

# An extension of Mercer theorem to matrix-valued measurable kernels

Ernesto De Vito<sup>a,\*</sup>, Veronica Umanità<sup>b</sup>, Silvia Villa<sup>b</sup>

<sup>a</sup>*DIMA, Università di Genova, Italy & I.N.F.N., Sezione di Genova, Italy*

<sup>b</sup>*DIMA, Università di Genova, Italy*

<sup>c</sup>*Istituto Italiano di Tecnologia, Italy & DIMA, Università di Genova, Italy*

---

## Abstract

We extend the classical Mercer theorem to reproducing kernel Hilbert spaces whose elements are functions from a measurable space  $X$  into  $\mathbb{C}^n$ . Given a finite measure  $\mu$  on  $X$ , we represent the reproducing kernel  $K$  as a convergent series in terms of the eigenfunctions of a suitable compact operator depending on  $K$  and  $\mu$ . Our result holds under the mild assumption that  $K$  is measurable and the associated Hilbert space is separable. Furthermore, we show that  $X$  has a natural second countable topology with respect to which the eigenfunctions are continuous and such that the series representing  $K$  uniformly converges to  $K$  on compact subsets of  $X \times X$ , provided that the support of  $\mu$  is  $X$ .

*Keywords:* Reproducing Kernel Hilbert Spaces, Integral Operators, Eigenvalues, Statistical Learning Theory

*2000 MSC:* 46E22, 47B32, 47B34, 47G10, 47A70, 68T05

---

## 1. Introduction

Reproducing kernel Hilbert spaces (RKHSs) are spaces of functions defined on an arbitrary set  $X$  and taking values into a normed vector space  $Y$  with the property that the evaluation operator at each point is continuous. Usually the output space  $Y$  is simply  $Y = \mathbb{R}$  or  $\mathbb{C}$ , but recently the vector-valued setting is becoming popular, especially in machine learning because

---

\*Corresponding author

*Email addresses:* `devito@dima.unige.it` (Ernesto De Vito),  
`umanita@dima.unige.it` (Veronica Umanità), `villa@dima.unige.it` (Silvia Villa)

of its generality and its good experimental performance in a variety of different domains [1, 2, 3]. The mathematical theory for vector-valued RKHS has been completely worked out in the seminal paper [4], which studies the Hilbert spaces that are continuously embedded into a locally convex topological vector space, see also [5]. If  $Y$  is itself a Hilbert space, the theory can be simplified as shown in [6, 7, 8, 9]. In particular, it remains true that the vector valued RKHSs are completely characterized by the corresponding reproducing kernel, which now takes value in the space of bounded operators on  $Y$ .

The focus of this paper is on Mercer theorem [10]. In the scalar setting, it provides a series representation, called  *Mercer representation* , for the reproducing kernel  $K$  under some suitable hypotheses. In the classical setting,  $X$  is assumed to be a compact metric space and the reproducing kernel  $K$  to be continuous. Hence, for a finite measure  $\mu$  on  $X$  whose support is  $X$ , the integral operator  $L_\mu$  with kernel  $K$  is a compact positive operator on  $L^2(X, \mu)$  and it admits an orthonormal basis  $\{f_i\}_{i \in I}$  of eigenfunctions with non-negative eigenvalues  $\{\sigma_i\}_{i \in I}$  such that each  $f_i$  with  $\sigma_i > 0$  is a continuous function. Mercer theorem states that

$$K(x, t) = \sum_{i \in I} \sigma_i f_i(t) \overline{f_i(x)} \quad \forall x, t \in X, \quad (1)$$

where the series is absolutely and uniformly convergent (see also [11]). In the following we refer to (1) as a Mercer representation of  $K$ .

The kind of representation for the reproducing kernel plays a special role in the applications. For example, since the family  $\{\sqrt{\sigma_i} f_i : \sigma_i > 0\}$  is an orthonormal basis of the corresponding RKHS  $\mathcal{H}_K$ , it provides a *canonical* feature map which relates the spectral properties of  $L_\mu$  and the structure of  $\mathcal{H}_K$ . This characterization has several consequences in the study of learning algorithms, since it allows to prove smoothing properties of kernels and to obtain error estimates, see for example [12, 13] and references therein. In addition, the Mercer representation is an important tool in the theory of stochastic processes [14, 15] and for dimensionality reduction methods, such as kernel PCA [16, 17].

However, in many applications, the “classical hypotheses” of Mercer theorem are not satisfied. For this reason, in the recent years there has been an increasing interest in Mercer representations under relaxed assumptions on the input space  $X$ , on the kernel  $K$  and on  $Y$ . A first group of results concerns scalar kernels. For example, [18] dealt with the case of a  $\sigma$ -compact

metric space  $X$  and a continuous kernel satisfying some natural integrability conditions. When  $X$  is an arbitrary measurable space endowed with a probability measure, and  $K$  is an  $L^2$ -integrable kernel, it is possible to obtain a Mercer representation of the kernel resorting to the spectral properties of the operator  $L_\mu$  [19]. The shortcoming of these results is that the corresponding series converges only almost everywhere. More stringent assumptions on the kernel, such as boundedness, allow to get convergence in  $L^\infty$ , which is still too weak to get a pointwise representation [20]. The preprint [21] contains a more general developments on the subject. In particular, a Mercer representation enjoying pointwise absolute convergence is obtained under less restrictive assumptions on the kernel. More precisely, given a finite Borel measure  $\mu$  on  $X$ , if the RKHS is separable and compactly embedded into  $L^2(X, \mu)$ , a Mercer representation almost everywhere pointwise convergent is recovered. Moreover, the series is pointwise absolutely convergent if and only if the embedding of  $\mathcal{H}_K$  into  $L^2(X, \mu)$  is injective. For vector valued RKHSs, [8] provides an (integral) Mercer representation under the conditions that  $K$  is square-integrable and  $Y$  is a (separable) Hilbert space.

In our paper we extend Mercer theorem under three aspects by assuming that

- i) the input space  $X$  is a measurable space;
- ii) the output space  $Y$  is a finite dimensional vector space;
- iii) the kernel  $K$  is a measurable function and the corresponding RKHS  $\mathcal{H}_K$  is separable.

The results in [8] strongly depend on the fact that the kernel is square-integrable, whereas we only assume that  $K$  is a measurable function. In the cited reference the Mercer decomposition is given with respect to the integral operator of kernel  $K$ , whose spectrum could have a continuous part. To take care of this problem, the proof in [8] is very technical. In our setting, with a suitable choice of the measure, the integral operator is always compact, so that the spectrum has only eigenvalues (up to zero, which does not play any role). Due to this compactness property, our proof is simpler and more straightforward.

Generalizing the ideas in [22, 23], we show that  $X$  has a natural second countable topology making  $K$  a continuous kernel. Moreover, given a finite measure  $\mu$  whose support is  $X$ , we construct a second measure  $\nu$  such that

the integral operator  $L_\nu$  of kernel  $K$  is compact on  $L^2(X, \nu; \mathbb{C}^n)$ . Hence, by using the singular value decomposition, we prove that the Mercer representation (1) holds true, where  $\{f_i\}_{i \in I}$  is any orthonormal basis of eigenfunctions of  $L_\nu$ ,  $\{\sigma_i\}_{i \in I}$  the corresponding family of eigenvalues and the series converges uniformly on compact subsets of  $X \times X$ . If the support of  $\mu$  is a proper subset of  $X$ , representation (1) still holds true provided that  $x, t \in \text{supp } \mu$ . Note that the assumption on  $Y$  can be relaxed allowing  $Y$  to be a separable Hilbert space provided that  $K(x, x)$  is a compact operator for all  $x \in X$ . However, for the sake of clarity we state our results only for finite dimensional output spaces and, by choosing a basis, we can further assume that  $Y = \mathbb{C}^n$ .

To extend our results to an infinite dimensional (separable) Hilbert space  $Y$ , one first identifies  $Y$  with  $\ell_2(\mathbb{N})$  by choosing a basis in  $Y$  and replaces  $\mathbb{C}^n$  with  $\ell_2(\mathbb{N})$  in our formulas. The assumption that for all  $x \in X$  the operator  $K(x, x)$  is compact implies that the integral operator (10) always has a basis of eigenfunctions, see Proposition 4.8 of [8], so that the content and the proof of Theorem 3.4 are the same with the only difference that the indexes  $j, l$  in (12) range in  $\mathbb{N}$ .

The paper is organized as follows. In Section 2 we introduce the notation and we recall some basic facts about vector-valued reproducing kernel Hilbert spaces. Section 3 contains the main results of the paper: given a measurable matrix-valued reproducing kernel  $K$ , Theorem 3.4 gives the Mercer representation of  $K$  and Proposition 3.5 studies the relation between  $K$  and the scalar reproducing kernels associated with the “diagonal blocks” of  $K$ , see (13). The proofs are given in Sections 4 and 5. In the former we prove the Mercer theorem for continuous matrix-valued kernels defined on metric spaces and satisfying a suitable integrability condition. Section 5 is devoted to the proof of Theorem 3.2 and Proposition 3.5. The appendix collects some properties of the associated integral operator.

## 2. Preliminaries and notation

For any integer  $n \geq 1$ , the Euclidean norm and the inner product on  $\mathbb{C}^n$  are denoted by  $\|\cdot\|$  and  $\langle \cdot, \cdot \rangle$ . The family  $\{e_j\}_{j=1}^n$  is the canonical basis of  $\mathbb{C}^n$  and  $M_n(\mathbb{C})$  is the space of complex  $n \times n$  matrices. For any matrix  $T \in M_n(\mathbb{C})$  we let  $\|T\| = \sup\{\|Ty\| : y \in \mathbb{C}^n, \|y\| \leq 1\}$  be the operator norm,  $T^*$  is the adjoint and  $\text{Tr } T = \sum_{j=1}^n T_{jj}$  the trace.

Given a set  $X$ ,  $\mathcal{F}(X, \mathbb{C}^n)$  denotes the vector space of functions from  $X$  into  $\mathbb{C}^n$ . When  $X$  is endowed with a  $\sigma$ -algebra  $\mathcal{A}$  and a positive finite measure

$\nu : \mathcal{A} \rightarrow [0, +\infty)$ , then  $L^2(X, \nu; \mathbb{C}^n)$  is the Hilbert space of (equivalence classes of)  $\nu$ -square-integrable functions from  $X$  into  $\mathbb{C}^n$ , with inner product  $\langle \cdot, \cdot \rangle_2$  and norm  $\|\cdot\|_2$ . If  $X$  has a topology,  $\mathcal{C}(X, \mathbb{C}^n)$  is the vector space of continuous functions from  $X$  into  $\mathbb{C}^n$  and  $\mathcal{B}(X)$  is the Borel  $\sigma$ -algebra.

In this paper we focus on reproducing kernel Hilbert spaces whose elements are functions from a set  $X$  taking values in  $\mathbb{C}^n$ . These Hilbert spaces are completely characterized by their reproducing kernel, which is a function from  $X \times X$  to  $M_n(\mathbb{C})$ , so that we consider the kernel as our primary object. We recall the following definition.

**Definition 2.1.** A map  $K : X \times X \rightarrow M_n(\mathbb{C})$  is called a  $M_n(\mathbb{C})$ -reproducing kernel if

- a) for all  $x, t \in X$ ,  $K(x, t)^* = K(t, x)$ ;
- b) for any  $m \geq 1$ ,  $x_1, \dots, x_m \in X$ ,  $y_1, \dots, y_m \in \mathbb{C}^n$

$$\sum_{i,j=1}^m \langle K(x_i, x_j) y_j, y_i \rangle \geq 0.$$

From now on we fix a  $M_n(\mathbb{C})$ -reproducing kernel  $K$  and, for any  $x \in X$  and  $j = 1, \dots, n$ , we denote by  $K_x^j$  the function in  $\mathcal{F}(X, \mathbb{C}^n)$  given by

$$K_x^j(t) := K(t, x) e_j, \quad t \in X.$$

We recall that  $K$  defines a unique RKHS  $\mathcal{H}_K$ , whose inner product and norm are denoted by  $\langle \cdot, \cdot \rangle_K$  and  $\|\cdot\|_K$ , respectively, such that  $\mathcal{H}_K$  is a vector subspace of  $\mathcal{F}(X, \mathbb{C}^n)$  and

$$\begin{aligned} K_x^j &\in \mathcal{H}_K, & \forall x \in X, j = 1, \dots, n \\ f(x) &= (\langle f, K_x^1 \rangle_K, \dots, \langle f, K_x^n \rangle_K), & \forall x \in X, f \in \mathcal{H}_K, \end{aligned} \quad (2)$$

see Proposition 2.1 of [8]. Furthermore, the following properties hold true

$$K(x, t)_{lj} = \langle K_t^j, K_x^l \rangle_K, \quad x, t \in X, j, l = 1, \dots, n \quad (3)$$

$$\mathcal{H}_K = \overline{\text{span}}\{K_x y : x \in X, y \in \mathbb{C}^n\} \quad (4)$$

$$f(x) = K_x^* f \quad x \in X$$

where  $K_x : \mathbb{C}^n \rightarrow \mathcal{H}_K$  is the (bounded) operator defined by  $K_x y = \sum y_j K_x^j$  for all  $y = (y^1, \dots, y^n) \in \mathbb{C}^n$ .

Finally, we recall that  $\mathcal{H}_K$  can be realized also as a closed subspace of a suitable Hilbert space by means of a particular feature map, as shown by the next result.

**Proposition 2.2** (Proposition 2.4 [8]). *Let  $\mathcal{H}$  be a Hilbert space and a map  $\gamma : X \rightarrow \mathcal{H}^n$ ,  $\gamma(x) = (\gamma_x^1, \dots, \gamma_x^n)$  be given. Then the operator  $W : \mathcal{H} \rightarrow \mathcal{F}(X; \mathbb{C}^n)$  defined by*

$$(Wu)(x) = (\langle u, \gamma_x^1 \rangle, \dots, \langle u, \gamma_x^n \rangle), \quad u \in \mathcal{H}, \quad x \in X, \quad (5)$$

*is a partial isometry from  $\mathcal{H}$  onto the reproducing kernel Hilbert space  $\mathcal{H}_K$  with reproducing kernel*

$$K(x, t)_{lj} = \langle \gamma_t^j, \gamma_x^l \rangle, \quad x, t \in X, \quad l, j = 1, \dots, n. \quad (6)$$

*Moreover,  $W^*W$  is the orthogonal projection onto*

$$\ker W^\perp = \overline{\text{span}}\{\gamma_x^i \mid x \in X, \quad i = 1, \dots, n\}.$$

### 3. Mercer theorem for measurable kernels

In this section we present the main result of the paper, namely a Mercer representation of a  $M_n(\mathbb{C})$ -reproducing kernel  $K$  under the assumptions that  $X$  is endowed with a finite measure  $\mu$  and  $K$  is measurable. The distinctive feature of our result with respect to already existing generalizations of Mercer theorem relies in the construction of an *ad hoc* topological structure on the space  $X$ , intrinsically defined by the kernel. Thanks to this topology and introducing a suitable measure related to  $\mu$ , we do not need to assume the space  $\mathcal{H}_K$  to be embedded in  $L^2(X, \mu; \mathbb{C}^n)$ , and we are nevertheless able to get a Mercer representation for the kernel and a strong convergence result on the series defining it. In particular, we recover uniform convergence on compact subsets with respect to the topology we introduce.

As in [22, 23], we note that the reproducing kernel  $K$  defines a pseudo-metric  $d$  on  $X$

$$d(x, t) = \sup_{\substack{y \in \mathbb{C}^n \\ \|y\| \leq 1}} \|K_x y - K_t y\|_K \quad x, t \in X, \quad (7)$$

which induces a (non-Hausdorff) topology  $\tau_K$  on  $X$ . A basis of  $\tau_K$  is provided by the family of open balls  $\{B(x, r) : x \in X, r > 0\}$  where

$$B(x, r) = \{t \in X : d(x, t) < r\}. \quad (8)$$

*Remark 3.1.* The pseudo-metric  $d$  can be also computed as follows.

$$\begin{aligned} d(x, t)^2 &= \sup_{\substack{y \in \mathbb{C}^n \\ \|y\| \leq 1}} \langle (K(x, x) - K(x, t) - K(t, x) + K(t, t))y, y \rangle \\ &= \|K(x, x) - K(x, t) - K(t, x) + K(t, t)\|, \end{aligned} \quad (9)$$

where  $\|A\| = \sup_{\|y\| \leq 1} \|Ay\|$  is the operator norm of the matrix  $A$  and the second equality holds since  $K(x, x) - K(x, t) - K(t, x) + K(t, t)$  is a positive-defined matrix. It is easy to check that  $d$  is equivalent to the pseudo-metric  $d'(x, t) = \sqrt{\sum_{j=1}^n \|K_x^j - K_t^j\|_K^2}$ , which gives rise to the same topology.

The following result states some properties of  $\tau_K$ .

**Theorem 3.2.** *Assume that  $\mathcal{H}_K$  is separable.*

- i) The space  $X$  endowed with the topology  $\tau_K$  is second countable and  $K$  is continuous;*
- ii) If  $\mathcal{A}$  is a  $\sigma$ -algebra on  $X$  with respect to which  $K$  is measurable, the Borel  $\sigma$ -algebra  $\mathcal{B}(X)$  generated by  $\tau_K$  is contained in  $\mathcal{A}$ ;*
- iii) If  $\mu : \mathcal{A} \rightarrow [0, +\infty)$  is a finite measure, then there exists a unique closed set  $C \subseteq X$ , namely the support of  $\mu$ , such that  $\mu(C) = \mu(X)$  and that, if  $C'$  is another closed subset with  $\mu(C') = \mu(X)$ , then  $C' \supseteq C$ .*

The support of  $\mu$  is denoted by  $\text{supp } \mu$  and it is, by its very definition, the smallest closed subset of  $X$  having full measure. The assumption that  $\mathcal{H}_K$  is separable is essential to prove its existence.

*Remark 3.3.* If  $X$  is endowed with its own topology  $\tau$  and the reproducing kernel is continuous with respect to  $\tau$ , then  $\tau_K \subseteq \tau$ . Indeed, the fact that  $K$  is  $\tau$ -continuous from  $X \times X$  into  $M_n(\mathbb{C})$ , endowed with the operator norm, implies that the map  $(x, t) \mapsto \|K(x, x) - K(x, t) - K(t, x) + K(t, t)\|$  is  $\tau$ -continuous. By Eq. (9), for any  $x \in X$  the map  $t \mapsto d(x, t)$  is also  $\tau$ -continuous. Since  $B(x, r) = \{t \in X \mid d(x, t) < r\}$  for all  $r > 0$ , then  $B(x, r) \in \tau$ , so that  $\tau_K \subseteq \tau$ . Though in general  $\tau_K \neq \tau$ , any function in  $\mathcal{H}_K$  is continuous with respect to both topologies.

From now on, we fix a  $\sigma$ -algebra  $\mathcal{A}$  on  $X$  and a finite measure  $\mu$  defined on  $\mathcal{A}$ . We assume that  $\mathcal{H}_K$  is separable and that  $K$  is measurable, and we regard  $X$  as a second countable topological space with respect to the

topology  $\tau_K$ . Though  $K$  is continuous, this condition does not ensure that the integral operator with kernel  $K$  is bounded on  $L^2(X, \mu; \mathbb{C}^n)$ . We overcome this problem by considering a second measure  $\nu$ , which is equivalent to  $\mu$ , such that the integral operator with kernel  $K$  is bounded on  $L^2(X, \mu; \mathbb{C}^n)$ . Indeed, define  $\nu : \mathcal{A} \rightarrow [0, +\infty)$  as

$$\nu(A) := \int_A \frac{1}{1 + \|K(x, x)\|} d\mu(x), \quad A \in \mathcal{A}. \quad (10)$$

Clearly  $\nu$  is a positive finite measure, which satisfies  $\text{supp } \nu = \text{supp } \mu$ . Furthermore, since  $\text{Tr } K(x, x) \leq n\|K(x, x)\|$ , the integral  $\int_X \text{Tr } K(x, x) d\nu(x)$  is finite and Theorem 6.1 in the appendix states that the integral operator with kernel  $K$

$$\begin{aligned} L_\nu : L^2(X, \nu; \mathbb{C}^n) &\rightarrow L^2(X, \nu; \mathbb{C}^n) \\ (L_\nu f)(x) &= \int_X K(x, t) f(t) d\nu(t), \end{aligned} \quad (11)$$

is well-defined, positive and compact<sup>1</sup>. The Hilbert-Schmidt theorem gives the existence of a basis of  $L^2(X, \nu; \mathbb{C}^n)$  of eigenfunctions of  $L_\nu$  and this basis provides a Mercer decomposition of  $K$ , as shown by the following result.

**Theorem 3.4.** *Let  $(X, \mathcal{A})$  be a measurable space endowed with a finite measure  $\mu$ . Assume that the reproducing kernel  $K : X \times X \rightarrow M_n(\mathbb{C})$  is measurable and  $\mathcal{H}_K$  is separable. Define  $\nu$  as in (10) and  $L_\nu$  as in (11). Then there exists a countable family  $\{f_i\}_{i \in I}$  in  $\mathcal{F}(X, \mathbb{C}^n)$  such that:*

- a) for all  $i \in I$  the function  $f_i$  is continuous with respect to  $\tau_K$ ,
- b) the family  $\{f_i\}_{i \in I}$  is an orthonormal basis of  $\ker L_\nu^\perp \subseteq L^2(X, \nu; \mathbb{C}^n)$  and, for all  $i \in I$ ,  $L_\nu f_i = \sigma_i f_i$  for some  $\sigma_i \in (0, +\infty)$ .

Given any family  $\{f_i\}_{i \in I}$  satisfying a) and b), then

- i) for all  $x, t \in \text{supp } \mu$  and  $j, l = 1, \dots, n$

$$K(x, t)_{lj} = \sum_{i \in I} \sigma_i f_i^j(t) \overline{f_i^l(x)}, \quad (12)$$

where the convergence is uniform on compact subsets of  $\text{supp } \mu \times \text{supp } \mu$ ;

---

<sup>1</sup>If  $Y$  is infinite dimensional and  $K(x, x)$  is compact for all  $x \in X$ , it is possible to prove that  $L_\nu$  is compact by Proposition 4.8 of [8].



- ii) the family  $\{\sqrt{\sigma_i}f_i\}_{i \in I}$  is orthonormal in  $\mathcal{H}_K$ ;
- iii) if  $\text{supp } \mu = X$ ,  $\{\sqrt{\sigma_i}f_i\}_{i \in I}$  is an orthonormal basis of  $\mathcal{H}_K$ .
- iv) for  $j = 1, \dots, n$ , the family  $\{\sqrt{\sigma_i}f_i^j\}_{i \in I}$  is a Parseval frame in the scalar reproducing kernel Hilbert space  $\mathcal{H}_{K_j}$  with reproducing kernel  $K_j$  given by

$$K_j(x, t) = K(x, t)_{jj} \quad x, t \in X. \quad (13)$$

We recall that  $\{\sqrt{\sigma_i}f_i^j\}_{i \in I}$  is a Parseval frame in  $\mathcal{H}_{K_j}$  if

$$\|f\|_{K_j}^2 = \sum_{i \in I} \sigma_i |\langle f, f_i^j \rangle_{K_j}|^2 \quad \forall f \in \mathcal{H}_{K_j}. \quad (14)$$

Item iv) of Theorem 3.4 provides a tool to construct  $M_n(\mathbb{C})$ -reproducing kernels as shown by the following result.

**Proposition 3.5.** *Let  $(X, \mathcal{A})$  be a measurable space endowed with a finite measure  $\mu$  such that  $\text{supp } \mu = X$ . Given a family  $K_1, \dots, K_n$  of  $n$  scalar measurable reproducing kernels on  $X$ , for each  $j = 1, \dots, n$  take a Parseval frame  $\{f_i^j\}_{i \in I}$  in the corresponding reproducing kernel Hilbert space  $\mathcal{H}_{K_j}$  with  $I$  countable, and define the function  $K : X \times X \rightarrow M_n(\mathbb{C})$  as*

$$K(x, t)_{lj} = \sum_{i \in I} f_i^j(t) \overline{f_i^l(x)} \quad \forall x, t \in X. \quad (15)$$

The map  $K$  is a measurable  $M_n(\mathbb{C})$ -reproducing kernel on  $X$  satisfying (13) and  $\mathcal{H}_K$  is separable.

The following two simple examples illustrate the machinery involved in the above results. First, we focus on the Mercer decomposition (12). The main difficult task is the computation of the spectral decomposition of  $L_\nu$ . To overcome this problem, we consider a continuous periodic kernel on  $\mathbb{R}$ .

**Example 3.6.** Fix  $n = 2$  and choose a continuous  $2\pi$ -periodic  $M_2(\mathbb{C})$ -valued kernel on the real line, *i.e.*

$$K(x, t) = \begin{bmatrix} a(t-x) & b(t-x) \\ \overline{b(x-t)} & c(t-x) \end{bmatrix} \quad (16)$$

where  $a(t-x) = \langle K_t^1, K_x^1 \rangle$ ,  $c(t-x) = \langle K_t^2, K_x^2 \rangle$  and  $\overline{b(t-x)} = \langle K_t^2, K_x^1 \rangle$  are continuous  $2\pi$ -periodic functions, with  $K(x, t)_{21} = \overline{b(x-t)}$  by condition a)

of Definition 2.1. Due to the periodicity, we set  $X = (-\pi, \pi]$ ,  $\mu = (2\pi)^{-1}dx$  the normalized Lebesgue measure, and  $\tau$  the natural topology of  $X$ . Since  $K$  is  $\tau$ -continuous,  $\tau_K \subseteq \tau$  by Remark 3.3 and, as a consequence, the support of  $\mu$  (with respect to  $\tau_K$ ) contains the support of  $\mu$  (with respect to  $\tau$ ), which is  $X$ . Hence  $\text{supp } \mu = X$ .

Since  $K$  satisfies condition b) of Definition 2.1,  $a$  and  $b$  are functions of positive type and  $b$  is a linear combination of four functions of positive type. Hence, a standard result on Fourier series implies that for all  $x \in X$

$$a(x) = \sum_{m \in \mathbb{Z}} \hat{a}_m e^{imx} \quad b(x) = \sum_{m \in \mathbb{Z}} \hat{b}_m e^{imx} \quad c(x) = \sum_{m \in \mathbb{Z}} \hat{c}_m e^{imx}, \quad (17)$$

where  $(\hat{a}_m)_{m \in \mathbb{Z}}$ ,  $(\hat{b}_m)_{m \in \mathbb{Z}}$  and  $(\hat{c}_m)_{m \in \mathbb{Z}}$  are in  $\ell_1(\mathbb{Z})$ . Furthermore condition b) of Definition 2.1 gives that for all  $m \in \mathbb{Z}$  the matrix

$$\hat{K}_m := \begin{bmatrix} \hat{a}_m & \hat{b}_m \\ \overline{\hat{b}_m} & \hat{c}_m \end{bmatrix}$$

is positive defined, that is,

$$\hat{a}_m \geq 0 \quad \hat{c}_m \geq 0 \quad \hat{a}_m \hat{c}_m \geq |\hat{b}_m|^2. \quad (18)$$

Conversely, any family of sequences  $(\hat{a}_m)_{m \in \mathbb{Z}}$ ,  $(\hat{b}_m)_{m \in \mathbb{Z}}$ ,  $(\hat{c}_m)_{m \in \mathbb{Z}}$  in  $\ell_1(\mathbb{Z})$  satisfying (18) defines a continuous  $2\pi$ -periodic  $M_2(\mathbb{C})$ -valued reproducing kernel on  $\mathbb{R}$  by means of Eqs. (16) and (17) (hence providing us with a set of explicit examples). For instance, the choice

$$K(t, x) = \begin{bmatrix} \frac{(\pi - |t-x|)^2}{2} & \frac{\pi}{8}(t-x)(\pi - |t-x|) \\ \frac{\pi}{8}(x-t)(\pi - |x-t|) & \frac{(\pi - |t-x|)^2}{2} \end{bmatrix} \quad -\pi < t-x \leq \pi$$

gives a  $M_2(\mathbb{C})$ -reproducing kernel since it is well known that

$$\hat{K}_0 = \begin{bmatrix} \frac{\pi^2}{6} & 0 \\ 0 & \frac{\pi^2}{6} \end{bmatrix} \quad \hat{K}_m = \begin{bmatrix} \frac{1}{m^2} & \frac{1-(-1)^m}{4im^3} \\ \frac{(-1)^m-1}{4im^3} & \frac{1}{m^2} \end{bmatrix} \quad m \neq 0,$$

so that condition (18) holds true. Note that this kernel can not be written as the product of a scalar kernel with a constant  $2 \times 2$ -matrix.

Since  $K$  is bounded, the integral operator  $L_\mu$  of kernel  $K$  acting on  $L^2(X, \mu, \mathbb{C}^2)$  is always compact, so we do not need to introduce the measure

$\nu$ . Moreover, the eigenfunctions of  $L_\mu$  can be easily computed by observing that  $L_\mu$  acts by convolution and, in the Fourier domain, as pointwise multiplication, *i.e.*

$$\widehat{L}_\mu \left( \begin{bmatrix} \hat{f}_m^1 \\ \hat{f}_m^2 \end{bmatrix} \right)_{m \in \mathbb{Z}} = \left( \hat{K}_m \begin{bmatrix} \hat{f}_m^1 \\ \hat{f}_m^2 \end{bmatrix} \right)_{m \in \mathbb{Z}}, \quad \left( \begin{bmatrix} \hat{f}_m^1 \\ \hat{f}_m^2 \end{bmatrix} \right)_{m \in \mathbb{Z}} \in \ell_2(\mathbb{Z})^2.$$

Since  $\hat{K}_m$  is positive defined,  $\hat{K}_m$  has two eigenvalues  $\sigma_{1,m}, \sigma_{2,m} \in [0, +\infty)$  with corresponding (normalized) eigenvectors  $v_{1,m}, v_{2,m} \in \mathbb{C}^2$ . Then, for all  $m \in \mathbb{Z}$  and  $\ell = 1, 2$ , the function  $f_{\ell,m}, f_{\ell,m}(x) = v_{\ell,m} e^{imx}$ , is an eigenfunction of  $L_\mu$  with eigenvalue  $\sigma_{\ell,m}$  and  $\{f_{\ell,m}\}_{\sigma_{\ell,m} > 0}$  is an orthonormal basis of  $\ker L_\mu^\perp$ . By item v) of Theorem 6.1 (with  $L_\nu$  replaced by  $L_\mu$ ), if  $\sigma_{\ell,m} > 0$  there exists a function  $g_{\ell,m} \in \mathcal{H}_K$  such that for  $\mu$ -almost all  $x \in X$  we have  $g_{\ell,m}(x) = f_{\ell,m}(x)$ . By Remark 3.3  $g_{\ell,m}$  is continuous with respect to  $\tau$ , hence for all  $x \in X$   $g_{\ell,m}(x) = f_{\ell,m}(x)$ , *i.e.*  $f_{\ell,m}$  is  $\tau_K$ -continuous. Item i) of Theorem 3.4 provides the following decomposition

$$\begin{aligned} K(x, t)_{ij} &= \sum_{\ell=1,2} \sum_{m \in \mathbb{Z}} \sigma_{\ell,m} f_{\ell,m}^j(t) \overline{f_{\ell,m}^i(x)} \\ &= \sum_{\ell=1,2} \sum_{m \in \mathbb{Z}} \sigma_{\ell,m} v_{\ell,m}^j \overline{v_{\ell,m}^i} e^{im(t-x)}, \end{aligned}$$

where the series converges absolutely for all  $x, t \in X$  and uniformly on  $X \times X$  (the series  $\sum \sigma_{\ell,m}$  is finite since it is equal to  $\sum(\hat{a}_m + \hat{c}_m)$ , which is convergent). This example extends to any Hilbert space valued RKHS on an abelian locally compact group, see Proposition 5.3. in [9].

**Example 3.7.** Using Proposition 3.5, we illustrate how to build a matrix-valued kernel starting from two scalar Gaussian kernels  $K_j : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  ( $j = 1, 2$ ) defined as usual as

$$K_1(x, t) = e^{-\frac{(x-t)^2}{2\sigma_1^2}}, \quad K_2(x, t) = e^{-\frac{(x-t)^2}{2\sigma_2^2}},$$

with  $\sigma_j > 0$ , for  $j = 1, 2$ . Set

$$f_i^j(x) = \frac{x^i}{\sqrt{i!}(\sigma_j)^i} e^{-\frac{x^2}{2\sigma_j^2}}, \quad j = 1, 2, \quad i \in \mathbb{N}, \quad x \in \mathbb{R},$$

then

$$\begin{aligned}
\sum_{i \in \mathbb{N}} f_i^j(x) \overline{f_i^j(t)} &= \sum_{i \in \mathbb{N}} \frac{x^i}{\sqrt{i!}(\sigma_j)^i} e^{-\frac{x^2}{2\sigma_j^2}} \frac{t^i}{\sqrt{i!}(\sigma_j)^i} e^{-\frac{t^2}{2\sigma_j^2}} \\
&= e^{\left(-\frac{x^2}{2\sigma_j^2} - \frac{t^2}{2\sigma_j^2}\right)} \sum_{i \in \mathbb{N}} \frac{x^i t^i}{i!(\sigma_j)^{2i}} \\
&= e^{\left(-\frac{x^2}{2\sigma_j^2} - \frac{t^2}{2\sigma_j^2}\right)} e^{\frac{xt}{\sigma_j^2}} = K_j(x, t).
\end{aligned}$$

Condition (15) is satisfied. It is known, see for example [13], that  $\{f_i^j\}_{i \in \mathbb{N}}$  is an orthonormal basis of  $\mathcal{H}_{K_j}$ , hence a Parseval frame (a direct proof can be obtained by considering the maps  $\gamma^i : X \rightarrow \ell_2$ ,  $\gamma(x) = \{f_i^j(x)\}_{i \in \mathbb{N}}$  and by applying Proposition 2.2). Therefore a matrix-valued kernel is defined by setting

$$K(x, t)_{lj} = \sum_{i \in \mathbb{N}} f_i^j(x) \overline{f_i^l(t)}.$$

Note that the off-diagonal blocks

$$K(x, t)_{12} = K(x, t)_{21} = e^{-\left(\frac{x}{\sqrt{2}\sigma_1} - \frac{t}{\sqrt{2}\sigma_2}\right)^2}$$

are not a Gaussian kernel, if  $\sigma_1 \neq \sigma_2$ .

#### 4. Continuous Mercer theorem on a metric space

The first step in order to show Theorem 3.4 is to prove Mercer theorem under the assumption that  $X$  is a metric space,  $K$  is continuous and  $\int_X \text{Tr} K(x, x) d\nu(x)$  is finite. For scalar kernels the result is well known, see [13]. However, our proof is elementary and it holds for matrix-valued kernels. As in [24, 25], it is based on the singular value decomposition of the embedding  $i_\nu : \mathcal{H}_K \rightarrow L^2(X, \nu; \mathbb{C}^n)$ , which is a compact operator. We will make use of some known properties of  $i_\nu$  collected in the appendix.

**Theorem 4.1.** *Let  $X$  be a separable metric space and  $\nu$  a finite measure defined on  $\mathcal{B}(X)$ . Assume  $K : X \times X \rightarrow M_n(\mathbb{C})$  to be a continuous reproducing kernel such that*

$$\int_X \text{Tr} K(x, x) d\nu(x) < +\infty. \quad (19)$$

Define the trace class operator  $L_\nu$  as in (11) and take an orthonormal basis  $\{f_i\}_{i \in I}$  of  $\ker L_\nu^\perp$  of continuous eigenvectors of  $L_\nu$  and let  $\{\sigma_i\}_{i \in I} \subseteq (0, +\infty)$  be the corresponding family of eigenvalues. Then the family  $\{\sqrt{\sigma_i} f_i\}_{i \in I}$  is orthonormal in  $\mathcal{H}_K$  and

$$K(x, t)_{lj} = \sum_{i \in I} \sigma_i f_i^j(t) \overline{f_i^l(x)} \quad \forall x, t \in \text{supp } \nu, \quad (20)$$

where the series converges uniformly on compact subsets of  $\text{supp } \nu \times \text{supp } \nu$ . If  $\text{supp } \nu = X$ ,  $\{\sqrt{\sigma_i} f_i\}_{i \in I}$  is an orthonormal basis of  $\mathcal{H}_K$ .

*Remark 4.2.* Item 4) of Theorem 6.1 in the appendix guarantees the existence of a basis  $\{f_i\}_{i \in I}$  of  $\ker L_\nu^\perp$  of continuous eigenvectors of  $L_\nu$ .

*Proof.* As in Theorem 6.1, we denote by  $i_\nu : \mathcal{H}_K \hookrightarrow L^2(X, \nu; \mathbb{C}^n)$  the canonical embedding. Its adjoint  $i_\nu^*$  is given by (29), so that  $L_\nu = i_\nu i_\nu^*$ , and we define the operator  $T_\nu : \mathcal{H}_K \rightarrow \mathcal{H}_K$  as  $T_\nu := i_\nu^* i_\nu$ . Take a family  $\{f_i\}_{i \in I}$  as in the statement of the theorem and, for all  $i \in I$ , define  $g_i = i_\nu^* f_i / \sqrt{\sigma_i}$ . The singular value decomposition of  $i_\nu^*$  gives that  $\{g_i\}_{i \in I}$  is an orthonormal basis of  $\ker T_\nu^\perp$  made of eigenvectors of  $T_\nu$ . We claim that, for all  $x \in \text{supp } \nu$  and  $j = 1, \dots, n$ ,  $K_x^j \in \ker T_\nu^\perp$ . Indeed, for any  $f \in \ker T_\nu$

$$0 = \langle T_\nu f, f \rangle_K = \langle i_\nu f, i_\nu f \rangle = \sum_{j=1}^n \int_X |f^j(x)|^2 d\nu(x).$$

Hence, for any  $j = 1, \dots, n$ , the map  $x \mapsto f^j(x) = \langle f, K_x^j \rangle_K$  is zero  $\nu$ -almost everywhere. Since  $\mathcal{H}_K \subseteq C(X, \mathbb{C}^n)$ , see item 1) of Theorem 6.1, the definition of support implies that  $\langle f, K_x^j \rangle_K = 0$  for all  $x \in \text{supp } \nu$ . Hence

$$K_t^j \in \ker T_\nu^\perp \quad \forall t \in \text{supp } \nu, \quad j = 1, \dots, n. \quad (21)$$

Furthermore, since  $\{g_i\}_{i \in I}$  is a basis of  $\ker T_\nu^\perp$ , for all  $x \in \text{supp } \nu$  and  $j = 1, \dots, n$

$$K_x^j = \sum_{i \in I} \langle K_x^j, g_i \rangle_K g_i.$$

Hence, the reproducing property gives that

$$K(x, t)_{lj} = \langle K_x^l, K_t^j \rangle_K = \sum_{i \in I} \langle K_x^l, g_i \rangle_K \langle g_i, K_t^j \rangle_K = \sum_{i \in I} \sigma_i f_i^j(t) \overline{f_i^l(x)}$$

for all  $x, t \in \text{supp } \nu$ .

Concerning the uniform convergence, suppose  $I = \mathbb{N}$ , fix two compact subsets  $C, C' \subseteq \text{supp } \nu$ , and consider the remainder

$$\sup_{(x,t) \in C \times C'} \left| \sum_{i=q}^{+\infty} \sigma_i f_i^j(t) \overline{f_i^l(x)} \right| \leq \sqrt{\sup_{x \in C} \sum_{i=q}^{+\infty} \sigma_i |f_i^l(x)|^2} \sqrt{\sup_{t \in C'} \sum_{i=q}^{+\infty} \sigma_i |f_i^j(t)|^2}. \quad (22)$$

The series of continuous functions  $\sum_{i=0}^{+\infty} \sigma_i |f_i^l(x)|^2$  converges pointwise to the continuous function  $K(x, x)_l$  on the compact set  $C$ , and therefore uniform convergence follows from Dini's theorem. Thus, relying on the bound in (22), we have

$$\lim_{q \rightarrow +\infty} \sup_{(x,t) \in C \times C'} \left| \sum_{i=q}^{+\infty} \sigma_i f_i^j(t) \overline{f_i^l(x)} \right| = 0.$$

Assume that  $\text{supp } \nu = X$ . Since, by (4),  $\{K_t^j : t \in X, j = 1, \dots, n\}$  is total in  $\mathcal{H}_K$ , (21) implies that  $\ker T_\nu = \{0\}$ . Hence the family  $\{\sqrt{\sigma_i} f_i\}_{i \in I}$  is an orthonormal basis of  $\mathcal{H}_K$ .  $\square$

## 5. Proofs

To prove the Mercer representation in the general setting of Theorem 3.4, we would like to define a metric  $d$  on  $X$  such that  $K$  becomes continuous. A natural choice would be the map  $d$  defined by (7). However,  $d$  is not a metric unless the map  $x \mapsto K_x$  is injective. To overcome this problem, we first introduce a suitable metric space  $\tilde{X}$  and a continuous kernel  $\tilde{K}$  such that the corresponding reproducing kernel  $\mathcal{H}_{\tilde{K}}$  is isomorphic to  $\mathcal{H}_K$  and, as a consequence, we prove Theorem 3.2. Afterwards, the Mercer representation of  $K$  is deduced by the corresponding representation (20) of  $\tilde{K}$  given by Theorem 4.1. From now on  $(X, \mathcal{A})$  is a measurable space endowed with a finite measure  $\mu$  and  $K$  is a  $\mathbb{C}^n$ -measurable reproducing kernel such that  $\mathcal{H}_K$  is separable.

Clearly  $d$  in (7) is a pseudo-metric. The symmetry property and the triangular inequality directly follow from the definition, while from  $d(x, t) = 0$  we get  $K_x = K_t$ , which as noted before in general does not imply  $x = t$ . However, the reproducing property (2) gives  $f(x) = f(t)$  for all  $f \in \mathcal{H}_K$ , which means that the functions in  $\mathcal{H}_K$  are not able to distinguish the points  $x$  and  $t$ . This suggests to define an equivalence relation  $\sim$  on  $X$  by setting

$$x \sim t \iff K_x = K_t. \quad (23)$$

Denote by  $\tilde{X} = X/\sim$  the corresponding quotient space and, given  $[x], [t] \in \tilde{X}$ , define the function  $\tilde{d}([x], [t]) := d(x, t)$ . Then  $\tilde{d}$  is a distance on  $\tilde{X}$  so that  $(\tilde{X}, \tilde{d})$  is a metric space.

We consider the pull-back topology  $\tau_K$  induced on  $X$  by the canonical projection  $\pi : X \rightarrow X/\sim$ , i.e.

$$\tau_K = \{\pi^{-1}(A) : A \text{ open in } (\tilde{X}, \tilde{d})\}.$$

It is clear that the family of open balls  $\{B(x, r) : x \in X, r > 0\}$  is a basis for  $\tau_K$ , see (8). Now, the proof of Theorem 3.2 is a consequence of the next proposition where  $\mathcal{L}(\mathbb{C}^n, \mathcal{H}_K)$  denotes the space of (bounded) linear operator from  $\mathbb{C}^n$  to  $\mathcal{H}_K$  endowed with the operator norm  $\|\cdot\|_{n,K}$  so that, for example,

$$d(x, t) = \|K_x - K_t\|_{n,K} = \sup_{\substack{y \in \mathbb{C}^n \\ \|y\| \leq 1}} \|K_x y - K_t y\|_K.$$

**Proposition 5.1.** *The following facts hold:*

- i) the map  $\Phi : \tilde{X} \rightarrow \mathcal{L}(\mathbb{C}^n, \mathcal{H}_K)$  given by  $\Phi([x]) = K_x$  is an isometry from  $(\tilde{X}, \tilde{d})$  into  $(\mathcal{L}(\mathbb{C}^n, \mathcal{H}_K), \|\cdot\|_{n,K})$ ;*
- ii) the spaces  $(\tilde{X}, \tilde{d})$  and  $(X, \tau_K)$  are second countable;*
- iii) the  $\sigma$ -algebra  $\mathcal{A}$  contains  $\mathcal{B}(X)$ , the Borel sets generated by  $\tau_K$ ;*
- iv) given a positive finite measure  $\nu$  on  $X$ , there exists  $\text{supp } \nu$ .*

*Proof.* Statement *i)* follows directly from the definition of the equivalence relation  $\sim$  and of the pseudo-distance  $d$ .

*ii)* Since  $\mathcal{H}_K$  is separable, the space  $\mathcal{L}(\mathbb{C}^n, \mathcal{H}_K)$  can be identified with  $\mathcal{H}_K^n$ , and then it is separable. Therefore, the set  $\Phi(\tilde{X}) \subseteq \mathcal{L}(\mathbb{C}^n, \mathcal{H}_K)$  is separable as well, and so is  $\tilde{X}$ , since  $\Phi$  is an isometry. Since  $\tilde{X}$  is a separable metric space, there exists a countable basis  $\{A_i\}_{i \in \mathbb{N}}$  of open subsets of  $\tilde{X}$ . Clearly,  $\{\pi^{-1}(A_i)\}_{i \in \mathbb{N}}$  is a countable basis for  $\tau_K$ , that is,  $\tau_K$  is second countable.

To show that *iii)* holds true, it is enough to prove that each element  $B(x, r)$  of the basis of  $\tau_K$  belongs to  $\mathcal{A}$ . Towards this end, if for a given  $x \in X$  we prove that the map  $G_x : (X, \mathcal{A}) \rightarrow [0, +\infty)$ ,  $G_x(y) = \|K_y - K_x\|_{n,K}$  is measurable we are done. Since  $G_x$  is the composition of the function  $X \ni y \mapsto K_y - K_x \in \mathcal{L}(\mathbb{C}^n, \mathcal{H}_K)$ , with  $\mathcal{L}(\mathbb{C}^n, \mathcal{H}_K) \ni A \mapsto \|A\|_{n,K} \in \mathbb{R}$ , and

the latter is continuous, it is enough to prove that the first one is measurable. This follows from separability of  $\mathcal{H}_K$  and Proposition 3.1 in [8].

Finally, to prove *iv*), define  $\text{supp } \nu$  as the intersection of all  $\tau_K$ -closed subsets  $C \subseteq X$  with  $\nu(C) = \nu(X)$ . Clearly  $\text{supp } \nu$  is closed, and we prove that  $\nu(\text{supp } \nu) = \nu(X)$ . Indeed, since  $\tau_K$  is second countable, there exists a sequence of closed sets  $\{C_j\}_{j \in \mathbb{N}}$  such that, for an arbitrary closed set  $C$ ,  $C = \bigcap_k C_{j_k}$  for a suitable subsequence  $\{C_{j_k}\}_{k \in \mathbb{N}}$ . Hence,

$$\begin{aligned} \nu(\text{supp } \nu) &= \nu \left( \bigcap_{\substack{C \text{ closed,} \\ \nu(C) = \nu(X)}} C \right) = \nu \left( \bigcap_{\substack{j \in \mathbb{N} \\ \nu(C_j) = \nu(X)}} C_j \right) \\ &= \lim_{\substack{j \in \mathbb{N} \\ \nu(C_j) = \nu(X)}} \nu(C_j) = \nu(X). \end{aligned}$$

□

Note that, since  $\tilde{X}$  is a second countable metric space, it is separable. We now define a continuous kernel  $\tilde{K}$  on the separable metric space  $(\tilde{X}, \tilde{d})$  in order to apply Theorem 4.1, once that a suitable measure  $\tilde{\nu}$  has been also introduced. Set

$$\tilde{K} : \tilde{X} \times \tilde{X} \rightarrow M_n(\mathbb{C}), \quad \tilde{K}([x], [t]) := K(x, t),$$

and denote by  $\mathcal{H}_{\tilde{K}}$  the RKHS associated to  $\tilde{K}$ . First of all, note that (3) and the definition of the equivalence classes in  $\tilde{X}$  guarantee that  $\tilde{K}$  is well-defined. The next proposition aims at clarifying some basic properties of this space and most of all the connections between  $\mathcal{H}_K$  and  $\mathcal{H}_{\tilde{K}}$ . In particular, as it will be made precise later, the two spaces roughly speaking coincide.

**Proposition 5.2.** *The following facts hold:*

- i)  $\tilde{K}$  is a continuous kernel and every  $f \in \mathcal{H}_{\tilde{K}}$  is a continuous function;*
- ii)  $\mathcal{H}_{\tilde{K}}$  is separable;*
- iii)  $\mathcal{H}_{\tilde{K}}$  and  $\mathcal{H}_K$  are unitarily equivalent by means of the unitary operator*

$$W : \mathcal{H}_{\tilde{K}} \rightarrow \mathcal{H}_K \quad (Wf)(x) := \tilde{f}([x]); \quad (24)$$



iv) given a sequence of functions  $(\tilde{f}_n)_{n \in \mathbb{N}}$  in  $\mathcal{H}_{\tilde{K}}$  such that  $\tilde{f}_n \rightarrow \tilde{f} \in \mathcal{H}_{\tilde{K}}$  uniformly on the compact sets of  $\tilde{X}$ , then  $W\tilde{f}_n \rightarrow W\tilde{f}$  uniformly on the compact sets of  $X$ .

*Proof.* i) Given  $x_0, t_0 \in X$  we prove that  $\tilde{K}$  is continuous in  $([x_0], [t_0])$ . For all  $x, t \in X$  we have

$$\begin{aligned} \|\tilde{K}([x], [t]) - \tilde{K}([x_0], [t_0])\| &\leq \|K_x^* K_t - K_x^* K_{t_0}\| + \|K_x^* K_{t_0} - K_{x_0}^* K_{t_0}\| \\ &\leq \|K_x^*\|_{K,n} \|K_t - K_{t_0}\|_{n,K} + \|K_x^* - K_{x_0}^*\|_{K,n} \|K_{t_0}\|_{n,K}. \end{aligned}$$

Since  $\|K_x^*\|_{K,n} \leq \|K_x^* - K_{x_0}^*\|_{K,n} + \|K_{x_0}^*\|_{K,n} = \|K_x - K_{x_0}\|_{n,K} + \|K_{x_0}\|_{n,K}$ , the continuity of  $\Phi$  gives the thesis.

The second part of statement i) follows by the reproducing formula  $f(x) = \tilde{K}_x^* f$  for all  $f \in \mathcal{H}_{\tilde{K}}$ .

ii) Since  $\tilde{X}$  and  $\mathbb{C}^n$  are separable (see Proposition 5.1.ii), the space  $\mathcal{H}_{\tilde{K}} = \overline{\text{span}}\{\tilde{K}_{[x]}y : x \in X, y \in \mathbb{C}^n\}$  is separable too.

iii) We apply Proposition 2.2 taking  $\mathcal{H} = \mathcal{H}_{\tilde{K}}$  and  $\gamma_x = \tilde{K}_{[x]}$ , so that

$$(W\tilde{f})(x) = \tilde{f}([x]) = f(x)$$

for all  $x \in X$  and  $\tilde{f} \in \mathcal{H}_{\tilde{K}}$ . Since

$$\gamma_x^* \gamma_t = \tilde{K}([x], [t]) = K(x, t),$$

Proposition 2.2 implies that the range of  $W$  is  $\mathcal{H}_K$  and that  $W$  is a partial isometry from  $\mathcal{H}_{\tilde{K}}$  into  $\mathcal{H}_K$ . Moreover,  $f = 0$  clearly implies  $\tilde{f} = 0$ , and so  $W$  is injective.

iv) Let  $C$  be a compact subset of  $(X, \tau_K)$ . Since by construction  $\pi : X \rightarrow \tilde{X}$  is continuous with respect to  $\tau_K$ ,  $\pi(C)$  is compact in  $\tilde{X}$  and therefore  $\sup_{[x] \in \pi(C)} |\tilde{f}_n([x]) - \tilde{f}([x])| \rightarrow 0$ . Since  $W\tilde{f}_n(x) = \tilde{f}_n([x])$  by definition, the thesis follows.  $\square$

In order to apply Theorem 4.1 to the kernel  $\tilde{K}$ , the last ingredient we need is a finite measure  $\tilde{\nu}$  on  $\tilde{X}$ . If  $\nu$  is defined as in (10), using the canonical projection we can set

$$\tilde{\nu}(A) := \nu(\pi^{-1}(A)) \quad \text{for all Borel set } A \text{ in } (\tilde{X}, \tilde{d}).$$

$\tilde{\nu}$  is well defined since  $\pi^{-1}(A) \in \mathcal{B}(X)$  by continuity of  $\pi$ , and  $\mathcal{B}(X) \subseteq \mathcal{A}$  thanks to Proposition 5.1.iii). Moreover, we clearly have

$$\text{supp } \mu = \text{supp } \nu = \pi^{-1}(\text{supp } \tilde{\nu}). \quad (25)$$

We are now ready to prove our main result.

*Proof of Theorem 3.4.* From the results collected so far, we know that  $\tilde{K}$  is a continuous kernel by Proposition 5.2, and  $(\tilde{X}, \tilde{d})$  is a separable metric space (see Proposition 5.1), which is endowed with a finite measure  $\tilde{\nu}$ . In order to apply Theorem 4.1, we need to show that the integrability condition (19) is met by  $\tilde{K}$ . From the definition of  $\tilde{X}$ ,  $\tilde{K}$  and  $\tilde{\nu}$ , taking into account that  $\tilde{K}([x], [x]) = K(x, x)$  for all  $x \in X$ , and using the change of variables formula, we have

$$\int_{\tilde{X}} \tilde{K}([x], [x]) d\tilde{\nu}([x]) = \int_X K(x, x) d\nu(x).$$

Therefore  $\tilde{X}$ ,  $\tilde{K}$  and  $\tilde{\nu}$  satisfy the assumptions of Theorem 4.1. Hence,  $\tilde{K}$  can be written component-wise as

$$\tilde{K}([x], [t])_{jl} = \sum_{i \in I} \sigma_i \tilde{f}_i^l([t]) \overline{\tilde{f}_i^j([x])} \quad (26)$$

where  $(\sqrt{\sigma_i} \tilde{f}_i)_{i \in I}$  is basis of  $\ker L_{\tilde{\nu}}^\perp$  of eigenvectors of the integral operator  $L_{\tilde{\nu}}$  whose kernel is  $\tilde{K}$ . Furthermore  $(\sqrt{\sigma_i} \tilde{f}_i)_{i \in I}$  is an orthonormal family of  $\mathcal{H}_{\tilde{K}}$  with  $\tilde{f}_i$  continuous on  $\tilde{X}$ . Then  $f_i := W \tilde{f}_i$  is a continuous function on  $X$  thanks to the definition of  $\tau_K$  and  $W$  (see (24)), and  $(\sqrt{\sigma_i} f_i)_{i \in I}$  is an orthonormal part of  $\mathcal{H}_K$  by Proposition 5.2.iii). Moreover, for all  $f, g \in \mathcal{H}_K$ ,

$$\int_{\tilde{X}} \tilde{f}([x]) \tilde{g}([x]) d\tilde{\nu}([x]) = \int_X f(x) g(x) d\nu(x)$$

holds by definition of  $\tilde{\nu}$  and the change of variables formula, and thus  $\{f_i\}_{i \in I}$  is an orthonormal family in  $L^2(X, \nu; \mathbb{C}^n)$  as well. Note that  $\{f_i\}$  is also a basis of eigenvectors of  $L_\nu$  since  $(L_\nu f_i)(x) = (L_{\tilde{\nu}} \tilde{f}_i)([x])$  for all  $i \in I$ . The definition of  $W$  and equation (25) entail

$$K(x, t)_{jl} = \sum_{i \in I} \sigma_i f_i^l(t) \overline{f_i^j(x)}$$

for all  $x, t \in \text{supp } \nu = \text{supp } \mu$ .

Since the series in (26) is uniformly convergent on the compact subsets of  $\text{supp } \tilde{\nu} \times \text{supp } \tilde{\nu}$ , by Proposition 5.2.iv the latter series is uniformly convergent on the compact subsets of  $\text{supp } \mu \times \text{supp } \mu$ .

The unitary equivalence between  $\mathcal{H}_K$  and  $\mathcal{H}_{\tilde{K}}$  (through  $W$ ) implies that  $(\sqrt{\sigma_i} f_i)_{i \in I}$  is an orthonormal basis of  $\mathcal{H}_K$  if and only if  $(\sqrt{\sigma_i} \tilde{f}_i)_{i \in I}$  is an orthonormal basis of  $\mathcal{H}_{\tilde{K}}$ . Hence item *iii*) is a consequence of Theorem 4.1 and (25).

Finally, we prove item *iv*). First of all note that it is straightforward to see that every  $K_j$  given by (13) is a scalar kernel on  $X$ . Moreover, it satisfies

$$K_j(x, t) = \sum_{i \in I} \sigma_i f_i^j(t) \overline{f_i^j(x)} \quad \forall x, t \in X$$

thanks to equation (12).

Fix  $j = 1, \dots, n$  and set  $\gamma_x = (\sqrt{\sigma_i} f_i^j(x))_{i \in I} \in \ell^2(I)$  for all  $x \in X$ . Since

$$\langle \gamma_t, \gamma_x \rangle = \sum_{i \in I} \sigma_i f_i^j(t) \overline{f_i^j(x)} = K_j(x, t),$$

the function defined by

$$(W^j c)(x) := \langle c, \gamma_x \rangle = \sum_{i \in I} \sqrt{\sigma_i} c_i \overline{f_i^j(x)}, \quad c \in \ell^2(I),$$

is a partial isometry onto  $\mathcal{H}_{K_j}$  by Proposition 2.2. Therefore we have

$$\|f\|_j^2 = \|W^* f\|_2^2 = \sum_{i \in I} \sigma_i |\langle f, f_i^j \rangle_{K_j}|^2 \quad \forall f \in \mathcal{H}_{K_j},$$

i.e.  $\{\sqrt{\sigma_i} f_i^j\}_{i \in I}$  is a Parseval frame in  $\mathcal{H}_{K_j}$ . □

*Proof of Proposition 3.5.* Fix  $j = 1, \dots, n$  and let  $\{f_i^j\}_{i \in I}$  be a Parseval frame in  $\mathcal{H}_{K_j}$ . The function  $K$  given by (15) is a  $M_n(\mathbb{C})$ -reproducing kernel on  $X$  since

$$\begin{aligned} \sum_{l, r=1}^m \langle K(x_l, x_r) y_r, y_l \rangle &= \sum_{l, r=1}^m \sum_{p, q=1}^n K(x_l, x_r)_{pq} y_r^q \overline{y_l^p} \\ &= \sum_{l, r=1}^m \sum_{p, q=1}^n \sum_{i \in I} y_r^q \overline{y_l^p} f_i^q(x_r) \overline{f_i^p(x_l)} \\ &= \sum_{i \in I} \left| \sum_{r=1}^m \sum_{q=1}^n y_r^q f_i^q(x_r) \right|^2 \geq 0 \end{aligned}$$

for all  $x_1, \dots, x_m \in X$ ,  $y_1, \dots, y_m \in \mathbb{C}^n$ ,  $m \geq 1$ . Finally, we have

$$K_j(x, x) = \|(K_j)_x\|_{K_j}^2 = \sum_{i \in I} |f_i^j(x)|^2 = K(x, x)_{jj}$$

for all  $x \in X$ , so that  $K_j(x, t) = K(x, t)_{jj}$  for all  $x, t \in X$  by polarization identity.

Since all  $K_j$  are measurable, so is  $K$ . The fact that  $I$  is countable implies that each  $\mathcal{H}_{K_j}$  is separable, as well as  $\mathcal{H}_K$ .  $\square$

## 6. Appendix

We recall some basic facts about the embedding of a reproducing kernel Hilbert space into  $L^2(X, \nu; \mathbb{C}^n)$ .

**Theorem 6.1.** *Let  $X$  be a separable metric space and  $\nu$  a finite measure on  $\mathcal{B}(X)$ . Assume  $K : X \times X \rightarrow M_n(\mathbb{C})$  to be a continuous  $M_n(\mathbb{C})$ -reproducing kernel such that*

$$\int_X \text{Tr} K(x, x) d\nu(x) < +\infty. \quad (27)$$

The following facts hold true:

- i) every function in  $\mathcal{H}_K$  is continuous and  $\mathcal{H}_K$  is separable;
- ii) the canonical embedding

$$i_\nu : \mathcal{H}_K \hookrightarrow L^2(X, \nu; \mathbb{C}^n) \quad (28)$$

is a well defined compact operator. Its adjoint  $i_\nu^* : L^2(X, \nu; \mathbb{C}^n) \rightarrow \mathcal{H}_K$  is given by

$$i_\nu^* f = \sum_{j=1}^n \int_X K_x^j f^j(x) d\nu(x), \quad (29)$$

where the integrals converge in  $\mathcal{H}_K$ ;

- iii) the composition  $i_\nu i_\nu^* : L^2(X, \nu; \mathbb{C}^n) \rightarrow L^2(X, \nu; \mathbb{C}^n)$  is a positive trace class operator given by

$$(i_\nu i_\nu^* f)(x) = \int_X K(x, t) f(t) d\nu(t) = (L_\nu f)(x);$$

iv) there exist a family  $\{f_i\}_{i \in I}$  in  $L^2(X, \nu; \mathbb{C}^n)$  and a sequence  $\{\sigma_i\}_{i \in I}$  in  $(0, +\infty)$  such that  $\{f_i\}_{i \in I}$  is an orthonormal basis of  $\ker L_\nu^\perp = \overline{\text{Ran} L_\nu}$  and

$$L_\nu f_i = \sigma_i f_i \quad \forall i \in I.$$

v) every eigenfunction of  $L_\nu$  with strictly positive eigenvalue admits a representative in  $\mathcal{H}_K$ .

*Proof.* Set  $M := \int_X \text{Tr} K(x, x) d\nu(x) \in \mathbb{R}_+$ .

i) Given  $f \in \mathcal{H}_K$ , by the reproducing property we have

$$f(x) = (\langle f, K_x^1 \rangle_K, \dots, \langle f, K_x^n \rangle_K).$$

Since the  $j$ -th component of  $f$  coincides with the composition of the inner product in  $\mathcal{H}_K$  with the map  $x \mapsto K_x^j$ , which is clearly continuous, it follows that  $\mathcal{H}_K \subseteq \mathcal{C}(X, \mathbb{C}^n)$ . Moreover, since  $X$  is separable, there exists a countable set  $X_0$  dense in  $X$ . Hence  $\mathcal{H}_K$  is separable since  $\mathcal{S} = \{K_x^j : x \in X_0, j = 1, \dots, n\}$  is total in  $\mathcal{H}_K$ . Indeed, take  $f \in \mathcal{S}^\perp$ , then the reproducing property gives that  $f(x)^j = \langle f, K_x^j \rangle_K = 0$  for all  $x \in X_0$  and  $j = 1, \dots, n$ . Since  $f$  is continuous and  $X_0$  dense, it follows that  $f = 0$ , so that the claim is proved. ii). If  $f \in \mathcal{H}_K$ , then the following chain of inequalities holds:

$$\int_X \|f(x)\|^2 d\nu(x) \leq \int_X \langle K_x K_x^* f, f \rangle_K^2 d\nu(x) \leq \int_X \|f\|_K^2 \text{Tr} K(x, x) d\nu(x) \leq M \|f\|_K^2, \quad (30)$$

and the last quantity is finite by hypothesis. Thus  $i_\nu$  is well-defined and bounded. Moreover, if  $f \in L^2(X, \nu; \mathbb{C}^n)$ , we get

$$\langle i_\nu^* f, g \rangle_K = \langle f, g \rangle_2 = \int_X \langle f(x), K_x^* g \rangle d\nu(x) = \left\langle \int_X K_x f(x) d\nu(x), g \right\rangle_K,$$

where the integral  $\int_X K_x f(x) d\nu(x)$  converges in  $\mathcal{H}_K$  by the Hölder inequality, since

$$\int_X \|K_x f(x)\| d\nu(x) \leq \int_X (\text{Tr} K(x, x))^{1/2} \|f(x)\| d\nu(x),$$

the map  $x \mapsto (\text{Tr} K(x, x))^{1/2}$  is in  $L^2(X, \nu; \mathbb{C}^n)$  and  $f \in L^2(X, \nu; \mathbb{C}^n)$ . The component-wise representation in equation (29) follows by (2). The fact that  $i_\nu$  is compact follows from the property that  $i_\nu^* i_\nu$  is a trace class operator,

whose proof is given in the following item (the reasoning relies only on the fact that  $i_\nu$  is a bounded operator and on assumption (27)).

*iii)* The formula for  $L_\nu$  follows immediately using the expression for  $i_\nu^*$  obtained in item 1 and the fact that  $i_\nu$  is the canonical embedding. In order to prove that  $L_\nu$  is a Hilbert-Schmidt operator, we prove that in fact it is a trace class operator. Fix  $\{\varphi_\ell\}_{\ell \in \mathbb{N}}$  an orthonormal basis of  $\mathcal{H}_K$  and note that

$$\begin{aligned}
\text{Tr } L_\nu &= \text{Tr}(i_\nu^* i_\nu) = \sum_{\ell \in \mathbb{N}} \|i_\nu \varphi_\ell\|_2^2 = \sum_{\ell \in \mathbb{N}} \int_X \|\varphi_\ell(x)\|^2 d\nu(x) \\
&= \sum_{\ell \in \mathbb{N}} \int_X \sum_{j=1}^n \langle \varphi_\ell, K_x^j \rangle_K^2 d\nu(x) \\
&= \int_X \sum_{j=1}^n \sum_{\ell \in \mathbb{N}} \langle \varphi_\ell, K_x^j \rangle_K^2 d\nu(x) \\
&= \int_X \sum_{j=1}^n \|K_x^j\|_K^2 d\nu(x) \\
&= \int_X \sum_{j=1}^n K(x, x)_{jj} d\nu(x) \\
&= M.
\end{aligned}$$

*iv)*  $L_\nu$  is compact by *iii)*. and positive by construction, so there exist a basis of eigenvectors  $\{f_i\}_{i \in \mathbb{N}} \subseteq L^2(X, \nu; \mathbb{C}^n)$  with the associated sequence of positive eigenvalues  $\{\sigma_i\}_{i \in \mathbb{N}}$ . If we denote by  $I$  the set of indices corresponding to strictly positive eigenvalues, we have that  $\{f_i\}_{i \in I}$  is a basis of  $\ker L_\nu^\perp = \overline{\text{Ran } L_\nu}$ .

*v)* Given an eigenfunction  $f$  of  $L_\nu$  with corresponding eigenvalue  $\sigma > 0$ , define  $g = i_\nu^* f / \sigma \in \mathcal{H}_K$ , then

$$(i_\nu g)(x) = \frac{(L_\nu f)(x)}{\sigma} = f(x) \tag{31}$$

for  $\nu$ -almost all  $x \in X$ , *i.e.*  $g$  is a representative of  $f$  with  $g \in \mathcal{H}_K$ . □

- [1] Y. Lee, Y. Lin, G. Wahba, Multicategory support vector machines: theory and application to the classification of microarray data and satellite radiance data, *J. Amer. Statist. Assoc.* 99 (465) (2004) 67–81.

- [2] T. Evgeniou, C. A. Micchelli, M. Pontil, Learning multiple tasks with kernel methods, *J. Mach. Learn. Res.* 6 (2005) 615–637 (electronic).
- [3] A. Caponnetto, E. De Vito, Optimal rates for the regularized least-squares algorithm, *Found. Comput. Math.* 7 (3) (2007) 331–368.
- [4] L. Schwartz, Sous-espaces hilbertiens d’espaces vectoriels topologiques et noyaux associés (noyaux reproduisants), *J. Analyse Math.* 13 (1964) 115–256.
- [5] G. Pedrick, Theory of reproducing kernels for hilbert spaces of vector valued functions, Tech. rep., Kansas Univ Lawrence (1957).
- [6] C. A. Micchelli, M. Pontil, Learning the kernel function via regularization, *J. Mach. Learn. Res.* 6 (2005) 1099–1125.
- [7] A. Caponnetto, C. A. Micchelli, M. Pontil, Y. Ying, Universal multi-task kernels, *J. Mach. Learn. Res.* 9 (2008) 1615–1646.
- [8] C. Carmeli, E. De Vito, A. Toigo, Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem, *Anal. Appl. (Singap.)* 4 (4) (2006) 377–408.
- [9] C. Carmeli, E. De Vito, A. Toigo, V. Umanità, Vector valued reproducing kernel Hilbert spaces and universality, *Anal. Appl. (Singap.)* 8 (1) (2010) 19–61.
- [10] J. Mercer, Functions of positive and negative type and their connection with the theory of integral equations, *Phil. Trans. R. Soc. Lon. A* 209 (441–458) (1909) 415–446.
- [11] N. Dunford, L. Schwartz, *Linear Operators Part II Spectral Theory: Self Adjoint Operators in Hilbert Space*, Pure and Applied Mathematics, Wiley, New York, 1963.
- [12] F. Cucker, D.-X. Zhou, *Learning theory: an approximation theory viewpoint*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2007.
- [13] I. Steinwart, A. Christmann, *Support vector machines*, Information Science and Statistics, Springer, New York, 2008.

- [14] A. Berlinet, C. Thomas-Agnan, Reproducing kernel Hilbert spaces in probability and statistics, Kluwer Academic Publishers, Boston, MA, 2004.
- [15] A. Maurer, Unsupervised slow subspace-learning from stationary processes, *Theoret. Comput. Sci.* 405 (3) (2008) 237–255.
- [16] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.* 10 (1998) 12991319.
- [17] J. Shawe-Taylor, C. Williams, N. Cristianini, J. Kandola, On the eigen-spectrum of the gram matrix and the generalisation error of kernel PCA, *IEEE Transactions on Information Theory* 51 (7) (2005) 2510 – 2522.
- [18] H. Sun, Mercer theorem for RKHS on noncompact sets., *J. Complexity* 21 (3) (2005) 337–349.
- [19] D. Werner, *Functional analysis.* , Berlin: Springer-Verlag, 1995.
- [20] H. König, *Eigenvalue distribution of compact operators.*, Birkhäuser, 1986.
- [21] I. Steinwart, C. Scovel, Mercer’s theorem on general domains: on the interaction between measures, kernels, and RKHSs, *Constr. Approx.* (2012) DOI: 10.1007/s00365-012-9153-3.
- [22] S. Smale, D.-X. Zhou, Geometry on probability spaces, *Constr. Approx.* 30 (3) (2009) 311–323.
- [23] E. De Vito, L. Rosasco, A. Toigo, Spectral regularization for support estimation, in: J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, A. Culotta (Eds.), *Advances in Neural Information Processing Systems*, NIPS Foundation, 2010, pp. 1–9.
- [24] R. R. Coifman, S. Lafon, Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions, *Appl. Comput. Harmon. Anal.* 21 (1) (2006) 31–52.
- [25] L. Rosasco, M. Belkin, E. De Vito, On learning with integral operators, *J. Mach. Learn. Res.* 11 (2010) 905–934.