,1)\left[k_2(1,1)a_{21}a_{11} + k_2(1,2)a_{21}a_{12} + k_2(2,1)a_{22}a_{11} + k_2(2,2)a_{22}a_{12}\right] \\
&\quad + k_1(2,2)\left[k_2(1,1)a_{21}a_{21} + k_2(1,2)a_{21}a_{22} + k_2(2,1)a_{22}a_{21} + k_2(2,2)a_{22}a_{22}\right] \\
&= \underbrace{\left(a_{11}^2 - 2a_{11}a_{12} + a_{12}^2\right)}_{(a_{11}-a_{12})^2} + \underbrace{\left(a_{21}^2 - 2a_{21}a_{22} + a_{22}^2\right)}_{(a_{21}-a_{22})^2} - 2\underbrace{\left(a_{11}a_{21} - a_{11}a_{22} - a_{12}a_{21} + a_{12}a_{22}\right)}_{(a_{11}-a_{12})(a_{21}-a_{22})} \\
&= (a_{11} - a_{12} - a_{21} + a_{22})^2.
\end{aligned} \tag{14}$$

Solving (13) and (14) yields

$$a_{11} + a_{22} = 0 \quad \text{and} \quad a_{12} + a_{21} = 0. \tag{15}$$

Step 2. Any $\mathbb{P} \in \mathcal{M}_1^+(\mathcal{X}_1 \times \mathcal{X}_2)$ can be parametrized as

$$\mathbb{P} = \sum_{i,j=1}^{2} p_{ij}\delta_{(i,j)}, \quad p_{ij} \geq 0,\, \forall (i,j) \quad \text{and} \quad \sum_{i,j=1}^{2} p_{ij} = 1. \tag{16}$$

Let $\mathbb{F} = \mathbb{P} - \mathbb{P}_1 \otimes \mathbb{P}_2 = \sum_{i,j=1}^{2} a_{ij}\delta_{(i,j)}$; for illustration see Table 1. It follows from step 1 that $\mathbb{F}$ satisfying (15) is equivalent to satisfying (12). Therefore, for the choice of $\mathbb{F} := \mathbb{P} - \mathbb{P}_1 \otimes \mathbb{P}_2$, we obtain

$$p_{11} - (p_{11} + p_{12})(p_{11} + p_{21}) + p_{22} - (p_{21} + p_{22})(p_{12} + p_{22}) = 0, \tag{17}$$

$$p_{12} - (p_{11} + p_{12})(p_{12} + p_{22}) + p_{21} - (p_{21} + p_{22})(p_{11} + p_{21}) = 0, \tag{18}$$

where $(p_{ij})_{i,j\in[2]}$ satisfy (16). Solving (16)–(18), we obtain

$$p_{11} = \frac{a[1-(a+b)]}{a+b}, \;\; p_{12} = \frac{b[1-(a+b)]}{a+b}, \;\; p_{21} = a \quad \text{and} \quad p_{22} = b,$$

with $0 \leq a, b \leq 1$, $a+b \leq 1$ and $(a,b) \neq \mathbf{0}$. The resulting distribution family with its marginals is summarized in Table 2. It can be seen that each member of this family (any a, b in the constraint set) factorizes: $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$. In other words, $\mathbb{F} = \mathbb{P} - \mathbb{P}_1 \otimes \mathbb{P}_2 = 0$; hence $k_1 \otimes k_2$ is

$\mathbb{P}$: $y\backslash x$	1	2	$\mathbb{P}_2$
1	p_{11}	p_{21}	$q_1 = p_{11} + p_{21}$
2	p_{12}	p_{22}	$q_2 = p_{12} + p_{22}$
$\mathbb{P}_1$	$p_1 = p_{11} + p_{12}$	$p_2 = p_{21} + p_{22}$	

$\Rightarrow$

$\mathbb{F} := \mathbb{P} - \mathbb{P}_1 \otimes \mathbb{P}_2$	1	2
1	$a_{11} = p_{11} - (p_{11} + p_{12})(p_{11} + p_{21})$	$a_{21} = p_{21} - (p_{21} + p_{22})(p_{11} + p_{21})$
2	$a_{12} = p_{12} - (p_{11} + p_{12})(p_{12} + p_{22})$	$a_{22} = p_{22} - (p_{21} + p_{22})(p_{12} + p_{22})$

Table 1: Joint ($\mathbb{P}$), joint minus product of the marginals ($\mathbb{P} - \mathbb{P}_1 \otimes \mathbb{P}_2$).

$\mathbb{P}$: $y\backslash x$	1	2	$\mathbb{P}_2$
1	$p_{11} = \frac{a[1-(a+b)]}{a+b}$	$p_{21} = a$	$q_1 = \frac{a}{a+b}$
2	$p_{12} = \frac{b[1-(a+b)]}{a+b}$	$p_{22} = b$	$q_2 = \frac{b}{a+b}$
$\mathbb{P}_1$	$p_1 = 1 - (a + b)$	$p_2 = a + b$	

Table 2: Family of probability distributions solving (16)–(18).

$\mathcal{I}$-characteristic.

Remark. We would like to mention that while k_1 and k_2 are characteristic, they are not universal. Since $\mathcal{X}$ is finite, the usual notion of universality (also called c-universality) matches with c_0-universality. Therefore, from (9), we have $\int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') \, \mathrm{d}\mathbb{F}(x) \, \mathrm{d}\mathbb{F}(x) = (a_1 - a_2)^2$ where $\mathbb{F} = a_1\delta_1 + a_2\delta_2$ for some $a_1, a_2 \in \mathbb{R}\backslash\{0\}$. Clearly, the choice of $a_1 = a_2$ establishes that there exists $\mathbb{F} \in \mathcal{M}_b(\mathcal{X})\backslash\{0\}$ such that $\int_{\mathcal{X}} \int_{\mathcal{X}} k(x, x') \, \mathrm{d}\mathbb{F}(x) \, \mathrm{d}\mathbb{F}(x) = 0$. Hence k is not universal. Note that the constraint in (8), which is need to verify the characteristic property of k is not need to verify its universality.

5.3 Proof of Example 2

Let $M = 3$, $\times_{m=1}^{M} \mathcal{X}_m = \{(i_1, i_2, i_3) : i_m = 1, 2, \, m \in [3]\}$, $k_m(x, x') = 2\delta_{x,x'} - 1$. Our goal is to show that $\otimes_{m=1}^{3} k_m$ is *not* $\mathcal{I}$-characteristic. The structure of the proof is as follows:

1. First we describe the equations of the non-characteristic property of $\otimes_{m=1}^{3} k_m$ with a general finite signed measure $\mathbb{F} = \sum_{i_1,i_2,i_3=1}^{2} a_{i_1,i_2,i_3} \delta_{(i_1,i_2,i_3)}$ on $\times_{m=1}^{3} \mathcal{X}_m$ where $a_{i_1,i_2,i_3} \in \mathbb{R}$ $(\forall i_1, i_2, i_3)$.
2. Next, we apply the $\mathbb{F} = \mathbb{P} - \otimes_{m=1}^{3} \mathbb{P}_m$ parameterization and show that there exists $\mathbb{P}$ that satisfies the equations of step 1 to conclude that $\otimes_{m=1}^{3} k_m$ is not $\mathcal{I}$-characteristic.

The details are as follows.

Step 1. The equations of non-characteristic property in terms of $\mathbf{A} = [a_{i_1,i_2,i_3}]_{(i_m)_{m=1}^3 \in [2]^3} \in \mathbb{R}^{2\times 2\times 2}$ are

$$\mathbb{F} \in \mathcal{M}_b \left(\times_{m=1}^{3} \mathcal{X}_m \right) \backslash \{0\} \Leftrightarrow \mathbf{A} \neq \mathbf{0},$$

$$0 = \mathbb{F}(\times_{m=1}^{3}\mathcal{X}_m) \Leftrightarrow 0 = \sum_{i_1,i_2,i_3=1}^{2} a_{i_1,i_2,i_3}, \tag{19}$$

$$\begin{aligned} 0 &= \int_{\times_{m=1}^{3}\mathcal{X}_m}\int_{\times_{m=1}^{3}\mathcal{X}_m} \underbrace{(\otimes_{m=1}^{3}k_m)\left((i_1,i_2,i_3),(i_1',i_2',i_3')\right)}_{\prod_{m=1}^{3}k_m(i_m,i_m')}\,\mathrm{d}\mathbb{F}(i_1,i_2,i_3)\,\mathrm{d}\mathbb{F}(i_1',i_2',i_3') \\ &= \sum_{i_1,i_2,i_3=1}^{2}\sum_{i_1',i_2',i_3'=1}^{2}\prod_{m=1}^{3}k_m(i_m,i_m')a_{i_1,i_2,i_3}a_{i_1',i_2',i_3'}. \end{aligned} \tag{20}$$

Solving (19) and (20) yields

$$a_{1,1,1} + a_{1,2,2} + a_{2,1,2} + a_{2,2,1} = 0 \quad \text{and} \quad a_{1,1,2} + a_{1,2,1} + a_{2,1,1} + a_{2,2,2} = 0.$$

Step 2. The equations of non $\mathcal{I}$-characteristic property can be obtained from step 1 by choosing $\mathbb{F} = \mathbb{P} - \otimes_{m=1}^{M}\mathbb{P}_m$, where

$$\mathbb{P} = \sum_{i_1,i_2,i_3=1}^{2} p_{i_1,i_2,i_3}\delta_{(i_1,i_2,i_3)} \quad \text{and} \quad \mathbf{P} = [p_{i_1,i_2,i_3}]_{(i_m)_{m=1}^{3}\in[2]^3} \in \mathbb{R}^{2\times2\times2}.$$

In other words, it is sufficient to obtain a $\mathbf{P}$ that solves the following system of equations for which $\mathbf{A} = \mathbf{A}(\mathbf{P}) \neq \mathbf{0}$:

$$\sum_{i_1,i_2,i_3=1}^{2} p_{i_1,i_2,i_3} = 1, \tag{21}$$

$$p_{i_1,i_2,i_3} \geq 0,\ \forall\,(i_1,i_2,i_3) \in [2]^3, \tag{22}$$

$$a_{1,1,1} + a_{1,2,2} + a_{2,1,2} + a_{2,2,1} = 0, \tag{23}$$

$$a_{1,1,2} + a_{1,2,1} + a_{2,1,1} + a_{2,2,2} = 0, \tag{24}$$

where

$$a_{i_1,i_2,i_3} = p_{i_1,i_2,i_3} - p_{1,i_1}p_{2,i_2}p_{3,i_3}, \tag{25}$$

and

$$p_{1,i_1} = \sum_{i_2,i_3=1}^{2} p_{i_1,i_2,i_3}, \quad p_{2,i_2} = \sum_{i_1,i_3=1}^{2} p_{i_1,i_2,i_3}, \quad p_{3,i_3} = \sum_{i_1,i_2=1}^{2} p_{i_1,i_2,i_3}. \tag{26}$$

One can get an analytical description for the solution of (21)–(26), where the solution $\mathbf{P}(\mathbf{z})$ is parameterized by $\mathbf{z} = (z_0,\ldots,z_5) \in \mathbb{R}^6$. For explicit expressions, we refer the reader to Appendix A. In the following, we present two examples of $\mathbf{P}$ that satisfy (21)–(26) such that $\mathbf{A} \neq \mathbf{0}$, thereby establishing the non $\mathcal{I}$-characteristic property of $\otimes_{m=1}^{3}k_m$.

1. $\mathbf{P}$:

$$\begin{aligned} p_{1,1,1} &= \frac{1}{5}, & p_{1,1,2} &= \frac{1}{10}, & p_{1,2,1} &= \frac{1}{10}, & p_{1,2,2} &= \frac{1}{10}, \\ p_{2,1,1} &= \frac{1}{5}, & p_{2,1,2} &= \frac{1}{10}, & p_{2,2,1} &= \frac{1}{10}, & p_{2,2,2} &= \frac{1}{10}, \end{aligned}$$

and **A**:

$$a_{1,1,1} = \frac{1}{50}, \qquad a_{1,1,2} = -\frac{1}{50}, \qquad a_{1,2,1} = -\frac{1}{50}, \qquad a_{1,2,2} = \frac{1}{50},$$
$$a_{2,1,1} = \frac{1}{50}, \qquad a_{2,1,2} = -\frac{1}{50}, \qquad a_{2,2,1} = -\frac{1}{50}, \qquad a_{2,2,2} = \frac{1}{50}.$$

2. **P**:

$$p_{1,1,1} = 0, \qquad p_{1,1,2} = \frac{1}{10}, \qquad p_{1,2,1} = \frac{1}{10}, \qquad p_{1,2,2} = \frac{1}{10},$$
$$p_{2,1,1} = \frac{1}{10}, \qquad p_{2,1,2} = \frac{1}{10}, \qquad p_{2,2,1} = \frac{3}{10}, \qquad p_{2,2,2} = \frac{1}{5},$$

and **A**:

$$a_{1,1,1} = -\frac{9}{200}, \qquad a_{1,1,2} = \frac{11}{200}, \qquad a_{1,2,1} = -\frac{1}{200}, \qquad a_{1,2,2} = -\frac{1}{200},$$
$$a_{2,1,1} = -\frac{1}{200}, \qquad a_{2,1,2} = -\frac{1}{200}, \qquad a_{2,2,1} = \frac{11}{200}, \qquad a_{2,2,2} = -\frac{9}{200}.$$

In fact these examples are obtained with the choices $\mathbf{z} = (\frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10})$ and $\mathbf{z} = (\frac{3}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{2}{10})$ respectively. See Appendix A for details.

5.4 Proof of Theorem 3

Define $\mathcal{H}_m := \mathcal{H}_{k_m}$.

(i) Suppose k_1 and k_2 are characteristic and that for some $\mathbb{F} = \mathbb{P} - \mathbb{P}_1 \otimes \mathbb{P}_2 \in \mathcal{I}$,

$$\mathcal{H}_1 \otimes \mathcal{H}_2 \ni \int_{\mathcal{X}_1 \times \mathcal{X}_2} (k_1 \otimes k_2)(\cdot, x)\, \mathrm{d}\mathbb{F}(x) = \int_{\mathcal{X}_1 \times \mathcal{X}_2} k_1(\cdot, x_1) \otimes k_2(\cdot, x_2)\, \mathrm{d}\mathbb{F}(x) = 0, \tag{27}$$

where $x = (x_1, x_2)$. We want to show that $\mathbb{F} = 0$, for which it is sufficient to prove that $\mathbb{F}(B_1 \times B_2) = 0$, $\forall B_m \in \mathcal{B}(\mathcal{X}_m)$, $m = 1, 2$. To this end, it follows from (27) that for all $h_2 \in \mathcal{H}_2$,

$$\mathcal{H}_1 \ni \int_{\mathcal{X}_1 \times \mathcal{X}_2} k_1(\cdot, x_1) h_2(x_2)\, \mathrm{d}\mathbb{F}(x) = \int_{\mathcal{X}_1} k_1(\cdot, x_1)\, \mathrm{d}\nu(x_1) = 0, \tag{28}$$

where

$$\nu(B_1) := \nu_{h_2}(B_1) = \int_{\mathcal{X}_1 \times \mathcal{X}_2} \chi_{B_1}(x_1) h_2(x_2)\, \mathrm{d}\mathbb{F}(x), \quad B_1 \in \mathcal{B}(\mathcal{X}_1).$$

Since k_1 is characteristic, (28) implies $\nu = 0$, provided that $|\nu|(\mathcal{X}_1) < \infty$ and $\nu(\mathcal{X}_1) = 0$. These two requirements hold:

$$\nu(\mathcal{X}_1) = \int_{\mathcal{X}_1 \times \mathcal{X}_2} h_2(x_2)\, \mathrm{d}\mathbb{F}(x) = \int_{\mathcal{X}_2} h_2(x_2)\, \mathrm{d}[\mathbb{P}_2 - \mathbb{P}_2](x_2) = 0;$$
$$|\nu|(\mathcal{X}_1) \leq \int_{\mathcal{X}_1 \times \mathcal{X}_2} \underbrace{|h_2(x_2)|}_{\left|\langle h_2, k_2(\cdot, x_2)\rangle_{\mathcal{H}_2}\right|} \mathrm{d}[\mathbb{P} + \mathbb{P}_1 \otimes \mathbb{P}_2](x_1, x_2)$$

$$\leq \|h_2\|_{\mathcal{H}_2} \int_{\mathcal{X}_1\times\mathcal{X}_2} \sqrt{k_2(x_2,x_2)}\,\mathrm{d}[\mathbb{P}+\mathbb{P}_1\otimes\mathbb{P}_2](x_1,x_2)$$
$$\leq 2\,\|h_2\|_{\mathcal{H}_2} \int_{\mathcal{X}_2} \sqrt{k_2(x_2,x_2)}\,\mathrm{d}\mathbb{P}_2(x_2) < \infty, \tag{29}$$

where the last inequality follows from the boundedness of k_2. The established $\nu = 0$ implies that for $\forall B_1 \in \mathcal{B}(\mathcal{X}_1)$ and $\forall h_2 \in \mathcal{H}_2$,

$$0 = \nu(B_1) = \left\langle h_2, \int_{\mathcal{X}_1\times\mathcal{X}_2} \chi_{B_1}(x_1) k_2(\cdot,x_2)\,\mathrm{d}\mathbb{F}(x) \right\rangle_{\mathcal{H}_2},$$

and hence

$$0 = \int_{\mathcal{X}_1\times\mathcal{X}_2} \chi_{B_1}(x_1) k_2(\cdot,x_2)\,\mathrm{d}\mathbb{F}(x) = \int_{\mathcal{X}_2} k_2(\cdot,x_2)\,\mathrm{d}\theta_{B_1}(x_2), \tag{30}$$

where

$$\theta_{B_1}(B_2) = \int_{\mathcal{X}_1\times\mathcal{X}_2} \chi_{B_1}(x_1)\chi_{B_2}(x_2)\,\mathrm{d}\mathbb{F}(x), \quad B_2 \in \mathcal{B}(\mathcal{X}_2).$$

Using the characteristic property of k_2, it follows from (30) that $\theta_{B_1} = 0$ for $\forall B_1 \in \mathcal{B}(\mathcal{X}_1)$, i.e.,

$$0 = \theta_{B_1}(B_2) = \mathbb{F}(B_1 \times B_2), \quad \forall B_1 \in \mathcal{B}(\mathcal{X}_1), \forall B_2 \in \mathcal{B}(\mathcal{X}_2)$$

provided that $\theta_{B_1}(\mathcal{X}_2) = 0$ and $|\theta_{B_1}|(\mathcal{X}_2) < \infty$. Both these conditions hold:

$$\theta_{B_1}(\mathcal{X}_2) = \int_{\mathcal{X}_1\times\mathcal{X}_2} \chi_{B_1}(x_1)\,\mathrm{d}\mathbb{F}(x) = \int_{\mathcal{X}_1} \chi_{B_1}(x_1)\,\mathrm{d}[\mathbb{P}_1-\mathbb{P}_1](x_1) = 0;$$
$$|\theta_{B_1}|(\mathcal{X}_2) \leq \int_{\mathcal{X}_1\times\mathcal{X}_2} \mathrm{d}[\mathbb{P}+\mathbb{P}_1\otimes\mathbb{P}_2](x) = 2.$$

(*ii*) Assume w.l.o.g. that k_1 is not characteristic. This means there exists $\mathbb{P}_1 \neq \mathbb{P}_1' \in \mathcal{M}_1^+(\mathcal{X}_1)$ such that $\mu_{k_1}(\mathbb{P}_1) = \mu_{k_1}(\mathbb{P}_1')$. Our goal is to construct an $\mathbb{F} \in \mathcal{M}_1^+\left(\times_{m=1}^M \mathcal{X}_m\right)$ such that

$$\mu_{\otimes_{m=1}^M k_m}\left(\mathbb{F} - \otimes_{m=1}^M \mathbb{F}_m\right) = \int_{\times_{m=1}^M} \otimes_{m=1}^M k_m(\cdot,x_m)\mathrm{d}\left[\mathbb{F} - \otimes_{m=1}^M \mathbb{F}_m\right] = 0, \text{ but } \mathbb{F} \neq \otimes_{m=1}^M \mathbb{F}_m.$$

Define $\mathbb{I} := \mathbb{F} - \otimes_{m=1}^M \mathbb{F}_m \in \mathcal{I}$. In other words we want to get a witness $\mathbb{I} \in \mathcal{I}$ proving that $\otimes_{m=1}^M k_m$ is not $\mathcal{I}$-characteristic. Let us take $z \neq z' \in \mathcal{X}_2$, which is possible since $|\mathcal{X}_2| \geq 2$. Let us define $\mathbb{F}$ as[6]

$$\mathbb{F} = \frac{\mathbb{P}_1\otimes\delta_z\otimes(\otimes_{m=3}^M\mathbb{Q}_m) + \mathbb{P}_1'\otimes\delta_{z'}\otimes(\otimes_{m=3}^M\mathbb{Q}_m)}{2} \in \mathcal{M}_1^+\left(\times_{m=1}^M\mathcal{X}_m\right).$$

It is easy to verify that

$$\mathbb{F}_1 = \frac{\mathbb{P}_1+\mathbb{P}_1'}{2},\ \mathbb{F}_2 = \frac{\delta_z+\delta_{z'}}{2} \text{ and } \mathbb{F}_m = \mathbb{Q}_m \quad (m = 3,\ldots,M),$$

where $\mathbb{Q}_3,\ldots,\mathbb{Q}_M$ are arbitrary probability measures on $\mathcal{X}_3,\ldots,\mathcal{X}_M$, respectively. First we check that $\mathbb{I} \neq 0$. Indeed it is the case since

[6]The $\mathbb{F}$ construction specializes to that of Lyons [2013, Proposition 3.15] in the $M = 2$ case; Lyons used it for distance covariances, which is known to be equivalent to HSIC [Sejdinovic et al., 2013].

- $z \neq z'$ and $\mathcal{X}_2$ is a Hausdorff space, there exists $B_2 \in \mathcal{B}(\mathcal{X}_2)$ such that $z \in B_2$, $z' \notin B_2$.
- $\mathbb{P}_1 \neq \mathbb{P}_1'$, $\mathbb{P}_1(B_1) \neq \mathbb{P}_1'(B_1)$ for some $B_1 \in \mathcal{B}(\mathcal{X}_1)$.

Let $S = B_1 \times B_2 \times \left(\times_{m=3}^{M} \mathcal{X}_m\right)$, and compare its measure under $\mathbb{F}$ and $\otimes_{m=1}^{M} \mathbb{F}_m$:

$$\begin{aligned}
\mathbb{F}(S) &= \frac{\mathbb{P}_1(B_1) \overbrace{\delta_z(B_2)}^{=1\,(z\in B_2)} \prod_{m=3}^{M} \overbrace{\mathbb{Q}_m(\mathcal{X}_m)}^{=1} + \mathbb{P}_1'(B_1) \overbrace{\delta_{z'}(B_2)}^{=0\,(z'\notin B_2)} \prod_{m=3}^{M} \overbrace{\mathbb{Q}_m(\mathcal{X}_m)}^{=1}}{2} \\
&= \frac{\mathbb{P}_1(B_1)}{2}, \\
\left(\otimes_{m=1}^{M} \mathbb{F}_m\right)(S) &= \prod_{m=1}^{M} \mathbb{F}_m(B_m) = \frac{\mathbb{P}_1(B_1) + \mathbb{P}_1'(B_1)}{2} \frac{\overbrace{\delta_z(B_2)}^{=1} + \overbrace{\delta_{z'}(B_2)}^{=0}}{2} \prod_{m=3}^{M} \overbrace{\mathbb{Q}_m(\mathcal{X}_m)}^{=1} \\
&= \frac{\mathbb{P}_1(B_1) + \mathbb{P}_1'(B_1)}{4} \neq \frac{\mathbb{P}_1(B_1)}{2},
\end{aligned}$$

where the last equality holds since $\mathbb{P}_1(B_1) \neq \mathbb{P}_1'(B_1)$. This shows that $\mathbb{I} = \mathbb{F} - \otimes_{m=1}^{M} \mathbb{F}_m \neq 0$ since $\mathbb{I}(S) \neq 0$.

Next we prove that $\mu_{\otimes_{m=1}^{M} k_m}\left(\mathbb{F} - \otimes_{m=1}^{M} \mathbb{F}_m\right) = 0$. Indeed,

$$\begin{aligned}
\mu_{\otimes_{m=1}^{M} k_m}(\mathbb{I}) &= \int_{\times_{m=1}^{M} \mathcal{X}_m} \otimes_{m=1}^{M} k_m(\cdot, x_m)\, \mathrm{d}\left[\mathbb{F} - \otimes_{m=1}^{M} \mathbb{F}_m\right](x_1, \ldots, x_M) \\
&= \int_{\times_{m=1}^{M} \mathcal{X}_m} \otimes_{m=1}^{M} k_m(\cdot, x_m)\, \mathrm{d}\left(\left[\frac{\mathbb{P}_1 \otimes \delta_z + \mathbb{P}_1' \otimes \delta_{z'}}{2} - \frac{\mathbb{P}_1 + \mathbb{P}_1'}{2} \otimes \frac{\delta_z + \delta_{z'}}{2}\right]\right. \\
&\qquad \left. \otimes \left(\otimes_{m=3}^{M} \mathbb{Q}_m\right)\right)(x_1, \ldots, x_M) \\
&= \int_{\times_{m=1}^{M} \mathcal{X}_m} \otimes_{m=1}^{M} k_m(\cdot, x_m)\, \mathrm{d}\left(\left[\frac{\mathbb{P}_1(x_1) \otimes \delta_z(x_2) + \mathbb{P}_1'(x_1) \otimes \delta_{z'}(x_2)}{2}\right.\right. \\
&\qquad - \frac{\mathbb{P}_1(x_1) \otimes \delta_z(x_2) + \mathbb{P}_1(x_1) \otimes \delta_{z'}(x_2)}{4} \\
&\qquad \left.\left. - \frac{\mathbb{P}_1'(x_1) \otimes \delta_z(x_2) + \mathbb{P}_1'(x_1) \otimes \delta_{z'}(x_2)}{4}\right] \otimes \left(\otimes_{m=3}^{M} \mathbb{Q}_m(x_m)\right)\right) \\
&\overset{(*)}{=} \left[\frac{\mu_{k_1}(\mathbb{P}_1) \otimes k_2(\cdot, z) + \mu_{k_1}(\mathbb{P}_1') \otimes k_2(\cdot, z')}{2}\right. \\
&\qquad - \frac{\mu_{k_1}(\mathbb{P}_1) \otimes k_2(\cdot, z) + \mu_{k_1}(\mathbb{P}_1) \otimes k_2(\cdot, z')}{4} \\
&\qquad \left. - \frac{\mu_{k_1}(\mathbb{P}_1') \otimes k_2(\cdot, z) + \mu_{k_1}(\mathbb{P}_1') \otimes k_2(\cdot, z')}{4}\right] \otimes \left[\otimes_{m=3}^{M} \mu_{k_m}\left(\mathbb{Q}_m\right)\right] \\
&= \underbrace{0}_{\in \mathcal{H}_{k_1 \otimes k_2}} \otimes \left[\otimes_{m=3}^{M} \mu_{k_m}\left(\mathbb{Q}_m\right)\right] = 0,
\end{aligned}$$

where we used $\mu_{k_1}(\mathbb{P}_1) = \mu_{k_1}\left(\mathbb{P}_1'\right)$ in $(*)$.

5.5 Proof of Theorem 4

It follows from (6) and Theorem 2 that $(v) \Rightarrow (iii) \Rightarrow (ii) \Leftrightarrow (i)$. It also follows from (6) and Theorem 3(ii) that $(v) \Rightarrow (iv) \Rightarrow (i)$. We now show that $(i) \Rightarrow (v)$ which establishes

the equivalence of (i)–(v). Suppose (i) holds. Then by Bochner's theorem [Wendland, 2005, Theorem 6.6], we have that for all $m \in [M]$,

$$k_m(x_m, y_m) = \int_{\mathbb{R}^{d_m}} e^{-\sqrt{-1}\langle \omega_m, x_m - y_m \rangle} \, \mathrm{d}\Lambda_m(\omega_m), \ x_m, y_m \in \mathbb{R}^{d_m},$$

where $(\Lambda_m)_{m=1}^M$ are finite non-negative Borel measures on $(\mathbb{R}^{d_m})_{m=1}^M$ respectively. This implies

$$\otimes_{m=1}^M k_m(x_m, y_m) = \otimes_{m=1}^M \int_{\mathbb{R}^{d_m}} e^{-\sqrt{-1}\langle \omega_m, x_m - y_m \rangle} \, \mathrm{d}\Lambda_m(\omega_m) = \int_{\mathbb{R}^d} e^{-\sqrt{-1}\langle \omega, x - y \rangle} \, \mathrm{d}\Lambda(\omega),$$

where $x = (x_1, \ldots, x_M) \in \mathbb{R}^d$, $y = (y_1, \ldots, y_M) \in \mathbb{R}^d$, $\omega = (\omega_1, \ldots, \omega_M) \in \mathbb{R}^d$, $d = \sum_{m=1}^M d_m$ and $\Lambda := \otimes_{m=1}^M \Lambda_m$. Sriperumbudur et al. [2010, Theorem 9] showed that k_m is characteristic iff $\mathrm{supp}\,(\Lambda_m) = \mathbb{R}^{d_m}$, where $\mathrm{supp}(\cdot)$ denotes the support of its argument. Since $\mathrm{supp}(\Lambda) = \mathrm{supp}\left(\otimes_{m=1}^M \Lambda_m\right) = \times_{m=1}^M \mathrm{supp}\,(\Lambda_m) = \times_{m=1}^M \mathbb{R}^{d_m} = \mathbb{R}^d$, it follows that $\otimes_{m=1}^M k_m$ is characteristic.

5.6 Proof of Theorem 5

The c_0-kernel property of k_m-s $(m = 1, \ldots, M)$ implies that of $\otimes_{m=1}^M k_m$. Moreover, $\mathcal{X}_m$-s are LCP spaces, hence $\times_{m=1}^M \mathcal{X}_m$ is also LCP.

$(\Leftarrow)$ Assume that $\otimes_{m=1}^M k_m$ is c_0-universal. Since $\otimes_{m=1}^M \mathcal{M}_b\left(\mathcal{X}_m\right) \subseteq \mathcal{M}_b\left(\times_{m=1}^M \mathcal{X}_m\right)$, we have that for all $\mathbb{F} = \otimes_{m=1}^M \mathbb{F}_m \in \otimes_{m=1}^M \mathcal{M}_b(\mathcal{X}_m) \backslash \{0\}$,

$$\begin{aligned} 0 &< \int_{\times_{m=1}^M \mathcal{X}_m} \int_{\times_{m=1}^M \mathcal{X}_m} \underbrace{\left(\otimes_{m=1}^M k_m\right)(x, x')}_{\prod_{m=1}^M k_m(x_m, x'_m)} \mathrm{d}\mathbb{F}(x) \, \mathrm{d}\mathbb{F}(x') \\ &= \prod_{m=1}^M \int_{\mathcal{X}_m \times \mathcal{X}_m} k_m(x_m, x'_m) \, \mathrm{d}\mathbb{F}_m(x_m) \, \mathrm{d}\mathbb{F}_m\left(x'_m\right), \end{aligned}$$

where $x = (x_1, \ldots, x_M)$ and $x' = (x'_1, \ldots, x'_M)$. The above inequality implies

$$\int_{\mathcal{X}_m \times \mathcal{X}_m} k_m(x_m, x'_m) \, \mathrm{d}\mathbb{F}_m(x_m) \, \mathrm{d}\mathbb{F}_m\left(x'_m\right) > 0, \ \forall\, m \in [M].$$

Since $\mathbb{F} \in \otimes_{m=1}^M \mathcal{M}_b\left(\mathcal{X}_m\right) \backslash \{0\}$ iff $\mathbb{F}_m \in \mathcal{M}_b(\mathcal{X}_m) \backslash \{0\}$ for all $m \in [M]$, the result follows.

$(\Rightarrow)$ Assume that k_m-s are c_0-universal. By the note above $\otimes_{m=1}^M k_m$ is c_0-kernel; its c_0-universality is equivalent to the injectivity of $\mu = \mu_{\otimes_{m=1}^M k_m}$ on $\mathcal{M}_b\left(\times_{m=1}^M \mathcal{X}_m\right)$. In other words, we want to prove that $\mu(\mathbb{F}) = 0$ implies $\mathbb{F} = 0$, where $\mathbb{F} \in \mathcal{M}_b\left(\times_{m=1}^M \mathcal{X}_m\right)$. We will use the shorthand $\mathcal{H}_m = \mathcal{H}_{k_m}$ below.

Suppose there exists $\mathbb{F} \in \mathcal{M}_b\left(\times_{m=1}^M \mathcal{X}_m\right)$ such that

$$\mu_{\mathbb{F}} = \int_{\times_{m=1}^M \mathcal{X}_m} \underbrace{\left(\otimes_{m=1}^M k_m\right)(\cdot, x)}_{\otimes_{m=1}^M k_m(\cdot, x_m)} \mathrm{d}\mathbb{F}(x) = 0 \ (\in \otimes_{m=1}^M \mathcal{H}_m). \tag{31}$$

In order to get $\mathbb{F} = 0$, it is sufficient to prove that

$$\mathbb{F}\left(\times_{m=1}^M B_m\right) = 0, \quad \forall B_m \in \mathcal{B}(\mathcal{X}_m), m \in [M].$$

We will prove by induction that for $m = 0, \ldots, M$

$$\begin{aligned}\left(\otimes_{j=m+1}^{M}\mathcal{H}_j \ni\right) 0 &= \int_{\times_{j=1}^{M}\mathcal{X}_j} \prod_{j=1}^{m} \chi_{B_j}(x_j) \otimes_{j=m+1}^{M} k_j(\cdot, x_j)\, \mathrm{d}\mathbb{F}(x) \\ &=: o(B_1, \ldots, B_m, k_{m+1}, \ldots, k_M), \forall B_j \in \mathcal{B}(\mathcal{X}_j),\, j \in [m], \end{aligned} \tag{32}$$

which

- $(*)$ reduces to (31) when $m = 0$ by definining $\prod_{j=1}^{0} \chi_{B_j}(x_j) := 1$;
- $(\dagger)$ for $m = M$, $\otimes_{m=M+1}^{M}\mathcal{H}_m$ is defined to be equal to $\mathbb{R}$ and $\otimes_{m=M+1}^{M} k_m(\cdot, x_m) := 1$, in which case $o(B_1, \ldots, B_M) = \mathbb{F}\left(\times_{j=1}^{M} B_j\right) = 0 \Rightarrow \mathbb{F} = 0$, the result we want to prove.

From the above, it is clear that (32) holds for $m = 0$. Assuming (32) holds for some m, we now prove that it holds for $m+1$. To this end, it follows from (32) that $\forall h_{m+2} \in \mathcal{H}_{m+2}, \ldots, \forall h_M \in \mathcal{H}_M$,

$$\begin{aligned}\left(\mathcal{H}_{m+1} \ni\right) 0 &= o(B_1, \ldots, B_m, k_{m+1}, \ldots, k_M)\,(h_{m+2}, \ldots, h_M) \\ &= \left[\int_{\times_{j=1}^{M}\mathcal{X}_j} \left(\prod_{j=1}^{m} \chi_{B_j}(x_j)\right) \otimes_{j=m+1}^{M} k_j(\cdot, x_j)\, \mathrm{d}\mathbb{F}(x)\right](h_{m+2}, \ldots, h_M) \\ &= \int_{\times_{j=1}^{M}\mathcal{X}_j} k_{m+1}(\cdot, x_{m+1}) \prod_{j=1}^{m} \chi_{B_j}(x_j) \prod_{j=m+2}^{M} h_j(x_j)\, \mathrm{d}\mathbb{F}(x) \\ &= \int_{\mathcal{X}_{m+1}} k_{m+1}(\cdot, x_{m+1})\, \mathrm{d}\nu(x_{m+1}),\end{aligned}$$

where

$$\begin{aligned}\nu(B) &:= \nu_{B_1,\ldots,B_m,h_{m+2},\ldots,h_M}(B) \\ &= \int_{\times_{j=1}^{M}\mathcal{X}_j} \left[\prod_{j=1}^{m} \chi_{B_j}(x_j)\right] \chi_B(x_{m+1}) \left[\prod_{j=m+2}^{M} h_j(x_j)\right] \mathrm{d}\mathbb{F}(x),\ B \in \mathcal{B}(\mathcal{X}_{m+1}).\end{aligned}$$

By the c_0-universality of k_{m+1},

$$\nu = 0 \text{ for } \forall h_{m+2} \in \mathcal{H}_{m+2}, \ldots, \forall h_M \in \mathcal{H}_M \tag{33}$$

provided that $\nu \in \mathcal{M}_b(\mathcal{X}_{m+1})$, in other words if $|\nu|(\mathcal{X}_{m+1}) < \infty$. This condition is met:

$$\begin{aligned}|\nu|(\mathcal{X}_{m+1}) &\leq \int_{\times_{j=1}^{M}\mathcal{X}_j} \prod_{j=m+2}^{M} \underbrace{\left|\langle h_j, k_j(\cdot, x_j)\rangle_{\mathcal{H}_j}\right|}_{\leq \|h_j\|_{\mathcal{H}_j}\sqrt{k_j(x_j,x_j)}} \mathrm{d}|\mathbb{F}|(x) \\ &\leq |\mathbb{F}|\left(\times_{m=1}^{M}\mathcal{X}_m\right) \prod_{j=m+2}^{M} \|h_j\|_{\mathcal{H}_j} \sup_{x\in\mathcal{X}_j, x'\in\mathcal{X}_j} \sqrt{k_j(x, x')} < \infty,\end{aligned} \tag{34}$$

where we used the boundedness of k_m-s in the last inequality. (33) implies that for $\forall B_1 \in \mathcal{B}(\mathcal{X}_1), \ldots, \forall B_{m+1} \in \mathcal{B}(\mathcal{X}_{m+1})$ and $\forall h_{m+2} \in \mathcal{H}_{m+2}, \ldots, \forall h_M \in \mathcal{H}_M$

$$\begin{aligned} 0 = \nu(B_{m+1}) &= \int_{\times_{j=1}^M \mathcal{X}_j} \left[\prod_{j=1}^{m+1} \chi_{B_j}(x_j)\right] \left[\prod_{j=m+2}^{M} h_j(x_j)\right] \mathrm{d}\mathbb{F}(x) \\ &= \left\langle \otimes_{j=m+2}^M h_j, \int_{\times_{j=1}^M \mathcal{X}_j} \left[\prod_{j=1}^{m+1} \chi_{B_j}(x_j)\right] \otimes_{j=m+2}^M k_j(\cdot, x_j)\, \mathrm{d}\mathbb{F}(x) \right\rangle_{\otimes_{j=m+2}^M \mathcal{H}_j}, \end{aligned}$$

and therefore

$$\begin{aligned} o(B_1, \ldots, B_{m+1}, k_{m+2}, \ldots, k_M) &= \int_{\times_{j=1}^M \mathcal{X}_j} \left[\prod_{j=1}^{m+1} \chi_{B_j}(x_j)\right] \otimes_{j=m+2}^M k(\cdot, x_j)\, \mathrm{d}\mathbb{F}(x) \\ &= 0 \left(\in \otimes_{j=m+2}^M \mathcal{H}_j\right) \end{aligned}$$

for $\forall B_1 \in \mathcal{B}(\mathcal{X}_1), \ldots, \forall B_{m+1} \in \mathcal{B}(\mathcal{X}_{m+1})$, i.e., (32) holds for $m+1$. Therefore, by induction, (32) holds for $m = M$ and the result follows from (†). To justify the convention in (†), consider the case of $m = M-1$ in which case (32) can be written as

$$\int_{\mathcal{X}_M} k_M(\cdot, x_M)\, \mathrm{d}\nu(x_M) = 0,$$

where

$$\nu(B) = \int_{\times_{j=1}^M \mathcal{X}_j} \left[\prod_{j=1}^{M-1} \chi_{B_j}(x_j)\right] \chi_B(x_M)\, \mathrm{d}\mathbb{F}(x),\ B \in \mathcal{B}(\mathcal{X}_M).$$

Then by the c_0-universal property of k_M, since

$$|\nu|(\mathcal{X}_M) \leq \int_{\times_{j=1}^M \mathcal{X}_j} 1\, \mathrm{d}|\mathbb{F}|(x) = |\mathbb{F}|\left(\times_{j=1}^M \mathcal{X}_j\right) < \infty$$

we obtain

$$\int_{\times_{j=1}^M \mathcal{X}_j} \prod_{j=1}^{M} \chi_{B_j}(x_j)\, \mathrm{d}\mathbb{F}(x) = \mathbb{F}\left(\times_{j=1}^M B_j\right) = 0, \forall B_1 \in \mathcal{B}(\mathcal{X}_1), \ldots, \forall B_M \in \mathcal{B}(\mathcal{X}_M).$$

Acknowledgments

A part of the work was carried out while BKS was visiting ZSz at CMAP, École Polytechnique. BKS is supported by NSF-DMS-1713011 and also thanks CMAP for their generous support.

A Analytical Solution to (21)–(26) in Example 2

The solution of (21)–(26) takes the form

$$
\begin{aligned}
p_{1,1,1} = -\frac{\begin{array}{l} z_2+z_1+z_4+z_5-3z_2z_1-4z_2z_4-4z_1z_4-z_2z_3-2z_2z_0-2z_1z_3-3z_2z_5 \\ -2z_4z_3-z_1z_0-3z_1z_5-2z_4z_0-4z_4z_5-z_3z_0-z_3z_5-z_0z_5+2z_2z_1^2+2z_2^2z_1 \\ +4z_2z_4^2+2z_2^2z_4+4z_1z_4^2+2z_1^2z_4+2z_2^2z_0+2z_1^2z_3+2z_2z_5^2+2z_2^2z_5+2z_4^2z_3 \\ +2z_1z_5^2+2z_1^2z_5+2z_4^2z_0+2z_4z_5^2+4z_4^2z_5-z_2^2-z_1^2-3z_4^2+2z_4^3-z_5^2 \\ +6z_2z_1z_4+2z_2z_1z_3+2z_2z_4z_3+2z_2z_1z_0+4z_2z_1z_5+4z_2z_4z_0+4z_1z_4z_3 \\ +6z_2z_4z_5+2z_1z_4z_0+6z_1z_4z_5+2z_2z_3z_0+2z_2z_3z_5+2z_1z_3z_0+2z_2z_0z_5 \\ +2z_1z_3z_5+2z_4z_3z_0+2z_4z_3z_5+2z_1z_0z_5+2z_4z_0z_5 \end{array}}{\begin{array}{l} 2z_2z_1-z_1-2z_4-z_3-z_0-2z_5-z_2+2z_2z_4+2z_1z_4+2z_2z_0+2z_1z_3+2z_2z_5 \\ +2z_4z_3+2z_1z_5+2z_4z_0+4z_4z_5+2z_3z_0+2z_3z_5+2z_0z_5+2z_4^2+2z_5^2 \end{array}}, \\
p_{1,1,2} &= z_2, \\
p_{1,2,1} &= z_1, \\
p_{1,2,2} &= z_4, \\
p_{2,1,1} = -\frac{\begin{array}{l} z_4+z_3+z_0+z_5-z_2z_1-z_2z_4-z_1z_4-z_2z_3-2z_2z_0-2z_1z_3-2z_2z_5 \\ -3z_4z_3-z_1z_0-2z_1z_5-3z_4z_0-4z_4z_5-3z_3z_0-4z_3z_5-4z_0z_5+2z_2z_0^2 \\ +2z_1z_3^2+2z_2z_5^2+2z_4z_3^2+2z_4^2z_3+2z_1z_5^2+2z_4z_0^2+2z_4^2z_0+4z_4z_5^2+2z_4^2z_5 \\ +2z_3z_0^2+2z_3^2z_0+4z_3z_5^2+2z_3^2z_5+4z_0z_5^2+2z_0^2z_5-z_4^2-z_3^2-z_0^2-3z_5^2 \\ +2z_5^3+2z_2z_1z_3+2z_2z_4z_3+2z_2z_1z_0+2z_2z_1z_5+2z_2z_4z_0+2z_1z_4z_3 \\ +2z_2z_4z_5+2z_1z_4z_0+2z_1z_4z_5+2z_2z_3z_0+2z_2z_3z_5+2z_1z_3z_0+4z_2z_0z_5 \\ +4z_1z_3z_5+4z_4z_3z_0+6z_4z_3z_5+2z_1z_0z_5+6z_4z_0z_5+6z_3z_0z_5 \end{array}}{\begin{array}{l} 2z_2z_1-z_1-2z_4-z_3-z_0-2z_5-z_2+2z_2z_4+2z_1z_4+2z_2z_0+2z_1z_3+2z_2z_5 \\ +2z_4z_3+2z_1z_5+2z_4z_0+4z_4z_5+2z_3z_0+2z_3z_5+2z_0z_5+2z_4^2+2z_5^2 \end{array}}, \\
p_{2,1,2} &= z_3, \\
p_{2,2,1} &= z_0, \\
p_{2,2,2} &= z_5,
\end{aligned}
$$

form, where $\mathbf{z} = (z_0, z_1, \ldots, z_5) \in \mathbb{R}^6$ satisfies

$$
\begin{aligned}
0 \leq\ & (2z_0z_2-z_1-z_2-z_3-2z_4-2z_5-z_0+2z_0z_3+2z_1z_2+2z_0z_4+2z_1z_3+2z_0z_5 \\
& +2z_1z_4+2z_1z_5+2z_2z_4+2z_2z_5+2z_3z_4+2z_3z_5+4z_4z_5+2z_4^2+2z_5^2) \times \\
& (z_0z_3-z_3-z_4-z_5-z_0z_1-z_0-z_1z_2+z_0z_5-2z_1z_4-z_2z_3-z_1z_5-2z_2z_4-z_2z_5 \\
& +z_3z_5+2z_0z_2^2+2z_1z_2^2+2z_1^2z_2+2z_0z_4^2+2z_1^2z_3+4z_1z_4^2+2z_1^2z_4+2z_1z_5^2+4z_2z_4^2 \\
& +2z_1^2z_5+2z_2^2z_4+2z_2z_5^2+2z_3z_4^2+2z_2^2z_5+2z_4z_5^2+4z_4^2z_5-z_1^2-z_2^2-z_4^2+2z_4^3+z_5^2 \\
& +2z_0z_1z_2+2z_0z_1z_3+2z_0z_1z_4+2z_0z_2z_3+2z_0z_1z_5+4z_0z_2z_4+2z_1z_2z_3+2z_0z_2z_5 \\
& +2z_0z_3z_4+6z_1z_2z_4+4z_1z_2z_5+4z_1z_3z_4+2z_0z_4z_5+2z_1z_3z_5+2z_2z_3z_4+6z_1z_4z_5 \\
& +2z_2z_3z_5+6z_2z_4z_5+2z_3z_4z_5),
\end{aligned}
$$

$$
\begin{aligned}
0 \le\ & (2z_0z_2 - z_1 - z_2 - z_3 - 2z_4 - 2z_5 - z_0 + 2z_0z_3 + 2z_1z_2 + 2z_0z_4 + 2z_1z_3 + 2z_0z_5 \\
& +2z_1z_4 + 2z_1z_5 + 2z_2z_4 + 2z_2z_5 + 2z_3z_4 + 2z_3z_5 + 4z_4z_5 + 2z_4^2 + 2z_5^2) \times \\
& (z_1z_2 - z_2 - z_4 - z_5 - z_0z_1 - z_0z_3 - z_1 - z_0z_4 - 2z_0z_5 + z_1z_4 - z_2z_3 + z_2z_4 \\
& -z_3z_4 - 2z_3z_5 + 2z_0^2z_2 + 2z_0z_3^2 + 2z_0^2z_3 + 2z_0z_4^2 + 2z_1z_3^2 + 2z_0^2z_4 + 4z_0z_5^2 + 2z_0^2z_5 \\
& +2z_1z_5^2 + 2z_2z_5^2 + 2z_3z_4^2 + 2z_3^2z_4 + 4z_3z_5^2 + 2z_3^2z_5 + 4z_4z_5^2 + 2z_4^2z_5 - z_0^2 - z_3^2 + z_4^2 \\
& -z_5^2 + 2z_5^3 + 2z_0z_1z_2 + 2z_0z_1z_3 + 2z_0z_1z_4 + 2z_0z_2z_3 + 2z_0z_1z_5 + 2z_0z_2z_4 + 2z_1z_2z_3 \\
& +4z_0z_2z_5 + 4z_0z_3z_4 + 6z_0z_3z_5 + 2z_1z_2z_5 + 2z_1z_3z_4 + 6z_0z_4z_5 + 4z_1z_3z_5 + 2z_2z_3z_4 \\
& +2z_1z_4z_5 + 2z_2z_3z_5 + 2z_2z_4z_5 + 6z_3z_4z_5)\,,
\end{aligned}
$$

$$
\begin{aligned}
& 2z_0z_2 + 2z_0z_3 + 2z_1z_2 + 2z_0z_4 + 2z_1z_3 + 2z_0z_5 + 2z_1z_4 + 2z_1z_5 + 2z_2z_4 + 2z_2z_5 \\
& + 2z_3z_4 + 2z_3z_5 + 4z_4z_5 + 2z_4^2 + 2z_5^2 \neq z_0 + z_1 + z_2 + z_3 + 2z_4 + 2z_5,
\end{aligned}
$$

$$
\begin{aligned}
& (2z_0z_2 - z_1 - z_2 - z_3 - 2z_4 - 2z_5 - z_0 + 2z_0z_3 + 2z_1z_2 + 2z_0z_4 + 2z_1z_3 + 2z_0z_5 \\
& +2z_1z_4 + 2z_1z_5 + 2z_2z_4 + 2z_2z_5 + 2z_3z_4 + 2z_3z_5 + 4z_4z_5 + 2z_4^2 + 2z_5^2) \times \\
& (z_1 + z_2 + z_4 + z_5 - z_0z_1 - 2z_0z_2 - z_0z_3 - 3z_1z_2 - 2z_0z_4 - 2z_1z_3 - z_0z_5 - 4z_1z_4 \\
& -z_2z_3 - 3z_1z_5 - 4z_2z_4 - 3z_2z_5 - 2z_3z_4 - z_3z_5 - 4z_4z_5 + 2z_0z_2^2 + 2z_1z_2^2 + 2z_1^2z_2 \\
& +2z_0z_4^2 + 2z_1^2z_3 + 4z_1z_4^2 + 2z_1^2z_4 + 2z_1z_5^2 + 4z_2z_4^2 + 2z_1^2z_5 + 2z_2^2z_4 + 2z_2z_5^2 \\
& +2z_3z_4^2 + 2z_2^2z_5 + 2z_4z_5^2 + 4z_4^2z_5 - z_1^2 - z_2^2 - 3z_4^2 + 2z_4^3 - z_5^2 + 2z_0z_1z_2 \\
& +2z_0z_1z_3 + 2z_0z_1z_4 + 2z_0z_2z_3 + 2z_0z_1z_5 + 4z_0z_2z_4 + 2z_1z_2z_3 + 2z_0z_2z_5 \\
& +2z_0z_3z_4 + 6z_1z_2z_4 + 4z_1z_2z_5 + 4z_1z_3z_4 + 2z_0z_4z_5 + 2z_1z_3z_5 + 2z_2z_3z_4 + 6z_1z_4z_5 \\
& + 2z_2z_3z_5 + 6z_2z_4z_5 + 2z_3z_4z_5) \le 0,
\end{aligned}
$$

$$
\begin{aligned}
& (2z_0z_2 - z_1 - z_2 - z_3 - 2z_4 - 2z_5 - z_0 + 2z_0z_3 + 2z_1z_2 + 2z_0z_4 + 2z_1z_3 + 2z_0z_5 \\
& +2z_1z_4 + 2z_1z_5 + 2z_2z_4 + 2z_2z_5 + 2z_3z_4 + 2z_3z_5 + 4z_4z_5 + 2z_4^2 + 2z_5^2) \times \\
& (z_0 + z_3 + z_4 + z_5 - z_0z_1 - 2z_0z_2 - 3z_0z_3 - z_1z_2 - 3z_0z_4 - 2z_1z_3 - 4z_0z_5 \\
& -z_1z_4 - z_2z_3 - 2z_1z_5 - z_2z_4 - 2z_2z_5 - 3z_3z_4 - 4z_3z_5 - 4z_4z_5 + 2z_0^2z_2 \\
& +2z_0z_3^2 + 2z_0^2z_3 + 2z_0z_4^2 + 2z_1z_3^2 + 2z_0^2z_4 + 4z_0z_5^2 + 2z_0^2z_5 + 2z_1z_5^2 + 2z_2z_5^2 \\
& +2z_3z_4^2 + 2z_3^2z_4 + 4z_3z_5^2 + 2z_3^2z_5 + 4z_4z_5^2 + 2z_4^2z_5 - z_0^2 - z_3^2 - z_4^2 - 3z_5^2 + 2z_5^3 \\
& +2z_0z_1z_2 + 2z_0z_1z_3 + 2z_0z_1z_4 + 2z_0z_2z_3 + 2z_0z_1z_5 + 2z_0z_2z_4 + 2z_1z_2z_3 \\
& +4z_0z_2z_5 + 4z_0z_3z_4 + 6z_0z_3z_5 + 2z_1z_2z_5 + 2z_1z_3z_4 + 6z_0z_4z_5 + 4z_1z_3z_5 \\
& +2z_2z_3z_4 + 2z_1z_4z_5 + 2z_2z_3z_5 + 2z_2z_4z_5 + 6z_3z_4z_5) \le 0,
\end{aligned}
$$

and $\quad 0 \le z_0, z_1, z_2, z_3, z_4, z_5 \le 1.$

The above analytic solution to (21)–(26) is obtained by symbolic math programming in MATLAB.

References

Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer, 2004.

Claudio Carmeli, Ernesto De Vito, Alessandro Toigo, and Veronica Umanitá. Vector valued reproducing kernel Hilbert spaces and universality. *Analysis and Applications*, 8:19–61, 2010.

Kenji Fukumizu, Arthur Gretton, Xiaohai Sun, and Bernhard Schölkopf. Kernel measures of conditional dependence. In *Neural Information Processing Systems (NIPS)*, pages 498–496, 2008.

Kenji Fukumizu, Francis Bach, and Michael Jordan. Kernel dimension reduction in regression. *The Annals of Statistics*, 37(4):1871–1905, 2009.

Kenji Fukumizu, Le Song, and Arthur Gretton. Kernel Bayes' rule: Bayesian inference with positive definite kernels. *Journal of Machine Learning Research*, 14:3753–3783, 2013.

Arthur Gretton. A simpler condition for consistency of a kernel independence test. Technical report, University College London, 2015. (`http://arxiv.org/abs/1501.06103`).

Arthur Gretton, Olivier Bousquet, Alexander Smola, and Bernhard Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic Learning Theory (ALT)*, pages 63–78, 2005.

Arthur Gretton, Kenji Fukumizu, Choon Hui Teo, Le Song, Bernhard Schölkopf, and Alexander J. Smola. A kernel statistical test of independence. In *Neural Information Processing Systems (NIPS)*, pages 585–592, 2008.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Genki Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. Persistence weighted Gaussian kernel for topological data analysis. In *International Conference on Machine Learning (ICML)*, pages 2004–2013, 2016.

James Robert Lloyd, David Duvenaud, Roger Grosse, Joshua B. Tenenbaum, and Zoubin Ghahramani. Automatic construction and natural-language description of nonparametric regression models. In *AAAI Conference on Artificial Intelligence*, pages 1242–1250, 2014.

Russell Lyons. Distance covariance in metric spaces. *The Annals of Probability*, 41:3284–3305, 2013.

Charles A. Micchelli, Yuesheng Xu, and Haizhang Zhang. Universal kernels. *Journal of Machine Learning Research*, 7:2651–2667, 2006.

Joris M. Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. Distinguishing cause from effect using observational data: Methods and benchmarks. *Journal of Machine Learning Research*, 17:1–102, 2016.

Krikamol Muandet, Kenji Fukumizu, Francesco Dinuzzo, and Bernhard Schölkopf. Learning from distributions via support measure machines. In *Neural Information Processing Systems (NIPS)*, pages 10–18, 2011.

Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1-2):1–141, 2017.

Mijung Park, Wittawat Jitkrittum, and Dino Sejdinovic. K2-ABC: Approximate Bayesian computation with kernel embeddings. In *International Conference on Artificial Intelligence and Statistics (AISTATS; PMLR)*, volume 51, pages 51:398–407, 2016.

Niklas Pfister, Peter Bühlmann, Bernhard Schölkopf, and Jonas Peters. Kernel-based tests for joint independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2017. ISSN 1467-9868.

Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.

Bernhard Schölkopf, Krikamol Muandet, Kenji Fukumizu, Stefan Harmeling, and Jonas Peters. Computing functions of random variables via reproducing kernel Hilbert space representations. *Statistics and Computing*, 25(4):755–766, 2015.

Dino Sejdinovic, Bharath K. Sriperumbudur, Arthur Gretton, and Kenji Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Annals of Statistics*, 41:2263–2291, 2013.

Carl-Johann Simon-Gabriel and Bernhard Schölkopf. Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions. Technical report, Max Planck Institute for Intelligent Systems, 2016. (`https://arxiv.org/abs/1604.05251`).

Alexander Smola, Arthur Gretton, Le. Song, and Bernhard Schölkopf. A Hilbert space embedding for distributions. In *Algorithmic Learning Theory (ALT)*, pages 13–31, 2007.

Le Song, Arthur Gretton, Danny Bickson, Yucheng Low, and Carlos Guestrin. Kernel belief propagation. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 707–715, 2011.

Le Song, Alexander Smola, Arthur Gretton, Justin Bedo, and Karsten Borgwardt. Feature selection via dependence maximization. *Journal of Machine Learning Research*, 13:1393–1434, 2012.

Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R.G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11:1517–1561, 2010.

Bharath K. Sriperumbudur, Kenji Fukumizu, and Gert R. G. Lanckriet. Universality, characteristic kernels and RKHS embedding of measures. *Journal of Machine Learning Research*, 12:2389–2410, 2011.

Ingo Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of Machine Learning Research*, 6(3):67–93, 2001.

Zoltán Szabó, Bharath Sriperumbudur, Barnabás Póczos, and Arthur Gretton. Learning theory for distribution regression. *Journal of Machine Learning Research*, 17(152):1–40, 2016.

Willem Waegeman, Tapio Pahikkala, Antti Airola, Tapio Salakoski, Michiel Stock, and Bernard De Baets. A kernel-based framework for learning graded relations from data. *IEEE Transactions on Fuzzy Systems*, 20:1090–1101, 2012.

Holger Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2005.

Makoto Yamada, Yuta Umezu, Kenji Fukumizu, and Ichiro Takeuchi. Post selection inference with kernels. Technical report, 2016. (`https://arxiv.org/abs/1610.03725`).

Kun Zhang, Bernhard Schölkopf, Krikamol Muandet, and Zhikun Wang. Domain adaptation under target and conditional shift. *Journal of Machine Learning Research*, 28(3):819–827, 2013.