Appl. Comput. Harmon. Anal. 34 (2013) 339–351

Contents lists available at SciVerse ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

An extension of Mercer theorem to matrix-valued measurable kernels

Ernesto De Vito [a,b,*], Veronica Umanità [a], Silvia Villa [c,a]

[a] DIMA, Università di Genova, Italy
[b] I.N.F.N., Sezione di Genova, Italy
[c] Istituto Italiano di Tecnologia, Italy

ARTICLE INFO

Article history:
Received 17 October 2011
Revised 23 April 2012
Accepted 8 June 2012
Available online 13 June 2012
Communicated by Ding-Xuan Zhou

Keywords:
Reproducing kernel Hilbert spaces
Integral operators
Eigenvalues
Statistical learning theory

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 μ 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 μ. 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 μ is X.

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 $\mathbb{R}$ or $\mathbb{C}$, but recently the vector-valued setting is becoming popular, especially in machine learning because of its generality and its good experimental performance in a variety of different domains [1–3]. The mathematical theory for vector-valued RKHSs has been completely worked out in the seminal paper [4], which more generally 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 [6–9]. In particular, as in the scalar case the vector-valued RKHSs are completely characterized by the corresponding reproducing kernel, which 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 μ on X whose support is X, the integral operator L_μ 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

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

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

* Corresponding author at: DIMA, Università di Genova, Italy.
E-mail addresses: devito@dima.unige.it (E. De Vito), umanita@dima.unige.it (V. Umanità), villa@dima.unige.it (S. Villa).

1063-5203/$ – see front matter
http://dx.doi.org/10.1016/j.acha.2012.06.001

The kind of representation for the reproducing kernel plays a special role in the applications. For example, since the family $\{\sqrt{\sigma_i} f_i\colon\ \sigma_i > 0\}$ is an orthonormal basis of the corresponding RKHS $\mathcal{H}_K$, it provides a *canonical* feature map which relates the spectral properties of L_μ 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. The first group of results concerns scalar kernels. For example, [18] deals with the case of a σ-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_μ [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^∞, which is still too weak to get a pointwise representation [20]. The paper [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 μ 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 μ whose support is X, we construct a second measure ν such that the integral operator L_ν 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_ν, $\{\sigma_i\}_{i\in I}$ the corresponding family of eigenvalues and the series converges uniformly on compact subsets of $X \times X$. If the support of μ is a proper subset of X, representation (1) still holds true provided that $x, t \in \operatorname{supp}\mu$. 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. Appendix A collects some properties of the associated integral operator.

2. Preliminaries and notation

For any integer $n \geqslant 1$, 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 $T = (T)_{ij} \in M_n(\mathbb{C})$ we let $\|T\| = \sup\{\|Ty\|\colon\ y \in \mathbb{C}^n,\ \|y\| \leqslant 1\}$ be the operator norm, T^* is the adjoint and $\operatorname{Tr} T = \sum_{j=1}^n T_{jj}$ 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 σ-algebra $\mathcal{A}$ and a positive finite measure $\nu : \mathcal{A} \to [0,+\infty)$, then $L^2(X,\nu;\mathbb{C}^n)$ is the Hilbert space of (equivalence classes of) ν-square-