



HAL
open science

Isometric representation of integral operators with positive-semidefinite kernels

Bertrand Gauthier

► **To cite this version:**

Bertrand Gauthier. Isometric representation of integral operators with positive-semidefinite kernels. 2023. hal-03848105v2

HAL Id: hal-03848105

<https://hal.science/hal-03848105v2>

Preprint submitted on 30 Apr 2023 (v2), last revised 7 Apr 2024 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Isometric representation of integral operators with positive-semidefinite kernels

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} . We discuss the properties of this coisometry, and show that trace-class integral operators defined by general measures and the reproducing kernel of \mathcal{H} always belong to its initial space. The images of such operators are the Riesz representations of integral functionals on \mathcal{G} , drawing a direct connection between the approximation of trace-class integral operators with PSD kernels and the approximation of integral functionals on RKHSs associated with squared-modulus kernels.

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

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

Contents

1	Introduction	2
2	Framework and notations	3
2.1	Hilbert-Schmidt operators and RKHSs	3
2.2	Self-adjoint operators and real-valuedness	6
2.3	Positive-semidefinite operators and nonnegativity	7
3	Trace-class integral operators with PSD kernels	7
3.1	Integral operators and kernel embeddings of measures	8
3.2	Approximation and kernel quadrature	10
3.3	Nonnegative measures and L^2 -embeddings of RKHSs	11

*gauthierb@cardiff.ac.uk

Cardiff University, School of Mathematics

Abacws, Senghennydd Road, Cardiff, CF24 4AG, United Kingdom

4	Projections and subspaces defined by measures	13
4.1	Further notations and general properties	13
4.2	Approximate measures and projections	15
4.3	Partial L^2 -embeddings	16
4.4	Column sampling for PSD-matrix approximation	20
5	Concluding discussion	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.

Among other interpretations and 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]. Let \mathcal{G} be the RKHS for which the squared-modulus kernel $|K|^2$ is reproducing. In this note, we show that 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 $\bar{\mu}$, the conjugate of μ . The operator L_μ is in this case trace-class, and g_μ is the kernel embedding, or potential, of the measure μ in the RKHS \mathcal{G} . Consequently, in the Hilbert space $\text{HS}(\mathcal{H})$ of all HS operators on \mathcal{H} , the characterisation of an approximate measure ν (defining together with K an integral operator L_ν on \mathcal{H}) for the approximation of an initial operator L_μ reduces to a quadrature problem for the approximation of integral functionals on the RKHS \mathcal{G} associated with $|K|^2$. More precisely, denoting by $B_{\mathcal{G}}$ the closed unit ball of \mathcal{G} , we have

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

so that the norm $\|L_\mu - L_\nu\|_{\text{HS}(\mathcal{H})}$ corresponds to a generalised *integral probability metric* (see for instance [16, 24, 2, 15, 27]) for the approximation of integral functionals on \mathcal{G} .

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 $\text{HS}(\mathcal{H})$ onto \mathcal{G} ; in particular, Γ maps self-adjoint operators to real-valued functions, and PSD operators to nonnegative functions (Section 2). Under general measurability conditions on K and assuming that the diagonal of K is integrable with respect to $|\mu|$, we show that the integral operator L_μ always belongs to the initial space of Γ , and that $\Gamma[L_\mu] = g_\mu$. We next describe the equivalence between the quadrature approximation of trace-class integral operators with PSD kernels and kernel quadrature in RKHSs with squared-modulus kernels (Section 3).

For an approximate measure ν , we denote by P_ν the orthogonal projection from \mathcal{H} onto \mathcal{H}_ν , the closure in \mathcal{H} of the range of $L_{|\nu|}$. We then discuss the extent to which the kernel-quadrature setting of Section 3 can be used as a surrogate for the characterisation of projection-based approximations of L_μ of the form $P_\nu L_\mu$, $L_\mu P_\nu$ or $P_\nu L_\mu P_\nu$. Further, when the initial measure μ is nonnegative, the integral operator defined by the kernel K and the measure μ can more generally be regarded as a HS operator from, and to, $L^2(\mu)$ or \mathcal{H} (four possibilities). Noticing that these integral operators are composites of the natural embedding ι_μ of \mathcal{H} in $L^2(\mu)$, we study the properties of the low-rank approximations induced in this framework by *partial embeddings* of the form $\iota_\mu P_\nu$. We in parallel illustrate the interest of the kernel-quadrature setting in terms of differentiability and numerical complexity, and describe the connections between of the considered framework and the low-rank approximation of PSD matrices through column sampling (Section 4).

2 Framework and notations

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 \bar{z} , $|z|$ and $\Re(z)$ the conjugate, modulus and real part of z , respectively, and $i \in \mathbb{C}$ is the imaginary unit. By analogy, for a complex-valued function f on a general set S , we denote by \bar{f} and $|f|$ the functions defined as $\bar{f}(s) = \overline{f(s)}$ 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 $A : H \rightarrow F$. The map A is an *isometry* if $A^*A = \text{id}_H$, the identity operator on H , and A is *coisometry* if A^* is an isometry (i.e. if $AA^* = \text{id}_F$). A coisometry A is a surjective *partial isometry* (i.e. $AA^*A = A$), 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 $\text{null}(A)$ the nullspace of A , and by $\text{range}(A)$ 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 Hilbert-Schmidt operators and RKHSs

Hereafter, we introduce the overall framework leading to the definition of a natural coisometry Γ from the Hilbert space $\text{HS}(\mathcal{H})$ of all HS operators on an underlying 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.

Underlying RKHS. Let \mathcal{H} be a separable RKHS of complex-valued functions on a general set \mathcal{X} , with reproducing kernel $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{C}$; see for instance

[1, 17]. For $t \in \mathcal{X}$, let $k_t \in \mathcal{H}$ be defined as $k_t(x) = K(x, t)$, with $x \in \mathcal{X}$. 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 similar notations for the inner products and norms of all the Hilbert spaces encountered in this note.

Hilbert-Schmidt space. Let $\text{HS}(\mathcal{H})$ be the Hilbert space of all HS operators on \mathcal{H} ; see e.g. [7, 11, 3]. For $T \in \text{HS}(\mathcal{H})$, 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 $x \in \mathcal{X}$; we use a similar convention for all function-valued operators. For a and $b \in \mathcal{H}$, let $T_{a,b} \in \text{HS}(\mathcal{H})$ 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 $T \in \text{HS}(\mathcal{H})$ admits a singular value decomposition of the form $T = \sum_{i \in \mathbb{N}} \sigma_i T_{u_i, v_i}$, $\mathbb{N} \subseteq \mathbb{N}$, where $\{\sigma_i\}_{i \in \mathbb{N}} \in \ell^2(\mathbb{N})$ is the set of all strictly positive singular values of T , and where $\{u_i\}_{i \in \mathbb{N}}$ and $\{v_i\}_{i \in \mathbb{N}}$ are two orthonormal systems (ONSs) in \mathcal{H} ; the series converges in $\text{HS}(\mathcal{H})$. \triangleleft

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 from \mathcal{H} to \mathcal{H}' (we may notice that $\xi_{\alpha h} = \bar{\alpha} \xi_h$, $\alpha \in \mathbb{C}$). 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 $\text{HS}(\mathcal{H})$ to the tensor Hilbert space $\mathcal{H} \otimes \mathcal{H}'$. \triangleleft

Conjugate RKHS. Let $\overline{\mathcal{H}}$ be the RKHS of complex-valued functions on \mathcal{X} associated with the conjugate kernel \overline{K} , with $\overline{K}(x, t) = \overline{K(x, t)}$, x and $t \in \mathcal{X}$. For all $h \in \mathcal{H}$, we have $\overline{h} \in \overline{\mathcal{H}}$ (i.e. 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 may notice that $\overline{k_t}(x) = \overline{K(x, t)} = \overline{k_t(x)}$, and that

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

We denote by Ψ the bijective linear isometry from $\text{HS}(\mathcal{H})$ 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 from \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 from $\mathcal{H} \otimes \mathcal{H}'$ to $\mathcal{H} \otimes \overline{\mathcal{H}}$; the composition of this isometry with the bijective isometry from $\text{HS}(\mathcal{H})$ to $\mathcal{H} \otimes \mathcal{H}'$ discussed in Remark 2.2 yields the isometry $\Psi : \text{HS}(\mathcal{H}) \rightarrow \mathcal{H} \otimes \overline{\mathcal{H}}$; see the diagram (5). \triangleleft

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

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

Let \mathcal{G} be the RKHS of complex-valued functions on \mathcal{X} for which $|K|^2$ is reproducing (\mathcal{G} 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 $C_\Delta : \mathcal{H} \otimes \overline{\mathcal{H}} \rightarrow \mathcal{G}$ the coisometry densely defined as $C_\Delta(a \otimes \overline{b}) = \overline{ab}$, a and $b \in \mathcal{H}$, where $\overline{ab} \in \mathcal{G}$ is the complex-valued function on \mathcal{X} given by

$$(\overline{ab})(x) = a(x)\overline{b(x)} = \langle k_x \otimes \overline{k_x} \mid a \otimes \overline{b} \rangle_{\mathcal{H} \otimes \overline{\mathcal{H}}}, \quad x \in \mathcal{X}.$$

For $Y \in \mathcal{H} \otimes \mathcal{H}'$, we more generally have

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

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

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

Natural coisometry from $\text{HS}(\mathcal{H})$ onto \mathcal{G} . We are now in the position to introduce the coisometry $\Gamma = C_\Delta \Psi : \text{HS}(\mathcal{H}) \rightarrow \mathcal{G}$; since this map plays a central role in our study, we gather some of its properties in the following Lemma 2.1.

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

$$\mathcal{I}(\Gamma) = \overline{\text{span}_{\mathbb{C}}\{S_{k_x} \mid x \in \mathcal{X}\}}^{\text{HS}(\mathcal{H})}. \quad (2)$$

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

$$\Gamma[T](x) = \langle S_{k_x} \mid T \rangle_{\text{HS}(\mathcal{H})} = \langle k_x \mid T[k_x] \rangle_{\mathcal{H}} = T[k_x](x), \quad x \in \mathcal{X}. \quad (3)$$

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

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

By definition of C_Δ and Ψ , we have $\Gamma[T_{a,b}] = C_\Delta[a \otimes \overline{b}] = \overline{ab}$, a and $b \in \mathcal{H}$. Noticing that $\Gamma^*[|k_x|^2] = \Psi^*[k_x \otimes \overline{k_x}] = S_{k_x}$, $x \in \mathcal{X}$, (2) follows by density of $\text{span}_{\mathbb{C}}\{|k_x|^2 \mid x \in \mathcal{X}\}$ in \mathcal{G} (see Remark 2.4). From the reproducing property in \mathcal{G} ,

we have $\Gamma[T](x) = \langle |k_x|^2 | \Gamma[T] \rangle_{\mathcal{G}}$, $T \in \text{HS}(\mathcal{H})$; as $\Gamma^*[|k_x|^2] = \mathcal{S}_{k_x}$, we obtain $\Gamma[T](x) = \langle \mathcal{S}_{k_x} | T \rangle_{\text{HS}(\mathcal{H})}$. We next notice that

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

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

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

$$\begin{array}{ccc} \text{HS}(\mathcal{H}) & \xrightarrow{\Gamma} & \mathcal{G} \\ \cong \downarrow & \searrow \Psi & \nearrow C_{\Delta} \\ \mathcal{H} \otimes \mathcal{H}' & \xrightarrow{\cong} & \mathcal{H} \otimes \overline{\mathcal{H}} \end{array} \quad (5)$$

Through Γ , the HS operators on \mathcal{H} belonging to $\mathcal{I}(\Gamma)$ are isometrically represented as functions in the RKHS \mathcal{G} associated with the squared-modulus kernel $|K|^2$.

2.2 Self-adjoint operators and real-valuedness

For $T \in \text{HS}(\mathcal{H})$, we have $T^* \in \text{HS}(\mathcal{H})$, and so, from (3), $\Gamma[T^*] = \overline{\Gamma[T]}$. Consequently, if T is self-adjoint, then the function $\Gamma[T]$ is real-valued. This observation is intrinsically related to the structures of $\text{HS}(\mathcal{H})$ and \mathcal{G} , as illustrated below.

Remark 2.5. Form the surjectivity of Γ , for all $g \in \mathcal{G}$, there exists $T \in \text{HS}(\mathcal{H})$ such that $\Gamma[T] = g$, and the function $\bar{g} = \Gamma[T^*]$ is thus also a vector of \mathcal{G} . \triangleleft

Squared-kernel real RKHS. The squared kernel $|K|^2$ being real-valued, it is the reproducing kernel of a real RKHS $\mathcal{G}_{\mathbb{R}}$ of real-valued functions on \mathcal{X} . The real RKHS $\mathcal{G}_{\mathbb{R}}$ can be regarded as a closed real-linear subspace of \mathcal{G} , and the complex RKHS \mathcal{G} is a *complexification* of $\mathcal{G}_{\mathbb{R}}$ (see e.g. [17, Chapter 5]; $\mathcal{G}_{\mathbb{R}}$ is referred to as a *standard subspace* of \mathcal{G}). More precisely, for all $g \in \mathcal{G}$, there exist g_r and $g_i \in \mathcal{G}_{\mathbb{R}} \subset \mathcal{G}$ such that $g = g_r + ig_i$, with $g_r = \frac{1}{2}(\bar{g} + g)$ and $g_i = \frac{1}{2i}(\bar{g} - g)$, and the decomposition $\mathcal{G} = \mathcal{G}_{\mathbb{R}} + i\mathcal{G}_{\mathbb{R}}$ holds. The functions g_r and g_i are the real and imaginary parts of g , and we have $\|g\|_{\mathcal{G}}^2 = \|g_r\|_{\mathcal{G}}^2 + \|g_i\|_{\mathcal{G}}^2$.

Self-adjoint HS operators. We denote by $\text{HS}_*(\mathcal{H}) \subset \text{HS}(\mathcal{H})$ the closed real-linear subspace of all self-adjoint HS operators on \mathcal{H} . For every $T \in \text{HS}(\mathcal{H})$, there exist T_r and $T_i \in \text{HS}_*(\mathcal{H})$ such that $T = T_r + iT_i$, with $T_r = \frac{1}{2}(T^* + T)$ and $T_i = \frac{1}{2i}(T^* - T)$. Endowed with the inner product of $\text{HS}(\mathcal{H})$, the real vector space $\text{HS}_*(\mathcal{H})$ is a real Hilbert space, and the decomposition $\text{HS}(\mathcal{H}) = \text{HS}_*(\mathcal{H}) + i\text{HS}_*(\mathcal{H})$ holds. By linearity of Γ , we may notice that if $\Gamma[T] = g$, then $\Gamma[T_r] = g_r$ and $\Gamma[T_i] = g_i$.

Remark 2.6. For all $T \in \text{HS}(\mathcal{H})$, we have $\|T\|_{\text{HS}(\mathcal{H})} \geq \|\Gamma[T]\|_{\mathcal{G}}$, with equality if and only if $T \in \mathcal{I}(\Gamma)$; as $\|T\|_{\text{HS}(\mathcal{H})} = \|T^*\|_{\text{HS}(\mathcal{H})}$ and $\|g\|_{\mathcal{G}} = \|\bar{g}\|_{\mathcal{G}}$, it follows that if $T \in \mathcal{I}(\Gamma)$, then $T^* \in \mathcal{I}(\Gamma)$, and also T_r and $T_i \in \mathcal{I}(\Gamma)$. \triangleleft

Remark 2.7. The diagram (5) is also well-defined when all the involved Hilbert spaces are real (the kernel K in this case verifies $K = \bar{K}$, that is, K is symmetric). We then have $\mathcal{I}(\Gamma) = \overline{\text{span}_{\mathbb{R}}\{S_{k_x} \mid x \in \mathcal{X}\}}^{\text{HS}(\mathcal{H})}$, and $\mathcal{I}(\Gamma)$ thus solely consists of self-adjoint operators. In the real case, as $\Gamma[T^*] = \Gamma[T]$, $T \in \text{HS}(\mathcal{H})$, we also obtain that if $T^* = -T$, then $\Gamma[T] = 0$; by comparison, in the complex case, if $T^* = -T$, then the function $\Gamma[T]$ is pure-imaginary. \triangleleft

2.3 Positive-semidefinite operators and nonnegativity

From (3), if $T \in \text{HS}(\mathcal{H})$ is PSD, then $\Gamma[T](x) = \langle k_x \mid T[k_x] \rangle_{\mathcal{H}} \geq 0$ for all $x \in \mathcal{X}$, so that the function $\Gamma[T]$ is nonnegative. This observation again directly relates to the structures of $\text{HS}(\mathcal{H})$ and \mathcal{G} .

Positive-semidefinite HS operators. We denote by $\text{HS}_*^+(\mathcal{H}) \subset \text{HS}_*(\mathcal{H})$ the closed real convex cone of all PSD HS operators on \mathcal{H} . For all $T \in \text{HS}_*(\mathcal{H})$, there exist T^+ and $T^- \in \text{HS}_*^+(\mathcal{H})$ such that $T = T^+ - T^-$, so that the cone $\text{HS}_*^+(\mathcal{H})$ is generating in $\text{HS}_*(\mathcal{H})$.

Nonnegative functions in \mathcal{G} . We denote by $\mathcal{G}_{\mathbb{R}}^+ \subset \mathcal{G}_{\mathbb{R}}$ the closed real convex cone of all nonnegative functions in $\mathcal{G}_{\mathbb{R}}$. For all $T \in \text{HS}_*^+(\mathcal{H})$, we have $\Gamma[T] \in \mathcal{G}_{\mathbb{R}}^+$, and since Γ is surjective, the cone $\mathcal{G}_{\mathbb{R}}^+$ is generating (that is, for all $g \in \mathcal{G}_{\mathbb{R}}$, there exist g^+ and $g^- \in \mathcal{G}_{\mathbb{R}}^+$ such that $g = g^+ - g^-$).

Remark 2.8. Let $T = \sum_{j \in \mathbb{I}} \lambda_j S_{\varphi_j}$ be a spectral expansion of $T \in \text{HS}_*^+(\mathcal{H})$, with $\{\lambda_j\}_{j \in \mathbb{I}}, \mathbb{I} \subseteq \mathbb{N}$, the set of all strictly-positive eigenvalues of T (repeated with multiplicity), and where $\{\varphi_j\}_{j \in \mathbb{I}} \subset \mathcal{H}$ is a set of associated eigenvectors, orthonormal in \mathcal{H} . We then have $\Gamma[T] = \sum_{j \in \mathbb{I}} \lambda_j \Gamma[S_{\varphi_j}] = \sum_{j \in \mathbb{I}} \lambda_j |\varphi_j|^2$, so that if $T \in \text{HS}(\mathcal{H})$ is PSD, then $\Gamma[T] = 0$ if and only if $T = 0$. \triangleleft

3 Trace-class integral operators with PSD kernels

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

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

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 $x \in \mathcal{X}$.

We in addition have $\Gamma[T] = \sum_{j=1}^n \omega_j |k_{s_j}|^2$, so that $\langle \Gamma[T] | g \rangle_{\mathcal{G}} = \sum_{j=1}^n \overline{\omega_j} g(s_j)$, $g \in \mathcal{G}$, 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 adequate measurability conditions, these observations more generally hold for all trace-class integral operators on \mathcal{H} defined by the reproducing kernel K of \mathcal{H} and general measures on \mathcal{X} .

3.1 Integral operators and kernel embeddings of measures

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

(A.1) for all $t \in \mathcal{X}$, the function $k_t \in \mathcal{H}$ is measurable on $(\mathcal{X}, \mathcal{A})$;

(A.2) the diagonal of K is measurable on $(\mathcal{X}, \mathcal{A})$.

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]. The three maps $t \mapsto k_t$, $t \mapsto |k_t|^2$ and $t \mapsto S_{k_t}$, $t \in \mathcal{X}$, are then weakly-measurable, and since the Hilbert spaces \mathcal{H} , \mathcal{G} and $\text{HS}(\mathcal{H})$ are separable, by the Pettis measurability theorem, these three maps are also strongly-measurable (see e.g. [8, 28]). \triangleleft

We denote by \mathcal{M}_+ , \mathcal{M} , and $\mathcal{M}_{\mathbb{C}}$ the set of all nonnegative, signed and complex measures¹ on $(\mathcal{X}, \mathcal{A})$, and we set $\mathcal{M}_{\mathbb{F}} = \mathcal{M} \cup \mathcal{M}_{\mathbb{C}}$ (we have $\mathcal{M}_+ \subset \mathcal{M}$). Noticing that $K(t, t) \geq 0$, $t \in \mathcal{X}$, from (A.2), we define

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

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 $\mathcal{T}_{\mathbb{F}}(K) = \mathcal{T}(K) \cup \mathcal{T}_{\mathbb{C}}(K)$. We may notice that $\mathcal{T}_{\mathbb{C}}(K)$ is a vector space, and that $\mathcal{T}_+(K)$ is a real convex cone. We recall that

$$K(t, t) = \|k_t\|_{\mathcal{H}}^2 = \|S_{k_t}\|_{\text{HS}(\mathcal{H})} = \||k_t|^2\|_{\mathcal{G}}, t \in \mathcal{X}.$$

Integral operators on \mathcal{H} with kernel K . By assumption, for $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, the integral $\int_{\mathcal{X}} \|S_{k_t}\|_{\text{HS}(\mathcal{H})} d|\mu|(t) = \tau_{\mu}$ 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]). We then set

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

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

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

and $L_{\mu} \in \text{HS}(\mathcal{H})$ can thus be regarded as an integral operator on \mathcal{H} defined by the kernel K and the measure μ (see also Remark 3.2).

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

Lemma 3.1. For all $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, we have

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

Proof. From the Cauchy-Schwarz (CS) inequality in \mathcal{H} , we have

$$\left| \int_{\mathcal{X}} \overline{h(t)} f(t) d\overline{\mu}(t) \right| \leq \int_{\mathcal{X}} |\langle k_t | h \rangle_{\mathcal{H}}| |\langle k_t | f \rangle_{\mathcal{H}}| d|\mu|(t) \leq \|f\|_{\mathcal{H}} \|h\|_{\mathcal{H}} \tau_{\mu},$$

so that the linear functional $\Theta_{h,\mu} : f \mapsto \int_{\mathcal{X}} \overline{h(t)} f(t) d\overline{\mu}(t)$, $f \in \mathcal{H}$, is bounded. Taking $f = k_x$, $x \in \mathcal{X}$, we obtain

$$\Theta_{h,\mu}(k_x) = \int_{\mathcal{X}} \overline{h(t)} K(t, x) d\overline{\mu}(t) = \overline{L_{\mu}[h](x)} = \langle L_{\mu}[h] | k_x \rangle_{\mathcal{H}},$$

so that $L_{\mu}[h]$ is the Riesz representation of $\Theta_{h,\mu}$. \square

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

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

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

We may also notice that the map $t \mapsto k_t$ is Bochner-integrable with respect to the measure $\mathbf{m}_{h,\mu}$ on $(\mathcal{X}, \mathcal{A})$ defined as $\mathbf{m}_{h,\mu}(A) = \int_A h(t) d\mu(t)$, $A \in \mathcal{A}$; we then have $L_{\mu}[h] = \int_{\mathcal{X}} k_t h(t) d\mu(t)$, with $\int_{\mathcal{X}} \|k_t\|_{\mathcal{H}} |h(t)| d|\mu|(t) \leq \|h\|_{\mathcal{H}} \tau_{\mu}$. \triangleleft

Lemma 3.2. 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 $t \in \mathcal{X}$, 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 have

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

so that $\text{trace}(|L_{\mu}|) \leq \tau_{\mu}$, with $|L_{\mu}| = (L_{\mu}^* L_{\mu})^{1/2}$ the modulus of L_{μ} . \square

Kernel embedding of measures in \mathcal{G} . As $\int_{\mathcal{X}} \| |k_t|^2 \|_{\mathcal{G}} d|\mu|(t) = \tau_{\mu}$, for $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, the map $t \mapsto |k_t|^2$ is Bochner-integrable with respect to μ . We then set

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

For all $g \in \mathcal{G}$, we have $\langle g_{\mu} | g \rangle_{\mathcal{G}} = \int_{\mathcal{X}} g(t) d\bar{\mu}(t)$, and g_{μ} is thus the Riesz representation of the linear functional $I_{\mu} : \mathcal{G} \rightarrow \mathbb{C}$ such that $I_{\mu}(g) = \int_{\mathcal{X}} g(t) d\bar{\mu}(t)$; we may notice that $|I_{\mu}(g)| \leq \|g\|_{\mathcal{G}} \tau_{\mu}$. 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 $T \in \text{HS}(\mathcal{H})$, we have

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

so that $\Gamma^*[g_{\mu}] = L_{\mu}$. □

Remark 3.3. Following Section 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; as expected. Also, following Section 2.3, 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 $L_{\delta_x} = S_{k_x}$, $x \in \mathcal{X}$. \triangleleft

3.2 Approximation and kernel quadrature

Let $B_{\mathcal{G}} = \{g \in \mathcal{G} | \|g\|_{\mathcal{G}} \leq 1\}$ be the closed unit ball of \mathcal{G} . For μ and $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, from the definition of g_{μ} and $g_{\nu} \in \mathcal{G}$ and by the CS inequality, we have

$$\begin{aligned} \mathfrak{M}_{\mathcal{G}}(\mu, \nu) &= \sup_{g \in B_{\mathcal{G}}} |\langle g_{\mu} - g_{\nu} | g \rangle_{\mathcal{G}}| = \sup_{g \in B_{\mathcal{G}}} \left| \int_{\mathcal{X}} g(t) d\bar{\mu}(t) - \int_{\mathcal{X}} g(t) d\bar{\nu}(t) \right| \\ &= \|g_{\mu} - g_{\nu}\|_{\mathcal{G}}. \end{aligned}$$

When the measures μ and ν are probability measures, the term $\mathfrak{M}_{\mathcal{G}}(\mu, \nu)$ is referred to as an *integral probability metric*, or *maximum mean discrepancy*; see for instance [16, 24, 23, 15, 27].

Remark 3.4. For all $g \in \mathcal{G}$, we have $\bar{g} \in \mathcal{G}$ and $\|\bar{g}\|_{\mathcal{G}} = \|g\|_{\mathcal{G}}$ (see Section 2.2), so that $\mathfrak{M}_{\mathcal{G}}(\mu, \nu) = \sup_{g \in B_{\mathcal{G}}} \left| \int_{\mathcal{X}} g(t) d\mu(t) - \int_{\mathcal{X}} g(t) d\nu(t) \right|$, μ and $\nu \in \mathcal{T}_{\mathbb{F}}(K)$. \triangleleft

The following Corollary 3.1 illustrates the connection between the quadrature approximation of trace-class integral operators with PSD kernels (*quadrature method*; see e.g. [13]) and the approximation of integral functionals on RKHSs with squared-modulus kernels (*kernel quadrature*; see e.g. [2]).

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

Proof. From 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 $\|g_\mu - g_\nu\|_{\mathcal{G}} = \|L_\mu - L_\nu\|_{\text{HS}(\mathcal{H})}$. \square

For a fixed initial measure $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, the map $D_\mu : \mathcal{T}_{\mathbb{F}}(K) \rightarrow \mathbb{R}_{\geq 0}$, with

$$D_\mu(\nu) = \|L_\mu - L_\nu\|_{\text{HS}(\mathcal{H})}^2 = \|g_\mu - g_\nu\|_{\mathcal{G}}^2, \quad \nu \in \mathcal{T}_{\mathbb{F}}(K), \quad (7)$$

is convex on any convex set $\mathcal{C} \subseteq \mathcal{T}_{\mathbb{F}}(K)$; the map $\nu \mapsto g_\nu$ is indeed linear on \mathcal{C} (that is, for $\xi = \nu + \alpha\eta \in \mathcal{C}$, with $\nu, \eta \in \mathcal{C}$ and $\alpha \in \mathbb{C}$, we have $g_\xi = g_\nu + \alpha g_\eta$), and the map $g \mapsto \|g_\mu - g\|_{\mathcal{G}}^2$ is convex on \mathcal{G} . Also, for ν and $\eta \in \mathcal{C}$, the directional derivative of D_μ at ν in the direction $\eta - \nu$ is

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

From Theorem 3.1, we may in particular notice that

$$\begin{aligned} \langle g_\nu \mid g_\mu \rangle_{\mathcal{G}} &= \langle L_\nu \mid L_\mu \rangle_{\text{HS}(\mathcal{H})} = \iint_{\mathcal{X}} |K(x, t)|^2 d\mu(t) d\bar{\nu}(x) \\ &= \int_{\mathcal{X}} g_\mu(x) d\bar{\nu}(x) = \int_{\mathcal{X}} \overline{g_\nu(t)} d\mu(t); \end{aligned} \quad (8)$$

and since $|K(x, t)|^2 = |\langle k_x \mid k_t \rangle_{\mathcal{H}}|^2 \leq \|k_x\|_{\mathcal{H}}^2 \|k_t\|_{\mathcal{H}}^2 = K(x, x)K(t, t)$, x and $t \in \mathcal{X}$ (from the CS inequality in \mathcal{H}), we have $|\langle g_\nu \mid g_\mu \rangle_{\mathcal{G}}| \leq \tau_\nu \tau_\mu$.

Remark 3.5. For $\mu \in \mathcal{T}(K)$ and $\nu \in \mathcal{T}_{\mathbb{C}}(K)$, and denoting by ν_r and $\nu_i \in \mathcal{T}(K)$ the real and imaginary parts of ν , we have $g_\mu - g_{\nu_r} \in \mathcal{G}_{\mathbb{R}}$ and $g_{\nu_i} \in \mathcal{G}_{\mathbb{R}}$. Following Section 2.2, we obtain $\|g_\mu - g_\nu\|_{\mathcal{G}}^2 = \|g_\mu - g_{\nu_r}\|_{\mathcal{G}}^2 + \|g_{\nu_i}\|_{\mathcal{G}}^2$, so that $D_\mu(\nu_r) \leq D_\mu(\nu)$. Consequently, when L_μ is self-adjoint, the search of an approximate measure ν for the approximation of L_μ by L_ν may be restricted to $\mathcal{T}(K)$. \triangleleft

3.3 Nonnegative measures and L^2 -embeddings of RKHSs

Hereafter and in anticipation of the forthcoming Section 4, we further discuss the 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

$$\int_{\mathcal{X}} |h(t)|^2 d\mu(t) = \int_{\mathcal{X}} |\langle k_t \mid h \rangle_{\mathcal{H}}|^2 d\mu(t) \leq \|h\|_{\mathcal{H}}^2 \tau_\mu, \quad h \in \mathcal{H}, \quad (9)$$

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

Lemma 3.3. *For all $\mu \in \mathcal{T}_+(K)$, the embedding ι_μ is HS and $L_\mu = \iota_\mu^* \iota_\mu$; we in addition have $\text{null}(L_\mu) = \{h \in \mathcal{H} \mid \iota_\mu[h] = 0\}$.*

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

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

so that ι_μ is HS. From (6), we also obtain

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

and so $L_\mu = \iota_\mu^* \iota_\mu$. Finally, noticing that $\langle L_\mu[h] | h \rangle_{\mathcal{H}} = \|\iota_\mu[h]\|_{L^2(\mu)}^2$ and that L_μ is PSD, we have $L_\mu[h] = 0$ if and only if $\iota_\mu[h] = 0$. \square

For $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we by definition have $|\nu| \in \mathcal{T}_+(K)$ and $\bar{\nu} \in \mathcal{T}_{\mathbb{F}}(K)$; from (6), we can in addition notice that

$$\langle L_\nu[h] | f \rangle_{\mathcal{H}} = \int_{\mathcal{X}} \overline{h(t)} f(t) d\bar{\nu}(t) = \langle h | L_{\bar{\nu}}[f] \rangle_{\mathcal{H}}, \quad h \text{ and } f \in \mathcal{H},$$

so that $L_\nu^* = L_{\bar{\nu}}$. The following relation (Lemma 3.4) holds between the range of L_ν and the range of $L_{|\nu|}$.

Lemma 3.4. *For all $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we have $\overline{\text{range}(L_\nu)}^{\mathcal{H}} \subseteq \overline{\text{range}(L_{|\nu|})}^{\mathcal{H}}$.*

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

$$\begin{aligned} |\langle L_{\bar{\nu}}[h] | f \rangle_{\mathcal{H}}| &= \left| \int_{\mathcal{X}} \overline{h(t)} f(t) d\nu(t) \right| \leq \int_{\mathcal{X}} |h(t)| |f(t)| d|\nu|(t) \\ &\leq \|\iota_{|\nu|}[h]\|_{L^2(|\nu|)} \|\iota_{|\nu|}[f]\|_{L^2(|\nu|)}, \quad h \text{ and } f \in \mathcal{H}. \end{aligned}$$

Lemma 3.3 then entails $\text{null}(L_{|\nu|}) \subseteq \text{null}(L_{\bar{\nu}})$, and so $\text{null}(L_{\bar{\nu}})^{\perp \mathcal{H}} \subseteq \text{null}(L_{|\nu|})^{\perp \mathcal{H}}$. Since for all $T \in \text{HS}(\mathcal{H})$, we have $\text{null}(T^*) = \text{range}(T)^{\perp \mathcal{H}}$, we conclude by noticing that $L_{\bar{\nu}}^* = L_\nu$ and $L_{|\nu|}^* = L_{|\nu|}$. \square

The following Lemma 3.5 illustrates that when ν is finitely-supported, then the range of $L_{|\nu|}$ is fully characterised by the support² of ν .

Lemma 3.5. *For $\nu = \sum_{i=1}^n v_i \delta_{s_i}$, with $v_i \in \mathbb{C}$, $v_i \neq 0$, and $s_i \in \mathcal{X}$, we have $\text{range}(L_{|\nu|}) = \text{span}_{\mathbb{C}}\{k_{s_1}, \dots, k_{s_n}\}$.*

Proof. Noticing that $|\nu| = \sum_{i=1}^n |v_i| \delta_{s_i} \in \mathcal{T}_+(K)$, from Lemma 3.3, we obtain $\text{null}(L_{|\nu|}) = \{h \in \mathcal{H} | \iota_{|\nu|}[h] = 0\} = \bigcap_{i=1}^n \{h \in \mathcal{H} | \langle k_{s_i} | h \rangle_{\mathcal{H}} = 0\}$. Since $L_{|\nu|}$ is self-adjoint, the result follows. \square

²Assuming that \mathcal{X} is a topological space and that the RKHS \mathcal{H} consists of continuous functions, a similar characterisation could be obtained for Radon measures, for instance. In this note, we merely focus on the finitely-supported case as it does not require any further assumption and is one of the most relevant in terms of practical applications.

4 Projections and subspaces defined by measures

Hereafter, we discuss the extent to which the kernel-quadrature setting described in Section 3.2 can be used as a surrogate for the characterisation of closed linear subspaces of \mathcal{H} for the approximation of trace-class integral operators with PSD kernels through projections.

4.1 Further 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 $t \in \mathcal{X}$.

Remark 4.1. The linear map $T \mapsto P_S T$ is the orthogonal projection from $\text{HS}(\mathcal{H})$ onto $\mathcal{R}(\mathcal{H}_S) = \{T \in \text{HS}(\mathcal{H}) \mid \text{range}(T) \subseteq \mathcal{H}_S\}$, the closed linear subspace of $\text{HS}(\mathcal{H})$ of all operators with range included in \mathcal{H}_S .

Further, the linear map $T \mapsto T P_S$ is the orthogonal projection from $\text{HS}(\mathcal{H})$ onto $\mathcal{Z}(\mathcal{H}_S) = \{T \in \text{HS}(\mathcal{H}) \mid \text{range}(T^*) \subseteq \mathcal{H}_S\}$. The two orthogonal projections $T \mapsto P_S T$ and $T \mapsto T P_S$ commute, and their composition $T \mapsto P_S T P_S$ is the orthogonal projection from $\text{HS}(\mathcal{H})$ onto $\mathcal{R}(\mathcal{H}_S) \cap \mathcal{Z}(\mathcal{H}_S)$. For all $T \in \text{HS}(\mathcal{H})$, we thus have

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

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

with $\|P_S T P_S\|_{\text{HS}(\mathcal{H})} \leq \|P_S T\|_{\text{HS}(\mathcal{H})} \leq \|T\|_{\text{HS}(\mathcal{H})}$; in particular, if T is self-adjoint, then so is $P_S T P_S$. Since $(P_S T)^* = T^* P_S$, 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 $P_S T P_S$. \triangleleft

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 $T \in \text{HS}(\mathcal{H})$, we have*

$$\|P_R T\|_{\text{HS}(\mathcal{H})} \leq \|P_S T\|_{\text{HS}(\mathcal{H})} \quad \text{and} \quad \|P_R T P_R\|_{\text{HS}(\mathcal{H})} \leq \|P_S T P_S\|_{\text{HS}(\mathcal{H})}.$$

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 $\langle P_R T \mid P_e T \rangle_{\text{HS}(\mathcal{H})} = \langle T P_R \mid T P_e \rangle_{\text{HS}(\mathcal{H})} = 0$, $T \in \text{HS}(\mathcal{H})$. It follows that $\|P_S T\|_{\text{HS}(\mathcal{H})}^2 = \|P_R T\|_{\text{HS}(\mathcal{H})}^2 + \|P_e T\|_{\text{HS}(\mathcal{H})}^2$, and that

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

completing the proof. \square

Remark 4.2. For a fixed initial operator $T \in \text{HS}(\mathcal{H})$ and a given rank, the subspaces leading to minimal values of $\|T - P_S T\|_{\text{HS}(\mathcal{H})}$ or $\|T - P_S T P_S\|_{\text{HS}(\mathcal{H})}$ correspond to spectral truncations of T ; see e.g. [12, Theorem 4.4.7]. In practical applications, a direct implementation of such optimal approximations is therefore limited to operators for which a singular value decomposition can be computed beforehand. \triangleleft

Orthogonal projections being bounded linear maps, for $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, we have $P_S L_\mu = \int_{\mathcal{X}} P_S S_{k_t} d\mu(t)$ (see e.g. [28]), that is

$$P_S L_\mu[h](x) = \int_{\mathcal{X}} K_S(x, t) h(t) d\mu(t), \quad h \in \mathcal{H} \text{ and } x \in \mathcal{X},$$

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

Lemma 4.2. 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

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

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

$$\begin{aligned} \|P_V L_\mu P_U\|_{\text{HS}(\mathcal{H})}^2 &= \sum_{j \in \mathbb{I}} \langle L_\mu P_U[h_j] | P_V L_\mu P_U[h_j] \rangle_{\mathcal{H}} \\ &= \sum_{j \in \mathbb{I}} \iint_{\mathcal{X}} P_U[h_j](t) \overline{P_U[h_j](x)} K_V(x, t) d\mu(t) d\bar{\mu}(x). \end{aligned} \quad (12)$$

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

$$\iint_{\mathcal{X}} \sum_{j \in \mathbb{I}} |P_U[h_j](t)| |P_U[h_j](x)| |K_V(x, t)| d|\mu|(t) d|\mu|(x) \leq \tau_\mu^2,$$

and the result follows from (12) and Fubini's theorem. \square

Remark 4.3. For $\mu \in \mathcal{T}_{\mathbb{F}}(K)$ and $x \in \mathcal{X}$, we have

$$\begin{aligned} \Gamma[P_S L_\mu](x) &= \langle k_x | P_S L_\mu[k_x] \rangle_{\mathcal{H}} = \int_{\mathcal{X}} K_S(x, t) K(t, x) d\mu(t), \text{ and} \\ \Gamma[P_S L_\mu P_S](x) &= \langle k_x | P_S L_\mu P_S[k_x] \rangle_{\mathcal{H}} = \int_{\mathcal{X}} |K_S(x, t)|^2 d\mu(t); \end{aligned}$$

however, for general linear subspaces $\mathcal{H}_S \subseteq \mathcal{H}$, the operators $P_S L_\mu$ and $P_S L_\mu P_S$ do not necessarily belong to $\mathcal{I}(\Gamma)$. \triangleleft

4.2 Approximate measures and projections

Motivated by Lemmas 3.4 and 4.1, for $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we introduce $\mathcal{H}_{\nu} = \overline{\text{range}(L_{|\nu|})}^{\mathcal{H}}$, and we denote by P_{ν} the orthogonal projection from \mathcal{H} onto \mathcal{H}_{ν} . We refer to \mathcal{H}_{ν} as the *subspace of \mathcal{H} defined by ν* .

Lemma 4.3. *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 3.4, we have $L_{\nu} = P_{\nu}L_{\nu}$. Noticing that $L_{\bar{\nu}} = P_{\nu}L_{\bar{\nu}}$, we also have $L_{\nu} = L_{\bar{\nu}}^* = (P_{\nu}L_{\bar{\nu}})^* = L_{\nu}P_{\nu}$, and thus $L_{\nu} = P_{\nu}L_{\nu}P_{\nu}$. \square

The following Lemma 4.4 illustrates the relation between the approximations L_{ν} , $P_{\nu}L_{\mu}$ and $P_{\nu}L_{\mu}P_{\nu}$ of L_{μ} characterised by an approximate measure ν .

Lemma 4.4. *For all μ and $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, we have*

$$\|L_{\mu} - L_{\nu}\|_{\text{HS}(\mathcal{H})}^2 = \|L_{\mu} - P_{\nu}L_{\mu}\|_{\text{HS}(\mathcal{H})}^2 + \|P_{\nu}L_{\mu} - L_{\nu}\|_{\text{HS}(\mathcal{H})}^2, \text{ and} \quad (13)$$

$$\|L_{\mu} - L_{\nu}\|_{\text{HS}(\mathcal{H})}^2 = \|L_{\mu} - P_{\nu}L_{\mu}P_{\nu}\|_{\text{HS}(\mathcal{H})}^2 + \|P_{\nu}L_{\mu}P_{\nu} - L_{\nu}\|_{\text{HS}(\mathcal{H})}^2. \quad (14)$$

Proof. Using the notations of Remark 4.1, Lemma 4.3 reads $L_{\nu} \in \mathcal{R}(\mathcal{H}_{\nu}) \cap \mathcal{Z}(\mathcal{H}_{\nu})$. Noticing 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 (13). 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 (14). \square

For $\mu \in \mathcal{T}_{\mathbb{F}}(K)$, we introduce the two maps C_{μ}^{P} and C_{μ}^{PP} : $\mathcal{T}_{\mathbb{F}}(K) \rightarrow \mathbb{R}_{\geq 0}$, with

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

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

$$\lim_{\rho \rightarrow 0^+} \frac{1}{\rho} [C_{\mu}^{\text{P}}(\nu + \rho(\eta - \nu)) - C_{\mu}^{\text{P}}(\nu)] \in \{-\infty, 0\}. \quad (15)$$

This statement stays true if we replace C_{μ}^{P} by C_{μ}^{PP} .

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 the value of $\rho \in (0, 1)$, we have $\text{null}(L_{\xi}) = \text{null}(L_{\nu}) \cap \text{null}(L_{\eta})$, and so $\mathcal{H}_{\xi} = \overline{\mathcal{H}_{\nu} + \mathcal{H}_{\eta}}^{\mathcal{H}}$. The two maps

$$\rho \mapsto C_{\mu}^{\text{P}}(\nu + \rho(\eta - \nu)) \quad \text{and} \quad \rho \mapsto C_{\mu}^{\text{PP}}(\nu + \rho(\eta - \nu))$$

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 $C_{\mu}^{\text{P}}(\xi) \leq C_{\mu}^{\text{P}}(\nu)$ and $C_{\mu}^{\text{P}}(\xi) \leq C_{\mu}^{\text{P}}(\eta)$; from (11), we also get $C_{\mu}^{\text{PP}}(\xi) \leq C_{\mu}^{\text{PP}}(\nu)$ and $C_{\mu}^{\text{PP}}(\xi) \leq C_{\mu}^{\text{PP}}(\eta)$, concluding the proof. \square

In view of Theorem 4.1, the maps C_μ^P and C_μ^{PP} are convex and piecewise-constant on the convex cone $\mathcal{T}_+(K)$. In contrast to the map D_μ defined in (7), the directional derivatives of C_μ^P and C_μ^{PP} are *noninformative*, in the sense that they do not provide any information on the local steepness of the landscape of these maps; see Figure 1 for an illustration. From Remark 4.1 and Lemma 4.4, we have

$$C_\mu^P(v) \leq C_\mu^{PP}(v) \leq D_\mu(v), v \in \mathcal{T}_\mathbb{F}(K). \quad (16)$$

Also, from a numerical standpoint and in view of Lemma 4.2, for $v \in \mathcal{T}_\mathbb{F}(K)$, the evaluation of $C_\mu^P(v)$ or $C_\mu^{PP}(v)$ requires a suitable characterisation of the reproducing kernel K_v of \mathcal{H}_v ; in practice, K_v is a priori unknown and needs to be computed from K and v . In comparison and in view of (8), the map D merely involves the kernel K . See Remark 4.7 for more details.

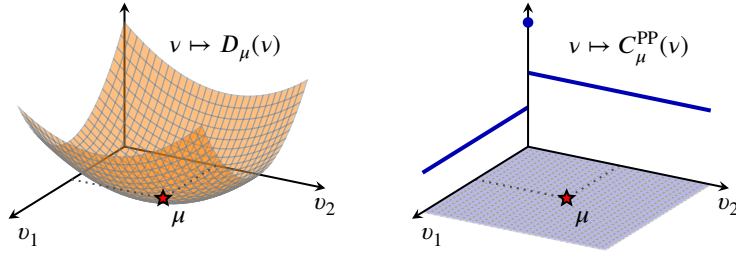


Figure 1: Graphical representation of the maps D_μ and C_μ^{PP} as functions of the weight parameters characterising an approximate measure $v \in \mathcal{T}_+(K)$. The measures μ and v are supported by the same set of points $\{x_1, x_2\} \subseteq \mathcal{X}$, and described by their weight parameters (ω_1, ω_2) and $(v_1, v_2) \in \mathbb{R}_{\geq 0}^2$, 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 $v = 0$, and the bold lines indicate the constant values taken by the map along the horizontal axes.

Remark 4.4. For $\mu \in \mathcal{T}_\mathbb{F}(K)$ and $s \in \mathcal{X}$, introducing $c_{\delta_s} = g_\mu(s)/|K(s, s)|^2$ if $K(s, s) > 0$, and $c_{\delta_s} = 0$ otherwise, a direct computation yields $P_{\delta_s} L_\mu P_{\delta_s} = c_{\delta_s} S_{k_s}$. We thus have $P_{\delta_s} L_\mu P_{\delta_s} \in \mathcal{I}(\Gamma)$ and $C_\mu^{PP}(\delta_s) = D_\mu(c_{\delta_s} \delta_s)$. Consequently, in Figure 1, the graph of C_μ^{PP} is tangent to the graph of D_μ along the horizontal axes. \triangleleft

4.3 Partial L^2 -embeddings

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

$$\langle k_x | \iota_\mu^*[f] \rangle_{\mathcal{H}} = \langle \iota_\mu[k_x] | f \rangle_{L^2(\mu)} = \int_{\mathcal{X}} K(x, t) f(t) d\mu(t).$$

Consequently, in addition to $L_\mu = \iota_\mu^* \iota_\mu \in \text{HS}(\mathcal{H})$, the three operators

$$\iota_\mu^* : L^2(\mu) \rightarrow \mathcal{H}, \quad \iota_\mu \iota_\mu^* : L^2(\mu) \rightarrow L^2(\mu), \quad \text{and} \quad \iota_\mu \iota_\mu^* \iota_\mu : \mathcal{H} \rightarrow L^2(\mu), \quad (17)$$

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 $\text{HS}(\mu, \mathcal{H})$, $\text{HS}(\mu)$ and $\text{HS}(\mathcal{H}, \mu)$ 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.

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

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

For $L_\mu = \iota_\mu^* \iota_\mu \in \text{HS}(\mathcal{H})$, we obtain the approximation $P_S \iota_\mu^* \iota_\mu P_S$ discussed in Section 4.2, and for the three operators defined in (17), we obtain the approximations

$$P_S \iota_\mu^* \in \text{HS}(\mu, \mathcal{H}), \quad \iota_\mu P_S \iota_\mu^* \in \text{HS}(\mu), \quad \text{and} \quad \iota_\mu P_S \iota_\mu^* \iota_\mu P_S \in \text{HS}(\mathcal{H}, \mu).$$

In what follows and for simplicity, we mainly focus on the approximations related to $\text{HS}(\mu, \mathcal{H})$ and $\text{HS}(\mu)$; the case of $\text{HS}(\mathcal{H}, \mu)$ is more briefly discussed in Remark 4.5.

Lemma 4.5. *Let \mathcal{H}_S be a closed linear subspace of \mathcal{H} . For $\mu \in \mathcal{T}_+(K)$, we have*

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

$$\|\iota_\mu \iota_\mu^* - \iota_\mu P_S \iota_\mu^*\|_{\text{HS}(\mu)}^2 = \iint_{\mathcal{X}} |K(x, t) - K_S(x, t)|^2 d\mu(t) d\mu(x). \quad (19)$$

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

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

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

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

so that (19) follows from Lemma 4.2. \square

The following inequality (Lemma 4.6) holds between the approximations in $\text{HS}(\mu)$ and $\text{HS}(\mathcal{H})$ defined by a subspace \mathcal{H}_S . We recall that $\|\iota_\mu \iota_\mu^*\|_{\text{HS}(\mu)} = \|\iota_\mu^* \iota_\mu\|_{\text{HS}(\mathcal{H})}$, and that $\|\iota_\mu^* \iota_\mu - P_S \iota_\mu^* \iota_\mu P_S\|_{\text{HS}(\mathcal{H})} \leq \|\iota_\mu^* \iota_\mu - P_S \iota_\mu^* \iota_\mu P_S\|_{\text{HS}(\mathcal{H})}$; see Remark 4.1.

Lemma 4.6. *Let \mathcal{H}_S be a closed linear subspace of \mathcal{H} . For all $\mu \in \mathcal{T}_+(K)$, we have*

$$\|\iota_\mu^* - \iota_\mu P_S \iota_\mu^*\|_{\text{HS}(\mu)} \leq \|\iota_\mu^* - P_S \iota_\mu^*\|_{\text{HS}(\mathcal{H})}.$$

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

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

so that $\|\iota_\mu^* - \iota_\mu P_S \iota_\mu^*\|_{\text{HS}(\mu)} \leq \|\iota_\mu^* - P_S \iota_\mu^*\|_{\text{HS}(\mathcal{H})}$. \square

Lemma 4.7. *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 $\mu \in \mathcal{T}_+(K)$, we have*

$$\|\iota_\mu^* - P_S \iota_\mu^*\|_{\text{HS}(\mu, \mathcal{H})} \leq \|\iota_\mu^* - P_R \iota_\mu^*\|_{\text{HS}(\mu, \mathcal{H})} \text{ and } \|\iota_\mu^* - \iota_\mu P_S \iota_\mu^*\|_{\text{HS}(\mu)} \leq \|\iota_\mu^* - \iota_\mu P_R \iota_\mu^*\|_{\text{HS}(\mu)}.$$

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

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

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 $\mathcal{H}_{0S} \subseteq \mathcal{H}_{0R}$, and Lemma 4.1 gives

$$\|\iota_\mu P_{0S} \iota_\mu^*\|_{\text{HS}(\mu)} = \|P_{0S} L_\mu P_{0S}\|_{\text{HS}(\mathcal{H})} \leq \|P_{0R} L_\mu P_{0R}\|_{\text{HS}(\mathcal{H})} = \|\iota_\mu P_{0R} \iota_\mu^*\|_{\text{HS}(\mu)},$$

which completes the proof. \square

Following Section 4.2 and considering subspaces of \mathcal{H} defined by measures, we introduce the maps C_μ^{tr} and $C_\mu^{\text{F}} : \mathcal{T}_{\mathbb{F}}(K) \rightarrow \mathbb{R}_{\geq 0}$, with

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

The notations C_μ^{tr} and C_μ^{F} are motivated by the relation existing between these maps and the trace and Frobenius norms for matrices; see Section 4.4.

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

Proof. Following 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 $\mathcal{H}_\xi = \overline{\mathcal{H}_\nu + \mathcal{H}_\eta}^{\mathcal{H}}$, and the result follows from Lemma 4.7. \square

Remark 4.5. 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

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

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

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

The approximation error $\|i_\mu i_\mu^* i_\mu - i_\mu P_S i_\mu^* i_\mu P_S\|_{\text{HS}(\mathcal{H}, \mu)}^2$ may hence be expressed as a triple integral involving the kernels K and K_S . \triangleleft

Through the partial embedding $i_\mu P_\nu$, a measure $\nu \in \mathcal{T}_{\mathbb{F}}(K)$ defines approximations of the integral operators (with kernel K and measure μ) induced by i_μ in the four Hilbert spaces $\text{HS}(\mathcal{H})$, $\text{HS}(\mu, \mathcal{H})$, $\text{HS}(\mu)$ and $\text{HS}(\mathcal{H}, \mu)$; since these approximations are all composite of $i_\mu P_\nu$, following Section 4.2, the map D_μ may more generally be used as a surrogate for the characterisation of low-rank approximations of these operators. From Lemma 4.6, we have $C_\mu^{\text{F}}(\nu) \leq C_\mu^{\text{P}}(\nu)$, $\nu \in \mathcal{T}_{\mathbb{F}}(K)$, adding to the sequence of inequalities (16). In addition to the informative nature of the directional derivatives of D_μ , the exploration of the kernel-quadrature approximation landscape does not require the characterisation of orthogonal projections onto subspaces of \mathcal{H} ; this feature is of interest from numerical-complexity standpoint, as illustrated in the following Remarks 4.6 and 4.7.

Remark 4.6. In the framework of Lemma 3.5, for a discrete measure ν supported by a set of $n \in \mathbb{N}$ points $\mathcal{S} = \{s_1, \dots, s_n\} \subseteq \mathcal{X}$, the reproducing kernel K_ν of \mathcal{H}_ν can be written as

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

where $\chi_{i, j}$ is the i, j entry of the pseudoinverse of the $n \times n$ kernel matrix with i, j entry $K(s_i, s_j)$; see (20) in Section 4.4 for an illustration. The numerical complexity of the evaluation of K_ν at $M \in \mathbb{N}$ distinct locations in $\mathcal{X} \times \mathcal{X}$ 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 \mathcal{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. \triangleleft

Remark 4.7. Assuming that μ and ν are supported by N and n points, respectively, from (8), the computational complexity of the evaluation of $D_\mu(\nu)$ up to the constant $\|g_\mu\|_{\mathcal{G}}^2$ is $\mathcal{O}(n^2 + nN)$. In comparison, following Remark 4.6 and assuming that the kernel K_ν is evaluated from K and ν , the complexity is $\mathcal{O}(n^3 + n^2 N)$ for $C_\mu^{\text{tr}}(\nu)$, and $\mathcal{O}(n^3 + n^2 N^2)$ for $C_\mu^{\text{P}}(\nu)$, $C_\mu^{\text{PP}}(\nu)$ and $C_\mu^{\text{F}}(\nu)$.

If we for instance in addition assume that \mathcal{X} is a discrete set of size N (and so, that $n \leq N$; see Section 4.4 for an illustration), then the numerical cost of multiple evaluations of D_μ may be reduced by beforehand computing and storing the values $\{g_\mu(x)\}_{x \in \mathcal{X}}$ of the target potential $g_\mu \in \mathcal{G}$; this operation has complexity $\mathcal{O}(N^2)$, but can nevertheless be easily parallelised. From (8), the evaluation of $\langle g_\nu | g_\mu \rangle_{\mathcal{G}}$ then reduces to the integration of g_μ with respect to $\bar{\nu}$, with complexity $\mathcal{O}(n)$. \triangleleft

4.4 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*³ of \mathbf{K} induced by the columns of \mathbf{K} with index in I is the $N \times N$ PSD matrix

$$\hat{\mathbf{K}}(I) = \mathbf{K}_{\cdot, I}(\mathbf{K}_{I, I})^\dagger \mathbf{K}_{I, \cdot}, \quad (20)$$

where $\mathbf{K}_{\cdot, I}$ is the $N \times n$ 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 $\mathbf{K}_{I, \cdot}$ consists of rows of \mathbf{K} ; it is the conjugate-transpose of $\mathbf{K}_{\cdot, I}$); see e.g. [9, 18, 14, 10].

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 $\mathcal{X} = [N]$. The j -th column of \mathbf{K} then corresponds to the function $k_j \in \mathcal{H}$, with $j \in \mathcal{X}$, and the subset I defines the linear subspace $\mathcal{H}_I = \text{span}_{\mathbb{C}}\{k_j | j \in I\} \subseteq \mathcal{H}$; 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 .

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 4.3, we can then notice that

- the trace norm $\|\mathbf{K} - \hat{\mathbf{K}}(I)\|_{\text{tr}}$ corresponds to (18), and
- the squared Frobenius norm $\|\mathbf{K} - \hat{\mathbf{K}}(I)\|_{\text{F}}^2$ 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 efficient approximation $\hat{\mathbf{K}}(I)$ of \mathbf{K}) is thus a special instance of the general framework discussed in Section 4.3. In particular, the support of an approximate measure ν on $\mathcal{X} = [N]$ defines a subset of columns of \mathbf{K} , and the kernel-quadrature setting may be used as surrogate for the characterisation of such measures.

5 Concluding discussion

We described the overall framework surrounding the isometric representation of trace-class integral operators with PSD kernel K as potentials in the RKHS \mathcal{G} associated with the kernel $|K|^2$, and illustrated the parallel between the quadrature approximation of such operators and kernel quadrature rules in \mathcal{G} . Through subspaces defined by measures and partial L^2 -embeddings, we discussed the extent to which the kernel-quadrature setting can be used as a surrogate to characterise projection-based approximations of general integral operators with PSD kernels. We also illustrated the connections between the considered framework and the low-rank approximation of PSD matrices through column sampling.

³In the machine-learning literature, Nyström approximation refers to the low-rank approximation of PSD matrices through column sampling; although related, this terminology should not be confused with the quadrature method for the numerical approximation of integral equations.

In contrast to the projection-based error maps C_μ^P , C_μ^{PP} , C_μ^{tr} and C_μ^F considered in Section 4, the quadrature-based error map D_μ defined in (7) is convex and admits informative directional derivatives on any convex set of approximate measures; to this extent, the map D_μ may be regarded as a *differentiable relaxation* of the piecewise-constant projection-based error maps of Section 4. Further, from a numerical standpoint and as opposed to the projection setting, the exploration of the kernel-quadrature approximation landscape does not require the characterisation of orthogonal projections onto subspaces of \mathcal{H} , and instead only involves the evaluation of integrals related to $|K|^2$. All together, these properties make the kernel-quadrature setting an interesting framework for the design of numerically-efficient problem-dependent sampling strategies for the low-rank approximation of trace-class integral operators with PSD kernels.

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] 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.
- [11] Israel Gohberg, Seymour Goldberg, and Marius A. Kaashoek. *Classes of Linear Operators*, volume 63. Birkhäuser, 2013.
- [12] Tailen Hsing and Randall Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons, 2015.
- [13] Rainer Kress. *Linear Integral Equations*, volume 82. Springer, 1989.
- [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] Kōsaku Yosida. *Functional Analysis*. Springer, 2012.