



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. 2022. hal-03848105v1

HAL Id: hal-03848105

<https://hal.science/hal-03848105v1>

Preprint submitted on 10 Nov 2022 (v1), 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 associated with the squared-modulus of the reproducing kernel of \mathcal{H} . We discuss the properties of this coisometry, and in particular 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 integral operators are the potentials of the underlying measures with respect to the squared-modulus kernel, drawing a direct connection between the approximation of integral operators with PSD kernels and the kernel embedding of measures in RKHSs associated with squared-modulus kernels.

Keywords: reproducing kernel Hilbert spaces, integral operators, low-rank approximation, maximum mean discrepancy.

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 [3, 17, 18, 15, 21, 16, 14]. 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. [17, 18]. 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 a *potential* g_μ in the RKHS \mathcal{G} associated with the squared-modulus kernel $|K|^2$. The operator L_μ is in this case trace-class, and g_μ is the Riesz-Fréchet representer of the linear functional on \mathcal{G} corresponding to the integration with respect to the measure $\bar{\mu}$, the

*gauthierb@cardiff.ac.uk

Cardiff University, School of Mathematics

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

conjugate of μ . The design of an approximate measure ν , defining together with the kernel K an integral operator L_ν for the approximation of an initial operator L_μ , thus reduces to the approximation in \mathcal{G} of a target potential g_μ by an approximate potential g_ν . In particular, the error criterion

$$\nu \mapsto D(\nu) = \|L_\mu - L_\nu\|_{\text{HS}(\mathcal{H})}^2 = \|g_\mu - g_\nu\|_{\mathcal{G}}^2$$

can then be interpreted as a generalised *maximum-mean-discrepancy* (MMD; see e.g. [19, 12, 22]) for the embedding of measures in the RKHS \mathcal{G} , drawing a direct connection between the approximation of integral operators and the design of representative samples in kernel-based statistical learning (see [6, 10] for illustrations).

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 the Hilbert space $\text{HS}(\mathcal{H})$ of all HS operators on \mathcal{H} and onto the squared-kernel RKHS \mathcal{G} ; in particular, Γ maps self-adjoint operators to real-valued functions, and PSD operators to nonnegative functions (Section 2). Under some measurability conditions on K and assuming integrability of the diagonal of K with respect to $|\mu|$, the integral operator L_μ always belongs to the initial space of Γ , and $\Gamma[L_\mu] = g_\mu$; both L_μ and g_μ can in this case be represented as Bochner integrals (Section 3). When the initial operator L_μ is self-adjoint, i.e. when μ is signed, we also discuss the extent to which the criterion D can be used as a surrogate for the characterisation of subspaces of \mathcal{H} for the approximation of L_μ through projection. In particular, when μ is nonnegative, the RKHS \mathcal{H} is compactly embedded in $L^2(\mu)$, and the integral operator defined by K and μ can be regarded as a HS operator from, and to, \mathcal{H} or $L^2(\mu)$ (i.e. four possibilities); we discuss the implications of the considered approximation framework in this setting, and illustrate its connection with the *column-sampling problem* for low-rank PSD-matrix approximation (see e.g. [5, 11, 7]; 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 and linear maps. Inner-products of complex Hilbert spaces are assumed to be linear with respect to their right argument. We denote by \bar{z} the conjugate of a complex number $z \in \mathbb{C}$, by $|z|$ its modulus, and by $\text{Re}(z)$ its real part; the imaginary unit is denoted by $i \in \mathbb{C}$. 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 $|f|(s) = |f(s)|$, with $s \in S$, respectively; we also use the notation $|f|^2$ to refer to the function $s \mapsto |f(s)|^2$. The adjoint of an operator A is denoted by A^* .

A linear map $A : H \rightarrow F$, with H and F two Hilbert spaces, is referred to as an *isometry* if $A^*A = \text{id}_H$, the identity operator on H . The map A is a *coisometry* if A^* is an isometry, i.e. if $AA^* = \text{id}_F$. A coisometry $A : H \rightarrow F$ is in particular a *surjective partial isometry* from H onto F (i.e. $AA^*A = A$), and A^*A is in this case the orthogonal projection from H onto the *initial space* $\mathcal{I}(A)$ of A , where $\mathcal{I}(A)$ is the orthogonal in H of the null-space of A .

2.1 Hilbert-Schmidt operators and RKHSs

Hereafter, we give a detailed description of the overall framework leading to the definition 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 considered construction is summarised in the diagram (4). In all this note, the terminology *natural* is used to emphasises that a given definition 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, 14]. For all $t \in \mathcal{X}$, 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} ; we denote by $\|\cdot\|_{\mathcal{H}}$ the corresponding norm on \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. [8, 2]. 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 operator given by

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

we also set $S_b = T_{b,b}$. For all $T \in \text{HS}(\mathcal{H})$, we have $\langle T_{a,b} | T \rangle_{\text{HS}(\mathcal{H})} = \langle a | T[b] \rangle_{\mathcal{H}}$, and in particular, for all $x \in \mathcal{X}$, $T[h](x) = \langle T_{k_x, h} | T \rangle_{\text{HS}(\mathcal{H})}$.

Remark 2.1. Any operator $T \in \text{HS}(\mathcal{H})$ admits a singular value decomposition $T = \sum_{i \in \mathbb{I}} \sigma_i T_{u_i, v_i}$, where $\{u_i\}_{i \in \mathbb{I}}$ and $\{v_i\}_{i \in \mathbb{I}}$, with $\mathbb{I} \subseteq \mathbb{N}$, are two orthonormal systems (ONSs) in \mathcal{H} , and where $\{\sigma_i\}_{i \in \mathbb{I}} \in \ell^2(\mathbb{I})$, with $\sigma_i > 0$ for all $i \in \mathbb{I}$, is the set of all strictly positive singular values of T ; 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 $\zeta_h \in \mathcal{H}'$ be the bounded linear functional such that $\zeta_h(f) = \langle h | f \rangle_{\mathcal{H}}$, for all $f \in \mathcal{H}$. Endowed with the inner product $\langle \zeta_f | \zeta_h \rangle_{\mathcal{H}'} = \langle h | f \rangle_{\mathcal{H}}$, \mathcal{H}' is a Hilbert space, and the *Riesz map* $h \mapsto \zeta_h$ is a natural bijective conjugate-linear isometry from \mathcal{H} to \mathcal{H}' . The linear map densely defined as $T_{a,b} \mapsto a \otimes \zeta_b$ (see Remark 2.1) then defines a bijective natural linear isometry from $\text{HS}(\mathcal{H})$ 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 to the conjugate kernel \overline{K} , with $\overline{K}(x, t) = \overline{K(x, t)}$, for all 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 natural bijective conjugate-linear isometry from \mathcal{H} to $\overline{\mathcal{H}}$. We may in particular 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 natural bijective linear isometry from $\text{HS}(\mathcal{H})$ to the Hilbert tensor product $\mathcal{H} \otimes \overline{\mathcal{H}}$, densely defined as $\Psi(T_{a,b}) = a \otimes \overline{b}$, for all a and $b \in \mathcal{H}$.

Remark 2.3. Following Remark 2.2, the map $\zeta_h \mapsto \overline{h}$ is a natural bijective linear isometry from \mathcal{H}' to $\overline{\mathcal{H}}$. Further, the linear map densely defined as $a \otimes \zeta_b \mapsto a \otimes \overline{b}$ is a natural bijective isometry from $\mathcal{H} \otimes \mathcal{H}'$ to $\mathcal{H} \otimes \overline{\mathcal{H}}$; the composition of this isometry with the natural bijective isometry from $\text{HS}(\mathcal{H})$ to the tensor Hilbert space $\mathcal{H} \otimes \mathcal{H}'$ yields the isometry $\Psi : \text{HS}(\mathcal{H}) \rightarrow \mathcal{H} \otimes \overline{\mathcal{H}}$; see also the diagram (4). \triangleleft

Squared-kernel RKHS. The kernels K and \overline{K} being PSD, by the Schur-product theorem, so is the kernel $|K|^2 = K\overline{K}$, with for all x and $t \in \mathcal{X}$,

$$|K|^2(x, t) = K(x, t)\overline{K(x, t)} = |K(x, t)|^2 = |k_t|^2(x).$$

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

Following [14, Chapter 5], we denote by $C_\Delta : \mathcal{H} \otimes \overline{\mathcal{H}} \rightarrow \mathcal{G}$ the coisometry densely defined as $C_\Delta(a \otimes \overline{b}) = \overline{ab}$, for all a and $b \in \mathcal{H}$, where $\overline{ab} \in \mathcal{G}$ is the complex-valued function on \mathcal{X} such that $(\overline{ab})(x) = \overline{a(x)b(x)}$. For all $Y \in \mathcal{H} \otimes \overline{\mathcal{H}}$ and all $x \in \mathcal{X}$, we have

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

The initial space $\mathcal{I}(C_\Delta)$ of C_Δ is the closure in $\mathcal{H} \otimes \overline{\mathcal{H}}$ of the linear space spanned by the tensors $k_x \otimes \overline{k_x}$, for all $x \in \mathcal{X}$, i.e. $\mathcal{I}(C_\Delta) = \overline{\text{span}_{\mathbb{C}} \{k_x \otimes \overline{k_x} | x \in \mathcal{X}\}}_{\mathcal{H} \otimes \overline{\mathcal{H}}}$; see Remark 2.4 for further details.

Remark 2.4. From (1), we have $C_\Delta^* [|k_x|^2] = k_x \otimes \overline{k_x}$, for all $x \in \mathcal{X}$; the linear subspace $\text{span}_{\mathbb{C}} \{|k_x|^2 | x \in \mathcal{X}\}$ being dense in \mathcal{G} , we have $C_\Delta C_\Delta^* = \text{id}_{\mathcal{G}}$, so that C_Δ^* is an isometry. \triangleleft

Natural coisometry from $\text{HS}(\mathcal{H})$ onto \mathcal{G} . As $\Psi\Psi^* = \text{id}_{\mathcal{H} \otimes \overline{\mathcal{H}}}$, the map

$$\Gamma = C_\Delta \Psi : \text{HS}(\mathcal{H}) \rightarrow \mathcal{G}$$

is a coisometry. From (1), for all $T \in \text{HS}(\mathcal{H})$ and all $x \in \mathcal{X}$, we have

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

and in particular $\Gamma[T_{a,b}] = \overline{ab}$, for all a and $b \in \mathcal{H}$. The initial space of Γ is

$$\mathcal{I}(\Gamma) = \overline{\text{span}_{\mathbb{C}} \{S_{k_x} | x \in \mathcal{X}\}}^{\text{HS}(\mathcal{H})} \subseteq \text{HS}(\mathcal{H}), \quad (3)$$

and $\Gamma^* [|k_x|^2] = S_{k_x}$; operators in $\mathcal{I}(\Gamma)$ are isometrically represented in \mathcal{G} .

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

$$\begin{array}{ccc}
 \text{HS}(\mathcal{H}) & \xrightarrow{\Gamma} & \mathcal{G} \\
 \cong \downarrow & \searrow \Psi & \nearrow C_\Delta \\
 \mathcal{H} \otimes \mathcal{H}' & \xrightarrow{\cong} & \mathcal{H} \otimes \overline{\mathcal{H}}
 \end{array} \tag{4}$$

2.2 Self-adjoint operators

In view of (2), for all $T \in \text{HS}(\mathcal{H})$, we have $\Gamma[T^*] = \overline{\Gamma[T]}$; so, if T is self-adjoint, then the function $\Gamma[T]$ is real-valued. This property is intrinsically related to the structure of $\text{HS}(\mathcal{H})$ and \mathcal{G} , as illustrated below. Form the surjectivity of Γ , we may notice that for all $g \in \mathcal{G}$, the function \bar{g} is also a vector of \mathcal{G} .

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} ; the complex RKHS \mathcal{G} is then a *complexification* of $\mathcal{G}_{\mathbb{R}}$ (see e.g. [14, Chapter 5]) and $\mathcal{G}_{\mathbb{R}}$ is a *standard subspace* of \mathcal{G} (see e.g. [13]). More precisely, for any $g \in \mathcal{G}$, there always exist two functions g_r and $g_i \in \mathcal{G}_{\mathbb{R}} \subset \mathcal{G}$ such that $g = g_r + ig_i$, with

$$g_r = \frac{1}{2}(\bar{g} + g) \quad \text{and} \quad g_i = \frac{1}{2}i(\bar{g} - g), \tag{5}$$

i.e. g_r and g_i are the real and imaginary parts of g . We have $\mathcal{G} = \mathcal{G}_{\mathbb{R}} + i\mathcal{G}_{\mathbb{R}}$, and for $f = f_r + if_i \in \mathcal{G}$, with f_r and $f_i \in \mathcal{G}_{\mathbb{R}}$,

$$\langle f | g \rangle_{\mathcal{G}} = \langle f_r | g_r \rangle_{\mathcal{G}_{\mathbb{R}}} + i\langle f_r | g_i \rangle_{\mathcal{G}_{\mathbb{R}}} - i\langle f_i | g_r \rangle_{\mathcal{G}_{\mathbb{R}}} + \langle f_i | g_i \rangle_{\mathcal{G}_{\mathbb{R}}}.$$

As $\langle f_r | g_r \rangle_{\mathcal{G}_{\mathbb{R}}} = \langle f_r | g_r \rangle_{\mathcal{G}}$, with a slight abuse of notation, we may use the same notation $\langle \cdot | \cdot \rangle_{\mathcal{G}}$ to refer to the inner-products of $\mathcal{G}_{\mathbb{R}}$ and \mathcal{G} .

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} . Endowed with the inner-product of $\text{HS}(\mathcal{H})$, the real vector space $\text{HS}_*(\mathcal{H})$ is a real Hilbert space. 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

$$T_r = \frac{1}{2}(T^* + T) \quad \text{and} \quad T_i = \frac{1}{2}i(T^* - T), \tag{6}$$

and the decomposition $\text{HS}(\mathcal{H}) = \text{HS}_*(\mathcal{H}) + i\text{HS}_*(\mathcal{H})$ holds.

Natural real-linear coisometry. With the above notations and from (5) and (6), if $\Gamma[T] = g$, then by linearity, we have $\Gamma[T_r] = g_r$ and $\Gamma[T_i] = g_i$. In particular, the restriction of Γ to $\text{HS}_*(\mathcal{H})$ defines a natural real-linear coisometry $\tilde{\Gamma}$ from the real Hilbert space $\text{HS}_*(\mathcal{H})$ onto the real RKHS $\mathcal{G}_{\mathbb{R}}$. The initial space of $\tilde{\Gamma}$ is

$$\mathcal{I}(\tilde{\Gamma}) = \mathcal{I}(\Gamma) \cap \text{HS}_*(\mathcal{H}) = \overline{\text{span}_{\mathbb{R}} \{S_{k_x} | x \in \mathcal{X}\}}^{\text{HS}(\mathcal{H})},$$

and $\mathcal{I}(\Gamma) = \mathcal{I}(\tilde{\Gamma}) + i\mathcal{I}(\tilde{\Gamma})$.

Remark 2.5. The diagram (4) is also well-defined when \mathcal{H} , $\text{HS}(\mathcal{H})$ and \mathcal{G} are real Hilbert spaces; the coisometry Γ is then real-linear. From (2), we in this case have $\Gamma[T^*] = \Gamma[T]$ for all $T \in \text{HS}(\mathcal{H})$, so that $\Gamma[T] = 0$ when $T^* = -T$; in the real case, the operators in $\mathcal{I}(\Gamma)$ are thus all self-adjoint. \triangleleft

2.3 Positive-semidefinite operators

If $T \in \text{HS}_*(\mathcal{H})$ is PSD, then $\Gamma[T](x) = \langle k_x | T[k_x] \rangle_{\mathcal{H}} \geq 0$, for all $x \in \mathcal{X}$, and the function $\Gamma[T] \in \mathcal{G}_{\mathbb{R}}$ is nonnegative; again, this property is intrinsically related to the structure 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 convex cone of all (self-adjoint) PSD operators on \mathcal{H} . For all $T \in \text{HS}_*(\mathcal{H})$, there exist T^+ and $T^- \in \text{HS}_*^+(\mathcal{H})$ such that $T = T^+ - T^-$, i.e. 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 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 (i.e. for all $g \in \mathcal{G}_{\mathbb{R}}$, there exist g^+ and $g^- \in \mathcal{G}_{\mathbb{R}}^+$ such that $g = g^+ - g^-$).

Remark 2.6. Let $T = \sum_{j \in \mathbb{I}} \lambda_j S_{\varphi_j}$ be a spectral expansion of $T \in \text{HS}_*^+(\mathcal{H})$, where $\{\lambda_j\}_{j \in \mathbb{I}}$, with $\mathbb{I} \subseteq \mathbb{N}$, is 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 $\Gamma[T] = 0$ if and only if $T = 0$. \triangleleft

3 Trace-class integral operators with PSD kernels

From (3), if $T \in \text{HS}(\mathcal{H})$ is of the form $T = \sum_{j=1}^n \omega_j S_{k_{s_j}}$, with $n \in \mathbb{N}$, $\{s_j\}_{j=1}^n \in \mathcal{X}^n$ and $\{\omega_j\}_{j=1}^n \in \mathbb{C}^n$, then $T \in \mathcal{I}(\Gamma)$ and $\Gamma[T] = \sum_{j=1}^n \omega_j |k_{s_j}|^2$. For all $h \in \mathcal{H}$ and all

$x \in \mathcal{X}$, we in this case have

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

and T can thus be regarded as an integral operator 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 $x \in \mathcal{X}$. Under some general measurability conditions, a larger collection of integral operators belonging to $\mathcal{I}(\Gamma)$ can be characterised.

We consider a σ -algebra \mathcal{A} of subsets of \mathcal{X} . We make the following assumptions on the kernel 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 $(\mathcal{X}, \mathcal{A})$.

We denote by \mathcal{M}_+ , \mathcal{M} , and $\mathcal{M}_{\mathbb{C}}$ the set of all nonnegative, signed and complex measures* on $(\mathcal{X}, \mathcal{A})$, respectively. For $\mu \in \mathcal{M} \cup \mathcal{M}_{\mathbb{C}}$ (with $\mathcal{M}_+ \subset \mathcal{M}$), we set

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

with $|\mu|$ the variation of μ ; we denote by $\bar{\mu}$ the conjugate measure of μ . We next define 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. We may notice that $\mathcal{T}_{\mathbb{C}}(K)$ is a vector space, and that $\mathcal{T}_+(K)$ is a real pointed convex cone; the inclusion $\mathcal{T}_+(K) \subset \mathcal{T}(K)$ holds.

For $\mu \in \mathcal{T}(K) \cup \mathcal{T}_{\mathbb{C}}(K)$, we denote by $L_{\mu} \in \text{HS}(\mathcal{H})$ the integral operator on \mathcal{H} defined by the kernel K and the measure μ , that is

$$L_{\mu}[h](x) = \int_{\mathcal{X}} K(x, t) h(t) d\mu(t), \text{ for all } h \in \mathcal{H} \text{ and all } x \in \mathcal{X};$$

see the proof of Theorem 3.1 and Remarks 3.1 and 3.2 for more details on L_{μ} .

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

$$g_{\mu}(x) = \int_{\mathcal{X}} |K(x, t)|^2 d\mu(t), \text{ for all } x \in \mathcal{X}.$$

Proof. From (A.1), for all $t \in \mathcal{X}$, the function $|k_t|^2 \in \mathcal{G}_{\mathbb{R}}$ is measurable on $(\mathcal{X}, \mathcal{A})$, so that the RKHS \mathcal{G} and the real RKHS $\mathcal{G}_{\mathbb{R}}$ both consist of measurable functions on $(\mathcal{X}, \mathcal{A})$; see e.g. [20, Lemma 4.24]. In particular, for all $T \in \text{HS}(\mathcal{H})$, the function $\Gamma[T]$ is measurable.

For all $\mu \in \mathcal{T}(K) \cup \mathcal{T}_{\mathbb{C}}(K)$, the operator $L_{\mu} \in \text{HS}(\mathcal{H})$ is the Riesz-Fréchet representer of the bounded linear functional $Z_{\mu} : \text{HS}(\mathcal{H}) \rightarrow \mathbb{C}$ such that

$$Z_{\mu}(T) = \int_{\mathcal{X}} \langle S_{k_t} | T \rangle_{\text{HS}(\mathcal{H})} d\bar{\mu}(t) = \int_{\mathcal{X}} \Gamma[T](t) d\bar{\mu}(t), \text{ for all } T \in \text{HS}(\mathcal{H}).$$

*We only consider finite complex measures, while nonnegative and signed measures might not necessarily be finite.

Indeed, as $\|S_{k_t}\|_{\text{HS}(\mathcal{H})} = \|k_t\|_{\mathcal{H}}^2 = K(t, t)$ for all $t \in \mathcal{X}$, we have

$$|Z_\mu(T)| \leq \int_{\mathcal{X}} \left| \langle S_{k_t} | T \rangle_{\text{HS}(\mathcal{H})} \right| d|\mu|(t) \leq \|T\|_{\text{HS}(\mathcal{H})} \tau_\mu,$$

and $Z_\mu(T_{k_x, h}) = \langle L_\mu | T_{k_x, h} \rangle_{\text{HS}(\mathcal{H})} = \overline{L_\mu[h]}(x)$, for all $h \in \mathcal{H}$ and all $x \in \mathcal{X}$. In a similar way, the function $g_\mu \in \mathcal{G}$ is the Riesz-Fréchet representer of the bounded linear functional $I_\mu : \mathcal{G} \rightarrow \mathbb{C}$ such that

$$I_\mu(g) = \int_{\mathcal{X}} g(t) d\bar{\mu}(t), \text{ for all } g \in \mathcal{G};$$

we indeed have $|I_\mu(g)| \leq \|g\|_{\mathcal{G}} \tau_\mu$, and $I_\mu(|k_x|^2) = \langle g_\mu | |k_x|^2 \rangle_{\mathcal{G}} = \overline{g_\mu(x)}$. For all $x \in \mathcal{X}$, we have

$$\Gamma[L_\mu](x) = \langle S_{k_x} | L_\mu \rangle_{\text{HS}(\mathcal{H})} = \overline{Z_\mu(S_{k_x})} = g_\mu(x),$$

so that $\Gamma[L_\mu] = g_\mu$. We then notice that

$$\|L_\mu\|_{\text{HS}(\mathcal{H})}^2 = Z_\mu(L_\mu) = \iint_{\mathcal{X}} |K(x, t)|^2 d\mu(t) d\bar{\mu}(x) = I_\mu(g_\mu) = \|g_\mu\|_{\mathcal{G}}^2,$$

and thus $L_\mu \in \mathcal{I}(\Gamma)$. □

Remark 3.1. For all $\mu \in \mathcal{T}(K) \cup \mathcal{T}_{\mathbb{C}}(K)$, the operator $L_\mu \in \text{HS}(\mathcal{H})$ and the function $g_\mu \in \mathcal{G}$ can be represented as Bochner integrals, that is

$$L_\mu = \int_{\mathcal{X}} S_{k_t} d\mu(t), \quad \text{and} \quad g_\mu = \int_{\mathcal{X}} |k_t|^2 d\mu(t);$$

the strong measurability of the maps $t \mapsto S_{k_t}$ and $t \mapsto |k_t|^2$ follows from the weak measurability condition (A.1) and the separability of $\text{HS}(\mathcal{H})$ and \mathcal{G} (Pettis measurability theorem), and we in addition have

$$\int_{\mathcal{X}} \|S_{k_t}\|_{\text{HS}(\mathcal{H})} d|\mu|(t) = \int_{\mathcal{X}} \| |k_t|^2 \|_{\mathcal{G}} d|\mu|(t) = \tau_\mu.$$

For all $h \in \mathcal{H}$, we may also notice that the vector $L_\mu[h] \in \mathcal{H}$ is the Riesz-Fréchet representer of the bounded linear functional $\Theta_{h, \mu} : \mathcal{H} \rightarrow \mathbb{C}$, with

$$\Theta_{h, \mu}(f) = \int_{\mathcal{X}} \overline{h(t)} f(t) d\bar{\mu}(t) = \langle L_\mu[h] | f \rangle_{\mathcal{H}}, \text{ for all } f \in \mathcal{H}; \quad (7)$$

we indeed have $|\Theta_{h, \mu}(f)| \leq \|f\|_{\mathcal{H}} \|h\|_{\mathcal{H}} \tau_\mu$, and $\Theta_{h, \mu}(k_x) = \overline{T_\mu[h]}(x)$, for all $x \in \mathcal{X}$. The vector $L_\mu[h]$ can also be represented as a Bochner integral, that is

$$L_\mu[h] = \int_{\mathcal{X}} k_t h(t) d\mu(t);$$

the strong measurability of the map $t \mapsto k_t$ follows from (A.1) and the separability of \mathcal{H} , and the inequality $\int_{\mathcal{X}} \|k_t\|_{\mathcal{H}} |h(t)| d|\mu|(t) \leq \|h\|_{\mathcal{H}} \tau_\mu$ holds. ◁

From Theorem 3.1 and following Sections 2.2 and 2.3, we can readily notice that if $\mu \in \mathcal{T}(K)$, then $g_\mu \in \mathcal{G}_\mathbb{R}$, and $\Gamma^*[g_\mu] = L_\mu \in \text{HS}_*(\mathcal{H})$; also, if $\mu \in \mathcal{T}_+(K)$, then $g_\mu \in \mathcal{G}_\mathbb{R}^+$, and $\Gamma^*[g_\mu] = L_\mu \in \text{HS}_*^+(\mathcal{H})$. As $\delta_t \in \mathcal{T}_+(K)$ and $L_{\delta_t} = S_{k_t}$ for all $t \in \mathcal{X}$, we may also notice that $\text{span}_\mathbb{C}\{L_\nu | \nu \in \mathcal{T}_+(K)\}$ is dense in $\mathcal{I}(\Gamma)$.

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

Proof. For all $\mu \in \mathcal{T}_+(K)$, the operator L_μ is PSD. Let $\{h_i\}_{i \in \mathbb{N}}$ be any orthonormal basis (ONB) of \mathcal{H} , with $\mathbb{N} \subseteq \mathbb{N}$; from (7) and by monotone convergence, we have

$$\sum_{i \in \mathbb{N}} \langle h_i | L_\mu[h_i] \rangle_{\mathcal{H}} = \sum_{i \in \mathbb{N}} \int_{\mathcal{X}} |\langle k_t | h_i \rangle_{\mathcal{H}}|^2 d\mu(t) = \int_{\mathcal{X}} \|k_t\|_{\mathcal{H}}^2 d\mu(t) = \tau_\mu, \quad (8)$$

so that L_μ is trace-class, with trace τ_μ .

For all $\mu \in \mathcal{T}(K)$, there exist μ^+ and $\mu^- \in \mathcal{M}_+$ such that $\mu = \mu^+ - \mu^-$ and $|\mu| = \mu^+ + \mu^-$; we thus have μ^+ and $\mu^- \in \mathcal{T}_+(K)$. For all $\mu \in \mathcal{T}_\mathbb{C}(K)$, we have $|\mu| = |\bar{\mu}|$, so that $\bar{\mu} \in \mathcal{T}_\mathbb{C}(K)$; we then have $\mu = \mu_r + i\mu_i$, with $\mu_r = \frac{1}{2}(\bar{\mu} + \mu)$ and $\mu_i = \frac{1}{2}i(\bar{\mu} - \mu) \in \mathcal{T}_\mathbb{C}(K) \cap \mathcal{M} \subseteq \mathcal{T}(K)$. For all $\mu \in \mathcal{T}(K) \cup \mathcal{T}_\mathbb{C}(K)$, the operator L_μ can therefore be written as a finite sum of trace-class operators. \square

Remark 3.2. For $\mu \in \mathcal{T}_+(K)$, let $L^2(\mu)$ be the Hilbert space of all square-integrable functions with respect to μ . For all $h \in \mathcal{H}$, we have $\int_{\mathcal{X}} |h(t)|^2 d\mu(t) \leq \|h\|_{\mathcal{H}}^2 \tau_\mu$, 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 $h \in \mathcal{H}$, is bounded (see for instance [21]). Let $\{h_i\}_{i \in \mathbb{N}}$ be any ONB of \mathcal{H} , with $\mathbb{N} \subseteq \mathbb{N}$; from (7) and (8), we have

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

and ι_μ is thus HS. From (7), we can in addition notice that

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

so that $L_\mu = \iota_\mu^* \iota_\mu$ for all $\mu \in \mathcal{T}_+(K)$; see Section 4.2 for a further discussion. \triangleleft

For $\mu \in \mathcal{T}(K) \cup \mathcal{T}_\mathbb{C}(K)$, we refer to $g_\mu = \int_{\mathcal{X}} |k_t|^2 d\mu(t)$ as the *potential of μ with respect to $|K|^2$* ; for all $g \in \mathcal{G}$, we have (see the proof of Theorem 3.1)

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

The squared norm $\|g_\mu\|_{\mathcal{G}}^2$ is sometimes referred to as the *energy of μ with respect to the kernel $|K|^2$* ; see for instance [4]. For μ and $\nu \in \mathcal{T}(K) \cup \mathcal{T}_\mathbb{C}(K)$, we have

$$\begin{aligned} \langle L_\nu | L_\mu \rangle_{\text{HS}(\mathcal{H})} &= \langle g_\nu | g_\mu \rangle_{\mathcal{G}} = \iint_{\mathcal{X}} |K(x, t)|^2 d\mu(t) d\bar{\nu}(t) \\ &= \int_{\mathcal{X}} g_\mu(t) d\bar{\nu}(t) = \int_{\mathcal{X}} \overline{g_\nu(t)} d\mu(t), \end{aligned} \quad (9)$$

and the criterion $D(\nu) = \|L_\mu - L_\nu\|_{\text{HS}(\mathcal{H})}^2$ thus corresponds to a MMD (see e.g. [19, 12]) for the kernel embedding of measures in the RKHS \mathcal{G} . In particular, the map D is convex on any convex set $\mathcal{C} \subseteq \mathcal{T}(K) \cup \mathcal{T}_{\mathbb{C}}(K)$, and for all ν and $\eta \in \mathcal{C}$, the directional derivative of D at ν in the direction $\eta - \nu$ is

$$\lim_{\rho \rightarrow 0^+} \frac{1}{\rho} \left[D(\nu + \rho(\eta - \nu)) - D(\nu) \right] = 2\text{Re}(\langle L_\nu - L_\mu \mid L_\eta - L_\nu \rangle_{\text{HS}(\mathcal{H})}).$$

From a numerical standpoint and as the double integral in (9) solely involves the kernel $|K|^2$, we may also notice that related quantities can in practice be efficiently evaluated or approximated; see [6, 10] for illustrations (see also Remark 4.3).

4 Subspaces defined by measures

The parallel between integral-operator approximation and the approximation of potentials suggests that strategies inspired from the kernel-embedding-of-measures literature may be applied to design measures for the approximation of integral operators with PSD kernels.

When L_μ is self-adjoint, i.e. for $\mu \in \mathcal{T}(K)$, the MMD setting described in Section 3 may more generally be used as a surrogate for the characterisation of closed linear subspaces of \mathcal{H} for the approximation of L_μ through projection; this framework in particular encompasses the approximation of PSD matrices via *column sampling*, as discussed hereafter.

4.1 Self-adjoint operators and projections

Let \mathcal{H}_S be a closed linear subspace of \mathcal{H} . We denote by K_S the reproducing kernel of \mathcal{H}_S , and by $P_{\mathcal{H}_S}$ the orthogonal projection from \mathcal{H} onto \mathcal{H}_S ; we then in particular have $P_{\mathcal{H}_S}[k_t](x) = K_S(x, t)$, for all x and $t \in \mathcal{X}$.

For $\mu \in \mathcal{T}(K)$, the operator $P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}$ is the orthogonal projection of L_μ onto the closed real-linear subspace \mathcal{R}_S of all operators in $\text{HS}_*(\mathcal{H})$ with range included in \mathcal{H}_S . We have $P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}[h](x) = \int_{\mathcal{X}} K_S(x, t) P_{\mathcal{H}_S}[h](t) d\mu(t)$, for all $h \in \mathcal{H}$ and all $x \in \mathcal{X}$, and the approximation error verifies

$$\|L_\mu - P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}\|_{\text{HS}(\mathcal{H})}^2 = \|L_\mu\|_{\text{HS}(\mathcal{H})}^2 - \|P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}\|_{\text{HS}(\mathcal{H})}^2, \quad (10)$$

with $\|P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}\|_{\text{HS}(\mathcal{H})}^2 = \iint_{\mathcal{X}} |K_S(x, t)|^2 d\mu(x) d\mu(t)$.

Remark 4.1. For a given rank, the subspaces leading to minimal values of (10) correspond to spectral truncations of L_μ (see e.g. [9, Theorem 4.4.7]). In practice, a direct implementation of such optimal approximations is thus limited to operators for which eigenvectors can be computed beforehand. \triangleleft

The operator L_μ being self-adjoint, for a complex measure $\nu \in \mathcal{T}_{\mathbb{C}}(K)$, with real part ν_r , we have $D(\nu_r) \leq D(\nu)$; hence, for $\mu \in \mathcal{T}(K)$, the search of a measure ν for

the approximation of L_μ by L_ν may be restricted to $\mathcal{T}(K)$. For $\nu \in \mathcal{T}(K)$, we denote by $\mathcal{H}_\nu \subseteq \mathcal{H}$ the closure in \mathcal{H} of the range of L_ν ; for instance, if $\nu = \sum_{i=1}^n v_i \delta_{s_i}$, with $v_i \in \mathbb{R}$, $v_i \neq 0$, and $s_i \in \mathcal{X}$, then $\mathcal{H}_\nu = \text{span}_{\mathbb{C}}\{k_{s_1}, \dots, k_{s_n}\}$. In addition to L_ν , the measure ν also defines the approximation $P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu}$ of L_μ .

Lemma 4.1. *For μ and $\nu \in \mathcal{T}(K)$, we have*

$$\|L_\mu - L_\nu\|_{\text{HS}(\mathcal{H})}^2 = \|L_\mu - P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu}\|_{\text{HS}(\mathcal{H})}^2 + \|P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu} - L_\nu\|_{\text{HS}(\mathcal{H})}^2. \quad (11)$$

Proof. Both L_ν and $P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu}$ have range in \mathcal{H}_ν . As $P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu}$ is the orthogonal projection of L_μ onto \mathcal{R}_ν , the closed real-linear subspace of all operators in $\text{HS}_*(\mathcal{H})$ with range in \mathcal{H}_ν , we have $\langle L_\mu - P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu} | P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu} - L_\nu \rangle_{\text{HS}(\mathcal{H})} = 0$. \square

The decomposition (11) may be interpreted as follows.

- The term $C_{\text{pr}}(\nu) = \|L_\mu - P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu}\|_{\text{HS}(\mathcal{H})}^2$ measures the similarity between L_μ and the approximation $P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu}$. This term corresponds to the component of $D(\nu)$ which only depends on the range of L_ν ; it is by definition constant for all measures $\eta \in \mathcal{T}(K)$ such that $\mathcal{H}_\eta = \mathcal{H}_\nu$. For instance, when ν is finitely supported, $C_{\text{pr}}(\nu)$ only depends on the support on ν .
- As $L_\nu = P_{\mathcal{H}_\nu} L_\nu P_{\mathcal{H}_\nu}$, the restrictions to \mathcal{H}_ν of $P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu}$ and L_ν correspond to the integral operators on \mathcal{H}_ν defined by the kernel K_ν and the measures μ and ν , respectively. Following Section 3 and noticing that $\mathcal{T}(K) \subseteq \mathcal{T}(K_\nu)$, the term $\|P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu} - L_\nu\|_{\text{HS}(\mathcal{H})}^2$ is thus a MMD for the kernel embedding of measures in the RKHS associated to $|K_\nu|^2$; by contrast with C_{pr} , this term allows D to discriminate among measures defining operators with same range.

The following Lemma 4.2 illustrates that the criterion C_{pr} can be regarded as a convex piecewise-constant map on the real convex cone $\mathcal{T}_+(K)$.

Lemma 4.2. *For $\mu \in \mathcal{T}(K)$, the map C_{pr} is convex on the real convex cone $\mathcal{T}_+(K)$, and for all ν and $\eta \in \mathcal{T}_+(K)$, we have*

$$\lim_{\rho \rightarrow 0^+} \frac{1}{\rho} [C_{\text{pr}}(\nu + \rho(\eta - \nu)) - C_{\text{pr}}(\nu)] \in \{-\infty, 0\}. \quad (12)$$

Denoting by $\mathcal{H}_{\eta \setminus \nu}$ the orthogonal of \mathcal{H}_ν in $\overline{\mathcal{H}_\nu + \mathcal{H}_\eta}^{\mathcal{H}}$, a sufficient condition for the directional derivative (12) to be null is the orthogonality between $\mathcal{H}_{\eta \setminus \nu}$ and \mathcal{H}_μ ; when μ is nonnegative, this sufficient condition is also necessary.

Proof. Let \mathcal{H}_S and \mathcal{H}_R be two closed linear subspaces of \mathcal{H} , with $\mathcal{H}_S \subseteq \mathcal{H}_R$; we denote by \mathcal{H}_e the orthogonal of \mathcal{H}_S in \mathcal{H}_R . Noticing that $P_{\mathcal{H}_R} = P_{\mathcal{H}_S} + P_{\mathcal{H}_e}$ and that $P_{\mathcal{H}_S} P_{\mathcal{H}_e} = P_{\mathcal{H}_e} P_{\mathcal{H}_S} = 0$, we have

$$\|P_{\mathcal{H}_R} L_\mu P_{\mathcal{H}_R}\|_{\text{HS}(\mathcal{H})}^2 = \|P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}\|_{\text{HS}(\mathcal{H})}^2 + \|P_{\mathcal{H}_e} L_\mu P_{\mathcal{H}_e}\|_{\text{HS}(\mathcal{H})}^2 + 2\|P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_e}\|_{\text{HS}(\mathcal{H})}^2$$

so that $\|P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}\|_{\text{HS}(\mathcal{H})}^2 \leq \|P_{\mathcal{H}_R} L_\mu P_{\mathcal{H}_R}\|_{\text{HS}(\mathcal{H})}^2$. If \mathcal{H}_e is orthogonal to \mathcal{H}_μ , we have $L_\mu P_{\mathcal{H}_e} = P_{\mathcal{H}_e} L_\mu = 0$, and the previous inequality is an equality; in particular, if L_μ is PSD, then $P_{\mathcal{H}_e} L_\mu P_{\mathcal{H}_e} = 0$ if and only if $L_\mu P_{\mathcal{H}_e} = P_{\mathcal{H}_e} L_\mu = 0$.

For ν and $\eta \in \mathcal{T}_+(K)$, and for $\rho \in (0, 1)$, 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 $\mathcal{H}_\xi = \overline{\mathcal{H}_\nu + \mathcal{H}_\eta}^{\mathcal{H}}$, so that the map $\rho \mapsto C_{\text{pr}}(\nu + \rho(\eta - \nu))$ is constant on the open interval $(0, 1)$. From (10) and as $\mathcal{H}_\nu \subseteq \mathcal{H}_\xi$, we have $C_{\text{pr}}(\xi) \leq C_{\text{pr}}(\nu)$, with equality when $\mathcal{H}_{\eta \setminus \nu}$ is orthogonal to \mathcal{H}_μ . In the same way, we have $C_{\text{pr}}(\xi) \leq C_{\text{pr}}(\eta)$, concluding the proof. \square

Lemma 4.1 illustrates the extent to which the criterion D may be used as a surrogate for the characterisation of subspaces $\mathcal{H}_\nu \subseteq \mathcal{H}$, defined by measures ν , for the approximation of L_μ by $P_{\mathcal{H}_\nu} L_\mu P_{\mathcal{H}_\nu}$. In view of Lemma 4.2 and in contrast to the MMD criterion D , we can notice that the directional derivatives of C_{pr} on any convex set of measures $\mathcal{C} \subseteq \mathcal{T}(K)$ are *noninformative*, in the sense that they do not provide any information on the local steepness of the landscape of C_{pr} . Also, from a numerical standpoint and in view of (10), for all $\nu \in \mathcal{T}(K)$, the computations underlying the evaluation of $C_{\text{pr}}(\nu)$ involve the reproducing kernel K_ν of \mathcal{H}_ν (or equivalently, the orthogonal projection $P_{\mathcal{H}_\nu}$); in practice, this kernel is a priori unknown and needs to be computed from K and ν ; see Section 4.3 for an illustration. In comparison, the criterion D solely involves the kernel K .

Remark 4.2. Let \mathcal{H}_S be a closed linear subspace of \mathcal{H} and let $\mu \in \mathcal{T}(K)$; we set $g_\mu^S = \Gamma[P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}] \in \mathcal{G}$. From (7), for all $x \in \mathcal{X}$, we have

$$g_\mu^S(x) = \langle P_{\mathcal{H}_S}[k_x] \mid L_\mu P_{\mathcal{H}_S}[k_x] \rangle_{\mathcal{H}} = \int_{\mathcal{X}} |K_S(x, t)|^2 d\mu(t),$$

and $\|P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}\|_{\text{HS}(\mathcal{H})}^2 = \langle g_\mu \mid g_\mu^S \rangle_{\mathcal{G}} \geq \|g_\mu^S\|_{\mathcal{G}}^2$. The operator $P_{\mathcal{H}_S} L_\mu P_{\mathcal{H}_S}$ does generally not belong to $\mathcal{I}(\Gamma)$; this occurs only in specific situations, e.g. for $\mathcal{H}_\mu \subseteq \mathcal{H}_S$, or for $\mathcal{H}_S = \mathcal{H}_{\delta_x}$, with $x \in \mathcal{X}$. \triangleleft

4.2 Nonnegative measures and approximate embeddings

Following Remark 3.2, for $\mu \in \mathcal{T}_+(K)$, the embedding $\iota_\mu : \mathcal{H} \rightarrow L^2(\mu)$ is HS. For all $f \in L^2(\mu)$ and all $x \in \mathcal{X}$, we have

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

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

$$\iota_\mu^* : L^2(\mu) \rightarrow \mathcal{H}, \quad \iota_\mu \iota_\mu^* : L^2(\mu) \rightarrow L^2(\mu), \quad \text{and} \quad \iota_\mu \iota_\mu^* \iota_\mu : \mathcal{H} \rightarrow L^2(\mu), \quad (13)$$

all correspond to different interpretations of an integral operator defined by K and $\mu \in \mathcal{T}_+(K)$. These four interpretations are *natural* in the sense that they are inherent

to the considered kernel and measure; see for instance [3, 18, 15, 21, 16]. In each case, the operator is HS.

For a closed linear subspace $\mathcal{H}_S \subseteq \mathcal{H}$, the embedding ι_μ can be approximated by $\iota_\mu P_{\mathcal{H}_S}$; for all $f \in L^2(\mu)$ and all $x \in \mathcal{X}$, we may notice that

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

For $L_\mu = \iota_\mu^* \iota_\mu \in \text{HS}(\mathcal{H})$, we obtain the approximation discussed in Section 4.1; and in the three cases considered in (13), we obtain the approximations

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

with $\text{HS}(\mu, \mathcal{H})$, $\text{HS}(\mu)$ 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. We for instance have

$$\|\iota_\mu^* - P_{\mathcal{H}_S} \iota_\mu^*\|_{\text{HS}(\mu, \mathcal{H})}^2 = \int_{\mathcal{X}} K(t, t) - K_S(t, t) d\mu(t), \quad (14)$$

$$\|\iota_\mu \iota_\mu^* - \iota_\mu P_{\mathcal{H}_S} \iota_\mu^*\|_{\text{HS}(\mu)}^2 = \iint_{\mathcal{X}} |K(x, t) - K_S(x, t)|^2 d\mu(x) d\mu(t), \quad \text{and} \quad (15)$$

$$\langle \iota_\mu \iota_\mu^* \iota_\mu | \iota_\mu P_{\mathcal{H}_S} \iota_\mu^* \iota_\mu P_{\mathcal{H}_S} \rangle_{\text{HS}(\mathcal{H}, \mu)} = \iiint_{\mathcal{X}} K(t, x) K_S(x, s) K_S(s, t) d\mu(t) d\mu(s) d\mu(x).$$

The following inequality holds between the approximations in $\text{HS}(\mathcal{H})$ and $\text{HS}(\mu)$.

Lemma 4.3. *We consider a measure $\mu \in \mathcal{T}_+(K)$ and a closed linear subspace \mathcal{H}_S of \mathcal{H} ; we have $\|\iota_\mu \iota_\mu^* - \iota_\mu P_{\mathcal{H}_S} \iota_\mu^*\|_{\text{HS}(\mu)}^2 \leq \|\iota_\mu \iota_\mu^* - P_{\mathcal{H}_v} \iota_\mu^* \iota_\mu P_{\mathcal{H}_v}\|_{\text{HS}(\mathcal{H})}^2$.*

Proof. We denote by \mathcal{H}_{0S} the orthogonal of \mathcal{H}_S in \mathcal{H} , and let $K_{0S} = K - K_S$ be the reproducing kernel of \mathcal{H}_{0S} . Using an ONB of \mathcal{H}_{0S} and Fubini's theorem, we have

$$\|P_{\mathcal{H}_S} \iota_\mu^* \iota_\mu P_{\mathcal{H}_{0S}}\|_{\text{HS}(\mathcal{H})}^2 = \iint_{\mathcal{X}} K_{0S}(x, t) \overline{K_S(x, t)} d\mu(x) d\mu(t) \geq 0,$$

and the result follows from (10) and (15). \square

As observed in Lemma 4.2 for the criterion C_{pr} , the three maps

$$\begin{aligned} \nu \mapsto \|\iota_\mu^* - P_{\mathcal{H}_v} \iota_\mu^*\|_{\text{HS}(\mu, \mathcal{H})}^2, \quad \nu \mapsto \|\iota_\mu \iota_\mu^* - \iota_\mu P_{\mathcal{H}_v} \iota_\mu^*\|_{\text{HS}(\mu)}^2 \quad \text{and} \\ \nu \mapsto \|\iota_\mu \iota_\mu^* \iota_\mu - \iota_\mu P_{\mathcal{H}_v} \iota_\mu^* \iota_\mu P_{\mathcal{H}_v}\|_{\text{HS}(\mathcal{H}, \mu)}^2, \end{aligned} \quad (16)$$

are convex and piecewise-constant on $\mathcal{T}_+(K)$. Through the closed linear subspace $\mathcal{H}_v \subseteq \mathcal{H}$, a measure $\nu \in \mathcal{T}(K)$ hence defines an approximation in the four HS spaces $\text{HS}(\mathcal{H})$, $\text{HS}(\mu, \mathcal{H})$, $\text{HS}(\mu)$ and $\text{HS}(\mathcal{H}, \mu)$. As these four approximations are all induced by the approximate embedding $\iota_\mu P_{\mathcal{H}_v}$, following Section 4.1, the MMD criterion D may more generally be used as a surrogate for the characterisation of low-rank approximations of any of the four operators induced by the embedding ι_μ .

4.3 Column sampling for PSD-matrix approximation

Let \mathbf{K} be a $N \times N$ PSD matrix, with $N \in \mathbb{N}$; the *Nyström approximation** of \mathbf{K} defined by a subset $I \subseteq \llbracket 1, N \rrbracket$, of size $n \leq N$, is the $N \times N$ PSD matrix

$$\hat{\mathbf{K}}(I) = \mathbf{K}_{\cdot, I}(\mathbf{K}_{I, I})^\dagger \mathbf{K}_{I, \cdot}, \quad (17)$$

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. [5, 11, 7].

For i and $j \in \llbracket 1, N \rrbracket$, 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 $\mathcal{X} = \llbracket 1, N \rrbracket$. 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 4.2, we can then notice that

- the trace norm $\|\mathbf{K} - \hat{\mathbf{K}}(I)\|_{\text{tr}}$ corresponds to (14), and
- the squared Frobenius norm $\|\mathbf{K} - \hat{\mathbf{K}}(I)\|_{\text{F}}^2$ corresponds to (15).

The column-sampling problem for the Nyström approximation of a PSD matrix \mathbf{K} is thus a special instance of the framework discussed in the previous Section 4.2. In particular, the support of an approximate measure ν on $\mathcal{X} = \llbracket 1, N \rrbracket$ characterises a subset of columns of \mathbf{K} , and the MMD setting of Section 3 may be used as surrogate for the design such measures; see [6, 10] for some illustrations.

Remark 4.3. For discrete measures μ and ν supported by N and n points, respectively, and from (9), the computational complexity of the evaluation of $D(\nu)$ up to the constant $\|L_\mu\|_{\text{HS}(\mathcal{H})}^2$ is $\mathcal{O}(n^2 + nN)$. In comparison, for $\mathcal{H}_S = \mathcal{H}_\nu$ and assuming that the kernel K_ν is evaluated from K and ν following (17), the computational complexity is $\mathcal{O}(n^3 + n^2N)$ for (14), and $\mathcal{O}(n^3 + n^2N^2)$ for $C_{\text{pr}}(\nu)$ and (15). \triangleleft

5 Conclusion

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 squared-modulus kernel $|K|^2$, and illustrated the parallel between the design of approximate measures for the approximation of such operators and the characterisation of low-MMD configurations for the kernel embedding of measures in \mathcal{G} . We also discussed the extent to which the considered MMD setting can be used

*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 be confused with the quadrature method for the numerical approximation of integral equations.

as a surrogate for the characterisation of projection-based approximations induced by subspaces defined by measures.

In contrast to the projection-based criteria C_{pr} and (16), the criterion D is convex and admits informative directional derivatives on any admissible convex set of measures; to this extent, the MMD criterion D may hence be regarded as a *differentiable relaxation* of the piecewise-constant criteria discussed in Section 4. Further, and as opposed to the criteria C_{pr} and (16), the numerical exploration of the MMD landscape does not require the characterisation of orthogonal projections onto subspaces of the underlying RKHS \mathcal{H} (i.e. the characterisation of the kernels of subspaces of \mathcal{H}), so that considering the MMD criterion D is of noticeable interest in terms of computational complexity (see for instance Remark 4.3).

References

- [1] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.
- [2] John B. Conway. *A course in functional analysis*, volume 96. Springer, 2019.
- [3] Felipe Cucker and Steve Smale. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39:1–49, 2002.
- [4] 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.
- [5] 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.
- [6] Bertrand Gauthier and Johan Suykens. Optimal quadrature-sparsification for integral operator approximation. *SIAM Journal on Scientific Computing*, 40:A3636–A3674, 2018.
- [7] 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.
- [8] Israel Gohberg, Seymour Goldberg, and Marius A. Kaashoek. *Classes of Linear Operators*, volume 63. Birkhäuser, 2013.
- [9] Tailen Hsing and Randall Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons, 2015.

- [10] Matthew Hutchings and Bertrand Gauthier. Local optimisation of Nyström samples through stochastic gradient descent. *To appear in Proceedings of the LOD 2022 Conference*, 2022.
- [11] Sanjiv Kumar, Mehryar Mohri, and Ameet Talwalkar. Sampling methods for the Nyström method. *Journal of Machine Learning Research*, 13:981–1006, 2012.
- [12] 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.
- [13] Karl-Hermann Neeb. On the geometry of standard subspaces. In *Contemporary Mathematics*, volume 714, pages 199–223. AMS, 2018.
- [14] Vern I. Paulsen and Mrinal Raghupathi. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. Cambridge University Press, 2016.
- [15] Lorenzo Rosasco, Mikhail Belkin, and Ernesto De Vito. On learning with integral operators. *Journal of Machine Learning Research*, 11:905–934, 2010.
- [16] Gabriele Santin and Robert Schaback. Approximation of eigenfunctions in kernel-based spaces. *Advances in Computational Mathematics*, 42:973–993, 2016.
- [17] Steve Smale and Ding-Xuan Zhou. Learning theory estimates via integral operators and their approximations. *Constructive Approximation*, 26:153–172, 2007.
- [18] Steve Smale and Ding-Xuan Zhou. Geometry on probability spaces. *Constructive Approximation*, 30:311–323, 2009.
- [19] 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.
- [20] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer, 2008.
- [21] 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.
- [22] 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.