Kernel embedding of measures and low-rank approximation of integral operators

Bertrand Gauthier

Kernel embedding of measures and low-rank approximation of integral operators

Bertrand GAUTHIER*

Abstract

We describe a natural coisometry from the Hilbert space of all Hilbert-Schmidt operators on a separable reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ and onto the RKHS $\mathcal{G}$ associated with the squared-modulus of the reproducing kernel of $\mathcal{H}$. Through this coisometry, trace-class integral operators defined by general measures and the reproducing kernel of $\mathcal{H}$ are isometrically represented as potentials in $\mathcal{G}$, and the quadrature approximation of these operators is equivalent to the approximation of integral functionals on $\mathcal{G}$. We then discuss the extent to which the approximation of potentials in RKHSs with squared-modulus kernels can be regarded as a differentiable surrogate for the characterisation of low-rank approximation of integral operators.

Keywords: reproducing kernel Hilbert spaces, integral operators, low-rank approximation, kernel embedding of measures, differentiable relaxation.

Mathematics Subject Classification: 46E22, 47B32, 47G10, 65F55.

Contents

1 Introduction **2**

2 Framework, notations and basic properties **3**
2.1 RKHSs and Hilbert-Schmidt operators 4
2.2 Natural coisometry from HS($\mathcal{H}$) onto $\mathcal{G}$ 5

3 Trace-class integral operators with PSD kernels **7**
3.1 Integral operators and kernel embedding of measures 7
3.2 Quadrature approximation 10
3.3 Further properties 10

*gauthierb@cardiff.ac.uk
Cardiff University, School of Mathematics
Abacws, Senghennydd Road, Cardiff, CF24 4AG, United Kingdom
ORCID: 0000-0001-5469-814X

4 Measures and projection-based approximations **11**
4.1 Additional notations and general properties 12
4.2 Projections defined by measures 13
4.3 Error maps on sets of measures 13

5 Nonnegative measures and L^2-embeddings **16**
5.1 Partial L^2-embeddings 16
5.2 Trace and Frobenius error maps 18
5.3 Column sampling for PSD-matrix approximation 19

6 Concluding discussion **19**

A Technical results **20**

1 Introduction

Integral operators with positive-semidefinite (PSD) kernels play a central role in the theory of reproducing kernel Hilbert spaces (RKHSs) and their applications; see for instance [4, 5, 19, 26, 20]. As an important instance, this class of operators encompasses the PSD matrices.

Under suitable conditions, an integral operator defined by a PSD kernel K and a measure μ can be regarded as a Hilbert-Schmidt (HS) operator L_μ on the RKHS $\mathcal{H}$ associated with K; see e.g. [21, 22, 20]. Let $\mathcal{G}$ be the RKHS for which the squared-modulus kernel $|K|^2$ is reproducing. Following [10], when the integral of the diagonal of K with respect to the variation of μ is finite, the HS operator L_μ on $\mathcal{H}$ can be isometrically represented as the Riesz representation $g_\mu \in \mathcal{G}$ of the integral functional on $\mathcal{G}$ defined by the measure $\overline{\mu}$, the conjugate of μ. The operator L_μ is in this case trace-class, and g_μ is the *potential*, or *kernel embedding*, of the measure μ in the RKHS $\mathcal{G}$. In the Hilbert space $\mathrm{HS}(\mathcal{H})$ of all HS operators on $\mathcal{H}$, the quadrature approximation of trace-class integral operators with kernel K is hence equivalent to the approximation of integral functionals on $\mathcal{G}$. Considering another measure ν and denoting by $B_\mathcal{G}$ the closed unit ball of $\mathcal{G}$, we more specifically have

$$\|L_\mu - L_\nu\|_{\mathrm{HS}(\mathcal{H})} = \|g_\mu - g_\nu\|_\mathcal{G} = \sup_{g \in B_\mathcal{G}} \left| \int g(t) \mathrm{d}\mu(t) - \int g(t) \mathrm{d}\nu(t) \right|,$$

so that the map $(\mu, \nu) \mapsto \|L_\mu - L_\nu\|_{\mathrm{HS}(\mathcal{H})}$ corresponds to a generalised *integral probability metric*, or *maximum mean discrepancy* (see e.g. [16, 24, 2, 15, 27]).

We give an overall description of the framework surrounding such an isometric representation, and illustrate that it follows from the definition of a *natural coisometry* Γ from $\mathrm{HS}(\mathcal{H})$ onto $\mathcal{G}$; this coisometry maps self-adjoint operators to real-valued functions, and PSD operators to nonnegative functions (Section 2). Under adequate measurability conditions on K and assuming that the diagonal of K is integrable with respect to $|\mu|$, we show that L_μ always belongs to the initial space of Γ, and

that $\Gamma[L_\mu] = g_\mu$. We then describe the equivalence between the quadrature approximation of integral operators with PSD kernels and the approximation of potentials in RKHSs with squared-modulus kernels (Section 3).

For an approximate measure ν, and denoting by $\mathcal{H}_\nu$ the closure in $\mathcal{H}$ of the range of $L_{|\nu|}$ (so that when ν is finitely-supported, $\mathcal{H}_\nu$ is fully characterised by the support of ν), we next investigate the extent to which the approximation of potentials in $\mathcal{G}$ can be used as a differentiable surrogate for the characterisation of approximations of L_μ of the form $P_\nu L_\mu$, $L_\mu P_\nu$ or $P_\nu L_\mu P_\nu$, with P_ν the orthogonal projection from $\mathcal{H}$ onto $\mathcal{H}_\nu$ (Section 4). When the measure μ is nonnegative, the operator L_μ admits the decomposition $L_\mu = \imath_\mu^* \imath_\mu$, with $\imath_\mu$ the natural embedding of $\mathcal{H}$ in $L^2(\mu)$. The three operators

$$\imath_\mu^* : L^2(\mu) \to \mathcal{H}, \quad \imath_\mu \imath_\mu^* : L^2(\mu) \to L^2(\mu), \quad \text{and} \quad \imath_\mu \imath_\mu^* \imath_\mu : \mathcal{H} \to L^2(\mu),$$

can then also be regarded as integral operators defined by the kernel K and the measure μ, and through the *partial embedding* $\imath_\mu P_\nu$, a measure ν characterises approximations of each of these operators. We study the properties of these approximations and further illustrate the connections between the low-rank approximation of integral operators with PSD kernels and the approximation of potentials in RKHSs with squared-modulus kernels. We also describe the link between the considered framework and the low-rank approximation of PSD matrices through column sampling (Section 5). The presentation ends with a concluding discussion (Section 6) and some technical results are gathered in appendix (Appendix A). The approximation schemes considered in this note should be apprehended from the the point of view of numerical strategies for discretisation or dimension reduction; in practical applications, approximations will generally be characterised by finitely-supported measures.

2 Framework, notations and basic properties

By default, all the Hilbert spaces considered in this note are complex; they are otherwise explicitly referred to as real Hilbert spaces; we use a similar convention for vector spaces. Inner products of complex Hilbert spaces are assumed to be linear with respect to their right argument. For $z \in \mathbb{C}$, we denote by $\overline{z}$, $|z|$ and $\mathfrak{R}(z)$ the conjugate, modulus and real part of z, respectively, and $\mathrm{i} \in \mathbb{C}$ is the imaginary unit. By analogy, for a complex-valued function f on a general set S, we denote by $\overline{f}$ and $|f|$ the functions defined as $\overline{f}(s) = \overline{f(s)}$ and $|f|(s) = |f(s)|$, $s \in S$; we also use the notation $|f|^2$ to refer to the function $s \mapsto |f(s)|^2$.

For two Hilbert spaces H and F, we denote by A^* the adjoint of a bounded linear operator $A : H \to F$. The map A is an *isometry* if $A^*A = \mathrm{id}_H$, the identity operator on H, and A is a *coisometry* if A^* is an isometry (and so $AA^* = \mathrm{id}_F$). A coisometry A is a surjective *partial isometry* (that is, $AA^*A = A$), and A^*A is then the orthogonal projection from H onto the *initial space* $\mathcal{I}(A)$ of A, with $\mathcal{I}(A)$ the orthogonal complement in H of the nullspace of A. We denote by $\mathrm{null}(A)$ the

nullspace of A, and by $\mathrm{range}(A)$ its range. Also, for a subset C of H, we denote by $C^{\perp_H}$ the orthogonal complement of C in H, and by $\overline{C}^H$ the closure of C in H.

2.1 RKHSs and Hilbert-Schmidt operators

Below, we introduce the various Hilbert spaces relevant to our study.

Underlying RKHS. Let $\mathcal{H}$ be a separable RKHS of complex-valued functions on a general set $\mathscr{X}$, with reproducing kernel $K : \mathscr{X} \times \mathscr{X} \to \mathbb{C}$; see e.g. [1, 17]. For $t \in \mathscr{X}$, let $k_t \in \mathcal{H}$ be defined as $k_t(x) = K(x,t)$, $x \in \mathscr{X}$. For all $h \in \mathcal{H}$, we have $\langle k_t \,|\, h \rangle_{\mathcal{H}} = h(t)$, where $\langle \cdot \,|\, \cdot \rangle_{\mathcal{H}}$ stands for the inner product of $\mathcal{H}$ (this equality is often referred to as the *reproducing property*); we denote by $\|\cdot\|_{\mathcal{H}}$ the norm of $\mathcal{H}$, and we use a similar convention for the inner products and norms of all the Hilbert spaces encountered in this note.

Hilbert-Schmidt space. Let $\mathrm{HS}(\mathcal{H})$ be the Hilbert space of all HS operators on $\mathcal{H}$; see e.g. [7, 3]. For $T \in \mathrm{HS}(\mathcal{H})$, we denote by $T[h] \in \mathcal{H}$ the image of $h \in \mathcal{H}$ through T, and by $T[h](x)$ the value of the function $T[h]$ at $x \in \mathscr{X}$; we use similar notations for all function-valued operators. For a and $b \in \mathcal{H}$, let $T_{a,b} \in \mathrm{HS}(\mathcal{H})$ be the rank-one operator given by

$$T_{a,b}[h] = a\langle b \,|\, h \rangle_{\mathcal{H}}, h \in \mathcal{H};$$

we also set $S_b = T_{b,b}$.

Remark 2.1. An operator $T \in \mathrm{HS}(\mathcal{H})$ always admits a singular value decomposition (SVD) of the form $T = \sum_{i \in \mathbb{I}} \sigma_i T_{u_i, v_i}$, $\mathbb{I} \subseteq \mathbb{N}$, where $\{\sigma_i\}_{i \in \mathbb{I}} \in \ell^2(\mathbb{I})$ is the set of all strictly-positive singular values of T, and where $\{u_i\}_{i \in \mathbb{I}}$ and $\{v_i\}_{i \in \mathbb{I}}$ are two orthonormal systems in $\mathcal{H}$; the series converges in $\mathrm{HS}(\mathcal{H})$. ◁

Remark 2.2. Let $\mathcal{H}'$ be the continuous dual of $\mathcal{H}$. For $h \in \mathcal{H}$, let $\xi_h \in \mathcal{H}'$ be the bounded linear functional such that $\xi_h(f) = \langle h \,|\, f \rangle_{\mathcal{H}}$, $f \in \mathcal{H}$. Endowed with the inner product $\langle \xi_f \,|\, \xi_h \rangle_{\mathcal{H}'} = \langle h \,|\, f \rangle_{\mathcal{H}}$, the vector space $\mathcal{H}'$ is a Hilbert space, and the *Riesz map* $h \mapsto \xi_h$ is a bijective conjugate-linear isometry form $\mathcal{H}$ to $\mathcal{H}'$ (we may notice that $\xi_{\alpha h} = \overline{\alpha}\xi_h$, $\alpha \in \mathbb{C}$). The linear map densely defined as $T_{a,b} \mapsto a \otimes \xi_b$ (see Remark 2.1) is then a bijective isometry from the Hilbert space $\mathrm{HS}(\mathcal{H})$ to the tensor Hilbert space $\mathcal{H} \otimes \mathcal{H}'$. ◁

Conjugate RKHS. Let $\overline{\mathcal{H}}$ be the RKHS of complex-valued functions on $\mathscr{X}$ associated with the conjugate kernel $\overline{K}$, with $\overline{K}(x,t) = \overline{K(x,t)}$, x and $t \in \mathscr{X}$. For all $h \in \mathcal{H}$, we have $\overline{h} \in \overline{\mathcal{H}}$ (that is, the function $\overline{h} : x \mapsto \overline{h(x)}$ is a vector of $\overline{\mathcal{H}}$), and the map $h \mapsto \overline{h}$ is a bijective conjugate-linear isometry from $\mathcal{H}$ to $\overline{\mathcal{H}}$. We have $\overline{k_t}(x) = \overline{K(x,t)} = \overline{k_t(x)}$, and

$$\langle \overline{k_x} \,|\, \overline{k_t} \rangle_{\overline{\mathcal{H}}} = \overline{K(x,t)} = K(t,x) = \langle k_t \,|\, k_x \rangle_{\mathcal{H}}.$$

We denote by Ψ the bijective linear isometry from $\mathrm{HS}(\mathcal{H})$ to the tensor Hilbert space $\mathcal{H} \otimes \overline{\mathcal{H}}$, densely defined as $\Psi(T_{a,b}) = a \otimes \overline{b}$, a and $b \in \mathcal{H}$.

Remark 2.3. Following Remark 2.2, the linear map $\xi_h \mapsto \overline{h}$ is a bijective isometry form $\mathcal{H}'$ to $\overline{\mathcal{H}}$. Further, the linear map densely defined as $a \otimes \xi_b \mapsto a \otimes \overline{b}$ is a bijective isometry form $\mathcal{H} \otimes \mathcal{H}'$ to $\mathcal{H} \otimes \overline{\mathcal{H}}$; the composition of this isometry with the bijective isometry from $\mathrm{HS}(\mathcal{H})$ to $\mathcal{H} \otimes \mathcal{H}'$ discussed in Remark 2.2 yields the isometry $\Psi : \mathrm{HS}(\mathcal{H}) \to \mathcal{H} \otimes \overline{\mathcal{H}}$; see the diagram (5). ◁

Squared-kernel RKHS. The kernels K and $\overline{K}$ being PSD, by the Schur-product theorem, so is the squared-modulus kernel $|K|^2 = K\overline{K}$, with

$$|K|^2(x,t) = K(x,t)\overline{K(x,t)} = |K(x,t)|^2 = |k_t|^2(x),\ x \text{ and } t \in \mathscr{X}.$$

Let $\mathcal{G}$ be the RKHS of complex-valued functions on $\mathscr{X}$ for which $|K|^2$ is reproducing ($\mathcal{G} = \mathcal{H} \odot \overline{\mathcal{H}}$ is the *product* of the two RKHSs $\mathcal{H}$ and $\overline{\mathcal{H}}$; see e.g. [1, 17]).

Following [17, Chapter 5], we denote by $C_\Delta : \mathcal{H} \otimes \overline{\mathcal{H}} \to \mathcal{G}$ the coisometry densely defined as $C_\Delta(a \otimes \overline{b}) = a\overline{b}$, a and $b \in \mathcal{H}$, where $a\overline{b} \in \mathcal{G}$ is the complex-valued function on $\mathscr{X}$ given by

$$\big(a\overline{b}\big)(x) = a(x)\overline{b(x)} = \langle |k_x|^2 \,|\, a\overline{b}\rangle_{\mathcal{G}} = \langle k_x \otimes \overline{k_x} \,|\, a \otimes \overline{b}\rangle_{\mathcal{H}\otimes\overline{\mathcal{H}}},\ x \in \mathscr{X}.$$

For $\Upsilon \in \mathcal{H} \otimes \overline{\mathcal{H}}$, we more generally have

$$C_\Delta[\Upsilon](x) = \langle |k_x|^2 \,|\, C_\Delta[\Upsilon]\rangle_{\mathcal{G}} = \langle k_x \otimes \overline{k_x} \,|\, \Upsilon\rangle_{\mathcal{H}\otimes\overline{\mathcal{H}}}. \tag{1}$$

The initial space of C_Δ is $\mathcal{I}(C_\Delta) = \overline{\mathrm{span}_{\mathbb{C}}\{k_x \otimes \overline{k_x} | x \in \mathscr{X}\}}^{\mathcal{H}\otimes\overline{\mathcal{H}}}$, the closure in $\mathcal{H} \otimes \overline{\mathcal{H}}$ of the linear space spanned by the simple tensors $k_x \otimes \overline{k_x}$, $x \in \mathscr{X}$.

Remark 2.4. From (1), for all $x \in \mathscr{X}$, we have $C_\Delta^*[|k_x|^2] = k_x \otimes \overline{k_x}$. The linear space $\mathrm{span}_{\mathbb{C}}\{|k_x|^2 | x \in \mathscr{X}\}$ being dense in $\mathcal{G}$ (see for instance [17, Chapter 2]), we have $C_\Delta C_\Delta^* = \mathrm{id}_{\mathcal{G}}$, so that C_Δ^* is an isometry. ◁

2.2 Natural coisometry from $\mathrm{HS}(\mathcal{H})$ onto $\mathcal{G}$

We can now define a natural coisometry from the Hilbert space $\mathrm{HS}(\mathcal{H})$ of all HS operators on a RKHS $\mathcal{H}$, and onto the RKHS $\mathcal{G}$ associated with the squared-modulus of the reproducing kernel of $\mathcal{H}$. The terminology *natural* is used to emphasise that the considered construction does not depend on the choice of any specific basis.

Lemma 2.1. *The linear map $\Gamma = C_\Delta \Psi : \mathrm{HS}(\mathcal{H}) \to \mathcal{G}$ is a coisometry, and its initial space is*

$$\mathcal{I}(\Gamma) = \overline{\mathrm{span}_{\mathbb{C}}\{S_{k_x} | x \in \mathscr{X}\}}^{\mathrm{HS}(\mathcal{H})}; \tag{2}$$

in addition, for all $T \in \mathrm{HS}(\mathcal{H})$, we have

$$\Gamma[T](x) = \langle S_{k_x} \,|\, T\rangle_{\mathrm{HS}(\mathcal{H})} = \langle k_x \,|\, T[k_x]\rangle_{\mathcal{H}} = T[k_x](x), x \in \mathscr{X}. \tag{3}$$

Proof. The linear isometry Ψ being bijective, we have $\Psi\Psi^* = \mathrm{id}_{\mathcal{H}\otimes\overline{\mathcal{H}}}$, and so

$$\Gamma\Gamma^* = C_\Delta \Psi\Psi^* C_\Delta^* = C_\Delta C_\Delta^* = \mathrm{id}_{\mathcal{G}}\,.$$

By definition of C_Δ and Ψ, we have $\Gamma^*[|k_x|^2] = \Psi^*[k_x \otimes \overline{k_x}] = S_{k_x}$, $x \in \mathscr{X}$, so that (2) follows from the density of $\mathrm{span}_{\mathbb{C}}\{|k_x|^2 | x \in \mathscr{X}\}$ in $\mathcal{G}$ (see Remark 2.4). The reproducing property in $\mathcal{G}$ then gives

$$\Gamma[T](x) = \langle |k_x|^2 \,|\, \Gamma[T]\rangle_{\mathcal{G}} = \langle S_{k_x} \,|\, T\rangle_{\mathrm{HS}(\mathcal{H})}, T \in \mathrm{HS}(\mathcal{H}).$$

We next observe that

$$\langle T_{a,b} \,|\, T\rangle_{\mathrm{HS}(\mathcal{H})} = \langle a \,|\, T[b]\rangle_{\mathcal{H}},\ a \text{ and } b \in \mathcal{H}; \tag{4}$$

indeed, as $T_{a,0} = 0$, equality (4) trivially holds for $b = 0$, and for $b \neq 0$, we have $\langle T_{a,b} \,|\, T\rangle_{\mathrm{HS}(\mathcal{H})} = \langle T_{a,b}[b] \,|\, T[b]\rangle_{\mathcal{H}} / \|b\|_{\mathcal{H}}^2$, with $T_{a,b}[b] = a\|b\|_{\mathcal{H}}^2$. Taking $a = b = k_x$ in (4) gives $\langle S_{k_x} \,|\, T\rangle_{\mathrm{HS}(\mathcal{H})} = \langle k_x \,|\, T[k_x]\rangle_{\mathcal{H}} = T[k_x](x)$, concluding the proof. □

The following diagram summarises the construction of Γ (the $\cong$ symbol refers to the two bijective linear isometries discussed in Remarks 2.2 and 2.3).

$$\begin{array}{ccc} \mathrm{HS}(\mathcal{H}) & \xrightarrow{\quad\Gamma\quad} & \mathcal{G} \\ \cong\downarrow & \searrow^{\Psi} & \uparrow{\scriptstyle C_\Delta} \\ \mathcal{H}\otimes\mathcal{H}' & \xrightarrow[\cong]{} & \mathcal{H}\otimes\overline{\mathcal{H}} \end{array} \tag{5}$$

Through Γ, the HS operators on $\mathcal{H}$ belonging to $\mathcal{I}(\Gamma)$ can be isometrically represented as functions in the RKHS $\mathcal{G}$ associated with the squared-modulus kernel $|K|^2$. In the framework of Remark 2.1, we may notice that if $T = \sum_{i\in\mathbb{I}} \sigma_i T_{u_i,v_i} \in \mathrm{HS}(\mathcal{H})$, then $\Gamma[T] = \sum_{i\in\mathbb{I}} \sigma_i u_i \overline{v_i}$.

Lemma 2.2. *The following assertions hold:*

(i) if $T \in \mathrm{HS}(\mathcal{H})$ is self-adjoint, then the function $\Gamma[T]$ is real-valued;

(ii) if $T \in \mathrm{HS}(\mathcal{H})$ is PSD, then the function $\Gamma[T]$ is nonnegative;

(iii) if $T \in \mathrm{HS}(\mathcal{H})$ is PSD and $\Gamma[T] = 0$, then $T = 0$; and

(iv) if $T \in \mathcal{I}(\Gamma)$, then $T^* \in \mathcal{I}(\Gamma)$.

Proof. Assertions *(i)* and *(ii)* follow directly form (3). We assume that $T \in \mathrm{HS}(\mathcal{H})$ is PSD, and we consider a spectral expansion $T = \sum_{j\in\mathbb{I}} \lambda_j S_{\varphi_j}$ of T, with $\lambda_j \geqslant 0$, $\varphi_j \in \mathcal{H}$ and $\mathbb{I} \subseteq \mathbb{N}$; observing that $\Gamma[S_{\varphi_j}] = |\varphi_j|^2$, $j \in \mathbb{I}$, we obtain *(iii)*. To prove assertion *(iv)*, we first observe that if $g \in \mathcal{G}$, then $\overline{g} \in \mathcal{G}$ (that is, the function $\overline{g}$ is a vector of $\mathcal{G}$); the map Γ is indeed surjective, and if $g = \Gamma[T]$, $T \in \mathrm{HS}(\mathcal{H})$, then $\overline{g} = \Gamma[T^*]$. By linearity, the real and imaginary parts of g are then also vectors of $\mathcal{G}$, and so $\|g\|_{\mathcal{G}} = \|\overline{g}\|_{\mathcal{G}}$ (see for instance [17, Chapter 5]; see also Remark 2.6). Since Γ is a partial isometry, for $T \in \mathrm{HS}(\mathcal{H})$, we have $\|T\|_{\mathrm{HS}(\mathcal{H})} \geqslant \|\Gamma[T]\|_{\mathcal{G}}$, with equality if and only if $T \in \mathcal{I}(\Gamma)$; as $\|T\|_{\mathrm{HS}(\mathcal{H})} = \|T^*\|_{\mathrm{HS}(\mathcal{H})}$, the result follows. □

Remark 2.5. The diagram (5) is also well-defined when the involved Hilbert spaces are real. We in this case have $\mathcal{I}(\Gamma) = \overline{\mathrm{span}_{\mathbb{R}}\{S_{k_x} | x \in \mathscr{X}\}}^{\mathrm{HS}(\mathcal{H})}$ and the operators in $\mathcal{I}(\Gamma)$ are self-adjoint; also if $T^* = -T$, then $\Gamma[T] = 0$. By comparison, in the complex case, if $T^* = -T$, then the function $\Gamma[T]$ is pure-imaginary. ◁

Remark 2.6. The PSD kernel $|K|^2$ being real-valued, it is the reproducing kernel of a real RKHS $\mathcal{G}_{\mathbb{R}}$ of real-valued functions on $\mathscr{X}$. The decomposition $\mathcal{G} = \mathcal{G}_{\mathbb{R}} + \mathrm{i}\mathcal{G}_{\mathbb{R}}$ holds, and $\mathcal{G}_{\mathbb{R}}$ is the real-linear subspace of all real-valued functions in $\mathcal{G}$. This decomposition mirrors the decomposition $\mathrm{HS}(\mathcal{H}) = \mathrm{HS}_{\mathbb{R}}(\mathcal{H}) + \mathrm{iHS}_{\mathbb{R}}(\mathcal{H})$, with $\mathrm{HS}_{\mathbb{R}}(\mathcal{H}) \subset \mathrm{HS}(\mathcal{H})$ the real-linear subspace of all self-adjoint HS operators on $\mathcal{H}$. Also, the real convex cone $\mathrm{HS}^+_{\mathbb{R}}(\mathcal{H}) \subset \mathrm{HS}_{\mathbb{R}}(\mathcal{H})$ of all PSD HS operators on $\mathcal{H}$ is generating in $\mathrm{HS}_{\mathbb{R}}(\mathcal{H})$, and the real convex cone $\mathcal{G}^+_{\mathbb{R}} \subset \mathcal{G}_{\mathbb{R}}$ of all nonnegative functions in $\mathcal{G}_{\mathbb{R}}$ is generating in $\mathcal{G}_{\mathbb{R}}$. ◁

Remark 2.7. Let $\mathcal{F}$ be another separable RKHS of complex-valued functions on $\mathscr{X}$, with reproducing kernel $J : \mathscr{X} \times \mathscr{X} \to \mathbb{C}$. We denote by $\mathrm{HS}(\mathcal{F}, \mathcal{H})$ the Hilbert space of all HS operators from $\mathcal{F}$ to $\mathcal{H}$, and let $\mathcal{H} \odot \overline{\mathcal{F}}$ be the product of the RKHSs $\mathcal{H}$ and $\overline{\mathcal{F}}$, that is, the RKHS with kernel $K\overline{J}$. Following (5), we can more generally define a natural coisometry from $\mathrm{HS}(\mathcal{F}, \mathcal{H})$ onto $\mathcal{H} \odot \overline{\mathcal{F}}$. ◁

3 Trace-class integral operators with PSD kernels

From Lemma 2.1, if $T \in \mathrm{HS}(\mathcal{H})$ is of the form $T = \sum_{j=1}^n \omega_j S_{k_{s_j}}$, with $n \in \mathbb{N}$, $s_j \in \mathscr{X}$ and $\omega_j \in \mathbb{C}$, then $T \in \mathcal{I}(\Gamma)$. We in this case have

$$T[h](x) = \sum_{j=1}^n \omega_j k_{s_j}(x)\langle k_{s_j} \mid h\rangle_{\mathcal{H}} = \sum_{j=1}^n \omega_j K(x, s_j)h(s_j), h \in \mathcal{H}, x \in \mathscr{X},$$

so that T can be regarded as an integral operator on $\mathcal{H}$ defined by the kernel K and the finitely-supported measure $\sum_{j=1}^n \omega_j \delta_{s_j}$, with δ_x the Dirac measure at $x \in \mathscr{X}$. We also have $\Gamma[T] = \sum_{j=1}^n \omega_j |k_{s_j}|^2$, so that $\langle \Gamma[T] \mid g\rangle_{\mathcal{G}} = \sum_{j=1}^n \overline{\omega_j} g(s_j)$, $g \in \mathcal{G}$, and $\Gamma[T]$ is thus the Riesz representation of the integral functional on $\mathcal{G}$ defined by the measure $\sum_{j=1}^n \overline{\omega_j}\delta_{s_j}$. Under measurability conditions, this observation holds for all trace-class integral operators on $\mathcal{H}$ defined by the reproducing kernel K of $\mathcal{H}$ and general measures on $\mathscr{X}$, as illustrated below.

3.1 Integral operators and kernel embedding of measures

Let $\mathcal{A}$ be a σ-algebra of subsets of $\mathscr{X}$. We consider the Borel σ-algebra of $\mathbb{C}$, and make the following assumptions on K and the measurable space $(\mathscr{X}, \mathcal{A})$:

(A.1) for all $t \in \mathscr{X}$, the function $k_t : \mathscr{X} \to \mathbb{C}$ is measurable;

(A.2) the diagonal of K is measurable.

We recall that $K(t,t) = \|k_t\|_{\mathcal{H}}^2 = \|S_{k_t}\|_{\mathrm{HS}(\mathcal{H})} = \||k_t|^2\|_{\mathcal{G}}, t \in \mathscr{X}$.

Remark 3.1. The RKHSs $\mathcal{H}$ and $\mathcal{G}$ being separable, (A.1) ensures that all the functions in $\mathcal{H}$ and $\mathcal{G}$ are measurable; see for instance [25, Lemma 4.24]. Consequently, under (A.1), the maps $t \mapsto k_t$, $t \mapsto |k_t|^2$ and $t \mapsto S_{k_t}$, $t \in \mathscr{X}$, are weakly-measurable, and since the Hilbert spaces $\mathcal{H}$, $\mathcal{G}$ and $\mathrm{HS}(\mathcal{H})$ are separable, by the Pettis measurability theorem, these maps are also strongly-measurable (see e.g. [8, 28]). ◁

We denote by $\mathcal{M}_+$, $\mathcal{M}$, and $\mathcal{M}_{\mathbb{C}}$ the set of all nonnegative, signed and complex measures[1] on $(\mathscr{X}, \mathcal{A})$, and we set $\mathcal{M}_{\mathbb{F}} = \mathcal{M} \cup \mathcal{M}_{\mathbb{C}}$ (we have $\mathcal{M}_+ \subset \mathcal{M}$). Noticing that $K(t,t) \geqslant 0$, $t \in \mathscr{X}$, from (A.2), we define

$$\tau_\mu = \int_{\mathscr{X}} K(t,t)\mathrm{d}|\mu|(t) \in \mathbb{R}_{\geqslant 0} \cup \{+\infty\}, \mu \in \mathcal{M}_{\mathbb{F}}.$$

We next introduce the sets $\mathcal{T}_+(K)$, $\mathcal{T}(K)$ and $\mathcal{T}_{\mathbb{C}}(K)$ of all measures μ in $\mathcal{M}_+$, $\mathcal{M}$, and $\mathcal{M}_{\mathbb{C}}$ such that τ_μ is finite, respectively; the inclusion $\mathcal{T}_+(K) \subset \mathcal{T}(K)$ holds, and we set $\mathcal{T}_{\mathbb{F}}(K) = \mathcal{T}(K) \cup \mathcal{T}_{\mathbb{C}}(K)$.

Integral operators on $\mathcal{H}$ with kernel K. By assumption, for $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, the integral $\int_{\mathscr{X}} \|S_{k_t}\|_{\mathrm{HS}(\mathcal{H})}\mathrm{d}|\mu|(t) = \tau_\mu$ is finite, and the map $t \mapsto S_{k_t}$ is thus Bochner-integrable with respect to μ (Bochner integrability criterion, see e.g. [8, 28]; see also Remark 3.1). We set

$$L_\mu = \int_{\mathscr{X}} S_{k_t}\mathrm{d}\mu(t) \in \mathrm{HS}(\mathcal{H}).$$

From (4), for $h \in \mathcal{H}$ and $x \in \mathscr{X}$, we have

$$L_\mu[h](x) = \langle T_{k_x,h} \mid L_\mu \rangle_{\mathrm{HS}(\mathcal{H})} = \int_{\mathscr{X}} \langle T_{k_x,h} \mid S_{k_t} \rangle_{\mathrm{HS}(\mathcal{H})}\mathrm{d}\mu(t) = \int_{\mathscr{X}} K(x,t)h(t)\mathrm{d}\mu(t),$$

so that $L_\mu \in \mathrm{HS}(\mathcal{H})$ can be regarded as an integral operator on $\mathcal{H}$ defined by the kernel K and the measure μ.

Remark 3.2. For $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, the operator $L_\mu \in \mathrm{HS}(\mathcal{H})$ is the Riesz representation of the bounded linear functional $Z_\mu : \mathrm{HS}(\mathcal{H}) \to \mathbb{C}$ given by

$$Z_\mu(T) = \int_{\mathscr{X}} \langle S_{k_t} \mid T \rangle_{\mathrm{HS}(\mathcal{H})}\mathrm{d}\overline{\mu}(t), T \in \mathrm{HS}(\mathcal{H}),$$

that is, $Z_\mu(T) = \langle L_\mu \mid T \rangle_{\mathrm{HS}(\mathcal{H})}$; from the CS inequality in $\mathrm{HS}(\mathcal{H})$, we in particular have $|Z_\mu(T)| \leqslant \int_{\mathscr{X}} |\langle S_{k_t} \mid T \rangle_{\mathrm{HS}(\mathcal{H})}|\mathrm{d}|\mu|(t) \leqslant \|T\|_{\mathrm{HS}(\mathcal{H})}\tau_\mu$.

By boundedness of the linear evaluation map $T \mapsto T[h]$ from $\mathrm{HS}(\mathcal{H})$ to $\mathcal{H}$, we obtain (see for instance [28, Chapter 5])

$$L_\mu[h] = \int_{\mathscr{X}} S_{k_t}[h]\mathrm{d}\mu(t) = \int_{\mathscr{X}} k_t h(t)\mathrm{d}\mu(t), h \in \mathcal{H},$$

[1] We only consider finite complex measures, while signed measures may not be finite.

with, from the CS inequality in $\mathcal{H}$, $\int_{\mathscr{X}} \|k_t\|_{\mathcal{H}} |h(t)| \mathrm{d}|\mu|(t) \leqslant \|h\|_{\mathcal{H}} \tau_\mu$. We also have

$$\langle L_\mu[h] \,|\, f \rangle_{\mathcal{H}} = \int_{\mathscr{X}} \overline{h(t)} f(t) \mathrm{d}\overline{\mu}(t), \ h \text{ and } f \in \mathcal{H}, \tag{6}$$

so that $L_\mu[h]$ is the Riesz representation of the bounded linear functional $\Theta_{h,\mu}$ on $\mathcal{H}$, with $\Theta_{h,\mu} : f \mapsto \int_{\mathscr{X}} \overline{h(t)} f(t) \mathrm{d}\overline{\mu}(t)$, $f \in \mathcal{H}$ (and $|\Theta_{h,\mu}(f)| \leqslant \|h\|_{\mathcal{H}} \|f\|_{\mathcal{H}} \tau_\mu$). ◁

Lemma 3.1. *For all $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, the operator L_μ is trace-class.*

Proof. Let $\{h_i\}_{i \in \mathbb{I}}$ be an orthonormal basis (ONB) of $\mathcal{H}$, with $\mathbb{I} \subseteq \mathbb{N}$. For all $t \in \mathscr{X}$, we have $k_t = \sum_{i \in \mathbb{I}} h_i \overline{h_i(t)}$, so that $\{h_i(t)\}_{i \in \mathbb{I}} \in \ell^2(\mathbb{I})$ and $\sum_{i \in \mathbb{I}} |h_i(t)|^2 = K(t,t)$; see e.g. [17, Chapter 2]. Let $\{f_i\}_{i \in \mathbb{I}}$ be another ONB of $\mathcal{H}$; from (6), and by monotone convergence and the CS inequality in $\ell^2(\mathbb{I})$, we obtain

$$\begin{aligned} \sum_{i \in \mathbb{I}} \left| \langle f_i \,|\, L_\mu[h_i] \rangle_{\mathcal{H}} \right| &\leqslant \sum_{i \in \mathbb{I}} \int_{\mathscr{X}} |f_i(t)| |h_i(t)| \mathrm{d}|\mu|(t) \\ &= \int_{\mathscr{X}} \sum_{i \in \mathbb{I}} |f_i(t)| |h_i(t)| \mathrm{d}|\mu|(t) \leqslant \int_{\mathscr{X}} \sqrt{K(t,t)} \sqrt{K(t,t)} \mathrm{d}|\mu|(t) = \tau_\mu, \end{aligned}$$

so that $\mathrm{trace}(|L_\mu|) \leqslant \tau_\mu$, with $|L_\mu| = (L_\mu^* L_\mu)^{1/2}$ the modulus of L_μ. □

Kernel embedding of measures in $\mathcal{G}$. By assumption again, for $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, the integral $\int_{\mathscr{X}} \||k_t|^2\|_{\mathcal{G}} \mathrm{d}|\mu|(t)$ is finite, and the map $t \mapsto |k_t|^2$ is therefore Bochner-integrable with respect to μ. We set

$$g_\mu = \int_{\mathscr{X}} |k_t|^2 \mathrm{d}\mu(t) \in \mathcal{G}.$$

We have $\langle g_\mu \,|\, g \rangle_{\mathcal{G}} = \int_{\mathscr{X}} g(t) \mathrm{d}\overline{\mu}(t)$, $g \in \mathcal{G}$, so that g_μ is the Riesz representation of the linear functional $I_\mu : \mathcal{G} \to \mathbb{C}$, with $I_\mu(g) = \int_{\mathscr{X}} g(t) \mathrm{d}\overline{\mu}(t)$; we may observe that $|I_\mu(g)| \leqslant \|g\|_{\mathcal{G}} \tau_\mu$, and that $g_\mu(x) = \int_{\mathscr{X}} |K(x,t)|^2 \mathrm{d}\mu(t)$, $x \in \mathscr{X}$. The vector g_μ is referred to as the *kernel embedding*, or *potential*, of the measure μ in the RHKS $\mathcal{G}$; see for instance [24, 6, 15].

Theorem 3.1. *For all $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, we have $L_\mu \in \mathcal{I}(\Gamma)$ and $\Gamma[L_\mu] = g_\mu$.*

Proof. From Lemma 2.1 and by definition of L_μ and g_μ, for all $T \in \mathrm{HS}(\mathcal{H})$, we have

$$\langle \Gamma^*[g_\mu] \,|\, T \rangle_{\mathrm{HS}(\mathcal{H})} = \langle g_\mu \,|\, \Gamma[T] \rangle_{\mathcal{G}} = \int_{\mathscr{X}} \langle S_{k_t} \,|\, T \rangle_{\mathrm{HS}(\mathcal{H})} \mathrm{d}\overline{\mu}(t) = \langle L_\mu \,|\, T \rangle_{\mathrm{HS}(\mathcal{H})},$$

so that $\Gamma^*[g_\mu] = L_\mu$. □

Remark 3.3. Following Lemma 2.2, for a signed measure $\mu \in \mathcal{T}(K)$, the function g_μ is real-valued, and the operator $L_\mu = \Gamma^*[g_\mu]$ is self-adjoint. Also, for a nonnegative measure $\mu \in \mathcal{T}_+(K)$, the function g_μ is nonnegative, and the operator L_μ is PSD. We may notice that $L_{\delta_x} = S_{k_x}$, $x \in \mathscr{X}$. ◁

From Theorem 3.1, for μ and $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, the following equalities hold:

$$\begin{aligned}\langle L_\nu \mid L_\mu \rangle_{\mathrm{HS}(\mathcal{H})} = \langle g_\nu \mid g_\mu \rangle_{\mathcal{G}} &= \iint_{\mathscr{X}} |K(x,t)|^2 \mathrm{d}\mu(t)\mathrm{d}\overline{\nu}(x) \\ &= \int_{\mathscr{X}} g_\mu(x)\mathrm{d}\overline{\nu}(x) = \int_{\mathscr{X}} \overline{g_\nu(t)}\mathrm{d}\mu(t),\end{aligned} \tag{7}$$

relating the evaluation of inner products in $\mathrm{HS}(\mathcal{H})$ between trace-class integral operators with kernel K to the integration of potentials in $\mathcal{G}$.

3.2 Quadrature approximation

Let $B_{\mathcal{G}} = \{g \in \mathcal{G} | \|g\|_{\mathcal{G}} \leqslant 1\}$ be the closed unit ball of $\mathcal{G}$. We set

$$\mathfrak{M}_{\mathcal{G}}(\mu,\nu) = \sup_{g \in B_{\mathcal{G}}} \left| \int_{\mathscr{X}} g(t)\mathrm{d}\mu(t) - \int_{\mathscr{X}} g(t)\mathrm{d}\nu(t) \right|, \mu \text{ and } \nu \in \mathcal{T}_{\mathbb{F}}(K).$$

The map $\mathfrak{M}_{\mathcal{G}}$ defines a pseudometric on $\mathcal{T}_{\mathbb{F}}(K)$; for probability measures, such pseudometrics are referred to as *integral probability metrics*, or *maximum mean discrepancies*; see for instance [16, 24, 23, 15, 27].

The following Corollary 3.1 describes the equivalence between the quadrature approximation of trace-class integral operators with PSD kernels and the approximation of integral functionals on RKHSs with squared-modulus kernels.

Corollary 3.1. *For all μ and $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we have $\|L_\mu - L_\nu\|_{\mathrm{HS}(\mathcal{H})} = \mathfrak{M}_{\mathcal{G}}(\mu,\nu)$.*

Proof. Form Theorem 3.1 and by linearity of Γ^*, we have $\Gamma^*[g_\mu - g_\nu] = L_\mu - L_\nu$. Since Γ^* is an isometry, it follows that $\|L_\mu - L_\nu\|_{\mathrm{HS}(\mathcal{H})} = \|g_\mu - g_\nu\|_{\mathcal{G}}$. The CS inequality in $\mathcal{G}$ and the definition of g_μ and g_ν then give

$$\|g_\mu - g_\nu\|_{\mathcal{G}} = \sup_{g \in B_{\mathcal{G}}} \left|\langle g_\mu - g_\nu \mid g\rangle_{\mathcal{G}}\right| = \sup_{g \in B_{\mathcal{G}}} \left| \int_{\mathscr{X}} g(t)\mathrm{d}\overline{\mu}(t) - \int_{\mathscr{X}} g(t)\mathrm{d}\overline{\nu}(t) \right|.$$

We conclude by observing that for all $g \in \mathcal{G}$, we have $\overline{g} \in \mathcal{G}$ and $\|\overline{g}\|_{\mathcal{G}} = \|g\|_{\mathcal{G}}$ (see the proof of Lemma 2.2), so that $\|g_\mu - g_\nu\|_{\mathcal{G}} = \mathfrak{M}_{\mathcal{G}}(\mu,\nu)$. □

3.3 Further properties

In this section and in anticipation of the forthcoming developments, we discuss some further properties verified by the integral operators considered in Theorem 3.1.

For $\mu \in \mathcal{T}_+(K)$, let $L^2(\mu)$ be the Hilbert space of all square-integrable functions with respect to μ. From the CS inequality in $\mathcal{H}$, we have

$$\int_{\mathscr{X}} |h(t)|^2 \mathrm{d}\mu(t) = \int_{\mathscr{X}} \left|\langle k_t \mid h\rangle_{\mathcal{H}}\right|^2 \mathrm{d}\mu(t) \leqslant \|h\|_{\mathcal{H}}^2 \tau_\mu, h \in \mathcal{H}, \tag{8}$$

so that the linear embedding $\imath_\mu : \mathcal{H} \to L^2(\mu)$, with $\imath_\mu[h]$ the equivalence class of all measurable functions μ-almost everywhere equal to h, is bounded (see e.g. [26]).

Lemma 3.2. *For all $\mu \in \mathcal{T}_+(K)$, the map $\imath_\mu$ is HS and $L_\mu = \imath_\mu^* \imath_\mu$.*

Proof. Let $\{h_i\}_{i\in\mathbb{I}}$ be an ONB of $\mathcal{H}$, with $\mathbb{I} \subseteq \mathbb{N}$. As $K(t,t) = \sum_{i\in\mathbb{I}} \left|h_i(t)\right|^2$, $t \in \mathscr{X}$ (see e.g. [17]), from (6) and by monotone convergence, we have

$$\sum_{i\in\mathbb{I}} \|\imath_\mu[h_i]\|^2_{L^2(\mu)} = \sum_{i\in\mathbb{I}} \langle L_\mu[h_i] \mid h_i\rangle_{\mathcal{H}} = \int_{\mathscr{X}} \sum_{i\in\mathbb{I}} |h_i(t)|^2 \mathrm{d}\mu(t) = \tau_\mu,$$

so that $\imath_\mu$ is HS. From (6), we also obtain

$$\langle L_\mu[h] \mid f\rangle_{\mathcal{H}} = \langle \imath_\mu[h] \mid \imath_\mu[f]\rangle_{L^2(\mu)} = \langle \imath_\mu^*\imath_\mu[h] \mid f\rangle_{\mathcal{H}},\ h \text{ and } f \in \mathcal{H},$$

and so $L_\mu = \imath_\mu^*\imath_\mu$. □

For $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we by definition have $|\nu| \in \mathcal{T}_+(K)$ and $\overline{\nu} \in \mathcal{T}_{\mathbb{F}}(K)$; from (6), we also have $L_\nu^* = L_{\overline{\nu}}$. The following relation (Lemma 3.3) holds between the range of L_ν and the range of $L_{|\nu|}$.

Lemma 3.3. *For all $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we have* $\overline{\mathrm{range}(L_\nu)}^{\mathcal{H}} \subseteq \overline{\mathrm{range}(L_{|\nu|})}^{\mathcal{H}}$.

Proof. From (6) and the CS inequality in $L^2(|\nu|)$, we obtain

$$\begin{aligned} \left|\langle L_{\overline{\nu}}[h] \mid f\rangle_{\mathcal{H}}\right| &= \left|\int_{\mathscr{X}} \overline{h(t)} f(t) \mathrm{d}\nu(t)\right| \leqslant \int_{\mathscr{X}} \left|h(t)\right| |f(t)| \mathrm{d}|\nu|(t) \\ &\leqslant \|\imath_{|\nu|}[h]\|_{L^2(|\nu|)} \|\imath_{|\nu|}[f]\|_{L^2(|\nu|)},\ h \text{ and } f \in \mathcal{H}, \end{aligned} \tag{9}$$

where (8) ensures that the embedding $\imath_{|\nu|}$ is well-defined. From Lemma 3.2, we have $\mathrm{null}(L_\mu) = \{h \in \mathcal{H} | \imath_\mu[h] = 0\}$; inequality (9) then entails $\mathrm{null}(L_{|\nu|}) \subseteq \mathrm{null}(L_{\overline{\nu}})$, and so $\mathrm{null}(L_{\overline{\nu}})^{\perp_{\mathcal{H}}} \subseteq \mathrm{null}(L_{|\nu|})^{\perp_{\mathcal{H}}}$. Recalling that $\mathrm{null}(T^*) = \mathrm{range}(T)^{\perp_{\mathcal{H}}}$, $T \in \mathrm{HS}(\mathcal{H})$, we conclude by noticing that $L_{\overline{\nu}}^* = L_\nu$ and $L_{|\nu|}^* = L_{|\nu|}$. □

Lemma 3.4 illustrates that when the measure ν is finitely-supported, the range of $L_{|\nu|}$ is fully characterised by the support of ν.

Lemma 3.4. *For $\nu = \sum_{i=1}^n \upsilon_i \delta_{s_i}$, with $n \in \mathbb{N}$, $\upsilon_i \in \mathbb{C}$, $\upsilon_i \neq 0$, and $s_i \in \mathscr{X}$, we have* $\mathrm{range}(L_{|\nu|}) = \mathrm{span}_{\mathbb{C}}\{k_{s_1}, \cdots, k_{s_n}\}$.

Proof. We have $|\nu| = \sum_{i=1}^n |\upsilon_i| \delta_{s_i} \in \mathcal{T}_+(K)$, and $L_{|\nu|}$ is PSD. From Lemma 3.2, we obtain that $\mathrm{null}(L_{|\nu|}) = \{h \in \mathcal{H} | \imath_{|\nu|}[h] = 0\} = \bigcap_{i=1}^n \{h \in \mathcal{H} | \langle k_{s_i} \mid h\rangle_{\mathcal{H}} = 0\}$, and so $\mathrm{null}(L_{|\nu|})^{\perp_{\mathcal{H}}} = \mathrm{span}_{\mathbb{C}}\{k_{s_1}, \cdots, k_{s_n}\}$. Since $L_{|\nu|}$ is self-adjoint, the result follows. □

4 Measures and projection-based approximations

In this section, we illustrate the extent to which the the approximation of potentials in $\mathcal{G}$ can be used as a surrogate for the characterisation of closed linear subspaces of $\mathcal{H}$ for the approximation of $L_\mu \in \mathrm{HS}(\mathcal{H})$ through projections (see Remark 4.1).

4.1 Additional notations and general properties

For a closed linear subspace $\mathcal{H}_S$ of $\mathcal{H}$, we denote by P_S the orthogonal projection from $\mathcal{H}$ onto $\mathcal{H}_S$. Endowed with the Hilbert structure of $\mathcal{H}$, the vector space $\mathcal{H}_S$ is a RKHS, and its reproducing kernel K_S verifies $K_S(x,t) = P_S[k_t](x)$, x and $t \in \mathscr{X}$.

Remark 4.1. The linear map $T \mapsto P_S T$ is the orthogonal projection from $\mathrm{HS}(\mathcal{H})$ onto $\mathcal{R}(\mathcal{H}_S) = \{T \in \mathrm{HS}(\mathcal{H}) | \operatorname{range}(T) \subseteq \mathcal{H}_S\}$, the closed linear subspace of $\mathrm{HS}(\mathcal{H})$ of all operators with range included in $\mathcal{H}_S$. Also, the linear map $T \mapsto TP_S$ is the orthogonal projection from $\mathrm{HS}(\mathcal{H})$ onto $\mathcal{Z}(\mathcal{H}_S) = \{T \in \mathrm{HS}(\mathcal{H}) | \operatorname{range}(T^*) \subseteq \mathcal{H}_S\}$. The two orthogonal projections $T \mapsto P_S T$ and $T \mapsto TP_S$ commute, and their composition, that is, the linear map $T \mapsto P_S T P_S$, is the orthogonal projection from $\mathrm{HS}(\mathcal{H})$ onto $\mathcal{R}(\mathcal{H}_S) \cap \mathcal{Z}(\mathcal{H}_S)$. As $(P_S T)^* = T^* P_S$, the orthogonal projections onto $\mathcal{R}(\mathcal{H}_S)$ and $\mathcal{Z}(\mathcal{H}_S)$ are intrinsically related; for this reason, in what follows, we mainly focus on approximations of the form $P_S T$ and $P_S T P_S$. By orthogonality, for all $T \in \mathrm{HS}(\mathcal{H})$, we have

$$\|T - P_S T\|^2_{\mathrm{HS}(\mathcal{H})} = \|T\|^2_{\mathrm{HS}(\mathcal{H})} - \|P_S T\|^2_{\mathrm{HS}(\mathcal{H})}, \text{ and} \tag{10}$$

$$\|T - P_S T P_S\|^2_{\mathrm{HS}(\mathcal{H})} = \|T\|^2_{\mathrm{HS}(\mathcal{H})} - \|P_S T P_S\|^2_{\mathrm{HS}(\mathcal{H})}, \tag{11}$$

with $\|P_S T P_S\|_{\mathrm{HS}(\mathcal{H})} \leqslant \|P_S T\|_{\mathrm{HS}(\mathcal{H})} \leqslant \|T\|_{\mathrm{HS}(\mathcal{H})}$; in particular, if T is self-adjoint, then so is $P_S T P_S$. ◁

Lemma 4.1. *Let $\mathcal{H}_S$ and $\mathcal{H}_R$ be two closed linear subspaces of $\mathcal{H}$, with $\mathcal{H}_R \subseteq \mathcal{H}_S$. For all $T \in \mathrm{HS}(\mathcal{H})$, we have*

$$\|P_R T\|_{\mathrm{HS}(\mathcal{H})} \leqslant \|P_S T\|_{\mathrm{HS}(\mathcal{H})} \quad \textit{and} \quad \|P_R T P_R\|_{\mathrm{HS}(\mathcal{H})} \leqslant \|P_S T P_S\|_{\mathrm{HS}(\mathcal{H})}.$$

Proof. We denote by $\mathcal{H}_e$ the orthogonal complement of $\mathcal{H}_R$ in $\mathcal{H}_S$. We then have $P_S = P_R + P_e$ and $\langle P_R T \mid P_e \tilde{T} \rangle_{\mathrm{HS}(\mathcal{H})} = \langle T P_R \mid \tilde{T} P_e \rangle_{\mathrm{HS}(\mathcal{H})} = 0$, T and $\tilde{T} \in \mathrm{HS}(\mathcal{H})$. We hence obtain $\|P_S T\|^2_{\mathrm{HS}(\mathcal{H})} = \|P_R T\|^2_{\mathcal{H}} + \|P_e T\|^2_{\mathrm{HS}(\mathcal{H})}$, and

$$\begin{aligned}\|P_S T P_S\|^2_{\mathrm{HS}(\mathcal{H})} = \|P_R T P_R\|^2_{\mathrm{HS}(\mathcal{H})} &+ \|P_e T P_e\|^2_{\mathrm{HS}(\mathcal{H})} \\ &+ \|P_R T P_e\|^2_{\mathrm{HS}(\mathcal{H})} + \|P_e T P_R\|^2_{\mathrm{HS}(\mathcal{H})},\end{aligned}$$

completing the proof. □

By boundedness of P_S, for $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, we have $P_S L_\mu = \int_{\mathscr{X}} P_S S_{k_t} \mathrm{d}\mu(t)$, and so

$$P_S L_\mu[h](x) = \int_{\mathscr{X}} K_S(x,t) h(t) \mathrm{d}\mu(t), \ h \in \mathcal{H}, x \in \mathscr{X}.$$

The operator $P_S L_\mu \in \mathrm{HS}(\mathcal{H})$ can thus be regarded as an integral operator on $\mathcal{H}$ defined by the kernel K_S and the measure μ. Since $K_S(t,t) \leqslant K(t,t)$, $t \in \mathscr{X}$, we may notice that $\mathcal{T}_{\mathbb{F}}(K) \subseteq \mathcal{T}_{\mathbb{F}}(K_S)$. We have

$$\|L_\mu - P_S L_\mu\|^2_{\mathrm{HS}(\mathcal{H})} = \iint_{\mathscr{X}} K(t,x)\big(K(x,t) - K_S(x,t)\big) \mathrm{d}\mu(t) \mathrm{d}\overline{\mu}(x), \text{ and} \tag{12}$$

$$\|L_\mu - P_S L_\mu P_S\|^2_{\mathrm{HS}(\mathcal{H})} = \iint_{\mathscr{X}} |K(x,t)|^2 - |K_S(x,t)|^2 \mathrm{d}\mu(t) \mathrm{d}\overline{\mu}(x); \tag{13}$$

see (10), (11) and Lemma A.1 in Appendix A for a detailed computation (see also Remark 4.2 for an alternative computation involving Γ).

Remark 4.2. Let $\mathcal{H}_U$ and $\mathcal{H}_V$ be two closed linear subspaces of $\mathcal{H}$. For $\mu \in \mathcal{T}_{\mathbb{F}}(K)$ and $x \in \mathscr{X}$, we have

$$\Gamma[P_V L_\mu P_U](x) = \left\langle k_x \,\middle|\, P_V L_\mu P_U[k_x] \right\rangle_{\mathcal{H}} = \int_{\mathscr{X}} K_U(t,x) K_V(x,t) \mathrm{d}\mu(t).$$

From Theorem 3.1 and the properties of Γ, we then obtain

$$\|P_V L_\mu P_U\|^2_{\mathrm{HS}(\mathcal{H})} = \langle \Gamma^* \Gamma[L_\mu] \,|\, P_V L_\mu P_U \rangle_{\mathrm{HS}(\mathcal{H})} = \langle g_\mu \,|\, \Gamma[P_V L_\mu P_U] \rangle_{\mathcal{G}}.$$

For general subspaces $\mathcal{H}_U$ and $\mathcal{H}_V$, the operator $P_V L_\mu P_U$ does not necessarily belong to $\mathcal{I}(\Gamma)$; see Remark 4.3 for an example where this situation occurs. ◁

4.2 Projections defined by measures

Motivated by Lemmas 3.3 and 4.1, for $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we set $\mathcal{H}_\nu = \overline{\mathrm{range}(L_{|\nu|})}^{\mathcal{H}}$, and we denote by P_ν the orthogonal projection from $\mathcal{H}$ onto $\mathcal{H}_\nu$.

Lemma 4.2. *For all $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we have $L_\nu = P_\nu L_\nu = L_\nu P_\nu = P_\nu L_\nu P_\nu$.*

Proof. From Lemma 3.3, we have $L_\nu = P_\nu L_\nu$ and $L_{\overline{\nu}} = P_\nu L_{\overline{\nu}}$. We then obtain $L_\nu = L^*_{\overline{\nu}} = (P_\nu L_{\overline{\nu}})^* = L_\nu P_\nu$, and so $L_\nu = P_\nu L_\nu P_\nu$. □

For an initial operator L_μ, with $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, through the orthogonal projection P_ν and in addition to L_ν, an approximate measure $\nu \in \mathcal{T}_{\mathbb{F}}(K)$ also defines the approximations $P_\nu L_\mu$, $L_\mu P_\nu$ or $P_\nu L_\mu P_\nu$ of L_μ.

Lemma 4.3. *For all μ and $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we have*

$$\|L_\mu - L_\nu\|^2_{\mathrm{HS}(\mathcal{H})} = \|L_\mu - P_\nu L_\mu\|^2_{\mathrm{HS}(\mathcal{H})} + \|P_\nu L_\mu - L_\nu\|^2_{\mathrm{HS}(\mathcal{H})}, \text{ and} \tag{14}$$

$$\|L_\mu - L_\nu\|^2_{\mathrm{HS}(\mathcal{H})} = \|L_\mu - P_\nu L_\mu P_\nu\|^2_{\mathrm{HS}(\mathcal{H})} + \|P_\nu L_\mu P_\nu - L_\nu\|^2_{\mathrm{HS}(\mathcal{H})}. \tag{15}$$

Proof. Using the notations of Remark 4.1, Lemma 4.2 reads $L_\nu \in \mathcal{R}(\mathcal{H}_\nu) \cap \mathcal{Z}(\mathcal{H}_\nu)$. Observing that $L_\mu - P_\nu L_\mu$ is orthogonal to $\mathcal{R}(\mathcal{H}_\nu)$ and that $P_\nu L_\mu - L_\nu \in \mathcal{R}(\mathcal{H}_\nu)$, we obtain (14). In the same way, $L_\mu - P_\nu L_\mu P_\nu$ is orthogonal to $\mathcal{R}(\mathcal{H}_\nu) \cap \mathcal{Z}(\mathcal{H}_\nu)$, and we have $P_\nu L_\mu P_\nu - L_\nu \in \mathcal{R}(\mathcal{H}_\nu) \cap \mathcal{Z}(\mathcal{H}_\nu)$, leading to (15). □

4.3 Error maps on sets of measures

In the framework of Sections 3.2 and 4.2, the characterisation of measures leading to accurate approximations of an initial operator L_μ, $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, relates to the minimisation of error maps measuring the accuracy of the approximations induced by a measure $\nu \in \mathcal{T}_{\mathbb{F}}(K)$.

Quadrature approximation. We define the error map $D_\mu : \mathcal{T}_{\mathbb{F}}(K) \to \mathbb{R}_{\geqslant 0}$, with

$$D_\mu(\nu) = \|L_\mu - L_\nu\|^2_{\mathrm{HS}(\mathcal{H})} = \|g_\mu - g_\nu\|^2_{\mathcal{G}},\ \nu \in \mathcal{T}_{\mathbb{F}}(K).$$

Lemma 4.4. *For $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, the map D_μ is convex on any convex set $\mathscr{C} \subseteq \mathcal{T}_{\mathbb{F}}(K)$. For ν and $\eta \in \mathscr{C}$, the directional derivative of D_μ at ν along $\eta - \nu$ is*

$$\lim_{\rho \to 0^+} \frac{1}{\rho}\big[D_\mu\big(\nu + \rho(\eta - \nu)\big) - D_\mu(\nu)\big] = 2\Re\big(\langle g_\nu - g_\mu \,|\, g_\eta - g_\nu\rangle_{\mathcal{G}}\big).$$

Proof. For $\xi = (1-\rho)\nu + \rho\eta$, ν and $\eta \in \mathscr{C}$, $\rho \in [0,1]$, we have $g_\xi = (1-\rho)g_\nu + \rho g_\eta$; the convexity of D_μ on $\mathscr{C}$ then follows from the convexity of the map $g \mapsto \|g_\mu - g\|^2_{\mathcal{G}}$ on $\mathcal{G}$. Next, the expansion of the squared norm $\|g_\mu - g_\nu - \rho(g_\eta - g_\nu)\|^2_{\mathcal{G}}$ provides the expected expression for the directional derivatives of D_μ. □

Projection-based approximation. We denote by C^{P}_μ and $C^{\mathrm{PP}}_\mu : \mathcal{T}_{\mathbb{F}}(K) \to \mathbb{R}_{\geqslant 0}$ the error maps defined as

$$C^{\mathrm{P}}_\mu(\nu) = \|L_\mu - P_\nu L_\mu\|^2_{\mathrm{HS}(\mathcal{H})} \text{ and } C^{\mathrm{PP}}_\mu(\nu) = \|L_\mu - P_\nu L_\mu P_\nu\|^2_{\mathrm{HS}(\mathcal{H})},\ \nu \in \mathcal{T}_{\mathbb{F}}(K);$$

we may notice that $C^{\mathrm{X}}_\mu(\nu) = C^{\mathrm{X}}_\mu(|\nu|)$, $\mathrm{X} \in \{\mathrm{P}, \mathrm{PP}\}$.

Theorem 4.1. *For $\mu \in \mathcal{T}_{\mathbb{F}}(K)$ and $\mathrm{X} \in \{\mathrm{P}, \mathrm{PP}\}$, the map C^{X}_μ is convex on the real convex cone $\mathcal{T}_+(K)$, and for all ν and $\eta \in \mathcal{T}_+(K)$, we have*

$$\lim_{\rho \to 0^+} \frac{1}{\rho}\big[C^{\mathrm{X}}_\mu\big(\nu + \rho(\eta - \nu)\big) - C^{\mathrm{X}}_\mu(\nu)\big] \in \{-\infty, 0\}.$$

Proof. For $\nu, \eta \in \mathcal{T}_+(K)$ and $\rho \in (0,1)$, we set $\xi = \nu + \rho(\eta - \nu) \in \mathcal{T}_+(K)$. The three operators L_ν, L_η and L_ξ being PSD, independently of $\rho \in (0,1)$, we have $\mathrm{null}(L_\xi) = \mathrm{null}(L_\nu) \cap \mathrm{null}(L_\eta)$, and so $\mathcal{H}_\xi = \overline{\mathcal{H}_\nu + \mathcal{H}_\eta}^{\mathcal{H}}$. The two maps

$$\rho \mapsto C^{\mathrm{P}}_\mu\big(\nu + \rho(\eta - \nu)\big) \quad \text{and} \quad \rho \mapsto C^{\mathrm{PP}}_\mu\big(\nu + \rho(\eta - \nu)\big)$$

are therefore constant on the open interval $(0,1)$. From Lemma 4.1 and (10), noticing that $\mathcal{H}_\nu \subseteq \mathcal{H}_\xi$ and $\mathcal{H}_\eta \subseteq \mathcal{H}_\xi$, we obtain $C^{\mathrm{P}}_\mu(\xi) \leqslant C^{\mathrm{P}}_\mu(\nu)$ and $C^{\mathrm{P}}_\mu(\xi) \leqslant C^{\mathrm{P}}_\mu(\eta)$; from (11), we also get $C^{\mathrm{PP}}_\mu(\xi) \leqslant C^{\mathrm{PP}}_\mu(\nu)$ and $C^{\mathrm{PP}}_\mu(\xi) \leqslant C^{\mathrm{PP}}_\mu(\eta)$, concluding the proof. □

In view of Theorem 4.1, the maps C^{P}_μ and C^{PP}_μ are akin to piecewise-constant functions. By contrast (see Lemma 4.4), the directional derivatives of the map D_μ are *informative*, in the sense that the landscape of D_μ can be explored through steepest descents. From Remark 4.1 and Lemma 4.3, we have

$$C^{\mathrm{P}}_\mu(\nu) \leqslant C^{\mathrm{PP}}_\mu(\nu) \leqslant D_\mu(\nu), \nu \in \mathcal{T}_{\mathbb{F}}(K), \tag{16}$$

with $C^{\mathrm{P}}_\mu(\mu) = D_\mu(\mu) = 0$ and $C^{\mathrm{P}}_\mu(0) = D_\mu(0) = \|g_\mu\|^2_{\mathcal{G}}$ (see also Remark 4.3). The quadrature-approximation error map D_μ may hence be regarded as a *differentiable relaxation* of the projection-based-approximation error maps C^{P}_μ and C^{PP}_μ; see Figure 1 for an illustration.

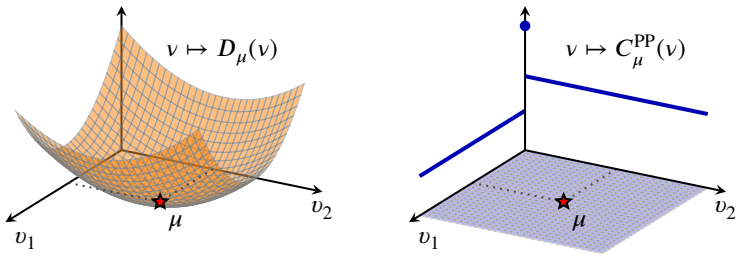

Figure 1: Graphical representation of the maps D_μ and C_μ^{PP} as functions of the weight parameters characterising an approximate measure $\nu \in \mathcal{T}_+(K)$. The measures μ and ν are supported by the same set of points $\{x_1, x_2\} \subseteq \mathscr{X}$, and described by their weight parameters (ω_1, ω_2) and $(\upsilon_1, \upsilon_2) \in \mathbb{R}^2_{\geqslant 0}$, respectively; the red star represents the weight parameters of $\mu = \omega_1 \delta_{x_1} + \omega_2 \delta_{x_2}$. The presented graphs correspond to the case $\omega_1 = \omega_2 = 1$, with K such that $K(x_1, x_1) = 1.225$, $K(x_2, x_2) = 0.894$ and $K(x_1, x_2) = 0.316$. In the graph of C_μ^{PP}, the point on the vertical axis indicates the value of the map at $\nu = 0$, and the bold lines indicate the constant values taken by the map along the horizontal axes (and following Remark 4.3, the graph of C_μ^{PP} is tangent to the graph of D_μ along the horizontal axes).

Remark 4.3. For $\mu \in \mathcal{T}_{\mathbb{F}}(K)$ and $s \in \mathscr{X}$, introducing $c_{\delta_s} = g_\mu(s)/|K(s,s)|^2$ if $K(s,s) > 0$, and $c_{\delta_s} = 0$ otherwise, we have $P_{\delta_s} L_\mu P_{\delta_s} = c_{\delta_s} S_{k_s}$. For $K(s,s) = 0$, we indeed have $k_s = 0$, and so $P_{\delta_s} = 0$, and for $K(s,s) > 0$,

$$P_{\delta_s} L_\mu P_{\delta_s}[h](x) = \frac{K(x,s)h(s)}{|K(s,s)|^2} \int_{\mathscr{X}} |K(s,t)|^2 \mathrm{d}\mu(t), \ h \in \mathcal{H} \text{ and } x \in \mathscr{X}.$$

We obtain $P_{\delta_s} L_\mu P_{\delta_s} \in \mathcal{I}(\Gamma)$ and $C_\mu^{\mathrm{PP}}(\delta_s) = D_\mu(c_{\delta_s}\delta_s)$, $s \in \mathscr{X}$. ◁

Remark 4.4. From a numerical standpoint, in view of (12) and (13), for $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, the evaluation of $C_\mu^{\mathrm{P}}(\nu)$ or $C_\mu^{\mathrm{PP}}(\nu)$ requires a suitable characterisation of the reproducing kernel K_ν of $\mathcal{H}_\nu$ (or equivalently, of the orthogonal projection P_ν); in practice, K_ν is a priori unknown and needs to be computed from K and ν (see Remark 4.5). In comparison and in view of (7), the error map D_μ only involves the kernel K; the *projection-free* nature of D_μ is of notable interest for numerical applications. ◁

Remark 4.5. Following Lemma 3.4, for a measure ν supported by $S = \{s_1, \cdots, s_n\}$, $n \in \mathbb{N}$, the reproducing kernel K_ν of $\mathcal{H}_\nu$ can be expressed as

$$K_\nu(x,t) = \sum_{i,j=1}^{n} K(x, s_i) \varkappa_{i,j} K(s_j, t), \ x \text{ and } t \in \mathscr{X},$$

where $\varkappa_{i,j}$ is the i,j entry of the pseudoinverse (Moore-Penrose inverse) of the $n \times n$ kernel matrix with i,j entry $K(s_i, s_j)$. The worst-case time complexity of the evaluation of K_ν at $M \in \mathbb{N}$ distinct locations in $\mathscr{X} \times \mathscr{X}$ is thus $\mathcal{O}(n^3 + n^2 M)$. The

term $\mathcal{O}(n^3)$ is related to the pseudoinversion of the kernel matrix defined by K and S, while the term $\mathcal{O}(n^2 M)$ corresponds to the evaluation, from this pseudoinverse and the kernel K, of K_ν at M different locations. ◁

5 Nonnegative measures and L^2-embeddings

Following Section 3.3, for $\mu \in \mathcal{T}_+(K)$, the embedding $\imath_\mu : \mathcal{H} \to L^2(\mu)$ is HS. For $f \in L^2(\mu)$ and $x \in \mathscr{X}$, we have

$$\langle k_x \mid \imath_\mu^*[f] \rangle_{\mathcal{H}} = \langle \imath_\mu[k_x] \mid f \rangle_{L^2(\mu)} = \int_{\mathscr{X}} K(x,t) f(t) \mathrm{d}\mu(t),$$

so that in addition to $L_\mu = \imath_\mu^* \imath_\mu \in \mathrm{HS}(\mathcal{H})$, the three operators

$$\imath_\mu^* : L^2(\mu) \to \mathcal{H}, \quad \imath_\mu \imath_\mu^* : L^2(\mu) \to L^2(\mu), \quad \text{and} \quad \imath_\mu \imath_\mu^* \imath_\mu : \mathcal{H} \to L^2(\mu), \tag{17}$$

can also be regarded as integral operators defined by the kernel K and the nonnegative measure μ. These four interpretations are inherent to K, which characterises $\mathcal{H}$, and μ, which characterises $L^2(\mu)$; see for instance [4, 22, 19, 26, 20] for illustrations. In each case, the corresponding operator is HS, and we denote by $\mathrm{HS}(\mu, \mathcal{H})$, $\mathrm{HS}(\mu)$ and $\mathrm{HS}(\mathcal{H}, \mu)$ the Hilbert spaces of all HS operators from $L^2(\mu)$ to $\mathcal{H}$, on $L^2(\mu)$, and from $\mathcal{H}$ to $L^2(\mu)$, respectively.

5.1 Partial L^2-embeddings

For a closed linear subspace $\mathcal{H}_S \subseteq \mathcal{H}$, the embedding $\imath_\mu$ can be approximated by the *partial embedding* $\imath_\mu P_S$. For $f \in L^2(\mu)$ and $x \in \mathscr{X}$, we have

$$\langle k_x \mid P_S \imath_\mu^*[f] \rangle_{\mathcal{H}} = \langle \imath_\mu P_S[k_x] \mid f \rangle_{L^2(\mu)} = \int_{\mathscr{X}} K_S(x,t) f(t) \mathrm{d}\mu(t),$$

so that $P_S \imath_\mu^*$ corresponds to an integral operator with kernel K_S. In the decomposition $L_\mu = \imath_\mu^* \imath_\mu \in \mathrm{HS}(\mathcal{H})$, substituting each $\imath_\mu$ with $\imath_\mu P_S$ gives the approximation $P_S \imath_\mu^* \imath_\mu P_S$ discussed in Section 4. For the operators defined in (17), a similar substitution yields the approximations

$$P_S \imath_\mu^* \in \mathrm{HS}(\mu, \mathcal{H}), \quad \imath_\mu P_S \imath_\mu^* \in \mathrm{HS}(\mu), \quad \text{and} \quad \imath_\mu P_S \imath_\mu^* \imath_\mu P_S \in \mathrm{HS}(\mathcal{H}, \mu);$$

see also Remark 5.1. In what follows, we mainly focus on the approximations related to $\mathrm{HS}(\mu, \mathcal{H})$ and $\mathrm{HS}(\mu)$; the case of $\mathrm{HS}(\mathcal{H}, \mu)$ is more briefly discussed in Remark 5.2.

Remark 5.1. In addition to $P_S \imath_\mu^* \imath_\mu P_S$, the approximation of $\imath_\mu$ by $\imath_\mu P_S$ gives rise to the approximations $P_S \imath_\mu^* \imath_\mu$ and $\imath_\mu^* \imath_\mu P_S$ of $\imath_\mu^* \imath_\mu \in \mathrm{HS}(\mathcal{H})$ (see Section 4). Similarly, for $\imath_\mu \imath_\mu^* \imath_\mu \in \mathrm{HS}(\mathcal{H}, \mu)$ and in addition to $\imath_\mu P_S \imath_\mu^* \imath_\mu P_S$, the approximations $\imath_\mu P_S \imath_\mu^* \imath_\mu$ and $\imath_\mu \imath_\mu^* \imath_\mu P_S$ of $\imath_\mu \imath_\mu^* \imath_\mu \in \mathrm{HS}(\mathcal{H}, \mu)$ may be considered. In these approximations, the substitution of $\imath_\mu$ with $\imath_\mu P_S$ is not applied invariably. ◁

Lemma 5.1. *Let $\mathcal{H}_S$ be a closed linear subspace of $\mathcal{H}$. For $\mu \in \mathcal{T}_+(K)$, we have*

$$\|\imath_\mu^* - P_S \imath_\mu^*\|^2_{\mathrm{HS}(\mu,\mathcal{H})} = \int_{\mathscr{X}} K(t,t) - K_S(t,t)\mathrm{d}\mu(t), \text{ and} \tag{18}$$

$$\|\imath_\mu \imath_\mu^* - \imath_\mu P_S \imath_\mu^*\|^2_{\mathrm{HS}(\mu)} = \iint_{\mathscr{X}} \big|K(x,t) - K_S(x,t)\big|^2 \mathrm{d}\mu(t)\mathrm{d}\mu(x). \tag{19}$$

Proof. Let $\mathcal{H}_{0S}$ be the orthogonal complement of $\mathcal{H}_S$ in $\mathcal{H}$; endowed with the Hilbert structure of $\mathcal{H}$, $\mathcal{H}_{0S}$ is a RKHS, and $K_{0S} = K - K_S$. For an ONB $\{h_j\}_{j\in\mathbb{I}}$ of $\mathcal{H}$, $\mathbb{I} \subseteq \mathbb{N}$, we have

$$\|\imath_\mu^* - P_S \imath_\mu^*\|^2_{\mathrm{HS}(\mu,\mathcal{H})} = \|\imath_\mu P_{0S}\|^2_{\mathrm{HS}(\mathcal{H},\mu)} = \sum_{j\in\mathbb{I}} \int_{\mathscr{X}} \big|P_{0S}[h_j](t)\big|^2 \mathrm{d}\mu(t);$$

since $\sum_{j\in\mathbb{I}} \big|P_{0S}[h_j](t)\big|^2 = K_{0S}(t,t)$, $t \in \mathscr{X}$ (see e.g. [17]), equality (18) follows by monotone convergence. We also have

$$\|\imath_\mu \imath_\mu^* - \imath_\mu P_S \imath_\mu^*\|^2_{\mathrm{HS}(\mu)} = \|\imath_\mu P_{0S} \imath_\mu^*\|^2_{\mathrm{HS}(\mu)} = \|P_{0S} L_\mu P_{0S}\|^2_{\mathrm{HS}(\mathcal{H})},$$

so that (19) follows from Lemma A.1 (we recall that $L_\mu = \imath_\mu^* \imath_\mu$). □

Remark 5.2. We consider four closed linear subspaces $\mathcal{H}_R, \mathcal{H}_S, \mathcal{H}_U$ and $\mathcal{H}_V$ of $\mathcal{H}$, and let $\{h_j\}_{j\in\mathbb{I}}$ be an ONB of $\mathcal{H}$. For $\mu \in \mathcal{T}_+(K)$, we have

$$\begin{aligned}&\langle \imath_\mu P_R \imath_\mu^* \imath_\mu P_S \mid \imath_\mu P_U \imath_\mu^* \imath_\mu P_V \rangle_{\mathrm{HS}(\mathcal{H},\mu)} \\ &\quad = \sum_{j\in\mathbb{I}} \iiint_{\mathscr{X}} P_V[h_j](t)\overline{P_S[h_j](s)} K_R(s,x) K_U(x,t) \mathrm{d}\mu(s)\mathrm{d}\mu(t)\mathrm{d}\mu(x),\end{aligned} \tag{20}$$

with, form the CS inequality in $\ell^2(\mathbb{I})$ and in $\mathcal{H}$,

$$\iiint_{\mathscr{X}} \sum_{j\in\mathbb{I}} \big|P_V[h_j](t)\big|\big|P_S[h_j](s)\big|\big|K_R(s,x)\big|\big|K_U(x,t)\big| \mathrm{d}\mu(s)\mathrm{d}\mu(t)\mathrm{d}\mu(x) \leqslant \tau_\mu^3.$$

From (20) and Fubini's theorem, we then for instance obtain

$$\begin{aligned}\|\imath_\mu \imath_\mu^* \imath_\mu - \imath_\mu P_S \imath_\mu^* \imath_\mu P_S\|^2_{\mathrm{HS}(\mathcal{H},\mu)} = \iiint_{\mathscr{X}} \Big[&K(t,s)K(s,x)K(x,t) \\ &+ K_S(t,s)K_S(s,x)\big(K_S(x,t) - 2K(x,t)\big)\Big] \mathrm{d}\mu(s)\mathrm{d}\mu(t)\mathrm{d}\mu(x),\end{aligned}$$

where we should observe that $\langle \imath_\mu \imath_\mu^* \imath_\mu \mid \imath_\mu P_S \imath_\mu^* \imath_\mu P_S \rangle_{\mathrm{HS}(\mathcal{H},\mu)} = \|\imath_\mu P_S \imath_\mu^* \imath_\mu\|^2_{\mathrm{HS}(\mathcal{H},\mu)}$. ◁

The following inequality (Lemma 5.2) holds between the approximations in $\mathrm{HS}(\mu)$ and $\mathrm{HS}(\mathcal{H})$ defined by a subspace $\mathcal{H}_S$. We recall that $\|\imath_\mu \imath_\mu^*\|_{\mathrm{HS}(\mu)} = \|\imath_\mu^* \imath_\mu\|_{\mathrm{HS}(\mathcal{H})}$, and that $\|\imath_\mu^* \imath_\mu - P_S \imath_\mu^* \imath_\mu\|_{\mathrm{HS}(\mathcal{H})} \leqslant \|\imath_\mu^* \imath_\mu - P_S \imath_\mu^* \imath_\mu P_S\|_{\mathrm{HS}(\mathcal{H})}$ (see Remark 4.1).

Lemma 5.2. *Let $\mathcal{H}_S$ be a closed linear subspace of $\mathcal{H}$. For all $\mu \in \mathcal{T}_+(K)$, we have*

$$\|\imath_\mu \imath_\mu^* - \imath_\mu P_S \imath_\mu^*\|_{\mathrm{HS}(\mu)} \leqslant \|\imath_\mu^* \imath_\mu - P_S \imath_\mu^* \imath_\mu\|_{\mathrm{HS}(\mathcal{H})}.$$

Proof. We have $\|\imath_\mu\imath_\mu^* - \imath_\mu P_S \imath_\mu^*\|_{\mathrm{HS}(\mu)} = \|P_{0S} L_\mu P_{0S}\|_{\mathrm{HS}(\mathcal{H})}$, with $P_{0S} = \mathrm{id}_\mathcal{H} - P_S$. The operators $P_{0S} L_\mu P_{0S}$ and $P_{0S} L_\mu P_S$ being orthogonal in $\mathrm{HS}(\mathcal{H})$, we obtain

$$\|P_{0S} L_\mu (P_{0S} + P_S)\|^2_{\mathrm{HS}(\mathcal{H})} = \|P_{0S} L_\mu P_{0S}\|^2_{\mathrm{HS}(\mathcal{H})} + \|P_{0S} L_\mu P_S\|^2_{\mathrm{HS}(\mathcal{H})},$$

and so $\|\imath_\mu\imath_\mu^* - \imath_\mu P_S \imath_\mu^*\|_{\mathrm{HS}(\mu)} \leqslant \|\imath_\mu^*\imath_\mu - P_S \imath_\mu^*\imath_\mu\|_{\mathrm{HS}(\mathcal{H})}$. □

5.2 Trace and Frobenius error maps

Following Section 4.3 and considering subspaces of $\mathcal{H}$ defined by measures, we introduce the error maps C^{tr}_μ and $C^{\mathrm{F}}_\mu : \mathcal{T}_\mathbb{F}(K) \to \mathbb{R}_{\geqslant 0}$, with

$$C^{\mathrm{tr}}_\mu(\nu) = \|\imath_\mu^* - P_\nu \imath_\mu^*\|^2_{\mathrm{HS}(\mu,\mathcal{H})} \text{ and } C^{\mathrm{F}}_\mu(\nu) = \|\imath_\mu\imath_\mu^* - \imath_\mu P_\nu \imath_\mu^*\|^2_{\mathrm{HS}(\mu)},\ \nu \in \mathcal{T}_\mathbb{F}(K).$$

The notations C^{tr}_μ and C^{F}_μ are motivated by the relation between these maps and the trace and Frobenius norms; see Section 5.3. As observed for C^{P}_μ and C^{PP}_μ, we may notice that $C^{\mathrm{X}}_\mu(\nu) = C^{\mathrm{X}}_\mu(|\nu|)$, $\mathrm{X} \in \{\mathrm{tr}, \mathrm{F}\}$

Theorem 5.1. *For $\mu \in \mathcal{T}_+(K)$, the statement of Theorem 4.1 also holds for the maps C^{tr}_μ and C^{F}_μ; that is, these two maps are convex on the real convex cone $\mathcal{T}_+(K)$, and their directional derivatives take values in the set $\{-\infty, 0\}$.*

Proof. We follow the same reasoning as in the proof of Theorem 4.1. For two measures ν and $\eta \in \mathcal{T}_+(K)$ and for $\rho \in (0,1)$, we set $\xi = \nu + \rho(\eta - \nu) \in \mathcal{T}_+(K)$. We then have $\mathcal{H}_\xi = \overline{\mathcal{H}_\nu + \mathcal{H}_\eta}^{\mathcal{H}}$ independently of $\rho \in (0,1)$. We conclude by combining the inclusions $\mathcal{H}_\nu \subseteq \mathcal{H}_\xi$ and $\mathcal{H}_\eta \subseteq \mathcal{H}_\xi$ with the inequalities provided in Lemma A.2 (Appendix A). □

As illustrated by Lemma 5.1 and Theorem 5.1, and as already observed for the error maps C^{P}_μ and C^{PP}_μ, the error maps C^{tr}_μ and C^{F}_μ are akin to piecewise-constant functions, and their evaluation requires a suitable characterisation of the kernel of subspaces of $\mathcal{H}$. For $\mu \in \mathcal{T}_+(K)$, the error maps C^{X}_μ, $\mathrm{X} \in \{\mathrm{tr}, \mathrm{F}, \mathrm{P}, \mathrm{PP}\}$ can be regarded as alternative ways to asses the accuracy of the approximation of $\imath_\mu$ by $\imath_\mu P_\nu$, $\nu \in \mathcal{T}_\mathbb{F}(K)$. From the relation between the error maps D_μ and C^{X}_μ, $\mathrm{X} \in \{\mathrm{P}, \mathrm{PP}\}$ (see Lemma 4.3) the approximation of potentials in $\mathcal{G}$ can hence more generally be regarded as a differentiable and projection-free surrogate for the characterisation of accurate partial embeddings. From Lemma 5.2, we may notice that $C^{\mathrm{F}}_\mu(\nu) \leqslant C^{\mathrm{P}}_\mu(\nu)$, $\nu \in \mathcal{T}_\mathbb{F}(K)$, extending the sequence of inequalities (16).

Remark 5.3. Let $\nu \in \mathcal{T}_\mathbb{C}(K)$ be a complex measure with real and imaginary parts $\nu_\mathbb{r}$ and $\nu_\mathbb{i} \in \mathcal{T}(K)$. For $\mu \in \mathcal{T}(K)$, the three operators L_μ, $L_{\nu_\mathbb{r}}$ and $L_{\nu_\mathbb{i}}$ are self-adjoint; we thus have $\|L_\mu - L_\nu\|^2_{\mathrm{HS}(\mathcal{H})} = \|L_\mu - L_{\nu_\mathbb{r}}\|^2_{\mathrm{HS}(\mathcal{H})} + \|L_{\nu_\mathbb{i}}\|^2_{\mathrm{HS}(\mathcal{H})}$, and so $D_\mu(\nu_\mathbb{r}) \leqslant D_\mu(\nu)$. Hence, when L_μ is self-adjoint, the search of an approximate measure ν for the approximation of L_μ by L_ν can be restricted to $\mathcal{T}(K)$. ◁

5.3 Column sampling for PSD-matrix approximation

Let $\mathbf{K}$ be a $N \times N$ PSD matrix, with $N \in \mathbb{N}$; we denote by $[N]$ the set of all integers between 1 and N. For a subset $I \subseteq [N]$ the *Nyström approximation*[2] of $\mathbf{K}$ induced by I is the $N \times N$ PSD matrix

$$\hat{\mathbf{K}}(I) = \mathbf{K}_{\bullet,I}(\mathbf{K}_{I,I})^{\dagger}\mathbf{K}_{I,\bullet}, \tag{21}$$

where $\mathbf{K}_{\bullet,I}$ is the matrix defined by the columns of $\mathbf{K}$ with index in I, and where $(\mathbf{K}_{I,I})^{\dagger}$ is the pseudoinverse of the principal submatrix of $\mathbf{K}$ defined by I (and $\mathbf{K}_{I,\bullet}$ consists of rows of $\mathbf{K}$); see e.g. [9, 18, 14, 11].

For i and $j \in [N]$, the i, j entry of $\mathbf{K}$ may be regarded as the value $K(i, j)$ of a PSD kernel K defined on the discrete set $\mathscr{X} = [N]$. The j-th column of $\mathbf{K}$ then corresponds to the function $k_j \in \mathcal{H}$, $j \in \mathscr{X}$, and the subset I defines the closed linear subspace $\mathcal{H}_I = \mathrm{span}_{\mathbb{C}}\{k_j | j \in I\} \subseteq \mathcal{H}$; in particular, the i, j entry of $\hat{\mathbf{K}}(I)$ is $K_I(i, j)$, with K_I the reproducing kernel of $\mathcal{H}_I$ (see e.g. [17], and Remark 4.5).

Introducing $\mu = \sum_{i=1}^{N} \delta_i$, the Hilbert space $L^2(\mu)$ can be identified with the Euclidean space $\mathbb{C}^N$; following Section 5.2, we then observe that

- the trace norm $\|\mathbf{K} - \hat{\mathbf{K}}(I)\|_{\mathrm{tr}}$ corresponds to (18), and
- the squared Frobenius norm $\|\mathbf{K} - \hat{\mathbf{K}}(I)\|_{\mathrm{F}}^2$ corresponds to (19).

The column-sampling problem for the Nyström approximation of a PSD matrix $\mathbf{K}$, that is, the search of a subset $I \subseteq [N]$ leading to an accurate approximation $\hat{\mathbf{K}}(I)$ of $\mathbf{K}$, is thus a special instance of the general framework discussed in Section 5.1. In particular, the support of an approximate measure ν on $\mathscr{X} = [N]$ defines a subset of columns of $\mathbf{K}$, and the approximation of potentials in the RKHS $\mathcal{G}$ may be used as surrogate for the characterisation of such measures. In the discrete setting, $\mathcal{G}$ corresponds to the RKHS defined by the $N \times N$ PSD matrix $\mathbf{S}$ with i, j entry $|\mathbf{K}_{i,j}|^2$ (that is, $\mathbf{S}$ is the element-wise product between $\mathbf{K}$ and $\overline{\mathbf{K}}$, the conjugate of $\mathbf{K}$).

6 Concluding discussion

We described the overall framework surrounding the isometric representation of integral operators with PSD kernels as potentials, and illustrated the equivalence between the quadrature approximation of such integral operators and the approximation of integral functionals on RKHSs with squared-modulus kernels. Through subspaces defined by measures and partial L^2-embeddings, we also discussed the extent to which the approximation of potentials in RKHSs with squared-modulus kernels can be used as a differentiable surrogate for the characterisation of projection-based approximation of integral operators with PSD kernels.

[2] In the machine-learning literature, Nyström approximation refers to the low-rank approximation of PSD matrices through column sampling; although related, this terminology should not to be confused with the quadrature method for the approximation of integral equations.

The link between integral-operator approximation and potential approximation may be leveraged to design sampling strategies for low-rank approximation (where approximations are characterised by sparse finitely-supported measures). The direct minimisation of D_μ under sparsity-inducing constraints is for instance considered in [10], while the possibility to locally optimised the support of approximate measures using particle-flow techniques is studied in [13]. Sequential approaches, where support points are added one-at-a-time on the basis of information provided by the directional derivatives of D_μ, are investigated in [12]. The present work aims at supporting this type of approaches by strengthening their theoretical underpinning.

A Technical results

Lemma A.1. *Let $\mathcal{H}_U$ and $\mathcal{H}_V$ be two closed linear subspaces of $\mathcal{H}$. For all $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, we have*

$$\|P_V L_\mu P_U\|^2_{\mathrm{HS}(\mathcal{H})} = \iint_{\mathscr{X}} K_U(t,x) K_V(x,t) \mathrm{d}\mu(t) \mathrm{d}\overline{\mu}(x).$$

Proof. We consider an ONB $\{h_j\}_{j\in\mathbb{I}}$ of $\mathcal{H}$, $\mathbb{I} \subseteq \mathbb{N}$. From (6), we have

$$\begin{aligned} \|P_V L_\mu P_U\|^2_{\mathrm{HS}(\mathcal{H})} &= \sum_{j\in\mathbb{I}} \langle L_\mu P_U[h_i] \mid P_V L_\mu P_U[h_j] \rangle_{\mathcal{H}} \\ &= \sum_{j\in\mathbb{I}} \iint_{\mathscr{X}} P_U[h_j](t) \overline{P_U[h_j](x)} K_V(x,t) \mathrm{d}\mu(t) \mathrm{d}\overline{\mu}(x). \end{aligned} \tag{22}$$

As $\sum_{j\in\mathbb{I}} P_U[h_j](t)\overline{P_U[h_j](x)} = K_U(t,x)$, x and $t \in \mathscr{X}$ (see e.g. [17]), and since $K_U(t,t) \leqslant K(t,t)$ and $K_V(t,t) \leqslant K(t,t)$, from the CS inequality in $\ell^2(\mathbb{I})$ and in $\mathcal{H}$, we obtain

$$\iint_{\mathscr{X}} \sum_{j\in\mathbb{I}} \big|P_U[h_j](t)\big|\big|P_U[h_j](x)\big|\big|K_V(x,t)\big| \mathrm{d}|\mu|(t) \mathrm{d}|\mu|(x) \leqslant \tau_\mu^2;$$

the result then follows form (22) and Fubini's theorem. □

Lemma A.2. *Let $\mathcal{H}_S$ and $\mathcal{H}_R$ be two closed linear subspaces of $\mathcal{H}$; we assume that $\mathcal{H}_R \subseteq \mathcal{H}_S$. For all $\mu \in \mathcal{T}_+(K)$, we have $\|\imath_\mu^* - P_S \imath_\mu^*\|_{\mathrm{HS}(\mu,\mathcal{H})} \leqslant \|\imath_\mu^* - P_R \imath_\mu^*\|_{\mathrm{HS}(\mu,\mathcal{H})}$ and $\|\imath_\mu \imath_\mu^* - \imath_\mu P_S \imath_\mu^*\|_{\mathrm{HS}(\mu)} \leqslant \|\imath_\mu \imath_\mu^* - \imath_\mu P_R \imath_\mu^*\|_{\mathrm{HS}(\mu)}$.*

Proof. We denote by $\mathcal{H}_e$ the orthogonal complement of $\mathcal{H}_R$ in $\mathcal{H}_S$. Noticing that $P_S = P_R + P_e$ and that $\langle \imath_\mu^* - P_R \imath_\mu^* \mid P_e \imath_\mu^* \rangle_{\mathrm{HS}(\mu,\mathcal{H})} = \|P_e \imath_\mu^*\|^2_{\mathrm{HS}(\mu,\mathcal{H})}$, we obtain

$$\|\imath_\mu^* - P_S \imath_\mu^*\|^2_{\mathrm{HS}(\mu,\mathcal{H})} = \|\imath_\mu^* - P_R \imath_\mu^*\|^2_{\mathrm{HS}(\mu,\mathcal{H})} - \|P_e \imath_\mu^*\|^2_{\mathrm{HS}(\mu,\mathcal{H})}.$$

Denoting by $\mathcal{H}_{0S}$ and $\mathcal{H}_{0R}$ the orthogonal complements of $\mathcal{H}_S$ and $\mathcal{H}_R$ in $\mathcal{H}$, respectively, we have $\mathcal{H}_{0S} \subseteq \mathcal{H}_{0R}$; Lemma 4.1 then gives

$$\|\imath_\mu P_{0S} \imath_\mu^*\|_{\mathrm{HS}(\mu)} = \|P_{0S} L_\mu P_{0S}\|_{\mathrm{HS}(\mathcal{H})} \leqslant \|P_{0R} L_\mu P_{0R}\|_{\mathrm{HS}(\mathcal{H})} = \|\imath_\mu P_{0R} \imath_\mu^*\|_{\mathrm{HS}(\mu)},$$

completing the proof. □

References

[1] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.

[2] Francis Bach. On the equivalence between kernel quadrature rules and random feature expansions. *Journal of Machine Learning Research*, 18(1):714–751, 2017.

[3] John B. Conway. *A Course in Functional Analysis*, volume 96. Springer, 2019.

[4] Felipe Cucker and Steve Smale. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39:1–49, 2002.

[5] Felipe Cucker and Ding Xuan Zhou. *Learning Theory: an Approximation Theory Viewpoint*. Cambridge University Press, 2007.

[6] Steven B. Damelin, Fred J. Hickernell, David L. Ragozin, and Xiaoyan Zeng. On energy, discrepancy and group invariant measures on measurable subsets of Euclidean space. *Journal of Fourier Analysis and Applications*, 16:813–839, 2010.

[7] Kenneth R. Davidson. *Nest Algebras: Triangular Forms for Operator Algebras on Hilbert Space*, volume 191. Longman Scientific and Technical, 1988.

[8] Joseph Diestel and John J. Uhl. *Vector Measures*, volume 15. American Mathematical Society, 1977.

[9] Petros Drineas and Michael W. Mahoney. On the Nyström method for approximating a Gram matrix for improved kernel-based learning. *Journal of Machine Learning Research*, 6:2153–2175, 2005.

[10] Bertrand Gauthier and Johan Suykens. Optimal quadrature-sparsification for integral operator approximation. *SIAM Journal on Scientific Computing*, 40:A3636–A3674, 2018.

[11] Alex Gittens and Michael W. Mahoney. Revisiting the Nyström method for improved large-scale machine learning. *Journal of Machine Learning Research*, 17:1–65, 2016.

[12] Matthew Hutchings and Bertrand Gauthier. Energy-based sequential sampling for low-rank PSD-matrix approximation. *Preprint,* `https://hal.science/hal-04102664`, 2023.

[13] Matthew Hutchings and Bertrand Gauthier. Local optimisation of Nyström samples through stochastic gradient descent. In *Machine Learning, Optimization, and Data Science - LOD 2022*, pages 123–140, 2023.

[14] Sanjiv Kumar, Mehryar Mohri, and Ameet Talwalkar. Sampling methods for the Nyström method. *Journal of Machine Learning Research*, 13:981–1006, 2012.

[15] Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: a review and beyond. *Foundations and Trends® in Machine Learning*, 10:1–141, 2017.

[16] Alfred Müller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29(2):429–443, 1997.

[17] Vern I. Paulsen and Mrinal Raghupathi. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. Cambridge University Press, 2016.

[18] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT press, Cambridge, MA, 2006.

[19] Lorenzo Rosasco, Mikhail Belkin, and Ernesto De Vito. On learning with integral operators. *Journal of Machine Learning Research*, 11:905–934, 2010.

[20] Gabriele Santin and Robert Schaback. Approximation of eigenfunctions in kernel-based spaces. *Advances in Computational Mathematics*, 42:973–993, 2016.

[21] Steve Smale and Ding-Xuan Zhou. Learning theory estimates via integral operators and their approximations. *Constructive Approximation*, 26:153–172, 2007.

[22] Steve Smale and Ding-Xuan Zhou. Geometry on probability spaces. *Constructive Approximation*, 30:311–323, 2009.

[23] Bharath K. Sriperumbudur, Kenji Fukumizu, Arthur Gretton, Bernhard Schölkopf, and Gert R.G. Lanckriet. On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics*, 6:1550–1599, 2012.

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

[25] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer, 2008.

[26] Ingo Steinwart and Clint Scovel. Mercer's theorem on general domains: on the interaction between measures, kernels, and RKHSs. *Constructive Approximation*, 35:363–417, 2012.

[27] Ingo Steinwart and Johanna F. Ziegel. Strictly proper kernel scores and characteristic kernels on compact spaces. *Applied and Computational Harmonic Analysis*, 51:510–542, 2021.

[28] Kösaku Yosida. *Functional Analysis*. Springer, 2012.