



# Construction and Monte Carlo Estimation of Wavelet Frames Generated by a Reproducing Kernel

Ernesto De Vito<sup>1</sup> · Zeljko Kereta<sup>2</sup> · Valeriya Naumova<sup>3</sup> · Lorenzo Rosasco<sup>4,5,6</sup> · Stefano Vigogna<sup>4</sup>

Received: 30 November 2020 / Revised: 2 March 2021 / Accepted: 2 March 2021 /  
Published online: 16 April 2021  
© The Author(s) 2021

## Abstract

We introduce a construction of multiscale tight frames on general domains. The frame elements are obtained by spectral filtering of the integral operator associated with a reproducing kernel. Our construction extends classical wavelets as well as generalized wavelets on both continuous and discrete non-Euclidean structures such as Riemannian manifolds and weighted graphs. Moreover, it allows to study the relation between continuous and discrete frames in a random sampling regime, where discrete frames can be seen as Monte Carlo estimates of the continuous ones. Pairing spectral regularization with learning theory, we show that a sample frame tends to its population counterpart, and derive explicit finite-sample rates on spaces of Sobolev and Besov regularity. Our results prove the stability of frames constructed on empirical data, in the sense that all stochastic discretizations have the same underlying limit regardless of the set of initial training samples.

---

✉ Stefano Vigogna  
vigogna@dibris.unige.it

Ernesto De Vito  
devito@dim.unige.it

Zeljko Kereta  
zeljko@simula.no

Valeriya Naumova  
valeriya@simula.no

Lorenzo Rosasco  
lorenzo.rosasco@unige.it

- 1 MaLGA - DIMA, Università di Genova, Via Dodecaneso 35, 16146 Genova, Italy
- 2 Simula Research Laboratory, Martin Linges Vei 25, 1364 Oslo, Norway
- 3 Machine Intelligence Department, SimulaMet, Pilestredet 52, 0167 Oslo, Norway
- 4 MaLGA - DIBRIS, Università di Genova, Via Dodecaneso 35, 16146 Genova, Italy
- 5 Center for Brains, Minds and Machines, MIT, Cambridge, MA, USA
- 6 Istituto Italiano di Tecnologia, Genova, Italy

**Keywords** Wavelets · Frames · Reproducing kernel Hilbert spaces · Regularization · Learning theory

**Mathematics Subject Classification** 42C15 · 42C40 · 65T60 · 46E22 · 47A52 · 68T05

## 1 Introduction

Wavelet systems have long been employed in time-frequency analysis and approximation theory to break the uncertainty principle and resolve local singularities against global smoothness. Nonlinear approximation over redundant families of localized waveforms has enabled the construction of efficient sparse representations, becoming common practice in signal processing, source coding, noise reduction, and beyond. Sparse dictionaries are also an important tool in machine learning, where the extraction of few relevant features can significantly enhance a variety of learning tasks, making them scale with enormous quantities of data. However, the role of wavelets in machine learning is still unclear, and the impact they had in signal processing has, by far, not been matched. One objective constraint to a direct application of classical wavelet techniques to modern data science is of geometric nature: real data are typically high-dimensional and inherently structured, often featuring or hiding non-Euclidean topologies. On the other hand, a representation built on empirical samples poses an additional problem of stability, accounted for by how well it generalizes to future data. In this paper, expanding upon the ideas outlined in [35], we introduce a data-driven construction of wavelet frames on non-Euclidean domains, and provide stability results in high probability.

Starting from Haar's seminal work [31] and since the founding contributions of Grossmann and Morlet [30], a general theory of wavelet transforms and a wealth of specific families of wavelets have rapidly arisen [10,14,23,39,41], first and foremost on  $\mathbb{R}^d$ , but soon thereafter also on non-Euclidean structures such as manifolds and graphs [12,13,18,20,26,28,33,44]. Generalized wavelets usually consist of frames with some kind of broad to tighter link to ideas from multi-resolution analysis. At the very least, elements of a wavelet frame ought to be associated with locations and scales, decomposing signals into a sum of local features in increasing resolution. On a basic conceptual level, many of these generalized constructions stem from a reinterpretation of the frequency domain as the spectrum of a differential operator. Indeed, wavelets on  $\mathbb{R}$  are commonly generated by dilating and translating a well-localized function  $\psi$ ,

$$\psi_{a,b}(x) = |a|^{-1/2} \psi\left(\frac{x-b}{a}\right) \quad a \neq 0, b \in \mathbb{R};$$

but taking the Fourier transform, they can be rewritten as

$$\psi_{a,b}(x) = \int |a|^{1/2} \widehat{\psi}(a\xi) e^{2\pi i(x-b)\xi} d\xi = \int G_a(\xi) \overline{v_\xi(b)} v_\xi(x) d\xi, \quad (1)$$

with  $G_a(\xi) = |a|^{1/2} \widehat{\psi}(a\xi)$  and  $v_\xi(x) = e^{2\pi i x \xi}$ . This allows to reinterpret the wavelet  $\psi_{a,b}(x)$  as a superposition of Fourier harmonics  $v_\xi(x)$ , modulated by a spectral filter

$G_a(\xi)$ . Moreover, each  $v_\xi$  can be seen as an eigenfunction of the Laplacian  $\Delta = -d^2/dx^2$ . Hence, in principle, we may retrace an analogous construction whenever some notion of Laplacian is at hand. In particular, Riemannian manifolds and weighted graphs are examples of spaces where this is possible, using the Laplace–Beltrami operator or the graph Laplacian. A more detailed overview of related work based on these or similar ideas is postponed to Sects. 2 and 6.

Thus far, the study of generalized wavelets on non-Euclidean domains has primarily focused on either the continuous or the discrete setting. It is nonetheless natural to investigate the relationship between the two. For instance, regarding a graph as a sample of a manifold, we may ask whether and in what sense the frame built on the graph tends to the one on the manifold. In this paper we present a unified framework for the construction and the comparison of continuous and discrete frames. Returning for a moment to the real line, let us consider the semigroup  $e^{-t\Delta}$  generated by the Laplacian. This defines an integral operator

$$e^{-t\Delta} f(x) = \int K_t(x, y) f(y) dy,$$

with  $K_t(x, y)$  being the heat kernel. Such a representation suggests that the generalized Fourier analysis, already revisited as spectral analysis of the Laplacian, can now be translated in terms of a corresponding integral operator (see e.g. [13,38]). With the attention shifting from the Laplacian to an integral kernel, our idea is to recast the above constructions inside a reproducing kernel Hilbert space. Exploiting the reproducing kernel, we will extend a discrete frame out of the given samples, and thus compare it to its natural continuous counterpart.

Our construction yields empirical frames  $\widehat{\Psi}^N$  on sets of  $N$  data. We will show that  $\widehat{\Psi}^N$  converges in high probability to a continuous frame  $\Psi$  associated to a reproducing kernel Hilbert space  $\mathcal{H}$  as  $N \rightarrow \infty$ , thus providing a proof of its stability in an asymptotic sense. The empirical frames  $\widehat{\Psi}^N$  can be seen as Monte Carlo estimates of  $\Psi$ . Repeated random sampling will in fact produce a sequence of frames  $\widehat{\Psi}^N$  on an increasing chain of finite dimensional reproducing kernel Hilbert spaces  $\widehat{\mathcal{H}}_N$

$$\begin{matrix} \widehat{\mathcal{H}}_N \subset \widehat{\mathcal{H}}_{N+1} \subset \dots \subset \mathcal{H} \\ \widehat{\Psi}^N \quad \widehat{\Psi}^{N+1} \quad \longrightarrow \quad \Psi \end{matrix},$$

which approximates  $\Psi$  on  $\mathcal{H}$  up to a desired sampling resolution quantifiable by finite sample bounds in high probability.

One may also look at our result as a form of stochastic discretization of continuous frames. Going from the continuum to the discrete setting is an important problem in frame theory and applications of coherent states. Given a continuous frame of a Hilbert space, the discretization problem [2, Chapter 17] asks to extract a discrete frame out of it. Originally motivated by the need of numerical implementations of coherent states arising in quantum mechanics [15,51], the problem was then generalized to continuous frames [1] and addressed in several theoretical efforts [21,24,29], until it found a complete yet not constructive characterization in [22]. Sampling the continuous frame

is tantamount to sampling the parameter space on which the frame is indexed. For a wavelet frame, this means the selection of a discrete set of scales and locations. While the discretization of the scales can be readily obtained by a dyadic parametrization, the difficult part is usually sampling locations, that is, the domain where the frame is defined. How to do this is known in many cases and consists in an attentive selection of nets of well covering but sufficiently separated points. Already sensitive in the Euclidean setting, this procedure can be hard to generalize and implement in more general geometries [13]. In this respect, our Monte Carlo frame estimation provides a randomized approach to frame discretization as opposed to a deterministic sampling design. Clearly, our Monte Carlo estimate is not solving the discretization problem in its original form, since it defines frames only on finite dimensional subspaces. It is rather providing an asymptotic approximate solution, computing frames on an invading sequence of subspaces  $\hat{\mathcal{H}}_N \subset \mathcal{H}$ . We should also remark that, due to covering properties, standard frame discretization always entails a loosening of the frame bounds; hence, in particular, only non-tight frames may be sampled, even when the starting continuous frame is Parseval. As a result, signal reconstruction with respect to the discretized frame will in general require the computation of a dual frame, which is a problem on its own. On the contrary, in our randomized construction we preserve the tightness, albeit at the expense of a (possibly large) loss of resolution power  $\mathcal{H} \setminus \hat{\mathcal{H}}_N$ .

The remainder of the paper is organized as follows. The general notation used throughout the paper is listed in Table 1. In Sect. 2 we relate our main contribution to recent constructions of wavelets on graphs. This is both a special case and a main motivation of the general theory developed in the subsequent sections. In Sect. 3 we introduce the general framework and define the fundamental objects used in our analysis. The focus is on kernels, reproducing kernel Hilbert spaces, and associated integral operators. In Sect. 4 we present our frame construction based on spectral calculus of the integral operator. Our theory encompasses continuous and discrete frames within a unified formalism, paving the way for a principled comparison of the two. In particular, in Sect. 5, interpreting discrete locations as samples from a probability distribution we propose a Monte Carlo method for the estimation of continuous frames. In Sect. 6 we compare and contrast our approach to the existing literature. In Sect. 7 we prove the consistency of our Monte Carlo wavelets and obtain explicit convergence rates under Sobolev regularity of the signals. This is done combining techniques borrowed from the theory of spectral regularization with bounds of concentration of measure. In Sect. 8 we study the convergence rates in Besov spaces. In Sect. 9 we draw our conclusions and point at some directions for future work.

## 2 Wavelets on Graphs and Their Stability

In this section we discuss how the framework introduced in the paper may be used to study the stability of typical constructions of wavelets on graphs. We first recall a few elementary concepts about graphs and set up some notation. After that, we outline a natural construction of wavelets based on the graph Laplacian, and observe that such a construction may be recast in terms of a reproducing kernel. Finally, we explain how this allows to establish the stability of wavelet frames in a suitable random graph model.

**Table 1** Notation

Symbol	Definition	Symbol	Definition
$\langle \cdot, \cdot \rangle_{\mathcal{H}}, \ \cdot\ _{\mathcal{H}}$	Inner product and norm in a RKHS $\mathcal{H}$	$P_S$	Orthogonal projection onto a closed subspace $S$
$\sigma(A)$	Spectrum of a linear operator $A$	$\text{supp}(\rho)$	Support of a measure $\rho$
$\ \cdot\ $	Operator norm	$\langle \cdot, \cdot \rangle_{\rho}, \ \cdot\ _{\rho}$	Inner product and norm in $L^2(\mathcal{X}, \rho)$
$\ \cdot\ _{\text{HS}}$	Hilbert–Schmidt norm	$\delta_x$	Dirac measure at $x$
$v \otimes w$	The operator $u \in \mathcal{H} \mapsto \langle u, v \rangle_{\mathcal{H}} w \in \mathcal{H}$	$\mathbf{v}[i]$	$i$ th component of a vector $\mathbf{v}$
$\text{span } S$	Linear span of a set $S$	$\mathbf{M}[i, j]$	$(i, j)$ th entry of a matrix $\mathbf{M}$
$S^{\perp}$	Orthogonal complement of a set $S$	$\mathbf{M}^+$	Pseudoinverse of a matrix $\mathbf{M}$
$\bar{S}$	Topological closure of a set $S$	$X \lesssim Y$	$X \leq CY$ for some constant $C > 0$
$S_1 \oplus S_2$	Direct sum of two subspaces $S_1$ and $S_2$	$X \asymp Y$	$X \lesssim Y$ and $Y \lesssim X$

### 2.1 Wavelets on Graphs

We start with some basics of spectral graph theory. We only review what is strictly necessary for our purposes, and refer to [11] for further details.

**Definition 2.1** (*Weighted graph*) An undirected graph is a pair  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is a finite discrete set of vertices  $\mathcal{V} := \{x_1, \dots, x_N\}$ , and  $\mathcal{E}$  is a set of unordered pairs  $\mathcal{E} \subset \{\{x_i, x_k\} : x_i, x_k \in \mathcal{V}\}$ , called edges. A weighted (undirected) graph is an undirected graph with an associated weight function  $w : \mathcal{E} \rightarrow (0, +\infty)$ .

Arguably, one of the most remarkable facts about graphs is that it is possible to define on such a minimal structure a consistent notion of Laplacian. Functions on the graph, more precisely functions  $f : \mathcal{V} \rightarrow \mathbb{R}$ , can be identified with vectors  $\mathbf{f} \in \mathbb{R}^N$  by  $\mathbf{f}_i := f(x_i)$ , and equipped with the standard inner product  $\mathbf{f}^{\top} \mathbf{g}$  for  $\mathbf{f}, \mathbf{g} \in \mathbb{R}^N$ . As an operator acting on functions, the graph Laplacian is thus defined by a matrix  $\mathbf{L} \in \mathbb{R}^{N \times N}$ .

**Definition 2.2** (*Graph Laplacian*) Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$  be a weighted graph. The weight matrix  $\mathbf{W} := [w_{i,k}]_{i,k=1}^N$  is defined by  $w_{i,k} := w(\{x_i, x_k\})$  for  $\{x_i, x_k\} \in \mathcal{E}$ , and  $w_{i,k} := 0$  otherwise. The degree matrix  $\mathbf{D} := \text{diag}(d_1, \dots, d_N)$  is defined by  $d_i := \sum_{k=1}^N w_{i,k}$ . The unnormalized graph Laplacian is the matrix

$$\mathbf{L} := \mathbf{D} - \mathbf{W}.$$

Assuming that  $\mathcal{G}$  is connected, hence  $d_i > 0$  for all  $i = 1, \dots, N$ , the symmetric normalized graph Laplacian is  $\mathbf{L}' := \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}$ . Several other variants are considered in the literature, including the random walk normalized graph Laplacian  $\mathbf{D}^{-1} \mathbf{L} = \mathbf{I} - \mathbf{D}^{-1} \mathbf{W}$ , which is not symmetric but conjugate to  $\mathbf{L}'$ . The

operators  $\mathbf{L}, \mathbf{L}'$  and further normalizations result from different definitions of Hilbert structures on the spaces of functions on  $\mathcal{V}$  and  $\mathcal{E}$  [34]. While each operator gives rise to a different analysis, the choice of one or the other does not have formal consequences in our construction, hence, for simplicity, we will generically use  $\mathbf{L}$ .

The matrix  $\mathbf{L}$  is positive semi-definite, hence it admits an orthonormal basis of eigenvectors with non-negative eigenvalues, customarily sorted in increasing order:

$$\mathbf{L}\mathbf{u}_i = \xi_i \mathbf{u}_i, \quad i = 0, \dots, N - 1, \quad 0 = \xi_0 \leq \xi_1 \leq \dots \leq \xi_{N-1}.$$

The spectrum of  $\mathbf{L}$  reveals several important topological properties of the graph. In particular, a graph has as many connected components as zero eigenvalues, with eigenfunctions being piecewise constant on the components. We assume from now on that the graph is connected, hence  $\xi_1 > 0$ .

The graph Laplacian can be seen as a discrete analog of the continuous Laplace operator. This analogy justifies the interpretation of the eigenvectors  $\mathbf{u}_i$  as Fourier harmonics, and the corresponding eigenvalues  $\xi_i$  as frequencies. Accordingly, the *graph Fourier transform* is defined by

$$\mathbf{F} := [\mathbf{u}_1 \cdots \mathbf{u}_N]^\top, \quad [\mathbf{F}\mathbf{f}]_i := \mathbf{u}_i^\top \mathbf{f}.$$

Note that the indexing is hiding that  $\mathbf{F}\mathbf{f}$  should be thought as a function on the frequencies  $\xi_i$ . Carrying the analogy forward, a family of *graph wavelets* can be constructed by spectral filtering of the Fourier basis as follows. Let  $\{H_j\}_{j \geq 0}$  be a family of functions  $H_j : [0, +\infty) \rightarrow [0, +\infty)$  satisfying

$$\begin{aligned} \sum_{j \geq 0} H_j(\xi)^2 &= 1 \quad \text{for all } \xi \in [0, +\infty), \\ \#\{H_j : H_j(\xi_i) \neq 0\} &< \infty \quad \text{for } i = 1, \dots, N. \end{aligned}$$

Then, the family

$$\varphi_{j,k} := \sum_{i=1}^N H_j(\xi_i) \mathbf{u}_i[k] \mathbf{u}_i \quad j \geq 0, k = 1, \dots, N \tag{2}$$

defines a Parseval frame on  $\mathcal{G}$  [28, Theorem 2].

Let  $\mathcal{H}_{\mathcal{G}} := \text{span}\{\mathbf{u}_0\}^\perp = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{N-1}\}$  the space of all non-constant signals on  $\mathcal{G}$ . The graph Laplacian defines an inner product on  $\mathcal{H}_{\mathcal{G}}$  by  $\langle \mathbf{f}, \mathbf{g} \rangle_{\mathcal{G}} := \mathbf{f}^\top \mathbf{L}\mathbf{g}$ , which is invariant under graph isomorphisms. The Hilbert space  $\mathcal{H}_{\mathcal{G}}$  has *reproducing kernel*

$$\mathbf{K} := \mathbf{L}^+.$$

The matrix  $\mathbf{K}$  on  $\mathcal{H}_{\mathcal{G}}$  has same eigenvectors  $\mathbf{u}_1, \dots, \mathbf{u}_{N-1}$  as  $\mathbf{L}$ , and eigenvalues

$$\lambda_1 = \xi_1^{-1} \geq \lambda_2 = \xi_2^{-1} \geq \dots \geq \lambda_{N-1} = \xi_{N-1}^{-1}.$$

Therefore, wavelets (2) can be as well defined starting from the spectral decomposition of the reproducing kernel  $\mathbf{K}$ , rather than the Laplacian  $\mathbf{L}$ . Conversely, given any reproducing kernel  $\mathbf{K}$ , a frame may be constructed, without any reference to a Laplacian matrix. Indeed, this is the point of view taken in this paper.

Besides the equivalence in defining the frame, starting from a kernel implies some technical differences, but also opens to new theoretical potential. First, note that the spectrum gets flipped, hence the eigenvalues of the kernel should be thought as inverses of Fourier frequencies. This seemingly irrelevant remark is actually important to correctly interpret the definitions of Sobolev and Besov spaces given in Sect. 8. Moreover, in light of this, the scale  $\tau$  in (27) can be understood as a frequency threshold, and the regularization  $\tau^{-1}$  in the regression problem (29) as keeping the low frequencies. Reasoning in reproducing kernel Hilbert spaces also suggests further definitions of filtering beyond typical band-pass of Example 4.5, employing regularization techniques from inverse problems, as exemplified in Table 2. Lastly, reproducing kernels naturally extend the wavelet functions out of the graph vertices, making possible to analyze the stability of the graph wavelet frame for different random realizations of the graph. We elaborate on this in the next section.

## 2.2 Stability of Wavelets on Random Graphs

By virtue of their generality, graphs can be used to model a variety of discrete objects with pairwise relations, as well as to approximate complex geometries in continuous domains. In both cases, complexity and uncertainty are often handled by assuming an underlying random model and studying statistics and asymptotic behavior of relevant variables. In particular, neighborhood graphs are often used to approximate the Riemannian structure of a manifold. In a neighborhood graph, vertices are sampled at random from the manifold, and edges are drawn connecting vertices in suitable neighborhoods, such as  $k$ -nearest neighborhoods or  $\epsilon$ -radius balls in the ambient Euclidean distance, or even putting weights using a global (possibly truncated) kernel function.

The convergence of the graph Laplacian to the Laplace–Beltrami operator has been studied and quantified in several settings, both as a pointwise [4,27,34,53,55] and as a spectral limit [3,25,37,46,54]. On the other hand, wavelets have been generalized to continuous non-Euclidean domains, notably Riemannian manifolds and spaces of homogenous type [13,20,26], and while the conceptual ingredients remain similar, the convergence of graph to manifold wavelets is hardly studied. We next describe how our theory provides a way to fill this gap.

Suppose we have a graph  $\mathcal{G}$  with vertices  $\{x_1, \dots, x_N\}$  and a positive definite kernel matrix  $\widehat{\mathbf{K}}$ . For instance, the matrix  $N\widehat{\mathbf{K}}$  may be the kernel associated with the graph Laplacian. Computing the eigenvalues  $\widehat{\lambda}_i$  and eigenvectors  $\widehat{\mathbf{u}}_i$  of  $\widehat{\mathbf{K}}$ , we can define, in analogy with (2), the family

$$\widehat{\varphi}_{j,k} := \sum_{i=1}^N F_j(\widehat{\lambda}_i) \widehat{\mathbf{u}}_i[k] \widehat{\mathbf{u}}_i \quad j \geq 0, k = 1, \dots, N, \quad (3)$$

for a suitable spectral filter  $F_j(\lambda)$ . By Proposition 4.7, (3) defines a Parseval frame on  $\mathcal{G}$ . Now, suppose that the vertices of our graph are sampled from a space  $\mathcal{X}$  with probability distribution  $\rho$  and reproducing kernel  $K$  satisfying the assumptions of Sects. 3 and 4. Furthermore, suppose that the kernel matrix  $\widehat{\mathbf{K}}$  is given by

$$\widehat{\mathbf{K}}[i, k] = N^{-1} K(x_i, x_k).$$

For example, the space  $\mathcal{X}$  may be a compact Riemannian manifold, in which case we could consider the heat kernel associated with the Laplace–Beltrami operator, and regard the kernel matrix as a discretization of the integral operator. As a discrete example, one may also think of  $\mathcal{X}$  as a supergraph of  $\mathcal{G}$ . Thanks to Proposition 4.7, the family of Monte Carlo wavelets

$$\widehat{\psi}_{j,k}(x) := \sum_i G_j(\widehat{\lambda}_i) \widehat{v}_i(x_k) \widehat{v}_i(x) \quad j \geq 0, k = 1, \dots, N$$

is a Parseval frame isomorphic to (3). Crucially, in this new representation, the frame functions are well-defined both on and off the graph  $\mathcal{G}$ , and thus the convergence of the frame can be studied on a test signal  $f : \mathcal{X} \rightarrow \mathbb{R}$ , as discussed in Sect. 7. The stability of the graph wavelets (3) can therefore be established by an application of Theorem 7.5 or 8.8.

Starting from the next section, we develop our theory in greater generality, but always bearing in mind the motivating setting just discussed.

### 3 Preliminaries

In this section we prepare the technical ground on which our results will built (see also [46]). Let  $\mathcal{X}$  be a locally compact, second countable topological space endowed with a Borel probability measure  $\rho$ . Given a continuous, positive semi-definite kernel

$$K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{C},$$

we denote the associated reproducing kernel Hilbert space (RKHS) by

$$\mathcal{H} := \overline{\text{span}}\{K_x : x \in \mathcal{X}\},$$

where  $K_x := K(\cdot, x) \in \mathcal{H}$ , and the closure is taken with respect to the inner product  $\langle K_x, K_y \rangle_{\mathcal{H}} := K(y, x)$ . Elements of  $\mathcal{H}$  are continuous functions satisfying the following reproducing property:

$$f(x) = \langle f, K_x \rangle_{\mathcal{H}} \quad \text{for all } f \in \mathcal{H}. \tag{4}$$



The space  $\mathcal{H}$  is separable, since  $\mathcal{X}$  is separable. We further assume  $K$  is bounded on  $\mathcal{X}$  and denote

$$\kappa := \sup_{x \in \mathcal{X}} \sqrt{K(x, x)} = \sup_{x \in \mathcal{X}} \|K_x\|_{\mathcal{H}} < \infty,$$

which implies that  $\mathcal{H}$  is continuously embedded into the space of bounded continuous functions on  $\mathcal{X}$ .

We define the (non-centered) covariance operator  $T : \mathcal{H} \rightarrow \mathcal{H}$  by

$$T := \int_{\mathcal{X}} K_x \otimes K_x d\rho(x), \tag{5}$$

where the integral converges strongly. The operator  $T$  is positive and trace-class (therefore compact) with  $\sigma(T) \subset [0, \kappa^2]$ . Hence, the spectral theorem ensures the existence of a countable orthonormal set  $\{v_i\}_{i \in \mathcal{I}_\rho \cup \mathcal{I}_0} \subset \mathcal{H}$  and a sequence  $(\lambda_i)_{i \in \mathcal{I}_\rho} \subset (0, \kappa^2]$  such that

$$Tv_i = \begin{cases} \lambda_i v_i & i \in \mathcal{I}_\rho \\ 0 & i \in \mathcal{I}_0 \end{cases}.$$

Let  $L^2(\mathcal{X}, \rho)$  be the space of square-integrable functions on  $\mathcal{X}$  with respect to the measure  $\rho$ , and denote  $\mathcal{X}_\rho := \text{supp}(\rho)$ . We define the integral operator  $L_K : L^2(\mathcal{X}, \rho) \rightarrow L^2(\mathcal{X}, \rho)$  by

$$L_K F(x) := \int_{\mathcal{X}} K(x, y)F(y) d\rho(y).$$

The spaces  $\mathcal{H}$  and  $L^2(\mathcal{X}, \rho)$  and the operators  $T$  and  $L_K$  are related through the inclusion operator  $S : \mathcal{H} \rightarrow L^2(\mathcal{X}, \rho)$  defined by

$$Sf(x) := \langle f, K_x \rangle_{\mathcal{H}}.$$

The adjoint operator  $S^* : L^2(\mathcal{X}, \rho) \rightarrow \mathcal{H}$  acts as the strongly converging integral

$$S^*F = \int_{\mathcal{X}} F(x)K_x d\rho(x).$$

We have  $T = S^*S$  and  $L_K = SS^*$ . Hence,  $\sigma(T) \setminus \{0\} = \sigma(L_K) \setminus \{0\}$ , and the eigenfunctions  $\{u_i\}_{i \in \mathcal{I}_\rho \cup \mathcal{I}_0} \subset L^2(\mathcal{X}, \rho)$  of  $L_K$  satisfy

$$Su_i = \begin{cases} \sqrt{\lambda_i} u_i & i \in \mathcal{I}_\rho \\ 0 & i \in \mathcal{I}_0 \end{cases}. \tag{6}$$

Mercer’s theorem gives

$$\begin{aligned}
 K(x, y) &= \sum_{i \in \mathcal{I}_\rho \cup \mathcal{I}_0} \overline{v_i(x)} v_i(y) \quad \text{for } x, y \in \mathcal{X}, \\
 K(x, y) &= \sum_{i \in \mathcal{I}_\rho} \lambda_i \overline{u_i(x)} u_i(y) \quad \text{for } x, y \in \mathcal{X}_\rho,
 \end{aligned}
 \tag{7}$$

where the series converge absolutely and uniformly on compact subsets.

Defining

$$\mathcal{H}_\rho := \overline{\text{span}}\{K_x : x \in \mathcal{X}_\rho\} = \overline{\text{span}}\{v_i : i \in \mathcal{I}_\rho\},$$

where the closure is taken in  $\mathcal{H}$ , we can identify  $\mathcal{H}_\rho$  as a (non-closed) subspace of  $L^2(\mathcal{X}, \rho)$ . The closure of  $\mathcal{H}_\rho$  in  $L^2(\mathcal{X}, \rho)$  is

$$\overline{\mathcal{H}}_\rho := \overline{\text{span}}\{u_i : i \in \mathcal{I}_\rho\},$$

and the following decompositions hold true:

$$\mathcal{H} = \mathcal{H}_\rho \oplus \ker S, \quad L^2(\mathcal{X}, \rho) = \overline{\mathcal{H}}_\rho \oplus \ker S^*.$$

For  $f \in \mathcal{H}_\rho$ , we can relate the norms in  $\mathcal{H}$  and  $L^2(\mathcal{X}, \rho)$  as

$$\|f\|_\rho = \|\sqrt{T}f\|_{\mathcal{H}}.
 \tag{8}$$

In other words,  $\sqrt{T}$  induces an isometric isomorphism between  $\overline{\mathcal{H}}_\rho$  and  $\mathcal{H}_\rho$ . We define the partial isometry  $U : \mathcal{H} \rightarrow L^2(\mathcal{X}, \rho)$ , such that  $U\mathcal{H}_\rho = \overline{\mathcal{H}}_\rho$ , by

$$Uf = \sum_{i \in \mathcal{I}_\rho} \langle f, v_i \rangle_{\mathcal{H}} u_i.$$

As examples of this setting, we may think of  $\mathcal{X}$  as  $\mathbb{R}^d$ , or a non-Euclidean domain such as a compact connected Riemannian manifold or a weighted graph. In these cases, we can take  $K$  as the heat kernel associated with the proper notion of Laplacian, be it the Laplace–Beltrami operator or the graph Laplacian.

### 4 Wavelet Frames by Reproducing Kernels

We now build Parseval frames in the RKHS  $\mathcal{H}$  and in  $L^2(\mathcal{X}, \rho)$ . Our construction is centered around eigenfunctions of the integral operator (5) and filters on the corresponding eigenvalues. Continuous frames emerged in the mathematical physics community from the study of coherent states, as a generalization of the more common notion of a discrete frame [2,23].

**Definition 4.1** (*Frame*) Let  $\mathcal{H}$  be a Hilbert space,  $\mathcal{A}$  a locally compact space and  $\mu$  a Radon measure on  $\mathcal{A}$  with  $\text{supp } \mu = \mathcal{A}$ . A family  $\Psi = \{\psi_a : a \in \mathcal{A}\} \subset \mathcal{H}$  is called a *frame* for  $\mathcal{H}$  if there exist constants  $0 < A \leq B < \infty$  such that, for every  $f \in \mathcal{H}$ , we have

$$A \|f\|_{\mathcal{H}}^2 \leq \int_{\mathcal{A}} |\langle f, \psi_a \rangle_{\mathcal{H}}|^2 d\mu(a) \leq B \|f\|_{\mathcal{H}}^2.$$

We say that  $\Psi$  is *tight* if  $A = B$ , and *Parseval* if  $A = B = 1$ .

In the above definition it is implicitly assumed that the map  $a \mapsto \langle \Psi_a, f \rangle_{\mathcal{H}}$  is measurable for all  $f \in \mathcal{H}$ . It is important to note that this definition depends on the choice of the measure  $\mu$ . In the case of a counting measure, we recover the standard definition of discrete frame.

### 4.1 Filters

To construct our wavelet frames, we first need to define filters, *i.e.* functions acting on the spectrum of  $T$  that satisfy a partition of unity condition.

**Definition 4.2** (*Filters*) A family  $\{G_j\}_{j \geq 0}$  of measurable functions  $G_j : [0, +\infty) \rightarrow [0, +\infty)$  such that

$$\lambda \sum_{j \geq 0} G_j(\lambda)^2 = 1 \quad \text{for all } \lambda \in (0, \kappa^2] \tag{9}$$

is called a family of *filters*.

By the spectral theorem,  $G_j(T)$  is a (possibly unbounded) positive operator on  $\mathcal{H}$  such that  $\sigma(G_j(T)) = G_j(\sigma(T))$ , with domain of definition

$$\mathcal{D}_j := \left\{ f \in \mathcal{H} : \sum_{i \in \mathcal{I}_\rho \cup \mathcal{I}_0} G_j(\lambda_i)^2 |\langle f, v_i \rangle_{\mathcal{H}}|^2 < \infty \right\}.$$

It follows that

$$\mathcal{D} := \text{span}\{v_i : i \in \mathcal{I}_\rho \cup \mathcal{I}_0\} \subset \mathcal{D}_j \quad \text{for all } j \geq 0,$$

and

$$G_j(T)v_i = \begin{cases} G_j(\lambda_i)v_i, & i \in \mathcal{I}_\rho \\ G_j(0)v_i, & i \in \mathcal{I}_0 \end{cases}.$$

An easy way to define filters is by differences of suitable spectral functions.

**Definition 4.3** (*Spectral functions*) A family  $\{g_j\}_{j \geq 0}$  of measurable functions  $g_j : [0, \infty) \rightarrow [0, \infty)$  satisfying

$$0 \leq g_j \leq g_{j+1}, \quad \lim_{j \rightarrow \infty} \lambda g_j(\lambda) = 1 \quad \text{for all } \lambda \in (0, \kappa^2] \tag{10}$$

is called a family of *spectral functions*.

Given a family of spectral functions  $\{g_j\}_{j \geq 0}$ , filters  $\{G_j\}_{j \geq 0}$  can be obtained setting

$$G_0(\lambda) := \sqrt{g_0(\lambda)}, \quad G_{j+1}(\lambda) := \sqrt{g_{j+1}(\lambda) - g_j(\lambda)} \quad \text{for } j \geq 0. \tag{11}$$

The filters thus defined give rise to a telescopic sum:

$$\sum_{j \leq \tau} G_j(\lambda)^2 = g_\tau(\lambda). \tag{12}$$

Taking the limit for  $\tau \rightarrow \infty$ , condition (9) is satisfied thanks to (10). Conversely, starting from a family of filters  $\{G_j\}_{j \geq 0}$ , we can define spectral functions  $\{g_j\}_{j \geq 0}$  by

$$g_j(\lambda) := \sum_{\ell \leq j} G_\ell(\lambda)^2 \quad \text{for } j \geq 0,$$

which enjoys (10) due to (9). Therefore, the notion of filter and that of spectral function are equivalent, and we will refer to them interchangeably.

The definition in (11) allows to find a wealth of filters by tapping into regularization theory [19]. In the forthcoming analysis, we will use the following notion of qualification.

**Definition 4.4** (*Qualification*) The *qualification* of a spectral function  $g_j : [0, \infty) \rightarrow [0, \infty)$  is the maximum constant  $\nu \in (0, \infty]$  such that

$$\sup_{\lambda \in (0, \kappa^2]} \lambda^\nu |1 - \lambda g_j(\lambda)| \leq C_\nu j^{-\nu} \quad \text{for all } j \geq 0,$$

where the constant  $C_\nu$  does not depend on  $j$ .

In the theory of regularization of ill-posed inverse problems [19], the qualification represents the limit within which a regularizer may exploit the regularity of the true solution. In particular, methods with finite qualification suffer from the so-called *satur-  
ation effect*.

Some standard examples of spectral functions, together with their qualifications, are listed in Table 2.

Additional examples of admissible filters widely used in the construction of wavelet frames (see e.g. [13,20]) are given by the following:

**Table 2** Spectral regularizers and their qualifications. Landweber iteration and Nesterov acceleration require  $\gamma < 1/\kappa^2$  and  $\beta \geq 1$ . In heavy ball,  $\alpha_j, \beta_j$  are suitably selected sequences depending on  $\nu$ , where  $\nu$  is any positive real (see [43])

Method	$g_j(\lambda)$	Qualification
Tikhonov regularization	$\frac{1}{\lambda + 1/j}$	1
Iterated Tikhonov ( $m$ iterations)	$\frac{(\lambda + 1/j)^m - (1/j)^m}{\lambda(\lambda + 1/j)^m}$	$m$
Landweber iteration	$\frac{1}{\lambda}(1 - (1 - \gamma\lambda)^j)$	$\infty$
Asymptotic regularization	$\frac{1}{\lambda}(1 - \exp(-j\lambda))$	$\infty$
Heavy ball ( $\nu$ -method)	$(1 - \alpha_j\lambda + \beta_j)g_{j-1}(\lambda) - \beta_j g_{j-2}(\lambda) + \alpha_j$	$\nu$
Nesterov acceleration	$(1 - \gamma\lambda)(g_{j-1}(\lambda) + \frac{j-2}{j-1+\beta}(g_{j-1}(\lambda) - g_{j-2}(\lambda))) + \gamma$	$\nu \geq 1/2$

**Example 4.5** (Localized filters) Let  $g \in C^\infty([0, \infty))$  such that  $\text{supp}(g) \subset (2^{-1}, \infty)$ ,  $0 \leq g \leq 1$ , and  $g(\lambda) = 1$  for all  $\lambda \geq 1$ . Define

$$\lambda g_j(\lambda) := g(2^j \lambda).$$

Then the family  $\{g_j\}_{j \geq 0}$  satisfies the properties (10). Furthermore, the corresponding filters (11) are localized, meaning that, defining  $F_j(\lambda) := \sqrt{\lambda}G_j(\lambda)$ , we have

$$\text{supp}(F_0) \subset (2^{-1}, \infty), \quad \text{supp}(F_j) \subset (2^{-j-1}, 2^{-j+1}) \quad \text{for } j \geq 1.$$

### 4.2 Frames

We are now ready to define our wavelet frames. We first form frame elements in  $\mathcal{H}$ , and then use the partial isometry  $U : \mathcal{H} \rightarrow L^2(\mathcal{X}, \rho)$  to obtain frames in  $L^2(\mathcal{X}, \rho)$ .

**Definition 4.6** (*Wavelets*) Let  $\{G_j\}_{j \geq 0}$  be a family of filters as in Definition 4.2, and assume

$$K_x \in \mathcal{D}_j \quad \text{for all } j \geq 0 \text{ and almost every } x \in \mathcal{X}_\rho. \tag{13}$$

We define the families of *wavelets*

$$\Psi := \{\psi_{j,x} : j \geq 0, x \in \mathcal{X}_\rho\} \subset \mathcal{H}, \quad \Phi := \{\varphi_{j,x} : j \geq 0, x \in \mathcal{X}_\rho\} \subset L^2(\mathcal{X}, \rho),$$

where

$$\psi_{j,x} := G_j(\mathbb{T})K_x, \quad \varphi_{j,x} := UG_j(\mathbb{T})K_x \quad \text{for } j \geq 0 \text{ and } x \in \mathcal{X}_\rho. \tag{14}$$

Observe that, since  $\psi_{j,x}$  and  $\varphi_{j,x}$  are defined for  $x \in \mathcal{X}_\rho$ , we actually have  $\Psi \subset \mathcal{H}_\rho \subset \mathcal{H}$ , and  $\Phi \subset \overline{\mathcal{H}}_\rho \subset L^2(\mathcal{X}, \rho)$ . In particular, the orthogonality of  $\mathcal{H}_\rho$  and  $\ker S$  entails

$\langle K_x, G_j(T)v_i \rangle_{\mathcal{H}} = 0$  for all  $i \in \mathcal{I}_0$ . By the reproducing property (4), condition (13) is thus equivalent to

$$\sum_{i \in \mathcal{I}_\rho} G_j(\lambda_i)^2 |v_i(x)|^2 < \infty \quad \text{for all } j \geq 0 \text{ and almost every } x \in \mathcal{X}_\rho.$$

If  $G_j$  is a bounded function, then  $G_j(T)$  is a bounded operator, hence  $\mathcal{D}_j = \mathcal{H}$ . In this case, which includes the spectral functions listed in Table 2, condition (13) is trivially satisfied.

Using the spectral decomposition of  $G_j(T)$  and the reproducing property, we obtain

$$\psi_{j,x}(y) = \sum_{i \in \mathcal{I}_\rho} G_j(\lambda_i) \overline{v_i(x)} v_i(y), \quad \varphi_{j,x}(y) = \sum_{i \in \mathcal{I}_\rho} \sqrt{\lambda_i} G_j(\lambda_i) \overline{u_i(x)} u_i(y). \quad (15)$$

These expressions allow to interpret  $\Psi$  and  $\Phi$  as families of wavelets, in the sense of (1). We interpret  $x$  as the location and  $j$  as the scale parameter; the functions  $K_x$  localize the signal in space, whereas the filters  $G_j$  regularize or localize in frequency. Note also the analogy with (7), in the light of which (15) may be seen as a filtered Mercer representation.

With the following proposition we show that (14) defines Parseval frames.

**Proposition 4.7** *Assume the setting in Sect. 3, and let  $\Psi, \Phi$  be defined as in Definition 4.6. Then, for every  $f \in \mathcal{H}$  we have*

$$\sum_{j \geq 0} \int_{\mathcal{X}} |\langle f, \psi_{j,x} \rangle_{\mathcal{H}}|^2 d\rho(x) = \|\mathbf{P}_{\mathcal{H}_\rho} f\|_{\mathcal{H}}^2, \quad (16)$$

and for any  $F \in L^2(\mathcal{X}, \rho)$  we have

$$\sum_{j \geq 0} \int_{\mathcal{X}} |\langle F, \varphi_{j,x} \rangle_{\rho}|^2 d\rho(x) = \|\mathbf{P}_{\overline{\mathcal{H}}_\rho} F\|_{\rho}^2. \quad (17)$$

**Proof** The equality (17) follows from (16) and the fact that  $U$  is unitary from  $\mathcal{H}_\rho$  to  $\overline{\mathcal{H}}_\rho$ . To establish (16), in view of Lemma A.1 it suffices to consider functions in the dense subspace  $\mathcal{D} \subset \mathcal{H}$ . Thus, let  $f \in \mathcal{D}$ . Since  $G_j(T)$  is self-adjoint on  $\mathcal{D}_j$ , and  $\mathcal{D} \subset \mathcal{D}_j$  for all  $j$ , we have

$$\langle f, \psi_{j,x} \rangle_{\mathcal{H}} = \langle f, G_j(T)K_x \rangle_{\mathcal{H}} = \langle G_j(T)f, K_x \rangle_{\mathcal{H}},$$

which integrated over  $x \in \mathcal{X}$  gives

$$\int_{\mathcal{X}} |\langle f, \psi_{j,x} \rangle_{\mathcal{H}}|^2 d\rho(x) = \langle \mathbf{T}G_j(T)f, G_j(T)f \rangle_{\mathcal{H}} = \langle \mathbf{T}G_j(T)^2 f, f \rangle_{\mathcal{H}}. \quad (18)$$

Summing over  $j \geq 0$  and using (9), we therefore obtain

$$\begin{aligned} \sum_{j \geq 0} \left( \text{T}G_j(\text{T})^2 f, f \right)_{\mathcal{H}} &= \sum_{i \in \mathcal{I}_\rho} \left( |\langle f, v_i \rangle_{\mathcal{H}}|^2 \sum_{j \geq 0} \lambda_i G_j(\lambda_i)^2 \right) \\ &= \sum_{i \in \mathcal{I}_\rho} |\langle f, v_i \rangle_{\mathcal{H}}|^2 = \|\text{P}_{\mathcal{H}_\rho} f\|_{\mathcal{H}}^2. \end{aligned}$$

□

The frame property can also be expressed as a resolution of the identity. Such a formulation will be particularly useful in Sect. 7.

**Proposition 4.8** *Under the assumptions of Proposition 4.7, there exists a positive bounded operator  $\text{T}_j : \mathcal{H} \rightarrow \mathcal{H}$  such that*

$$\text{T}_j = \int_{\mathcal{X}} \psi_{j,x} \otimes \psi_{j,x} \, d\rho(x), \tag{19}$$

where the integral converges weakly. Furthermore,

$$\text{T}_j = \text{T}G_j(\text{T})^2, \tag{20}$$

$$\sum_{j \leq \tau} \text{T}_j = \text{T}g_\tau(\text{T}), \tag{21}$$

and the following resolution of the identity holds true:

$$\text{P}_{\mathcal{H}_\rho} = \sum_{j \geq 0} \text{T}_j. \tag{22}$$

**Proof** From (18) we have, for all  $f \in \mathcal{D}$ ,

$$\int_{\mathcal{X}} |\langle f, \psi_{j,x} \rangle_{\mathcal{H}}|^2 \, d\rho(x) \leq \|\text{T}G_j(\text{T})^2\| \|f\|_{\mathcal{H}}^2,$$

where  $\text{T}G_j(\text{T})^2$  is bounded since  $\lambda G_j(\lambda)^2 \leq 1$  by (9). Hence, thanks to Lemma A.1, there exists a positive bounded operator  $\text{T}_j$  as in (19). Moreover, (18) implies (20) by the density of  $\mathcal{D}$ . The equality (21) follows from (20) and (12). Lastly, (22) is a reformulation of (16). □

Depending on the choice of the measure  $\rho$ , Proposition 4.7 gives the frame property for either a continuous or a discrete setting. Namely, consider a discrete set  $\{x_1, \dots, x_N\}$ , and let

$$\widehat{\rho}_N := \frac{1}{N} \sum_{k=1}^N \delta_{x_k}.$$

With the choice of the discrete measure  $\widehat{\rho}_N$ , (5) defines the discrete (non-centered) covariance operator  $\widehat{T} : \mathcal{H} \rightarrow \mathcal{H}$  by

$$\widehat{T} := \frac{1}{N} \sum_{k=1}^N K_{x_k} \otimes K_{x_k}.$$

Furthermore, Definition 4.6 produces the family of wavelets

$$\widehat{\psi}_{j,k} := G_j(\widehat{T})K_{x_k} \quad \text{for } j \geq 0 \text{ and } k = 1, \dots, N,$$

which, by Proposition 4.7, constitutes a discrete Parseval frame on

$$\widehat{\mathcal{H}}_N := \mathcal{H}_{\widehat{\rho}_N} = \text{span}\{K_{x_k} : k = 1, \dots, N\} \simeq \mathbb{C}^N.$$

In Sect. 5 we will make reference to this construction to define Monte Carlo wavelets, where the points  $x_1, \dots, x_N$  are drawn at random from  $\mathcal{X}_\rho$ .

### 4.3 Two Generalizations

We discuss here two generalizations of the framework presented in Sect. 4.2. First, one may readily consider more general scale parameterizations. Namely, let  $\Omega$  be a locally compact, second countable topological space, endowed with a measure  $\mu$  defined on the Borel  $\sigma$ -algebra of  $\Omega$ , finite on compact subsets, and such that  $\text{supp } \mu = \Omega$ . Adjusting the definitions accordingly, such as replacing the sums over all non-negative integers  $j$  in (9) and (16) with integrals over  $\Omega$  with respect to  $\mu$ , the proof of Proposition 4.7 follows along the same steps. In this context, Definition 4.2 can be seen as a special case where  $\Omega$  is countable and  $\mu$  is the counting measure. Second, the assumption that the kernel  $K$  is bounded, implying that  $L_K$  admits an orthonormal basis of eigenvectors, is not necessary for our construction of Parseval frames. Indeed, it is enough to assume that

$$\int_{\mathcal{X}} |f(x)|^2 d\rho(x) < +\infty \quad \text{for all } f \in \mathcal{H}.$$

This implies that  $\mathcal{H}$  is a subspace of  $L^2(\mathcal{X}; \rho)$  and the inclusion operator  $S$  is bounded. The integral (5) converges now in the weak operator topology, and the covariance operator  $T$  is positive and bounded. Thus, the Riesz–Markov theorem entails that, for all  $f \in \mathcal{H}$ , there is a unique finite measure  $\nu_f$  on  $[0, +\infty)$  such that  $\nu_f([0, +\infty)) = \|f\|_{\mathcal{H}}^2$  and

$$\langle Tf, f \rangle_{\mathcal{H}} = \int_{[0, +\infty)} \lambda d\nu_f(\lambda).$$



By spectral calculus, there exists a unique positive operator  $G_j(T) : \mathcal{D}_j \rightarrow \mathcal{H}$  such that

$$\langle G_j(T)f, f \rangle_{\mathcal{H}} = \int_{[0,+\infty)} G_j(\lambda) dv_f(\lambda),$$

where now

$$\mathcal{D}_j := \left\{ f \in \mathcal{H} : \int_{[0,+\infty)} G_j(\lambda)^2 dv_f(\lambda) < \infty \right\}.$$

Assume further that

$$\mathcal{D}_\infty := \{ f \in \mathcal{H} : f \in \text{dom } G_j(T)^2 \text{ for all } j \geq 0 \}$$

is a dense subset of  $\mathcal{H}$ . Assumption (13) and Definition 4.6 are still valid. Moreover, the proof of Proposition 4.7 remains essentially unchanged. The only difference is in the following lines of equalities: for a given  $f \in \mathcal{D}_\infty$ , we have

$$\begin{aligned} \sum_{j \geq 0} \langle G_j(T)^2 T f, f \rangle_{\mathcal{H}} &= \sum_{j \geq 0} \left( \int_{[0,+\infty)} \lambda G_j(\lambda)^2 dv_f(\lambda) \right) \\ &= \int_{(0,+\infty)} \left( \sum_{j \geq 0} \lambda G_j(\lambda)^2 \right) d\mu_f(\lambda) \\ &= \int_{(0,+\infty)} 1 d\mu_f(\lambda) = \|P_{\mathcal{H}_\rho} f\|_{\mathcal{H}}^2, \end{aligned}$$

where the second equality is due to Tonelli’s theorem.

### 5 Monte Carlo Wavelets

We are finally ready to define our Monte Carlo wavelets. In the following, we adopt notations, definitions and assumptions of Sects. 3 and 4. For the sake of simplicity, we further assume  $\text{supp}(\rho) = \mathcal{X}$ , so that  $\mathcal{H}_\rho = \mathcal{H}$ . By Proposition 4.7, the family  $\Psi$  defined in (14) describes a Parseval frame on the entire Hilbert space  $\mathcal{H}$ .

**Definition 5.1** (*Monte Carlo wavelets*) Suppose we have  $N$  independent and identically distributed samples  $x_1, \dots, x_N \sim \rho$ . Consider the empirical covariance operator  $\widehat{T} : \mathcal{H} \rightarrow \mathcal{H}$  defined by

$$\widehat{T} := \frac{1}{N} \sum_{k=1}^N K_{x_k} \otimes K_{x_k}.$$

Let  $\{G_j\}_{j \geq 0}$  be a family of filters as in Definition 4.2. We call

$$\widehat{\Psi}^N := \{ \widehat{\psi}_{j,k} := G_j(\widehat{T})K_{x_k} : j \geq 0 \text{ and } k = 1, \dots, N \}$$

a family of *Monte Carlo wavelets*.

The family  $\widehat{\Psi}^N$  of Definition 5.1 corresponds to the family  $\Psi$  of Definition 4.6 with respect to the empirical measure  $\widehat{\rho}_N := \frac{1}{N} \sum_{k=1}^N \delta_{x_k}$ . Hence, thanks to Proposition 4.7,  $\widehat{\Psi}^N$  defines a discrete Parseval frame on the finite dimensional space

$$\widehat{\mathcal{H}}_N := \text{span}\{K_{x_k} : k = 1, \dots, N\}.$$

Now, let  $\Psi$  be the family of wavelets in the sense of Definition 4.6 with respect to the (continuous) measure  $\rho$ . Again by Proposition 4.7,  $\Psi$  is a (continuous) Parseval frame on the (infinite dimensional) space  $\mathcal{H}$ . Taking more and more samples, we obtain a sequence of frames  $\widehat{\Psi}^N$  on a chain of nested subspaces of increasing dimension:

$$\widehat{\mathcal{H}}_N \subset \widehat{\mathcal{H}}_{N+1} \subset \dots \subset \mathcal{H}.$$

We thus interpret  $\widehat{\Psi}^N$  as a Monte Carlo estimate of  $\Psi$ . In this view, we are interested in studying the asymptotic behavior of  $\widehat{\Psi}^N$  as  $N \rightarrow \infty$ , and, in particular, the convergence of  $\widehat{\Psi}^N$  to  $\Psi$ .

Notice that, despite being finite-dimensional, the frame  $\widehat{\Psi}^N$  consists of functions that are well-defined on the entire space  $\mathcal{X}$ . In particular, for any signal  $f$  in the reproducing kernel Hilbert space  $\mathcal{H}$ , we can study the wavelet expansion

$$f \approx \sum_{j \leq \tau} \sum_{k=1}^N \langle f, \widehat{\psi}_{j,k} \rangle_{\mathcal{H}} \widehat{\psi}_{j,k}. \tag{23}$$

This series approximates  $f$  up to a resolution  $\tau$  and a sampling rate  $N$ . Our main result (Theorem 7.5) states that, cutting off the frequencies at a threshold  $\tau = \tau(N)$  and letting  $N$  go to infinity, the error of (23) goes to zero,

$$\left\| f - \sum_{j \leq \tau(N)} \sum_{k=1}^N \langle f, \widehat{\psi}_{j,k} \rangle_{\mathcal{H}} \widehat{\psi}_{j,k} \right\|_{\mathcal{H}} \xrightarrow{N \rightarrow \infty} 0,$$

at a rate that depends on the regularity of the signal  $f$ . In other words, the frame constructed on the sample space  $\{x_1, \dots, x_N\}$  is asymptotically resolving the signal defined on the space  $\mathcal{X}$ . This result will be derived as a finite-sample bound in high probability.

*Deterministic discretization vs random sampling* Discretization is a classical problem in frame theory, harmonic analysis and applied mathematics *tout court*. While the construction of reproducing representations may usefully exploit rich topological, algebraic and measure theoretical properties of a continuous parameter space, discretization is eventually required when it comes to numerical implementation. Starting from a continuous frame  $\{\psi_a : a \in \mathcal{A}\}$  in a Hilbert space  $\mathcal{H}$ , frame discretization selects a countable subset of parameters  $\mathcal{A}' \subset \mathcal{A}$  so that the corresponding subfamily

$\{\psi_a : a \in \mathcal{A}'\}$  preserves the frame property. This typically involves a deterioration of the frame bounds, which grows with the sparsity of  $\mathcal{A}'$ .

A possible interpretation of our Monte Carlo wavelets is as a randomized approximate frame discretization. Random sampling may be useful when the topology of the parameter space is complex or unknown. On the other hand, our discrete frame is not a frame on the original space  $\mathcal{H}$ , but only on a finite dimensional approximation  $\widehat{\mathcal{H}}$  of  $\mathcal{H}$ . Notice though that our frame preserves the tightness, and the signal loss  $\mathcal{H} \setminus \widehat{\mathcal{H}}$  is asymptotically zero. Moreover, the numerical implementation of any discretized frame on  $\mathcal{H}$  would still require truncation at finitely many terms, resulting in fact in a loss of the global frame property. Lastly, when the space is unknown and we can only access signals through finite samples, going beyond the given sampling resolution might per se not be significant, while our results characterize how the frame parameters may be chosen adaptively to the given sampling rate.

*Numerical implementation* The representation of  $\widehat{\psi}_{j,k}$  in Definition 5.1 is remarkably compact, but hardly suitable for computation. We next provide an implementable formula of our Monte Carlo wavelets, using the Mercer representation (15) along with the singular value decomposition (6). Let  $\widehat{\mathbf{T}}\widehat{v}_i = \widehat{\lambda}_i\widehat{v}_i$  be the eigendecomposition of  $\widehat{\mathbf{T}}$ . Then (15) reads as

$$\widehat{\psi}_{j,k}(x) = \sum_{i=1}^N G_j(\widehat{\lambda}_i)\overline{\widehat{v}_i(x_k)}\widehat{v}_i(x) \quad j \geq 0, k = 1, \dots, N,$$

where the eigenpairs  $(\widehat{\lambda}_i, \widehat{v}_i)$  can be computed from the *kernel matrix*

$$\mathbf{K}[i, k] := K(x_i, x_k) \quad i, k = 1, \dots, N. \tag{24}$$

Indeed, we have  $\widehat{\mathbf{T}} = \widehat{\mathbf{S}}^*\widehat{\mathbf{S}}$  and  $N^{-1}\mathbf{K} = \widehat{\mathbf{S}}\widehat{\mathbf{S}}^*$ , where  $\widehat{\mathbf{S}}$  is the *sampling operator*

$$\widehat{\mathbf{S}} : \mathcal{H} \rightarrow \mathbb{C}^N, \quad (\widehat{\mathbf{S}}f)[i] = f(x_i) \quad i = 1, \dots, N, \tag{25}$$

and  $\widehat{\mathbf{S}}^*$  is the *out-of-sample extension*

$$\widehat{\mathbf{S}}^* : \mathbb{C}^N \rightarrow \mathcal{H}, \quad (\widehat{\mathbf{S}}^*\mathbf{u})(x) = \frac{1}{N} \sum_{\ell=1}^N K(x, x_\ell)\mathbf{u}[\ell] \quad x \in \mathcal{X}. \tag{26}$$

Thus, the eigenvalues  $\widehat{\lambda}_i$  of  $\widehat{\mathbf{T}}$  are exactly the eigenvalues of  $N^{-1}\mathbf{K}$ . Moreover, in view of (6), the eigenfunctions  $\widehat{v}_i$  can be obtained from the eigenvectors  $\widehat{\mathbf{u}}_i$  of  $N^{-1}\mathbf{K}$  by

$$\widehat{v}_i = \widehat{\lambda}_i^{-1/2}\widehat{\mathbf{S}}^*\widehat{\mathbf{u}}_i = \widehat{\lambda}_i^{-1/2} \frac{1}{N} \sum_{\ell=1}^N \widehat{\mathbf{u}}_i[\ell]K_{x_\ell},$$

which evaluated at  $x_k$  gives

$$\widehat{v}_i(x_k) = \widehat{\lambda}_i^{-1/2} \frac{1}{N} \sum_{\ell=1}^N K(x_k, x_\ell) \widehat{\mathbf{u}}_i[\ell] = \widehat{\lambda}_i^{-1/2} N^{-1} (\mathbf{K} \widehat{\mathbf{u}}_i)[k] = \widehat{\lambda}_i^{1/2} \widehat{\mathbf{u}}_i[k].$$

We therefore obtain the computable formula

$$\widehat{\psi}_{j,k}(x) = \frac{1}{N} \sum_{i,\ell=1}^N G_j(\widehat{\lambda}_i) \overline{\widehat{\mathbf{u}}_i[k]} \widehat{\mathbf{u}}_i[\ell] K(x, x_\ell) \quad j \geq 0, k = 1, \dots, N.$$

For what concerns the Monte Carlo wavelet transform of a signal  $f \in \mathcal{H}$ , it is easy to see that

$$\langle f, \widehat{\psi}_{j,k} \rangle_{\mathcal{H}} = \mathbf{U} G_j(\Lambda) \mathbf{U}^* f(x_k),$$

where  $N^{-1} \mathbf{K} = \mathbf{U} \Lambda \mathbf{U}^*$  expresses the eigendecomposition of  $N^{-1} \mathbf{K}$  in matrix form.

*Computational considerations* The bottleneck in the implementation of our Monte Carlo wavelets is the eigendecomposition of the kernel matrix, which requires in general  $\mathcal{O}(N^3)$  operations and is therefore impractical in typical large scale scenarios. This is in fact a common problem for virtually all spectral based constructions of frames (see e.g. [28,33,38]). A possible solution is approximating the filters by low order polynomials, thus simplifying the functional calculus to repeated matrix-vector multiplication, which scales well in the case of sparse graphs [33]. While kernel matrices are typically dense, such an approach may still be useful for compactly supported kernels [58], although their real applicability is mostly limited to the low-dimensional regime. Besides sparsity, a more reasonable property to leverage is fast eigenvalue decay, which opens onto a variety of methods for truncated approximate SVD. Deterministic methods allow to compute an  $r$ -rank approximation in  $\mathcal{O}(rN^2)$  [52], whereas randomized methods can further reduce the complexity to  $\mathcal{O}(\log rN^2 + r^2N)$  [32,40].

We also remark that the actual Monte Carlo approximation of a given signal is in principle a different problem than the computation of the frame itself, and as such may in some cases be more tractable. For example, for some specific filters as in Table 2, the computation of (23) boils down to the implementation of some regularized inversion or minimization procedure, for which several approaches based on sketching, random projections, hierarchical decompositions and early stopping may be profitably used [7,9,17,47–50,59]. An efficient implementation of Monte Carlo wavelets is out of the scope of this paper and will be subject of future work.

## 6 Comparison with Other Frame Constructions

The approach we adopt in Sect. 4 differs from the existing literature in several crucial aspects. We now give an overview of similarities and differences. As argued in Sect.

1, many techniques for the analysis of signals on non-Euclidean domains, such as manifolds and graphs, are based on spectral filtering of some suitable operator. There are, generally speaking, two distinct yet related perspectives.

A first type of methods builds frames for function spaces on compact differentiable manifolds associated with certain positive operators (predominantly the Laplace–Beltrami operator). In [13,26], filter functions  $g_j$  are applied to the given operator  $L$ , giving  $g_j(\sqrt{L})$  for  $j \geq 0$ . One then needs to ensure that this defines an integral operator with a corresponding kernel  $\psi_j(\sqrt{L})(x, y)$ , which often poses a technical challenge, and relies on the relationship between the operator  $L$  and local metric properties of the manifold. We avoid this by using a positive definite kernel from the start. The next step is to sample points  $\{x_k^j\}_{k=1}^{m_j}$  from the manifold for each scale  $j$ , in such a way that they form a  $\delta_j$ -net and satisfy a cubature rule for functions in the desired space. Frame elements are then defined by  $C_{j,k} \psi_j(\sqrt{L})(x_k^j, \cdot)$ , for some suitable weights  $C_{j,k}$ . The resulting family of functions constitutes a non-tight frame on the entire function space. On the contrary, our sampled frames are Parseval frames on finite-dimensional subspaces. As we are going to show in the next section, in order to establish convergence we do not require a stringent selection of points; instead, we sample at random, which allows for a straightforward algorithmic approach, independent of the specific geometry of the underlying space.

In a different line of research [38,42,57], frames are built on an arbitrary orthonormal basis  $\{w_i\}_{i \geq 0}$  of a separable Hilbert space of functions defined on a quasi-metric measure space, together with a suitable sequence of positive reals  $(l_i)_{i \geq 0}$ . Based on these data, a kernel-like function  $K_H(x, \cdot) := \sum_{i \geq 0} H(l_i) w_i(x) w_i$  is constructed. This mirrors the basis expansion of frame elements (15), but in our case a specific orthonormal basis is taken, that is, the eigenbasis of the integral operator, and  $(l_i)_{i \geq 0}$  are the corresponding eigenvalues. Due to the use of an arbitrary basis and sequence, an additional effort (or a set of assumptions) needs to be made in order to ensure the desired properties, such as the decay of the approximation error as the number of eigenvalues resolved by the function  $H$  increases. Some of the results are similar to those in our paper, albeit estimation errors or sample bounds have not been established in this context.

On the other hand, starting from a discrete setting, graph signal processing considers a weight (or adjacency) matrix to define a certain graph operator  $L$ , such as the graph Laplacian [28,33] or a diffusion operator [12]. The frame elements are then defined in the spectral domain as  $\psi_{j,x} := g_j(L) \delta_x$ , where  $g$  is an admissible wavelet kernel,  $j$  a scale parameter, and  $\delta_x$  the indicator function of a vertex  $x$ . This is conceptually similar to (14), though there are also several distinctions. First, following [28], our construction results in *Parseval* frames. This simplifies the computational effort, since Parseval frames are canonically self-dual, and thus signal reconstruction does not require the computation of a dual frame. Moreover, to localize the frame in space we use the continuous kernel function  $K_x$ , instead of the impulse  $\delta_x$ . Since in our setting the kernel  $K$  is used both to define the underlying integral operator and to localize the frame elements, we can use the theory of RKHS to establish a connection between continuous and discrete frames, as we will show in Sect. 7. In typical constructions of

frames on graphs, a more judicious effort is usually required to elaborate analogous convergence results.

### 7 Stability of Monte Carlo Wavelets

In this section we study the relationship between continuous and discrete frames, regarding the latter as Monte Carlo estimates of the former. We begin by restricting our attention to  $\mathcal{H}$ , and we will then extend the analysis to  $L^2(\mathcal{X}, \rho)$ . Let

$$T_j := \int_{\mathcal{X}} \psi_{j,x} \otimes \psi_{j,x} d\rho(x), \quad \widehat{T}_j := \frac{1}{N} \sum_{k=1}^N \widehat{\psi}_{j,k} \otimes \widehat{\psi}_{j,k}$$

be the frame operators associated with the scale  $j$ , and its empirical counterpart. By Proposition 4.8, we have

$$\text{Id}_{\mathcal{H}} = \sum_{j \geq 0} T_j, \quad \text{Id}_{\widehat{\mathcal{H}}_N} = \sum_{j \geq 0} \widehat{T}_j.$$

For  $f \in \mathcal{H}$ , given a threshold scale  $\tau \in \mathbb{N}$  and a sample size  $N$ , we let

$$\widehat{f}_{\tau,N} := \sum_{j=0}^{\tau} \widehat{T}_j f \tag{27}$$

be the empirical approximation of  $f$  using the first  $\tau$  scales of the frame  $\widehat{\Psi}^N$ . The reconstruction error of  $\widehat{f}_{\tau,N}$  can be decomposed into

$$\|f - \widehat{f}_{\tau,N}\|_{\mathcal{H}} \leq \left\| \sum_{j > \tau} T_j f \right\|_{\mathcal{H}} + \left\| \sum_{j=0}^{\tau} (T_j - \widehat{T}_j) f \right\|_{\mathcal{H}}. \tag{28}$$

The first term is the *approximation error*, arising from the truncation of the resolution of the identity. The second term is the *estimation error*, which stems from estimating the measure by means of empirical samples. Next, we derive quantitative error bounds for both terms, and then balance the resolution  $\tau$  in terms of sample size  $N$  to obtain our convergence result.

*Approximation error* Note that Proposition 4.7 already implies

$$\left\| \sum_{j > \tau} T_j f \right\|_{\mathcal{H}} \xrightarrow{\tau \rightarrow \infty} 0,$$

being the tail of a convergent series. To quantify the speed of convergence with respect to  $\tau$ , approximation theory suggests that  $f$  has to obey some notion of regularity. In

the following we assume a smoothness of Sobolev kind (see [20] and Sect. 8), also known in statistical learning theory as the *source condition* (see [8]):

$$f = T^\alpha h \text{ for some } h \in \mathcal{H} \text{ and } \alpha > 0.$$

**Proposition 7.1** *Assume that  $g_j$  has qualification  $\nu \in (0, \infty]$  and  $f \in \text{range}(T^\alpha)$  for some  $\alpha > 0$ . Let  $\beta := \min\{\nu, \alpha\}$ . Then*

$$\left\| \sum_{j>\tau} T_j f \right\|_{\mathcal{H}} \lesssim \|T^{-\alpha} f\|_{\mathcal{H}} \kappa^{2(\alpha-\beta)} \tau^{-\beta}.$$

**Proof** By (21) we have  $\sum_{j>\tau} T_j = \text{Id}_{\mathcal{H}} - T g_\tau(T)$ . Hence,

$$\begin{aligned} \left\| \sum_{j>\tau} T_j f \right\|_{\mathcal{H}}^2 &= \sum_{i \in \mathcal{I}_\rho} |1 - \lambda_i g_\tau(\lambda_i)|^2 |\langle f, v_i \rangle_{\mathcal{H}}|^2 \\ &= \sum_{i \in \mathcal{I}_\rho} \left( \lambda_i^\beta |1 - \lambda_i g_\tau(\lambda_i)| \right)^2 |\langle T^{-\beta} f, v_i \rangle_{\mathcal{H}}|^2 \\ &\leq \left( \sup_{i \in \mathcal{I}_\rho} \lambda_i^\beta |1 - \lambda_i g_\tau(\lambda_i)| \right)^2 \sum_{i \in \mathcal{I}_\rho} |\langle T^{-\beta} f, v_i \rangle_{\mathcal{H}}|^2 \\ &\lesssim \tau^{-2\beta} \kappa^{4(\alpha-\beta)} \|T^{-\alpha} f\|_{\mathcal{H}}^2. \end{aligned}$$

□

*Estimation error* To bound the second term in (28), we rely on concentration results for covariance operators [46].

**Proposition 7.2** *Assume that  $\lambda \mapsto \lambda g_\tau(\lambda)$  is Lipschitz continuous on  $[0, \kappa^2]$  with Lipschitz constant  $L(\tau)$ . Then, for every  $f \in \mathcal{H}$  and  $t > 0$ , with probability at least  $1 - 2e^{-t}$  we have*

$$\left\| \sum_{j=0}^{\tau} (T_j - \widehat{T}_j) f \right\|_{\mathcal{H}} \lesssim \|f\|_{\mathcal{H}} \kappa^2 \sqrt{t} L(\tau) N^{-1/2}.$$

**Proof** Using (21) and Lemma A.2 we have

$$\begin{aligned} \left\| \sum_{j=0}^{\tau} (T_j - \widehat{T}_j) f \right\|_{\mathcal{H}} &= \left\| (T g_\tau(T) - \widehat{T} g_\tau(\widehat{T})) f \right\|_{\mathcal{H}} \\ &\leq \left\| T g_\tau(T) - \widehat{T} g_\tau(\widehat{T}) \right\|_{\text{HS}} \|f\|_{\mathcal{H}} \\ &\leq L(\tau) \|T - \widehat{T}\|_{\text{HS}} \|f\|_{\mathcal{H}}. \end{aligned}$$

Bounding  $\|T - \widehat{T}\|_{HS}$  with the concentration estimate [46, Theorem 7] we obtain

$$\|T - \widehat{T}\|_{HS} \lesssim \kappa^2 \sqrt{t} N^{-1/2}$$

with probability no lower than  $1 - 2e^{-t}$ . □

All examples of filters given in Sect. 4.1 satisfy the Lipschitz condition required in Proposition 7.2.

**Lemma 7.3** *Let  $g_j$  be a spectral function from Table 2. Then the function  $\lambda \mapsto \lambda g_\tau(\lambda)$  is Lipschitz continuous on  $[0, \kappa^2]$ , with Lipschitz constant  $L(\tau) \lesssim \tau$  for the first four spectral functions, and  $L(\tau) \lesssim \tau^2$  for the last two. Moreover, let  $g_j$  be defined as in Example 4.5, with  $|g'| \leq B$ . Then the function  $\lambda \mapsto \lambda g_\tau(\lambda)$  is Lipschitz continuous on  $[0, \kappa^2]$ , with Lipschitz constant  $L(\tau) \leq B2^\tau$ .*

**Proof** For the first four spectral functions of Table 2, the claim follows by bounding the explicit derivative of  $\lambda \mapsto \lambda g_\tau(\lambda)$ ; for the last two, from an application of Markov brothers' inequality (see [43, Supplemental, Lemma 1]). For filters of Example 4.5, we differentiate  $\lambda \mapsto g(2^\tau \lambda)$  and use  $|g'| \leq B$ . □

**Remark 7.4** In this paper we are not interested in the constants. We rely on the Hilbert norm since it provides both a simple bound on  $\|T - \widehat{T}\|_{HS}$  and, by the Lipschitz assumption, the stability bound  $\|Tg_\tau(T) - \widehat{T}g_\tau(\widehat{T})\|_{HS} \|f\|_{\mathcal{H}} \leq L(\tau) \|T - \widehat{T}\|_{HS}$ . Our result can be improved by using the sharper bound

$$\|T - \widehat{T}\| \leq C \|T\| \max \left\{ \sqrt{\frac{r(T)}{N}}, \frac{r(T)}{N}, \sqrt{\frac{t}{N}}, \frac{r(T)}{N} \right\},$$

where  $r(T) = \frac{\text{trace}(T)}{\|T\|}$  (see Theorem 9 in [36] and the techniques in the proof of Theorem 3.4 in [6] to bound  $\|Tg_\tau(T) - \widehat{T}g_\tau(\widehat{T})\|$ ).

**Reconstruction error and convergence** Combining Propositions 7.1 and 7.2, we can finally prove the convergence of our Monte Carlo wavelets. In order to balance approximation and estimation error, we need to tune the resolution  $\tau$  with the number of samples  $N$  and the smoothness  $\alpha$  of the signal, in so far as the qualification  $\nu$  of the filter allows.

**Theorem 7.5** *Assume that  $g_\tau$  has qualification  $\nu \in (0, \infty]$ ,  $f \in \text{range}(T^\alpha)$  for some  $\alpha > 0$ , and  $\lambda \mapsto \lambda g_\tau(\lambda)$  is Lipschitz continuous on  $[0, \kappa^2]$  with Lipschitz constant  $L(\tau) \lesssim \tau^p$ ,  $p \geq 1$ . Let  $\beta := \min\{\alpha, \nu\}$  and set*

$$\tau := \lceil N^{\frac{1}{2(\beta+p)}} \rceil.$$

Then, for every  $t > 0$ , with probability at least  $1 - 2e^{-t}$  we have

$$\|f - \widehat{f}_{\tau, N}\|_{\mathcal{H}} \lesssim \|T^{-\alpha} f\|_{\mathcal{H}} (\kappa^{2(\alpha-\beta)} + \kappa^{2\alpha+2} \sqrt{t}) N^{-\frac{\beta}{2(\beta+p)}}.$$



**Table 3** Error rates for signals  $f \in \text{range}(T^\alpha)$  and several spectral regularizers

Method	Error rate in $\ \cdot\ _{\mathcal{H}}$	Error rate in $\ \cdot\ _\rho$
Tikhonov regularization	$N^{-\frac{\min\{\alpha,1\}}{2\min\{\alpha,1\}+2}}$	$N^{-\frac{\min\{\alpha+1/2,1\}}{2\min\{\alpha+1/2,1\}+2}}$
iterated Tikhonov ( $m$ )	$N^{-\frac{\min\{\alpha,m\}}{2\min\{\alpha,m\}+2}}$	$N^{-\frac{\min\{\alpha+1/2,m\}}{2\min\{\alpha+1/2,m\}+2}}$
Landweber iteration	$N^{-\frac{\alpha}{2\alpha+2}}$	$N^{-\frac{\alpha+1/2}{2\alpha+3}}$
asymptotic regularization	$N^{-\frac{\alpha}{2\alpha+2}}$	$N^{-\frac{\alpha+1/2}{2\alpha+3}}$
heavy ball ( $\nu$ )	$N^{-\frac{\min\{\alpha,\nu\}}{2\min\{\alpha,\nu\}+4}}$	$N^{-\frac{\min\{\alpha+1/2,\nu\}}{2\min\{\alpha+1/2,\nu\}+4}}$
Nesterov acceleration	$N^{-\frac{\min\{\alpha,\nu \geq 1/2\}}{2\min\{\alpha,\nu \geq 1/2\}+4}}$	$N^{-\frac{\min\{\alpha+1/2,\nu \geq 1/2\}}{2\min\{\alpha+1/2,\nu \geq 1/2\}+4}}$

**Proof** Starting from the decomposition (28), we bound the two terms by Propositions 7.1 and 7.2. The approximation error is  $\mathcal{O}(\tau^{-\beta})$ , while the estimation error is  $\mathcal{O}(\tau^\rho N^{-1/2})$ . We thus choose  $\tau$  to balance them out, and collect the constants.  $\square$

If  $\text{supp } \rho \neq \mathcal{X}$ , we have instead a frame on  $\mathcal{H}_\rho$ , and the corresponding resolution of the identity  $\text{Id}_{\mathcal{H}_\rho} = \sum_{j \geq 0} T_j$ . The reconstruction error would thus include an additional bias term:

$$\|f - \widehat{f}_{\tau,N}\|_{\mathcal{H}} \leq \|\text{P}_{\ker S} f\|_{\mathcal{H}} + \left\| \sum_{j > \tau} T_j f \right\|_{\mathcal{H}} + \left\| \sum_{j=0}^{\tau} (T_j - \widehat{T}_j) f \right\|_{\mathcal{H}}.$$

Classical spectral functions from Table 2 satisfy the assumptions of Theorem 7.5. We report the explicit rates in Table 3. A convergence result for filters of Example 4.5 will be provided at the end of Sect. 8.

*Convergence in  $L^2(\mathcal{X}, \rho)$*  Error rates in  $L^2(\mathcal{X}, \rho)$  can be extracted using the isometry between  $\overline{\mathcal{H}}_\rho$  and  $\mathcal{H}_\rho$ . Suppose again for simplicity that  $\text{supp } \rho = \mathcal{X}$ . In view of (8), for  $f \in \mathcal{H}_\rho = \mathcal{H}$  we have

$$\|f - \widehat{f}_{\tau,N}\|_\rho = \|\sqrt{T}(f - \widehat{f}_{\tau,N})\|_{\mathcal{H}}.$$

Decomposing the error into its approximation and estimation components, we can repeat the same analysis as in the proof of Theorem 7.5. The estimation bound simply gets an additional  $\kappa$  factor. Assuming  $f \in T^\alpha \mathcal{H}$  with  $\alpha > 0$ , for the approximation term we have

$$\begin{aligned} \|\sqrt{T} \sum_{j > \tau} T_j f\|_{\mathcal{H}} &\leq \sup_{i \in \mathcal{I}_\rho} (\lambda_i^\beta (1 - \lambda_i g_\tau(\lambda_i))) \sum_{i \in \mathcal{I}_\rho} |\langle T^{1/2-\beta} f, v_i \rangle_{\mathcal{H}}| \\ &\lesssim \|T^{-\alpha} f\|_{\mathcal{H}} \kappa^{2(\alpha-\beta)+1} \tau^{-\beta}, \end{aligned}$$

with  $\beta := \min(\alpha + 1/2, \nu)$ . Therefore, the approximation rate increases by 1/2 (qualification permitting). Combining all together, we obtain the following bound in  $L^2(\mathcal{X}, \rho)$ .

**Corollary 7.6** Assume that  $g_\tau$  has qualification  $\nu \in (0, \infty]$ ,  $f \in \text{range}(\mathbb{T}^\alpha)$  for some  $\alpha > 0$ , and  $\lambda \mapsto \lambda g_\tau(\lambda)$  is Lipschitz continuous on  $[0, \kappa^2]$  with Lipschitz constant  $L(\tau) \lesssim \tau^p$ ,  $p \geq 1$ . Let  $\beta := \min\{\alpha + 1/2, \nu\}$  and set

$$\tau := \lceil N^{\frac{1}{2(\beta+p)}} \rceil.$$

Then, for every  $t > 0$ , with probability at least  $1 - 2e^{-t}$  we have

$$\|f - \widehat{f}_{\tau,N}\|_\rho \lesssim \|\mathbb{T}^{-\alpha} f\|_{\mathcal{H}} \left( \kappa^{2(\alpha-\beta)+1} + \kappa^{2\alpha+3} \sqrt{t} \right) N^{-\frac{\beta}{2(\beta+p)}}.$$

See Table 3 for specific rates regarding spectral functions from Table 2.

*Monte Carlo wavelet approximation as noiseless kernel ridge regression* We conclude this section with an observation that draws a link between Monte Carlo wavelets and regression analysis. Let  $nb\widehat{f}_{\tau,N}$  be the Monte Carlo wavelet approximation (27) of  $f \in \mathcal{H}$  at resolution  $\tau$  given samples  $x_1, \dots, x_N$ . Then

$$\widehat{f}_{\tau,N} = \sum_{j=0}^{\tau} G_j(\widehat{\mathbb{T}})^2 \widehat{\mathbb{T}} f = g_\tau(\widehat{\mathbb{T}}) \widehat{\mathbb{T}} f.$$

With the choice of the Tikhonov filter  $g_j(\lambda) = (\lambda + \tau^{-1})^{-1}$  (Table 2), recalling (24), (25) and (26), and defining

$$\mathbf{y} = [f(x_1), \dots, f(x_N)]^\top, \quad \boldsymbol{\alpha} = \left( \mathbf{K} + \frac{N}{\tau} \mathbf{I} \right)^{-1} \mathbf{y},$$

we have

$$\begin{aligned} \widehat{f}_{\tau,N} &= \left( \widehat{\mathbb{T}} + \frac{1}{\tau} \text{Id}_{\mathcal{H}} \right)^{-1} \widehat{\mathbb{T}} f = \left( \widehat{\mathbf{S}}^* \widehat{\mathbf{S}} + \frac{1}{\tau} \text{Id}_{\mathcal{H}} \right)^{-1} \widehat{\mathbf{S}}^* \widehat{\mathbf{S}} f = \left( \widehat{\mathbf{S}}^* \widehat{\mathbf{S}} + \frac{1}{\tau} \text{Id}_{\mathcal{H}} \right)^{-1} \widehat{\mathbf{S}}^* \mathbf{y} \\ &= \widehat{\mathbf{S}}^* \left( \widehat{\mathbf{S}} \widehat{\mathbf{S}}^* + \frac{1}{\tau} \mathbf{I} \right)^{-1} \mathbf{y} = \frac{1}{N} \sum_{i=1}^N K(\cdot, x_i) \left[ \left( \frac{1}{N} \mathbf{K} + \frac{1}{\tau} \mathbf{I} \right)^{-1} \mathbf{y} \right][i] \\ &= \sum_{i=1}^N K(\cdot, x_i) \left[ \left( \mathbf{K} + \frac{N}{\tau} \mathbf{I} \right)^{-1} \mathbf{y} \right][i] = \sum_{i=1}^N \boldsymbol{\alpha}[i] K(\cdot, x_i). \end{aligned}$$

This is the (unique) solution to the kernel regularized least squares problem

$$\min_{\widehat{f} \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^N |y_i - \widehat{f}(x_i)|^2 + \lambda \|\widehat{f}\|_{\mathcal{H}}^2, \tag{29}$$

where  $y_i = \mathbf{y}[i]$  and  $\lambda = \tau^{-1}$ . Therefore,  $\widehat{f}_{\tau,N}$  is the kernel ridge estimator for the noiseless regression problem

$$y_i = f(x_i) \quad i = 1, \dots, N,$$

and the squared reconstruction error  $\|f - \widehat{f}_{\tau,N}\|_{\rho}^2$  is the generalization error of  $\widehat{f}_{\tau,N}$ .

Contrasting this with the optimal rate (in the minimax sense) for kernel ridge regression [8] entails that the rate in Table 3 is suboptimal for Tikhonov regularization, and presumably for all other regularizers. This is well expected from the crude Lipschitz bound used in Proposition 7.2. The scope of the present work was to establish a first result of convergence of randomly sampled frames, rather than identifying the optimality of the convergence rates. Refinement of our bounds will be object of future investigation (see also Remark 7.4).

### 8 Sobolev and Besov Spaces in RKHS

The convergence rates of the frame reconstruction error in Theorem 7.5 depend on the approximation rates in Proposition 7.1, hence on the regularity of the original signal  $f$ , as quantified by the condition  $f \in \text{range}(T^\alpha)$ . Thinking of  $T$  as the inverse square root of the Laplacian allows to interpret  $\text{range}(T^\alpha)$  as a Sobolev space. The theory of smoothness function spaces [56] plays a critical role in harmonic analysis, and serves also as a base for the definition of statistical priors in learning theory [5]. In this section we examine general notions of regularity and their effect on the reconstruction error. Many of the reported results on Besov spaces are well known [20], but we nonetheless include them here to be self contained and to adapt them to our setting and notation. In particular, as already observed in Sect. 2, it should be borne in mind that the spectrum of the integral operator  $T$  has inverse trend compared to that of a Laplace operator, and therefore all the spectral definitions of the generalized Besov spaces must take this into account in order to preserve the consistency with their classical counterparts. As in the previous section, we assume  $\text{supp}(\rho) = \mathcal{X}$ .

*Sobolev spaces as domains of powers of a positive operator* By virtue of the spectral theorem, for every  $\alpha > 0$ ,  $T^\alpha$  is a positive, bounded, injective operator on  $\mathcal{H}$ , with  $\sigma(T^\alpha) \subset (0, \kappa^{2\alpha}]$ . Thus,  $T^{-\alpha}$  is a positive, closed, densely-defined, injective operator with  $\sigma(T^{-\alpha}) \subset [\kappa^{-2\alpha}, \infty)$ . We put the following

**Definition 8.1** (*Sobolev spaces*) For  $\alpha > 0$ , we define the *Sobolev space*  $\mathcal{H}^\alpha$  by

$$\mathcal{H}^\alpha := \text{dom}(T^{-\alpha}) = \text{range}(T^\alpha),$$

equipped with the norm

$$\|v\|_{\mathcal{H}^\alpha} := \|T^{-\alpha}v\|_{\mathcal{H}}.$$

$\mathcal{H}^\alpha$  is a Hilbert space. Moreover, we have

$$\mathcal{H}^\alpha = \left\{ f \in \mathcal{H} : \sum_{i \in \mathcal{I}_\rho} \lambda_i^{-2\alpha} |\langle f, v_i \rangle_{\mathcal{H}}|^2 < \infty \right\},$$

which expresses  $\mathcal{H}^\alpha$  in terms of the speed of decay of the Fourier coefficients, thus generalizing the standard Sobolev spaces  $H^\alpha = W^{\alpha,2}$ . Theorem 7.5 establishes the convergence of Monte Carlo wavelets for signals in the class  $\mathcal{H}^\alpha$ .

*Besov spaces as approximation spaces* Besov spaces on Euclidean domains are traditionally defined by the decay of the modulus of continuity. A characterization that is best suited to generalize to arbitrary domains, and to which we also adhere, is through approximation and interpolation spaces [20,45,56]. We begin with the approximation perspective by defining a scale of Paley–Wiener spaces.

**Definition 8.2** (*Paley–Wiener spaces*) For  $\omega > 0$ , the *Paley–Wiener space*  $\mathbf{PW}(\omega)$  is defined by

$$\mathbf{PW}(\omega) := \left\{ f \in \mathcal{H} : \langle f, v_i \rangle_{\mathcal{H}} = 0 \text{ for } \lambda_i < \omega^{-1} \right\} = \overline{\text{span}} \left\{ v_i : \lambda_i \geq \omega^{-1} \right\}.$$

The associated approximation error for  $f \in \mathcal{H}$  is

$$\mathcal{E}(f, \omega) := \inf_{g \in \mathbf{PW}(\omega)} \|f - g\|_{\mathcal{H}} = \|\mathbf{P}_{\mathbf{PW}(\omega)^\perp} f\|_{\mathcal{H}} = \left( \sum_{\lambda_i < \omega^{-1}} |\langle f, v_i \rangle_{\mathcal{H}}|^2 \right)^{1/2}.$$

The space  $\mathbf{PW}(\omega)$  is a closed subspace of  $\mathcal{H}$ , and  $\bigcup_{\omega > 0} \mathbf{PW}(\omega)$  is dense in  $\mathcal{H}$ . Note that  $\mathcal{E}(f, \omega) \xrightarrow{\omega \rightarrow 0} \|f\|_{\mathcal{H}}$  and  $\mathcal{E}(f, \omega) \xrightarrow{\omega \rightarrow \infty} 0$ . Approximation spaces classify functions in  $\mathcal{H}$  according to the rate of decay of their approximation error.

**Definition 8.3** (*Besov spaces*) For  $s > 0$  and  $q \in [1, \infty)$ , we define the *Besov space*  $\mathcal{B}_q^s$  as the approximation space

$$\mathcal{B}_q^s := \left\{ f \in \mathcal{H} : \left( \int_0^\infty (\omega^s \mathcal{E}(f, \omega))^q \frac{d\omega}{\omega} \right)^{1/q} < \infty \right\},$$

equipped with the norm

$$\|f\|_{\mathcal{B}_q^s} := \|f\|_{\mathcal{H}} + \left( \int_0^\infty (\omega^s \mathcal{E}(f, \omega))^q \frac{d\omega}{\omega} \right)^{1/q}. \tag{30}$$

The space  $\mathcal{B}_\infty^s$  is defined with the usual adjustment.

Discretizing the integral in (30), we obtain the equivalent norm

$$\|f\|_{\mathcal{H}} + \left( \sum_{j \geq 0} \left( 2^{js} \mathcal{E}(f, 2^j) \right)^q \right)^{1/q} \asymp \|f\|_{\mathcal{B}_q^s}. \tag{31}$$

In particular, a function  $f \in \mathcal{B}_q^s$  if and only if the sequence  $(2^{js} \mathcal{E}(f, 2^j))_{j \geq 0} \in \ell^q$ . It is easy to see that the scale of spaces  $\mathcal{B}_q^s$  obeys the following lexicographical order

[45, Proposition 3]:

$$\begin{aligned} \mathcal{B}_q^s &\supset \mathcal{B}_p^t \quad \text{for } s < t, \\ \mathcal{B}_q^s &\subset \mathcal{B}_p^s \quad \text{for } q < p. \end{aligned} \tag{32}$$

*Besov spaces as interpolation spaces* The Sobolev space  $\mathcal{H}^\alpha$  is continuously embedded into  $\mathcal{B}_q^s$  for every  $\alpha > s$ . Indeed, for  $f \in \mathcal{H}^\alpha$  we have the Jackson-type inequality  $\mathcal{E}(f, \omega) \leq \omega^{-\alpha} \|f\|_{\mathcal{H}^\alpha}$ , hence

$$\sum_{j \geq 0} (2^{js} \mathcal{E}(f, 2^j))^q \leq \|f\|_{\mathcal{H}^\alpha}^q \sum_{j \geq 0} 2^{-jq(\alpha-s)} < \infty.$$

Furthermore,  $\mathcal{B}_q^s$  interpolates between  $\mathcal{H}^\alpha$  and  $\mathcal{H}$ .

**Definition 8.4** (*Interpolation spaces*) For quasi-normed spaces  $\mathbf{E}$  and  $\mathbf{F}$ ,  $\theta \in (0, 1)$  and  $q \in (0, \infty)$ , the quasi-normed *interpolation space*  $(\mathbf{E}, \mathbf{F})_{\theta,q}$  is defined by

$$(\mathbf{E}, \mathbf{F})_{\theta,q} := \left\{ f \in \mathbf{E} + \mathbf{F} : \int_0^\infty (t^{-\theta} \mathcal{K}(f, t))^q \frac{dt}{t} < \infty \right\},$$

where  $\mathcal{K}(f, t)$  is Peetre’s  $K$ -functional

$$\mathcal{K}(f, t) := \inf_{\substack{f_0+f_1=f \\ f_0 \in \mathbf{E}, f_1 \in \mathbf{F}}} \|f_0\|_{\mathbf{E}} + t \|f_1\|_{\mathbf{F}}.$$

The space  $(\mathbf{E}, \mathbf{F})_{\theta,\infty}$  is defined with the usual adjustment.

Standard interpolation theory [20,56] gives

$$\mathcal{B}_q^s = (\mathcal{H}, \mathcal{H}^\alpha)_{\frac{s}{\alpha},q} \quad \text{for } s \in (0, \alpha) \text{ and } q \in [1, \infty], \tag{33}$$

with

$$\|f\|_{\mathcal{B}_q^s} \asymp \|f\|_{\mathcal{H}} + \left( \int_0^\infty (t^{-\theta} \mathcal{K}(f, t))^q \frac{dt}{t} \right)^{1/q}. \tag{34}$$

In the next proposition we show that, as in the Euclidean setting, the Besov space  $\mathcal{B}_2^s$  coincides with the Sobolev space  $\mathcal{H}^s$  of the same order. As in the classical setting, this is particular to the case  $q = 2$ . This is probably a known fact, but we could find neither a proof nor a statement.

**Proposition 8.5** *For every  $s > 0$ ,  $\mathcal{B}_2^s = \mathcal{H}^s$  with equivalent norms.*

**Proof** Let  $\alpha = 2s$ . Then (33) and (34) give  $\mathcal{B}_2^s = (\mathcal{H}, \mathcal{H}^\alpha)_{\frac{s}{\alpha},2} = (\mathcal{H}, \mathcal{H}^{2s})_{\frac{1}{2},2}$  and

$$\|f\|_{\mathcal{B}_2^s}^2 \asymp \|f\|_{\mathcal{H}}^2 + \int_0^\infty t^{-1} \mathcal{K}(f, t)^2 \frac{dt}{t}. \tag{35}$$

Let  $A : \mathcal{H}^\alpha \rightarrow \mathcal{H}$  denote the canonical embedding  $Ag = g$ . Then, for  $f \in \mathcal{H}$  and  $t > 0$  we have

$$\begin{aligned} \mathcal{K}(f, t)^2 &= \inf_{\substack{f_0 + Ag = f \\ f_0 \in \mathcal{H}, g \in \mathcal{H}^\alpha}} (\|f_0\|_{\mathcal{H}} + t \|g\|_{\mathcal{H}^\alpha})^2 \\ &= \inf_{g \in \mathcal{H}^\alpha} (\|f - Ag\|_{\mathcal{H}} + t \|g\|_{\mathcal{H}^\alpha})^2 \asymp \mathcal{G}(f, t^2), \end{aligned} \tag{36}$$

with

$$\mathcal{G}(f, \lambda) := \inf_{g \in \mathcal{H}^\alpha} \|f - Ag\|_{\mathcal{H}}^2 + \lambda \|g\|_{\mathcal{H}^\alpha}^2.$$

This infimum is attained by  $g = (A^*A + \lambda \text{Id}_{\mathcal{H}^\alpha})^{-1}A^*f$ . Since

$$(A^*A + \lambda \text{Id}_{\mathcal{H}^\alpha})^{-1}A^* = A^*(AA^* + \lambda \text{Id}_{\mathcal{H}})^{-1},$$

defining  $B := AA^* : \mathcal{H} \rightarrow \mathcal{H}$  we obtain

$$A(A^*A + \lambda \text{Id}_{\mathcal{H}^\alpha})^{-1}A^* = B(B + \lambda \text{Id}_{\mathcal{H}})^{-1}.$$

Let  $A^* = U(AA^*)^{1/2} = UB^{1/2}$  be the polar decomposition of  $A^*$ , where  $U : \mathcal{H} \rightarrow \mathcal{H}^\alpha$  is unitary. We have

$$\mathcal{G}(f, \lambda) = \left\| (\text{Id}_{\mathcal{H}^\alpha} - B(B + \lambda \text{Id}_{\mathcal{H}})^{-1})f \right\|_{\mathcal{H}}^2 + \lambda \|UB^{1/2}(B + \lambda \text{Id}_{\mathcal{H}})^{-1}f\|_{\mathcal{H}^\alpha}^2.$$

Since  $(\text{Id}_{\mathcal{H}} - B(B + \lambda \text{Id}_{\mathcal{H}})^{-1})(B + \lambda \text{Id}_{\mathcal{H}}) = \lambda \text{Id}_{\mathcal{H}}$ , it follows that

$$\begin{aligned} \mathcal{G}(f, \lambda) &= \lambda^2 \| (B + \lambda \text{Id}_{\mathcal{H}})^{-1}f \|_{\mathcal{H}}^2 + \lambda \| B^{1/2}(B + \lambda \text{Id}_{\mathcal{H}})^{-1}f \|_{\mathcal{H}}^2 \\ &= \lambda \left[ \lambda \langle (B + \lambda \text{Id}_{\mathcal{H}})^{-2}f, f \rangle_{\mathcal{H}} + \langle B(B + \lambda \text{Id}_{\mathcal{H}})^{-2}f, f \rangle_{\mathcal{H}} \right] \\ &= \lambda \langle (B + \lambda \text{Id}_{\mathcal{H}})^{-2}(\lambda \text{Id}_{\mathcal{H}} + B), f \rangle_{\mathcal{H}} \\ &= \lambda \langle (B + \lambda \text{Id}_{\mathcal{H}})^{-1}f, f \rangle_{\mathcal{H}}. \end{aligned} \tag{37}$$

Plugging (36) and (37) into (35) we get

$$\begin{aligned} \int_0^\infty t^{-1} \mathcal{K}(f, t)^2 \frac{dt}{t} &\asymp \int_0^\infty t^{-1} \mathcal{G}(f, t^2) \frac{dt}{t} \\ &= \int_0^\infty \left\langle (B + t^2 \text{Id}_{\mathcal{H}})^{-1}f, f \right\rangle_{\mathcal{H}} dt = \int_0^\infty \int_0^\infty \frac{1}{\sigma + t^2} \langle d\pi_B(\sigma)f, f \rangle dt, \end{aligned}$$

where  $\pi_B$  is the spectral measure of  $B$ . By Fubini we have

$$\begin{aligned} \int_0^\infty \int_0^\infty \frac{1}{\sigma + t^2} dt \langle d\pi_B(\sigma) f, f \rangle &= \int_0^\infty \frac{1}{\sqrt{\sigma}} \arctan \left( \frac{t}{\sqrt{\sigma}} \right) \Big|_0^\infty \langle d\pi_B(\sigma) f, f \rangle \\ &\asymp \int_0^\infty \sigma^{-1/2} \langle d\pi_B(\sigma) f, f \rangle = \langle B^{-1/2} f, f \rangle_{\mathcal{H}} = \|B^{-1/4} f\|_{\mathcal{H}}^2. \end{aligned}$$

Therefore,  $f \in \mathcal{B}_2^s$  if and only if  $f \in \text{dom}(B^{-1/4})$ . It now suffices to show  $B^{-1/4} = T^{-s}$ , whence  $\|B^{-1/4} f\|_{\mathcal{H}}^2 = \|f\|_{\mathcal{H}^s}^2$ . For any  $f \in \mathcal{H}$  and  $g \in \mathcal{H}^\alpha$  we have

$$\langle f, Ag \rangle_{\mathcal{H}} = \langle A^* f, g \rangle_{\mathcal{H}^\alpha} = \langle T^{-\alpha} AA^* f, T^{-\alpha} Ag \rangle_{\mathcal{H}} = \langle T^{-2\alpha} B f, g \rangle_{\mathcal{H}}.$$

Since  $\mathcal{H}^\alpha$  is dense in  $\mathcal{H}$ , this implies  $T^{-2\alpha} B = \text{Id}_{\mathcal{H}}$ . Hence,  $B = T^{2\alpha} = T^{4s}$ , which completes the proof. □

*Besov spaces by wavelets coefficients* The Besov norm can also be expressed by means of wavelet coefficients. Let

$$F_j(\lambda) := \sqrt{\lambda} G_j(\lambda),$$

where  $G_j$  is a filter as in Definition 4.2. The partition of unity (9) becomes

$$\sum_{j \geq 0} F_j(\lambda)^2 = 1 \quad \text{for all } \lambda \in (0, \kappa^2]. \tag{38}$$

Moreover, in view of (18), for a frame  $\Psi$  as in Definition 4.6 we have

$$\| \langle f, \psi_{j \cdot} \rangle \|_{L^2(\mathcal{X}, \rho)} = \| F_j(T) f \|_{\mathcal{H}},$$

and the frame property (16) can be rewritten as

$$\| f \|_{\mathcal{H}}^2 = \sum_{j \geq 0} \| F_j(T) f \|_{\mathcal{H}}^2. \tag{39}$$

If we further assume the localization property (cf. Example 4.5)

$$\text{supp}(F_0) \subset (2^{-1}, \infty), \quad \text{supp}(F_j) \subset (2^{-j-1}, 2^{-j+1}) \quad \text{for } j \geq 1, \tag{40}$$

a weighted  $\ell^q$ -norm of the sequence  $(\| F_j(T) f \|_{\mathcal{H}})_{j \geq 0}$  gives an equivalent characterization of the space  $\mathcal{B}_q^s$ .

**Proposition 8.6** ([20, Theorem 3.18]) *Let  $\{F_j\}_{j \geq 0}$  be a family of measurable functions  $F_j : [0, \infty) \rightarrow [0, \infty)$  satisfying (38) and (40). Then, for every  $f \in \mathcal{B}_q^s$  we have*

$$\| f \|_{\mathcal{B}_q^s} \asymp \| f \|_{\mathcal{H}} + \left( \sum_{j \geq 0} \left( 2^{js} \| F_j(T) f \|_{\mathcal{H}} \right)^q \right)^{1/q}.$$

**Proof** We upper and lower bound the discretized norm in (31). Using (39) (which holds thanks to (38)) and (40), we have

$$\begin{aligned} \mathcal{E}(f, 2^\ell)^2 &= \|\mathbf{P}_{\mathbf{PW}(2^\ell)^\perp} f\|_{\mathcal{H}}^2 = \sum_{j \geq 0} \|F_j(\mathbf{T})\mathbf{P}_{\mathbf{PW}(2^\ell)^\perp} f\|_{\mathcal{H}}^2 \\ &= \sum_{j \geq 0} \sum_{i \in \mathcal{I}_\rho} \left| \langle F_j(\mathbf{T})\mathbf{P}_{\mathbf{PW}(2^\ell)^\perp} f, v_i \rangle_{\mathcal{H}} \right|^2 = \sum_{j \geq 0} \sum_{i \in \mathcal{I}_\rho} \left| \langle \mathbf{P}_{\mathbf{PW}(2^\ell)^\perp} f, F_j(\mathbf{T})v_i \rangle_{\mathcal{H}} \right|^2 \\ &= \sum_{j \geq 0} \sum_{\substack{\lambda_i < 2^{-\ell} \\ \lambda_i \in (2^{-j-1}, 2^{-j+1})}} F_j(\lambda_i)^2 |\langle f, v_i \rangle_{\mathcal{H}}|^2 = \sum_{j \geq \ell} \sum_{\lambda_i \in (2^{-j-1}, 2^{-j+1})} F_j(\lambda_i)^2 |\langle f, v_i \rangle_{\mathcal{H}}|^2 \\ &= \sum_{j \geq \ell} \sum_{i \in \mathcal{I}_\rho} F_j(\lambda_i)^2 |\langle f, v_i \rangle_{\mathcal{H}}|^2 = \sum_{j \geq \ell} \|F_j(\mathbf{T})f\|_{\mathcal{H}}^2. \end{aligned}$$

Thus, by the discrete Hardy inequality (Lemma A.3), we get

$$\begin{aligned} \left( \sum_{\ell \geq 0} (2^{\ell s} \mathcal{E}(f, 2^\ell))^q \right)^{1/q} &\leq \left( \sum_{\ell \geq 0} \left( 2^{\ell s} \sum_{j \geq \ell} \|F_j(\mathbf{T})f\|_{\mathcal{H}} \right)^q \right)^{1/q} \\ &\leq C_{sq} \left( \sum_{j \geq 0} \left( 2^{js} \|F_j(\mathbf{T})f\|_{\mathcal{H}} \right)^q \right)^{1/q}, \end{aligned}$$

with  $C_{sq} = \frac{2^{sq}}{2^{sq}-1}$ . Conversely,  $F_j(\mathbf{T})g = 0$  for every  $g \in \mathbf{PW}(2^j)$ , and therefore

$$\|F_j(\mathbf{T})f\|_{\mathcal{H}} = \|F_j(\mathbf{T})(f - g)\|_{\mathcal{H}} \leq \|F_j(\mathbf{T})\|_{\mathcal{H}} \|f - g\|_{\mathcal{H}} \leq \|f - g\|_{\mathcal{H}},$$

whence

$$\|F_j(\mathbf{T})f\|_{\mathcal{H}} \leq \inf_{g \in \mathbf{PW}(2^j)} \|f - g\|_{\mathcal{H}} = \mathcal{E}(f, 2^j).$$

□

*Convergence of spectrally-localized Monte Carlo wavelets* Proposition 8.6 can be used to obtain approximation bounds for frames built with filters satisfying the localization property (40).

**Proposition 8.7** *Under the conditions of Proposition 8.6, for every  $f \in \mathcal{B}_q^s$  and  $\epsilon \in (0, s)$ , we have*

$$\left\| \sum_{j > \tau} \mathbf{T}_j f \right\|_{\mathcal{H}} \lesssim \begin{cases} \|f\|_{\mathcal{B}_q^s} 2^{-\tau s} & \text{for } q \in [1, 2] \\ \|f\|_{\mathcal{B}_2^{s-\epsilon}} 2^{-\tau(s-\epsilon)} & \text{for } q \in (2, \infty] \end{cases}.$$

**Proof** By Proposition 8.6, we have

$$\sum_{j > \tau} \|F_j(\mathbf{T})f\|_{\mathcal{H}}^q = \sum_{j > \tau} 2^{-jsq} \left( 2^{js} \|F_j(\mathbf{T})f\|_{\mathcal{H}} \right)^q \lesssim 2^{-(\tau+1)sq} \|f\|_{\mathcal{B}_q^s}^q.$$



Also, (38) implies  $\left| \sum_j F_j(\lambda_i)^2 \right|^2 \leq \sum_j F_j(\lambda_i)^2$ . Hence, for  $q \leq 2$  we obtain

$$\begin{aligned} \left\| \sum_{j>\tau} T_j f \right\|_{\mathcal{H}}^2 &= \sum_{i \in \mathcal{I}_\rho} \left| \sum_{j>\tau} F_j(\lambda_i)^2 \right|^2 |\langle f, v_i \rangle_{\mathcal{H}}|^2 \leq \sum_{j>\tau} \sum_{i \in \mathcal{I}_\rho} F_j(\lambda_i)^2 |\langle f, v_i \rangle_{\mathcal{H}}|^2 \\ &= \sum_{j>\tau} \|F_j(\mathbf{T})f\|_{\mathcal{H}}^2 = \left\| \left( \|F_{\tau+j}(\mathbf{T})f\| \right)_{j \geq 1} \right\|_{\ell^2}^2 \\ &\leq \left\| \left( \|F_{\tau+j}(\mathbf{T})f\| \right)_{j \geq 1} \right\|_{\ell^q}^2 \leq \left( 2^{-(\tau+1)s} \|f\|_{\mathcal{B}_q^s} \right)^2. \end{aligned}$$

If  $q > 2$ , then  $\mathcal{B}_q^s \subset \mathcal{B}_2^{s-\epsilon}$  for every  $\epsilon \in (0, s)$ , thanks to (32), and the claim follows. □

Putting together Propositions 8.7 and 7.2 yields a convergence result for Monte Carlo wavelets with localized filters.

**Theorem 8.8** *Assume that  $F_j$  satisfies (40),  $f \in \mathcal{B}_q^s$  with  $q \in [1, 2]$ , and  $\lambda \mapsto \lambda g_\tau(\lambda)$  is Lipschitz continuous on  $[0, \kappa^2]$  with Lipschitz constant  $L(\tau) \lesssim 2^\tau$ . Set*

$$\tau = \lceil \frac{1}{2s+2} \log_2(N) \rceil.$$

Then, for every  $t > 0$ , with probability at least  $1 - 2e^{-t}$  we have

$$\|f - \widehat{f}_{\tau, N}\|_{\mathcal{H}} \lesssim \|f\|_{\mathcal{B}_q^s} (1 + \kappa^2 \sqrt{t}) N^{-\frac{s}{2s+2}}.$$

Compared to Theorem 7.5, Theorem 8.8 requires the resolution  $\tau$  to grow only logarithmically with respect to the sample size  $N$ . Note that the conditions of Theorem 8.8 exclude the spectral functions of Table 2, since they do not satisfy (40). Examples of admissible filters are given instead by Example 4.5, which have local support (40) but exponential Lipschitz constant.

### 9 Concluding Remarks and Future Directions

We presented a construction of tight frames which extends wavelets on general domains based on spectral filtering of a reproducing kernel. Depending on the measure considered, our construction leads to continuous or discrete frames, covering non-Euclidean structures such as Riemannian manifolds and weighted graphs. Besides standard frequency-localized filters commonly used in wavelet frames, we defined admissible spectral filters resorting to methods from regularization theory, such as Tikhonov regularization and Landweber iteration. Regarding discrete measures as empirical measures arising from independent realizations of a continuous density, we interpreted discrete frames as Monte Carlo estimates of continuous frames. We proved that the Monte Carlo frame converges to the corresponding deterministic continuous frame, and provided finite-sample bounds in high probability, with rates that depend

on the Sobolev or Besov class of the reproduced signal. This demonstrates the stability of empirical frames built on sampled data.

In future work we intend to study the numerical implementation of our Monte Carlo wavelets, along with possible applications in graph signal processing, regression analysis and denoising. Further theoretical investigation may include  $L^p$  Banach frame extensions, sparse representations, nonlinear approximation rates, Lipschitz bound refinements, and explicit localization properties for specific families of kernels.

**Acknowledgements** Part of this work has been carried out at the Machine Learning Genoa (MaLGA) center, Università di Genova (IT). Ernesto De Vito is part of the Machine Learning Genoa Center (MaLGA) and he is a member of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM). LR and SV acknowledge the financial support of the European Research Council (Grant SLING 819789), the AFOSR projects FA9550-17-1-0390 and BAA-AFRL-AFOSR-2016-0007 (European Office of Aerospace Research and Development), and the EU H2020-MSCA-RISE project NoMADS - DLV-777826. VN and ZK acknowledge the support from RCN-funded FunDaHD Project No. 251149/O70.

**Funding** Open access funding provided by Università degli Studi di Genova within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A

We recall the following result, whose proof can be collected from [23].

**Lemma A.1** *Let  $(\Omega; \mu)$  be a measure space and  $\mathcal{H}$  a Hilbert space. Given a weakly measurable mapping  $\omega \mapsto \Psi_\omega$  from  $\Omega$  to  $\mathcal{H}$ , assume there exists a dense subset  $\mathcal{D} \subset \mathcal{H}$ , and a constant  $C > 0$ , such that, for every  $f \in \mathcal{D}$ ,*

$$\int_{\Omega} |\langle f, \Psi_\omega \rangle_{\mathcal{H}}|^2 d\mu(\omega) \leq C \|f\|_{\mathcal{H}}^2. \quad (41)$$

*Then (41) holds for every  $f \in \mathcal{H}$ . Furthermore, there exists a positive bounded operator  $A : \mathcal{H} \rightarrow \mathcal{H}$  such that, for every  $f, g \in \mathcal{H}$ ,*

$$\langle Af, g \rangle_{\mathcal{H}} = \int_{\Omega} \langle f, \Psi_\omega \rangle_{\mathcal{H}} \langle \Psi_\omega, g \rangle_{\mathcal{H}} d\mu(\omega).$$

**Proof** For  $f \in \mathcal{H}$ , define the measurable mapping

$$\mathbf{V}f : \Omega \rightarrow \mathbb{C} \quad \mathbf{V}f(\omega) := \langle f, \Psi_\omega \rangle_{\mathcal{H}}.$$

Let  $\mathcal{S} := \{f \in \mathcal{H} : \mathbf{V}f \in L^2(\Omega; \mu)\}$ . The subspace  $\mathcal{S}$  is dense in  $\mathcal{H}$  since  $\mathcal{S} \supset \mathcal{D}$ , and the operator  $\mathbf{V} : \mathcal{S} \rightarrow L^2(\Omega; \mu)$  is closed. Indeed, fix a sequence  $(f_n) \subset \mathcal{S}$  converging to  $f \in \mathcal{H}$  and such that  $(\mathbf{V}f_n)$  converges to  $F \in L^2(\Omega; \mu)$ . Then, possibly passing to a subsequence, there is a subset  $E \subset \Omega$  of measure zero such that, for all  $\omega \notin E$ ,

$$F(\omega) = \lim_{n \rightarrow \infty} \mathbf{V}f_n(\omega) = \lim_{n \rightarrow \infty} \langle f_n, \Psi_\omega \rangle_{\mathcal{H}} = \langle f, \Psi_\omega \rangle_{\mathcal{H}}.$$

Then  $f \in \mathcal{S}$  and  $F = \mathbf{V}f$ . Moreover,

$$\|\mathbf{V}f\|_{L^2(\Omega; \mu)}^2 = \lim_{n \rightarrow \infty} \|\mathbf{V}f_n\|_{L^2(\Omega; \mu)}^2 \leq C \lim_{n \rightarrow \infty} \|f_n\|_{\mathcal{H}}^2 = C \|f\|_{\mathcal{H}}^2.$$

Thus,  $\mathbf{V}$  is a bounded operator, and the closed graph theorem implies  $\mathcal{S} = \mathcal{H}$ , i.e. (41) holds for all  $f \in \mathcal{H}$ . The second statement follows by defining  $\mathbf{A} := \mathbf{V}^*\mathbf{V}$ .  $\square$

The simple proof of the following bound is due to A. Maurer.

**Lemma A.2** *Let  $\mathbf{A}, \mathbf{B}$  be self-adjoint operators on a separable Hilbert space  $\mathcal{H}$ , and let  $F : \mathbb{R} \rightarrow \mathbb{C}$  be a Lipschitz continuous function with Lipschitz constant  $L$ . Then*

$$\|F(\mathbf{A}) - F(\mathbf{B})\|_{\text{HS}} \leq L\|\mathbf{A} - \mathbf{B}\|_{\text{HS}}.$$

**Proof** Let  $\{e_i\}_{i \in \mathcal{I}}$  and  $\{f_j\}_{j \in \mathcal{J}}$  be orthonormal bases of  $\mathcal{H}$  such that  $\mathbf{A}e_i = \lambda_i e_i$  and  $\mathbf{B}f_j = \mu_j f_j$ . Then

$$\begin{aligned} \|F(\mathbf{A}) - F(\mathbf{B})\|_{\text{HS}}^2 &= \sum_{i \in \mathcal{I}, j \in \mathcal{J}} | \langle (F(\mathbf{A}) - F(\mathbf{B}))e_i, f_j \rangle_{\mathcal{H}} |^2 \\ &= \sum_{i \in \mathcal{I}, j \in \mathcal{J}} |F(\lambda_i) - F(\mu_j)|^2 | \langle e_i, f_j \rangle_{\mathcal{H}} |^2 \\ &\leq L^2 \sum_{i \in \mathcal{I}, j \in \mathcal{J}} |\lambda_i - \mu_j|^2 | \langle e_i, f_j \rangle_{\mathcal{H}} |^2 = L^2 \|\mathbf{A} - \mathbf{B}\|_{\text{HS}}. \end{aligned}$$

$\square$

We include a proof of the discrete Hardy inequality [16, Eq. 5.2] where we explicitly compute the Hardy constant.

**Lemma A.3** (Hardy inequality) *Let  $(b_j)_{j \geq 0}$  and  $(a_j)_{j \geq 0}$  be two sequences such that*

$$|b_j| \leq \left( \sum_{k \geq j} |a_k|^p \right)^{1/p} \text{ for } 0 < p \leq q.$$

*Then, for every  $s > 0$ , we have*

$$\sum_{j \geq 0} \left( 2^{js} |b_j| \right)^q \leq \frac{2^{sq}}{2^{sq} - 1} \sum_{j \geq 0} \left( 2^{js} |a_j| \right)^q,$$

provided all the sums are finite.

**Proof** Let  $\alpha = \frac{q}{p}$ , and let  $\beta$  be such that  $sp > \beta > 0$ . Since  $p \leq q$ , we have  $\|\cdot\|_{\ell_q} \leq \|\cdot\|_{\ell_p}$ , hence

$$\begin{aligned} \sum_{j \geq 0} \left(2^{js} |b_j|\right)^q &\leq \sum_{j \geq 0} 2^{jsq} \left\| (a_{j+k})_{k \geq 0} \right\|_{\ell_q}^q \\ &\leq \sum_{j \geq 0} 2^{jsq} \left\| (a_{j+k})_{k \geq 0} \right\|_{\ell_p}^q = \sum_{j \geq 0} 2^{jsq} \left( \sum_{k \geq j} 2^{-k\beta} 2^{k\beta} |a_k|^p \right)^\alpha. \end{aligned}$$

Assume now  $\alpha \in (1, \infty)$ . Applying the Hölder inequality with  $1/\alpha + 1/\alpha' = 1$  we have

$$\begin{aligned} \sum_{k \geq j} \left(2^{-k\beta}\right) \left(2^{k\beta} |a_k|^p\right) &\leq \left(\sum_{k \geq j} 2^{-k\beta\alpha'}\right)^{1/\alpha'} \left(\sum_{k \geq j} 2^{k\beta\alpha} |a_k|^{\alpha p}\right)^{1/\alpha} \\ &= \frac{2^\beta}{(2^{\beta\alpha'} - 1)^{1/\alpha'}} 2^{-j\beta} \left(\sum_{k \geq j} 2^{k\beta\alpha} |a_k|^{\alpha p}\right)^{1/\alpha}. \end{aligned}$$

Plugging this in and using  $\alpha p = q$  we get

$$\begin{aligned} \sum_{j \geq 0} \left(2^{js} |b_j|\right)^q &\leq C_1 \sum_{j \geq 0} 2^{jsq} 2^{-j\beta\alpha} \left(\sum_{k \geq j} 2^{k\beta\alpha} |a_k|^{\alpha p}\right) \\ &= C_1 \sum_{j \geq 0} 2^{j(sq-\beta\alpha)} \left(\sum_{k \geq j} 2^{k\beta\alpha} |a_k|^q\right), \end{aligned}$$

with

$$C_1 := \frac{2^{\alpha\beta}}{(2^{\beta\alpha'} - 1)^{\alpha/\alpha'}}.$$

Changing the order of summation we get

$$\begin{aligned} \sum_{j \geq 0} \left(2^{js} |b_j|\right)^q &\leq C_1 \sum_{j \geq 0} 2^{j\beta\alpha} |a_j|^q \sum_{k \leq j} 2^{k(sq-\beta\alpha)} \\ &\leq C_1 C_2 \sum_{j \geq 0} 2^{jsq} |a_j|^q = C_1 C_2 \sum_{j \geq 0} \left(2^{js} |a_j|\right)^q, \end{aligned}$$

with

$$C_2 := \frac{2^{sq-\beta\alpha}}{2^{sq-\beta\alpha} - 1},$$

since

$$\sum_{k \leq j} 2^{k(sq - \beta\alpha)} = \frac{1}{2^{sq - \beta\alpha} - 1} \left( 2^{(j+1)(sq - \beta\alpha)} - 1 \right) \leq \frac{2^{sq - \beta\alpha}}{2^{sq - \beta\alpha} - 1} 2^{j(sq - \beta\alpha)}.$$

We have

$$C_1 C_2 = \frac{2^{sq}}{(2^{\beta\alpha'} - 1)^{\alpha'/\alpha'} (2^{sq - \beta\alpha} - 1)}.$$

If  $\alpha = 1$  ( $p = q$ ), then  $\alpha' = \infty$ , in which case  $C_1 = 1$  and therefore

$$C_1 C_2 = \frac{2^{sq - \beta}}{2^{sq - \beta} - 1}$$

for all  $\beta \in (0, sq)$ . Thus, we may set  $C := \frac{2^{sq}}{2^{sq} - 1}$ . □

## References

1. Ali, S., Antoine, J., Gazeau, J.: Continuous frames in Hilbert space. *Ann. Phys.* **222**(1), 1–37 (1993)
2. Ali, S., Antoine, J., Gazeau, J.: *Coherent States, Wavelets, and Their Generalizations*. Theoretical and Mathematical Physics. Springer, New York (2013)
3. Belkin, M., Niyogi, P.: Convergence of Laplacian eigenmaps. *Adv. Neural Inf. Process. Syst.* **19**, 129–136 (2007)
4. Belkin, M., Niyogi, P.: Towards a theoretical foundation for Laplacian-based manifold methods. *J. Comput. Syst. Sci.* **74**(8), 1289–1308 (2008)
5. Binev, P., Cohen, A., Dahmen, W., Devore, R.A., Temlyakov, V.N.: Universal algorithms for learning theory Part I: piecewise constant functions. *J. Mach. Learn. Res.* **6**, 1297–1321 (2005)
6. Blanchard, G., Mücke, N.: Optimal rates for regularization of statistical inverse learning problems. *Found. Comput. Math.* **18**(4), 971–1013 (2018)
7. Camoriano, R., Angles, T., Rudi, A., Rosasco, L.: NYTRO: when subsampling meets early stopping. In: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, vol. 51, pp. 1403–1411 (2016)
8. Caponnetto, A., De Vito, E.: Optimal rates for the regularized least-squares algorithm. *Found. Comput. Math.* **7**, 331–368 (2007)
9. Chen, J., Avron, H., Sindhvani, V.: Hierarchically compositional kernels for scalable nonparametric learning. *J. Mach. Learn. Res.* **18**(66), 1–42 (2017)
10. Chui, C.K.: *An Introduction to Wavelets, Wavelet Analysis and its Applications*, vol. 1. Academic Press, Boston (1992)
11. Chung, F.R.K.: *Spectral Graph Theory*. American Mathematical Society, Providence (1997)
12. Coifman, R.R., Maggioni, M.: Diffusion wavelets. *Appl. Comput. Harmonic Anal.* **21**, 53–94 (2006)
13. Coulhon, T., Kerkycharian, G., Petrushev, P.: Heat kernel generated frames in the setting of Dirichlet spaces. *J. Fourier Anal. Appl.* **18**(5), 995–1066 (2012)
14. Daubechies, I.: *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia (1992)
15. Daubechies, I., Grossmann, A., Meyer, Y.: Painless nonorthogonal expansions. *J. Math. Phys.* **27**, 1271–1283 (1986)
16. DeVore, R.A., Popov, V.A.: Interpolation of Besov spaces. *Trans. Am. Math. Soc.* **305**(1), 397–414 (1988)
17. Ding, Y., Kondor, R., Eskreis-Winkler, J.: Multiresolution kernel approximation for Gaussian process regression. *Adv. Neural Inf. Process. Syst.* **30**, 3740–3748 (2017)

18. Dong, B.: Sparse representation on graphs by tight wavelet frames and applications. *Appl. Comput. Harmonic Anal.* **42**(3), 452–479 (2017)
19. Engl, H.W., Hanke, M., Neubauer, A.: *Regularization of Inverse Problems, Mathematics and Its Applications*, vol. 375. Kluwer, Dordrecht (1996)
20. Feichtinger, H.G., Führ, H., Pesenson, I.: Geometric space-frequency analysis on manifolds. *J. Fourier Anal. Appl.* **22**(6), 1294–1355 (2016)
21. Fornasier, M., Rauhut, H.: Continuous frames, function spaces, and the discretization problem. *J. Fourier Anal. Appl.* **11**(3), 245–287 (2005)
22. Freeman, D., Speegle, D.: The discretization problem for continuous frames. *Adv. Math.* **345**, 784–813 (2019)
23. Führ, H.: *Abstract Harmonic Analysis of Continuous Wavelet Transforms. Lecture Notes in Mathematics*. Springer, Berlin (2005)
24. Führ, H., Gröchenig, K.: Sampling theorems on locally compact groups from oscillation estimates. *Math. Z.* **255**(1), 177–194 (2007)
25. García Trillos, N., Gerlach, M., Hein, M., Slepčev, D.: Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator. *Found. Comput. Math.* **20**(4), 827–887 (2020)
26. Geller, D., Pesenson, I.Z.: Band-limited localized Parseval frames and Besov spaces on compact homogeneous manifolds. *J. Geom. Anal.* **21**, 334–371 (2011)
27. Giné, E., Koltchinskii, V.: Empirical graph Laplacian approximation of Laplace–Beltrami operators: large sample results. *High Dimens. Probab.* **51**, 238–259 (2006)
28. Göbel, F., Blanchard, G., von Luxburg, U.: Construction of tight frames on graphs and application to denoising. In: Härdle, W.K., Lu, H.H.S., Shen, X. (eds.) *Handbook of Big Data Analytics*, pp. 503–522. Springer, Berlin (2018)
29. Gröchenig, K.: Describing functions: atomic decompositions versus frames. *Monat. Math.* **112**, 1–42 (1991)
30. Grossmann, A., Morlet, J.: Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.* **15**, 723–736 (1984)
31. Haar, A.: Zur Theorie der orthogonalen Funktionensysteme.(Erste Mitteilung). *Math. Ann.* **69**, 331–371 (1910)
32. Haiko, N., Martinsson, P.G., Tropp, J.A.: Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* **53**(42), 217–288 (2011)
33. Hammond, D.K., Vandergheynst, P., Gribonval, R.: Wavelets on graphs via spectral graph theory. *Appl. Comput. Harmonic Anal.* **30**(2), 129–150 (2011)
34. Hein, M., Audibert, J., von Luxburg, U.: From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians. In: *Proceedings of the 18th Conference on Learning Theory*, pp. 470–485 (2005)
35. Kereta, Z., Vigogna, S., Naumova, V., Rosasco, L., De Vito, E.: Monte Carlo wavelets: a randomized approach to frame discretization. In: *International Conference on Sampling Theory and Applications*, vol. 13 (2019)
36. Koltchinskii, V., Lounici, K.: Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23**(1), 110–133 (2017)
37. von Luxburg, U., Belkin, M., Bousquet, O.: Consistency of spectral clustering. *Ann. Stat.* **36**(2), 555–586 (2008)
38. Maggioni, M., Mhaskar, H.N.: Diffusion polynomial frames on metric measure spaces. *Appl. Comput. Harmonic Anal.* **24**(3), 329–353 (2008)
39. Mallat, S.: *A Wavelet Tour of Signal Processing*. Elsevier, Amsterdam (1999)
40. Martinsson, P.G., Tropp, J.A.: *Randomized numerical linear algebra: foundations & algorithms* (2020). [arXiv:2002.01387](https://arxiv.org/abs/2002.01387)
41. Meyer, Y.: *Wavelets and Operators. Cambridge Studies in Advanced Mathematics*, vol. 37. Cambridge University Press, Cambridge (1992)
42. Mhaskar, H.N.: Eignets for function approximation on manifolds. *Appl. Comput. Harmonic Anal.* **29**(1), 63–87 (2010)
43. Pagliana, N., Rosasco, L.: Implicit regularization of accelerated methods in Hilbert spaces. *Adv. Neural Inf. Process. Syst.* **32**, 14481–14491 (2019)

44. Pesenson, I., Le Gia, Q.T., Mayeli, A., Mhaskar, H., Zhou, D.X. (eds.): *Frames and Other Bases in Abstract and Function Spaces. Novel Methods in Harmonic Analysis. Applied and Numerical Harmonic Analysis*, vol. 1. Birkhäuser Basel, Boston (2017)
45. Pietsch, A.: Approximation spaces. *J. Approx. Theory* **32**, 115–134 (1981)
46. Rosasco, L., Belkin, M., De Vito, E.: On learning with integral operators. *J. Mach. Learn. Res.* **11**, 905–934 (2010)
47. Rudi, A., Camoriano, R., Rosasco, L.: Less is more: Nyström computational regularization. *Adv. Neural Inf. Process. Syst.* **28**, 1657–1665 (2015)
48. Rudi, A., Rosasco, L.: Generalization properties of learning with random features. *Adv. Neural Inf. Process. Syst.* **30**, 3215–3225 (2017)
49. Rudi, A., Carratino, L., Rosasco, L.: FALKON: an optimal large scale kernel method. *Adv. Neural Inf. Process. Syst.* **30**, 3888–3898 (2017)
50. Schäfer, F., Sullivan, T.J., Owhadi, H.: Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity (2017). [arXiv:1706.02205](https://arxiv.org/abs/1706.02205)
51. Schrödinger, E.: Der stetige übergang von der mikro- zur makromechanik. *Naturwissenschaften* **14**(28), 664–666 (1926)
52. Shishkin, S.L., Shalaginov, A., Bopardikar, S.D.: Fast approximate truncated SVD. *Numer. Linear Algebra Appl.* **26**(4), e2246 (2019)
53. Singer, A.: From graph to manifold Laplacian: the convergence rate. *Appl. Comput. Harmonic Anal.* **21**(1), 128–134 (2006)
54. Singer, A., Wu, H.T.: Spectral convergence of the connection Laplacian from random samples. *Inf. Inference* **6**, 58–123 (2017)
55. Ting, D., Huang, L., Jordan, M.I.: An analysis of the convergence of graph Laplacians. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pp. 1079–1086 (2010)
56. Triebel, H.: *Theory of Function Spaces II. Monographs in Mathematics*. Birkhäuser, Basel (1992)
57. Wang, Y.G., Zhuang, X.: Tight framelets and fast framelet filter bank transforms on manifolds. *Appl. Comput. Harmonic Anal.* **48**(1), 64–95 (2020)
58. Wendland, H.: Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.* **4**, 389–396 (1995)
59. Yao, Y., Rosasco, L., Caponnetto, A.: On early stopping in gradient descent learning. *Construct. Approx.* **26**, 289–315 (2007)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.