



ACADÉMIE  
DES SCIENCES  
INSTITUT DE FRANCE

# *Comptes Rendus*

---

## *Mathématique*

Fatima-Zahrae El-Boukkouri, Josselin Garnier and Olivier Roustant

**General reproducing properties in RKHS with application to derivative and integral operators**

Volume 364 (2026), p. 549-561

Online since: 27 July 2026

<https://doi.org/10.5802/crmath.847>



This article is licensed under the  
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.  
<http://creativecommons.org/licenses/by/4.0/>



*The Comptes Rendus. Mathématique are a member of the  
Mersenne Center for open scientific publishing*  
[www.centre-mersenne.org](http://www.centre-mersenne.org) — e-ISSN : 1778-3569



Research article  
Statistics

# General reproducing properties in RKHS with application to derivative and integral operators

Fatima-Zahrae El-Boukkouri<sup>\*,a</sup>, Josselin Garnier<sup>\*,b</sup> and Olivier Roustant<sup>\*,a</sup>

<sup>a</sup> Institut de Mathématiques de Toulouse, Université de Toulouse, INSA, 31077 Toulouse, France

<sup>b</sup> CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France

E-mails: el-boukkouri@insa-toulouse.fr, josselin.garnier@polytechnique.edu, roustant@insa-toulouse.fr

**Abstract.** In this paper, we consider the reproducing property in Reproducing Kernel Hilbert Spaces (RKHS). We establish a reproducing property for the closure of the class of combinations of composition operators under minimal conditions. This allows to revisit the sufficient conditions for the reproducing property to hold for the derivative operator, as well as for the existence of the mean embedding function. These results provide a framework of application of the representer theorem for regularized learning algorithms that involve data for function values, gradients, or any other operator from the considered class.

**Keywords.** RKHS, reproducing property, linear operator, derivative, mean embedding.

**2020 Mathematics Subject Classification.** 47B32.

**Funding.** This research has been done in the frame of the Chair PILEarnWater, part of the AI Cluster ANITI, funded by the France 2030 program under the Grant agreement ANR-23-IACL-0002.

*Manuscript received 7 April 2025, revised 26 January 2026, accepted 9 July 2026, online since 27 July 2026.*

## Introduction

Machine learning algorithms often involve penalized regression problems of the form

$$\min_{h \in \mathcal{H}} \sum_{i=1}^n (Lh(x_i) - y_i)^2 + \lambda \|h\|^2.$$

Here  $(x_i, y_i)_{1 \leq i \leq n}$  is a dataset where  $x_i$  belong to some set  $\mathbb{X}$  and  $y_i \in \mathbb{R}$ ,  $\lambda \in \mathbb{R}_+$  is a penalty coefficient,  $\mathcal{H}$  is a Hilbert space of real-valued functions on  $\mathbb{X}$ , and  $L$  is a linear operator from  $\mathcal{H}$  into  $\mathcal{F}(X, \mathbb{R})$ , the space of real-valued functions on  $\mathbb{X}$ . When  $\mathcal{H}$  is a reproducing kernel Hilbert space (RKHS), it turns out that the solution lives in a finite-dimensional space, a famous result known as “representer theorem” [1, 14, 17].

\*Corresponding author

A key property to prove this result is the generalized “reproducing property”

$$\forall x \in \mathbb{X}, \exists \tilde{L}(x) \in \mathcal{H} \text{ s.t. } \forall h \in \mathcal{H}, Lh(x) = \langle h, \tilde{L}(x) \rangle. \quad (1)$$

Note that, when  $\tilde{L}(x)$  exists, it is necessarily unique. Notice that in the particular case when  $L$  has its values in  $\mathcal{H}$  and admits an adjoint  $L^*$ , we simply have  $\tilde{L}(x) = L^*(K(x, \cdot))$ . When  $L$  is the identity operator, we have  $\tilde{L}(x) = K(x, \cdot)$  where  $K$  is the kernel associated to  $\mathcal{H}$ ,

$$\forall x \in \mathbb{X}, \forall h \in \mathcal{H}: \quad h(x) = \langle h, K(x, \cdot) \rangle,$$

which is the original reproducing property of RKHS [1]. This immediately extends to finite linear combinations, i.e. when  $L$  has the form

$$Lf(x) = \sum_{i=1}^q \alpha_i f(v_i(x)),$$

where  $\alpha_i \in \mathbb{R}$  and  $v_i(x) \in \mathbb{X}$ . In that case,  $\tilde{L}(x) = \sum_{i=1}^q \alpha_i K(v_i(x), \cdot)$ . However, the generalization to more complex operators such as derivative or integral operators is not straightforward as it involves a passage to the limit (see Section 1).

This question naturally arises in machine learning problems where the function to be learned is known to satisfy linear constraints. Typical examples include learning functions with prescribed mean or approximating solutions of linear partial differential equations from data [10]. In such settings, observations involve linear operators acting on the unknown function rather than pointwise evaluations, making the generalized reproducing property (1) a key theoretical tool.

In this paper, we consider a broad class of operators corresponding to limits of linear combinations and we give a necessary and sufficient condition on  $K$  for (1) to hold under the assumption that (1) remains true when passing to the limit. This allows us to revisit the reproducing properties in RKHS for derivative and integral operators. Firstly, we focus on the derivative operator, and show that the reproducing property holds if the cross derivative of the kernel exists and is continuous on the diagonal of  $\mathbb{X} \times \mathbb{X}$ . Thereby we retrieve the result presented in [13,15] by a different approach. We prove that this condition is less restrictive than the well-known  $C^2$  class condition [6,18] and we exhibit a counterexample.

Secondly, we consider the mean embedding in RKHS within the frame of (improper) Riemann integrals. We prove that it is properly defined if the kernel is absolutely integrable. This is similar to the result of [2,4,12] obtained for Lebesgue integrals. Then, we go a step further and show that, under the same condition, the reproducing property holds, which does not seem to be reported in the literature. We show that it is less restrictive than the condition that  $x \mapsto \sqrt{K(x, x)}$  is integrable [7,11] and we exhibit a counterexample.

## 1. Reproducing property for linear operators

A RKHS  $\mathcal{H}$  is a Hilbert space of real-valued functions defined on a set  $\mathbb{X}$  such that for all  $x \in \mathbb{X}$ , the evaluation  $\mathcal{H} \ni h \mapsto h(x)$  is continuous. Giving a RKHS is equivalent to giving a positive semidefinite function, or kernel,  $K$ . The link between  $\mathcal{H}$  and  $K$  is given by the reproducing property:

$$\forall f \in \mathcal{H}, \forall x \in \mathbb{X}: \quad f(x) = \langle f, K(x, \cdot) \rangle_{\mathcal{H}}.$$

This implies, by choosing  $f = K(y, \cdot)$ , that  $K(x, y) = \langle K(x, \cdot), K(y, \cdot) \rangle_{\mathcal{H}}$  for all  $x, y \in \mathbb{X}$ .

### 1.1. Main result

Before stating our main result in Theorem 4, we first recall the definition of so-called combination of composition operators, introduced in [9].

**Definition 1.** Let  $v: \mathbb{X} \rightarrow \mathbb{X}$  be an arbitrary function. The composition operator  $T_v$ , with function  $v$ , is defined by:

$$T_v: \mathcal{H} \longrightarrow \mathcal{F}(\mathbb{X}, \mathbb{R}), \quad T_v(f) := f \circ v.$$

We call a combination of composition operators with functions  $v_i$  and weights  $\alpha_i \in \mathbb{R}$  for  $1 \leq i \leq q$  the operator:

$$L = \sum_{i=1}^q \alpha_i T_{v_i}.$$

The class of combination of composition operators on functions on  $\mathcal{H}$  is denoted  $\mathcal{CC}(\mathcal{H})$ .

Let  $L = \sum_{i=1}^q \alpha_i T_{v_i} \in \mathcal{CC}(\mathcal{H})$ . Notice that for all  $x \in \mathbb{X}$ ,  $Lf(x) = \sum_{i=1}^q \alpha_i f(v_i(x))$ . Thus, the linear form  $\mathcal{H} \ni f \mapsto Lf(x)$  is continuous as a linear combination of evaluations of  $\mathcal{H}$ . Following [3, Section 4.4], we will denote by  $\tilde{L}(x) \in \mathcal{H}$  its representer, i.e. the unique element of  $\mathcal{H}$  verifying

$$Lf(x) = \langle f, \tilde{L}(x) \rangle. \quad (2)$$

We have explicitly

$$\tilde{L}(x) = \sum_{i=1}^q \alpha_i K(v_i(x), \cdot). \quad (3)$$

Mind that  $\tilde{L}(x) \neq L(K(x, \cdot)) = \sum_{i=1}^q \alpha_i K(x, v_i(\cdot))$ .

The class  $\mathcal{CC}(\mathcal{H})$  is limited to finite linear combinations. We now define a natural extension of this class by considering its closure with respect to pointwise convergence.

**Definition 2.** We define the class of operators

$$\mathcal{C}(\mathcal{H}) = \left\{ L: \mathcal{H} \rightarrow \mathcal{F}(\mathbb{X}, \mathbb{R}) : \exists (L_n)_{n \in \mathbb{N}} \in \mathcal{CC}(\mathcal{H})^{\mathbb{N}} \text{ s.t. } L(f) = \lim_{n \rightarrow +\infty} L_n(f) \ \forall f \in \mathcal{H} \right\},$$

where the limit is in the pointwise sense.

If  $L \in \mathcal{C}(\mathcal{H})$ , writing  $L_n = \sum_{i=1}^{q_n} \alpha_{i,n} T_{v_{i,n}}$  (where  $q_n \in \mathbb{N}$ ,  $\alpha_{i,n} \in \mathbb{R}$ ,  $v_{i,n} \in \mathcal{F}(\mathbb{X}, \mathbb{X})$ ), we have for all  $f \in \mathcal{H}$  and all  $x \in \mathbb{X}$ ,

$$L(f)(x) = \lim_{n \rightarrow +\infty} \sum_{i=1}^{q_n} \alpha_{i,n} f(v_{i,n}(x)).$$

As an example, when  $\mathbb{X} = \mathbb{R}$ , the class  $\mathcal{C}(\mathcal{H})$  includes derivative and integral operators:

- if the functions of  $\mathcal{H}$  are differentiable, the derivative operator  $L: f \mapsto f'$  is written as  $\lim_{n \rightarrow +\infty} L_n$  where  $L_n = nT_{v_{2,n}} - nT_{v_{1,n}}$  with  $v_{2,n}: x \mapsto x + \frac{1}{n}$  and  $v_{1,n}: x \mapsto x$ ; here  $\tilde{L}_n(x) = nK(v_{2,n}(x), \cdot) - nK(v_{1,n}(x), \cdot)$ ;
- if the functions of  $\mathcal{H}$  are continuous, the averaging operator  $L: f \mapsto x^{-1} \int_0^x f(t) dt$  (with  $L(f)(0) = f(0)$ ) is written as  $\lim_{n \rightarrow +\infty} L_n$  where  $L_n = \sum_{i=1}^n \frac{1}{n} T_{v_{i,n}}$  with  $v_{i,n}: x \mapsto \frac{i}{n}x$ ; here  $\tilde{L}_n(x) = \sum_{i=1}^n \frac{1}{n} K(v_{i,n}(x), \cdot)$ .

We now recall the Loève criterion for convergence of sequences in Hilbert spaces.

**Proposition 3 (Loève criterion).** Let  $\mathcal{H}$  be a Hilbert space with inner product  $\langle \cdot, \cdot \rangle$ . Let  $(h_n)_{n \in \mathbb{N}}$  be a sequence of  $\mathcal{H}$ . The following statements are equivalent:

- $(h_n)_{n \in \mathbb{N}}$  converges in  $\mathcal{H}$ ;
- the double sequence  $(\langle h_n, h_m \rangle)_{n,m \in \mathbb{N}}$  has a finite limit as  $n, m$  tend to  $+\infty$ .

In the whole paper the convergence of double sequences  $(u_{n,m})_{n,m \in \mathbb{N}}$  — such as the one appearing in (ii), is in the uniform sense (also known as Pringsheim sense [16]): if  $c$  is its limit,  $\forall \varepsilon > 0$ ,  $\exists n_0 \in \mathbb{N}$ ,  $\forall n, m \geq n_0$ ,  $|u_{n,m} - c| \leq \varepsilon$ .

**Proof.** The proof can be found in [5, Section 3.5] when  $\mathcal{H}$  is a space of square-integrable random variables. For completeness, we recall here the main arguments.

The direct sense (i)  $\Rightarrow$  (ii) is obvious by continuity of the scalar product.

Let us now prove (ii)  $\Rightarrow$  (i). Denote by  $c$  the limit of  $\langle h_n, h_m \rangle$  when  $n, m$  tend to infinity. We have, for all  $n, m \in \mathbb{N}$ ,

$$\|h_n - h_m\|^2 = \langle h_n, h_n \rangle + \langle h_m, h_m \rangle - 2\langle h_n, h_m \rangle. \quad (4)$$

This implies that  $\|h_n - h_m\|^2 \rightarrow c + c - 2c = 0$  when  $n, m \rightarrow +\infty$ . This proves that  $(h_n)$  is a Cauchy sequence, and thus converges in  $\mathcal{H}$ .  $\square$

For all operator  $L$  and for all function  $F: \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$  denote

$$L^\ell F: \mathbb{X} \times \mathbb{X} \longrightarrow \mathbb{R}, \quad (x, y) \longmapsto L(F(\cdot, y))(x),$$

and

$$L^r F: \mathbb{X} \times \mathbb{X} \longrightarrow \mathbb{R}, \quad (x, y) \longmapsto L(F(x, \cdot))(y).$$

Notice that with these notations, for all  $x, y \in \mathbb{X}$  we have  $(L^\ell F)(\cdot, y) = L(F(\cdot, y))$  and  $(L^r F)(x, \cdot) = L(F(x, \cdot))$ . We will also simply write  $L^\ell L^r$  instead of  $L^\ell \circ L^r$ .

**Theorem 4.** Let  $\mathcal{H}$  be a RKHS with kernel  $K$ . Let  $(L_n)_{n \in \mathbb{N}}$  be a sequence of  $\mathcal{CC}(\mathcal{H})$ . We denote by  $\widetilde{L}_n(x) \in \mathcal{H}$  the representer of the linear form  $\mathcal{H} \ni f \mapsto L_n f(x)$  (see (2)). The following conditions are equivalent:

- (i)  $(L_n^\ell L_m^r K(x, x))_{n, m}$  converges when  $n, m$  tend to  $+\infty$  for all  $x \in \mathbb{X}$ ;
- (ii)  $(\widetilde{L}_n(x))_{n \in \mathbb{N}}$  converges in  $\mathcal{H}$  for all  $x \in \mathbb{X}$ .

In that case, denote  $\widetilde{L}(x) := \lim_{n \rightarrow +\infty} \widetilde{L}_n(x) \in \mathcal{H}$ . Then, for all  $f \in \mathcal{H}$ ,  $L_n f$  converges pointwise. Define the operator  $L \in \mathcal{C}(\mathcal{H})$  by  $Lf(x) = \lim_{n \rightarrow +\infty} L_n f(x)$  (for all  $f \in \mathcal{H}$ ,  $x \in \mathbb{X}$ ). Then, for all  $x \in \mathbb{X}$  the mapping  $\mathcal{H} \ni f \mapsto Lf(x)$  is continuous and the reproducing property holds:

$$\forall x \in \mathbb{X}, \forall f \in \mathcal{H}: \quad Lf(x) = \langle f, \widetilde{L}(x) \rangle.$$

Finally,  $\|\widetilde{L}(x)\|^2 = L^\ell L^r K(x, x)$ .

**Proof.** Let  $x \in \mathbb{X}$ . We first prove that

$$\langle \widetilde{L}_n(x), \widetilde{L}_m(x) \rangle = L_n^\ell L_m^r K(x, x). \quad (5)$$

Let  $y \in \mathbb{X}$ . Denote  $F_m = L_m^r K$ . Using the definition (2) of  $\widetilde{L}_m(y)$  (with  $f = K(x, \cdot)$ ), we have

$$F_m(x, y) = (L_m K(x, \cdot))(y) = \langle K(x, \cdot), \widetilde{L}_m(y) \rangle = \widetilde{L}_m(y)(x). \quad (6)$$

Thus,  $F_m(\cdot, y) = \widetilde{L}_m(y)$  by uniqueness. Similarly, noting that  $F_m(\cdot, y) \in \mathcal{H}$  and using the definition (2) of  $\widetilde{L}_n(x)$  (with  $f = F_m(\cdot, y)$ ), we have

$$L_n^\ell F_m(x, y) = (L_n F_m(\cdot, y))(x) = \langle F_m(\cdot, y), \widetilde{L}_n(x) \rangle. \quad (7)$$

Finally, we get  $L_n^\ell L_m^r K(x, y) = \langle \widetilde{L}_m(y), \widetilde{L}_n(x) \rangle$ , which gives (5) when  $y = x$ .

From Equation (5), by the Loève criterion (Proposition 3), the sequence  $\widetilde{L}_n(x)$  converges in the Hilbert space  $\mathcal{H}$  for all  $x \in \mathbb{X}$  if and only if

$$\lim_{n, m \rightarrow \infty} L_n^\ell L_m^r K(x, x)$$

exists at each point  $x \in \mathbb{X}$ . This proves that (i) and (ii) are equivalent.

Suppose now that (i) or (ii) are verified and denote  $\widetilde{L}(x) := \lim_{n \rightarrow +\infty} \widetilde{L}_n(x) \in \mathcal{H}$ . Writing the equality (2) defining  $\widetilde{L}_n(x)$ , for a given  $f \in \mathcal{H}$  and  $x \in \mathbb{X}$ ,

$$L_n f(x) = \langle f, \widetilde{L}_n(x) \rangle,$$

and taking the limit when  $n$  tends to infinity, we obtain that  $L_n f(x)$  converges. Denoting by  $Lf(x)$  its limit, we thus have:

$$Lf(x) = \langle f, \widetilde{L}(x) \rangle.$$

This ensures that the mapping  $\mathcal{H} \ni f \mapsto Lf(x)$  is continuous.

Finally, by taking the limit in (5), we get:

$$\|\tilde{L}(x)\|^2 = L^\ell L^r K(x, x).$$

Indeed, let  $F = L^r K$  and  $x, y \in \mathbb{X}$ . By taking the limit in (6), we get  $F(\cdot, y) = \tilde{L}(y) \in \mathcal{H}$ . Then, taking the limit in (7), we obtain  $L^\ell F(x, y) = \langle F(\cdot, y), \tilde{L}(x) \rangle$ , or equivalently,

$$\langle \tilde{L}(x), \tilde{L}(y) \rangle = L^\ell L^r K(x, y). \quad \square$$

## 2. Derivative reproducing property

The reproducing property for the derivative operator has been extensively studied in the literature under various conditions. For example, in [18], this result is established for Mercer kernels of class  $C^2$ . A weaker condition is given in [13, Section 2] and [15, Corollary 4.36], namely that the cross derivative  $\frac{\partial^2 K}{\partial x_1 \partial x_2}$  exists and is continuous on  $\mathbb{X} \times \mathbb{X}$ . In what follows, we obtain a similar result, demanding the continuity at the neighborhood of the diagonal of  $\mathbb{X} \times \mathbb{X}$ , by following a different route with the class of compositions of operators. We argue that the continuity assumption, sometimes omitted [5,8], is important when the RKHS is not supposed a priori to contain differentiable functions.

**Theorem 5.** *Let  $\mathcal{H}$  be a RKHS of real-valued functions defined on a non-empty open set  $\mathbb{X} \subset \mathbb{R}$ , with reproducing kernel  $K$ . If  $K$  is of class  $C^1$  and  $\frac{\partial^2 K}{\partial x_1 \partial x_2}$  exists and is continuous on a neighborhood of the diagonal, then:*

- (a) *for all  $x \in \mathbb{X}$ ,  $\frac{\partial K}{\partial x_1}(x, \cdot)$  is in  $\mathcal{H}$ , and  $\left\| \frac{\partial K}{\partial x_1}(x, \cdot) \right\|^2 = \frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x)$ ;*
- (b) *all functions of  $\mathcal{H}$  are differentiable, the mapping  $\mathcal{H} \ni f \mapsto f'(x)$  is continuous, and the derivative reproducing property holds:*

$$\forall x \in \mathbb{X}, \forall f \in \mathcal{H}: \quad f'(x) = \left\langle f, \frac{\partial K}{\partial x_1}(x, \cdot) \right\rangle.$$

We will need the well-known fact that convergence in  $\mathcal{H}$  implies pointwise convergence (see [3]).

**Lemma 6.** *Let  $\mathcal{H}$  a RKHS on  $\mathbb{X}$  with kernel  $K$ , and let  $(h_x)_{x \in \mathbb{X}}$  be a family of functions of  $\mathcal{H}$ . If for some  $x_0 \in \mathbb{X}$ ,  $h_x$  converges in  $\mathcal{H}$  when  $x \rightarrow x_0$ , then  $h_x$  converges pointwise to the same limit.*

**Proof.** Let  $h_\infty \in \mathcal{H}$  be the limit of  $h_x$  when  $x \rightarrow x_0$ . Let  $y \in \mathbb{X}$ . By the reproducing property and the Cauchy–Schwarz inequality,

$$|h_x(y) - h_\infty(y)| = |\langle h_x - h_\infty, K(y, \cdot) \rangle| \leq \|h_x - h_\infty\| \|K(y, \cdot)\|.$$

The result follows.  $\square$

**Proof of Theorem 5.** Suppose that  $K$  is of class  $C^1$  and  $\frac{\partial^2 K}{\partial x_1 \partial x_2}$  exists and is continuous on a neighborhood of the diagonal.

Let  $(u_n)$  be a real-valued sequence converging to 0 as  $n$  tends to infinity with  $u_n \neq 0 \forall n \in \mathbb{N}$ .

Let  $n \in \mathbb{N}$  and  $x \in \mathbb{X}$  and consider the operator  $L_n \in \mathcal{CC}(\mathcal{H})$  defined by

$$L_n = \frac{T_{v_{2,n}} - T_{v_{1,n}}}{u_n}, \quad \text{where} \quad v_{2,n}(t) = t + u_n, \quad v_{1,n}(t) = t.$$

We have

$$L_n(f)(x) = \frac{f(x + u_n) - f(x)}{u_n} = \langle f, \tilde{L}_n(x) \rangle,$$

with

$$\tilde{L}_n(x) = \frac{K(x + u_n, \cdot) - K(x, \cdot)}{u_n}.$$

Notice that

$$L_n^\ell L_m^r K(x, x) = \frac{K(x + u_n, x + u_m) - K(x + u_n, x) - K(x, x + u_m) + K(x, x)}{u_n u_m}.$$

As  $\frac{\partial^2 K}{\partial x_1 \partial x_2}$  exists and is continuous on a neighborhood of the diagonal, the function  $(u', v') \mapsto \frac{\partial^2 K}{\partial x_1 \partial x_2}(x + u', x + v')$  is continuous on a neighborhood of  $(0, 0)$ . This justifies that there exists  $n_0 \in \mathbb{N}$  such that for all  $n, m \geq n_0$ ,

$$L_n^\ell L_m^r K(x, x) = \frac{\int_0^{u_n} \int_0^{u_m} \frac{\partial^2 K}{\partial x_1 \partial x_2}(x + u', x + v') du' dv'}{u_n u_m},$$

which can be obtained by two successive integrations with respect to  $u'$  and  $v'$  respectively. As  $\frac{\partial^2 K}{\partial x_1 \partial x_2}$  is continuous at  $(x, x)$ , then for a given  $\epsilon > 0$ , there exists  $r > 0$  such that for all  $u', v' \in [-r, r]$ ,

$$\left| \frac{\partial^2 K}{\partial x_1 \partial x_2}(x + u', x + v') - \frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x) \right| < \epsilon.$$

As the sequence  $(u_n)$  converges to 0, there exists  $N_0 \in \mathbb{N}$ , such that for all  $n \geq N_0$ ,  $u_n \in [-r, r]$ . Then for all  $n, m \geq \max(n_0, N_0)$ ,

$$\left| L_n^\ell L_m^r K(x, x) - \frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x) \right| \leq \left| \int_0^{u_n} \int_0^{u_m} \frac{\left| \frac{\partial^2 K}{\partial x_1 \partial x_2}(x + u', x + v') - \frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x) \right| du' dv'}{u_n u_m} \right| \leq \epsilon.$$

This implies that  $(L_n^\ell L_m^r K(x, x))_{n, m}$  converges when  $n, m$  tend to  $+\infty$  to  $\frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x)$ .

Then using Theorem 4, we conclude that  $(\widetilde{L}_n(x))_{n \in \mathbb{N}}$  converges in  $\mathcal{H}$ . Denoting by  $\widetilde{L}(x)$  its limit, we have by Lemma 6,

$$\frac{\partial K}{\partial x_1}(x, \cdot) = \widetilde{L}(x).$$

As the limit of  $\widetilde{L}_n(x)$  does not depend on the sequence  $u_n$ , we can conclude that  $\frac{K(x+t, \cdot) - K(x, \cdot)}{t}$  converges in  $\mathcal{H}$  when  $t$  tends to 0 to  $\frac{\partial K}{\partial x_1}(x, \cdot)$ .

To conclude the proof of (a), let us show that

$$\left\| \frac{\partial K}{\partial x_1}(x, \cdot) \right\|^2 = \frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x).$$

From Theorem 4, for all  $f \in \mathcal{H}$ ,  $L_n f$  converges pointwise, and with  $L f(x) = \lim_{n \rightarrow +\infty} L_n f(x)$ , we have

$$\|\widetilde{L}(x)\|^2 = L^\ell L^r K(x, x).$$

Recall that  $(L_n^\ell L_m^r K(x, x))_{n, m}$  converges to  $\frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x)$  when  $n, m$  tend to  $+\infty$ . In particular,  $(L_n^\ell L_n^r K(x, x))$  converges to  $\frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x)$  when  $n$  tends to  $+\infty$ . This implies that  $L^\ell L^r K(x, x) = \frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x)$ . And thus,

$$\left\| \frac{\partial K}{\partial x_1}(x, \cdot) \right\|^2 = \frac{\partial^2 K}{\partial x_1 \partial x_2}(x, x).$$

It remains to show (b). Let  $f \in \mathcal{H}$  and  $x \in \mathbb{X}$ . We know that  $\frac{\partial K}{\partial x_1}(x, \cdot)$  is in  $\mathcal{H}$ . We have

$$\frac{f(x+t) - f(x)}{t} = \left\langle f, \frac{K(x+t, \cdot) - K(x, \cdot)}{t} \right\rangle.$$

As  $\frac{K(x+t, \cdot) - K(x, \cdot)}{t}$  converges to  $\frac{\partial K}{\partial x_1}(x, \cdot)$  in  $\mathcal{H}$  when  $t$  tends to 0, we obtain that  $f$  is differentiable at  $x$  and

$$f'(x) = \left\langle f, \frac{\partial K}{\partial x_1}(x, \cdot) \right\rangle.$$

This implies that the mapping  $\mathcal{H} \ni f \mapsto f'(x)$  is continuous.  $\square$

**Remark 7.** Theorem 5 implies that if  $Y$  is a centered second-order random process with covariance function  $K$ , then  $Y$  is everywhere differentiable in quadratic mean (q.m.), the second-order cross derivative  $\frac{\partial^2 K}{\partial x_1 \partial x_2}$  exists everywhere and  $\text{cov}(Y'(s), Y'(t)) = \frac{\partial^2 K}{\partial x_1 \partial x_2}(s, t)$  [8].

**Remark 8.** It may be not sufficient to assume only the existence of the cross derivative on the diagonal (without assuming its continuity on a neighborhood of the diagonal). Indeed, in the proof of Theorem 5, we can write

$$L_n^\ell L_m^r K(x, x) = \frac{\frac{K(x+u_n, x+u_m) - K(x+u_n, x)}{u_m} - \frac{K(x, x+u_m) - K(x, x)}{u_m}}{u_n},$$

which leads to the fact that  $\lim_{n \rightarrow +\infty} \lim_{m \rightarrow +\infty} L_n^\ell L_m^r K(x, x)$  exists. Here, the limit is taken sequentially firstly in  $m$  and secondly in  $n$ . This does not imply that the limit of this double sequence exists when  $n, m$  tend to infinity, which is required by the Loève criterion (Proposition 3).

Finally, notice that the continuity of the cross derivative can be omitted if one assumes in addition from the outset that all functions in  $\mathcal{H}$  are differentiable, as proved in [2, Lemma 4].

Theorem 5 gives a sufficient condition on  $K$  for a derivative reproducing property to hold, namely the existence of the cross-derivative and its continuity on a neighborhood of the diagonal. This is less restrictive than the  $C^2$  condition on  $K$  found in the literature. Indeed, the following example exhibits a kernel  $K$  that is not of class  $C^2$  but such that  $\frac{\partial^2 K}{\partial x_1 \partial x_2}$  exists and is continuous.

**Example 9 (Non- $C^2$  kernel with finite cross derivative).** Consider the function:

$$f: \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) = \begin{cases} x^4 \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

We notice that  $f$  is twice differentiable on  $(-\infty, 0)$  and  $(0, \infty)$ . For  $x \in \mathbb{R} \setminus \{0\}$ , the first and second derivatives are:

$$f'(x) = 4x^3 \sin\left(\frac{1}{x}\right) - x^2 \cos\left(\frac{1}{x}\right)$$

and

$$f''(x) = 12x^2 \sin\left(\frac{1}{x}\right) - 6x \cos\left(\frac{1}{x}\right) - \sin\left(\frac{1}{x}\right).$$

Observe that

$$\frac{f(x) - f(0)}{x} = \frac{x^4 \sin\left(\frac{1}{x}\right)}{x} = x^3 \sin\left(\frac{1}{x}\right),$$

which proves that  $f$  is differentiable at  $x = 0$  with  $f'(0) = 0$ . Similarly,

$$\frac{f'(x) - f'(0)}{x} = \frac{f'(x)}{x} = \left(4x^2 \sin\left(\frac{1}{x}\right) - x \cos\left(\frac{1}{x}\right)\right),$$

which shows that  $f$  is twice differentiable at  $x = 0$  with  $f''(0) = 0$ . We conclude that  $f$  is twice differentiable on  $\mathbb{R}$ , with the second derivative given by:

$$f''(x) = \begin{cases} 12x^2 \sin\left(\frac{1}{x}\right) - 6x \cos\left(\frac{1}{x}\right) - \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

However,  $f''$  has no limit at 0, due to the term  $\sin\left(\frac{1}{x}\right)$ , which means that  $f$  is not  $C^2$ .

Now, consider the rank-one kernel  $K$ :

$$K: \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}, \quad K(x, y) = f(x)f(y).$$

$K$  is not of class  $C^2$  since for all  $y \in \mathbb{R}$ , the function  $x \mapsto \frac{\partial^2 K}{\partial x_1^2}(x, y) = f''(x)f(y)$  is not continuous at  $x = 0$ . Nevertheless, the cross derivative  $\frac{\partial^2 K}{\partial x_1 \partial x_2}(x, y) = f'(x)f'(y)$  exists for all  $x, y \in \mathbb{R}$  and is continuous.



### 3. Mean embedding in RKHS

In this subsection, we consider a random variable  $X$  defined on  $\mathbb{X} \subset \mathbb{R}$  with probability density function  $p$  and a RKHS  $\mathcal{H}$  of real-valued functions defined on  $\mathbb{X}$  with kernel  $K$ . We seek to establish the minimal condition under which there exists a function  $\mu_p \in \mathcal{H}$  satisfying:

$$\mathbb{E}_{X \sim p} f(X) = \langle f, \mu_p \rangle, \quad \forall f \in \mathcal{H}.$$

The function  $\mu_p$  is called the mean embedding of  $p$ .

Fukumizu et al. [7] proved that  $\mu_p$  exists if  $\mathbb{E}_{X \sim p} \sqrt{K(X, X)} < \infty$  in the sense of the Lebesgue integral. However, the function  $\mu_p$  exists under a less restrictive condition as we will demonstrate.

In what follows, our results will use the integral in the Riemann sense. To demonstrate it, we will need to use some notions related to the improper Riemann integral [19].

**Definition 10 (Admissible set).** A set  $E \subset \mathbb{R}^d$  is admissible if it is bounded and its boundary  $\partial E$  has Lebesgue measure zero. The integral of a function  $f$  over  $E$  is defined as:

$$\int_E f(x) dx = \int_I f(x) \mathbb{1}_E(x) dx,$$

where  $I$  is some interval in  $\mathbb{R}^d$  (i.e. a set of the form  $\{x \in \mathbb{R}^d, a^i \leq x_i \leq b^i, i = 1, \dots, d\}$ , for  $a^i < b^i \in \mathbb{R}$ ) such that  $E \subset I$ .

If the integral on the right-hand side of this equality exists, then we say that  $f$  is Riemann integrable over  $E$ . Lebesgue's criterion [19, Section 11.1.2, Theorem 1] for the existence of the Riemann integral over an interval implies that a function  $f: E \rightarrow \mathbb{R}$  is integrable over an admissible set  $E$  if and only if it is bounded and continuous at almost all points of  $E$  [19, Section 11.2.2, Theorem 1].

**Definition 11 (Exhaustion).** An exhaustion of a set  $E \subset \mathbb{R}^d$  is a sequence of admissible sets  $(E_s)_{s \in \mathbb{N}}$  such that:

$$E_s \subset E_{s+1} \subset E \quad \text{for all } s \in \mathbb{N},$$

and

$$\bigcup_{s=1}^{\infty} E_s = E.$$

The notion of exhaustion defines the improper integral over a set  $E \subset \mathbb{R}^d$ .

**Definition 12 (Riemann improper integral).** Let  $(E_s)_{s \in \mathbb{N}}$  be an exhaustion of the set  $E$  and suppose the function  $f: E \rightarrow \mathbb{R}$  is Riemann integrable on the  $E_s$  for all  $s \in \mathbb{N}$ . If the sequence  $\int_{E_s} f(x) dx$  converges and has a limit independent of the choice of the exhaustion, then  $f$  admits an improper integral over  $E$  defined by

$$\int_E f(x) dx = \lim_{s \rightarrow \infty} \int_{E_s} f(x) dx.$$

**Proposition 13 (Comparison test for improper integrals, [19]).** Let  $f$  and  $g$  be functions defined on  $E$  and integrable over exactly the same admissible subsets of  $E$ , and suppose  $|f| \leq g$  on  $E$ . If the improper integral  $\int_E g(x) dx$  exists, then the improper integrals  $\int_E |f(x)| dx$  and  $\int_E f(x) dx$  also exist.

Using improper integrals, we now establish the condition for the existence of the mean embedding  $\mu_p$  in  $\mathcal{H}$  when  $\mathbb{X}$  is not necessarily an interval. Note that  $\mathbb{X}$  can be unbounded and  $K$  can be unbounded.

**Theorem 14.** Let  $\mathcal{H}$  be a RKHS of real-valued functions defined on  $\mathbb{X}$ , with reproducing kernel  $K$ . Assume that  $K$  and  $p$  are continuous almost everywhere and locally bounded on  $\mathbb{X} \times \mathbb{X}$  and  $\mathbb{X}$  respectively, i.e.  $K$  (resp.  $p$ ) is bounded on all bounded subsets of  $\mathbb{X} \times \mathbb{X}$  (resp.  $\mathbb{X}$ ). Assume that the Riemann improper integral  $\int_{\mathbb{X} \times \mathbb{X}} |K(x, y)| p(x) p(y) dx dy$  exists. Then:

- (1) the mean embedding  $\mu_p: \mathbb{X} \ni x \mapsto \int_{\mathbb{X}} K(x, y) p(y) dy$  exists and is in  $\mathcal{H}$ ; moreover,  $\|\mu_p\|^2 = \iint_{\mathbb{X} \times \mathbb{X}} K(x, y) p(x) p(y) dx dy$ ;
- (2) for all function  $f \in \mathcal{H}$ , the function  $x \mapsto f(x) p(x)$  admits an improper integral on  $\mathbb{X}$  and the mapping  $\mathcal{H} \ni f \mapsto \mathbb{E}_{X \sim p}[f(X)] = \int_{\mathbb{X}} f(x) p(x) dx$  is continuous; furthermore, the reproducing property holds:

$$\forall f \in \mathcal{H}, \quad \mathbb{E}_{X \sim p}[f(X)] = \langle f, \mu_p \rangle.$$

**Proof.** Let  $(E_s)_{s \in \mathbb{N}}$  be an exhaustion of  $\mathbb{X}$ . For all  $s \in \mathbb{N}$ , as  $E_s$  is bounded, there exists an interval  $I_s$  in  $\mathbb{R}$  such that for all  $f \in \mathcal{H}$ ,

$$\int_{E_s} f(x) p(x) dx = \int_{I_s} f(x) \mathbb{1}_{E_s}(x) p(x) dx.$$

Let  $s \in \mathbb{N}$  and consider for all  $n \in \mathbb{N}^*$ , a partition  $(v_{i,n}^s)_{0 \leq i \leq n}$  of the interval  $I_s$  composed of  $n$  nodes. Suppose that the mesh of this partition  $\lambda_n$  defined as  $\lambda_n = \max_{1 \leq i \leq n} (v_{i,n}^s - v_{i-1,n}^s)$  converges to 0 as  $n$  tends to  $\infty$ . Let  $n \in \mathbb{N}^*$  and consider the following operator in  $\mathcal{CC}(\mathcal{H})$ ,

$$L_{s,n} = \sum_{i=1}^n (v_{i,n}^s - v_{i-1,n}^s) p(v_{i,n}^s) \mathbb{1}_{E_s}(v_{i,n}^s) T_{v_{i,n}^s}.$$

Thus, by (3),

$$\forall x \in \mathbb{X}, \quad \widetilde{L_{s,n}}(x) = \sum_{i=1}^n (v_{i,n}^s - v_{i-1,n}^s) p(v_{i,n}^s) \mathbb{1}_{E_s}(v_{i,n}^s) K(v_{i,n}^s, \cdot).$$

From now on, as  $L_{s,n}f(x)$  and  $\widetilde{L_{s,n}}(x)$  do not depend on  $x$ , we will omit  $x$  and simply write  $L_{s,n}f$  and  $\widetilde{L_{s,n}}$ .

Notice that for all  $n, m \in \mathbb{N}^*$ , and all  $x \in \mathbb{X}$ ,

$$L_{s,n}^\ell L_{s,m}^r K(x, x) = \sum_{i=1}^n \sum_{j=1}^m (v_{i,n}^s - v_{i-1,n}^s) p(v_{i,n}^s) (v_{j,m}^s - v_{j-1,m}^s) p(v_{j,m}^s) \mathbb{1}_{E_s \times E_s}(v_{i,n}^s, v_{j,m}^s) K(v_{i,n}^s, v_{j,m}^s).$$

From the assumptions on  $K$  and  $p$ , the function  $(x, y) \mapsto K(x, y) p(x) p(y)$  is continuous and bounded on the bounded set  $E_s \times E_s$ . Then, by [19, Section 11.2.2, Theorem 1] the Riemann integral  $\iint_{E_s \times E_s} K(x, y) p(x) p(y) dx dy$  exists. As the function  $x, y \mapsto K(x, y) p(x) p(y) \mathbb{1}_{E_s \times E_s}(x, y)$  is continuous almost everywhere on the interval  $I_s$ , the double sum above converges as  $n, m$  tend to  $+\infty$  to  $\int_{I_s \times I_s} K(x, y) p(x) p(y) \mathbb{1}_{E_s \times E_s}(x, y) dx dy$ , which is equal by Definition 10 to  $\int_{E_s \times E_s} K(x, y) p(x) p(y) dx dy$ . Then using Theorem 4 on the sequence  $(L_{s,n})_{n \in \mathbb{N}}$ , we conclude that  $(\widetilde{L_{s,n}})_{n \in \mathbb{N}}$  converges in  $\mathcal{H}$ . Denote by  $\widetilde{L_s}$  its limit. Furthermore, for all  $f \in \mathcal{H}$ ,  $L_{s,n}f$  converges pointwise with respect to  $n$ , and we denote  $L_s f = \lim_{n \rightarrow \infty} L_{s,n} f$ . Finally (still by Theorem 4),

$$\forall f \in \mathcal{H}, \quad L_s f = \langle f, \widetilde{L_s} \rangle. \quad (8)$$

For all  $x \in \mathbb{X}$ , the function  $y \mapsto K(x, y) p(y)$  is continuous almost everywhere and bounded on the bounded set  $E_s$ . Thus, it is Riemann integrable on  $E_s$  [19, Section 11.2.2, Theorem 1]. Then, for all  $x \in \mathbb{X}$ ,

$$\widetilde{L_s}(x) = \lim_{n \rightarrow \infty} \sum_{i=1}^n (v_{i,n}^s - v_{i-1,n}^s) p(v_{i,n}^s) \mathbb{1}_{E_s}(v_{i,n}^s) K(v_{i,n}^s, x) = \int_{E_s} K(x, y) p(y) dy.$$

Thus, by Lemma 6,

$$\widetilde{L_s} = \int_{E_s} K(\cdot, y) p(y) dy. \quad (9)$$

From (8), we obtain that for all  $f \in \mathcal{H}$ ,

$$L_s f = \lim_{n \rightarrow \infty} \sum_{i=1}^n (v_{i,n}^s - v_{i-1,n}^s) p(v_{i,n}^s) \mathbb{1}_{E_s}(v_{i,n}^s) f(v_{i,n}^s) = \langle f, \widetilde{L_s} \rangle.$$

As this limit does not depend on the subdivision  $(v_{i,n}^s)_{0 \leq i \leq n, n \in \mathbb{N}}$ , this implies that all functions of  $\mathcal{H}$  are  $p$ -integrable on  $E_s$  and,

$$\forall f \in \mathcal{H}, \quad \int_{E_s} f(x) p(x) dx = \langle f, \widetilde{L}_s \rangle.$$

In particular, for  $t \in \mathbb{N}$ , choosing  $f = \widetilde{L}_t$  (which is in  $\mathcal{H}$ ) and using (9), we have

$$\langle \widetilde{L}_s, \widetilde{L}_t \rangle = \int_{E_s} \left( \int_{E_t} K(x, y) p(x) p(y) dy \right) dx.$$

By the same argument used above, the Riemann integral  $\int_{E_s \times E_t} K(x, y) p(x) p(y) dx dy$  exists. By Fubini's theorem for Riemann integrals [19, Section 11.4], we have

$$\langle \widetilde{L}_s, \widetilde{L}_t \rangle = \int_{E_s \times E_t} K(x, y) p(x) p(y) dx dy. \quad (10)$$

Let us show that when  $s, t$  tend to  $+\infty$ ,  $\langle \widetilde{L}_s, \widetilde{L}_t \rangle$  converges to  $\int_{\mathbb{X} \times \mathbb{X}} K(x, y) p(x) p(y) dx dy$ . Let  $s, t \in \mathbb{N}$ , as  $x, y \mapsto |K(x, y)| p(x) p(y)$  is integrable on  $\mathbb{X} \times \mathbb{X}$  and using Proposition 13 with  $f(x, y) = K(x, y) p(x) p(y) \mathbb{1}_{\mathbb{X} \times \mathbb{X} \setminus E_s \times E_t}(x, y)$  and  $g(x, y) = |K(x, y)| p(x) p(y)$  for all  $x, y \in \mathbb{X}$ , then  $\int_{\mathbb{X} \times \mathbb{X} \setminus E_s \times E_t} K(x, y) p(x) p(y) dx dy$  and  $\int_{\mathbb{X} \times \mathbb{X} \setminus E_s \times E_t} |K(x, y)| p(x) p(y) dx dy$  exist. Let us denote  $I(s, t) = \int_{E_s \times E_t} K(x, y) p(x) p(y) dx dy$  and  $I = \int_{\mathbb{X} \times \mathbb{X}} K(x, y) p(x) p(y) dx dy$ . We then have,

$$\begin{aligned} |I(s, t) - I| &= \left| \int_{\mathbb{X} \times \mathbb{X} \setminus E_s \times E_t} K(x, y) p(x) p(y) dx dy \right| \\ &\leq \int_{\mathbb{X} \times \mathbb{X} \setminus E_s \times E_t} |K(x, y)| p(x) p(y) dx dy \\ &= \int_{\mathbb{X} \times \mathbb{X}} |K(x, y)| p(x) p(y) dx dy - \int_{E_s \times E_t} |K(x, y)| p(x) p(y) dx dy. \end{aligned}$$

As above, since  $E_{\min(s,t)} \times E_{\min(s,t)} \subset E_s \times E_t$ , Proposition 13 implies that

$$\int_{E_{\min(s,t)} \times E_{\min(s,t)}} |K(x, y)| p(x) p(y) dx dy$$

exists and

$$|I(s, t) - I| \leq \int_{\mathbb{X} \times \mathbb{X}} |K(x, y)| p(x) p(y) dx dy - \int_{E_{\min(s,t)} \times E_{\min(s,t)}} |K(x, y)| p(x) p(y) dx dy.$$

As  $(E_s \times E_s)_{s \in \mathbb{N}}$  is an exhaustion of  $\mathbb{X} \times \mathbb{X}$ ,  $\int_{E_{\min(s,t)} \times E_{\min(s,t)}} |K(x, y)| p(x) p(y) dx dy$  converges to  $\int_{\mathbb{X} \times \mathbb{X}} |K(x, y)| p(x) p(y) dx dy$  as  $s, t$  tend to  $+\infty$ . Thus,  $I(s, t) = \langle \widetilde{L}_s, \widetilde{L}_t \rangle$  converges to  $I = \int_{\mathbb{X} \times \mathbb{X}} K(x, y) p(x) p(y) dx dy$  as  $s, t$  tend to  $+\infty$ .

By applying Loève criterion (Proposition 3) to the sequence  $(\widetilde{L}_s)_{s \in \mathbb{N}}$ , we conclude that  $(\widetilde{L}_s)$  converges in  $\mathcal{H}$ . Let  $\mu_p$  denote its limit. Notice that for all  $x \in \mathbb{X}$ , we have

$$\mu_p(x) = \lim_{s \rightarrow \infty} \int_{E_s} K(x, y) p(y) dy.$$

Thus, for all  $x \in \mathbb{X}$ , the sequence  $(\int_{E_s} K(x, y) p(y) dy)$  converges as  $s$  tends to  $+\infty$  and its limit does not depend on the exhaustion  $(E_s)$ . This implies that  $\int_{\mathbb{X}} K(x, y) p(y) dy$  is finite for all  $x \in \mathbb{X}$ , and by Lemma 6, we deduce that

$$\mu_p = \int_{\mathcal{X}} K(\cdot, y) p(y) dy.$$

To conclude the proof of (1), recall that  $I(s, t) \rightarrow I$  as  $s$  and  $t$  tend to  $+\infty$ . This implies that  $I(s, s) \rightarrow I$  when  $s$  tends to  $+\infty$ . As  $\widetilde{L}_s$  converges in  $\mathcal{H}$  to  $\mu_p$ , we conclude that

$$\|\mu_p\|^2 = \int_{\mathbb{X} \times \mathbb{X}} K(x, y) p(x) p(y) dx dy.$$

It remains to prove (2). Let  $f \in \mathcal{H}$ . From the proof of (1), we have

$$\int_{E_s} f(x)p(x) dx = \langle f, \widetilde{L}_s \rangle.$$

As  $(\widetilde{L}_s)$  converges in  $\mathcal{H}$  to  $\mu_p$  when  $s$  tends to  $+\infty$ , we obtain that  $(\int_{E_s} f(x)p(x) dx)$  converges to  $\langle f, \mu_p \rangle$  as  $s$  tends to  $+\infty$ . As the limit does not depend on the chosen exhaustion, we conclude that  $f p$  admits an improper integral on  $\mathbb{X}$  and,

$$\int_{\mathbb{X}} f(x)p(x) dx = \langle f, \mu_p \rangle.$$

This implies that the mapping  $\mathcal{H} \ni f \mapsto \int_{\mathbb{X}} f(x)p(x) dx$  is continuous.  $\square$

**Remark 15.** Theorem 14 contains two statements. The first one, (1), is that the function  $\mu_p: \mathbb{X} \ni x \mapsto \int_{\mathbb{X}} K(x, y)p(y) dy$  exists and belongs to  $\mathcal{H}$ . This result is the analogue for improper Riemann integrals of that established for Lebesgue integrals (see [2, Proposition 2.3] and [12]). Under the same condition, we further demonstrate in (2) that the reproducing property holds, with a less restrictive condition than in [7, 11], as we will show in Proposition 16. Notice that (2) is not simply deduced from (1) by continuity of the scalar product in  $\mathcal{H}$  since the interchange of limit and integral is not guaranteed in general.

**Proposition 16.** *Let  $\mathcal{H}$  be a RKHS of real-valued function defined on  $\mathbb{X}$ , with reproducing kernel  $K$ . Assume that  $K$  and  $p$  are continuous almost everywhere and locally bounded on  $\mathbb{X} \times \mathbb{X}$ . If the standard variation  $\int_{\mathbb{X}} \sqrt{K(x, x)}p(x) dx$  exists, then  $\iint_{\mathbb{X} \times \mathbb{X}} |K(x, y)|p(x)p(y) dx dy$  is also finite.*

**Proof.** Let  $(E_n)$  be an exhaustion of  $\mathbb{X}$ . Recall that  $K(x, y) = \langle K(x, \cdot), K(y, \cdot) \rangle_{\mathcal{H}}$  for all  $x, y \in \mathbb{X}$ . Using Cauchy–Schwarz inequality, we then obtain:

$$\int_{E_n \times E_n} |K(x, y)|p(x)p(y) dx dy \leq \int_{E_n \times E_n} \sqrt{K(x, x)}p(x)\sqrt{K(y, y)}p(y) dx dy.$$

Since  $E_n \times E_n$  is an admissible subset of  $\mathbb{X} \times \mathbb{X}$ , applying Fubini's theorem for Riemann integrals (cf. [19, Section 11.4]) gives:

$$\begin{aligned} \int_{E_n \times E_n} \sqrt{K(x, x)}p(x)\sqrt{K(y, y)}p(y) dx dy &= \int_{E_n} \int_{E_n} \sqrt{K(x, x)}p(x)\sqrt{K(y, y)}p(y) dx dy \\ &= \left( \int_{E_n} \sqrt{K(x, x)}p(x) dx \right)^2. \end{aligned}$$

Thus, if  $\int_{\mathbb{X}} \sqrt{K(x, x)}p(x) dx$  is finite, it follows that  $\iint_{\mathbb{X} \times \mathbb{X}} |K(x, y)|p(x)p(y) dx dy$  is also finite.  $\square$

The minimal requirement for the existence of the mean embedding  $\mu_p$  is that  $\iint_{\mathbb{X} \times \mathbb{X}} |K(x, y)|p(x)p(y) dx dy$  is finite. This is weaker than the criterion found in the literature that requires  $\int_{\mathbb{X}} \sqrt{K(x, x)}p(x) dx$  to be finite, as shown by the previous proposition. Note, however, that the (almost sure) continuity of the kernel is also required in Theorem 14, which is a mild assumption. Finally, the following example exhibits a kernel  $K$  for which  $\int_{\mathbb{X}} \sqrt{K(x, x)}p(x) dx$  is not finite but such that  $\iint_{\mathbb{X} \times \mathbb{X}} |K(x, y)|p(x)p(y) dx dy$  is finite.

**Example 17 (Kernel with non-finite standard deviation and well-defined mean embedding).**

Let  $p: x \mapsto \frac{1}{x^2} \mathbb{1}_{x \geq 1}$  and let us consider the following function:

$$K: [1, \infty) \times [1, \infty) \longrightarrow \mathbb{R}, \quad K(x, y) = xy e^{-(x-y)^2}.$$

It is a positive semidefinite kernel as it is the covariance function of the process  $xZ(x)$  where  $Z$  is a stationary process with mean zero and Gaussian covariance function  $\exp(-(x-y)^2)$ . The standard deviation is clearly infinite:

$$\mathbb{E}_{X \sim p} [\sqrt{K(X, X)}] = \int_1^{+\infty} \frac{1}{x} dx = +\infty.$$

Let us now show that  $I := \int_{[1,\infty)^2} |K(x, y)| p(x) p(y) dx dy$  exists.

As the function  $x, y \mapsto K(x, y) p(x) p(y)$  is positive, it is enough to prove that for only one specific exhaustion  $(E_n)$  of the set  $[1, \infty)$ , the sequence  $I_n := \iint_{E_n^2} K(x, y) p(x) p(y) dx dy$  converges to  $I$ . Indeed, the positivity guarantees that the result will then be valid for all exhaustions [19, Section 11.6, Proposition 1]. Here, we choose  $E_n = [1, n]$ .

To compute  $I_n$ , let us split the cubic domain  $E_n \times E_n$  into two triangles:

$$I_n = \int_1^n \int_1^n \frac{1}{xy} e^{-(x-y)^2} dx dy = \int_1^n \int_y^n \frac{1}{xy} e^{-(x-y)^2} dx dy + \int_1^n \int_1^y \frac{1}{xy} e^{-(x-y)^2} dx dy.$$

Using the symmetry of the integrand, we can see that the two terms are equal, and thus

$$I_n = 2 \int_1^n \int_y^n \frac{1}{xy} e^{-(x-y)^2} dx dy.$$

Now let us apply the change of variables  $T: \mathbb{R}^2 \ni (x, y) \mapsto (w = x - y, y)$ . As for all  $x, y \in \mathbb{R}$ ,

$$1 \leq y \leq n \quad \text{and} \quad y \leq x \leq n \quad \Longleftrightarrow \quad 0 \leq w \leq n - 1 \quad \text{and} \quad 1 \leq y \leq n - w,$$

we have

$$I_n = 2 \int_0^{n-1} \int_1^{n-w} \frac{1}{y(w+y)} e^{-w^2} dy dw.$$

Now, for all  $w \in [0, n-1]$ , we have

$$0 \leq \int_1^{n-w} \frac{1}{y(w+y)} dy \leq \int_1^{n-w} \frac{1}{y^2} dy = \left[ -\frac{1}{y} \right]_1^{n-w} = 1 - \frac{1}{n-w} \leq 1.$$

Therefore,

$$I_n \leq 2 \int_0^{n-1} e^{-w^2} dw \leq 2 \int_0^{+\infty} e^{-w^2} dw = \sqrt{\pi}.$$

Finally, the sequence  $(I_n)_{n \geq 1}$ , which is monotonically increasing, is bounded. Hence it converges as  $n \rightarrow +\infty$ . Therefore,  $\iint_{[1,\infty)^2} |K(x, y)| p(x) p(y) dx dy$  is finite, even though  $\mathbb{E}_{X \sim p} [\sqrt{K(X, X)}]$  is not finite.

## Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

## References

- [1] N. Aronszajn, “Theory of reproducing kernels”, *Trans. Am. Math. Soc.* **68** (1950), pp. 337–404.
- [2] A. Barp, C.-J. Simon-Gabriel, M. Girolami and L. Mackey, “Targeted separation and convergence with kernel discrepancies”, *J. Mach. Learn. Res.* **25** (2024), no. 378, pp. 1–50.
- [3] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*, Kluwer Academic Publishers, 2004.
- [4] C. Carmeli, E. De Vito and A. Toigo, “Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem”, *Anal. Appl., Singap.* **4** (2006), no. 04, pp. 377–408.
- [5] H. Cramér and M. R. Leadbetter, *Stationary and related stochastic processes. Sample function properties and their applications*, Courier Corporation, 2013.
- [6] J. C. Ferreira and V. A. Menegatto, “Positive definiteness, reproducing kernel Hilbert spaces and beyond”, *Ann. Funct. Anal.* **4** (2013), no. 1, pp. 64–88.

- [7] K. Fukumizu, F. R. Bach and M. I. Jordan, “Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces”, *J. Mach. Learn. Res.* **5** (2003/04), pp. 73–99.
- [8] C. Gaetan and X. Guyon, “Second-order spatial models and geostatistics”, in *Spatial statistics and modeling*, Springer Series in Statistics, Springer, 2010, pp. 1–52.
- [9] D. Ginsbourger, N. Durrande and O. Roustant, “Kernels and designs for modelling invariant functions: from group invariance to additivity”, in *mODa 10 — Advances in model-oriented design and analysis: Proceedings of the 10th International Workshop in Model-Oriented Design and Analysis held in Łagów Lubuski, Poland, June 10–14, 2013* (D. Ucinski, A. C. Atkinson and M. Patan, eds.), Contributions to Statistics, Springer, 2013, pp. 107–115.
- [10] F. Lindgren, D. Bolin and H. Rue, “The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running”, *Spatial Stat.* **50** (2022), article no. 100599 (29 pages).
- [11] K. Muandet, K. Fukumizu, B. Sriperumbudur and B. Schölkopf, “Kernel mean embedding of distributions: a review and beyond”, *Found. Trends Mach. Learn.* **10** (2017), no. 1–2, pp. 1–141.
- [12] C. J. Oates, “Minimum kernel discrepancy estimators”, in *Proceedings of the 15th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (A. Hinrichs, P. Kritzer and F. Pillichshammer, eds.), Springer, 2022, pp. 133–161.
- [13] S. Saitoh and Y. Sawano, *Theory of reproducing kernels and applications*, Developments in Mathematics, vol. 44, Springer, 2016, pp. 71–78.
- [14] B. Schölkopf, R. Herbrich and A. J. Smola, “A generalized representer theorem”, in *Computational learning theory (Amsterdam, 2001)* (D. Helmbold and B. Williamson, eds.), Lecture Notes in Computer Science, vol. 2111, Springer, 2001, pp. 416–426.
- [15] I. Steinwart and A. Christmann, *Support vector machines*, Information Science and Statistics, Springer, 2008.
- [16] B. Tripathy and B. C. Tripathy, “On I-convergent double sequences”, *Soochow J. Math.* **31** (2005), no. 4, pp. 549–560.
- [17] G. Wahba, *Spline models for observational data*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 59, Society for Industrial and Applied Mathematics, 1990.
- [18] D.-X. Zhou, “Derivative reproducing properties for kernel methods in learning theory”, *J. Comput. Appl. Math.* **220** (2008), no. 1–2, pp. 456–463.
- [19] V. A. Zorich, *Mathematical analysis. II*, Universitext, Springer, 2016.