Institut für Numerische Simulation

Rheinische Friedrich-Wilhelms-Universität Bonn

M. Griebel, C. Rieger and B. Zwicknagl

Regularized kernel based reconstruction in generalized Besov spaces

Regularized kernel based reconstruction in generalized Besov spaces

Michael Griebel · Christian Rieger[*] · Barbara
Zwicknagl

Abstract We present a theoretical framework for reproducing kernel based reconstruction methods in certain generalized Besov spaces based on positive, essentially self-adjoint operators. An explicit representation of the reproducing kernel is given in terms of an infinite series. We provide stability estimates for the kernel, including inverse Bernstein-type estimates for kernel-based trial spaces, and we give condition estimates for the interpolation matrix. Then, a deterministic error analysis for regularized reconstruction schemes is presented by means of sampling inequalities. In particular, we provide error bounds for a regularized reconstruction scheme based on a numerically feasible approximation of the kernel. This allows us to derive explicit coupling relations between the series truncation, the regularization parameters and the data set.

Keywords Reproducing kernels · A priori error analysis · Generalized Besov spaces · Feasible reconstruction schemes · Spline smoothing

Mathematics Subject Classification (2010) 41A17 · 41A25 · 41A58 · 42A82 · 62G08

1 Introduction

In this article, we develop an analysis for numerically feasible reproducing kernel based reconstruction methods in the general setting of metric measure spaces with heat kernel induced geometries; see [10,17]. Such spaces are of practical importance in various machine learning applications. There, high dimensional reconstruction problems in the Euclidean setting suffer from the so-called *curse of dimension*; see [15,26]. One reason why reconstruction is, however, sometimes still feasible is that the underlying metric structure of the data is highly non-Euclidean. Over the last years, there has been large interest in exploiting this fact; see for instance [16,61]. Moreover, practical data analysis and machine learning tasks are usually formulated in the setting of probability spaces, i.e. measures spaces; see for instance [52]. A further analytical challenge for the numerical analysis of

Michael Griebel

Fraunhofer-Institut für Algorithmen und Wissenschaftliches Rechnen SCAI, Schloss Birlinghoven, 53754 Sankt Augustin, Germany

Michael Griebel · Christian Rieger

Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6-8, 53115 Bonn, Germany E-mail: {griebel, rieger}@ins.uni-bonn.de Tel: +49 228 7360462, Fax: +49 228 7360462

Barbara Zwicknagl

Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany E-mail: zwicknagl@iam.uni-bonn.de

machine learning problems is the fact that the underlying domain, in which the data is contained, might not be an easily recognizable (sub-)manifold of the ambient Euclidean space but might rather have a discrete structure such as a graph or a tree; see [5,6], and see [14] for reproducing kernels on Riemannian manifolds. The advantage of the general framework presented here, is that both, continuous Riemannian manifolds and discrete structures can be treated simultaneously. One focus of recent research in this direction has been on diffusion polynomials, see for instance [30]. Moreover, approximation problems in such spaces have also gained attention. Here, to the best of our knowledge, most works have been devoted to wavelet-type approximation schemes, see [33], and also [38] for the sphere and [8] for the torus.

Furthermore, many successful algorithms in machine learning make use of a reproducing kernel Hilbert space structure, see [50]. A practical benefit of those methods is that they have an energy optimization principle in the background with a solution which, though nominally the solution of an infinite dimensional optimization problem, can be expressed in terms of finite linear combinations of the kernel. The weights can usually be obtained by solving a finite dimensional optimization problem whose dimension is linear in the number of data points. Therefore, it is desirable to have reproducing kernels in this general framework available, see also [52].

To this end, generalized Besov spaces $\mathcal{B}_{p,q}^{\sigma}(M;\mathcal{D})$ were introduced in [10], which are based on an essentially self-adjoint operator $\mathcal{D}$ and its associated heat kernel defined on a rather general metric measure space (M, ρ, μ). This setting covers in particular uniformly elliptic operators in divergence form, Laplace-Beltrami operators on Riemannian manifolds with non-negative Ricci curvatures, heat kernels generated by Jacobi operators, but also Schrödinger-type operators, see [19], and graph Laplacians if the graph satisfies a relative Faber-Krahn inequality, see [9] and also [24]. The embedding properties derived in [10] show that the generalized Besov spaces $\mathcal{B}_{p,q}^{\sigma}(M;\mathcal{D})$ are indeed reproducing kernel Hilbert spaces for $p = q = 2$ and $\sigma > 0$ large enough. For those spaces, we derive an explicit multi-scale representation of the associated reproducing kernel, see Theorem 2, i.e.,

$$K^{(\sigma)}(x,y) := \sum_{\ell=0}^{\infty} b^{-2\sigma\ell} \boldsymbol{S}^{(\ell;\sigma)}(x,y), \tag{1}$$

where $b > 1$ is a fixed parameter, $\sigma > 0$ determines the smoothness, and $\boldsymbol{S}^{(\ell;\sigma)}$ are integral kernels to an operator defined via smooth functional calculus in terms of the operator $\mathcal{D}$, see Sections 2 and 3 for the precise definitions. We show how the analysis from [18,41,42] for kernel-based approximation with kernels given in series form can be extended to this setting, and we show improved results using properties of kernels of the specific multiscale form (1).

Then, our main focus lies on the error analysis of reconstruction processes of the following form. Given data $y_n = f(x_n)$, $x_n \in X_N \subset M$, generated by an unknown function $f \in \mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$ we consider minimizers of the regression functional

$$J_{\mathbf{y};\alpha;X_N}(g) = \sum_{n=1}^{N} (g(x_n) - y_n)^2 + \alpha \|g\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}^2 \tag{2}$$

with a regularization parameter $\alpha > 0$. It is well known that the minimizer lies in

$$\mathcal{L}_{X_N} := \operatorname{span}\{K^{(\sigma)}(x_1,\cdot),\,\ldots,\,K^{(\sigma)}(x_N,\cdot)\}. \tag{3}$$

The main technical difficulty here is that the kernel $K^{(\sigma)}$ of (3) is usually not available in closed form, and hence, further numerical approximations have to be incorporated to obtain a numerically feasible approximation. In particular, we make use of the multi-scale representation (1) and address both, a careful truncation of the infinite series and the numerical errors in the approximative

computation of the spectral projections. Thus, we first consider in the spirit of [18] the error made when, instead of $\mathcal{L}_{X_N}$ from (3), one uses

$$\mathcal{L}^L_{X_N} := \operatorname{span}\{K^{(\sigma,L)}(x_1,\cdot),\,\ldots,\,K^{(\sigma,L)}(x_N,\cdot)\} \quad \text{with} \quad K^{(\sigma,L)}(x,y) := \sum_{\ell=0}^{L} b^{-2\sigma\ell}\boldsymbol{S}^{(\ell;\sigma)}(x,y). \quad (4)$$

In view of the Mairhuber-Curtis theorem [12,31], the truncation parameter L has to be coupled to the point set X_N in order to maintain unisolvency. To this end, we derive an explicit lower bound on L depending on X_N, which guarantees the existence of a quasi-optimal interplant from $\mathcal{L}^L_{X_N}$ at X_N for arbitrary data. The lower bound takes the form $L \geq \ln(cq_{X_N}^{-1})$ with the separation distance $q_{X_N} := \inf_{x_n \neq x_m \in X_N} \rho(x_n, x_m)$ and a generic constant $c > 0$. Furthermore, in practical applications the mere truncation (4) of the kernel is not yet sufficient since, in general, the eigenfunctions and eigenvectors of $\mathcal{D}$ which enter the definition of $\boldsymbol{S}^{(\ell;\sigma)}$ are not analytically available and must be approximated properly. Therefore, we take into account also the numerical error made by approximating these eigenfunctions with prescribed accuracy.

Subsequently, the discretization error is addressed by means of sampling inequalities for functions from generalized Besov spaces $\mathcal{B}^\sigma_{p,q}(M;\mathcal{D})$. Such inequalities have been studied over the last years for classical Sobolev spaces, see e.g. [2–4,18,29,40,45,47,55,59] and have proven to be useful for the error analysis of stable and consistent approximation schemes, including spline smoothing and support vector regression algorithms, see [45,46,59]. A typical example takes the form

$$\|f\|_{L^\infty(M;d\mu)} \leq C\left(h^{\sigma-d/p}_{X_N,M}\|f\|_{\mathcal{B}^\sigma_{p,q}(M;\mathcal{D})} + \|f|_{X_N}\|_{\ell^\infty(X_N)}\right), \quad (5)$$

where $\sigma > d/p$, and $X_N = \{x_1,\ldots,x_N\} \subset M$ is a discrete point set with sufficiently small fill distance $h_{X_N,M} := \sup_{z\in M}\inf_{x_n\in X_N}\rho(z,x_n) \leq h_0$, compare also Theorem 6. Here, d stands for a parameter of the metric measure space M which generalizes the notion of dimension, see Section 2. Such inequalities can be used to obtain approximation error bounds for a large class of reconstruction methods in the following way, see [45] and the references given there: Suppose that a reconstruction process assigns to any function f sampled at the discrete points X_N an approximation $\mathcal{R}f$ for which a stability property $\|f - \mathcal{R}f\|_{\mathcal{B}^\sigma_{p,q}(M;\mathcal{D})} \leq C\|f\|_{\mathcal{B}^\sigma_{p,q,}(M;\mathcal{D})}$ and a consistency property $\max_{x_n\in X_N}|f(x_n) - \mathcal{R}f(x_n)| \leq G(f)$ holds. Then, by (5) applied to the residual $f - \mathcal{R}f$ we obtain the error estimate

$$\|f - \mathcal{R}f\|_{L^\infty(M;d\mu)} \leq C\left(h^{\sigma-d/p}_{X_N,M}\|f\|_{\mathcal{B}^\sigma_{p,q}(M;\mathcal{D})} + G(f)\right).$$

Thus applied to a numerically feasible approximation of $J_{\mathbf{y};\alpha;X_N}$, we obtain an upper bound on the approximation error which is explicit in the problem parameters and hence suggests coupling conditions of the parameters to ensure asymptotic convergence.

As a side product of our analysis, we derive two important stability properties, namely inverse Bernstein-type inequalities for $\mathcal{L}_{X_N}$, and lower bounds on the smallest eigenvalues $\lambda_{\min}$ of interpolation matrices based on $K^{(\sigma)}$. Precisely, we show (see Proposition 7), that for all point sets X_N with sufficiently small separation distance q_{X_N}, it holds

$$\|f\|_{\mathcal{B}^\sigma_{2,2}(M;\mathcal{D})} \leq Cq_{X_N}^{-\sigma}\|f\|_{L^2(M;d\mu)} \qquad \text{for all} \quad f \in \mathcal{L}_{X_N}, \quad (6)$$

and (see Proposition 8)

$$\lambda_{\min}\left(K^{(\sigma)}_{X_N,X_N}\right) \geq Cq_{X_N}^{2\sigma+d}, \quad (7)$$

where $K^{(\sigma)}_{X_N,X_N} := \left(K^{(\sigma)}(x_n,x_m)\right)_{x_n,x_m\in X_N}$ denotes the Gramian matrix. Proving Bernstein estimates of the form (6) for kernel-based trial spaces has been an active area of research during the

last years (see, e.g., [21, 32, 34, 36, 37, 44, 48, 56]). If combined with sampling inequalities of the form (5), such inverse estimates can be used to prove stability of reconstruction schemes (see [45]). We point out that we obtain inverse estimates for a variety of domains, including such with boundaries. In our case, the kernel depends explicitly on the domain, while often the kernel is defined on the whole space, and inverse estimates are proven for subdomains.

On the other hand, eigenvalue estimates of the form (7) can be used to get upper and lower bounds on the covering numbers which are frequently used in learning theory [11, 62].

The remainder of this paper is organized as follows: In Section 2, we introduce the necessary technical framework of metric measure spaces and heat kernels. In Section 3, we show how certain generalized Besov spaces carry the additional structure of a reproducing kernel Hilbert space and we characterize their kernels as infinite series (see Theorem 2). In Section 4, we present sampling inequalities for Besov spaces. Here we discuss two approaches. First, to derive bounds on arbitrary $L^p(M; d\mu)$-norms of Besov functions also on measure spaces with infinite mass, we consider maximal δ-nets as sampling points (see Theorem 3). Second, for the special case of finite volume measure spaces, we derive bounds on the strongest $L^\infty(M; d\mu)$-norm for more general data sets, see Theorem 6. In Section 5, the truncation of the kernel series expansion (1) is discussed. We derive an explicit lower bound for the truncation parameter L depending on the point set X_N which guarantees the existence of a quasi-optimal interpolant to arbitrary given data at X_N, see Theorem 7. The essential step in the proof is the derivation of a Riesz basis involving the integral kernels $\boldsymbol{S}^{(\ell;\sigma)}$, see Lemma 7. Subsection 5.2 is devoted to the proof of two stability properties of kernel-based approximation. Here we show inverse Bernstein-type estimates for $\mathcal{L}_{X_N}$ as defined in (3), see Proposition 7, and derive bounds on the condition of the interpolation matrix based on the kernel $K^{(\sigma)}$, see Proposition 8. Finally, in section 6, we analyze a regularized reconstruction method using a numerically feasible approximation of the kernel function. Here, besides the discretization error, we take into account the series truncation error and the error that stems from the numerical approximation of the spectrum of $\mathcal{D}$, see Theorem 9.

2 Notation and auxiliary results

In the sequel we denote by C and c generic positive constants that may change from line to line and from expression to expression. We denote by $C^\infty([0, \infty))$ the space of smooth functions $[0, \infty) \to \mathbb{R}$, and for $R > 0$ we denote by $C_c^\infty([0, R])$ the subspace of $C^\infty([0, \infty))$ of functions with compact support in $[0, R]$.

2.1 Measurable metric spaces

Let us use the framework of [10], and briefly recall basic definitions and assumptions. Suppose (M, ρ, μ) is a metric measure space with the following properties:

(i) (M, ρ) is a locally compact separable metric space with distance $\rho : M \times M \to [0, \infty)$. Furthermore, μ is a positive Radon measure with the *volume doubling property*, i.e., there is a constant $d > 0$ such that

$$0 < \mu(B(x, 2r)) \leq 2^d \mu(B(x, r)) < \infty \quad \text{for all } x \in M \text{ and } r > 0. \tag{8}$$

Here $B(x, r) := \{y \in M : \rho(x, y) < r\}$ denotes the open ball with radius r around x, and the constant d is a parameter of the space (M, ρ), which generalizes the notion of dimension.

(ii) The *reverse doubling condition* holds, i.e., there is a constant $\beta > 0$ such that

$$\mu(B(x, 2r)) \geq 2^\beta \mu(B(x, r)) \quad \text{for all } x \in M \text{ and all } 0 < r < \operatorname{diam} M/3. \tag{9}$$

Note that in general β can be different from d. The reverse doubling condition follows from (8) if M is connected (see [10, Proposition 2.2]).

(iii) The *non-collapsing condition* holds, i.e., there is a constant $c > 0$ such that

$$\inf_{x \in M} \mu\left(B\left(x, 1\right)\right) \geq c > 0 \quad \text{for all } x \in M. \tag{10}$$

It follows from (8) that for $x, y \in M$ and $r > 0$ (see [10, (2.2)])

$$\mu\left(B\left(y, r\right)\right) \leq 2^d \left(1 + \frac{\rho\left(x, y\right)}{r}\right)^d \mu\left(B\left(x, r\right)\right). \tag{11}$$

If M is connected and $\mu(M) < \infty$, then (8) implies (9) and (10), see [10]. We refer to [25] for a discussion of the parameters and their connection to the Assoud dimension.

2.2 Heat kernels on metric measure spaces

We recall that a family $(p_t)_{t>0}$ of kernel functions $p_t : M \times M \to \mathbb{R}$ is called *heat kernel* if for almost all $x, y \in M$, all $s, t > 0$ and all $f \in L^2\left(M; d\mu\right)$

$$p_t(x, y) \geq 0, \quad p_t(x, y) = p_t(y, x), \quad \int_M p_t(x, y) d\mu(y) \leq 1,$$

$$p_{t+s}(x, y) = \int_M p_t(x, z) p_s(y, z) d\mu(z), \quad L^2-\lim_{t \to 0^+} \int_M p_t(x, z) f(z) d\mu(z) = f(x), \tag{12}$$

where we use the notation $L^2-\lim_{t \to 0} a_t = b$ if $\lim_{t \to 0} \|a_t - b\|_{L^2(M; d\mu)} = 0$. As in [10], we impose the following additional conditions on the heat kernel:

(i) *Small time Gaussian upper bound*: For all $0 < t \leq 1$ and $x, y \in M$

$$|p_t\left(x, y\right)| \leq C \frac{\exp\left(-c\frac{\rho^2(x, y)}{t}\right)}{\sqrt{\mu\left(B\left(x, \sqrt{t}\right)\right) \mu\left(B\left(y, \sqrt{t}\right)\right)}}. \tag{13}$$

(ii) *Hölder continuity with exponent* $\alpha_H > 0$: For all $0 < t \leq 1$ and $x, y, \tilde{y} \in M$ with $\rho(y, \tilde{y}) \leq \sqrt{t}$,

$$|p_t\left(x, y\right) - p_t\left(x, \tilde{y}\right)| \leq C_1 \left(\frac{\rho\left(y, \tilde{y}\right)}{\sqrt{t}}\right)^{\alpha_H} \frac{\exp\left(-c\frac{\rho^2(x, y)}{t}\right)}{\sqrt{\mu\left(B\left(x, \sqrt{t}\right)\right) \mu\left(B\left(y, \sqrt{t}\right)\right)}}. \tag{14}$$

(iii) *Markov property*: For all $t > 0$

$$\int_M p_t\left(x, y\right) d\mu\left(y\right) \equiv 1. \tag{15}$$

A heat kernel gives rise to a family of operators

$$P_t : L^2\left(M; d\mu\right) \to L^2\left(M; d\mu\right), \qquad \left(P_t f\right)\left(x\right) := \begin{cases} f(x) & \text{if } t = 0, \\ \int_M p_t\left(x, y\right) f\left(y\right) d\mu\left(y\right) & \text{if } t > 0. \end{cases}$$

As worked out in [20], the conditions on the heat kernel ensure that the family of operators $\{P_t\}_{t>0}$ is a strongly continuous, symmetric Markovian semigroup in $L^2\left(M; d\mu\right)$. The associated infinitesimal generator $\mathcal{D}$ is defined by

$$\mathcal{D}f := L^2-\lim_{t \to 0} \frac{f - P_t f}{t}.$$

Moreover, the domain $\operatorname{Dom}\mathcal{D}$ of $\mathcal{D}$, i.e., the subspace of $L^2(M;d\mu)$ for which the limit exists, is a dense subspace of $L^2(M;d\mu)$. By construction, $\mathcal{D}$ is a self-adjoint and positive definite operator. Furthermore, there is a unique associated spectral resolution of the identity, denoted by $\{\mathbb{E}_\lambda\}_{\lambda\in[0,\infty)}$, such that $\mathbb{E}_\lambda$ is a bounded linear operator $L^2(M;d\mu) \to L^2(M;d\mu)$ for every $\lambda > 0$, and

$$\mathcal{D}f = \int_0^\infty \lambda\, d\mathbb{E}_\lambda f \qquad \text{for all} \quad f \in \operatorname{Dom}\mathcal{D} = \left\{ f \in L^2(M;d\mu) \ : \ \int_0^\infty \lambda^2 d\left(\mathbb{E}_\lambda f, f\right)_{L^2(M;d\mu)} < \infty \right\}.$$

The spectral resolution can be used to define for continuous $\boldsymbol{t}:\mathbb{R}\to[0,\infty)$ the operator

$$\boldsymbol{t}\left(\mathcal{D}\right)f := \int_0^\infty \boldsymbol{t}(\lambda)d\mathbb{E}_\lambda f \qquad \text{for all} \quad f \in L^2(M;d\mu) \quad \text{with} \quad \int_0^\infty \boldsymbol{t}(\lambda)^2 d\left(\mathbb{E}_\lambda f, f\right)_{L^2(M;d\mu)} < \infty. \tag{16}$$

Such operators are often integral operators. In [13, Theorem 6], it is shown that an operator $\mathfrak{g}: L^1(M;d\mu) \to L^\infty(M;d\mu)$ is bounded if and only if there is an integral kernel $G: M \times M \to \mathbb{R}$ satisfying $G \in L^\infty(M \times M)$ and

$$(\mathfrak{g}f)(x) = \int_M G(x,y)f(y)\,d\mu(y) \qquad \text{for almost every} \quad x \in M.$$

From now on, we always consider a metric measure space (M,ρ,μ) with the above outlined properties, in particular (8), (9) and (10), and an essentially self-adjoint positive operator $\mathcal{D}$ on $L^2(M;d\mu)$ with corresponding heat kernel satisfying (13), (14) and (15). Throughout the remainder of this paper, we employ the following notation.

Notation 1 *We denote smooth functions $[0,\infty) \to [0,\infty)$ that are used to define operators via spectral calculus by bold lowercase letters, and the corresponding (integral) kernels by the corresponding bold uppercase letters. Thus, for $\boldsymbol{t}:\mathbb{R}\to\mathbb{R}$ and $\delta > 0$ such that $\boldsymbol{t}(\delta\sqrt{\mathcal{D}})$ given by (16) is bounded as operator $L^1(M;d\mu) \to L^\infty(M;d\mu)$, we denote the associated integral kernel by $\boldsymbol{T}_\delta$. Precisely, for $0 < \delta \le 1$ and a smooth function $\boldsymbol{t}:[0,\infty)\to[0,\infty)$, we set*

$$\boldsymbol{t}(\delta\sqrt{\mathcal{D}})u(x) =: \int_M \boldsymbol{T}_\delta(x,y)u(y)d\mu(y) \qquad \text{for almost every} \quad x \in M. \tag{17}$$

Moreover we set $\boldsymbol{T} := \boldsymbol{T}_1$.

We note that under the above assumptions the kernel $\boldsymbol{T}_\delta(x,y)$ is real-valued, see [27, Sec. 2.5].

2.3 Smooth cut-off functions

Next, we will need several smooth cut-off functions with various properties that we collect in this subsection. Our computations follow the lines of [10].

Definition 1 Let $b > 1$, $\zeta \ge 1$, $1 > c_2 \ge c_1 > 0$ and $c > 0$.

(a) We set

$$\mathcal{G}(b,\zeta) := \left\{ \boldsymbol{t} \in C^\infty([0,\infty);[0,\infty)) : \ \left.\frac{d^\ell}{du^\ell}\boldsymbol{t}(u)\right|_{u=0} = 0 \text{ for all } 1 \le \ell \le \zeta, \text{ and } \operatorname{supp}\boldsymbol{t} \subset [0,b] \right\}. \tag{18}$$

Elements of $\mathcal{G}(b,\zeta)$ are called smooth cut-off functions of order ζ with support b.
We set

$$\mathcal{G}(b) := \bigcap_{\zeta\ge 1} \mathcal{G}(b,\zeta). \tag{19}$$

A function $\boldsymbol{t} \in \mathcal{G}(b,\zeta)$ is called normalized if $0 \le \boldsymbol{t} \le 1$.

(b) We define $\mathcal{A}(b, c_1, c_2)$ to consist of those normalized $\boldsymbol{t} \in \mathcal{G}(b)$ for which
 (i) $\operatorname{supp} \boldsymbol{t} \subset [0, b]$,
 (ii) $\boldsymbol{t} \equiv 1$ on $[0, 1]$,
 (iii) $\boldsymbol{t} \geq c_1 > 0$ on $[0, b^{3/4}]$, and
 (iv) $\boldsymbol{t} \leq c_2$ on $[b^{1/4}, b]$.
(c) We define $\mathcal{E}(b, c)$ to consist of those normalized $\boldsymbol{t} \in \mathcal{G}(b)$ for which
 (v) $\operatorname{supp} \boldsymbol{t} \subset [b^{-1}, b]$,
 (vi) $\boldsymbol{t} \geq c > 0$ on $[b^{-3/4}, b^{3/4}]$.
(d) A pair $(\boldsymbol{t}, \boldsymbol{p}) \in \mathcal{G}(b) \times \mathcal{G}(b)$ is called a partition of unity if

$$\boldsymbol{t}(u) + \sum_{\ell=1}^{\infty} \boldsymbol{p}(b^{-\ell} u) = 1 \qquad \text{for all } u \in [0, \infty). \tag{20}$$

Remark 1 Let $b > 1$ and $c > 0$. If $(\boldsymbol{t}, \boldsymbol{p}) \in \mathcal{G}(b) \times \mathcal{G}(b)$ is a partition of unity, then $\boldsymbol{t}(0) = 1$ and $\boldsymbol{p}(0) = 0$.

We introduce some abbreviations that will be used for the rest of this paper. From now on, we assume that $b > 1$ is fixed, and the generic constants c and C may depend on b without further mentioning.

Notation 2 *For $b > 1$ and $\boldsymbol{t} \in C^{\infty}([0, \infty))$, we set*

$$\Delta \boldsymbol{t}(\cdot) := \boldsymbol{t}(\cdot) - \boldsymbol{t}(b\cdot), \tag{21}$$

and, for $\ell \in \mathbb{N}$

$$\Delta_\ell \boldsymbol{t}(u) := \begin{cases} \boldsymbol{t}(u) & \text{if } \ell = 0, \\ \Delta \boldsymbol{t}\left(b^{-\ell} u\right) & \text{if } \ell \geq 1. \end{cases} \tag{22}$$

Lemma 1 *Let $\boldsymbol{t} \in C^{\infty}([0, \infty))$ with $\boldsymbol{t}(0) = 1$. Then*

$$\sum_{\ell=0}^{L} \Delta_\ell \boldsymbol{t}(u) = \boldsymbol{t}\left(b^{-L} u\right) \quad \text{for all } u \in [0, \infty) \tag{23}$$

and $(\boldsymbol{t}, \Delta \boldsymbol{t})$ form a partition of unity.

Proof We compute the partial sums for $L \in \mathbb{N}$. By a telescopic sum argument, we have for every $u \in [0, \infty)$,

$$\boldsymbol{t}(u) + \sum_{\ell=1}^{L} \Delta \boldsymbol{t}(b^{-\ell} u) = \boldsymbol{t}(b^{-L} u) \to \boldsymbol{t}(0) = 1 \text{ as } L \to \infty.$$

$\square$

We collect some more properties of the cut-off functions that will be used later on.

Remark 2 Let $b > 1$.

(i) Suppose that $\tau > 2d + 4$ and $\boldsymbol{t} \in \mathcal{G}(b)$. By [10, Theorem 3.4 and (2.7)], $\boldsymbol{t}(\delta\sqrt{\mathcal{D}})$ for $0 < \delta \leq 1$ is an integral operator and there exist $c_\tau > 0$ and $c_\tau' > 0$ such that, for every $0 < \delta \leq 1$, the associated integral kernel (compare (17)) satisfies

$$\left| \boldsymbol{T}_\delta(x, y) \right| \leq c_\tau \mu\left(B(x, \delta)\right)^{-1} \left(1 + \frac{\rho(x, y)}{\delta}\right)^{d/2 - \tau}; \tag{24}$$

and, if $\rho(x, x') \leq \delta$, then with α_H from (14) it holds

$$\left| \boldsymbol{T}_\delta(x,y) - \boldsymbol{T}_\delta(x',y) \right| \leq c'_\tau \left(\frac{\rho(x,x')}{\delta} \right)^{\alpha_H} \mu \left(B(x,\delta) \right)^{-1} \left(1 + \frac{\rho(x,y)}{\delta} \right)^{d/2 - \tau}. \tag{25}$$

Let $1 \leq p \leq \infty$. Then, by [10, Corollary 3.6], there is $C > 0$ such that for all $0 < \delta \leq 1$ and all $f \in L^p(M; d\mu)$

$$\|\boldsymbol{t}(\delta\sqrt{\mathcal{D}})f\|_{L^p(M;d\mu)} \leq C\|f\|_{L^p(M;d\mu)}. \tag{26}$$

(ii) Littlewood-Paley-type decomposition: Suppose further that $(\boldsymbol{t}, \boldsymbol{p}) \in \mathcal{G}(b) \times \mathcal{G}(b)$ is a partition of unity. Then, by [10, Corollary 3.9] it holds for all $f \in L^p(M; d\mu)$ with $1 \leq p < \infty$

$$f = \boldsymbol{t}(\sqrt{\mathcal{D}})f + \sum_{\ell=1}^{\infty} \boldsymbol{p}(b^{-\ell}\sqrt{\mathcal{D}})f. \tag{27}$$

(iii) Suppose that $\boldsymbol{f}, \boldsymbol{g} \in C_c^\infty([0,R])$ for some $R > 0$ and $0 \leq \boldsymbol{f} \leq \boldsymbol{g}$. Then, by [10, (3.43)], it holds for the associated integral kernels with $\delta = 1$ that

$$0 \leq \boldsymbol{F}(x,x) \leq \boldsymbol{G}(x,x) \qquad \text{for every} \quad x \in M.$$

(iv) By [10, Lemma 3.19 (b)], there exists $\tilde{b} > 1$ such that for all $r \geq \max\{1, 3/\operatorname{diam} M\}$ and all $x \in M$

$$C_1 \mu \left(B(x, r^{-1}) \right)^{-1} \leq \boldsymbol{1}_{[r, \tilde{b}r]}(x,x) \leq C_2 \mu \left(B(x, r^{-1}) \right)^{-1}, \tag{28}$$

where $\boldsymbol{1}_{[r, \tilde{b}r]}$ denotes the integral kernel associated to the operator $\boldsymbol{1}_{[r, \tilde{b}r]}(\sqrt{\mathcal{D}})$ with the characteristic function $\boldsymbol{1}$. The constants C_1 and C_2 depend only on the parameters of the space.

The parameter $\tilde{b}$ from Remark 2 (iv) will play a crucial role in our analysis, see also [27]. Therefore, we present a lower bound on $\tilde{b}$ in the Euclidean case $M = \mathbb{R}^d$ in the appendix. In particular, it turns out that $\tilde{b}$ depends on the space dimension d and $\tilde{b}(d) \to 1$ as $d \to \infty$.

2.4 Spectral spaces

Spectral spaces are usually defined as invariant sets under integral operators. Precisely, for $1 \leq p \leq \infty$ and a compact set $K \subset [0, \infty)$, we define the associated spectral space Σ_K^p, see [10, Definition 3.10],

$$\Sigma_K^p := \left\{ g \in L^p(M; d\mu) \ : \ \boldsymbol{t}(\sqrt{\mathcal{D}})g = g \text{ for all } \boldsymbol{t} \in C^\infty([0, \infty)), \boldsymbol{t} \equiv 1 \text{ on } K \right\}. \tag{29}$$

We will need the following result proven in [10, Proposition 3.12].

Proposition 1 *Let $1 \leq p \leq q \leq \infty$. If $R \geq 1$, then $\Sigma_{[0,R]}^p \subset \Sigma_{[0,R]}^q$ and there is a constant $C > 0$ such that*

$$\|g\|_{L^q(M;d\mu)} \leq CR^{d(1/p-1/q)}\|g\|_{L^p(M;d\mu)} \qquad \text{for all } g \in \Sigma_{[0,R]}^p \tag{30}$$

and

$$|g(x) - g(y)| \leq C(R\rho(x,y))^{\alpha_H}\|g\|_{L^\infty(M;d\mu)} \qquad \text{for all } g \in \Sigma_{[0,R]}^\infty \text{ and all } x, y \in M \tag{31}$$

with α_H from (14).

Remark 3 Note that Proposition 1 implies in particular that functions in $\Sigma_{[0,R]}^p$ for $R \geq 1$ and $1 \leq p \leq \infty$ have continuous representatives.

3 Generalized Besov spaces as reproducing kernel Hilbert spaces

In this section, we recall the notion of generalized Besov spaces. Our main result is Proposition 4, in which we explicitly identify the reproducing kernels for such spaces and derive a multiscale decomposition of it. To introduce Besov-type reproducing kernel Hilbert spaces, we define cut-off functions controlling the spectral decay in the kernel expansion. Here, we restrict ourselves to the case $\delta_\ell := b^{-\ell}$ with a fixed $b > 1$. We essentially follow the notion of Besov spaces based on spectral decompositions as introduced in [10, Section 6] and [27, Section 6], which in turn build on [43, 53, 54].

Now, we briefly recall the setting (see [10, Definition 6.1]) in the case $p, q \geq 1$ and as in [27], allow for different normalizations of the support ($b = 2$ in the notation of [10]):

Definition 2 Let $\sigma > 0$, $1 \leq p \leq \infty$, and $1 \leq q \leq \infty$. Suppose that $\phi \in \mathcal{A}(b, c_1, c_2)$, and $\psi \in \mathcal{E}(b, c_1)$ for some $0 < c_1 \leq c_2 < 1$. The Besov space $\mathcal{B}_{p,q}^\sigma(M; \mathcal{D})$ is defined as

$$\mathcal{B}_{p,q}^\sigma(M; \mathcal{D}) := \left\{ f \in L^p(M; d\mu) \ : \ \|f\|_{\mathcal{B}_{p,q}^\sigma(M;\mathcal{D})} < \infty \right\}$$

equipped with the norm

$$\|f\|_{\mathcal{B}_{p,q}^\sigma(M;\mathcal{D})} := \left(\left\| \phi(\sqrt{\mathcal{D}})f \right\|_{L^p(M;d\mu)}^q + \sum_{\ell=1}^\infty \left(b^{\sigma\ell} \|\psi(b^{-\ell}\sqrt{\mathcal{D}})f\|_{L^p(M;d\mu)} \right)^q \right)^{1/q}.$$

For $\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})$ we denote the associated inner product by

$$(f, g)_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} := \left(\phi(\sqrt{\mathcal{D}})f, \phi(\sqrt{\mathcal{D}})g \right)_{L^2(M;d\mu)} + \sum_{\ell=1}^\infty b^{2\sigma\ell} \left(\psi(b^{-\ell}\sqrt{\mathcal{D}})f, \psi(b^{-\ell}\sqrt{\mathcal{D}})g \right)_{L^2(M;d\mu)}.$$

$$(32)$$

The space $\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})$ and its topology do not depend on the specific choice of the functions ψ and ϕ (see [10, Section 6]). As in [27, Proof of Proposition 6.5] we will use such cut-off functions ϕ and ψ for which additionally (ϕ^2, ψ^2) build a partition of unity and $\phi(u) = 1$ for $u \in [0, 1]$. We now build on [10] to show a relation between the Besov spaces and Bessel potential spaces (see Theorem 1).

Lemma 2 Let $\sigma > 0$. Fix ϕ and ψ that satisfy the assumptions of Definition 2, are such that (ϕ^2, ψ^2) form a partition of unity, and satisfy $\phi(u) = 1$ for $u \in [0, 1]$. We set

$$\tilde{\phi} := b^\sigma \phi, \quad and \quad \tilde{\psi}(u) := u^\sigma \psi(u).$$

$$(33)$$

Then $\tilde{\phi}$ and $\tilde{\psi}$ satisfy the conditions of Definition 2 and it holds

$$\frac{1}{2}(u^{2\sigma} + 1) \leq \tilde{\phi}^2(u) + \sum_{\ell=1}^\infty b^{2\sigma\ell} \tilde{\psi}^2(b^{-\ell}u) = u^{2\sigma} + (b^{2\sigma} - u^{2\sigma})\phi^2(u) \leq b^{2\sigma}(u^{2\sigma} + 1)$$

$$(34)$$

for all $u \in [0, \infty)$.

Proof We check the conditions on $\tilde{\phi}$ first. We have $\operatorname{supp} \tilde{\phi} = \operatorname{supp} \phi \subset [0, b]$, and $\frac{d^\nu}{du^\nu} \tilde{\phi}(u) \Big|_{u=0} = b^\sigma \frac{d^\nu}{du^\nu} \phi(u) \Big|_{u=0} = 0$. Furthermore, if $u \in [0, b^{3/4}]$, then $\left| \tilde{\phi}(u) \right| = b^\sigma \phi(u) \geq b^\sigma c > 0$. Similarly for

$\tilde{\psi}$: We have supp $\tilde{\psi} = \text{supp}\,\psi \subset [b^{-1}, b]$; and if $u \in [b^{-3/4}, b^{3/4}]$, then $|\tilde{\psi}(u)| = u^\sigma \psi(u) \geq u^\sigma c \geq b^{-3\sigma/4} c > 0$. To show (34), note first that, since (ϕ^2, ψ^2) is a partition of unity, there holds

$$\sum_{\ell=1}^\infty b^{2\sigma\ell} \tilde{\psi}^2(b^{-\ell} u) = \sum_{\ell=1}^\infty b^{2\sigma\ell} \left((b^{-\ell}u)^\sigma \psi(b^{-\ell}u)\right)^2 = \sum_{\ell=1}^\infty u^{2\sigma} \psi^2(b^{-\ell}u) = u^{2\sigma}(1 - \phi^2(u)),$$

which implies

$$\tilde{\phi}^2(u) + \sum_{\ell=1}^\infty b^{2\sigma\ell} \tilde{\psi}^2(b^{-\ell}u) = u^{2\sigma} + (b^{2\sigma} - u^{2\sigma})\phi^2(u). \tag{35}$$

To prove the upper bound in (34), observe that, for $u \in [0, \infty)$, we have

$$u^{2\sigma} + (b^{2\sigma} - u^{2\sigma})\phi^2(u) = u^{2\sigma} + b^{2\sigma}\phi^2(u) - u^{2\sigma}\phi^2(u) \leq u^{2\sigma} + b^{2\sigma} \leq b^{2\sigma}(1 + u^{2\sigma}).$$

To show the lower bound, note that, for $u \leq 1$, we have

$$u^{2\sigma} + (b^{2\sigma} - u^{2\sigma})\phi^2(u) = b^{2\sigma} \geq \frac{1}{2}(u^{2\sigma} + 1),$$

while, for $1 \leq u$, we have

$$u^{2\sigma} + (b^{2\sigma} - u^{2\sigma})\phi^2(u) \geq u^{2\sigma} \geq \frac{1}{2}(u^{2\sigma} + 1).$$

This concludes the proof. □

Lemma 2 immediately implies that the Besov spaces $\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})$ are (norm-)equivalent to Bessel potential spaces based on $\mathcal{D}$. Moreover, we have the following result for the Hilbert space case $p = q = 2$.

Theorem 1 *Let $b > 1$. Then there are constants c, $C > 0$ such that*

$$c\|(\mathcal{D}^\sigma + \text{Id})f\|_{L^2(M;d\mu)} \leq \|f\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} \leq C\|(\mathcal{D}^\sigma + \text{Id})f\|_{L^2(M;d\mu)} \qquad \text{for all } f \in \mathcal{B}_{2,2}^\sigma(M;\mathcal{D}).$$

Proof Consider more generally smooth cut-off functions $\boldsymbol{s}$, $\boldsymbol{t} \in \mathcal{G}(b)$, and suppose that $\boldsymbol{s} \leq \boldsymbol{t}$ holds pointwise. Then we have $\|\boldsymbol{s}(\sqrt{\mathcal{D}})f\|_{L^2(M;d\mu)} \leq \|\boldsymbol{t}(\sqrt{\mathcal{D}})f\|_{L^2(M;d\mu)}$ for all $f \in L^2(M; d\mu)$ since

$$\|\boldsymbol{s}(\sqrt{\mathcal{D}})f\|_{L^2(M;d\mu)}^2 = \int_0^\infty |\boldsymbol{s}(\lambda)|^2 d\langle \mathbb{E}_\lambda f, f\rangle$$

$$\leq \int_0^\infty |\boldsymbol{t}(\lambda)|^2 d\langle \mathbb{E}_\lambda f, f\rangle = \|\boldsymbol{t}(\sqrt{\mathcal{D}})f\|_{L^2(M;d\mu)}^2.$$

Hence the assertion follows by (34). □

3.1 The reproducing kernel and its multiscale decomposition

We fix some cut-off functions that will be used throughout the remainder of this paper. First, we keep ϕ and ψ fixed that satisfy the assumptions of Lemma 2. (This corresponds to the choice of the cut-off functions used in the proof of [27, Prop. 6.5].)

Notation 3 *We set*

$$\boldsymbol{k}^{(\sigma)} : [0, \infty) \to [0, \infty), \qquad \boldsymbol{k}^{(\sigma)}(u) := (u^{2\sigma} + (b^{2\sigma} - u^{2\sigma})\boldsymbol{\phi}^2(u))^{-1}. \tag{36}$$

Furthermore, for $\ell \geq 1$, we set

$$\boldsymbol{g}^{(\ell;\sigma)} : [0, \infty) \to [0, \infty), \qquad \boldsymbol{g}^{(\ell;\sigma)}(u) := \left(u^{2\sigma}(1 - \boldsymbol{\phi}^2(b^\ell u)) + b^{2\sigma(1-\ell)}\boldsymbol{\phi}^2(b^\ell u)\right)^{-1};$$

and (recall Notation 2) we define

$$\boldsymbol{w}_\ell^{(\sigma)}(u) := \begin{cases} \left(\boldsymbol{k}^{(\sigma)} \cdot \boldsymbol{\phi}\right)(u) & \text{if } \ell = 0, \\ \left(\boldsymbol{g}^{(\ell;\sigma)} \cdot \Delta\boldsymbol{\phi}\right)(u) & \text{if } \ell \geq 1, \end{cases} \tag{37}$$

and

$$\boldsymbol{s}^{(\ell;\sigma)}(u) := \boldsymbol{w}_\ell^{(\sigma)}(b^{-\ell}u). \tag{38}$$

Lemma 3 *We have $\boldsymbol{w}_\ell^{(\sigma)} \equiv \boldsymbol{w}_2^{(\sigma)}$ for all $\ell \geq 2$. Furthermore, there is a $c > 0$ such that for all $\ell \geq 1$, we have the lower bound*

$$\boldsymbol{s}^{(\ell;\sigma)}(u) \geq c \quad \text{for all } u \in \left[b^\ell, b^{\ell+\frac{3}{4}}\right],$$

and in particular $\boldsymbol{s}^{(\ell;\sigma)} \geq c\mathbf{1}_{[b^\ell, b^{\ell+\frac{3}{4}}]}$. Moreover,

$$\sum_{\ell=0}^\infty b^{-2\sigma\ell}\boldsymbol{s}^{(\ell;\sigma)} = \boldsymbol{k}^{(\sigma)}. \tag{39}$$

Proof We note that $b^\ell u \in \operatorname{supp}\boldsymbol{\phi} \subset [0, b]$ implies that $u \in [0, b^{-\ell+1}]$. Recall also that for $\ell \geq 1$, $\operatorname{supp}\boldsymbol{w}^{(\ell;\sigma)} \subset \operatorname{supp}\Delta\boldsymbol{\phi} \subset [b^{-1}, b]$. Therefore, if $\ell \geq 2$ and $u \in \operatorname{supp}\boldsymbol{w}^{(\ell;\sigma)}$, then $b^\ell u \geq b^2 b^{-1} = b$, and thus $\boldsymbol{\phi}(b^\ell u) = 0$. Hence, for all $\ell \geq 2$,

$$\boldsymbol{w}_\ell^{(\sigma)}(u) = \left(u^{2\sigma}(1 - \boldsymbol{\phi}^2(b^\ell u)) + b^{2\sigma(1-\ell)}\boldsymbol{\phi}^2(b^\ell u)\right)^{-1}\Delta\boldsymbol{\phi}(u) = u^{-2\sigma}\Delta\boldsymbol{\phi}(u).$$

This shows in particular that $\boldsymbol{w}_\ell^{(\sigma)} \equiv \boldsymbol{w}_2^{(\sigma)}$ for all $\ell \geq 2$.

It remains to consider $\boldsymbol{s}^{(\ell;\sigma)}$. We note that by definition there holds

$$b^{-2\sigma\ell}\boldsymbol{s}^{(\ell;\sigma)}(u) = b^{-2\sigma\ell}\boldsymbol{w}_\ell^{(\sigma)}\left(b^{-\ell}\right) = b^{-2\sigma\ell}\boldsymbol{g}^{(\ell;\sigma)}(b^{-\ell}u)\Delta_\ell\boldsymbol{\phi}(u)$$

$$= \left(u^{2\sigma}(1 - \boldsymbol{\phi}^2(u)) + b^{2\sigma}\boldsymbol{\phi}^2(u)\right)^{-1}\Delta_\ell\boldsymbol{\phi}(u) = \Delta_\ell\boldsymbol{\phi}(u)\boldsymbol{k}^{(\sigma)}(u). \tag{40}$$

Consider $u \in [b^\ell, b^{\ell+\frac{3}{4}}]$. First, we get with (34)

$$\left(u^{2\sigma}(1 - \boldsymbol{\phi}^2(u)) + b^{2\sigma}\boldsymbol{\phi}^2(u)\right)^{-1} \geq \left(b^{2\sigma}(u^{2\sigma} + 1)\right)^{-1} \geq b^{-2\sigma}(2u^{2\sigma})^{-1} \geq \frac{1}{2}b^{-2\sigma}b^{-(\ell+\frac{3}{4})2\sigma}$$

$$\geq \frac{1}{2}b^{-\frac{7}{2}\sigma}b^{-2\sigma\ell}. \tag{41}$$

Second, we have

$$\Delta_\ell\boldsymbol{\phi}(u) = \boldsymbol{\phi}(b^{-\ell}u) - \boldsymbol{\phi}(b^{-\ell+1}u) \geq c \quad \text{if } u \in [b^\ell, b^{\ell+\frac{3}{4}}]. \tag{42}$$

Putting (40), (41) and (42) together, we obtain

$$\boldsymbol{s}^{(\ell;\sigma)}(u) \geq \frac{c}{2}b^{-\frac{7}{2}\sigma} \quad \text{for all } u \in \left[b^\ell, b^{\ell+\frac{3}{4}}\right] \quad \text{and all } \ell \geq 1.$$

Furthermore, note that $\phi(u) - \phi(bu) \geq 0$ for all $u \in \mathbb{R}^+$ since $\phi(u) = 1 \geq \phi(bu)$ if $u \leq 1$, and $\phi(bu) = 0$ if $u \geq 1$, and thus, $\boldsymbol{s}^{(\ell;\sigma)} \geq c\mathbf{1}_{[b^\ell, b^{\ell+\frac{3}{4}}]}$. Finally, to show (39), we observe that by Lemma 1 since $\phi(0) = 1$, we have

$$
\sum_{\ell=0}^{\infty} b^{-2\sigma\ell} \boldsymbol{s}^{(\ell;\sigma)}(u) = \boldsymbol{k}^{(\sigma)}\phi(u) + \sum_{\ell=1}^{\infty} b^{-2\sigma\ell}\left((b^{-\ell}u)^{2\sigma}(1-\phi^2(u)) + b^{2\sigma(1-\ell)}\phi^2(u)\right)^{-1}\Delta_\ell\phi(u)
$$

$$
= \boldsymbol{k}^{(\sigma)}(u)\sum_{\ell=0}^{\infty}\Delta_\ell\phi(u) = \boldsymbol{k}^{(\sigma)}(u).
$$

$\square$

Next, we aim to show that the Besov spaces $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$ are reproducing kernel Hilbert spaces if the index σ is sufficiently large. We recall that $K^{(\sigma)} : M \times M \to \mathbb{R}$ is called reproducing kernel for $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$ if $K^{(\sigma)}(x,\cdot) \in \mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$ for all $x \in M$ and $f(x) = \left(f, K^{(\sigma)}(x,\cdot)\right)_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}$ for all $x \in M$ and all $f \in \mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$ holds. A necessary condition is that point evaluations are continuous linear functionals on $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$. Recall that, for classical Sobolev spaces $W^{2,\sigma}(\Omega)$ on Lipschitz domains $\Omega \subset \mathbb{R}^d$, the Sobolev embedding theorem guarantees continuous point evaluations if $\sigma > \frac{d}{2}$, see [1]. In our abstract setting, we do not have a natural meaning of dimensionality, but, as already mentioned, the number d in (8) plays the role of space dimension. Moreover the parameter $\sigma > 0$ resembles the smoothness parameter of classical Sobolev spaces. Roughly speaking, the larger σ, the smoother the respective kernel, and thus the smaller the reproducing kernel Hilbert space. Here *small* is meant in terms of embeddings, see [10].

In the following proposition, we show that the reproducing kernel possesses a multilevel decomposition which in the sequel will be used to approximate the kernel. This is in the spirit of [7,8,18, 35,41,42]. In the following, we shall use $\boldsymbol{s}^{(\ell;\sigma)}$ and $\boldsymbol{k}^{(\sigma)}$ introduced in Notation 3. Recall that we denote the associated integral kernels by $\boldsymbol{S}^{(\ell;\sigma)}$ and $\boldsymbol{K}^{(\sigma)}$, respectively.

Proposition 2 *Let $\sigma > d/2$ and let $b > 1$. Then, the Besov space $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$ with cut-off functions $\tilde{\phi}$ and $\tilde{\psi}$ from Lemma 2 is a reproducing kernel Hilbert space with kernel*

$$
K^{(\sigma)}(x,y) := \sum_{\ell=0}^{\infty} b^{-2\sigma\ell}\boldsymbol{S}^{(\ell;\sigma)}(x,y) = \boldsymbol{K}^{(\sigma)}(x,y). \tag{43}
$$

Proof By [10, Proposition 6.7], we have the same embedding properties as for usual Besov spaces, namely $\mathcal{B}_{p,q}^{\sigma}(M;\mathcal{D}) \subset \mathcal{B}_{p_1,q_1}^{\sigma_1}(M;\mathcal{D})$ if $1 \leq p \leq p_1 \leq \infty$, $0 < q \leq q_1 \leq \infty$, $0 < \sigma \leq \sigma_1 < \infty$ and $\frac{\sigma}{d} - \frac{1}{p} = \frac{\sigma_1}{d} - \frac{1}{p_1}$. In particular, for $p = q = 2$, $p_1 = q_1 = \infty$ and $\sigma_1 = \sigma - \frac{d}{2} > 0$, we obtain $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D}) \subset \mathcal{B}_{\infty,\infty}^{\sigma_1}(M;\mathcal{D})$. Furthermore, by [10, Proposition 6.4(b)], we have $\mathcal{B}_{\infty,\infty}^{\sigma_1}(M;\mathcal{D}) \subset \mathrm{Lip}\,\sigma_1$ for $0 < \sigma_1 < \alpha_H$ with α_H from (14), where, for $L > 0$, $\mathrm{Lip}\,L$ denotes the space of functions f for which $\|f\|_{L^\infty(M;d\mu)} + \sup_{x \neq y}\frac{|f(x)-f(y)|}{\rho^L(x,y)}$ is finite. Since we also have the trivial embedding $\mathcal{B}_{p,q}^{\sigma}(M;\mathcal{D}) \subset \mathcal{B}_{p,q}^{\tau}(M;\mathcal{D})$ if $\tau \leq \sigma$, we deduce that $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D}) \subset \mathrm{Lip}\,\sigma_1$ for some $0 < \sigma_1 < \alpha_H$. Hence point evaluations are well-defined on $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$. We show the reproducing property

$$
(K^{(\sigma)}(x,\cdot), f)_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})} = f(x) \qquad \text{for all} \quad x \in M \text{ and all } f \in \mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D}). \tag{44}
$$

For that, we compute for $f \in \mathcal{B}_{2,2}^\sigma(M;\mathcal{D})$ and $x \in M$, using the notation introduced in (17)

$$\left(K^{(\sigma)}(x,\cdot), f\right)_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})}$$

$$= \left(\tilde{\phi}(\sqrt{\mathcal{D}})K^{(\sigma)}(x,\cdot), \tilde{\phi}(\sqrt{\mathcal{D}})f,\right)_{L^2(M;d\mu)} + \sum_{\ell=1}^{\infty} b^{2\sigma\ell}\left(\tilde{\psi}(b^{-\ell}\sqrt{\mathcal{D}})K^{(\sigma)}(x,\cdot), \tilde{\psi}(b^{-\ell}\sqrt{\mathcal{D}})f,\right)_{L^2(M;d\mu)}$$

$$= \left(\int_M \tilde{\boldsymbol{\Phi}}(y,\cdot)K^{(\sigma)}(x,y)\,d\mu(y), \tilde{\phi}(\sqrt{\mathcal{D}})f\right)_{L^2(M;d\mu)} +$$

$$+ \sum_{\ell=1}^{\infty} b^{2\sigma\ell}\left(\int_M \tilde{\boldsymbol{\Psi}}_{b^{-\ell}}(y,\cdot)K^{(\sigma)}(x,y)\,d\mu(y), \tilde{\psi}(b^{-\ell}\sqrt{\mathcal{D}})f\right)_{L^2(M;d\mu)}$$

$$= \int_M K^{(\sigma)}(x,y)\int_M \tilde{\phi}(\sqrt{\mathcal{D}})f(z)\tilde{\boldsymbol{\Phi}}(y,z)\,d\mu(z)\,d\mu(y)+$$

$$+ \sum_{\ell=1}^{\infty} b^{2\sigma\ell}\int_M K^{(\sigma)}(x,y)\int_M \tilde{\psi}(b^{-\ell}\sqrt{\mathcal{D}})f(z)\tilde{\boldsymbol{\Psi}}_{b^{-\ell}}(y,z)\,d\mu(z)\,d\mu(y)$$

$$= \left(K^{(\sigma)}(x,\cdot), \left(\tilde{\phi}(\sqrt{\mathcal{D}})\right)^2 f\right)_{L^2(M;d\mu)} + \left(K^{(\sigma)}(x,\cdot), \sum_{\ell=1}^{\infty} b^{2\sigma\ell}\left(\tilde{\psi}(b^{-\ell}\sqrt{\mathcal{D}})\right)^2 f\right)_{L^2(M;d\mu)}. \quad (45)$$

Next, recall at this point that $\boldsymbol{S}^{(\ell;\sigma)}$ is as usual the integral kernel associated to the operator $\boldsymbol{s}^{(\ell;\sigma)}(\sqrt{\mathcal{D}}) = \boldsymbol{g}^{(\ell;\sigma)}(b^{-\ell}\sqrt{\mathcal{D}})\Delta_\ell\phi(\sqrt{\mathcal{D}})$. If we denote the integral kernels associated to $\boldsymbol{g}^{(\ell;\sigma)}(b^{-\ell}\sqrt{\mathcal{D}})$ and $\Delta_\ell\phi(\sqrt{\mathcal{D}})$ by $\boldsymbol{G}_{b^{-\ell}}^{(\ell;\sigma)}$ and $\Delta_\ell\boldsymbol{\Phi}$, respectively, we obtain

$$\int_M \boldsymbol{S}^{(\ell;\sigma)}(x,y)f(y)\,d\mu(y) = \boldsymbol{s}^{(\ell;\sigma)}(\sqrt{\mathcal{D}})f(x) = \boldsymbol{g}^{(\ell;\sigma)}(b^{-\ell}\sqrt{\mathcal{D}})\Delta_\ell\phi(\sqrt{\mathcal{D}})f(x)$$

$$= \boldsymbol{g}^{(\ell;\sigma)}(b^{-\ell}\sqrt{\mathcal{D}})\int_M \Delta_\ell\boldsymbol{\Phi}(x,y)f(y)\,d\mu(y)$$

$$= \int_M \boldsymbol{G}_{b^{-\ell}}^{(\ell;\sigma)}(x,z)\int_M \Delta_\ell\boldsymbol{\Phi}(z,y)f(y)\,d\mu(y)\,d\mu(z),$$

that is,

$$\boldsymbol{S}^{(\ell;\sigma)}(x,y) = \int_M \boldsymbol{G}_{b^{-\ell}}^{(\ell;\sigma)}(x,z)\Delta_\ell\boldsymbol{\Phi}(z,y)\,d\mu(z). \quad (46)$$

Inserting this representation into (45), we obtain using (35)

$$\left(K^{(\sigma)}(x,\cdot),\ f\right)_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} = \left(\sum_{\ell=0}^{\infty} b^{-2\ell\sigma}\boldsymbol{S}^{(\sigma)}(x,y),\ (\tilde{\phi}(\sqrt{\mathcal{D}}))^2 f + \sum_{j=1}^{\infty} b^{2\sigma j}(\tilde{\psi}(b^{-j}\sqrt{\mathcal{D}}))^2 f\right)_{L^2(M;d\mu)}$$

$$= \left(\sum_{\ell=0}^{\infty} b^{-2\sigma\ell}\int_M \boldsymbol{G}_{b^{-\ell}}^{(\ell;\sigma)}(x,z)\Delta_\ell\boldsymbol{\Phi}(z,\cdot)\,d\mu(z),\ (\boldsymbol{k}^{(\sigma)}(\sqrt{\mathcal{D}}))^{-1}f\right)_{L^2(M;d\mu)}$$

$$= \left(\sum_{\ell=0}^{\infty} b^{-2\sigma\ell}\boldsymbol{g}^{(\ell;\sigma)}(b^{-\ell}\sqrt{\mathcal{D}})\Delta_\ell\boldsymbol{\Phi}(x,\cdot),\ (\boldsymbol{k}^{(\sigma)}(\sqrt{\mathcal{D}}))^{-1}f\right)_{L^2(M;d\mu)}$$

$$= \left(\sum_{\ell=0}^{\infty} b^{-2\sigma\ell}(b^{-2\sigma\ell}\mathcal{D}^\sigma(1-\phi^2(\sqrt{\mathcal{D}})) + b^{-2\ell\sigma+2\sigma}\phi^2(\sqrt{\mathcal{D}}))^{-1}\Delta_\ell\boldsymbol{\Phi}(x,\cdot), (\boldsymbol{k}^{(\sigma)}(\sqrt{\mathcal{D}}))^{-1}f\right)_{L^2(M;d\mu)}$$

$$= \left(\boldsymbol{k}^{(\sigma)}(\sqrt{\mathcal{D}})\sum_{\ell=0}^{\infty}\Delta_\ell\boldsymbol{\Phi}(x,\cdot), (\boldsymbol{k}^{(\sigma)}(\sqrt{\mathcal{D}}))^{-1}f\right)_{L^2(M;d\mu)} = \left(\sum_{\ell=0}^{\infty}\Delta_\ell\boldsymbol{\Phi}(x,\cdot),\ f\right)_{L^2(M;d\mu)} = f(x), \quad (47)$$

where the last step follows from Lemma 1 and Remark 2 (ii), see [10, Cor. 3.9]. The second equality in (43) follows from (39). □

So far, we have identified the reproducing kernel $K^{(\sigma)}(x,y)$ for the Hilbert space $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$. In the remaining sections, we shall focus on kernel-based approximation schemes based on these reproducing kernels.

4 Sampling inequalities

In this section, we prove sampling inequalities. They provide a systemic tool for the deterministic error analysis of stable and consistent processes for the reconstruction of functions $f \in \mathcal{B}_{p,q}^{\sigma}(M;\mathcal{D})$ from given point values at discrete locations $X \subset M$, as outlined in the introduction. We will present two approaches. First, we consider maximal δ-nets as sampling points which allows in special cases also for estimates if $\mu(M) = \infty$ (see Corollary 1). Then, we specialize to the case of finite volume and derive estimates in the $L^{\infty}(M;d\mu)$-norm using a norming set approach. The latter technique allows for more general scattered sampling points. Note that we derive sampling inequalities for general spaces $\mathcal{B}_{p,q}^{\sigma}(M;\mathcal{D})$ which can be continuously embedded into the space of bounded continuous functions, while in the applications we focus on kernel-based methods, and hence on Hilbert spaces $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$.

4.1 Sampling inequalities based on maximal δ-nets

To describe appropriate sampling points, we use the notion of maximal nets (see [10, Definition 2.4]). Let $\delta > 0$. A discrete (possibly infinite) set $X \subset M$ is called δ-net on M if $q_X \geq \delta$, where

$$q_X := \inf_{x_n \neq x_m \in X} \rho(x_n, x_m) \tag{48}$$

denotes the separation distance of X. A δ-net X is called a maximal δ-net on M if it cannot be enlarged, i.e., if for every $z \in M \setminus X$ there is an $x_n \in X$ such that $\rho(x_n, z) \leq \delta$. Note, that in this case

$$h_{X,M} := \sup_{z \in M} \inf_{x_n \in X} \rho(z, x_n) \leq \delta, \tag{49}$$

where $h_{X,M}$ denotes the fill distance of X in M. We use the following result from [10, Proposition 2.5]. Under the assumptions on (M, μ, ρ) outlined in Section 2, for every $\delta > 0$, there exists a maximal δ-net X_{δ}, which consists of at most countably many points. Furthermore, there exists a family of pairwise disjoint, measurable sets A_n, $x_n \in X_{\delta}$, such that $M = \bigcup_{x_n \in X_{\delta}} A_n$, and $B(x_n, \frac{\delta}{2}) \subset A_n \subset B(x_n, \delta)$.

It turns out that maximal nets are norming sets for the spectral spaces (29). Precisely, we have the following result from [10, Theorem 4.2].

Theorem 2 *Let $0 < \gamma < 1$ be such that, with the constants α and C_1 from (14), there holds*

$$C_1 (2d+1) 2^{2d+1} \gamma^{\alpha} = \frac{1}{2}. \tag{50}$$

For $R \geq b$ set $\delta := \frac{\gamma}{R}$ and consider a maximal δ-net $X_{\delta} \subset M$ with associated disjoint cover $\{A_n\}_{x_n \in X_{\delta}}$. Then, for every $f_R \in \Sigma_{[0,R]}^{p}$ with $1 \leq p < \infty$, we have

$$\frac{1}{2} \|f_R\|_{L^p(M;d\mu)} \leq \left(\sum_{x_n \in X_{\delta}} \mu(A_n) |f_R(x_n)|^p \right)^{1/p} \leq 2 \|f_R\|_{L^p(M;d\mu)} \tag{51}$$

and for $p = \infty$, we get

$$\frac{1}{2} \|f_R\|_{L^\infty(M;d\mu)} \leq \sup_{x_n \in X_\delta} |f_R(x_n)| \leq 2 \|f_R\|_{L^\infty(M;d\mu)}. \tag{52}$$

Our main contribution in this section is the following sampling inequality.

Theorem 3 *Let $\mu(M) < \infty$ and $0 < \gamma < 1$ be such that condition (50) holds. Let $r, p \in [1, \infty]$, $q \in (0, \infty]$, and $\sigma > \max\{d(1/p - 1/r)_+, d/r\}$, where $(x)_+ := \max\{x, 0\}$ for $x \in \mathbb{R}$. Then there is a constant $c > 0$ with the following property: For every $\delta \in (0, \gamma/b]$ and for every maximal δ-net X_δ with associated disjoint partition A_n of M, $x_n \in X_\delta$, we have for all $f \in \mathcal{B}_{r,q}^\sigma(M; \mathcal{D})$*

$$\|f\|_{L^p(M;d\mu)} \leq c \left(\delta^{\sigma - d(1/r - 1/p)_+} + \delta^{\sigma - d/r} \right) \|f\|_{\mathcal{B}_{p,q}^\sigma(M;\mathcal{D})} + 2 \left(\sum_{x_n \in X_\delta} \mu(A_n) |f(x_n)|^p \right)^{1/p}, \tag{53}$$

where we set $\left(\sum_{x_n \in X_\delta} \mu(A_n) |f(x_n)|^p \right)^{1/p} := \sup_{x_n \in X_\delta} |f(x_n)|$ if $p = \infty$.

Proof We note that, by [10, Proposition 6.7], $\mathcal{B}_{p,q}^\sigma(M; \mathcal{D}) \subset \mathcal{B}_{\infty,\infty}^\tau(M; \mathcal{D})$ with $\tau = \sigma - d/r > 0$, and thus, by [10, Proposition 6.4(b)], every element of $\mathcal{B}_{r,q}^\sigma(M; \mathcal{D})$ has a continuous representative, cf. also the proof of Theorem 2. Therefore, the last term in (53) is well-defined. We proceed similarly to the proof of classical sampling inequalities (see, e.g., [18,40,59]). Set $R := \gamma/\delta \geq b$. Let $f \in \mathcal{B}_{r,q}^\sigma(M; \mathcal{D})$ and $f_R \in \Sigma_{[0,R]}^p$ be arbitrary. Then, for every maximal δ-net X_δ with associated partition A_n, $x_n \in X_\delta$, we have for $p < \infty$ by (51) and with $\sum_{x_n \in X_\delta} \mu(A_n) = \mu(M)$,

$$\|f\|_{L^p(M;d\mu)} \leq \|f - f_R\|_{L^p(M;d\mu)} + \|f_R\|_{L^p(M;d\mu)}$$

$$\leq \|f - f_R\|_{L^p(M;d\mu)} + 2 \left(\sum_{x_n \in X_\delta} \mu(A_n) |f_R(x_n)|^p \right)^{1/p}$$

$$\leq \|f - f_R\|_{L^p(M;d\mu)} + 2 \left(\sum_{x_n \in X_\delta} \mu(A_n) |f(x_n) - f_R(x_n)|^p \right)^{1/p} + 2 \left(\sum_{x_n \in X_\delta} \mu(A_n) |f(x_n)|^p \right)^{1/p}$$

$$\leq \|f - f_R\|_{L^p(M;d\mu)} + 2\mu(M)^{1/p} \|f - f_R\|_{L^\infty(M;d\mu)} + 2 \left(\sum_{x_n \in X_\delta} \mu(A_n) |f(x_n)|^p \right)^{1/p}, \tag{54}$$

and similarly, using (52),

$$\|f\|_{L^\infty(M;d\mu)} \leq 3\|f - f_R\|_{L^\infty(M;d\mu)} + 2 \sup_{x_n \in X_\delta} |f(x_n)|. \tag{55}$$

We will show in Theorem 4 below that for every admissible p, r, q and σ, there is $C > 0$ such that for every $f \in \mathcal{B}_{r,q}^\sigma(M; \mathcal{D})$, we have

$$\inf_{f_R \in \Sigma_{[0,R]}^p} \|f - f_R\|_{L^p(M;d\mu)} \leq CR^{d(1/r - 1/p)_+ - \sigma} \|f\|_{\mathcal{B}_{r,q}^\sigma(M;\mathcal{D})}, \tag{56}$$

which finishes the proof. □

Remark 4 Note that for $p = \infty$, we obtain in Theorem 4 the anticipated optimal rate[1] $\delta^{d/r - \sigma}$.

---

[1] For $p \neq \infty$, the convergence rate is $\delta^{d/r - \sigma}$ instead of the anticipated rate $\delta^{d(1/r - 1/p)_+ - \sigma}$, where $(x)_+ = \max\{x, 0\}$ (cf. [59] for the case of classical Sobolev spaces). This is most likely due to the fact that we work with the global estimates (51) and (52) instead of local estimates on a cover (see also [40,59]).

Corollary 1 *In the case $p = \infty$, we can skip the assumption $\mu(M) < \infty$ in Theorem 3 and we get that there is a constant $c > 0$ with the following property: For every $\delta \in (0, \gamma/b)$ and for every maximal δ-net X_δ with associated disjoint partition A_n of M, $x_n \in X_\delta$, we have for all $f \in \mathcal{B}_{r,q}^\sigma(M; \mathcal{D})$*

$$\|f\|_{L^\infty(M;d\mu)} \le c\delta^{\sigma - \frac{d}{r}} \|f\|_{\mathcal{B}_{r,q}^\sigma(M;\mathcal{D})} + 2 \sup_{x_n \in X_\delta} |f(x_n)|. \tag{57}$$

It remains to prove the best approximation error estimate (56). The following theorem generalizes results from [10, Theorem 3.15].

Theorem 4 *Let $R \ge b > 1$. Suppose that $1 \le r$, $p \le \infty$, $0 < q \le \infty$ and $\sigma > d(1/p - 1/r)_+$, where again $(x)_+ = \max\{x, 0\}$. If $r \le p$, then there is a constant $c > 0$ such that for every $f \in \mathcal{B}_{r,q}^\sigma(M; \mathcal{D})$*

$$\inf_{g \in \Sigma_{[0,R]}^p} \|f - g\|_{L^p(M;d\mu)} \le cR^{d(1/r - 1/p) - \sigma} \|f\|_{\mathcal{B}_{r,q}^\sigma(M;\mathcal{D})}. \tag{58}$$

If $\mu(M) < \infty$, then for all $1 \le p, r \le \infty$ (not necessarily $r \le p$), there is a constant such that for every $f \in \mathcal{B}_{r,q}^\sigma(M; \mathcal{D})$

$$\inf_{g \in \Sigma_{[0,R]}^p} \|f - g\|_{L^p(M;d\mu)} \le cR^{d(1/r - 1/p)_+ - \sigma} \|f\|_{\mathcal{B}_{r,q}^\sigma(M;\mathcal{D})}. \tag{59}$$

Proof We follow the lines of the proof of [10, Theorem 3.15]. Here we employ different truncation functions ϕ and ψ, which yield an equivalent norm to the Besov norm corresponding to the reproducing kernel $K^{(\sigma)}$. Precisely, let $\phi \in \mathcal{A}(b, c_1, c_2)$, and set $\psi := \Delta\phi$. Then ϕ and ψ satisfy the assumptions of Definition 2, and (ϕ, ψ) form a partition of unity. Now choose $L_R \in \mathbb{N}$, such that $b^{L_R+1} \le R \le b^{L_R+2}$. Note that this is possible since $R \ge b$. Then by Lemma 1, we have for all $f \in L^p(M; d\mu)$

$$\phi(\sqrt{\mathcal{D}})f + \sum_{\ell=1}^{L_R} \psi(b^{-\ell}\sqrt{\mathcal{D}})f = \phi(b^{-L_R}\sqrt{\mathcal{D}})f \in \Sigma_{[0,b^{(L_R+1)}]}^p \subset \Sigma_{[0,R]}^p.$$

Since $\psi(b^{-\ell}\sqrt{\mathcal{D}})f \in \Sigma_{[0,b^{\ell+1}]}^p$, we have by (27) and (30)

$$\inf_{g \in \Sigma_{[0,R]}^p} \|f - g\|_{L^p(M;d\mu)} \le \left\| f - \left(\phi(\sqrt{\mathcal{D}})f + \sum_{\ell=1}^{L_R} \psi(b^{-\ell}\sqrt{\mathcal{D}})f \right) \right\|_{L^p(M;d\mu)}$$

$$\le \sum_{\ell=L_R+1}^\infty \left\| \psi(b^{-\ell}\sqrt{\mathcal{D}})f \right\|_{L^p(M;d\mu)} \le \sum_{\ell=L_R+1}^\infty C(b^\ell)^{d(1/r-1/p)} \|\psi(b^{-\ell}\sqrt{\mathcal{D}})f\|_{L^r(M;d\mu)}$$

$$= Cb^{(L_R+1)d(1/r-1/p)} \sum_{\ell=0}^\infty b^{\ell d(1/r-1/p)} \left\| \psi(b^{-(\ell+L_R+1)}\sqrt{\mathcal{D}})f \right\|_{L^r(M;d\mu)}.$$

Note that, since $\phi \ge 0$,

$$\|f\|_{\mathcal{B}_{r,q}^\sigma(M;\mathcal{D})}^q \ge c \sum_{\ell=1}^\infty b^{q\sigma\ell} \left\| \psi(b^{-\ell}\sqrt{\mathcal{D}})f \right\|_{L^r(M;d\mu)}^q \ge b^{\sigma m q} \left\| \psi(b^{-m}\sqrt{\mathcal{D}})f \right\|_{L^r(M;d\mu)}^q$$

for every $m \in \mathbb{N}$. Thus, we can further estimate (recall that $b^{L_R+1} \le R$)

$$\inf_{g \in \Sigma_{[0,R]}^p} \|f - g\|_{L^p(M;d\mu)} \le Cb^{(L_R+1)d(1/r-1/p)} \sum_{\ell=0}^\infty b^{\ell d(1/r-1/p)} b^{-\sigma(\ell+L_R+1)} \|f\|_{\mathcal{B}_{r,q}^\sigma(M;\mathcal{D})}$$

$$\le b^{(L_R+1)(d(1/r-1/p)-\sigma)} \|f\|_{\mathcal{B}_{r,q}^\sigma(M;\mathcal{D})} \sum_{\ell=0}^\infty b^{\ell(d(1/r-1/p)-\sigma)} \le CR^{d(1/r-1/p)-\sigma} \|f\|_{\mathcal{B}_{r,q}^\sigma(M;\mathcal{D})}, \tag{60}$$

where the geometric series converges because of the condition $\sigma > d\left(1/r - 1/p\right)$. This concludes the proof of (58).

If $\mu(M) < \infty$ and $r \geq p$, we proceed similarly to the proof of [10, Proposition 3.20]. We use Hölder's inequality $(1/p = 1/r + (r-p)/(pr))$ to obtain for every $\ell \in \mathbb{N}$

$$\left\| \psi(b^{-\ell}\sqrt{\mathcal{D}})f \right\|_{L^p(M;d\mu)} \leq \left\| \psi(b^{-\ell}\sqrt{\mathcal{D}})f \right\|_{L^r(M;d\mu)} \mu(M)^{(r-p)/(pr)}$$

$$= Cb^{-\ell\sigma} \|f\|_{\mathcal{B}^\sigma_{r,q}(M;\mathcal{D})},$$

and similarly for ϕ. This estimate replaces the Nikloskii-type inequality (30) in (60), and the rest of the proof follows as above. □

4.2 $L^\infty(M;d\mu)$-estimates for the case $\mu(M) < \infty$

We now follow the lines of [59,18] and derive a generalized polynomial reproduction, where the spectral spaces $\Sigma^\infty_{[0,R]}$ play the role of the classical polynomial spaces. To prove reproduction formulas, we use norming sets (see [23,57]), and proceed along the lines of [57] (see also [58]). In the case $\mu(M) < \infty$, we will always work with finite discrete point sets $X_N \subset M$. To stress this fact, we will indicate the number of points with a subscript $N \in \mathbb{N}$.

Proposition 3 *There is a constant $\tilde{C} > 0$ such that for all finite sets $X_N = \{x_1, \ldots, x_N\} \subset M$ with $h_{X_N,M} \leq \tilde{C}$ and all $R \leq \tilde{C}/h_{X_N,M}$, the sampling operator $T : \Sigma^\infty_{[0,R]} \to \mathbb{R}^N$, $f \mapsto (f(X_N))$ is injective, and $\|T^{-1}\| \leq 2$, where $\Sigma^\infty_{[0,R]}$ and $\mathbb{R}^{X_N}$ are equipped with the sup-norms.*

Proof Let $f \in \Sigma^\infty_{[0,R]}$ with $\|f\|_{L^\infty(M;d\mu)} = 1$. Then, there exists $x^* \in M$ such that $|f(x^*)| \geq 3/4$. We need to show that there exists $x_n \in X_N$ with $|f(x_n)| \geq 1/2$. By definition of the fill distance, there exists $x_n \in X$ such that $\rho(x_n, x^*) \leq h_{X_N, \Omega}$. We get from (31) that

$$|f(x^*) - f(x_n)| \leq C(Rh_{X_N,M})^{\alpha_H} \leq C\tilde{C}^\alpha \leq 1/4,$$

where α_H is from (14) and the last inequality holds if $\tilde{C} > 0$ is chosen small enough. Therefore,

$$|f(x_n)| \geq |f(x^*)| - |f(x^*) - f(x_n)| \geq 3/4 - 1/4 = 1/2.$$

This concludes the argument. □

Furthermore, we will use the following result, which is a special case of [58, Theorem 3.4].

Theorem 5 *Suppose V is a finite dimensional normed linear space and let $\{x_1, \ldots, x_N\}$ be such that $T : V \to \mathbb{R}^N$, $v \mapsto (v(x_1), \ldots, v(x_N))^T$ is injective. Then, for every $\varphi \in V^*$ there exists a vector $u \in \mathbb{R}^N$ such that $\varphi(v) = \sum_{n=1}^N u_n v(x_n)$ for every $v \in V$, and $\|u\|_{\mathbb{R}^{N*}} \leq \|\varphi\|_{V^*} \|T^{-1}\|_{T(V) \to V}$.*

Combining Theorem 5 and Proposition 3 gives the following result.

Proposition 4 *There is a constant $\tilde{C} > 0$ such that for all point sets $X_N = \{x_1, \ldots, x_N\} \subset M$ with $h_{X_N,M} \leq \tilde{C}$, there exist $a_n : M \to \mathbb{R}$, $n = 1, \ldots, N$ such that for $R \leq \tilde{C}/h_{X_N,M}$*

(i) $\sum_{n=1}^N a_n(x) f_R(x_n) = f_R(x)$ for all $x \in M$ and all $f_R \in \Sigma^\infty_{[0,R]}$; and
(ii) $\sum_{n=1}^N |a_n(x)| \leq 2$ for all $x \in M$.

With Proposition 4 at hand, we can proceed essentially verbatim along the lines of the proof of Theorem 3 and obtain finally the following result.

Theorem 6 *Let $1 \leq \tilde{p} \leq \infty$, $0 < q \leq \infty$ and $\sigma > d/\tilde{p}$. There are $C, h_0 > 0$ with the following property: For every set $X_N = \{x_1, \ldots, x_N\} \subset M$ with $h_{X_N,M} \leq h_0$, we have for all $f \in \mathcal{B}^\sigma_{\tilde{p},q}(M;\mathcal{D})$*

$$\|f\|_{L^\infty(M;d\mu)} \leq C\left(h^{\sigma-d/\tilde{p}}_{X_N,M}\|f\|_{\mathcal{B}^\sigma_{\tilde{p},q}(M;\mathcal{D})} + \|f|_{X_N}\|_{\ell^\infty(X_N)}\right).$$

5 Truncation of the kernel

In this section, we derive several technical estimates concerning the approximation of the kernel $K^{(\sigma)}$. Here, we closely follow [8]. The estimates are practically relevant, since linear combinations of truncated kernels lie in finite dimensional spectral spaces. Elements of spectral spaces will take the role of polynomials in classical analysis for Sobolev spaces on Euclidean domains. Furthermore, from now on we assume $\sup_{x \in M} \mu(B(x,r)) \leq C(r)$ for all $r > 0$. Note that this is trivially true if $\mu(M) < \infty$.

Lemma 4 *There is $C > 0$ such that for all N, $\ell \in \mathbb{N}$, all $X_N = \{x_1, \ldots, x_N\} \subset M$, and all $a_1, \ldots, a_N \in \mathbb{R}$,*

$$\Big\| \sum_{n=1}^{N} a_n b^{-2\sigma\ell} \boldsymbol{S}^{(\ell;\sigma)}(x_n, \cdot) \Big\|_{L^p(M;d\mu)} \leq C \Big\| \sum_{n=1}^{N} a_n K^{(\sigma)}(x_n, \cdot) \Big\|_{L^p(M;d\mu)}.$$

Proof Recall from (40) that $b^{-2\sigma\ell} \boldsymbol{s}^{(\ell;\sigma)}(u) = \Delta_\ell \phi(u) \boldsymbol{k}^{(\sigma)}(u)$. Hence,

$$b^{-2\sigma\ell} \boldsymbol{S}^{(\ell;\sigma)}(\cdot, z) = \Delta_\ell \phi(\sqrt{\mathcal{D}}) \boldsymbol{K}^{(\sigma)}(\cdot, z) \qquad \text{for all } z \in M.$$

Thus, by the triangle inequality and (26), we obtain

$$\left\| \sum_{n=1}^{N} a_n b^{-2\sigma\ell} \boldsymbol{S}^{(\ell;\sigma)}(\cdot, x_n) \right\|_{L^p(M;d\mu)} \leq \left\| \Delta_\ell \phi(\sqrt{\mathcal{D}}) \sum_{n=1}^{N} a_n K^{(\sigma)}(\cdot, x_n) \right\|_{L^p(M;d\mu)}$$

$$\leq 2 \sup_{0 < \delta \leq 1} \left\| \phi(\delta\sqrt{\mathcal{D}}) \sum_{n=1}^{N} a_n K^{(\sigma)}(\cdot, x_n) \right\|_{L^p(M;d\mu)} \leq 2C \left\| \sum_{n=1}^{N} a_n K^{(\sigma)}(\cdot, x_n) \right\|_{L^p(M;d\mu)}.$$

This concludes the argument. $\qquad\square$

Now, we use the following notation. For $X_N = \{x_1, \ldots, x_N\} \subset M$, $x \in M$ and $k \in \mathbb{N}_0$, we set

$$\mathcal{A}_k(x) := B\left(x, (k+1) q_{X_N}\right) \setminus B\left(x, k q_{X_N}\right), \tag{61}$$

where q_{X_N} denotes the separation distance of X_N as defined in (48). We need a combinatorial estimate that we briefly discuss. Note that in the special case of quasi-uniform point sets in $\mathbb{R}^d$, there is a constant $C > 0$ such that for all $X_N \subset \mathbb{R}^d$, every $k \in \mathbb{N}$ and every $x \in \mathbb{R}^d$, we have

$$\#(X_N \cap \mathcal{A}_k(x)) \leq C k^{d-1}.$$

In the general setting, we have by (8) and (11) for all $X_N \subset M$, every $k \in \mathbb{N}_0$, every $x \in M$ and every $x_n \in X_N \cap \mathcal{A}_k(x)$

$$\mu\left(B(x, (k+2)q_{X_N})\right) \leq 2^{2d}(k+2)^d \mu\left(B(x, q_{X_N}/2)\right) \leq 2^{3d}(k+2)^d \left(1 + 2\rho(x_n,x)/q_{X_N}\right)^d \mu\left(B(x_n, q_{X_N}/2)\right)$$

$$\leq 2^{4d}(k+2)^{2d} \mu\left(B(x_n, q_{X_N}/2)\right).$$

Furthermore, by definition of the separation distance q_{X_N}, we have $B(x_n, q_{X_N}/2) \cap B(x_m, q_{X_N}/2) = \emptyset$ if $x_n \neq x_m \in X_N$. Thus, since $B(x_n, q_{X_N}/2) \subset B(x, (k+2)q_{X_N})$ for every $x_n \in \mathcal{A}_k(x)$, we have the rough estimate

$$\mu\left(B(x, (k+2)q_{X_N})\right) \geq \sum_{x_n \in X_N \cap \mathcal{A}_k(x)} \mu\left(B(x_n, q_{X_N}/2)\right)$$

$$\geq 2^{-4d}(k+2)^{-2d} \mu\left(B(x, (k+2)q_{X_N})\right) \#(X_N \cap \mathcal{A}_k(x)),$$

which implies that

$$\#(X_N \cap \mathcal{A}_k) \leq 2^{4d}(k+2)^{2d}. \tag{62}$$

In particular, there is some $\tau > 2d + 4$, which depends only on d such that there is a constant $C_1 > 0$ with

$$\sum_{k=1}^{\infty} \# \left(X_N \cap \mathcal{A}_k(x) \right) k^{d/2-\tau} < C_1 \qquad \text{for all } x \in M \quad \text{and all } X_N \subset M. \tag{63}$$

Next, we will need the following auxiliary lemma, which is in the spirit of [8, Proposition 4.2].

Lemma 5 *Let $b > 1$, $\tau > 2d + 4$ be such that (63) holds, and let $e \in \mathcal{G}(b)$. There are constants C, $C_1 > 0$ such that, for all $X_N \subset M$, all $\delta < \min\{1, q_{X_N}\}$ and all $x \in M$, we have for the associated integral kernel $\boldsymbol{E}_\delta$ the estimates*

$$\sum_{\substack{x_n \in X_N \\ \rho(x,x_n) \geq q_{X_N}}} |\boldsymbol{E}_\delta(x,x_n)| \leq C |B(x,\delta)|^{-1} (q_{X_N}/\delta)^{d/2-\tau}, \quad \text{and}$$

$$\sum_{\substack{x_n \in X_N \\ \rho(x,x_n) < q_{X_N}}} |\boldsymbol{E}_\delta(x,x_n)| \leq C |B(x,\delta)|^{-1}. \tag{64}$$

The constant C depends only on the parameters of the space, the constant c_τ from (24), and the constant C_1 from (63).

Proof We show the two inequalities separately. For all $X_N \subset M$, all $x \in M$ and all $\delta \leq \min\{q_{X_N}, 1\}$, we have by (24), (63) and since $d/2 - \tau < 2d - \tau < 0$

$$\sum_{\substack{x_n \in X_N \\ \rho(x,x_n) \geq q_{X_N}}} \left| \boldsymbol{E}_\delta(x,x_n) \right| \leq C |B(x,\delta)|^{-1} \sum_{\substack{x_n \in X_N \\ \rho(x,x_n) > q_{X_N}}} (1 + \rho(x,x_n)/\delta)^{d/2-\tau}$$

$$\leq C |B(x,\delta)|^{-1} \sum_{k=1}^{\infty} \sum_{x_n \in \mathcal{A}_k} \left(1 + (kq_{X_N})/\delta \right)^{d/2-\tau}$$

$$\leq C |B(x,\delta)|^{-1} \left(q_{X_N}/\delta \right)^{d/2-\tau} \sum_{k=1}^{\infty} \# (X_N \cap \mathcal{A}_k) k^{d/2-\tau} \leq CC_1 |B(x,\delta)|^{-1} (q_{X_N}/\delta)^{d/2-\tau}.$$

To show the second assertion, we note first that by (62) with $k = 0$ we have

$$\# X_N \cap B(x, q_{X_N}) \leq 2^{6d} \qquad \text{for all} \ \ x \in M \ \text{and all} \ \ X_N \subset M. \tag{65}$$

Thus, we obtain by (24) and (65)

$$\sum_{\substack{x_n \in X_N \\ \rho(x,x_n) \leq q_{X_N}}} |\boldsymbol{E}_\delta(x,x_n)| \leq C\mu(B(x,\delta))^{-1} \sum_{\substack{x_n \in X_N \\ \rho(x,x_n) \leq q_{X_N}}} \left(1 + \frac{\rho(x,x_n)}{\delta} \right)^{d/2-\tau}$$

$$\leq C\mu(B(x,\delta))^{-1} \# (X_N \cap B(x,q_{X_N})) \leq C\mu(B(x,\delta))^{-1},$$

which is exactly (64). $\qquad \square$

For $g \in \Sigma_{[0,R]}^p$ and $X_N = \{x_1, \ldots, x_N\}$ set

$$\|g\|_{\ell^p(X_N)} := \begin{cases} \left(\sum_{n=1}^N |g\,(x_n)|^p\right)^{1/p} & \text{if } 1 \le p < \infty, \\ \max_{n=1,\ldots,N} |g\,(x_n)| & \text{if } p = \infty. \end{cases} \tag{66}$$

We will now, under suitable assumptions, derive explicit norm equivalence constants for $\|\cdot\|_{\ell^p(X_N)}$ and $\|\cdot\|_{L^p(M;d\mu)}$ on $\Sigma_{[0,R]}^p$. The following lemma is in the spirit of [8, Theorem 4.3].

Lemma 6 *Suppose that $1 \le p \le \infty$, $R > 1$, and let $\tau > 2d + 4$ be such that (63) holds. Then there is $C > 0$ such that, for all $\delta \le 1$, all $g \in \Sigma_{[0,R/\delta]}^p$ and all $X_N \subset M$,*

$$\|g\|_{\ell^p(X_N)} \le C(R) \begin{cases} \sup_{y \in M} \mu\,(B\,(y,\delta))^{-1/p} \left(1 + (q_{X_N}/\delta)^{d/2-\tau}\right)^{1/p} \|g\|_{L^p(M;d\mu)} & \text{if } 1 \le p < \infty, \\ \|g\|_{L^\infty(M;d\mu)} & \text{if } p = \infty. \end{cases} \tag{67}$$

Proof We follow the lines of [8, Theorem 4.3]. Let $e \in C^\infty\,([0,\infty))$ with $e|_{[0,R]} \equiv 1$. Then $g = e(\delta\sqrt{\mathcal{D}})g$ for all $g \in \Sigma_{[0,R/\delta]}^p$, and we thus get from Lemma 5

$$\|g\|_{\ell^1(X_N)} = \sum_{n=1}^N |g\,(x_n)| = \sum_{n=1}^N \left|e(\delta\sqrt{\mathcal{D}})g\,(x_n)\right| = \sum_{n=1}^N \left|\int_M \boldsymbol{E}_\delta\,(x_n,y)\,g(y)\,d\mu\,(y)\right|$$

$$\le \sup_{y \in M} \sum_{n=1}^N |\boldsymbol{E}_\delta(x_n,y)|\,\|g\|_{L^1(M;d\mu)} \le C \sup_{y \in M} |B\,(y,\delta)|^{-1} \left(1 + (q_{X_N}/\delta)^{d/2-\tau}\right) \|g\|_{L^1(M;d\mu)}.$$

Since $X_N \subset M$, we have $\|g\|_{\ell^\infty(X_N)} \le \|g\|_{L^\infty(M;d\mu)}$. The assertion then follows by Riesz-Thorin interpolation. $\qquad\square$

Now we are in the position to show that the translates $\boldsymbol{S}^{(\ell;\sigma)}\,(\cdot,x_n)$ for points $x_n \in X_N$ build a weighted Riesz basis in $L^p\,(M;d\mu)$. The following lemma is similar to a result from [22] which holds for a less general class of compact metric measure spaces.

Lemma 7 *Suppose that $b \ge \tilde{b}^{4/3}$ with $\tilde{b} > 1$ from Remark 2 and let $1 \le p, p' \le \infty$ such that $\frac{1}{p} + \frac{1}{p'} = 1$. There are constants $c', c, C > 0$ and $q_0 > 0$ such that, for all $N \in \mathbb{N}$, all $X_N \subset M$ with $q_{X_N} \le q_0$, all $\ell \ge \log_b\left(c'q_{X_N}^{-1}\right)$, all $\mathbf{a} = (a_1, \ldots, a_N)^T \in \mathbb{R}^N$ and all $\sigma > 0$, there holds*

$$c \sup_{x \in M} \mu\left(B\left(x,b^{-\ell}\right)\right)^{1/p'} \left\|\sum_{n=1}^N a_n \boldsymbol{S}^{(\ell;\sigma)}\,(\cdot,x_n)\right\|_{L^p(M;d\mu)} \le \|\mathbf{a}\|_{\ell^p(N)}$$

$$\le C \max_{1 \le n \le N} \sup_{x \in M} \frac{\mu\left(B\left(x_n,b^{-\ell}\right)\right)}{\mu\,(B(x,b^{-\ell}))^{1/p}} \left\|\sum_{n=1}^N a_n \boldsymbol{S}^{(\ell;\sigma)}\,(\cdot,x_n)\right\|_{L^p(M;d\mu)}, \tag{68}$$

and similarly for $p = \infty$.

Proof We prove the two inequalities separately. Recall from (38) that $\boldsymbol{s}^{(\ell;\sigma)}(u) = \boldsymbol{w}^{(\ell;\sigma)}(b^{-\ell}u)$. Since $\boldsymbol{w}_\ell^{(\sigma)} \equiv \boldsymbol{w}_2^{(\sigma)}$ for all $\ell \ge 2$ (see Lemma 3), we have by Lemma 5 with $\delta = b^{-\ell}$

$$\left\|\sum_{n=1}^N a_n \boldsymbol{S}^{(\ell;\sigma)}\,(\cdot,x_n)\right\|_{L^\infty(M;d\mu)} \le \|\mathbf{a}\|_{\ell^\infty(N)} \sup_{x \in M} \sum_{n=1}^N |\boldsymbol{S}^{(\ell;\sigma)}(x,x_n)|$$

$$\le C\,\|\mathbf{a}\|_{\ell^\infty} \sup_{x \in M} \mu\left(B\left(x,b^{-\ell}\right)\right)^{-1} \left(1 + \left(q_{X_N}b^\ell\right)^{d/2-\tau}\right), \tag{69}$$

where C is independent of ℓ. Furthermore, from (26) we get

$$\left\| \sum_{n=1}^{N} a_n \boldsymbol{S}^{(\ell;\sigma)} (\cdot, x_n) \right\|_{L^1(M;d\mu)} \leq \|\mathbf{a}\|_{\ell^1(N)} \sup_{x \in M} \int_M \left| \boldsymbol{S}^{(\ell;\sigma)}(y,x) \right| d\mu(y)$$

$$= \|\mathbf{a}\|_{\ell^1(N)} \|\boldsymbol{s}^{(\ell;\sigma)}(\sqrt{\mathcal{D}})\|_{L^\infty(M;d\mu) \to L^\infty(M;d\mu)} \leq C\|\mathbf{a}\|_{\ell^1(N)}. \tag{70}$$

Thus, combining (70) and (69), we obtain by Riesz-Thorin interpolation

$$\left\| \sum_{n=1}^{N} a_n \boldsymbol{S}^{(\ell;\sigma)} (\cdot, x_n) \right\|_{L^p(M;d\mu)} \leq \sup_{x \in M} \mu\left(B\left(x, b^{-\ell}\right)\right)^{-1/p'} \left(1 + \left(q_{X_N} b^\ell\right)^{d(2p'-\tau/p')}\right) \|\mathbf{a}\|_{\ell^p(N)}$$

$$\leq c \sup_{x \in M} \mu\left(B\left(x, b^{-\ell}\right)\right)^{-1/p'} \|\mathbf{a}\|_{\ell^p(N)}.$$

This finishes the proof of the first inequality.

Since $\rho(x_n, x_m) \geq q_{X_N}$ for $x_n \neq x_m \in X_N$, we get with Lemma 5

$$\sum_{\substack{n=1 \\ n \neq m}}^{N} \left| \boldsymbol{S}^{(\ell;\sigma)}(x_m, x_n)\right| \leq c\mu\left(B\left(x_m, b^{-\ell}\right)\right)^{-1} \left(q_{X_N} b^\ell\right)^{d/2 - \tau}. \tag{71}$$

Furthermore, Lemma 3 implies that $\boldsymbol{s}^{(\ell;\sigma)} \geq c\mathbf{1}_{[b^\ell, b^{\ell+3/4}]}$, and we thus obtain from Remark 2 (note that for $q_0 > 0$ small enough, the assumptions are satisfied)

$$\boldsymbol{S}^{(\ell;\sigma)}(x_m, x_m) \geq c\mathbf{1}_{[b^\ell, b^{\ell+3/4}]}(x_m, x_m) = c\mathbf{1}_{[b^\ell, b^{3/4} \cdot b^\ell]}(x_m, x_m)$$

$$\geq c\mathbf{1}_{[b^\ell, \tilde{b} \cdot b^\ell]}(x_m, x_m) \geq C\mu\left(B\left(x_m, b^{-\ell}\right)\right)^{-1}. \tag{72}$$

Inserting (72) into (71) yields for $\ell \geq \log_b\left(c' q_{X_N}^{-1}\right)$

$$\sum_{\substack{n=1 \\ n \neq m}}^{N} \left| \boldsymbol{S}^{(\ell;\sigma)}(x_m, x_n)\right| \leq C\left(q_{X_N} b^\ell\right)^{d/2 - \tau} \boldsymbol{S}^{(\ell;\sigma)}(x_m, x_m) \leq \frac{1}{2}\boldsymbol{S}^{(\ell;\sigma)}(x_m, x_m) \tag{73}$$

for some constant $c' > 0$ independent of ℓ. We now proceed similarly to [8]. Consider the Gramian matrix $G := \left(\boldsymbol{S}^{(\ell;\sigma)}(x_n, x_m)\right)_{n,m=1,\ldots,N} \in \mathbb{R}^{N \times N}$. Note that G is a symmetric matrix and satisfies (see (73)) $\sum_{\substack{n=1 \\ n \neq m}}^{N} \left| \boldsymbol{S}^{(\ell;\sigma)}(x_m, x_n)\right| \leq \frac{1}{2}\boldsymbol{S}^{(\ell;\sigma)}(x_m, x_m)$ for all $m = 1, \ldots N$. By (72) and (10) we have

$$\lambda := \min_{1 \leq n \leq N} \boldsymbol{S}^{(\ell;\sigma)}(x_n, x_n) \geq c \min_{1 \leq n \leq N} \left(\mu\left(B(x_n, b^{-\ell})\right)\right)^{-1} > 0. \tag{74}$$

Hence, it follows from [8, Proposition 6.1] that G is invertible, and

$$\left\| G^{-1} \right\|_{\ell^p(N) \to \ell^p(N)} \leq \left((1 - 1/2)\lambda\right)^{-1} \qquad \text{for all } 1 \leq p \leq \infty. \tag{75}$$

We use (75) and (67) with $R := b$ and $\delta := b^{-\ell}$ to deduce for $1 \le p < \infty$

$$\|\mathbf{a}\|_{\ell^p(N)} \le \|G^{-1}\|_{\ell^p(N)\to\ell^p(N)} \left\|\sum_{n=1}^N a_n \boldsymbol{S}^{(\ell;\sigma)}(\cdot, x_n)\right\|_{\ell^p(X_N)}$$

$$\le C \sup_{x\in M} \max_{1\le n\le N} \frac{\mu\left(B\left(x_n, b^{-\ell}\right)\right)}{(\mu\left(B(x, b^{-\ell})\right))^{1/p}} \left(1 + (q_{X_N} b^\ell)^{d/2-\tau}\right)^{1/p} \left\|\sum_{n=1}^N a_n \boldsymbol{S}^{(\ell;\sigma)}(\cdot, x_n)\right\|_{L^p(M;d\mu)}$$

$$\le C \sup_{x\in M} \max_{1\le n\le N} \frac{\mu\left(B\left(x_n, b^{-\ell}\right)\right)}{\mu\left(B\left(x, b^{-\ell}\right)\right)^{1/p}} \left\|\sum_{n=1}^N a_n \boldsymbol{S}^{(\ell;\sigma)}(\cdot, x_n)\right\|_{L^p(M;d\mu)} .$$

The assertion follows similarly for $p = \infty$. $\square$

Remark 5 Since by (10) and (8) we have $\mu\left(B\left(y, r\right)\right) \ge c r^d$ for all $y \in M$ and all $0 < r < r_0$, we can estimate the quantity on the right-hand side of (68) by

$$\sup_{x\in M} \max_{1\le n\le N} \frac{\mu\left(B\left(x_n, b^{-\ell}\right)\right)}{(\mu\left(B(x, b^{-\ell})\right))^{1/p}} \le C b^{\ell d/p} \max_{1\le n\le N} \mu\left(B\left(x_n, b^{-\ell}\right)\right) .$$

If additionally $\mu\left(B\left(y, r\right)\right) \le C r^d$ for all $y \in M$ and $r > 0$, then (68) reduces to

$$c b^{-\ell d/p'} \left\|\sum_{n=1}^N a_n \boldsymbol{S}^{(\ell;\sigma)}(\cdot, x_n)\right\|_{L^p(M;d\mu)} \le \|\mathbf{a}\|_{\ell^p(N)} \le C b^{-\ell d/p'} \left\|\sum_{n=1}^N a_n \boldsymbol{S}^{(\ell;\sigma)}(\cdot, x_n)\right\|_{L^p(M;d\mu)} .$$

Note that the condition $\ell \ge \log_b(c' q_{X_N})$ in Lemma 7 couples the truncation index L to the sampling points X_N. Thus, we are in the position to analyze approximation schemes using the properly truncated kernel. This is the first step towards a numerically feasible method for kernels that are not given analytically but merely by an infinite series as in (1). Moreover, since the $\boldsymbol{S}^{(\ell;\sigma)}$ are usually also not analytically given but must be approximated themselves, there is a second step needed where a proper approximation of the truncated kernel is computed. This will be discussed in Section 6. We now proceed along the general lines of [18,39]. We use the following short-hand notation.

Notation 4 *For $\sigma > d/2$, N, $L \in \mathbb{N}$, $\mathbf{a} \in \mathbb{R}^N$, and $X_N = \{x_1, \ldots, x_N\} \subset M$ set*

$$\mathcal{I}_{\sigma;\mathbf{a};X_N}(\cdot) := \sum_{n=1}^N a_n K^{(\sigma)}(\cdot, x_n) = \sum_{n=1}^N a_n \sum_{\ell=0}^\infty b^{-2\sigma\ell} \boldsymbol{S}^{(\ell;\sigma)}(\cdot, x_n), \qquad and$$

$$\mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)}(\cdot) := \sum_{n=1}^N a_n \sum_{\ell=0}^L b^{-2\sigma\ell} \boldsymbol{S}^{(\ell;\sigma)}(\cdot, x_n) = \sum_{n=1}^N a_n \phi\left(b^{-L}\sqrt{\mathcal{D}}\right) \boldsymbol{K}^{(\sigma)}(\cdot, x_n), \qquad (76)$$

where the last equation follows as in (47), i.e., using Lemma 1 in the last line, we have

$$\sum_{\ell=0}^L b^{-2\sigma\ell} \boldsymbol{S}^{(\ell;\sigma)}(\cdot, x_n) = \sum_{\ell=0}^L b^{-2\sigma\ell} \boldsymbol{g}^{(\ell;\sigma)}(b^{-\ell}\sqrt{\mathcal{D}}) \boldsymbol{\Delta_\ell \Phi}(\cdot, x_n)$$

$$= \sum_{\ell=0}^L \boldsymbol{k}^{(\sigma)}(\sqrt{\mathcal{D}}) \boldsymbol{\Delta_\ell \Phi}(\cdot, x_n) = \sum_{\ell=0}^L \Delta_\ell \phi(\sqrt{\mathcal{D}}) \boldsymbol{K}^{(\sigma)}(\cdot, x_n)$$

$$= \phi(b^{-L}\sqrt{\mathcal{D}}) \boldsymbol{K}^{(\sigma)}(x_n, \cdot).$$

Lemma 8 *Suppose that $b \geq \tilde{b}^{4/3}$. Let $2\sigma > d/p'$, $1 \leq p \leq \infty$, and $0 < \gamma < 1$. Then there are constants $C, q_0 > 0$ such that for all discrete sets $X_N = \{x_1, \ldots, x_N\} \subset M$ with separation distance $q_{X_N} \leq q_0$, all $L \geq C \log_b \left(C q_{X_N}^{-1} \right)$ and all $\mathbf{a} \in \mathbb{R}^N$, we have*

$$\left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} - \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)} \right\|_{L^p(M;d\mu)} \leq \gamma \left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{L^p(M;d\mu)}. \tag{77}$$

Proof By definition, we have

$$\left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} - \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)} \right\|_{L^p(M;d\mu)} = \left\| \sum_{n=1}^{N} a_n \sum_{\ell=L+1}^{\infty} b^{-2\sigma\ell} \mathbf{S}^{(\ell;\sigma)}(\cdot, x_n) \right\|_{L^p(M;d\mu)}$$

$$\leq \sum_{\ell=L+1}^{\infty} b^{-2\sigma\ell} \left\| \sum_{n=1}^{N} a_n \mathbf{S}^{(\ell;\sigma)}(\cdot, x_n) \right\|_{L^p(M;d\mu)}. \tag{78}$$

If we choose $C > c'$, we have by Lemma 7, for all $\ell \geq L+1$

$$\left\| \sum_{n=1}^{N} a_n \mathbf{S}^{(\ell;\sigma)}(\cdot, x_n) \right\|_{L^p(M;d\mu)} \leq C \sup_{x \in M} \mu \left(B(x, b^{-\ell}) \right)^{-1/p'} \|\mathbf{a}\|_{\ell^p(N)} \leq \tilde{C} b^{\ell d/p'} \|\mathbf{a}\|_{\ell^p(N)}, \tag{79}$$

where in the last step we used (10) and (8) which imply that $\mu \left(B(x, b^{-\ell}) \right) \geq C b^{-\ell d}$. Using (79) to further estimate (78), we obtain

$$\left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} - \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)} \right\|_{L^p(M;d\mu)} \leq C \sum_{\ell=L+1}^{\infty} b^{(d/p'-2\sigma)\ell} \|\mathbf{a}\|_{\ell^p(N)} \leq C b^{-L(2\sigma-d/p')} \|\mathbf{a}\|_{\ell^p(N)}. \tag{80}$$

Choose now $\tilde{\ell} \in \mathbb{N}$ such that $\log_b(c' q_{X_N}^{-1}) \leq \tilde{\ell} \leq 2 \log_b(c' q_{X_N}^{-1})$. Note that this is possible if q_0 is small enough. Then we obtain by Lemmata 7 and 4, and Remark 5

$$\|\mathbf{a}\|_{\ell^p(N)} \leq C b^{\tilde{\ell}d/p} \sup_{x \in M} \mu \left(B \left(x, b^{-\tilde{\ell}} \right) \right) \left\| \sum_{n=1}^{N} a_n \mathbf{S}^{(\tilde{\ell};\sigma)}(\cdot, x_n) \right\|_{L^p(M;d\mu)}$$

$$\leq C b^{\tilde{\ell}d/p} \sup_{x \in M} \mu \left(B \left(x, b^{-\tilde{\ell}} \right) \right) b^{2\sigma\tilde{\ell}} \left\| \sum_{n=1}^{N} a_n K^{(\sigma)}(\cdot, x_n) \right\|_{L^p(M;d\mu)}$$

$$\leq C q_X^{-(2\sigma+d/p)} \left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{L^p(M;d\mu)} \tag{81}$$

since $\mu \left(B(x, c q_{X_N}) \right) \leq C$. Using this to further estimate (80), we obtain

$$\left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} - \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)} \right\|_{L^p(M;d\mu)} \leq C b^{L(d/p'-2\sigma)} q_{X_N}^{-(2\sigma+d/p)} \left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{L^p(M;d\mu)} \leq \gamma \left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{L^p(M;d\mu)},$$

if $L \geq \frac{2\sigma+d/p}{2\sigma-d/p'} \log_b(C q_X^{-1})$ and if $q_0 > 0$ is chosen small enough. $\qquad \square$

5.1 Interpolation with truncated kernel

We now focus on the Hilbert space $\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})$ and suppose that $\sigma > d/2$. We are interested in the analysis of reconstruction methods based on truncated kernels. To this end, for given data $f(x_n)$ at some scattered locations $x_n \in X_N = \{x_1, \ldots, x_N\} \subset M$, we consider trial spaces of the form (cf. also (4))

$$\mathcal{L}_{X_N}^{(L)} := \text{span}\left\{\phi(b^{-L}\sqrt{\mathcal{D}})\boldsymbol{K}^{(\sigma)}(\cdot, x_n) : \ x_n \in X_N\right\} = \text{span}\{K^{(\sigma,L)}(x_1, \cdot), \ldots, K^{(\sigma,L)}(x_N, \cdot)\}. \quad (82)$$

Note that truncated kernels of the form $\phi(b^{-L+1}\sqrt{\mathcal{D}})\boldsymbol{K}^{(\sigma)}$ are in general only positive semi-definite, so that the existence of an interpolant from $\mathcal{L}_{X_N}^{(L)}$ to arbitrary given data is not obvious.

Remark 6 Note that $K^{(\sigma,L)}$ is the reproducing kernel of $\left(\Sigma_{[0,b^L]}, (\cdot, \cdot)_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}\right)$.

We follow the general lines of [18] and first derive a lower bound on the truncation index L (depending on the point set X_N) such that, for arbitrary values, there always exists an interpolant from $\mathcal{L}_{X_N}^{(L)}$. The following lemma corresponds to [18, Proposition 3.3], but in contrast to the situation considered there, we do not exploit orthogonality properties here. We will use the notation introduced in (76).

Lemma 9 *Suppose that $b \geq \tilde{b}^{4/3}$. There are constants $C > 0$, $q_0 > 0$ and $\kappa > 1$ such that, for all N, $L \in \mathbb{N}$, all $X_N = \{x_1, \ldots, x_N\} \subset M$ with $q_{X_N} \leq q_0$ and $L \geq \log_b(Cq_{X_N}^{-1})$, and for all $\mathbf{a} \in \mathbb{R}^N$, we have*

$$\|z\|_{(\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D}))'} \leq \kappa \left\|z|_{\Sigma_{[0,b^L]}^2}\right\|_{\left(\Sigma_{[0,b^L]}^2\right)'}, \qquad \text{for } z := \sum_{n=1}^{N} a_n \delta_{x_n}, \quad (83)$$

where again $\|\cdot\|_{(\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D}))'}$ denotes the dual norm.

Proof By Riesz' representation theorem, we have $\|z\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})'} = \left\|\mathcal{I}_{\sigma;\mathbf{a};X_N}\right\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}$ and furthermore $\left\|z|_{\Sigma_{[0,b^L]}^2}\right\|_{\left(\Sigma_{[0,b^L]}^2\right)'} = \left\|\mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)}\right\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}$. To estimate the Besov-norms, we employ a smoothness shift in the spirit of [8]. Precisely, we have (see the proof of Proposition 2)

$$\left\|\mathcal{I}_{\sigma;\mathbf{a};X_N}\right\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}^2 = \sum_{n,m=1}^{N} a_n a_m \left(\boldsymbol{k}^{(\sigma)}(\sqrt{\mathcal{D}}) \sum_{\ell=0}^{\infty} \Delta_\ell \boldsymbol{\Phi}(\cdot, x_n), \ \sum_{\ell=0}^{\infty} \Delta_\ell \boldsymbol{\Phi}(\cdot, x_m)\right)_{L^2(M;d\mu)}$$

$$= \left\|(\boldsymbol{k}^{(\sigma)}(\sqrt{\mathcal{D}}))^{1/2} \sum_{\ell=0}^{\infty} \Delta_\ell \sum_{n=1}^{N} a_n \boldsymbol{\Phi}(\cdot, x_n)\right\|_{L^2(M;d\mu)}^2. \quad (84)$$

Note that we have by (34)

$$\frac{1}{2\sqrt{2}}(u^{\sigma} + 1) \leq (\boldsymbol{k}^{(\sigma)})^{-1/2}(u) \leq b^{\sigma}(u^{\sigma} + 1) \qquad \text{for all } u \in [0, \infty),$$

and thus there are constants $c_\sigma, C_\sigma > 0$ such that $c_\sigma \boldsymbol{k}^{(\sigma/2)}(u) \leq (\boldsymbol{k}^{(\sigma)}(u))^{1/2} \leq C_\sigma \boldsymbol{k}^{(\sigma/2)}(u)$ for all $u \in [0, \infty)$. Thus, we can further estimate the left hand side in (84)

$$\left\|\mathcal{I}_{\sigma;\mathbf{a};X_N}\right\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}^2 \sim \left\|\boldsymbol{k}^{(\sigma/2)}(\sqrt{\mathcal{D}}) \sum_{\ell=0}^{\infty} \Delta_\ell \sum_{n=1}^{N} a_n \boldsymbol{\Phi}(x_n, \cdot)\right\|_{L^2(M;d\mu)}^2 = \left\|\mathcal{I}_{\sigma/2;\mathbf{a};X_N}\right\|_{L^2(M;d\mu)}^2, \quad (85)$$

where the notation $\sim$ indicates that the two norms are equivalent with norm equivalence constants that depend only on σ and the parameters of the space. Similarly,

$$\left\| \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)} \right\|_{\mathcal{B}_{2,2}^{\sigma}(M;d\mu)} \sim \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N}^{(L)} \right\|_{L^2(M;d\mu)}.$$

Therefore, to prove (83), it suffices to show that there is a constant $\tilde{\kappa} > 1$ such that

$$\left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)} \leq \tilde{\kappa} \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N}^{(L)} \right\|_{L^2(M;d\mu)}. \tag{86}$$

By Lemma 8, we have (note that $2\sigma/2 > d/2$ and $L \geq \log_b(c' q_{X_N}^{-1})$)

$$\begin{aligned}
\left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)} &\leq \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N} - \mathcal{I}_{\sigma/2;\mathbf{a};X_N}^{(L)} \right\|_{L^2(M;d\mu)} + \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N}^{(L)} \right\|_{L^2(M;d\mu)} \\
&\leq \gamma \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)} + \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N}^{(L)} \right\|_{L^2(M;d\mu)},
\end{aligned}$$

with $\gamma \in (0,1)$ from (77), and thus with $\tilde{\kappa} := \frac{1}{1-\gamma}$

$$\left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)} \leq \tilde{\kappa} \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N}^{(L)} \right\|_{L^2(M;d\mu)},$$

i.e., we obtain (86). This concludes the proof. □

Following [18], we now invoke an abstract result from [37, Proposition 3.1].

Proposition 5 *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'*

$$\left\| z' \right\|_{\mathcal{Y}'} \leq \hat{\gamma} \left\| z'|_{\mathcal{V}} \right\|_{\mathcal{V}'}, \tag{87}$$

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

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

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

We now apply Proposition 5 with

$$\mathcal{Y} = \mathcal{B}_{2,2}^{\sigma}(M; \mathcal{D}), \quad \mathcal{V} = \Sigma_{[0,b^L]}^2 \quad \text{and} \quad \mathcal{Z}' := \{\delta_{x_n} : x_n \in X_N\} \tag{88}$$

and use Lemma 9 to obtain the following result.

Theorem 7 *Suppose that $b \geq \tilde{b}^{4/3}$. There are $q_0 > 0$ and $c' > 0$ such that, for every N, $L \in \mathbb{N}$, $X_N = \{x_1, \ldots, x_N\} \subset M$ with $q_{X_N} \leq q_0$ and $L \geq \log_b(c' q_{X_N}^{-1})$, and for every $f \in \mathcal{B}_{2,2}^{\sigma}(M; \mathcal{D})$, there is an $f_L \in \Sigma_{[0,b^L]}^2$ satisfying*

$$f(x_n) = f_L(x_n) \quad \text{for all } x_n \in X_N, \quad \text{and} \quad \|f - f_L\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})} \leq (1 + 2\gamma) \operatorname{dist}_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}(f, \Sigma_{[0,b^L]}^2) \tag{89}$$

Note that the result states that there always exists an interpolant which is also a quasi-optimal approximant. We can then follow the lines of [18, Theorem 4.1] and obtain that under the assumptions of Theorem 7, there exists a quasi-optimal interpolant from $\mathcal{L}_{X_N}^{(L)}$. Precisely, we have the following result.

Proposition 6 *Suppose that $b \geq \tilde{b}^{4/3}$. There exists $C > 0$ with the following property: Let $q_0 > 0$ and $c' > 0$ be as in Theorem 7. Then, for every $N, L \in \mathbb{N}$, $X_N = \{x_1, \ldots, x_N\} \subset M$ with $q_{X_N} \leq q_0$ and $L \geq \log_b(c' q_{X_N}^{-1})$, and for every $f \in \mathcal{B}_{2,2}^\sigma(M; \mathcal{D})$ there is $\mathbf{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^N$ such that $\mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)}$ satisfies*

$$\mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)}(x_n) = f(x_n) \quad \text{for all } x_n \in X_N \qquad \text{and} \qquad \left\| f - \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)} \right\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} \leq C \|f\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})}.$$

Proof The proof follows as in [18, Theorem 4.1], and we recall it only for completeness. By Theorem 7, there exists $f_L \in \Sigma_{[0,b^L]}^2$ such that (89) holds. In particular, we can view the data $f(x_n)$ as generated by f_L. Thus, since $K^{(L,\sigma)}$ is the reproducing kernel of $\Sigma_{[0,b^L]}^2$, there exists a kernel based interpolant $\mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)}$ such that

$$\mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)}(x_n) = f_L(x_n) = f(x_n) \quad \text{for all } x_n \in X_N, \text{ and} \quad \left\| f_L - \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)} \right\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} \leq \|f_L\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})}.$$

The assertion then follows by triangle inequality since $\|f - \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)}\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} \leq \|f - f_L\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} + \|f_L - \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)}\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})}$ and $\|f_L - \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)}\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} \leq \|f_L\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} \leq \|f - f_L\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} + \|f\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})}$. □

5.2 Stability of kernel based methods

In this section we give two important stability properties for the approximation with trial spaces $\mathcal{L}_{X_N}$. Both results can be seen as corollaries to Lemma 8 but are, in our opinion, interesting in their own. The first one is a Bernstein estimate on the finite dimensional trial space $\mathcal{L}_{X_N}$. Such estimates are crucial to the analysis of various kinds of unsymmetric reconstruction methods (see, e.g., [45]). We continue to use the notation introduced in (76).

Proposition 7 *Suppose that $b \geq \tilde{b}^{4/3}$. There are constants $q_0 > 0$ and $C > 0$ such that, for all discrete point sets $X_N = \{x_1, \ldots, x_N\} \subset M$ with $q_{X_N} \leq q_0$ and all $\mathbf{a} \in \mathbb{R}^N$, we have*

$$\left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} \leq C q_{X_N}^{-\sigma} \left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)}.$$

Proof We choose $L \geq \log_b(c' q_{X_N}^{-1})$. From (34) and (86), it follows that

$$\left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} \sim \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)} \leq C \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N}^{(L)} \right\|_{L^2(M;d\mu)}$$

$$\sim \left\| (\mathrm{Id} + \mathcal{D}^{\sigma/2}) \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)} \right\|_{L^2(M;d\mu)}. \tag{90}$$

On $\Sigma_{[0,b^L]}^2$, there is an inverse estimate from [10, Theorem 3.13]. Precisely, it holds

$$\left\| (\mathrm{Id} + \mathcal{D}^{\sigma/2}) f \right\|_{L^2(M;d\mu)} \leq C b^{\sigma L} \|f\|_{L^2(M;d\mu)} \qquad \text{for all } f \in \Sigma_{[0,b^L]}^2. \tag{91}$$

Since $\mathcal{I}_{\sigma;\mathbf{a},X_N}^{(L)} \in \Sigma_{[0,b^L]}^2$ we obtain, inserting (91) into (90)

$$\left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})} \leq C b^{\sigma L} \left\| \mathcal{I}_{\sigma;\mathbf{a};X_N}^{(L)} \right\|_{L^2(M;d\mu)} \leq C b^{\sigma L} \left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)}$$

$$\leq C q_{X_N}^{-\sigma} \left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)}.$$

This concludes the proof. □

The second result concerns lower bounds on the smallest eigenvalue of Gramian matrices $K_{X_N,X_N}^{(\sigma)} = \left(K^{(\sigma)}(x_n, x_m) \right)_{n,m=1,\ldots,N}$. Such bounds are of importance in practical considerations since the condition number of the Gramian matrix is mainly controlled by the smallest eigenvalue. Hence lower bounds on the smallest eigenvalue imply upper bounds on the condition number of the Gramian $K_{X_N,X_N}^{(\sigma)}$.

Proposition 8 *Let $\sigma > d/2$. There are constants $q_0 > 0$ and $C > 0$ such that for all sets $X_N = \{x_1, \ldots, x_N\} \subset M$ with $q_{X_N} \leq q_0$, we have*

$$\lambda_{\min} \left(K_{X_N,X_N}^{(\sigma)} \right) \geq C q_{X_N}^{2\sigma+d}. \tag{92}$$

If $\mu(B(y,r)) \leq Cr^d$ for all $y \in M$ and all $r > 0$, then

$$\lambda_{\min} \left(K_{X_N,X_N}^{(\sigma)} \right) \geq C q_{X_N}^{2\sigma-d}. \tag{93}$$

Proof We closely follow the lines of [8, Theorem 4.5]. Note that we get from (85) that we have for all $\mathbf{a} \in \mathbb{R}^N$

$$\mathbf{a}^T \left(K_{X_N,X_N}^{(\sigma)} \right) \mathbf{a} = \left\| \mathcal{I}_{\sigma;\mathbf{a};X_N} \right\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}^2 \geq C \left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)}^2.$$

Then, by Lemmata 4 and 7 with $p = 2$ and $\ell \sim \log(c' q_{X_N}^{-1})$, we have using $Cr^d \leq \mu B(y,r) \leq C$

$$\begin{aligned}
\left\| \mathcal{I}_{\sigma/2;\mathbf{a};X_N} \right\|_{L^2(M;d\mu)}^2 &\geq Cb^{-2\sigma\ell} \left\| \sum_{n=1}^N a_n \mathbf{S}^{(\ell;(\sigma/2))}(x_n, \cdot) \right\|_{L^2(M;d\mu)}^2 \\
&\geq \|\mathbf{a}\|_{\ell^2(N)}^2 \, b^{-2\sigma\ell} \left(\sup_{x \in M} \max_{1 \leq n \leq N} \frac{\mu\left(B(x_n, b^{-\ell})\right)}{\mu\left(B(x, b^{-\ell})\right)^{1/2}} \right)^{-2} \\
&\geq \|\mathbf{a}\|_{\ell^2(N)}^2 \, b^{-\ell(2\sigma+d)} \geq C q_{X_N}^{2\sigma+d} \|a\|_{\ell^2(N)}^2
\end{aligned} \tag{94}$$

This concludes the proof since

$$\lambda_{\min} \left(K_{X_N,X_N}^{(\sigma)} \right) = \min_{\|\mathbf{a}\|_{\ell^2(N)}=1} \mathbf{a}^T \left(K_{X_N,X_N}^{(\sigma)} \right) \mathbf{a} \geq C q_{X_N}^{2\sigma+d}.$$

The proof of (93) works analogously, using in (94) the bound

$$\sup_{x \in M} \max_{1 \leq n \leq N} \frac{\mu\left(B(x_n, b^{-\ell})\right)}{\mu\left(B(x, b^{-\ell})\right)^{1/2}} \sim b^{-\ell d/2}.$$

□

5.3 Relation to complexity estimates in statistical learning theory

At this point, let us recall the concept of a covering number from statistical learning theory [11, Definition 3.1].

Definition 3 Let M be a metric space and $\eta > 0$. The η-covering number $\mathcal{N}(M, \eta)$ is defined as

$$\mathcal{N}(M, \eta) = \min \left\{ n \in \mathbb{N} \cup \{\infty\} : \text{ there exists } Y_n = \{y_1, \dots, y_n\} \subset M \text{ such that } \bigcup_{j=1}^{n} B(y_j, \eta) \supset M \right\}. \tag{95}$$

For a reproducing kernel Hilbert space $\mathcal{H}_K(M)$, we denote the norm-balls by

$$B_K(R) := \{f \in \mathcal{H}_K(M) : \|f\|_{\mathcal{H}_K} \leq R\} \tag{96}$$

and we denote the embedding $j_K : \mathcal{H}_K(M) \hookrightarrow C(M)$. There is a result [62, Theorem 1] which holds for M being a compact metric space and μ a Borel measure on it and which states that for $0 \leq \eta \leq \frac{R}{2}$

$$\ln \left(\mathcal{N} \left(\overline{j_K(B_K(R))}, \eta \right) \right) \leq \#Y_n \ln \left(8 \|K\|_{L^\infty(M \times M)}^{\frac{3}{2}} \#Y_n \left\| K_{Y_n, Y_n}^{-1} \right\|_{\ell^2(N) \to \ell^2(N)} \frac{R}{\eta} \right) \tag{97}$$

where $Y_n \subset M$ with $n \in \mathbb{N}$ is such that

$$\left\| P_{K; Y_n} \right\|_{L^\infty(M)} \leq \frac{\eta}{2R}. \tag{98}$$

Here, the quantity

$$P_{K; Y_n}(x) := \min_{a(x) \in \mathbb{R}^n} \left\| \delta_x - \sum_{\xi \in Y_n} a_\xi(x) \delta_\xi \right\|_{\mathcal{H}_K'} \tag{99}$$

is the so called power function [49]. If unique interpolation with the kernel is possible, the optimal functions $a_y(x)$ can be computed and are given by the cardinal functions. Since the quantity (99) measures the interpolation error, we can invoke our sampling inequality to obtain with Theorem 6, $\mathcal{H}_K = \mathcal{B}_{2,2}^\sigma(M; \mathcal{D})$ and $K = K^{(\sigma)}$

$$\left\| P_{K; Y_n} \right\|_{L^\infty(M)} \leq C h_{Y_n, M}^{\sigma - d/2}. \tag{100}$$

Also by (92), we have

$$\left\| K_{Y_n, Y_n}^{-1} \right\|_{\ell^2(N) \to \ell^2(N)} \leq \lambda_{\min}^{-1} \left(K_{Y_n, Y_n}^{(\sigma)} \right) \leq C q_{Y_n}^{-2\sigma - d}$$

We assume quasi-uniformity, i.e., there is a constant $c < 1$ such that $c h_{Y_n, M} \leq q_{Y_n} \leq h_{Y_n, M}$ holds. Here, we use the notation $a \cong b$ to indicate that there are constants $c, C > 0$ such that $ca \leq b \leq Ca$ and similarly for $\gtrsim$. For any quasi-uniform set of points Y_n such that $h_{Y_n, M}^{\sigma - d/2} \cong \frac{\eta}{2R}$, we have

$$\#Y_n \left\| K_{Y_n, Y_n}^{-1} \right\|_{\ell^2(N) \to \ell^2(N)}^{-1/2} \lesssim h_{Y_n, M}^{-d} \left\| K_{Y_n, Y_n}^{-1} \right\|_{\ell^2(N) \to \ell^2(N)}^{-1/2} \lesssim h_{Y_n, M}^{\sigma - d/2} \cong \frac{\eta}{2R},$$

using $\#Y_n \cong h_{Y_n, M}^{-d}$. Hence, we get

$$\frac{R}{\eta} \cong h_{Y_n, M}^{-\sigma + d/2} \lesssim (\#Y_n)^{-1} \left\| K_{Y_n, Y_n}^{-1} \right\|_{\ell^2(N) \to \ell^2(N)}^{1/2}$$

Inserting this into the bound (97), we arrive at[2]

$$\ln\left(\mathcal{N}\left(\overline{j_K(B_K(R))},\eta\right)\right) \lesssim h_{Y_n,M}^{-d}\ln\left(h_{Y_n,M}^{\frac{3}{2}(-2\sigma-d)}\right) \lesssim \left(\frac{R}{\eta}\right)^{\frac{2d}{2\sigma-d}}\ln\left(\frac{R}{\eta}\right). \tag{101}$$

6 Regularized reconstruction

From now on, we assume that $\mu(M) < \infty$. Since in this case the spectral spaces $\Sigma_{[0,R]}^p$ are equivalent for all $1 \leq p \leq \infty$, we skip the parameter p. We consider the reconstruction of functions $f \in \mathcal{B}_{2,2}^\sigma(M;\mathcal{D})$, $\sigma > d/2$, from given data $y_n \approx f(x_n)$ at discrete points $X_N = \{x_1,\ldots,x_N\} \subset M$. An approximant $\hat{g}^{(\mathbf{y};\alpha;X_N)}$ is given as a minimizer of the smoothing-type functional from (2), i.e.,

$$J_{\mathbf{y};\alpha;X_N} : \Sigma_{[0,b^L]} \to \mathbb{R}, \quad J_{\mathbf{y};\alpha;X_N}(g) = \sum_{n=1}^N \left(g(x_n) - y_n\right)^2 + \alpha\|g\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})}^2 \tag{102}$$

with a regularization parameter $\alpha \geq 0$. For the ease of notation, we restrict ourselves to the case of exact data

$$y_n = f(x_n) \quad \text{for all } n = 1,\ldots N. \tag{103}$$

In this section, we follow the general lines of [45, 46, 59] but take into account also the difficulty that the (truncated) kernel might be available only approximately. If the given data is corrupted by some additive noise, one can use the same machinery outlined below to derive error estimates which are explicit in terms of the noise (see [46]). Also, we restrict ourselves to the spline-smoothing regularization but point out that various kinds of additional regularization terms can be treated analogously (see [46, 45]). For some results in this direction using weak data we refer to [45].
We shall show by a representer theorem that for L large enough, a minimizer $\hat{g}^{(\mathbf{y};\alpha;X_N)}$ of (2) can be computed by solving a finite linear system which is built of translates of the truncated kernel. Furthermore, a deterministic error analysis will be derived by means of the sampling inequalities proven in Section 4. We first shall consider the (special) case of exact spectral projections, the case of numerically computed, approximated eigenfunctions of $\mathcal{D}$ will be dealt with afterwards.

We consider the spectral spaces $\Sigma_{[0,b^L]}$ (see (29)) as reproducing kernel Hilbert spaces. Precisely, we denote the ordered eigenvalues of $\mathcal{D}$ by $0 < \lambda_1 \leq \lambda_2 \leq \ldots$ and the associated orthonormal eigenfunctions $\varphi_m^{(\ell)}$, $m = 1,\ldots,\dim V_\ell$, $\ell = 1,2,\ldots$, where V_ℓ denotes the eigenspace to λ_ℓ. On $\Sigma_{[0,b^L]}$, we have the norm (recall that $\operatorname{supp}\psi(b^{-\ell}\cdot) \subset [b^{\ell-1},b^\ell]$, see Definition 2)

$$\|f\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})}^2 = \sum_{\ell=0}^L b^{2\sigma\ell} \sum_{\ell'=0}^\infty \psi_\ell^2(\sqrt{\lambda_\ell}) \sum_{i=1}^{\dim(V_{\ell'})} \left|(f,\varphi_i^{(\ell')})_{L^2(M;d\mu)}\right|^2, \quad f \in \Sigma_{[0,b^L]}.$$

For $\sigma > d/2$, the reproducing kernel of $\Sigma_{[0,b^L]}$ can be expressed as

$$K^{(\sigma,L)}(x,y) = \sum_{\ell=0}^L b^{-2\sigma\ell}\boldsymbol{S}^{(\ell;\sigma)}(x,y) = \sum_{\ell=0}^\infty \left(\boldsymbol{k}^{(\sigma)}(\sqrt{\lambda_\ell})\right)^{-1}\phi(b^{-L}\sqrt{\lambda_\ell}) \sum_{i=1}^{\dim(V_\ell)} \varphi_i^{(\ell)}(x)\varphi_i^{(\ell)}(y).$$

We are now in the position to give a representer theorem and a priori error estimates for minimizers of (2).

---
[2] There are also lower bounds for the covering number cf. [11, Theorem 5.21].

Theorem 8 *Let $\sigma > d/2$. There are constants c', C, $h_0 > 0$ such that, for all $X_N = \{x_1, \ldots, x_N\} \subset M$ with $h_{X_N, M} \leq h_0$, all $L \in \mathbb{N}$ with $L \geq \log_b\left(c' q_{X_N}^{-1}\right)$ and all $f \in \mathcal{B}_{2,2}^\sigma(M; \mathcal{D})$, there exists a minimizer $\hat{g}^{(\mathbf{y}; \alpha; X_N)}$ of (2) with data (103) such that*

$$\hat{g}^{(\mathbf{y}; \alpha; X_N)} \in span\{K^{(\sigma, L)}(x_1, \cdot), \ldots, K^{(\sigma, L)}(x_N, \cdot)\}. \tag{104}$$

Furthermore, there is the a priori estimate

$$\left\| f - \hat{g}^{(\mathbf{y}; \alpha; X_N)} \right\|_{L^\infty(M; d\mu)} \leq C \left(h_{X_N, M}^{\sigma - d/2} + \sqrt{\alpha} \right) \|f\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})}. \tag{105}$$

Proof By Theorem 7, there exists $f_L \in \Sigma_{[0, b^L]}$ such that

$$f_L(x_n) = f(x_n) \qquad \text{for all } x_n \in X_N, \qquad \text{and} \qquad \|f_L\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})} \leq C \|f\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})}.$$

Thus, we can view the data $y_n = f(x_n) = f_L(x_n)$ to be generated by a function $f_L \in \Sigma_{[0, b^L]}$. By a classical representer theorem (see, e.g., [50]) applied to the reproducing kernel Hilbert space $\Sigma_{[0, b^L]}$, there is a solution $\hat{g}^{(\mathbf{y}; \alpha; X_N)} \in \text{span}\{K^{(\sigma, L)}(x_1, \cdot), \ldots, K^{(\sigma, L)}(x_N, \cdot)\}$ to the optimization problem (102). We furthermore get the following stability and consistency estimates: First, we have

$$\alpha \left\| \hat{g}^{(\mathbf{y}; \alpha; X_N)} \right\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})}^2 \leq J_{\mathbf{y}; \alpha; X_N}\left(\hat{g}^{(\mathbf{y}; \alpha; X_N)} \right) \leq J_{\mathbf{y}; \alpha; X_N}(f_L) = \alpha \|f_L\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})}^2 \leq \alpha \|f\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})}^2,$$

which shows that

$$\|f - \hat{g}^{(\mathbf{y}; \alpha; X_N)}\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})} \leq \|f\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})} + \|\hat{g}^{(\mathbf{y}; \alpha; X_N)}\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})} \leq 2\|f\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})}. \tag{106}$$

Second, we get similarly

$$\max_{1 \leq n \leq N} \left| (f - \hat{g}^{(\mathbf{y}; \alpha; X_N)})(x_n) \right|^2 \leq \sum_{n=1}^N \left| f(x_n) - \hat{g}^{(y; \alpha; X_N)}(x_n) \right|^2 \leq J_{\mathbf{y}; \alpha; X_N}\left(\hat{g}^{(\mathbf{y}; \alpha; X_N)} \right)$$

$$\leq J_{\mathbf{y}; \alpha; X_N}(f_L) = \alpha \|f_L\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})}^2 \leq \alpha \|f\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})}^2.$$

Therefore, we obtain by Theorem 6 the estimate

$$\left\| f - \hat{g}^{(\mathbf{y}; \alpha; X_N)} \right\|_{L^\infty(M; d\mu)} \leq C \left(h_{X_N, M}^{\sigma - d/2} \left\| f - \hat{g}^{(\mathbf{y}; \alpha; X_N)} \right\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})} + \left\| (f - \hat{g}^{(\mathbf{y}; \alpha; X_N)})|_{X_N} \right\|_{\ell^\infty(X_N)} \right)$$

$$\leq C \left(h_{X_N, M}^{\sigma - \frac{d}{2}} + \sqrt{\alpha} \right) \|f\|_{\mathcal{B}_{2,2}^\sigma(M; \mathcal{D})}.$$

$$\square$$

The fact that the solution to a possibly infinite optimization problem is contained in a finite-dimensional linear space is usually called representer theorem. Thus (104) is indeed a representer theorem and hence a minimizer of (2) can be computed by solving the linear system with $\mathbf{y} = f|_{X_N} \in \mathbb{R}^N$

$$\left(K_{X_N, X_N}^{(\sigma, L)} + \alpha \, \text{Id} \right) \hat{\mathbf{c}} = \mathbf{y} \tag{107}$$

for the coefficient vector $\hat{\mathbf{c}}$ of $\hat{g}^{(\mathbf{y}; \alpha; X_N)}(\cdot) = \sum_{n=1}^N \hat{c}_n K^{(\sigma, L)}(\cdot, x_n)$. Here, an entry of the Gramian matrix $K_{X_N, X_N}^{(\sigma, L)} \in \mathbb{R}^{N \times N}$ is given by

$$\left(K_{X_N, X_N}^{(\sigma, L)} \right)_{n, m} = K^{(\sigma, L)}(x_n, x_m) = \sum_{\ell=0}^\infty \left(\mathbf{k}^{(\sigma)}(\sqrt{\lambda_\ell}) \right)^{-1} \sum_{\ell'=0}^L \Delta_{b; \ell'} \phi(\sqrt{\lambda_\ell}) \sum_{i=1}^{\dim(V_\ell)} \varphi_i^{(\ell)}(x_n) \varphi_i^{(\ell)}(x_m).$$

To set up this matrix, one needs to compute the values $\varphi_m^{(\ell)}(x_i)$ and λ_ℓ.

Unfortunately, the eigenfunctions $\varphi_m^{(\ell)}$ and eigenvalues λ_ℓ of $\mathcal{D}$ are typically not available in analytically closed form, but can only be numerically computed up to some error. We therefore assume that there are level dependent errors $\varepsilon_\ell \geq 0$ such that a corrupted kernel $K_\varepsilon^{(\sigma,L)}$ is at hand, for which

$$\left| (K_\varepsilon^{(\sigma,L)} - K^{(\sigma,L)})(x_n, x_m) \right| \leq \sum_{\ell=0}^{\infty} \left(\boldsymbol{k}^{(\sigma)}(\sqrt{\lambda_\ell}) \right)^{-1} \phi\left(b^{-L}\sqrt{\lambda_\ell} \right) \varepsilon_\ell^2 \dim(V_\ell),$$

for all $n, m = 1, \ldots, N$, where some prescribed accuracy $\varepsilon_{\max}$ is respected, i.e.,

$$\sum_{\ell=0}^{\infty} \left(\boldsymbol{k}^{(\sigma)}(\sqrt{\lambda_\ell}) \right)^{-1} \phi\left(b^{-L}\sqrt{\lambda_\ell} \right) \varepsilon_\ell^2 \dim(V_\ell) \leq \varepsilon_{\max}. \tag{108}$$

Thus, instead of working with the practically unavailable optimal system (107), we compute an approximant

$$\hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)}(\cdot) = \sum_{n=1}^{N} \hat{c}_n^{(\varepsilon)}(f) K_\varepsilon^{(\sigma,L)}(\cdot, x_n) \tag{109}$$

with coefficients $\hat{\mathbf{c}}^\varepsilon \in \mathbb{R}^N$ given as solution of

$$\left(K_{X_N,X_N}^{(\sigma,L,\varepsilon)} + \alpha \,\mathrm{Id} \right) \hat{\mathbf{c}}^{(\varepsilon)} = \left(K_{X_N,X_N}^{(\sigma,L)} + R_\varepsilon + \alpha \,\mathrm{Id} \right) \hat{\mathbf{c}}^\varepsilon = \mathbf{y} \qquad \text{with} \quad |(R_\varepsilon)_{n,m}| \leq \varepsilon_{\max}, \tag{110}$$

where $(K_{X_N,X_N}^{(\sigma,L,\varepsilon)})_{n,m} := K_\varepsilon^{(\sigma,L)}(x_n, x_m)$ for $n, m = 1 \ldots N$ and with $\mathbf{y} = f|_{X_N} \in \mathbb{R}^N$. We obtain the following deterministic error estimate for this approximant.

Theorem 9 *There is a constant $C > 0$ such that, for all sets $X_N = \{x_1, \ldots, x_N\} \subset M$, all $L \geq \log_b\left(Cq_{X_N}^{-1} \right)$, all $\alpha > N\varepsilon_{\max}$ and every $f \in \mathcal{B}_{2,2}^\sigma(M;\mathcal{D})$, we have for $\hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)}$ given by (109) and (110) with data $\mathbf{y} \in \mathbb{R}^N$ given by (103)*

$$\left\| f - \hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)} \right\|_{L^\infty(M;d\mu)} \leq C \left(h_{X_N,M}^{\sigma-d/2} + \sqrt{\alpha} \right) \|f\|_{\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})}$$

$$+ \varepsilon_{\max} \left(N\alpha^{-1} + \tilde{C}\frac{N^2\alpha^{-2}}{1 - N\varepsilon_{\max}\alpha^{-1}} \right) \frac{1}{\sqrt{N}} \|f|_{X_N}\|_{\ell^2(X_N)}. \tag{111}$$

Proof In view of Theorem 8, it suffices to estimate the error $\left\| \hat{g}^{(\mathbf{y};\alpha;X_N)} - \hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)} \right\|_{L^\infty(M;d\mu)}$. For that, we proceed in two steps and first estimate the error in the coefficients and deduce subsequently the error in the function values. We employ the following classical result of perturbation theory (see, e.g., [60]). Suppose that the matrix $A \in \mathbb{R}^{N \times N}$ has full rank, and let $\delta A \in \mathbb{R}^{N \times N}$ be such that $\|A^{-1}\|_{\ell^2(N) \to \ell^2(N)} \|\delta A\|_{\ell^2(N) \to \ell^2(N)} < 1$. For a given vector $\mathbf{b} \in \mathbb{R}^N \setminus \{0\}$ let the vectors $\mathbf{z}, \delta\mathbf{z} \in \mathbb{R}^N$ be such that $A\mathbf{z} = \mathbf{b}$ and $(A + \delta A)(\mathbf{z} + \delta\mathbf{z}) = \mathbf{b}$. Then

$$\frac{\|\delta\mathbf{z}\|_{\ell^2(N)}}{\|\mathbf{z}\|_{\ell^2(N)}} \leq \frac{\|A^{-1}\|_{\ell^2(N) \to \ell^2(N)} \|\delta A\|_{\ell^2(N) \to \ell^2(N)}}{1 - \|A^{-1}\|_{\ell^2(N) \to \ell^2(N)} \|\delta A\|_{\ell^2(N) \to \ell^2(N)}}. \tag{112}$$

We apply this result with $A := K_{X_N,X_N}^{(\sigma,L)} + \alpha \,\mathrm{Id}$, $\delta A := R_\varepsilon$, $\delta\mathbf{z} = \hat{\mathbf{c}} - \hat{\mathbf{c}}^\epsilon$ and $\mathbf{z} = \hat{\mathbf{c}}$. If we denote by $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ the largest and smallest eigenvalues of a matrix A, respectively, then

$$\left\| \left(K_{X_N,X_N}^{(\sigma,L)} + \alpha \,\mathrm{Id} \right)^{-1} \right\|_{\ell^2(N) \to \ell^2(N)} \leq \lambda_{\max}\left(K_{X_N,X_N}^{(\sigma,L)} + \alpha \,\mathrm{Id} \right)^{-1} \leq \frac{1}{\lambda_{\min}(K_{X_N,X_N}^{(\sigma,L)}) + \alpha} \leq \alpha^{-1}.$$

Since furthermore $\|R_\varepsilon\|_{\ell^2(N)\to\ell^2(N)} \leq N\varepsilon_{\max}$, the condition $\|A^{-1}\|_{\ell^2(N)\to\ell^2(N)}\|\delta A\|_{\ell^2(N)\to\ell^2(N)} < 1$ is satisfied if $N\varepsilon_{\max} < \alpha$. Hence by (112) and the definition of $\hat{\mathbf{c}}$, we obtain

$$\left\|\hat{\mathbf{c}}^{(\varepsilon)} - \hat{\mathbf{c}}\right\|_{\ell^2(N)} = \|\delta\mathbf{z}\|_{\ell^2(N)} \leq \frac{N\varepsilon_{\max}\alpha^{-1}}{1 - N\varepsilon_{\max}\alpha^{-1}}\|\hat{\mathbf{c}}\|_{\ell^2(N)}$$

$$\leq \frac{N\varepsilon_{\max}\alpha^{-1}}{1 - N\varepsilon_{\max}\alpha^{-1}}\left\|\left(K_{X_N,X_N}^{(\sigma,L)} + \alpha\,\mathrm{Id}\right)^{-1}f|_{X_N}\right\|_{\ell^2(X_N)} \leq \frac{N\varepsilon_{\max}\alpha^{-2}}{1 - N\varepsilon_{\max}\alpha^{-1}}\|f|_{X_N}\|_{\ell^2(X_N)}. \quad (113)$$

We now turn to the approximants and note that by the triangle inequality

$$\left\|\hat{g}^{(\mathbf{y};\alpha;X_N)} - \hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)}\right\|_{L^\infty(M;d\mu)} \leq \left\|\sum_{n=1}^N \left(\hat{c}_n^{(\varepsilon)} - \hat{c}_n\right)K_\varepsilon^{(\sigma,L)}(\cdot,x_n)\right\|_{L^\infty(M;d\mu)} +$$

$$+ \left\|\sum_{n=1}^N \hat{c}_n\left(K^{(\sigma,L)}(\cdot,x_n) - K_\varepsilon^{(\sigma,L)}(\cdot,x_n)\right)\right\|_{L^\infty(M;d\mu)}. \quad (114)$$

We estimate the two terms of the right hand side of (114) separately. First, by Hölder's inequality and (113) we have

$$\left\|\sum_{n=1}^N \left(\hat{c}_n^{(\varepsilon)} - \hat{c}_n\right)K_\varepsilon^{(\sigma,L)}(\cdot,x_n)\right\|_{L^\infty(M;d\mu)}$$

$$\leq \max_{1\leq n\leq N}\left\|K_\varepsilon^{(\sigma,L)}(\cdot,x_n)\right\|_{L^\infty(M;d\mu)}\sum_{n=1}^N\left|\hat{c}_n^{(\varepsilon)} - \hat{c}_n\right|$$

$$\leq \sqrt{N}\left\|\hat{\mathbf{c}}^{(\varepsilon)} - \hat{\mathbf{c}}\right\|_{\ell^2(N)}\max_{1\leq n\leq N}\left|K_\varepsilon^{(\sigma,L)}(x_n,x_n)\right| \leq C\frac{N^{\frac{3}{2}}\varepsilon_{\max}\alpha^{-2}}{1 - N\varepsilon_{\max}\alpha^{-1}}\|f|_{X_N}\|_{\ell^2(X_N)}, \quad (115)$$

and second,

$$\left\|\sum_{n=1}^N \hat{c}_n\left(K_\varepsilon^{(\sigma,L)}(\cdot,x_n) - K^{(\sigma,L)}(\cdot,x_n)\right)\right\|_{L^\infty(M;d\mu)}$$

$$\leq \sum_{n=1}^N|\hat{c}_n|\max_{1\leq n\leq N}\left\|K_\varepsilon^{(\sigma,L)}(\cdot,x_n) - K^{(\sigma,L)}(\cdot,x_n)\right\|_{L^\infty(M;d\mu)} \leq \sqrt{N}\|\hat{c}\|_{\ell^2}\varepsilon_{\max}$$

$$\leq \sqrt{N}\left\|\left(K_{X_N,X_N}^{(\sigma,L)} + \alpha\,\mathrm{Id}\right)^{-1}f|_{X_N}\right\|_{\ell^2(X_N)}\varepsilon_{\max} \leq \sqrt{N}\alpha^{-1}\varepsilon_{\max}\|f|_{X_N}\|_{\ell^2(X_N)}. \quad (116)$$

Putting together (114), (115) and (116) we obtain

$$\left\|\hat{g}^{(\mathbf{y};\alpha;X_N)} - \hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)}\right\|_{L^\infty(M;d\mu)} \leq \left(C\frac{N^{\frac{3}{2}}\varepsilon_{\max}\alpha^{-2}}{1 - N\varepsilon_{\max}\alpha^{-1}} + \sqrt{N}\alpha^{-1}\varepsilon_{\max}\right)\|f|_{X_N}\|_{\ell^2(X_N)},$$

and the assertion then follows by Theorem 8 since

$$\left\|f - \hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)}\right\|_{L^\infty(M;d\mu)} \leq \left\|\hat{g}^{(\mathbf{y};\alpha;X_N)} - \hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)}\right\|_{L^\infty(M;d\mu)} + \left\|f - \hat{g}^{(\mathbf{y};\alpha;X_N)}\right\|_{L^\infty(M;d\mu)}.$$

This concludes the proof. $\qquad\square$

Note that Theorem 9 provides an error estimate which is explicit in the problem parameters $\varepsilon_{\max}$ and α. We shall now outline how to choose these parameters in order to ensure asymptotic convergence[3].

Corollary 2 *Suppose that the assumptions of Theorem 9 are valid and suppose additionally that for $N \to \infty$ it holds that $1/\sqrt{N} \, \|f|_{X_N}\|_{\ell^2(X_N)} \to \|f\|_{L^2(M;d\mu)}$. If one chooses*

$$\alpha := h_{X_N,M}^{2(\sigma-d/2)} \qquad and \qquad \varepsilon_{\max} := \frac{h_{X_N,M}^{5(\sigma-d/2)}}{N^2} \frac{\|f\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}}{\|f\|_{L^2(M;\mathcal{D})}},$$

then

$$\left\| f - \hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)} \right\|_{L^\infty(M;d\mu)} \leq C h_{X_N,M}^{\sigma-d/2} \|f\|_{\mathcal{B}_{2,2}^{\sigma}(M;\mathcal{D})}.$$

6.1 Relation to error analysis for statistical regression theory

Here, we closely follow [51], see also [28] for recent applications of spherical needlet kernels in learning theory. The problem can be formulated as follows: We are given a probability measure $\mathbb{P}$ on $M \times \mathbb{R}$ where $M \subset \mathbb{R}^n$ is supposed to be a compact set and try to recover a function $f : M \to \mathbb{R}$ from its sampled values $(x_i, y_i) \in M \times \mathbb{R}$ for $1 \leq i \leq N$ and $x_i \neq x_j$ for $i \neq j$. Moreover, we assume K to be the reproducing kernel of $B_{2,2}^{\sigma}(M;\mathcal{D})$ for $\sigma > d/2$. We define the expressions [51, Eq.s 2.1 & 2.2]

$$\mathcal{E}(f) := \int_{M \times \mathbb{R}} (f(x) - y)^2 \, d\mathbb{P}(x,y), \quad \mathcal{E}_y(f) := \frac{1}{N} \sum_{i=1}^{N} (f(x_i) - y_i)^2,$$

which are called least squares error and empirical error. We denote their respective minimizers by

$$f^\star(x) := \int_{\mathbb{R}} y \, d\mathbb{P}(y|x)$$

$$\mathcal{I}_{y,X_N} := \arg \min_{f \in \overline{j_K(B_K(R))}} \mathcal{E}_y(f) = \sum_{i=1}^{N} y_i a_{x_i},$$

with the cardinal functions a_{x_i} as in the definition of the power function (99). The fact that $f^\star$ is indeed minimizing the least squares problem follows from [11, Prop. 1.8], where for any $f : M \to \mathbb{R}$ being square-integrable the decomposition

$$\mathcal{E}(f) = \int_M \left(f(x) - f^\star(x) \right)^2 d\mu(x) + \mathcal{E}(f^\star),$$

with μ being the marginal measure on M is shown. Furthermore, we define the target function from [51]

$$f^R := \arg \min_{f \in \overline{j_K(B_K(R))}} \mathcal{E}(f) = \inf_{f \in B_K(R)} \mathcal{E}(f)$$

with $B_K(R)$ from (96). We point out here, that both functions $f^\star$ and $\mathcal{I}_{y,X_N}$ are not at our disposal. The first one is the unknown global solution and the second one is not available since we cannot

---

[3] For practical consideration other parameter choices could be more useful. We do not give the details here, but leave those considerations to the reader since we work in a very general framework and hence do not have a model for the numerical costs for realizing $\varepsilon_{\max}$. In many specific application estimate for these cots are available and can be employed in an exhaustive cost-benefit discussion.

evaluate the kernel directly. We therefore use the function $\hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)}$ as an approximation for the latter. The aim is to give an upper bound of the error

$$\int_M \left(f^\star(x) - \hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)}(x)\right)^2 d\mu(x) \lesssim \int_M \left(f^\star(x) - \mathcal{I}_{y,X_N}(x)\right)^2 d\mu(x)$$
$$+ \int_M \left(\hat{g}_\varepsilon^{(\mathbf{y};\alpha;X_N)}(x) - \mathcal{I}_{y,X_N}(x)\right)^2 d\mu(x). \qquad (117)$$

The latter error is already bounded in Corollary 2 if M and μ satisfy all conditions given there. The first term is the usual error in statistical learning theory. It is decomposed into the sampling error and the approximation error, i.e.,

$$\int_M \left(f^\star(x) - \mathcal{I}_{y,X_N}(x)\right)^2 d\mu(x) = \mathcal{E}(\mathcal{I}_{y,X_N}) - \mathcal{E}(f^\star)$$
$$= \left(\mathcal{E}(\mathcal{I}_{y,X_N}) - \mathcal{E}(\mathcal{I}_{y,X_N}^R)\right) + \int_M \left(f^R(x) - f^\star(x)\right)^2 d\mu(x)$$

where the first summand, i.e., the content of the first bracket is called sampling error and the second term is called approximation error (see [51]). Under the assumption that there is a constant $V > 0$ such that $|f(x) - y| \leq V$, we have the bound [51, Eq. (2.6)]

$$\text{Prob}\left\{\left(\mathcal{E}(\mathcal{I}_{y,X_N}) - \mathcal{E}(\mathcal{I}_{y,X_N}^R)\right) \leq \eta\right\} \geq 1 - \mathcal{N}\left(\overline{j_K(B_K(R))}, \frac{\eta}{24V}\right) \exp\left(-\frac{N\eta}{288V^2}\right)$$
$$\gtrsim 1 - \exp\left(\left(24V\frac{R}{\eta}\right)^{\frac{2d}{2\sigma-d}} \ln\left(24V\frac{R}{\eta}\right)\right) \exp\left(-\frac{N\eta}{288V^2}\right)$$

for all $\eta > 0$, were we used (101) in the last step. For the approximation error, we can use the following result from interpolation theory, [51, Thm. 3.1]: We have that $L^2(M;d\mu)$ is a dense subspace of $B_{2,2}^\sigma(M;\mathcal{D})$ such that $\|f\|_{L^2(M;d\mu)} \leq c\|f\|_{B_{2,2}^\sigma(M;\mathcal{D})}$. For $0 < \theta < 1$ we assume $f \in \left(L^2(M;d\mu), B_{2,2}^\sigma(M;\mathcal{D})\right)_{\theta,\infty}$. Then we have [51, Thm. 3.1]

$$\inf_{g \in B_K(R)} \|f - g\|_{L^2(M;d\mu)} \leq R^{-\frac{\theta}{1-\theta}} \|f\|_{\left(L^2(M;d\mu), B_{2,2}^\sigma(M;\mathcal{D})\right)_{\theta,\infty}}$$

By the equivalence of the Besov spaces $\mathcal{B}_{2,2}^\sigma(M;\mathcal{D})$ to the Bessel potential spaces $\text{dom}(\text{Id} + \sqrt{\mathcal{D}})^{2\sigma}$ and [10, Proposition 6.2 & Theorem 3.16] we get $\left(L^2(M;d\mu), B_{2,2}^{\sigma/\theta}(M;\mathcal{D})\right)_{\theta,\infty} \approx B_{2,\infty}^\sigma(M;\mathcal{D})$. The final error bounds for (117) now stem from an optimization with respect to R, see [51] for an example of such calculations. For the sake of brevity, we do not give the lengthy computations here.

7 Concluding remarks

We derived an explicit representation of the reproducing kernel for Besov spaces in the abstract framework of metric measure spaces, see Proposition 2. As fundamental step towards a priori error estimates for reconstruction schemes in such Besov spaces we proved sampling inequalities, see Theorem 6. Such sampling inequalities quantify the observation that a reconstruction scheme with small residuals has small global error as long as it is stable in the Besov space norm. In order to design numerically feasible approximation schemes, we discussed the truncation of the infinite series representation of the kernel in Theorem 7. We gave an explicit condition to couple the truncation parameters to the discrete point set which guarantees well-posedness and quasi-optimality of the reconstruction process. Furthermore, sampling inequalities lead also to a priori error estimates for

regularized reconstruction schemes, see Theorem 9. The resulting error bounds are explicit in the regularization parameters, the discretization error and the modeling error, and therefore allow for a balancing these terms. Finally, we explained how the well-established machinery of error estimates for statistical learning can be applied here.

Appendix

In this appendix, we give an explicit bound on the constant $\tilde{b}$ of Remark 2 for the Euclidean space $\mathbb{R}^d$, where we closely follow the lines of the proof of the statement in [10, Lemma 3.19]. We recall the general strategy first, and then perform the necessary estimates in our setting. For measurable sets $\Omega \subset \mathbb{R}^d$, we set $|\Omega| := \mu(\Omega)$. In this case, (8) and (9) hold with $\beta = d$, i.e.,

$$0 < |B(x, 2t)| = 2^d |B(x, t)| \qquad \text{for all} \quad x \in \mathbb{R}^d, \qquad \text{and all} \quad t > 0.$$

Furthermore,

$$|B(x, \sqrt{t})| = \frac{(\pi t)^{d/2}}{\Gamma(\frac{d}{2} - 1)}, \qquad \text{and} \qquad p_t(x, y) = \frac{1}{(4\pi t)^{d/2}} \exp\left(-\|x - y\|^2 / (4t)\right).$$

Consequently,

$$p_t(x, x) = (4\pi t)^{-d/2} = \frac{1}{2^d \Gamma(d/2 + 1)} |B(x, \sqrt{t})|^{-1},$$

and in particular, if $\mathbf{1}$ denotes the characteristic function,

$$\mathbf{1}_{[0,\tau]}(\sqrt{\mathcal{D}})(x, x) \leq e \cdot p_{\tau^{-2}}(x, x) = \frac{e}{2^d \Gamma(d/2 + 1)} |B(x, \tau^{-1})|^{-1} \qquad \text{for all} \quad \tau > 0. \qquad (118)$$

It is shown in [10, Lemma 3.19] that for $\tau > 0$ and $r \in \mathbb{N}$, we can set $\tau\sqrt{t} = 2^r$ such that

$$\frac{2^{-rd}}{|B(x, \tau^{-1})|} \left(c' - 2^d c_4 \sum_{k \geq r} \exp(-2^{2k}) 2^{kd}\right) \leq \mathbf{1}_{[0,\tau]}(x, x),$$

holds, where the constants $c_4 := \frac{e}{2^d \Gamma(d/2+1)} =: ec'$ can be obtained from (118). Hence, to make the lower bound positive, we need to choose $r \in \mathbb{N}$ large enough such that

$$\exp(-1) > \sum_{k \geq r} \exp(-2^{2k}) 2^{kd}. \qquad (119)$$

Once we have an appropriate $r \in \mathbb{N}$ at hand, we follow the argument in [10] and set (see [10, (3.44)])

$$c_3 := \frac{2^{-r}}{2^d \Gamma(d/2 + 1)} \left(1 - \sum_{k \geq r} \exp(-2^{2k}) 2^{(k+1)d}\right) > 0,$$

and choose a $\ell > 0$ large enough such that

$$0 < c_3 2^{d\ell} - c_4. \qquad (120)$$

Then, following the proof of [10, Lemma 3.19], we may set $\tilde{b} := 2^\ell$.

Lemma 10 *If we choose $r(d) \in \mathbb{N}$ as the smallest integer such that*

$$r(d) \geq \frac{7}{2\ln(2)} \ln\left(\frac{d\ln(2) + 1}{2\ln(2)}\right),$$

then (119) *holds.*

Proof We determine $r \in \mathbb{N}$ such that

$$\exp(-2^{2k})2^{(k+1)d} < e^{-k} \qquad \text{for all} \quad k \geq r, \tag{121}$$

since then

$$\sum_{k \geq r} \exp(-2^{2k})2^{kd} < \sum_{k \geq r} \exp(-k) = \exp(-(r+1))/(1-e^{-1}) \leq 2\exp(-(r+1))$$
$$< 2\exp(-2). \tag{122}$$

To determine $r := r(d)$ such that (121) holds, we set

$$h_d(x) := -\exp(2x\ln(2)) + d(x+1)\ln(2) + x.$$

Then $h_d'(x) = -2\ln(2)\exp(2x\ln(2)) + d\ln(2) + 1$, and, since $h_d(x) \to -\infty$ as $x \to +\infty$, h_d has a unique global maximum at

$$\tilde{x} := \frac{1}{2\ln(2)} \ln\left(\frac{d\ln(2)+1}{2\ln(2)}\right).$$

Note that $h_d(\tilde{x}) > 0$. Therefore, we look for $r \geq \tilde{x}$ such that $h_d(r) < 0$, and then (121) follows. We make the ansatz $r =: s\tilde{x}$ with $s \geq 1$, and use the abbreviation $a_d := d\ln(2)$ and $b := 2\ln(2)$. Then

$$h_d(s\tilde{x}) = -\left(\frac{a_d+1}{b}\right)^s + \frac{s}{b}\ln\left(\frac{a_d+1}{b}\right)(a_d+1) + a_d. \tag{123}$$

If $d = 1, \ldots, 10$, it suffices to choose $s = 6$. If $d \geq 10$, we set

$$\tilde{h}_d(s) := -\left(\frac{a_d+1}{b}\right)^s + \frac{(a_d+1)}{b}\left(s\ln\left(\frac{a_d+1}{b}\right) + b\right) \geq h_d(s\tilde{x}).$$

Now we set

$$s(d) := \max\left\{2, \frac{1+2\ln(d/2)}{\ln(d/2)-1}\right\},$$

and estimate very roughly as follows: Since $\ln(\frac{a_d+1}{b}) \leq \frac{a_d+1}{b}$ and $b \leq 2\frac{a_d+1}{b}$, we have

$$\tilde{h}(s) \leq -\left(\frac{a_d+1}{b}\right)^2\left[-\left(\frac{a_d+1}{b}\right)^{s-2} + s + 2\right] < 0,$$

since by the choice of s,

$$\ln\left(\frac{a_d+1}{b}\right) \geq \ln(d/2) \geq \frac{s+1}{s-2} \geq \frac{\ln(s+2)}{s-2}.$$

Note that for $d \geq 10$, $s(d)$ is monotonically decreasing. Therefore, $s(d) \leq s(10) \leq 7$, and with $r(d) \geq 7\tilde{x}$ the assertion follows. $\qquad \square$

We now turn to (121) and use r as obtained in Lemma 10. By (122), it suffices to choose $\ell > 0$ such that

$$2^{d\ell - r(d)}(1 - \frac{2}{e}) > e,$$

which holds for

$$\ell \geq \frac{1}{d}\left[\frac{3 - \ln(e^2-2)}{\ln(2)} + r(d)\right].$$

In particular $\ell \to 0$ as $d \to \infty$, and thus $\tilde{b} \to 1$.

Acknowledgements

We are grateful for the comments and suggestions of the anonymous referees. The authors acknowledge support of the *Deutsche Forschungsgemeinschaft (DFG)* through the *Sonderforschungsbereich 1060: The Mathematics of Emergent Effects.*

References

1. R. A. ADAMS AND J. J. F. FOURNIER, *Sobolev Spaces*, Academic Press, Oxford (UK), 2003.
2. R. ARCANGÉLI, M. C. L. DI SILANES, AND J. 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.
3. R. ARCANGÉLI, M. C. L. DI SILANES, AND J. J. TORRENS, *Estimates for functions in Sobolev spaces defined on unbounded domains*, J. Approx. Theory, 161 (2009), pp. 198 – 212.
4. ———, *Extension of sampling inequalities to Sobolev semi-norms of fractional order and derivative data*, Numer. Math,, 121 (2012), pp. 587–608.
5. L. BOYTSOV AND B. NAIDAN, *Learning to prune in metric and non-metric spaces*, in Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States., C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger, eds., 2013, pp. 1574–1582.
6. T. BOZKAYA AND M. OZSOYOGLU, *Distance-based indexing for high-dimensional metric spaces*, in Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data, SIGMOD '97, New York, NY, USA, 1997, ACM, pp. 357–368.
7. S. CHANDRASEKARAN, K. JAYARAMAN, AND H. MHASKAR, *Minimum Sobolev norm interpolation with trigonometric polynomials on the torus*, Journal of Computational Physics, 249 (2013), pp. 96 – 112.
8. S. CHANDRASEKARAN, K. R. JAYARAMAN, AND H. N. MHASKAR, *Minimum Sobolev norm interpolation with trigonometric polynomials on the torus*, J. Comput. Phys, 249 (2013), pp. 96 – 112.
9. T. COULHON AND A. GRIGOR'YAN, *Random walks on graphs with regular volume growth*, Geom. Funct. Anal., 8 (1998), pp. 656–701.
10. T. COULHON, G. KERKYACHARIAN, AND P. PETRUSHEV, *Heat kernel generated frames in the setting of Dirichlet spaces*, J. Fourier Anal. Appl., 18 (2012), pp. 995–1066.
11. F. CUCKER AND D.-X. ZHOU, *Learning Theory: An Approximation Theory Viewpoint*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2007.
12. P. C. CURTIS, *n-parameter families and best approximation*, Pac. J. Math., 9 (1959), p. 10131027.
13. N. DUNFORD AND J. SCHWARTZ, *Linear Operators, Part I: Genral Theory*, Interscience Publishers, New York, 1958.
14. N. DYN, F. J. NARCOWICH, AND J. D. WARD, *Variational principles and Sobolev-type estimates for generalized interpolation on a Riemannian manifold*, Constr. Approx, 15 (1999), pp. 175–208.
15. P. F. EVANGELISTA, M. J. EMBRECHTS, AND B. K. SZYMANSKI, *Taming the curse of dimensionality in kernels and novelty detection*, in Applied Soft Computing Technologies: The Challenge of Complexity, A. Abraham, B. de Baets, M. Köppen, and B. Nickolay, eds., vol. 34 of Advances in Soft Computing, Springer Berlin Heidelberg, 2006, pp. 425–438.
16. A. GLOBERSON AND S. ROWEIS, *Metric learning by collapsing classes*, Adv. Neural Inf. Process Syst., 18 (2006), pp. 451–458.
17. M. GORDINA, T. KUMAGAI, L. SALOFF-COSTE, AND K.-T. STURM, *Heat kernels, stochastic processes and functional inequalities*, Oberwolfach Reports, 10 (2013), pp. 1359–1443.
18. M. GRIEBEL, C. RIEGER, AND B. ZWICKNAGL, *Multiscale approximation and reproducing kernel Hilbert space methods*, SIAM J. Numer. Anal., 53 (2015), pp. 852–873.
19. A. GRIGOR'YAN, *Heat kernels on weighted manifolds and applications*, Cont. Math., 398 (2006), pp. 93–191.
20. ———, *Heat Kernel and Analysis on Manifolds*, vol. 47 of AMS/IP Studies in Advanced Mathematics, American Mathematical Society, USA, 2009.
21. T. HANGELBROEK, F. NARCOWICH, C. RIEGER, AND J. WARD, *An inverse theorem for compact Lipschitz regions in $\mathbb{R}^d$ using localized kernel bases.* available at http://arxiv.org/abs/1508.02952, 2015.
22. T. HANGLEBROEK, F. NARCOWICH, X. SUN, AND J. WARD, *Kernel approximation on manifolds ii: the l^∞ norm of the l^2 projector*, SIAM J. Math. Anal., 43 (2011), pp. 662–684.
23. K. JETTER, J. STÖCKLER, AND J. D. WARD, *Norming sets and scattered data approximation on spheres*, in Approximation Theory IX, Vol. II: Computational Aspects, Vanderbilt University Press, 1998, pp. 137 – 144.
24. P. JORGENSEN AND F. TIAN, *Frames and factorization of graph Laplacians.* available at http://arxiv.org/abs/1404.1424, 2014.
25. A. KAENMAKI, J. LEHRBACK, AND M. VUORINEN, *Dimensions, Whitney covers, and tubular neighborhoods*, Indiana Univ. Math. J., 62 (2013), pp. 1861–1889.

26. E. Keogh and A. Mueen, *Curse of dimensionality*, in Encyclopedia of Machine Learning, C. Sammut and G. I. Webb, eds., Springer US, 2010, pp. 257–258.
27. G. Kerkyacharian and P. Petrushev, *Heat kernel based decomposition of spaces of distributions in the framework of dirichlet spaces*, Trans. Amer. Math. Soc., 367 (2015), pp. 121–189.
28. S. Lin, *Nonparametric regression using needlet kernels for spherical data*, ArXiv e-prints, (2015).
29. W. Madych, *An estimate for multivariate approximation II*, J. Approx. Theory, 142 (2006), pp. 116–128.
30. M. Maggioni and H. N. Mhaskar, *Diffusion polynomial frames on metric measure spaces*, Appl. Comp. Harm. Anal., 24 (2008), pp. 329 – 353.
31. J. Mairhuber, *On Haar's theorem concerning Chebysheff problems having unique solutions*, Proc. Amer. Math. Soc., 7 (1956), p. 609615.
32. H. N. Mhaskar, *A Markov-Bernstein inequality for Gaussian networks*, in Trends and Applications in Constructive Approximation, vol. 151 of Internat. Ser. Numer. Math., Birkhäuser, Basel, 2005, pp. 165–180.
33. ——, *Eignets for function approximation on manifolds*, Appl. Comp. Harm. Anal., 29 (2010), pp. 63 – 87.
34. H. N. Mhaskar, F. J. Narcowich, J. Prestin, and J. D. Ward, *L^p Bernstein estimates and approximation by spherical basis functions*, Math. Comp., 79 (2010), pp. 1647–1679.
35. F. Narcowich, P. Petrushev, and J. Ward, *Decomposition of Besov and Triebel-Lizorkin spaces on the sphere* , Journal of Functional Analysis, 238 (2006), pp. 530 – 564.
36. F. Narcowich, J. Sun, X.and Ward, and H. Wendland, *Direct and inverse Sobolev error estimates for scattered data interpolation via spherical basis functions*, Found. Comput. Math., 7 (2007), pp. 369–390.
37. F. Narcowich, J. Ward, and H. Wendland, *Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions*, Constr. Approx., 24 (2006), pp. 175–186.
38. F. J. Narcowich, P. Petrushev, and J. D. Ward, *Localized Tight Frames on Spheres*, SIAM J. Math. Anal., 38 (2006), pp. 574–594.
39. F. J. Narcowich and J. D. Ward, *Scattered-data interpolation on $\mathbb{R}^n$: Error estimates for radial basis and band-limited functions*, SIAM J. Math. Anal., 36 (2004), pp. 284–300.
40. F. J. Narcowich, J. D. 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.
41. R. Opfer, *Multiscale kernels*, Adv. Comp. Math., 25 (2006), pp. 357–380.
42. ——, *Tight frame expansions of multiscale reproducing kernels in Sobolev spaces*, Appl. Comput. Harm. Anal., 20 (2006), pp. 357–374.
43. P. Petrushev and Y. Xu, *Decomposition of spaces of distributions induced by Hermite expansions*, J. Fourier Anal. Appl., 14 (2008), pp. 371–414.
44. C. Rieger, *Sampling inequalities and applications*, PhD thesis, Fakultät für Mathematik und Informatik, University of Göttingen, 2008.
45. 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.
46. 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.
47. ——, *Sampling inequalities for infinitely smooth functions, with applications to interpolation and machine learning*, Adv. Comp. Math., 32(1) (2010), pp. 103–129.
48. R. Schaback and H. Wendland, *Inverse and saturation theorems for radial basis function interpolation*, Math. Comp., 71 (2002), pp. 669–681.
49. R. Schaback and H. Wendland, *Kernel techniques: From machine learning to meshless methods*, Acta Numerica, 15 (2006), pp. 543–639.
50. B. Schölkopf and A. J. Smola, *Learning with kernels - Support Vector Machines, Regularisation, and Beyond*, MIT Press, Cambridge, Massachusetts, 2002.
51. S. Smale and D.-X. Zhou, *Estimating the approximation error in learning theory*, Anal. Appl., 01 (2003), pp. 17–41.
52. B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. G. Lanckriet, *Hilbert space embeddings and metrics on probability measures*, J. Mach. Learn. Res., 11 (2010), pp. 1517–1561.
53. H. Triebel, *Interpolation theory, function spaces, differential operators*, North-Holland Mathematical Library, Amsterdam-New York, 1978.
54. ——, *Theory of function spaces*, vol. 78 of Monographs in Math., Birkhäuser Verlag, Basel, 1983.
55. F. Utreras, *Convergence rates for multivariate smoothing spline functions*, J. Approx. Theory, 52 (1988), pp. 1–27.
56. J. Ward, *L^p Bernstein inequalities and inverse theorems for RBF approximation on R^d*. preprint, arXiv:1010.4554v1 [math.CA], 2010.
57. H. Wendland, *Local polynomial reproduction and moving least squares approximation*, IMA J Numer. Anal., 21 (2001), pp. 285–300.
58. ——, *Scattered Data Approximation*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2005.
59. H. Wendland and C. Rieger, *Approximate interpolation with applications to selecting smoothing parameters*, Numer. Math., 101 (2005), pp. 729–748.

60. J. Werner, *Numerische Mathematik 1. Lineare und nichtlineare Gleichungssysteme, Interpolation, numerische Integration*, Vieweg, Braunschweig-Wiesbaden, 1992.
61. W. Zhang, X. Xue, Z. Sun, Y. F. Guo, and H. Lu, *Optimal dimensionality of metric space for classification*, in ICML '07: Proceedings of the 24th international conference on Machine learning, New York, NY, USA, 2007, ACM Press, pp. 1135–1142.
62. D.-X. Zhou, *The covering number in learning theory*, J. of Complexity, 18 (2002), pp. 739 – 767.