Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn
Wegelerstraße 6 • 53115 Bonn • Germany
phone +49 228 73-3427 • fax +49 228 73-7527
www.ins.uni-bonn.de

M. Griebel, C. Rieger and B. Zwicknagl

# Multiscale approximation and reproducing kernel Hilbert space methods

INS Preprint No. 1312

2013

# MULTISCALE APPROXIMATION AND REPRODUCING KERNEL HILBERT SPACE METHODS

MICHAEL GRIEBEL*, CHRISTIAN RIEGER†, AND BARBARA ZWICKNAGL‡

**Abstract.** We consider reproducing kernels $K : \Omega \times \Omega \to \mathbb{R}$ in multiscale series expansion form, i.e., kernels of the form $K(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\ell \in \mathbb{N}} \lambda_\ell \sum_{j \in I_\ell} \phi_{\ell,j}(\boldsymbol{x}) \phi_{\ell,j}(\boldsymbol{y})$ with weights $\lambda_\ell$ and structurally simple basis functions $\{\phi_{\ell,i}\}$. Here, we deal with basis functions such as polynomials or frame systems, where, for $\ell \in \mathbb{N}$, the index set $I_\ell$ is finite or countable. We derive relations between approximation properties of spaces based on basis functions $\{\phi_{\ell,j} : 1 \leq \ell \leq L, j \in I_\ell\}$ and spaces spanned by translates of the kernel span$\{K(\boldsymbol{x}_1, \cdot), \ldots, K(\boldsymbol{x}_N, \cdot)\}$ with $X_N := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$ if the truncation index $L$ is appropriately coupled to the discrete set $X_N$. An analysis of a numerically feasible approximation from trial spaces span$\{K^L(\boldsymbol{x}_1, \cdot), \ldots, K^L(\boldsymbol{x}_N, \cdot)\}$ based on finitely truncated series kernels of the form $K^L(\boldsymbol{x}, \boldsymbol{y}) := \sum_{\ell=1}^{L} \lambda_\ell \sum_{j \in I_\ell} \phi_{\ell,j}(\boldsymbol{x}) \phi_{\ell,j}(\boldsymbol{y})$ is provided where the truncation index $L$ is chosen sufficiently large depending on the point set $X_N$. Furthermore, Bernstein-type inverse estimates and derivative-free sampling inequalities for kernel based spaces are obtained from estimates for spaces based on the basis functions $\{\phi_{\ell,j} : 1 \leq \ell \leq L, j \in I_\ell\}$.

**Key words.** Reproducing kernel Hilbert spaces, multiscale expansion, a priori error estimates, Bernstein estimates

**AMS subject classifications.** 41A05, 41A17, 41A25, 41A58, 65D05

**1. Introduction.** In many applications, one seeks to reconstruct or to approximate a function from given (measured) strong or weak data. Kernel-based methods have proven a reliable and useful tool in a large variety of such problems, including meshless methods for the solution of partial differential equations, surface reconstruction and machine learning algorithms (see [4, 5, 6, 7, 24, 25, 28]).

In kernel-based meshless methods one typically considers trial spaces that are based on a finite discrete set $X_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$ of data points. They are spanned by translates of a given kernel function $K : \Omega \times \Omega \to \mathbb{R}$, i.e., we use

$$\mathcal{L}_{X_N} = \text{span}\{K(\boldsymbol{x}_1, \cdot), \ldots, K(\boldsymbol{x}_N, \cdot)\}. \tag{1.1}$$

The approximation properties of such spaces are well understood and various estimates for interpolation and approximation with $\mathcal{L}_{X_N}$ can be found in the literature, see [28] and the references cited therein.

Typically, one assumes a closed form representation of the kernel by a simple analytical formula. Popular examples are the Gaussian function and surface splines. However, to our opinion, the restriction to only kernels in closed form is a severe limitation. Kernels as infinite series expansions are important as well and indeed appear in many practical applications.

First, mathematical modeling naturally introduces a function space for which one wants to construct the associated reproducing kernel. While for certain function spaces, including spaces which are isomorphic to Sobolev spaces over product domains, there are closed form expressions for the kernel available (see [28] and the references therein), this is no longer the case for more general situations involving arbitrary domains and boundary conditions.

Second, kernels are sometimes introduced to characterize certain properties or features of the function space under consideration. Typical examples are kernels induced by feature maps in machine learning (see [25]) or wavelet-type function systems (see [16, 17]) often used in numerical analysis. Thus, the kernel is a priori given as an infinite (multilevel) series expansion by means of multiscale basis functions. Then, it is unlikely that a closed analytical formula can be found for such a series representation.

*Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6-8, 53115 Bonn, Germany
†Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6-8, 53115 Bonn, Germany
‡Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany

In this paper we thus consider kernels of the form

$$K(\boldsymbol{x},\boldsymbol{y}) = \sum_{\ell\in\mathbb{N}} \lambda_\ell \sum_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x})\,\phi_{\ell,j}(\boldsymbol{y}), \quad \boldsymbol{x},\boldsymbol{y}\in\Omega\subset\mathbb{R}^d, \tag{1.2}$$

with functions $\phi_{\ell,j}$ and weights $\lambda_\ell \geq 0$, where the $I_\ell$ denote suitable finite or countably infinite index sets for $\ell\in\mathbb{N}$. The first index $\ell$ refers to the level while the second index $j$ labels the basis functions within a level. Here, we assume that the weights $\lambda_\ell$ depend only on the level $\ell$ but not on $j$.[1] Furthermore, we assume that $\left\{\sqrt{\lambda_\ell}\phi_{\ell,j}\right\}$ forms a frame in a Hilbert space $\mathcal{H}$. This covers most of the interesting standard function spaces and includes especially Sobolev spaces of continuous functions (see [15]). Our analysis reproduces the known results in that case.

Since the infinite series in (1.2) cannot be dealt with in a practical application, it must be truncated somehow. Here, the choice of the truncation parameter is a crucial issue. Although there is some numerical experience (see [8]), to the best of our knowledge, so far no theoretical justification has been provided of how to perform the truncation properly. This issue will be the main subject of this article. Based on a careful truncation, we will first derive direct Jackson-type approximation theorems for kernel-based interpolation and approximation in the limit of dense data locations. Then, we will consider approximation and interpolation with truncated kernels and we will balance the associated truncation error with the interpolation error that stems from the data $X_N$. To this end, for a kernel $K$ with decomposition (1.2), we consider truncated index sets with associated trial spaces

$$\mathcal{H}_L := \Big\{ f = \sum_{\ell=1}^{L}\sum_{j\in I_\ell} a_{\ell,j}\lambda_\ell^{1/2}\phi_{\ell,j} : \sum_{\ell=1}^{L}\sum_{j\in I_\ell} |a_{\ell,j}|^2 < \infty \Big\}. \tag{1.3}$$

The truncated trial spaces are associated to the kernel $K^L$ with truncated series expansion

$$K^L(\boldsymbol{x},\boldsymbol{y}) := \sum_{\ell=1}^{L} \lambda_\ell \sum_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x})\,\phi_{\ell,j}(\boldsymbol{y}), \tag{1.4}$$

which involves all basis functions of the first $L$ levels.

We will show that for kernels of the form (1.2) there is a subtle interplay between the approximation properties of $\mathcal{H}_L$ and $\mathcal{L}_{X_N}$ if the truncation index $L\in\mathbb{N}$ is appropriately coupled to the set of data points $X_N$. This coupling is roughly given by (see Lemma 2.1)

$$\sum_{\ell=L+1}^{\infty} \lambda_\ell \lesssim \frac{\sigma_{\min}\left(G_{X_N,X_N}\right)}{N}, \tag{1.5}$$

where $\sigma_{\min}\left(G_{X_N,X_N}\right)$ denotes the smallest eigenvalue of the Gramian matrix which results from the full kernel K of (1.2) by evaluation in the points of $X_N$, i.e., $\left(G_{X_N,X_N}\right)_{i,j} = K(\boldsymbol{x}_n,\boldsymbol{x}_m)$ for $1\leq n,m\leq N$. Note that the left-hand side of (1.5) depends only on the decay of the level-weights $\lambda_\ell$, whereas the right-hand side depends only on the kernel and the point set $X_N$.

To be precise, following techniques from [13] we derive the existence of interpolants based on the kernel (1.4) for data on $X_N\subset\Omega$ if the truncation index $L$ is chosen according to (1.5). We show furthermore that the interpolation error is increased only by a constant factor compared to that obtained with the full series

[1] Note that $I_\ell$ is allowed to consist of one element only. Then, the double sum in (1.2) reduces to a single sum in $\ell$.

kernel (1.2). Note here that the evaluation costs of the kernel $K^L$ depend on $L$ which in turn depends on the data points $X_N$ and can hence not be regarded to be constant. This dependency, however, can be properly controlled as we will show in the following sections.[2]

In the important case of Sobolev spaces, we will obtain the error estimate

$$\left\| f - s^L_{f,X_N}(f) \right\|_{L^2(\Omega)} \leq C h^{\tau}_{X_N,\Omega} \left\| f \right\|_{W^{\tau}_2(\Omega)},$$

for the interpolation based on the truncated kernel (1.4), where $h_{X_N,\Omega} = \sup_{\boldsymbol{x} \in \Omega} \min_{\boldsymbol{x}_j \in X_N} \|\boldsymbol{x} - \boldsymbol{x}_j\|_2$ denotes the fill-distance of the discrete set $X_N$ in $\Omega$.

We furthermore give a priori error estimates for general reconstruction processes. As a result, we will derive derivative-free sampling inequalities for $\mathcal{H}$. They state that for $f \in L_\infty(\Omega)$

$$\|f\|_{L_\infty(\Omega)} \leq C \Big( \inf_{\phi \in \mathcal{H}_L} \|f - \phi\|_{L_\infty(\Omega)} + \|f\|_{\ell_\infty(X_N)} \Big)$$

holds if the truncation index $L$ is properly related to $X_N$. In order to also allow for unsymmetric reconstruction methods, we present Bernstein estimates for trial spaces of the form (1.1). These estimates are based on a careful truncation result as well, and can be seen as generalization of classical sampling inequalities (cf. [1, 2, 3, 11, 19, 29]). When applied to a residual function, such estimates open the way for a deterministic error analysis of a large class of reconstruction problems (compare also [19]). Jackson estimates then allow to derive error rates in terms of $L$ which in turn depends on $X_N$.

The remainder of this paper is organized as follows: In Section 2 we give basic definitions and derive a condition on the coupling of the discrete set of points $X_N$ and the truncation index $L$. In Section 3, we discuss interpolation with the truncated kernel $K^L$ and derive corresponding error estimates. Then, we show derivative-free sampling inequalities for general approximations in Section 4. Section 5 addresses Bernstein estimates for trial space of the form (1.1). We give some concluding remarks in Section 6.

**2. Notation and auxiliary results.** Throughout this paper, we denote by $C$ generic constants that may change from expression to expression. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ we set $|\boldsymbol{x} - \boldsymbol{y}| := \|\boldsymbol{x} - \boldsymbol{y}\|_{\ell_2(\mathbb{R}^d)}$. We denote by $(\mathcal{H}, (\cdot,\cdot)_{\mathcal{H}})$ a real separable Hilbert space of functions $f : \Omega \to \mathbb{R}$ where $\Omega \subset \mathbb{R}^d$ is an open, bounded domain.

**2.1. Hilbert spaces and frames.** Denote by $\mathcal{I} \subset \mathbb{N} \times \mathbb{N}$ an index set with multilevel structure, i.e.

$$\mathcal{I} = \bigotimes_{\ell=1}^{\infty} I_\ell := \left\{ (\ell, j) : \ell \in \mathbb{N},\ j \in I_\ell \right\}, \quad \text{with} \quad m_\ell := \# I_\ell \in \mathbb{N}_0 \cup \{\infty\}.$$

For $(\ell, j) \in \mathcal{I}$, let the $\lambda_\ell$ be positive weights, and let $\phi_{\ell,j} \in \mathcal{H}$ be such that the collection of functions $\left\{ \lambda_\ell^{1/2} \phi_{\ell,j} \right\}_{(\ell,j) \in \mathcal{I}}$ is a tight[3] frame for $\mathcal{H}$, i.e.

$$\|f\|^2_{\mathcal{H}} = \sum_{(\ell,j) \in \mathcal{I}} \lambda_\ell \left| (f, \phi_{\ell,j})_{\mathcal{H}} \right|^2 \quad \text{for all } f \in \mathcal{H}.$$

[2]Note that, for analytically given kernels, often quite complicated function like e.g., Bessel functions for Sobolev splines must be evaluated. This, however, involves a truncated series expansion within the respective numerical library which introduces an error. This way, non–constant costs are implicitly introduced as well.

[3]The assumption on the frame $\{\lambda_\ell^{1/2} \phi_{\ell,j}\}$ to be tight is made here only for simplicity of notation. If $\{\lambda_\ell^{1/2} \phi_{\ell,j}\}$ is not tight but satisfies $A\|f\|^2_{\mathcal{H}} \leq \sum_{(\ell,j) \in \mathcal{I}} \lambda_\ell \left| (f, \phi_{\ell,j})_{\mathcal{H}} \right|^2 \leq B\|f\|^2_{\mathcal{H}}$ for all $f \in \mathcal{H}$ with positive frame bounds $A$ and $B$, then there is an equivalent norm $\|f\|_{\tilde{\mathcal{H}}} := \sum_{(\ell,j) \in \mathcal{I}} \lambda_\ell \left| (f, \phi_{\ell,j})_{\mathcal{H}} \right|^2$ on $\mathcal{H}$ such that $\{\lambda_\ell^{1/2} \phi_{\ell,j}\}$ is tight with respect to it.

Note that the norm in $\mathcal{H}$ can be written as (see [14, Lemma 2.4 and Corollary 2.5])

$$\|f\|_{\mathcal{H}}^2 := \min\left\{ \sum_{(\ell,j)\in\mathcal{I}} |a_{\ell,j}|^2 : \{a_{\ell,j}\} \in \ell_2,\ f = \sum_{(\ell,j)\in\mathcal{I}} a_{\ell,j}\lambda_\ell^{1/2}\phi_{\ell,j} \right\} = \sum_{(\ell,j)\in\mathcal{I}} \left| \left(f, \mathcal{S}^{-1}\lambda_\ell^{1/2}\phi_{\ell,j}\right)_{\mathcal{H}} \right|^2$$

where the so-called frame operator

$$\mathcal{S} : \mathcal{H} \to \mathcal{H},\ \mathcal{S}(f) = \sum_{(\ell,j)\in\mathcal{I}} \lambda_\ell \left(f, \phi_{\ell,j}\right)_{\mathcal{H}} \phi_{\ell,j}$$

is invertible due to the tightness assumption.
We further assume that the weights $\lambda_\ell$ are positive and summable, i.e.

$$\sum_{\ell=0}^{\infty} \lambda_\ell < \infty, \tag{2.1}$$

and that the functions $\phi_{\ell,j}$ are level-wise uniformly square-summable, that is

$$\sup_{\boldsymbol{x}\in\Omega} \sup_{\ell\in\mathbb{N}} \sum_{j=1}^{m_\ell} |\phi_{\ell,j}(\boldsymbol{x})|^2 \le A. \tag{2.2}$$

Then in particular for some $\tilde{A} > 0$,

$$\sum_{(\ell,j)\in\mathcal{I}} \lambda_\ell |\phi_{\ell,j}(\boldsymbol{x})|^2 \le \tilde{A} \quad \text{for all } \boldsymbol{x} \in \Omega. \tag{2.3}$$

We point out that the latter condition (2.3) plays a crucial role in our analysis. In particular, an important parameter will be the approximation error made when approximating the series by partial sums of the form $\sup_{\boldsymbol{x}\in\Omega} \sum_{\ell=1}^{L} \lambda_\ell \sum_{j\in I_\ell} |\phi_{\ell,j}(x)|^2$. On the other hand, (2.1) and (2.2) provide only a normalization of $\lambda_\ell$ and $\phi_{\ell,j}$, which we choose in view of wavelet-type decompositions, see [14, 15]. Precisely, given a frame system $\{\mu^{1/2}\psi_{\ell,j}\}_{(\ell,j)\in\mathcal{I}}$ with positive weights $\mu_\ell$ such that the summability condition (2.3) holds, i.e., $\sup_{x\in\Omega} \sum_{(\ell,j)\in\mathcal{I}} \mu_\ell |\psi_{\ell,j}(\boldsymbol{x})|^2 \le \tilde{A}$, we may introduce the rescaled quantities

$$\phi_{\ell,j} := \begin{cases} \psi_{\ell,j}\left(\sup_{\boldsymbol{x}\in\Omega} \sum_{j=1}^{m_\ell} |\psi_{\ell,j}(\boldsymbol{x})|^2\right)^{-1/2}, & \text{if } \psi_{\ell,j} \not\equiv 0, \\ 0, & \text{if } \psi_{\ell,j} \equiv 0, \end{cases} \qquad \text{and} \qquad \lambda_\ell := \mu_\ell \sup_{\boldsymbol{x}\in\Omega} \sum_{j=1}^{m_\ell} |\psi_{\ell,j}(\boldsymbol{x})|^2.$$

Then $\lambda_\ell^{1/2}\phi_{\ell,j} = \mu^{1/2}\psi_{\ell,j}$, that is, they yield the same frame system. Thus, by construction, $\{\lambda_\ell\}$ and $\{\phi_{\ell,j}\}$ satsify all required assumptions in particular (2.1) and (2.2), since

$$\sum_{\ell=0}^{\infty} \lambda_\ell = \sup_{\boldsymbol{x}\in\Omega} \sum_{(\ell,j)\in\mathcal{I}} \mu_\ell |\phi_{\ell,j}(\boldsymbol{x})|^2 \le \tilde{A},$$

and

$$\sup_{\boldsymbol{x}\in\Omega} \sup_{\ell\in\mathbb{N}} \sum_{j=1}^{m_\ell} |\phi_{\ell,j}(\boldsymbol{x})|^2 = \sup_{\boldsymbol{x}\in\Omega} \sum_{j=1}^{m_\ell} \frac{|\psi_{\ell,j}(\boldsymbol{x})|^2}{\sup_{\boldsymbol{x}\in\Omega} \sum_{j=1}^{m_\ell} |\psi_{\ell,j}(\boldsymbol{x})|^2} = 1.$$

Next, we define

$$\mathcal{H}_L := \Big\{ f_L = \sum_{\ell=1}^{L} \sum_{j \in I_\ell} a_{\ell,j} \lambda_\ell^{1/2} \phi_{\ell,j} : \sum_{(\ell,j)} |a_{\ell,j}|^2 < \infty \Big\}. \tag{2.4}$$

Furthermore, we assume that for all $L \in \mathbb{N}$ the set of functions $\Big\{ \lambda_\ell^{\frac{1}{2}} \phi_{\ell,j} : 1 \leq \ell \leq L,\ j \in I_\ell \Big\}$ is a tight frame for $(\mathcal{H}_L, (\cdot,\cdot)_{\mathcal{H}})$, i.e., we have for $f_L \in \mathcal{H}_L \subset \mathcal{H}$

$$\|f_L\|_{\mathcal{H}}^2 = \min \Big\{ \sum_{\ell=1}^{L} \sum_{j \in I_\ell} |a_{\ell,j}|^2 : f_L = \sum_{\ell=1}^{L} \sum_{j \in I_\ell} a_{\ell,j} \lambda^{1/2} \phi_{\ell,j} \Big\}. \tag{2.5}$$

Prominent examples of decompositions of the form (1.2) arise from Mercer's theorem, where the weights $\lambda_\ell$ and the functions $\{\phi_{\ell,j}\}$ denote the eigenvalues and eigenfunctions of the associated compact integral operator, respectively. Also for frame systems arising from stable splittings, the assumption (2.5) is satisfied (see [16, 17]). For orthonormal systems $\{\phi_{\ell,j}\}$, the equation (2.5) directly holds.

**2.2. Kernel-based trial spaces.** Denote by $C(\Omega)$ the space of continuous real functions on $\Omega$. If there is a continuous embedding $\mathcal{H} \subset C(\Omega)$, then $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS), i.e. there is a reproducing kernel $K : \Omega \times \Omega \to \mathbb{R}$ such that
(i) $K(\cdot, \boldsymbol{x}) \in \mathcal{H}$ for all $\boldsymbol{x} \in \Omega$, and
(ii) $f(\boldsymbol{x}) = (f, K(\cdot,\boldsymbol{x}))_{\mathcal{H}}$ for all $f \in \mathcal{H}$ and all $\boldsymbol{x} \in \Omega$.
By (2.3), the kernel function

$$K : \Omega \times \Omega \to \mathbb{R}, \quad K(\boldsymbol{x}, \boldsymbol{y}) = \sum_{(\ell,j) \in \mathcal{I}} \lambda_\ell \, \phi_{\ell,j}(\boldsymbol{x}) \, \phi_{\ell,j}(\boldsymbol{y}) \tag{2.6}$$

is well defined. It is indeed the reproducing kernel in $\mathcal{H}$ (see [14, Lemma 2.6]). Since $K$ is a reproducing kernel, it is positive semi-definite (cf. [28, Theorem 10.4]), but we will restrict ourselves to positive definite kernels $K$ for reasons of simplicity. If the set $\{\phi_{\ell,j}\}$ is point-separating, then $K$ is positive definite, see [23]. Now, let a set $X_N := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$ of points $\boldsymbol{x}_n$ to be given. Associated with it are two natural density measures, namely the fill distance $h_{X_N,\Omega}$ and the separation distance $q_{X_N}$, given by

$$h_{X_N,\Omega} := \sup_{\boldsymbol{x} \in \Omega} \min_{\boldsymbol{x}_n \in X_N} \|\boldsymbol{x} - \boldsymbol{x}_n\|_2 \quad \text{and} \quad q_{X_N} := \frac{1}{2} \min_{\substack{\boldsymbol{x}_n \neq \boldsymbol{x}_m, \\ \boldsymbol{x}_n, \boldsymbol{x}_m \in X_N}} \|\boldsymbol{x}_n - \boldsymbol{x}_m\|_2 . \tag{2.7}$$

Clearly, $h_{X_N,\Omega} \leq q_{X_N}$ for all point sets $X_N \subset \Omega$, and $h_{X_N,\Omega} \sim q_{X_N} \sim N^{-1/d}$ for quasi-uniformly distributed sets. Furthermore, for $X_N \subset \Omega$, consider the associated kernel-based trial space $\mathcal{L}_{X_N}$ of (1.1). Elements of $\mathcal{L}_{X_N}$ are functions of the form $\sum_{n=1}^{N} \alpha_n K(\cdot, \boldsymbol{x}_n)$ with coefficient vectors $\boldsymbol{\alpha} \in \mathbb{R}^N$. We use the notation

$$s_{\boldsymbol{\alpha},X_N}(\cdot) := \sum_{n=1}^{N} \alpha_n K(\cdot, \boldsymbol{x}_n) = \sum_{(\ell,j) \in \mathcal{I}} \left( \sum_{n=1}^{N} \alpha_n \lambda_\ell \phi_{\ell,j}(\boldsymbol{x}_n) \right) \phi_{\ell,j}(\cdot) .$$

There is a strong connection between the reproduction property of the kernel and the underlying tight frame. Of particular interest to us is the fact that, for functions from $\mathcal{L}_{X_N}$, the frame coefficients can be easily computed,[4] i.e., the frame coefficients of $s_{\boldsymbol{\alpha},X_N}$ are (see [15, Theorem 3.2])

$$\left(s_{\boldsymbol{\alpha},X_N}, \mathcal{S}^{-1}\lambda_\ell^{1/2}\phi_{\ell,j}\right)_{\mathcal{H}} = \sum_{n=1}^{N} \alpha_n \lambda_\ell^{1/2} \phi_{\ell,j}\left(\boldsymbol{x}_n\right) \quad \text{for all } (\ell,j) \in \mathcal{I}. \tag{2.8}$$

Next, for a coefficient vector $\boldsymbol{\alpha} \in \mathbb{R}^N$ and a truncation index $L \in \mathbb{N}$ we use the notation

$$s^L_{\boldsymbol{\alpha},X_N}(\cdot) := \sum_{n=1}^{N} \alpha_n K^L(\cdot, \boldsymbol{x}_n) := \sum_{n=1}^{N} \alpha_n \sum_{\ell=1}^{L} \lambda_\ell \sum_{j \in I_\ell} \phi_{\ell,j}\left(\boldsymbol{x}_n\right) \phi_{\ell,j}\left(\cdot\right). \tag{2.9}$$

Finally, denote the discrete analysis operator by

$$S_{X_N} : \mathcal{H} \to \ell_2\left(X_N\right) \quad , \quad f \mapsto S_{X_N} f := \left(\left(f, K\left(\cdot, \boldsymbol{x}_1\right)\right)_{\mathcal{H}}, \ldots, \left(f, K\left(\cdot, \boldsymbol{x}_N\right)\right)_{\mathcal{H}}\right)^T \tag{2.10}$$

the discrete synthesis operator by

$$S^\star_{X_N} : \ell_2\left(X_N\right) \to \mathcal{H}, \quad \boldsymbol{\alpha} = \left(\alpha_1, \ldots, \alpha_N\right)^T \mapsto S^\star_{X_N}\boldsymbol{\alpha} := \sum_{n=1}^{N} \alpha_j K(\cdot, \boldsymbol{x}_n), \tag{2.11}$$

and the discretized version of the frame operator by

$$S_{X_N} S^\star_{X_N} : \ell_2\left(X_N\right) \to \ell_2\left(X_N\right), \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} \mapsto \begin{pmatrix} K(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & K(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ K(\boldsymbol{x}_N, \boldsymbol{x}_1) & \cdots & K(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix}. \tag{2.12}$$

For further details, see [22]. We will refer to this Gramian matrix as $K_{X_N,X_N} := S_{X_N} S^\star_{X_N} \in \mathbb{R}^{N \times N}$.

**2.3. Auxiliary truncation estimate.** For a given point set $X_N \subset \Omega$, we will now derive a sufficient condition on the number of levels of the kernel $K^L$ such that for all functions $s_{\boldsymbol{\alpha},X_N} \in \operatorname{span}\{K(\boldsymbol{x}_j, \cdot) : \boldsymbol{x}_j \in X_N\}$ the Hilbert space norm of $s_{\boldsymbol{\alpha},X_N}$ is uniformly bounded from above by the Hilbert space norm of $s^L_{\boldsymbol{\alpha},X_N}$. Roughly speaking, the more dense $X_N$ becomes in $\Omega$, the more levels $L$ are needed. The essential parameter entering the coupling condition between $X_N$ and $L$ turns out to be the smallest eigenvalue of the Gramian matrix $K_{X_N,X_N}$ in (2.12). Since $K$ is assumed to be positive definite, this eigenvalue is positive. We recall that for many popular kernels, the latter can be estimated in terms of the separation distance of $X_N$ in $\Omega$, see [28], the references therein, and also definition (2.7) in Section 5. We have the following lemma.

LEMMA 2.1. *Let $\gamma > 1$, and let $\{\lambda_\ell^{1/2}\phi_{\ell,j}, \ell \in \mathbb{N}$ and $j \in I_\ell\}$ be a tight frame in $\mathcal{H}$ such that* (2.1)*,* (2.2)*, and* (2.5) *are satisfied. For $X_N \subset \Omega$ denote by $\sigma_{\min}\left(S_{X_N} S^\star_{X_N}\right)$ the smallest eigenvalue of the Gram matrix $K_{X_N,X_N} = S_{X_N} S^\star_{X_N}$ in* (2.12)*. If L is large enough such that with A from* (2.2) *it holds*

$$\Lambda_L := \sum_{\ell=L+1}^{\infty} \lambda_\ell \leq \frac{1}{\gamma^2} \frac{\sigma_{\min}\left(S_{X_N} S^\star_{X_N}\right)}{N\ A}, \tag{2.13}$$

*then for all $s_{\boldsymbol{\alpha},X_N} \in \mathcal{L}_{X_N}$*

$$\|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{H}} \leq \frac{\gamma}{\gamma - 1} \left\|s^L_{\boldsymbol{\alpha},X_N}\right\|_{\mathcal{H}}. \tag{2.14}$$

[4] Recall that, for general functions from a Hilbert space $\mathcal{H}$, the computation of the frame coefficients is usually a difficult task.

*Proof.* We follow the lines of [13]. First let us show

$$\left\| s_{\boldsymbol{\alpha},X_N} - s^L_{\boldsymbol{\alpha},X_N} \right\|_{\mathcal{H}} \leq \frac{1}{\gamma} \left\| s_{\boldsymbol{\alpha},X_N} \right\|_{\mathcal{H}} .$$

A lower bound for $\|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{H}}$ is universal in all reproducing kernel Hilbert spaces, i.e., we have

$$\begin{aligned} \|s_{\boldsymbol{\alpha},X_N}\|^2_{\mathcal{H}} &= \left( \sum_{n=1}^N \alpha_n K(\cdot, \boldsymbol{x}_n), \sum_{m=1}^N \alpha_m K(\cdot, \boldsymbol{x}_m) \right)_{\mathcal{H}} \\ &= \left( S^\star_{X_N}(\boldsymbol{\alpha}), S^\star_{X_N}(\boldsymbol{\alpha}) \right)_{\mathcal{H}} \geq \sigma_{\min}\left( S_{X_N} S^\star_{X_N} \right) \|\boldsymbol{\alpha}\|^2_{\ell_2} . \end{aligned} \tag{2.15}$$

Using the level-wise summability assumption (2.2), we further estimate

$$\begin{aligned} &\left\| s_{\boldsymbol{\alpha},X_N} - s^L_{\boldsymbol{\alpha},X_N} \right\|^2_{\mathcal{H}} = \min\left\{ \|\boldsymbol{g}\|^2_{\ell_2} : s_{\boldsymbol{\alpha},X_N} - s^L_{\boldsymbol{\alpha},X_N} = \sum g_{\ell,j} \lambda_\ell^{1/2} \phi_{\ell,j} \right\} \\ &\leq \sum_{\ell=L+1}^{\infty} \sum_{j \in I_\ell} \left( \sum_{n=1}^N \alpha_n \lambda_\ell^{\frac{1}{2}} \phi_{\ell,j}(\boldsymbol{x}_n) \right)^2 \leq \sum_{\ell=L+1}^{\infty} \lambda_\ell \|\boldsymbol{\alpha}\|^2_{\ell_2} \sum_{j \in I_\ell} \sum_{n=1}^N |\phi_{\ell,j}(\boldsymbol{x}_n)|^2 \\ &\leq NA \|\boldsymbol{\alpha}\|^2_{\ell_2} \sum_{\ell=L+1}^{\infty} \lambda_\ell = NA \|\boldsymbol{\alpha}\|^2_{\ell_2} \Lambda_L, \end{aligned} \tag{2.16}$$

where $\Lambda_L$ is given in (2.13). Thus by (2.15) and (2.16) we obtain

$$\left\| s_{\boldsymbol{\alpha},X_N} - s^L_{\boldsymbol{\alpha},X_N} \right\|_{\mathcal{H}} \leq \frac{1}{\gamma} \|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{H}}.$$

With the identity $s_{\boldsymbol{\alpha},X_N} = s_{\boldsymbol{\alpha},X_N} - s^L_{\boldsymbol{\alpha},X_N} + s^L_{\boldsymbol{\alpha},X_N}$ and the triangle inequality

$$\|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{H}} \leq \left\| s_{\boldsymbol{\alpha},X_N} - s^L_{\boldsymbol{\alpha},X_N} \right\|_{\mathcal{H}} + \left\| s^L_{\boldsymbol{\alpha},X_N} \right\|_{\mathcal{H}}$$

we have $\|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{H}} \leq \frac{1}{\gamma} \|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{H}} + \left\| s^L_{\boldsymbol{\alpha},X_N} \right\|_{\mathcal{H}}$. Since $\gamma > 1$, this is equivalent to

$$\frac{\gamma - 1}{\gamma} \|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{H}} \leq \left\| s^L_{\boldsymbol{\alpha},X_N} \right\|_{\mathcal{H}}$$

which concludes the proof. ☐

Lemma 2.1 requires only weak assumptions on the frame. Moreover, the condition (2.13) on the truncation index $L$ is quite general. It can be further relaxed if the frame system has additional decay properties, such as an off-diagonal decay.

**3. Interpolation with truncated kernels.** We now consider the reconstruction problem to find an approximation to an unknown function $f \in \mathcal{H}$ that interpolates the values of $f$ at some scattered locations $X_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$. Since the kernel $K$ is positive definite, there exist unique coefficients $(\alpha_1(f), \ldots, \alpha_N(f))^T$ such that $s_{f,X_N}|_{X_N} = f|_{X_N}$, where

$$s_{f,X_N}(\cdot) = \sum_{n=1}^N \alpha_n(f) K(\cdot, \boldsymbol{x}_n) = \sum_{\ell \in \mathbb{N}} \lambda_\ell \sum_{j \in I_\ell} \left( \sum_{n=1}^N \alpha_n(f) \phi_{\ell,j}(\boldsymbol{x}_n) \right) \phi_{\ell,j}(\cdot).$$

The coefficients $(\alpha_1, \ldots, \alpha_N) = (\alpha_1(f), \ldots, \alpha_N(f))$ are determined by the linear system

$$\begin{pmatrix} \sum\limits_{\ell\in\mathbb{N}} \lambda_\ell \sum\limits_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_1)\phi_{\ell,j}(\boldsymbol{x}_1) & \ldots & \sum\limits_{\ell\in\mathbb{N}} \lambda_\ell \sum\limits_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_1)\phi_{\ell,j}(\boldsymbol{x}_N) \\ \vdots & \vdots & \vdots \\ \sum\limits_{\ell\in\mathbb{N}} \lambda_\ell \sum\limits_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_N)\phi_{\ell,j}(\boldsymbol{x}_1) & \ldots & \sum\limits_{\ell\in\mathbb{N}} \lambda_\ell \sum\limits_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_N)\phi_{\ell,j}(\boldsymbol{x}_N) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix} \tag{3.1}$$

with right-hand side $(f_1, \ldots, f_N)^T = (f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N))^T$ . This yields the interpolant with minimal Hilbert space norm (see [28]). The linear system to determine the coefficients can however be ill-conditioned. Furthermore, if there is no closed form expression available for the kernel then it is in general impossible to evaluate the infinite sums and to build the matrix in (3.1) in the first place. Thus, instead of $K$ one is led to consider the truncated kernel $K^L$ introduced in (2.9), i.e.,

$$K^L : \Omega \times \Omega \to \mathbb{R}, \quad K^L(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\ell=1}^{L} \lambda_\ell \sum_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x})\phi_{\ell,j}(\boldsymbol{y}). \tag{3.2}$$

The kernel $K^L$ is positive semi-definite for every $L \in \mathbb{N}$, see [14, 18]. Indeed, if $N \in \mathbb{N}$, $X_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$, and $\boldsymbol{\alpha} \in \mathbb{R}^N$, then

$$\sum_{n,m=1}^{N} \alpha_n \alpha_m K^L(\boldsymbol{x}_n, \boldsymbol{x}_m) = \sum_{\ell=1}^{L} \lambda_\ell \sum_{j\in I_\ell} \left( \sum_{n=1}^{N} \alpha_n \phi_{\ell,j}(\boldsymbol{x}_n) \right)^2 \geq 0.$$

Fortunately, the Hilbert space theory is well-established also for positive semi-definite kernels, see [28]. The associated reproducing kernel Hilbert space of $K^L$ is

$$\mathcal{H}_L := \left\{ \sum_{\ell=1}^{L} \sum_{j\in I_\ell} a_{\ell,j} \lambda_\ell^{1/2} \phi_{\ell,j}(\cdot) : a_{\ell,j} \in \mathbb{R} \text{ and } \sum_{\ell=1}^{L} \sum_{j\in I_\ell} |a_{\ell,j}|^2 < \infty \right\} \tag{3.3}$$

with norm

$$\|f\|_{\mathcal{H}_L}^2 := \min \left\{ \sum_{\ell=1}^{L} \sum_{j\in I_\ell} |a_{\ell,j}|^2 : f = \sum_{\ell=1}^{L} \lambda_\ell \sum_{j\in I_\ell} a_{\ell,j} \phi_{\ell,j} \right\}.$$

Note that by assumption (2.5) we have $\|\cdot\|_{\mathcal{H}_L} = \|\cdot\|_{\mathcal{H}}$ on $\mathcal{H}_L$ .

Now we consider the problem to interpolate a function $f \in \mathcal{H}$ by translates of the truncated kernel $K^L$. Assume for the moment that such an interpolant exists. We denote it by

$$s_{f,X_N}^L = \sum_{n=1}^{N} \alpha_n^{(L)}(f) K^L(\cdot, \boldsymbol{x}_n), \tag{3.4}$$

where the coefficients $\left(\alpha_1^{(L)}, \ldots, \alpha_N^{(L)}\right) = \left(\alpha_1^{(L)}(f), \ldots, \alpha_N^{(L)}(f)\right)$ are determined by the linear system

$$\begin{pmatrix} \sum\limits_{\ell=1}^{L} \lambda_\ell \sum_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_1)\phi_{\ell,j}(\boldsymbol{x}_1) & \ldots & \sum\limits_{\ell=1}^{L} \lambda_\ell \sum\limits_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_1)\phi_{\ell,j}(\boldsymbol{x}_N) \\ \vdots & \vdots & \vdots \\ \sum\limits_{\ell=1}^{L} \lambda_\ell \sum\limits_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_N)\phi_{\ell,j}(\boldsymbol{x}_1) & \ldots & \sum\limits_{\ell=1}^{L} \lambda_\ell \sum\limits_{j\in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_N)\phi_{\ell,j}(\boldsymbol{x}_N) \end{pmatrix} \begin{pmatrix} \alpha_1^{(L)} \\ \vdots \\ \alpha_N^{(L)} \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix}. \tag{3.5}$$

Note here that the cost for the evaluation of a properly truncated kernel which is needed to evaluate an entry of the truncated Gramian in (3.5) is no longer a constant independent of the degrees of freedom $N$. The cost of evaluating $K^L(\boldsymbol{x}, \boldsymbol{y})$ from (3.2) will be at least linear in $L$, hence, we have to determine the dependence of $L$ on $N$. To this end, we proceed in two steps: We first show that there exists a function from $\mathcal{H}_L$ that interpolates the given data, and we subsequently show that the system (3.5) possesses a solution with the associated interpolant being a quasi-optimal approximant. We use the following *general* result proven in [13, Proposition 3.1].

PROPOSITION 3.1. *Let $\mathcal{Y}$ be a Banach space, $\mathcal{V} \subset \mathcal{Y}$ a subspace, and $\mathcal{Z}'$ a finite-dimensional subspace of the dual space $\mathcal{Y}'$. If for every $z' \in \mathcal{Z}'$ and some $\hat{\gamma} > 1$ independent of $z'$*

$$\|z'\|_{\mathcal{Y}'} \leq \hat{\gamma} \, \|z'|_{\mathcal{V}}\|_{\mathcal{V}'}, \tag{3.6}$$

*then, for any $y \in \mathcal{Y}$ there exists $v = v(y) \in \mathcal{V}$ such that $v(y)$ interpolates $y$ on $\mathcal{Z}'$, that is, $z'(y) = z'(v(y))$ for all $z' \in \mathcal{Z}'$. In addition $v(y)$ approximates $y$ in the sense that*

$$\|v(y) - y\|_{\mathcal{Y}} \leq (1 + 2\hat{\gamma}) \operatorname{dist}_{\mathcal{Y}}(y, \mathcal{V}),$$

*where* $\operatorname{dist}_{\mathcal{Y}}(y, \mathcal{V}) := \inf_{v \in \mathcal{V}} \|y - v\|_{\mathcal{Y}}$.

Now, we apply Proposition 3.1 for a discrete set $X_N \subset \Omega$ with

$$\mathcal{Y} = \mathcal{H}, \quad \mathcal{V} = \mathcal{H}_L \quad \text{and} \quad \mathcal{Z}' := \left\{\delta_{\boldsymbol{x}_j} : \boldsymbol{x}_j \in X_N\right\} \tag{3.7}$$

and obtain the following result.

THEOREM 3.2. *Let $X_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$ be a discrete set of points. Suppose that $\{\lambda_\ell^{1/2}\phi_{\ell,j}\colon \ell \in \mathbb{N}$ and $j \in I_\ell\}$ satisfies* (2.1), (2.2) *and* (2.5) *and that the truncation index $L \in \mathbb{N}$ is large enough such that* (2.13) *holds. Then, for every $f \in \mathcal{H}$, there is an $f_L \in \mathcal{H}_L$ such that*

$$f|_{X_N} = f_L|_{X_N} \quad \textit{and} \quad \|f - f_L\|_{\mathcal{H}} \leq (1 + 2\hat{\gamma}) \operatorname{dist}_{\mathcal{H}}(f, \mathcal{H}_L),$$

*where $\hat{\gamma} := \gamma\,(\gamma - 1)^{-1}$ with $\gamma$ as in Lemma 2.1.*

*Proof*. We follow [13, Lemma 3.3 and Theorem 3.4]. By Lemma 2.1 it suffices to show that (2.14) implies (3.6) with $\hat{\gamma} := \gamma\,(\gamma - 1)^{-1}$ for the spaces specified in (3.7). The claim then follows from Proposition 3.1. To this end, let $z' = \sum_{n=1}^N \alpha_n \delta_{\boldsymbol{x}_n}$ be an arbitrary element of $\mathcal{Z}' = \operatorname{span}\{\delta_{\boldsymbol{x}_n} : \boldsymbol{x}_n \in X_N\}$. Then, by Riesz' representation and the reproducing property, we have

$$\|z'\|_{\mathcal{H}'} = \Big\| \sum_{n=1}^{N} \alpha_n K(\cdot, \boldsymbol{x}_n) \Big\|_{\mathcal{H}} = \|s_{\boldsymbol{\alpha}, X_N}\|_{\mathcal{H}},$$

where $\|\cdot\|_{\mathcal{H}'}$ denotes the norm dual to $\|\cdot\|_{\mathcal{H}}$. It remains to show that

$$\|z'|_{\mathcal{H}_L}\|_{\mathcal{H}'} = \left\|s^L_{\boldsymbol{\alpha}, X_N}\right\|_{\mathcal{H}},$$

since (3.6) then follows from Lemma 2.1 via

$$\|z'\|_{\mathcal{H}'} = \|s_{\boldsymbol{\alpha}, X_N}\|_{\mathcal{H}} \leq \frac{\gamma}{\gamma - 1} \|s^L_{\boldsymbol{\alpha}, X_N}\|_{\mathcal{H}} = \frac{\gamma}{\gamma - 1} \|z'|_{\mathcal{H}_L}\|_{\mathcal{H}'_L}.$$

Consider an arbitrary element $g_L \in \mathcal{H}_L$, i.e., $g_L = \sum_{\ell=1}^L \lambda_\ell^{1/2} \sum_{j \in I_\ell} \beta_{\ell,j}\phi_{\ell,j}$, set $\mathcal{I}_L := \otimes_{\ell=1}^L I_\ell$ and

$$\boldsymbol{s}_L := \{s_{\ell,j}\}_{(\ell,j) \in \mathcal{I}_L} = \left\{ \sum_{n=1}^{N} \alpha_n \lambda_\ell^{1/2} \phi_{\ell,j}(\boldsymbol{x}_n) \right\}_{(\ell,j) \in \mathcal{I}_L} \in \ell_2(\mathcal{I}_L). \tag{3.8}$$

Then

$$|z'(g_L)| = \left|\sum_{\ell=1}^{L}\sum_{j\in I_\ell}\beta_{\ell,j}\sum_{n=1}^{N}\alpha_n\lambda_\ell^{1/2}\phi_{\ell,j}(\boldsymbol{x}_n)\right| = \left|\sum_{\ell=1}^{L}\sum_{j\in I_\ell}\beta_{\ell,j}s_{\ell,j}\right| = \left|(\boldsymbol{s}_L,\boldsymbol{\beta})_{\ell_2}\right|.$$

Recall that the $\{s_{\ell,j}\}$ of (3.8) are the frame coefficients of $s^L_{\boldsymbol{\alpha},X_N} \in \mathcal{H}_L$ with respect to the truncated frame. Now, consider the operator $T : \ell_2(\mathcal{I}_L) \to \mathcal{H}_L$, $\eta \mapsto \sum_{\ell=1}^{L}\sum_{j\in I_\ell}\eta_{\ell,j}\lambda_\ell^{1/2}\phi_{\ell,j}$ and denote its null space by $\mathcal{N}$. Then $\boldsymbol{s}_L \in \mathcal{N}^\perp$ since for every $\boldsymbol{\eta} \in \mathcal{N}$

$$(\boldsymbol{s}_L,\boldsymbol{\eta})_{\ell_2} = \sum_{\ell=1}^{L}\sum_{j\in I_\ell}\Big(\sum_{n=1}^{N}\alpha_n\lambda_\ell^{1/2}\phi_{\ell,j}(\boldsymbol{x}_n)\Big)\eta_{\ell,j} = \sum_{n=1}^{N}\alpha_n\sum_{\ell=1}^{L}\sum_{j\in I_\ell}\eta_{\ell,j}\lambda_\ell^{1/2}\phi_{\ell,j}(\boldsymbol{x}_n) = 0.$$

In particular, $\boldsymbol{s}_L = P_{\mathcal{N}^\perp}(\boldsymbol{s}_L)$, where $P_{\mathcal{N}^\perp} : \ell_2(\mathcal{I}_L) \to \ell_2(\mathcal{I}_L)$ denotes the orthogonal projection onto $\mathcal{N}^\perp$. Since projections are self-adjoint, we have

$$\begin{aligned}|z'(g_L)| &= \left|(\boldsymbol{s}_L,\boldsymbol{\beta})_{\ell_2(\mathcal{I}_L)}\right| = \left|(P_{\mathcal{N}^\perp}(\boldsymbol{s}_L),\boldsymbol{\beta})_{\ell_2(\mathcal{I}_L)}\right| = \left|(P_{\mathcal{N}^\perp}(\boldsymbol{\beta}),\boldsymbol{s}_L)_{\ell_2(\mathcal{I}_L)}\right| \\ &\le \|P_{\mathcal{N}^\perp}(\boldsymbol{\beta})\|_{\ell_2(\mathcal{I}_L)}\|\boldsymbol{s}_L\|_{\ell_2(\mathcal{I}_L)}.\end{aligned}$$

Furthermore, by assumption (2.5), it holds $\|g_L\|_{\mathcal{H}} = \|P_{\mathcal{N}^\perp}(\boldsymbol{\beta})\|_{\ell_2}$ which implies

$$\|z'|_{\mathcal{H}_L}\|_{\mathcal{H}'} \le \|\boldsymbol{s}_L\|_{\ell_2(\mathcal{I}_L)} = \left\|s^L_{\boldsymbol{\alpha},X_N}(\cdot)\right\|_{\mathcal{H}}.$$

Finally, the equality $\|z'|_{\mathcal{H}_L}\|_{\mathcal{H}'} = \left\|s^L_{\boldsymbol{\alpha},X_N}\right\|_{\mathcal{H}}$ follows by choosing $g_L = s^L_{\boldsymbol{\alpha},X_N}$. ∎

Theorem 3.2 shows the existence of an interpolant from $\mathcal{H}_L$ which is even quasi-optimal. This is a strong statement: Typically in approximation theory, the interpolation error is worse than the best approximation error by a factor referred to as the Lebesgue constant. In Theorem 3.2 the Lebesgue constant is bounded by the constant $(1+2\hat{\gamma})$. Recall that the parameter $\hat{\gamma} = \frac{\gamma}{\gamma-1}$ is linked to the truncation parameter $L$ via the parameter $\gamma$ from Lemma 2.1. Roughly speaking, if $\gamma$ is chosen close to 1, then the truncation index may be chosen rather small and, consequently, the Lebesgue constant is large, while if $\gamma$ is chosen large, then also the truncation index $L$ is large and the Lebesgue constant is close to $1 + 2\cdot 1 = 3$. The proof of Theorem 3.2 is not constructive but it lays the foundation to show the existence of a quasi-optimal interpolant based on the truncated kernel which can be computed by (3.5).

**4. Error estimates for approximation with truncated kernels.** In this section we want to give an a priori error analysis for stable reconstruction processes using the truncated kernel. Precisely, for a discrete set $Y_M := \{y_1,\dots,y_M\} \subset \Omega$, we denote by $\mathcal{R}_{Y_M} : S_{Y_M}(\mathcal{H}) \to \mathcal{H}$ a reconstruction operator which uses standard information obtained by the analysis operator $S_{Y_M}$ to produce an approximation to the function $f \in \mathcal{H}$. Naturally such a reconstruction operator should satisfy a consistency and a stability estimate in order to provide a convergent reconstruction scheme. The consistency property states that, for a second set $X_N$, there is a number $\varepsilon = \varepsilon(M,N) > 0$ with $\lim_{N,M\to\infty}\varepsilon(M,N) = 0$ and

$$\|S_{X_N}(\mathcal{R}_{Y_M}(S_{Y_M}(f))) - S_{X_N}(f)\|_{\ell_\infty(X_N)} \le \varepsilon(M,N). \tag{4.1}$$

For classical interpolation processes, we assume $X_N = Y_M$ and we have $\varepsilon(N,N) \equiv 0$ for all $N$. The stability property takes the form

$$\|\mathcal{R}_{Y_M}(S_{Y_M}(f))\|_{\mathcal{H}} \le C\,\|f\|_{\mathcal{H}}. \tag{4.2}$$

Both conditions are satisfied for symmetric kernel-based interpolation problems, i.e., for $X_N = Y_M$, but also for a larger variety of machine learning algorithms with, in general, $X_N \neq Y_M$, see [9, 10, 28, 29, 20]. For kernel-based interpolation, we can express the reconstruction operator in terms of the discrete analysis operator (2.10) and its adjoint (2.11). We have for $X_N = Y_M$ and $f \in \mathcal{H}$

$$\mathcal{R}_{X_N}^{\text{Interpolation}} S_{X_N} f(\boldsymbol{x}) = S_{X_N}^{\star} \left(S_{X_N} S_{X_N}^{\star}\right)^{-1} S_{X_N} f(\boldsymbol{x}).$$

From this equation we also see that the case $X_N \neq Y_M$ needs additional arguments since the associated Gramian is a rectangular matrix and hence not invertible.

The error analysis for the case of symmetric interpolation follows by standard Jackson-type estimates. Here, in *general*, one considers a normed linear space $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ and a nested sequence of subspaces $\emptyset \subset \mathcal{V}_1 \subset \ldots \mathcal{V}_j \subset \ldots$ of $\mathcal{Y}$ endowed with a weaker norm $\|\cdot\|_{\mathcal{V}}$. We say that the sequence of subspaces $(\mathcal{V}_j)_j$ satisfies a Jackson-type estimate if there is a strictly decreasing function $h : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{j\to\infty} h(j) = 0$ and

$$\inf_{v^{(j)} \in \mathcal{V}_j} \left\| v - v^{(j)} \right\|_{\mathcal{V}} \leq h(j) \|v\|_{\mathcal{Y}} \tag{4.3}$$

holds for all $v \in \mathcal{Y}$.

For the special case of kernel based interpolation, we employ a Jackson-type estimate which is related to the space $\mathcal{L}_{X_N}$ as defined in (1.1). We assume $\mathcal{V}_j = \mathcal{L}_{X_{N_j}} \subset \mathcal{Y} = \mathcal{H}$ with $X_{N_j}$ getting dense in $\Omega$, which necessarily implies $N_j \to \infty$ for $j \to \infty$ and $\mathcal{L}_{X_{N_j}} \to \mathcal{L}$. Moreover, we set $\|\cdot\|_{\mathcal{L}} = \|\cdot\|_{\mathcal{V}}$, i.e., the weak norm on $\mathcal{L}_{X_{N_j}}$ is denoted by $\|\cdot\|_{\mathcal{L}}$. We assume that for interpolation with the reproducing kernel $K$ there is a Jackson-type estimate of the form

$$\|f - s_{f,X_N}\|_{\mathcal{L}} \leq \rho(N)\|f\|_{\mathcal{H}} = \rho(N, \mathcal{L}, \mathcal{H})\|f\|_{\mathcal{H}} \tag{4.4}$$

for all $f \in \mathcal{H}$. We use the notation $\rho(N, \mathcal{L}, \mathcal{H})$ to stress the dependence on the weak norm. The following theorem shows that we obtain, up to a constant, the same rates $\rho(N, \mathcal{L}, \mathcal{H})$ if we employ a properly truncated kernel $K^L$ instead of $K$.

THEOREM 4.1. *Suppose that* $\{\lambda_\ell^{1/2}\phi_{\ell,j} \,:\, \ell \in \mathbb{N},\, j \in I_\ell\}$ *satisfies* (2.1)*,* (2.2) *and that assumption* (2.5) *holds. Then there is* $C > 0$ *such that for all* $X_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$*, for all* $L \in \mathbb{N}$ *large enough such that* (2.13) *holds, and for every* $f \in \mathcal{H}$*, there exists an interpolant* $s_{f,X_N}^L$ *as in* (3.4) *based on the truncated kernel* $K^L$ *such that* (4.4) *implies*

$$\|f - s_{f,X_N}^L\|_{\mathcal{L}} \leq C\rho(X_N, \mathcal{L}, \mathcal{H})\|f\|_{\mathcal{H}},$$

*where* $\rho(X_N, \mathcal{L}, \mathcal{H})$ *is the same as in* (4.4).

*Proof.* By Theorem 3.2 there is a $f^L \in \mathcal{H}_L$ such that $f|_{X_N} = f^L|_{X_N}$ for every $f \in \mathcal{H}$ which implies $s_{f,X_N}^L = s_{f^L,X_N}^L$. In particular, we can view that data generated by $f \in \mathcal{H}$ as data generated by $f^L \in \mathcal{H}_L$. Hence, since $f^L$ lies in the reproducing kernel Hilbert space of $K^L$ the system (3.5) is solvable, i.e. $s_{f,X_N}^L$ exists for all $f \in \mathcal{H}$, and $s_{f,X_N}$ is also an interpolant to $s_{f,X_N}^L$. Thus, we may view $s_{f,X_N}^L$ as a function in $\mathcal{H}$ generating the data. By (4.4) we obtain

$$\left\|f - s_{f,X_N}^L\right\|_{\mathcal{L}} \leq \|f - s_{f,X_N}\|_{\mathcal{L}} + \left\|s_{f,X_N} - s_{f,X_N}^L\right\|_{\mathcal{L}} \leq 2\rho(X_N, \mathcal{L}, \mathcal{H}) \left(\|f\|_{\mathcal{H}} + \left\|s_{f,X_N}^L\right\|_{\mathcal{H}}\right).$$

It remains to bound $\|s^L_{f,X_N}\|_{\mathcal{H}}$ from above. Since $s^L_{f,X_N}$ is the best approximation to $f^L \in \mathcal{H}_L$ from $\operatorname{span}\{K^L(\boldsymbol{x}_j,\cdot):\ \boldsymbol{x}_j \in X_N\}$ we obtain by Theorem 3.2 the estimate

$$\begin{aligned}
\left\|s^L_{f,X_N}\right\|_{\mathcal{H}} &\leq \left\|s^L_{f,X_N} - f^L\right\|_{\mathcal{H}} + \|f^L - f\|_{\mathcal{H}} + \|f\|_{\mathcal{H}} \\
&\leq \left\|s^L_{f,X_N} - f^L\right\|_{\mathcal{H}} + (2+2\hat{\gamma})\,\|f\|_{\mathcal{H}} \\
&\leq \left\|f^L\right\|_{\mathcal{H}} + (2+2\hat{\gamma})\,\|f\|_{\mathcal{H}} \\
&\leq \|f_L - f\|_{\mathcal{H}} + \|f\|_{\mathcal{H}} + (1+2\hat{\gamma})\,\|f\|_{\mathcal{H}} \\
&\leq 2(2+2\hat{\gamma})\,\|f\|_{\mathcal{H}}\,,
\end{aligned}$$

which concludes the proof with $C := 2(1+2(2+\hat{\gamma}))$. □

In order to analyze general reconstruction processes, we need more flexible error estimates. Such estimates will have the form of a classical sampling inequality [19]. To this end, we first outline how Jackson-type estimates for sections of the frame imply Jackson estimates for kernel-based trial spaces. These results again introduce a certain coupling of the truncation index $L$ and the discrete set of centers $X_N$. In order to derive the coupling of the truncation index $L$ and the discrete set of centers $X_N$ we proceed as follows. Slightly abusing the notation from (2.10), we consider discrete analysis operators for discrete sets $X_N = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\} \subset \Omega$ restricted to subspaces $\mathcal{H}_L \subset \mathcal{H}$, i.e.,

$$S^{\infty}_{X_N} : (\mathcal{H}_L, \|\cdot\|_{L_\infty}) \to \left(\mathbb{R}^N, \|\cdot\|_{\ell_\infty(X_N)}\right), \quad f \mapsto (f(\boldsymbol{x}_1),\ldots,f(\boldsymbol{x}_N))^T\,. \tag{4.5}$$

If $S_{X_N}$ is injective, we denote its left-inverse by $\left(S^{\infty}_{X_N}\right)^{-1} : S^{\infty}_{X_N}(\mathcal{H}_L) \to \mathcal{H}_L$. The associated operator norms are given respectively as

$$\|S^{\infty}_{X_N}\| := \sup_{\substack{f\in\mathcal{H}_L \\ \|f\|_{L_\infty}=1}} \left\|S^{\infty}_{X_N}(f)\right\|_{\ell_\infty(\mathbb{R}^N)}\,, \qquad \text{and} \qquad \left\|\left(S^{\infty}_{X_N}\right)^{-1}\right\| := \sup_{\substack{\boldsymbol{\beta}\in S^{\infty}_{X_N}(\mathcal{H}_L) \\ \|\boldsymbol{\beta}\|_{\ell_\infty(\mathbb{R}^N)}=1}} \left\|\left(S^{\infty}_{X_N}\right)^{-1}(\boldsymbol{\beta})\right\|_{L_\infty(\mathbb{R}^N)}\,.$$

Note here that we now employ the $\|\cdot\|_{L_\infty}$-norm instead of the $\|\cdot\|_{\mathcal{H}}$-norm as in (2.10). Next, we fix $0 < c < 1$, and let the truncation index $L$ and the set $X_N \subset \Omega$ be such that

$$\omega(h_{X_N,\Omega}, f^L)_{L_\infty(\Omega)} \leq (1-c)\left\|f^L\right\|_{L_\infty(\Omega)} \quad \text{for all } f^L \in \mathcal{H}_L \tag{4.6}$$

where the modulus of continuity is defined as

$$\omega(h_{X_N,\Omega}, f)_{L_\infty(\Omega)} := \sup_{\substack{|\boldsymbol{t}|\leq h_{X_N,\Omega} \\ \boldsymbol{x},\boldsymbol{x}+\boldsymbol{t}\in\Omega}} |f(\boldsymbol{x}) - f(\boldsymbol{x}+\boldsymbol{t})|\,. \tag{4.7}$$

The condition (4.6) couples the truncation index $L$ and the point set $X_N$ via the fill distance $h_{X_N,\Omega}$. Roughly speaking, the larger $L$, that is, the larger $\mathcal{H}_L$, the smaller the fill distance of $X_N$ has to be. Thus the coupling condition (4.6) can be interpreted as a Bernstein-type inequality.[5] Note that if $\Omega$ is compact, then, by Arzela-Ascoli's theorem, (4.6) implies that $\mathcal{H}_L$ is finite dimensional since its unit ball with respect to $\|\cdot\|_{L_\infty}$ is

[5] In the univariate setting one formally has

$$\omega(h_{X_N,\Omega}, f^L)_{L_\infty(\Omega)} \leq \sup_{|t|\leq h_{X_N,\Omega}} \left|\int_{x+t}^{x} \frac{d}{dt} f^L(t)dt\right| \leq h_{X_N,\Omega}\left\|\frac{d}{dt} f^L(t)\right\|_{L_\infty(\Omega)}.$$

compact. However, we will show an even stronger bound on the norm of the inverse of $S^{\infty}_{X_N}$ in the following Proposition 4.2.

PROPOSITION 4.2. *For all* $L \in \mathbb{N}$ *and all* $X_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$ *such that* (4.6) *holds, the associated discrete analysis operator* (4.5) *satisfies*

$$\left\| S^{\infty}_{X_N} \right\| \leq 1 \quad \textit{and} \quad \left\| \left( S^{\infty}_{X_N} \right)^{-1} \right\| \leq \frac{1}{c}. \tag{4.8}$$

*Proof.* The bound on $\left\| S^{\infty}_{X_N} \right\|$ follows directly from the definition of the operator norm since

$$\left\| S^{\infty}_{X_N} \right\| = \sup_{f^L \in \mathcal{H}_L \setminus \{0\}} \frac{\left\| S^{\infty}_{X_N} \left( f^L \right) \right\|_{\ell_\infty(X_N)}}{\|f^L\|_{L_\infty(\Omega)}} = \sup_{f^L \in \mathcal{H}_L \setminus \{0\}} \frac{\max_{\boldsymbol{x}_n \in X_N} \left| f^L \left( \boldsymbol{x}_n \right) \right|}{\|f^L\|_{L_\infty(\Omega)}} \leq 1.$$

To show the second assertion we follow [28, Theorem 3.8]. For $f^L \in \mathcal{H}_L$ there is a point $\overline{\boldsymbol{x}} \in \overline{\Omega}$ with $\left| f^L(\overline{\boldsymbol{x}}) \right| = \|f^L\|_{L_\infty(\overline{\Omega})}$ since $\overline{\Omega}$ is compact. By definition of the fill-distance there is $\boldsymbol{x}_n \in X_N$ such that $|\overline{\boldsymbol{x}} - \boldsymbol{x}_n| \leq h_{X_N,\Omega}$. Hence by (4.6) and (4.7)

$$\left| f^L(\overline{\boldsymbol{x}}) - f^L(\boldsymbol{x}_n) \right| \leq \sup_{|t| \leq h_{X_N,\Omega}} \left| f^L(\overline{\boldsymbol{x}}) - f^L(\overline{\boldsymbol{x}} + t) \right| = \omega(h_{X_N,\Omega}, f^L)_{L_\infty(\Omega)} \leq (1-c) \left\| f^L \right\|_{L_\infty(\Omega)},$$

and thus, using the triangle inequality, we have

$$\left\| f^L \right\|_{L_\infty(\Omega)} = \left| f^L(\overline{\boldsymbol{x}}) \right| \leq \left| f^L(\overline{\boldsymbol{x}}) - f^L(\boldsymbol{x}_n) \right| + \left| f^L(\boldsymbol{x}_n) \right| \leq (1-c) \left| f^L(\overline{\boldsymbol{x}}) \right| + \left| f^L(\boldsymbol{x}_n) \right|$$

which directly yields $\left| f^L \left( \boldsymbol{x}_n \right) \right| \geq c \left\| f^L \right\|_{L_\infty(\Omega)}$. Therefore

$$\left\| \left( S^{\infty}_{X_N} \right)^{-1} \right\| = \sup_{\boldsymbol{\alpha} \in S_{X_N}(\mathcal{H}_L) \setminus \{0\}} \frac{\left\| \left( S^{\infty}_{X_N} \right)^{-1} \boldsymbol{\alpha} \right\|_{L_\infty(\Omega)}}{\|\boldsymbol{\alpha}\|_{\ell_\infty(\mathbb{R}^N)}} = \sup_{f^L \in \mathcal{H}_L \setminus \{0\}} \frac{\|f^L\|_{L_\infty(\Omega)}}{\left\| S^{\infty}_{X_N} f^L \right\|_{\ell_\infty(\mathbb{R}^N)}} \leq \frac{1}{c}.$$

This concludes the proof. $\square$

If the inverses of the discrete analysis operators are bounded then a generalized reproduction formula follows. To this end, there is the following *general* result, see [28, Theorem 3.4].

PROPOSITION 4.3. *Suppose that* $\mathcal{V}$ *is a finite dimensional normed linear space and that* $\mathcal{Z}'$ *is a set of functionals, i.e.,* $\mathcal{Z}' = \{z_1, \ldots, z_N\} \subset \mathcal{V}'$ *is such that the discrete analysis operator* $S_{\mathcal{Z}'} : \mathcal{V} \to S_{\mathcal{Z}'}(\mathcal{V}) \subset \mathbb{R}^N$ *defined by* $S_{\mathcal{Z}'}(v) := (z_1(v), \ldots, z_N(v))^T$ *is injective. Then for every* $\psi \in \mathcal{V}'$*, there exists a vector* $\boldsymbol{u} \in \mathbb{R}^N$ *depending only on* $\psi$ *such that for every* $w \in \mathcal{V}$ *it holds*

$$\psi(f) = \sum_{n=1}^{N} u_j z_j(f) \quad \textit{and} \quad \|\boldsymbol{u}\|_{(\mathbb{R}^N)'} \leq \|\psi\|_{\mathcal{V}'} \|S^{-1}_{\mathcal{Z}'}\|.$$

Note here that Proposition 4.3 can be seen as dual to Proposition 3.1 in the following sense: The condition (3.6) in Proposition 3.1 can always be satisfied by increasing the dimension of $\mathcal{V}$ with respect to the set $\mathcal{Z}'$ while in Proposition 4.3 it is just the other way around, i.e., it is possible to make the operator $S_{\mathcal{Z}'}$ injective by oversampling which means to enlarge the set $\mathcal{Z}'$ properly.

In particular, we now set $\mathcal{V} = \mathcal{H}_L$, $\psi = \delta_{\boldsymbol{x}}$ and $\mathcal{Z}' = \{\delta_{\boldsymbol{x}_1}, \ldots, \delta_{\boldsymbol{x}_N}\}$ and $S_{\mathcal{Z}'} = S_{X_N}$. Recall here that $\mathcal{H}_L$ is embedded into $C(\overline{\Omega})$, and thus the point evaluation functional $\delta_{\boldsymbol{x}}$ for $\boldsymbol{x} \in \Omega$ is uniformly bounded by a constant $C > 0$ since for all $L \in \mathbb{N}$ and all $\boldsymbol{x} \in \overline{\Omega}$

$$\|\delta_{\boldsymbol{x}}\|_{\mathcal{H}_L'}^2 = \left\|K^L(\boldsymbol{x}, \cdot)\right\|_{\mathcal{H}_L}^2 = K^L(\boldsymbol{x}, \boldsymbol{x}) = \sum_{\ell=1}^{L} \sum_{j \in I_\ell} \phi_{\ell,j}^2(\boldsymbol{x}) \leq \sup_{\boldsymbol{y} \in \overline{\Omega}} \sum_{\ell=1}^{\infty} \sum_{j \in I_\ell} \phi_{\ell,j}^2(\boldsymbol{y}) =: C.$$

Then, we obtain from Propositions 4.2 and 4.3 the following result:

LEMMA 4.4. *Let $L \in \mathbb{N}$. There are functions $u_k : \Omega \to \mathbb{R}$, $k = 1, \ldots, N$, such that for all $f^L \in \mathcal{H}_L$, all $X_N \subset \Omega$ with* (4.6) *and all $\boldsymbol{x} \in \Omega$*

$$f^L(\boldsymbol{x}) = \sum_{n=1}^{N} u_n(\boldsymbol{x}) f^L(\boldsymbol{x}_n), \quad \textit{and} \quad \sum_{n=1}^{N} |u_n(\boldsymbol{x})| \leq \frac{1}{c} \|\delta_{\boldsymbol{x}}|_{\mathcal{H}_L}\|_{\mathcal{H}_L'} \leq \frac{C}{c}. \tag{4.9}$$

Lemma 4.4 can be seen as a generalized reproduction formula with uniformly bounded Lebesgue constant $\sup_{\boldsymbol{x} \in \Omega} \sum_{n=1}^{N} |u_n(\boldsymbol{x})|$. This reproduction formula is the main step to prove generalized (derivative-free) sampling inequalities in the following.

THEOREM 4.5. *Let $L \in \mathbb{N}$. Then for all $f \in L_\infty(\Omega)$ and all sets $X_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ such that* (4.6) *holds, we have*

$$\|f\|_{L_\infty(\Omega)} \leq \left(1 + \frac{C}{c}\right) \inf_{g^L \in \mathcal{H}_L} \left\|f - g^L\right\|_{L_\infty(\Omega)} + \frac{C}{c} \left\|S_{X_N}(f)\right\|_{\ell_\infty(X_N)}. \tag{4.10}$$

*Proof.* Let $f \in L_\infty(\Omega)$. We proceed similarly as for classical sampling inequalities [12, 29, 21]. For arbitrary $g^L \in \mathcal{H}_L$ with the functions $\{u_n\}_{n=1}^N$ from Lemma 4.4, we have

$$\begin{aligned}
\|f\|_{L_\infty(\Omega)} &\leq \|f - g^L\|_{L_\infty(\Omega)} + \|g^L\|_{L_\infty(\Omega)} = \|f - g^L\|_{L_\infty(\Omega)} + \left\|\sum_{n=1}^{N} g^L(\boldsymbol{x}_n) u_n\right\|_{L_\infty(\Omega)} \\
&\leq \|f - g^L\|_{L_\infty(\Omega)} + \left\|\sum_{n=1}^{N} \left(g^L(\boldsymbol{x}_n) - f(\boldsymbol{x}_n)\right) u_n\right\|_{L_\infty(\Omega)} + \left\|\sum_{n=1}^{N} f(\boldsymbol{x}_n) u_n\right\|_{L_\infty(\Omega)} \\
&\leq \left(1 + \left\|\sum_{n=1}^{N} |u_n|\right\|_{L_\infty(\Omega)}\right) \left\|f - g^L\right\|_{L_\infty(\Omega)} + \left\|\sum_{n=1}^{N} |u_n|\right\|_{L_\infty(\Omega)} \left\|S_{X_N}(f)\right\|_{\ell_\infty(X_N)} \\
&\leq \left(1 + \frac{C}{c}\right) \left\|f - g^L\right\|_{L_\infty(\Omega)} + \frac{C}{c} \left\|S_{X_N}(f)\right\|_{\ell_\infty(X_N)}.
\end{aligned}$$

Since this estimate holds for an arbitrary $g^L \in \mathcal{H}_L$, the result follows. ∎

The estimate (4.10) contains a best approximation error which can be further quantified by means of level-wise Jackson inequalities, see [16, Example 7] and also compare (4.3) where we set $\mathcal{V}_j = \mathcal{H}_L$, $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_{L_\infty(\Omega)}$ and $\mathcal{Y} = \mathcal{H}$. This results directly in the following lemma.

LEMMA 4.6. *Suppose that the pair $(\mathcal{H}, L_\infty(\Omega))$ satisfies a Jackson inequality with respect to the family $\{\mathcal{H}_L\}$, i.e. there are numbers $h(\mathcal{H}_L)$ such that for all $f \in \mathcal{H}$ and all $L \in \mathbb{N}$*

$$\inf_{g^L \in \mathcal{H}_L} \left\|f - g^L\right\|_{L_\infty(\Omega)} \leq h(\mathcal{H}_L) \|f\|_{\mathcal{H}}.$$

*If a discrete set $X_N \subset \Omega$ and a truncation index $L \in \mathbb{N}$ satisfy* (4.6)*, then for all $f \in L_\infty(\Omega)$,*

$$\|f\|_{L_\infty(\Omega)} \leq \left(1 + \frac{C}{c}\right) h(\mathcal{H}_L)\|f\|_{\mathcal{H}} + \frac{C}{c}\left\|S_{X_N}(f)\right\|_{\ell_\infty(X_N)}. \tag{4.11}$$

Theorem 4.5 and Lemma 4.6 might be seen as a derivative-free generalization of the sampling inequalities [19]. Roughly speaking classical sampling inequalities correspond to the choice $\mathcal{H} = W_2^\tau(\Omega)$ for $\tau > d/2$, and $\mathcal{H}_L := \pi_L(\Omega)$ where $\pi_L$ denotes the space of $d$-variate polynomials of degree at most $L$.

We now apply (4.11) to the residuum $R_{N,M}(f) := f - \mathcal{R}_{Y_M}(S_{X_N}(f))$. Let $L$ be such that (4.6) holds. Then there are constants such that for all $f \in \mathcal{H}$

$$\begin{aligned}
\|R_{N,M}(f)\|_{L_\infty(\Omega)} &\leq \left(1 + \frac{C}{c}\right) h(\mathcal{H}_L)\|R_{N,M}(f)\|_{\mathcal{H}} + \frac{C}{c}\|S_{X_N}(R_{N,M}(f))\|_{\ell_\infty(X_N)} \\
&\leq \left(1 + \frac{C}{c}\right) h(\mathcal{H}_L)\|R_{N,M}(f)\|_{\mathcal{H}} + \frac{C}{c}\varepsilon(M,N) \\
&\leq \left(1 + \frac{C}{c}\right) h(\mathcal{H}_L)\|f\|_{\mathcal{H}} + \frac{C}{c}\varepsilon(M,N),
\end{aligned}$$

where we used (4.1) in the last step. In particular, if $\varepsilon(M,N) \leq h(\mathcal{H}_L)\|f\|_{\mathcal{H}}$ then the error of the reconstruction process is of the same order as that of the best approximation from $\mathcal{H}_L$.

We now consider the practically most relevant example of Sobolev spaces in more detail. It is well-known that

$$K(\boldsymbol{x},\boldsymbol{y}) := \sum_{\ell=0}^{\infty} 2^{\ell(d-2\tau)}\left(\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \phi(2^\ell\boldsymbol{x}-\boldsymbol{k})\phi(2^\ell\boldsymbol{y}-\boldsymbol{k})\right) \tag{4.12}$$

reproduces in a space which is norm equivalent to the Sobolev space $W_2^\tau$ if $\phi \in W_2^\tau(\mathbb{R})$ is $r$-regular with $r > \tau > \frac{d}{2}$, for details see [14]. We then have the following approximation result, see [28, Cor. 11.33] and [14].

THEOREM 4.7. *Suppose that $\Omega \subset \mathbb{R}^d$ is bounded with a Lipschitz boundary and satisfies an interior cone condition. Let $K$ be the reproducing kernel for the Sobolev space $W_2^\tau(\Omega)$ with $\tau > d/2$. Then there exist constants $C > 0$ and $h_0 > 0$ such that for all sets $X_N \subset \Omega$ with $h_{X_N,\Omega} \leq h_0$ and all $f \in W_2^\tau(\Omega)$,*

$$\|f - s_{f,X_N}\|_{L^2(\Omega)} \leq C h_{X_N,\Omega}^{\tau}\|f\|_{W_2^\tau(\Omega)} \quad \text{and} \quad \|f - s_{f,X_N}\|_{L^\infty(\Omega)} \leq C h_{X_N,\Omega}^{\tau-\frac{d}{2}}\|f\|_{W_2^\tau(\Omega)},$$

*where $h_{X_N}$ denotes the fill-distance (cf.* (2.7)*) of the point set $X_N$ in $\Omega$.*

This can be seen as a Jackson-type estimate (4.3) with $\mathcal{V}_j = \mathcal{L}_{X_{N_j}}$ as defined in (1.1) with weak norm $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_{L^2(\Omega)}$ or $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_{L^\infty(\Omega)}$, respectively, and $\mathcal{Y} = \mathcal{H} = W_2^\tau(\Omega)$ with norm $\|\cdot\|_{\mathcal{H}}$. As a consequence, we finally obtain the following corollary of Theorem 4.1.

LEMMA 4.8. *Let $L$ satisfy the coupling condition* (2.13)*, and denote by $s_{X_N}^L$ the interpolation with the truncated kernel $K^L$ from* (3.2) *with the specific kernel $K$ from* (4.12)*. Then, we have*

$$\left\|f - s_{X_N}^L(f)\right\|_{L^2(\Omega)} \leq C h_{X_N,\Omega}^{\tau}\|f\|_{W_2^\tau(\Omega)} \quad \text{and} \quad \left\|f - s_{X_N}^L(f)\right\|_{L^\infty(\Omega)} \leq C h_{X_N,\Omega}^{\tau-\frac{d}{2}}\|f\|_{W_2^\tau(\Omega)}$$

*for all $f \in W_2^\tau(\Omega)$.*

Note that, for (2.13), we expect in this case $\sigma_{\min}\left(S_{X_N}S_{X_N}^\star\right) \sim q_{X_N}^{\tau-\frac{d}{2}}$, see [28], with the separation distance of the point set $X_N$. Altogether, we obtain the same convergence rates $\tau$ if we employ the truncated kernel $K^L$ instead of $K$.

At this point, let us state that evaluation of the truncated kernel $K^L$ involves certain costs. They are at least linear in $L$ and hence depend on the coupling condition (2.13). In order to give a quantitative estimate of these costs, we need to take the dependence of $\Lambda_L$ on $L$ into account. To this end, we consider two special cases for the decaying weights $\lambda_\ell$ of the kernel expansion (2.6). The first example assumes polynomially decaying weights, i.e., $\lambda_\ell \sim \ell^{-s}$ for some $s > 1$. In this case, we obtain

$$\Lambda_L = \sum_{\ell=L+1}^{\infty} \lambda_\ell \sim \sum_{\ell=L+1}^{\infty} \ell^{-s} \le \int_L^\infty x^{-s}\,dx \lesssim L^{1-s},$$

and thus the coupling condition

$$L \gtrsim \left(\frac{\sigma_{\min}\left(S_{X_N}S_{X_N}^\star\right)}{N}\right)^{\frac{1}{1-s}}.$$

The second example addresses geometrically decaying weights, i.e., $\lambda_\ell \sim \eta^\ell$ for some $\eta < 1$. In this case, we obtain

$$\Lambda_L = \sum_{\ell=L+1}^{\infty} \lambda_\ell \sim \sum_{\ell=L+1}^{\infty} \eta^\ell = \sum_{\ell=0}^{\infty}\eta^\ell - \sum_{\ell=0}^{L}\eta^\ell = \frac{1}{1-\eta} - \frac{1-\eta^{L+1}}{1-\eta} = \frac{\eta}{1-\eta}\eta^L \sim \lambda_L.$$

For the Sobolev kernel (4.12), we have $\eta = \frac{1}{2^{2\tau-d}}$, i.e., $\lambda_\ell = \left(\frac{1}{2^{2\tau-d}}\right)^\ell$. In this case, we expect for the smallest eigenvalue of the Gramian an estimate of the form $\sigma_{\min} \sim q_{X_N}^{2\tau-d}$ with the separation distance $q_{X_N}$ as defined in (2.7). For quasi-uniform points we have $q_{X_N} \sim N^{-\frac{1}{d}}$, which yields $\frac{\sigma_{\min}\left(S_{X_N}S_{X_N}^\star\right)}{N} \sim q_{X_N}^{2\tau}$. Hence, we obtain as final condition for the Sobolev kernel (4.12),

$$L \gtrsim \frac{2\tau}{\ln\left(2^{2\tau-d}\right)} \ln\left(q_{X_N}^{-1}\right) \sim \frac{2\tau}{d\ln\left(2^{2\tau-d}\right)} \ln\left(N\right),$$

and the costs for evaluating one entry of $K^L$ thus depend only logarithmically on the number of degrees of freedom $N$.

**5. Bernstein inequalities.** As already mentioned, Bernstein estimates are crucial for unsymmetric reconstruction processes. We recall the consistency condition (4.1)

$$\left\|S_{X_N}\left(\mathcal{R}_{Y_M}\left(S_{Y_M}(f)\right)\right) - S_{X_N}(f)\right\|_{\ell_\infty(X_N)} \le \varepsilon(M,N).$$

If $Y_M \neq X_N$, we cannot invoke classical interpolation techniques to ensure the existence of a feasible solution with $\varepsilon(M,N) = 0$, since the Gramian $K_{N,M}$ is unsymmetric and hence not invertible. The best we can expect is a full rank of the Gramian. This means that if a function $s_M$ from set $\mathcal{L}_{Y_M}$ vanishes on $X_N$ it vanishes everywhere, i.e., it is the trivial zero function. Such a result can be shown by a coupling of sampling inequalities and Bernstein estimates, for further details see [19]. To outline the idea, denote by $\|\cdot\|_{\mathcal{L}}$ the weak norm and by $\|\cdot\|_{\mathcal{H}}$ the strong norm and assume the following two inequalities:

$$\|f\|_{\mathcal{L}} \le h(N)\,\|f\|_{\mathcal{H}} + B\,\|S_{X_N}(f)\|_{\ell\infty(X_N)} \quad \text{for all } f \in \mathcal{H} \tag{5.1}$$

$$\|s_M\|_{\mathcal{H}} \le D\left(M\right)\|s_M\|_{\mathcal{L}} \quad \text{for all } s_M \in \mathcal{L}_{Y_M}, \tag{5.2}$$

with some constant $B > 0$. They key ingredient of our analysis relies on the fact that the first inequality only depends on $X_N$ while the second depends only on $Y_M$. Applying (5.1) to $f = s_M$ in the first step and using (5.2) in the second step yields

$$\|s_M\|_{\mathcal{L}} \leq h(N)\|s_M\|_{\mathcal{H}} + B\|S_{X_N}s_M\|_{\ell^\infty(X_N)} \leq h(N)D(M)\|s_M\|_{\mathcal{L}} + B\|S_{X_N}s_M\|_{\ell^\infty(X_N)}.$$

Hence, if $h(N)D(M) < 1$ holds, we have

$$\|s_M\|_{\mathcal{L}} \leq B(1 - h(N)D(M))^{-1}\|S_{X_N}s_M\|_{\ell^\infty(X_N)}$$

and the continuous norm $\|\cdot\|_{\mathcal{L}}$ is bounded by the discrete data on $X_N$. Altogether, for the numerical analysis of unsymmetric reconstruction processes, Bernstein estimates like (5.2) are urgently needed.

Thus, we finally also derive Bernstein inequalities for our setting using the truncated kernels $K^L$. In the following, we assume that Bernstein-type inequalities hold for sections of the frame. To this end, suppose that $\mathcal{H}$ is compactly embedded in the space $\mathcal{L}$ such that

$$\|v\|_{\mathcal{H}} \leq E(L)\|v\|_{\mathcal{L}} \quad \text{for all } v \in \mathcal{H}_L \text{ and all } L \in \mathbb{N}.$$

By the inclusion $\mathcal{H}_L \subset \mathcal{H}_{L+1}$, the term $E(L)$ is expected to grow with $L$. Motivated by dyadic wavelets and, more generally, by frames arising from stable subspace splittings (see [16, 17]), we suppose that $E(L)$ depends on $L$ via the weights $\lambda_\ell$. Note that if the weights $\lambda_\ell$ decrease geometrically then $\lambda_{L+1} \sim \lambda_L \sim \Lambda_L$ with constants independent of $L$, where $\Lambda_L$ is defined in (2.13). Hence, there is a function $\kappa : [0,\infty) \to [0,\infty)$ such that

$$\|v\|_{\mathcal{H}} \leq \kappa(\Lambda_{L-1})\|v\|_{\mathcal{L}} \quad \text{for all } v \in \mathcal{H}_L \text{ and all } L \in \mathbb{N}. \tag{5.3}$$

Since $\Lambda_L$ is monotonically decreasing as a function of $L$, the function $\kappa$ is supposed to be also monotonically decreasing. Assume further that with the notation from (2.9)

$$\left\|s^L_{\boldsymbol{\alpha},X_N}\right\|_{\mathcal{L}} \leq \|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{L}} \quad \text{for all } L \in \mathbb{N}, \quad \text{all } X_N = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}, \quad \text{and all } \boldsymbol{\alpha} \in \mathbb{R}^N. \tag{5.4}$$

Indeed, if the frame and its finite sections are also frames with respect to the $\mathcal{L}$-norm with frame constants independent of the level then (5.4) is satisfied. This is for example the case for frames that lead to reproducing kernels in Sobolev spaces $W^k_2(\Omega)$ on domains $\Omega$ such that the sections are also frames for $\mathcal{L} := L_2(\Omega)$ (see [17]). We than have the following lemma.

LEMMA 5.1. *Let* $\{\lambda_\ell^{1/2}\phi_{\ell,j} : \ell \in \mathbb{N}$ *and* $j \in I_\ell\}$ *be a tight frame in* $\mathcal{H}$ *satisfying* (5.3), (5.4) *and* (2.2), *and choose the minimal* $L > 1$ *such that* (2.13) *holds. Then, for all* $s_{\boldsymbol{\alpha},X_N} \in \mathcal{L}_{X_N}$, *we have*

$$\|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{H}} \leq \frac{\gamma}{\gamma - 1}\kappa\left(\frac{\sigma_{\min}\left(S_{X_N}S^*_{X_N}\right)}{NA\gamma^2}\right)\|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{L}}.$$

*Proof.* By Lemma 2.1, (5.3) and (5.4), we obtain

$$\|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{H}} \leq \frac{\gamma}{\gamma - 1}\left\|s^L_{\boldsymbol{\alpha},X_N}\right\|_{\mathcal{H}} \leq \gamma\kappa(\Lambda_{L-1})\left\|s^L_{\boldsymbol{\alpha},X_N}\right\|_{\mathcal{L}} \leq \gamma\kappa(\Lambda_{L-1})\|s_{\boldsymbol{\alpha},X_N}\|_{\mathcal{L}},$$

for all $s_{\boldsymbol{\alpha},X_N} \in \mathcal{L}_{X_N}$. Furthermore, we have $\Lambda_{L-1} > \frac{\sigma_{\min}(S_{X_N}S^*_{X_N})}{NA\gamma^2}$ by minimality of $L$. Thus since $\kappa$ is monotonically decreasing, we get

$$\kappa(\Lambda_{L-1}) \leq \kappa\left(\frac{\sigma_{\min}\left(S_{X_N}S^*_{X_N}\right)}{NA\gamma^2}\right),$$

which concludes the proof. ∎

In order to apply Lemma 5.1 one has to estimate $\sigma_{\min}\left(S_{X_N}S^\star_{X_N}\right)$ from below. We now give an estimate in terms of the weights of the frame which is a straight-forward generalization of [18, Lemma 5.23]. The proof can be found in the Appendix.

LEMMA 5.2. *Suppose that there is $a>1$ and $C_1$, $C_2>0$ such that for all $\ell\in\mathbb{N}$ and all $j\in I_\ell$*

$$C_1 a^{-\ell}\le \operatorname{diam}\operatorname{supp}\phi_{\ell,j}\le C_2 a^{-\ell}. \tag{5.5}$$

*Suppose further that there is a uniform level-wise lower bound*

$$\inf_{\ell\in\mathbb{N}}\sum_{j\in I_\ell}\phi_{\ell,j}(\boldsymbol{x})^2\ge\beta_0>0\quad\textit{for all } \boldsymbol{x}\in\Omega. \tag{5.6}$$

*Then, for all $X_N\subset\Omega$, it holds that*

$$\sigma_{\min}\left(S_{X_N}S^\star_{X_N}\right)\ge\beta_0\lambda_M\quad\textit{with}\quad M>\ln_a\left(\frac{C_2}{q_{X_N}}\right),$$

*where $\lambda_M$ denotes the weight of the kernel expansion* (1.2).

The assumptions of Lemma 5.2 are natural for many typical examples such as dyadic wavelets where the assumption (5.5) holds with $a=2$. Note furthermore that (5.6) is satisfied for instance for partition of unity functions occuring in modern discretizations schemes for partial differential equations, see [26, 27] for an overview. We only need the following two ingredients:

(i) There is $\gamma_0$ with the property: For all $\ell\in\mathbb{N}$ and all $j\in I_\ell$ the number of functions $\phi_{\ell,k}$, $k\in I_\ell$, such that the supports of $\phi_{\ell,k}$ and $\phi_{\ell,j}$ have an intersection of positive measure is bounded by $\gamma_0$, i.e.

$$\sup_{\ell\in\mathbb{N}}\sup_{j\in I_\ell}\left(\#\left\{k:\left|\operatorname{supp}\left(\phi_{\ell,j}\right)\cap\operatorname{supp}\left(\phi_{\ell,k}\right)\right|>0\right\}\right)\le\gamma_0.$$

(ii) For all $\ell\in\mathbb{N}$, the functions $\{\phi_{\ell,j}:j\in I_\ell\}$ are non-negative and build a partition of unity, i.e.

$$\sum_{j\in I_\ell}\phi_{\ell,j}(\boldsymbol{x})=1\quad\text{for all } \boldsymbol{x}\in\Omega.$$

Indeed if (i) and (ii) are satisfied then for all $\ell\in\mathbb{N}$

$$1\le\sum_{j\in I_\ell}|\phi_{\ell,j}(\boldsymbol{x})|=\sum_{\substack{j\in I_\ell\\ \operatorname{supp}\phi_{\ell,j}\cap\{\boldsymbol{x}\}\neq\emptyset}}|\phi_{\ell,j}(\boldsymbol{x})|\le\gamma_0^{\frac12}\left(\sum_{j\in I_\ell}|\phi_{\ell,j}(\boldsymbol{x})|^2\right)^{\frac12}.$$

As an example, let us finally consider the case of the Sobolev space $W_2^\tau$ again with its reproducing kernel (4.12). There is a constant $F>0$ such that for all $\ell\in\mathbb{N}$ for all $v_\ell\in\mathcal{H}_\ell$ (see [17])

$$\|v_\ell\|_{W_2^\tau}\le F2^{\ell\tau}\|v_\ell\|_{L^2}.$$

By Lemma 5.2 we have with $M\sim\log_2(1/q_{X_N})$ that $\sigma_{\min}\gtrsim\lambda_M\sim q^{\tau-\frac12}$. Following Lemma 2.1 we choose $L\in\mathbb{N}$ such that $\Lambda_L\sim\lambda_L\sim\frac{\sigma_{\min}}{N}\gtrsim q^{\tau+\frac12}$. Hence we obtain that (see Lemma 5.1) there is a constant $G>0$ such that

$$\|s_{\boldsymbol{\alpha},X_N}\|_{W_2^\tau}\le Gq^{\frac{1+2\tau}{1-2\tau}}\|s_{\boldsymbol{\alpha},X_N}\|_{L^2}. \tag{5.7}$$

for all $s_{\boldsymbol{\alpha},X_N} \in \mathcal{L}_{X_N}$.

Note that the condition in Lemma 5.1 on a minimal $L > 1$ such that $\Lambda_L$ satisfies (2.13) is not very restrictive. Otherwise the Gramian matrix from (2.12) would have a lowest eigenvalue independent of $X_N$ which is known to be true only for very coarse point sets. This makes it applicable to a variety of examples. The price for this generality is that, in some cases, the norm equivalence constants are not optimal with respect to the point set $X_N$. Moreover, the the rate $q^{\frac{1+2\tau}{1-2\tau}}$ of the inverse estimate (5.7) is inferior to the rate $q^{-\tau}$ expected for Sobolev spaces (see [28, Chap. 12]). However, such better estimates have, to the best of our knowledge, not yet been shown for kernels from (4.12). We would obtain the better order in $q_{X_N}$ if we could choose $\Lambda_L \sim \sigma_{\min}$ which is possible under certain further assumptions on the frame system, as will be shown elsewhere.

**6. Concluding remarks.** In conclusion, we showed that there is a theoretical justification for using kernel which are given as a infinite series expansion of the form (1.2). The fundamental step was to derive an explicit coupling of the discrete point set to the decay of the summation weights in (1.2). We furthermore proved that we do not loose accuracy (up to constants) if we employ a properly truncated kernel instead of the infinite expansion (1.2). Moreover, we showed that the finite dimensional spaces spanned by the first terms of the expansion provide good approximation spaces. This observation was formalized by deriving a sampling inequality based on these finite dimensional spaces. Our estimates can be employed in the error analysis of many regularized reconstruction algorithms. Since our sampling inequality is based on the kernel expansion only, our results can be applied where classical function spaces are not available, as it is often the case in machine learning applications.

**Appendix.** In this appendix we now give for reasons of completeness the proof of Lemma 5.2. It is a straight-forward generalization of [18, Lemma 5.23].

*Proof.* The Gramian matrix based on $X_N = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ can be decomposed into

$$\sum_{\ell \in \mathbb{N}} \lambda_\ell \left( \sum_{j \in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_n) \phi_{\ell,j}(\boldsymbol{x}_m) \right)_{n,m} =: \sum_{\ell \in \mathbb{N}} \lambda_\ell \boldsymbol{A}_\ell(X_N).$$

For fixed $\ell \geq 0$ we denote by $\sigma_{\min}(\boldsymbol{A}_\ell(X_N))$ the smallest eigenvalue of the matrix $\boldsymbol{A}_\ell(X_N)$. Note that the matrices $\boldsymbol{A}_\ell(X_N)$ are positive semi-definite since

$$\boldsymbol{A}_\ell(X_N) := \sum_{j \in I_\ell} \left( \phi_{\ell,j}(X_N) \otimes \phi_{\ell,j}(X_N) \right) := \sum_{j \in I_\ell} \begin{pmatrix} \phi_{\ell,j}(\boldsymbol{x}_1) \\ \vdots \\ \phi_{\ell,j}(\boldsymbol{x}_N) \end{pmatrix} \left( \phi_{\ell,j}(\boldsymbol{x}_1), \dots, \phi_{\ell,j}(\boldsymbol{x}_N) \right)$$

is a sum of rank-one matrices with non-negative spectrum. Hence, for all $M \in \mathbb{N}$ and all $\boldsymbol{\alpha} \in \mathbb{R}^N$, we obtain

$$\begin{aligned}\left(\boldsymbol{\alpha}, S_{X_N} S_{X_N}^\star \boldsymbol{\alpha}\right)_{\ell_2(X_N)} &= \sum_{\ell \in \mathbb{N}} \lambda_\ell \left(\boldsymbol{\alpha}, \boldsymbol{A}_\ell(X_N)\boldsymbol{\alpha}\right)_{\ell_2(X_N)} \geq \lambda_M \left(\boldsymbol{\alpha}, \boldsymbol{A}_M(X_N)\boldsymbol{\alpha}\right)_{\ell_2(X_N)} \\ &\geq \lambda_M \sigma_{\min}\left(\boldsymbol{A}_M(X_N)\right) \|\boldsymbol{\alpha}\|^2_{\ell_2(X_N)}.\end{aligned}$$

This implies in particular

$$\sigma_{\min}\left(S_{X_N} S_{X_N}^\star\right) \geq \lambda_M \sigma_{\min}\left(\boldsymbol{A}_M(X_N)\right) \quad \text{for all} \quad M \in \mathbb{N}. \tag{6.1}$$

Thus, we have to find a lower bound on $\sigma_{\min}\left(\boldsymbol{A}_M(X_N)\right)$ for a fixed $M \in \mathbb{N}$. To this end, we choose $M$ large enough such that for all $j \in I_M$

$$q_{X_N} > C_2 a^{-M} \geq \operatorname{diam}\operatorname{supp}\phi_{M,j} \quad \text{or equivalently} \quad M > \ln_a\left(\frac{C_2}{q_{X_N}}\right).$$

Then, by (5.6), we obtain

$$\sum_{j \in I_\ell} \phi_{\ell,j}(\boldsymbol{x}_n)\phi_{\ell,j}(\boldsymbol{x}_m) = \sum_{j \in I_\ell} \phi_{\ell,j}^2(\boldsymbol{x}_n) \geq \beta_0 \quad \text{for all } \ell \geq M,$$

that is, the matrix $\boldsymbol{A}_\ell\left(X_N\right)$ is diagonal and the entries on the diagonal are bounded below by $\beta_0 > 0$. Hence, $\sigma_{\min}\left(\boldsymbol{A}_M(X_N)\right) \geq \beta_0$, and, by (6.1), we finally get

$$\sigma_{\min}\left(S_{X_N} S_{X_N}^\star\right) \geq \lambda_M \beta_0.$$

This yields the desired lower bound. $\square$

**Acknowledgements.** This work was supported in part by the Collaborative Research Centre (SFB) 1060 of the Deutsche Forschungsgemeinschaft. Some of the results are contained in Barbara Zwicknagl's PhD thesis defended at the University of Bonn (see [30]). Michael Griebel and Christian Rieger acknowledge partial support by the DFG through the Collaborative Research Centers (SFB) 611 and the Hausdorff Center for Mathematics. Barbara Zwicknagl thanks the Center for Nonlinear Analysis (NSF Grant No. DMS-0635983), where part of this research was carried out. Her research was partly funded by a postdoctoral fellowship of the National Science Foundation under Grant No. DMS-0905778.

## REFERENCES

[1] R. Arcangéli, M. L. di Silanes, and J. Torrens, *An extension of a bound for functions in Sobolev spaces, with applications to $(m,s)$-spline interpolation and smoothing*, Numer. Math., 107(2) (2007), pp. 181–211.
[2] ———, *Estimates for functions in Sobolev spaces defined on unbounded domains*, Journal of Approximation Theory, 161 (2009), pp. 198 – 212.
[3] ———, *Extension of sampling inequalities to Sobolev semi-norms of fractional order and derivative data*, Numerische Mathematik, 121 (2012), pp. 587–608.
[4] S. Atluri and S. Shen, *The Meshless Local Petrov-Galerkin (MLPG) Method*, Tech Science Press, Encino, CA, 2002.
[5] M. Buhmann, *Radial Basis Functions: Theory and Implementations*, Cambridge University Press, Cambridge, 2003.
[6] N. Christianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge Univ. Press, 2000.
[7] G. Fasshauer, *Meshfree Approximation methods with MATLAB*, Interdisciplinary Mathematical Sciences - Vol. 6, World Scientific Publishers, Singapore, 2006.
[8] G. Fasshauer and M. McCourt, *Stable evaluation of Gaussian radial basis function interpolants*, SIAM Journal on Scientific Computing, 34 (2012), pp. A737–A762.
[9] J. Krebs, *Support vector regression for the solution of linear integral equations*, Inverse Problems, 27 (2011), p. 065007.
[10] J. Krebs, A. Louis, and H. Wendland, *Sobolev error estimates and a priori parameter selection for semi-discrete Tikhonov regularization*, Journal of Inverse and Ill-Posed Problems, 17 (2009), pp. 845–869.
[11] W. Madych, *An estimate for multivariate approximation II*, J. Approx. Theory, 142 (2006), pp. 116–128.
[12] F. Narcowich, J. Ward, and H. Wendland, *Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting*, Math. Comp., 74 (2005), pp. 743–763.
[13] ———, *Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions*, Constr. Approx., 24 (2006), pp. 175–186.
[14] R. Opfer, *Multiscale kernels*, Adv. Comp. Math., 25 (2006), pp. 357–380.
[15] ———, *Tight frame expansions of multiscale reproducing kernels in Sobolev spaces*, Appl. Comput. Harm. Anal., 20 (2006), pp. 357–374.
[16] P. Oswald, *Frames and space splittings in Hilbert spaces*. Survey Lecture given at the Universities Bonn and Lancaster in 1997 available at http://www.faculty.jacobs-university.de/poswald, 1997.

[17] ———, *Multilevel frames and Riesz bases in Sobolev spaces*. Survey Lecture given at the Universities Bonn and Lancaster in 1997 available at http://www.faculty.jacobs-university.de/poswald, 1997.

[18] R. A. RESSEL, *Unsymmetrische Kollokationsverfahren mittels Multiskalenkernen*, diploma thesis, Institut für Numerische und Angewandte Mathematik, Universität Göttingen, 2008.

[19] C. RIEGER, R. SCHABACK, AND B. ZWICKNAGL, *Sampling and stability*, in Mathematical Methods for Curves and Surfaces, vol. 5862 of Lecture Notes in Computer Science, Springer, New York, 2010, pp. 347–369.

[20] C. RIEGER AND B. ZWICKNAGL, *Deterministic error analysis of support vector machines and related regularized kernel methods*, J. Mach. Learn. Res., 10 (2009), pp. 2115–2132.

[21] ———, *Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning*, Adv. Comp. Math., 32(1) (2010), pp. 103–129.

[22] L. ROSASCO, M. BELKIN, AND E. DE VITO, *On learning with integral operators*, J. Mach. Learn. Res., 11 (2010), pp. 905–934.

[23] R. SCHABACK AND H. WENDLAND, *Approximation by positive definite kernels*, in Advanced Problems in Constructive Approximation, vol. 142, International Series in Numerical Mathematics, 2002, pp. 203–221.

[24] R. SCHABACK AND H. WENDLAND, *Kernel techniques: From machine learning to meshless methods*, Acta Numerica, 15 (2006), pp. 543–639.

[25] B. SCHÖLKOPF AND A. SMOLA, *Learning with Kernels - Support Vector Machines, Regularisation, and Beyond*, MIT Press, Cambridge, Massachusetts, 2002.

[26] M. A. SCHWEITZER, *A Parallel Multilevel Partition of Unity Method for Elliptic Partial Differential Equations*, vol. 29 of Lecture Notes in Computational Science and Engineering, Springer, 2003.

[27] M. A. SCHWEITZER, *Meshfree and Generalized Finite Element Methods*, habilitation, Institute for Numerical Simulation, University of Bonn, 2008.

[28] H. WENDLAND, *Scattered Data Approximation*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2005.

[29] H. WENDLAND AND C. RIEGER, *Approximate interpolation with applications to selecting smoothing parameters*, Numer. Math., 101 (2005), pp. 729–748.

[30] B. ZWICKNAGL, *Mathematical analysis of microstructures and low hysteresis shape memory alloys*, phd thesis, Institute for Applied Mathematics, University of Bonn, 2011.