

Full Length Article

Metric entropy, n -widths, and sampling of functions on manifolds

Martin Ehler^{a,*}, Frank Filbir^{b,c}

^a University of Vienna, Department of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria

^b Faculty of Mathematics, Technische Universität München, Boltzmannstrasse 3, 85748, Garching, Germany

^c Helmholtz Zentrum München, Ingolstädter Landstrasse 1, D-85764 Neuherberg, Germany

Received 5 March 2015; received in revised form 13 August 2017; accepted 6 September 2017

Available online 14 September 2017

Communicated by H. Mhaskar

Dedicated to Hrushikesh N. Mhaskar on occasion of his 60th birthday.

Abstract

We first investigate on the asymptotics of the Kolmogorov metric entropy and nonlinear n -widths of approximation spaces on some function classes on manifolds and quasi-metric measure spaces. Secondly, we develop constructive algorithms to represent those functions within a prescribed accuracy. The constructions can be based on either spectral information or scattered samples of the target function. Our algorithmic scheme is asymptotically optimal in the sense of nonlinear n -widths and asymptotically optimal up to a logarithmic factor with respect to the metric entropy.

© 2017 Elsevier Inc. All rights reserved.

Keywords: Diffusion measure space; Diffusion polynomials; Sampling; Metric entropy

1. Introduction

In classical computational mathematics, it is customary to represent a function by using finitely many parameters, e.g., the coefficients of some truncated series expansion. Representing a function in terms of binary bits rather than a sequence of real numbers is the problem of

* Corresponding author.

E-mail addresses: martin.ehler@univie.ac.at (M. Ehler), filbir@helmholtz-muenchen.de (F. Filbir).

quantization. The integers thus obtained are represented as a bit string to which coding techniques can be applied to achieve a final compression. In wireless communication, for instance, one needs to transform an analogue signal into a stream of bits, from which the original signal should be recovered at the receiving end with a minimal distortion.

The theory of bit representation of functions pre-dates these modern requirements and was already studied by Kolmogorov. The notion of metric entropy in the sense of Kolmogorov gives a measurement of the minimal number of bits needed to represent an arbitrary function from a compact subset of a function space. Babenko, Kolmogorov, Tikhomirov, Vitushkin, and Yerokhin have given many estimates on the metric entropy for several compact subsets of the standard function spaces, cf. [21], [40, Chapter 2, 3].

Constructive algorithms were derived in [26] to represent functions in suitably defined Besov spaces on the sphere using asymptotically the same number of bits as the metric entropy of these classes, except for a logarithmic factor. A generalization was obtained for compact smooth Riemannian manifolds \mathbb{X} and global approximation spaces in [11]. Periodic function classes were considered in [8,37]. Related measures of complexity are n -widths [20] and were studied for some classical function spaces in [32,33], see [6] for the concept of nonlinear n -widths. We also refer to [1, Chapter 1], [10, Sections 1.3, 3, 4], [19, Sections 6.3, 6.4] for related results.

Both concepts, metric entropy and n -widths, are important complexity measures for the analysis of functions on high-dimensional datasets occurring in biology, medicine, and related areas. Many computational schemes are categorized into the field of manifold learning, where functions need to be learned from finitely many training data that are assumed to lie on some (unknown) manifold [2,4,30,35,36,38]. While much of the recent research in this direction focuses on the understanding of data geometry, approximation theory methods were introduced in [12,13,25,27,28].

The purpose of the present paper is to generalize results on metric entropy and n -widths to the context of sampled functions on quasi-metric measure spaces and covering a larger scale of approximation spaces. Indeed, we determine the asymptotics of the metric entropy and nonlinear n -widths for global approximation spaces. We also provide a computational approximation method, and this scheme is asymptotically optimal with respect to the nonlinear n -widths and asymptotically optimal up to a logarithmic factor in the sense of the metric entropy. In addition to obtaining theoretical bounds on the metric entropy and n -widths, our results have the following notable features:

- The computational scheme is based on a linear approximation operator to asymptotically match the optimal bounds in the sense of n -widths.
- We give explicit schemes for converting the target function into a near minimal number of bits by combining the linear approximation operator with linear quantization, and we derive a reconstruction scheme from such bits to a prescribed accuracy.
- Our constructions can deal with both, spectral information as well as finitely many training data consisting of function evaluations at scattered data points.

In addition, we shall investigate on the asymptotics of the metric entropy of local approximation spaces.

The outline of this paper is as follows: In Section 2 we introduce the setting and define metric entropy and n -widths. The asymptotics of the metric entropy of global (and under additional assumptions also local) approximation spaces is determined in Section 2.3. In Section 3, we introduce our approximation schemes for global approximation spaces and compute the asymptotics of the n -widths of global approximation spaces. In Section 4 we verify that linear

quantization of the approximation scheme leads to optimal bit representations up to a logarithmic factor for the global approximation space. Local versions of approximation spaces are considered in Section 5. For the readers convenience, [Appendix](#) contains a list and brief discussion of the technical assumptions used for the main results of the present paper.

2. Approximation spaces and their metric entropy and n -widths

2.1. Diffusion measure space

We first fix the setting and introduce some technical assumptions used throughout the paper. Let (\mathbb{X}, ρ) be a quasi-metric space, i.e., a space with a nonnegative symmetric map ρ satisfying $\rho(x, y) = 0$ if and only if $x = y$, and the triangle inequality holds at least up to a constant factor $c > 0$, i.e.,

$$\rho(x, y) \leq c(\rho(x, z) + \rho(z, y)), \quad \text{for all } x, y, z \in \mathbb{X}.$$

The quasi-metric ρ induces a topology and we assume that \mathbb{X} is endowed with a Borel probability measure μ . The system $\{\varphi_k\}_{k=0}^\infty \subset L_2(\mathbb{X}, \mu)$ is supposed to be an orthonormal basis of continuous real-valued functions with $\varphi_0 \equiv 1$ and our results also involve a sequence of nondecreasing real numbers $\{\lambda_k\}_{k=0}^\infty$ such that $\lambda_0 = 0$ and $\lambda_k \rightarrow \infty$ as $k \rightarrow \infty$. Let N be a positive integer and we shall restrict ourselves to $N = 2^n$, where n is some nonnegative integer. The space of *diffusion polynomials* up to degree N is $\Pi_N := \text{span}\{\varphi_k : \lambda_k \leq N\}$, and the *generalized heat kernel* is

$$G_t(x, y) = \sum_{k=0}^{\infty} \exp(-\lambda_k^2 t) \varphi_k(x) \varphi_k(y), \quad t > 0. \quad (1)$$

We use the symbols \lesssim when the left-hand side is bounded by a generic positive constant times the right-hand side, \gtrsim is used analogously, and \asymp means that both \lesssim and \gtrsim hold. Now, we can summarize the technical assumptions that are related to so-called upper and lower Gaussian bounds on the generalized heat kernel:

Definition 2.1 ([3]). Under the above notation, a quasi-metric space \mathbb{X} is called a *diffusion measure space* if there is some $\alpha > 0$ such that each of the following properties is satisfied:

- (i) For each $x \in \mathbb{X}$ and $t > 0$, the closed ball $B_t(x)$ of radius t at x is compact and

$$\mu(B_t(x)) \lesssim t^\alpha, \quad x \in \mathbb{X}, \quad t > 0.$$

- (ii) There is $c > 0$ such that

$$|G_t(x, y)| \lesssim t^{-\alpha/2} \exp\left(-c \frac{\rho(x, y)^2}{t}\right), \quad x, y \in \mathbb{X}, \quad 0 < t \leq 1.$$

- (iii) We have $t^{-\alpha/2} \lesssim G_t(x, x)$, $x \in \mathbb{X}$, $0 < t < 1$.

From here on, we suppose that \mathbb{X} is a diffusion measure space throughout the present paper. It is also noteworthy that the conditions of a diffusion measure space imply that $\mu(B_t(x)) \asymp t^\alpha$, for all $0 < t < 1$, cf. [12]. Thus, the volume of small balls essentially behaves as in \mathbb{R}^α . The above conditions also imply the following estimate on the Christoffel function (or spectral function),

$$\sum_{\lambda_k \leq t} |\varphi_k(x)|^2 \asymp t^\alpha, \quad x \in \mathbb{X}, \quad t \geq 1, \quad (2)$$

see [3,12,13] for a discussion and references. By integrating over \mathbb{X} , we obtain

$$\dim(\Pi_t) \asymp t^\alpha. \quad (3)$$

Remark 2.2. It was pointed out in [3] that all technical assumptions are satisfied when $\mathbb{X} \subset \mathbb{R}^d$ is an α -dimensional compact, connected, Riemannian manifold without boundary, with non-negative Ricci curvature, geodesic distance ρ , and μ being the Riemannian volume measure on \mathbb{X} normalized with $\mu(\mathbb{X}) = 1$, $\{\varphi_k\}_{k=0}^\infty$ are the eigenfunctions of the Laplace–Beltrami operator on \mathbb{X} , and $\{-\lambda_k\}_{k=0}^\infty$ are the corresponding eigenvalues arranged in nonincreasing order, see also [18]. In this case, (3) is consistent with Weyl’s law. For further discussions, we refer to [12,13,27].

Given an arbitrary normed space Z and a subset $Y \subset Z$, we define, for $f \in Z$,

$$E(f, Y, Z) := \inf_{g \in Y} \|f - g\|_Z. \quad (4)$$

As we have pointed out already, we assume \mathbb{X} is a diffusion measure space throughout. The notation $|B$ in the following means restricting functions to the set B .

Definition 2.3. For a nonempty ball $B \subset \mathbb{X}$ and $1 \leq p \leq \infty$, the *approximation space* of order $s > 0$ is defined by

$$\mathcal{A}^s(L_p(B)) = \{f \in L_p(B) : \|f\|_{\mathcal{A}^s(L_p(B))} < \infty\}, \quad (5)$$

where the associated norm is given by

$$\|f\|_{\mathcal{A}^s(L_p(B))} := \|f\|_{L_p(B)} + \sup_{N \geq 1} N^s E(f, \Pi_{N|B}, L_p(B)).$$

The unit ball in $\mathcal{A}^s(L_p(B))$ is denoted by

$$\overline{\mathcal{A}^s(L_p(B))} := \{f \in L_p(B) : \|f\|_{\mathcal{A}^s(L_p(B))} \leq 1\}. \quad (6)$$

For a nonempty ball $B \subset \mathbb{X}$, let us denote the closure of $\bigcup_N \Pi_{N|B}$ in $L_p(B)$ by $X_p(B)$. We can simply switch from $L_p(B)$ to $X_p(B)$ in the definition of approximation space without changing anything, so that we observe $\mathcal{A}^s(L_p(B)) = \mathcal{A}^s(X_p(B))$.

Remark 2.4. In classical situations, the smoothness of a function is related to the accuracy of its approximation, and for a more classical characterization of the approximation spaces in terms of pseudo-differential operators and K-functionals, we refer to [25]. In turn it has also become customary to consider the accuracy of approximation itself as a measurement of smoothness. For classical results on approximation spaces, we refer to [7, Chapter 7] and references therein.

2.2. Kolmogorov metric entropy and nonlinear n -widths

Metric entropy as studied in [23] refers to the minimal number of bits needed to represent a function f up to precision ε . This number determines the maximal compression when loss of information is bounded by ε . For a more stringent mathematical exposition, let Y be a compact subset of a metric space (Z, ϱ) . Given $\varepsilon > 0$, let $\mathcal{N}_\varepsilon(Y)$ be the ε -covering number of Y in Z , i.e., the minimal number of balls of radius ε that cover Y . Suppose that $g_1, \dots, g_{\mathcal{N}_\varepsilon(Y)}$ is a list of centers of these balls. Given any $f \in Y$, there is g_j such that $\varrho(f, g_j) \leq \varepsilon$. We may then

represent f using the binary representation of j , and use g_j as the reconstruction of f based on this representation. Any binary enumeration of these centers takes $\lceil \log_2(\mathcal{N}_\varepsilon(Y)) \rceil$ many bits, which somewhat measures the complexity of Y and where $\lceil \cdot \rceil$ denotes the ceiling function.

Definition 2.5. Let Y be a compact subset of a metric space (Z, ϱ) and, for $\varepsilon > 0$, let $\mathcal{N}_\varepsilon(Y)$ be the ε -covering number of Y in Z . Then

$$H_\varepsilon(Y, Z) := \log_2(\mathcal{N}_\varepsilon(Y)) \quad (7)$$

is called the *metric entropy* of Y in Z .

Thus, the smallest integer not less than the metric entropy is the minimal number of bits necessary to represent any f with precision ε .

Let us also introduce some alternative notions of complexity. Given some Banach space Z , let $n \geq 1$ be an integer, and let $M_n : \mathbb{R}^n \rightarrow Z$ be some mapping, for which $Z_n := M_n(\mathbb{R}^n)$ denotes its range. For a compact subset $Y \subset Z$, we define the worst case error of approximation, consistently with (4), by

$$E(Y, Z_n, Z) := \sup_{f \in Y} E(f, Z_n, Z).$$

Endow Y with some quasi-norm and consider continuous maps $\tilde{a} : Y \rightarrow \mathbb{R}^n$. For fixed \tilde{a} , the term $M_n(\tilde{a}(f)) \in Z$ is an approximation of f from Z_n . The quantity

$$E(Y, \tilde{a}, M_n, Z) := \sup_{f \in Y} \|f - M_n(\tilde{a}(f))\|_Z$$

is the error of approximation by the nonlinear method of approximation $M_n(\tilde{a}(\cdot)) : Y \rightarrow Z$. The *continuous nonlinear n -width* is defined in [6] by

$$d_n(Y, Z) := \inf_{\tilde{a}, M_n} E(Y, \tilde{a}, M_n, Z),$$

where the infimum runs over all continuous maps $\tilde{a} : Y \rightarrow \mathbb{R}^n$ and all mappings $M_n : \mathbb{R}^n \rightarrow Z$. For further considerations on n -widths, we refer to [6, 9, 14, 29, 31, 39].

In the subsequent sections we shall compute the metric entropy (7) and the continuous nonlinear n -widths of the approximation ball $\overline{\mathcal{A}}^s(X_p(\mathbb{X}))$ of radius 1 given by (6). We refer to [17] for related results.

2.3. The metric entropy of approximation spaces

We shall determine the asymptotics of the metric entropy of global and local approximation spaces, so let $B \subset \mathbb{X}$ be some nonempty ball, which is allowed to coincide with \mathbb{X} . Since $\mathcal{A}^s(X_p(B))$ is not finite-dimensional, $\overline{\mathcal{A}}^s(X_p(B))$ is not compact in the approximation space. Here, we consider $\mathcal{A}^s(X_p(B))$ as a subspace of $X_p(B)$, in which it is compact, see [11] for $B = \mathbb{X}$ and the case $B \subsetneq \mathbb{X}$ can be proven analogously.

The following result for the approximation space extends findings in [26] from the sphere to diffusion measure spaces.

Theorem 2.6. Let \mathbb{X} be a diffusion measure space and suppose we are given a nonempty ball $B \subset \mathbb{X}$ such that $\{\varphi_{k|B}\}_{j=0}^\infty$ are linearly independent. If $s > 0$ is fixed, then

$$H_\varepsilon(\overline{\mathcal{A}}^s(X_p(B)), X_p(B)) \asymp (1/\varepsilon)^{\alpha/s}, \quad \text{for all } 0 < \varepsilon \leq 1, \quad (8)$$

where α is the constant in Definition 2.1.

This result was verified in [11] provided that $B = \mathbb{X}$ and $p = \infty$. Note that the linear independence condition is trivially satisfied for $B = \mathbb{X}$ since $\{\varphi_k\}_{j=0}^\infty$ is even an orthonormal basis. In case B is a proper subset of \mathbb{X} , the linear independence condition may be hard to check in general. However, if \mathbb{X} is a compact, connected, real-analytic Riemannian manifold (equipped with a real analytic Riemannian metric), then the eigenfunctions $\{\varphi_k\}_{j=0}^\infty$ of the Laplace–Beltrami operator are real analytic due to the real analytic hypoellipticity of elliptic partial differential operators, cf. [22, Section 5.3]. In this situation, the global linear independence implies the linear independence of the restrictions. For results related to Theorem 2.6, see [9, 31, 32].

The proof of Theorem 2.6 is based on a general Banach space result, also used in [26, Theorem 4.1]. Let Z be a Banach space and $\{\phi_k\}_{k=1}^\infty \subset Z$ be a sequence of linearly independent elements whose linear span is dense in Z , and define $Z_k := \text{span}\{\phi_1, \dots, \phi_k\}$ with $Z_0 = \{0\}$. Let $\{\delta_k\}_{k=0}^\infty$ be a nonincreasing sequence of positive numbers with $\lim_{k \rightarrow \infty} \delta_k = 0$ and define

$$\mathcal{A}(Z; \{\delta_k\}_{k=0}^\infty, \{\phi_k\}_{k=1}^\infty) := \{f \in Z : E(f, Z_k, Z) \leq \delta_k, \text{ for } k = 0, 1, \dots\}. \quad (9)$$

The following result goes back to Lorentz in [23]:

Theorem 2.7 (Theorem 3.3 in [24]). *Let $\{\delta_k\}_{k=0}^\infty$ be a nonincreasing sequence of positive numbers such that $\delta_{2k} \leq c\delta_k$, for $k = 1, 2, \dots$ and some constant $c \in (0, 1)$. For $\ell \geq 0$, let $M_\ell := \min\{k : \delta_k \leq e^{-\ell}\}$, then we have, for $0 < \varepsilon \leq 1$,*

$$H_\varepsilon(\mathcal{A}(Z; \{\delta_k\}_{k=0}^\infty, \{\phi_k\}_{k=1}^\infty), Z) \asymp \sum_{\ell=1}^L M_\ell, \quad (10)$$

where $L := 2 + \lfloor \log(1/\varepsilon) \rfloor$.

Proof of Theorem 2.6. We aim to apply Theorem 2.7 with the function system $\{\varphi_{k|B}\}_{k=0}^\infty$ and $Z = X_p(B)$. There, the index set is supposed to start with $k = 1$, so we set $\phi_k = \varphi_{k-1|B}$, $k = 1, 2, \dots$. To define the sequence $\{\delta_k\}_{k=0}^\infty$, we need some preparation. The linear independence assumption yields that (3) implies $\dim(\Pi_{N|B}) \asymp N^\alpha$. By using $Z_k := \text{span}\{\varphi_{0|B}, \dots, \varphi_{k-1|B}\}$, we derive, for $N^\alpha \leq k \leq (2N)^\alpha$,

$$(2N)^s E(f, \Pi_{2N|B}, X_p(B)) \lesssim k^{s/\alpha} E(f, Z_k, X_p(B)) \lesssim N^s E(f, \Pi_{N|B}, X_p(B)).$$

Therefore, there are constants $C_i \geq 1$, for $i = 1, 2$, such that the definitions $\delta_{1;0} = \frac{1}{2}$, $\delta_{1;k} := (2C_1)^{-1} k^{-s/\alpha}$, and $\delta_{2;0} = C_2$, $\delta_{2;k} := C_2 k^{-s/\alpha}$, lead to

$$\mathcal{A}(X_p(B); \{\delta_{1;k}\}_{k=0}^\infty, \{\phi_k\}_{k=1}^\infty) \subset \overline{\mathcal{A}}^s(X_p(B)) \subset \mathcal{A}(X_p(B); \{\delta_{2;k}\}_{k=0}^\infty, \{\phi_k\}_{k=1}^\infty),$$

which also yields

$$\begin{aligned} \mathcal{N}_\varepsilon(\mathcal{A}(X_p(B); \{\delta_{1;k}\}_{k=0}^\infty, \{\phi_k\}_{k=1}^\infty)) &\leq \mathcal{N}_\varepsilon(\overline{\mathcal{A}}^s(X_p(B))) \\ &\leq \mathcal{N}_\varepsilon(\mathcal{A}(X_p(B); \{\delta_{2;k}\}_{k=0}^\infty, \{\phi_k\}_{k=1}^\infty)). \end{aligned}$$

Since $\delta_{i;2k} \leq c\delta_{i;k}$, for $c := 2^{-s/\alpha} \in (0, 1)$, we can apply Theorem 2.7. According to [26, Lemma 4.1], $\sum_{\ell=1}^L M_\ell \asymp e^{L\alpha/s}$, so that the choice of L in (10) implies (8). \square

It is obvious that increased precision requires more bits, and smoother functions can be represented with fewer bits. The exact growth condition (8) reflects these thoughts in a quantitative fashion and that (8) serves as a benchmark for function representation on diffusion measure spaces. The remaining part of the present work is dedicated to develop a scheme that matches the optimality bound at least up to a logarithmic factor.

3. Global function approximation

This section is dedicated to introduce our approximation scheme, to discuss its characterization of certain function spaces, and to determine the asymptotics of the n -widths.

3.1. Localized kernels

We now collect few ingredients for our approximation scheme.

Definition 3.1. We call an infinitely often differentiable and non-increasing function $H : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ a *low-pass filter* if $H(t) = 1$ for $t \leq 1/2$ and $H(t) = 0$ for $t \geq 1$.

For any low-pass filter H , the kernel

$$K_N(x, y) := \sum_{k=0}^{\infty} H\left(\frac{\lambda_k}{N}\right) \varphi_k(x) \varphi_k(y), \quad (11)$$

cf. [12, 25, 27], is localized, i.e., for fixed $r > \alpha$ and all $x \neq y$ with $N = 1, 2, \dots$,

$$|K_N(x, y)| \lesssim \frac{N^{\alpha-r}}{\rho(x, y)^r}. \quad (12)$$

Alternatively, we also have for fixed $r > \alpha$ and all x, y with $N = 1, 2, \dots$,

$$|K_N(x, y)| \lesssim \frac{N^{\alpha}}{\max(1, (N^r \rho(x, y)^r))}. \quad (13)$$

We find in [12, Inequality (3.12)] that

$$\sup_{y \in \mathbb{X}} \int_{\mathbb{X}} |K_N(x, y)| d\mu(x) \lesssim 1 \quad (14)$$

holds.

For some signed Borel measure ν on \mathbb{X} and $f \in L_1(\mathbb{X}, |\nu|)$, we can define, for $N = 1, 2, \dots$,

$$\sigma_N(f, \nu) := \sum_{k=0}^{\infty} H\left(\frac{\lambda_k}{N}\right) \int_{\mathbb{X}} f(x) \varphi_k(x) \varphi_k d\nu(x) = \int_{\mathbb{X}} f(x) K_N(x, \cdot) d\nu(x). \quad (15)$$

3.2. Quadrature measures

This section is dedicated to introduce further ingredients to develop our approximation scheme. We first aim to replace the integral over diffusion polynomials with a finite sum or at least with an integral over a “simpler” measure.

Definition 3.2. A signed Borel measure ν on \mathbb{X} is called a *quadrature measure* of order N if

$$\int_{\mathbb{X}} f(x) g(x) d\mu(x) = \int_{\mathbb{X}} f(x) g(x) d\nu(x), \quad \text{for all } f, g \in \Pi_N.$$

While Π_N is finite dimensional, the collection of products is so as well and, hence, there do exist finitely supported quadrature measure, cf. [5], [34, Section 1.5]. Note that we request exact integration of products from functions in Π_N , which leads us to the following product assumption on the structure of diffusion polynomials, which is also used in [13]: for $f \in L_p(\mathbb{X})$, let

$$\text{dist}(f, \Pi_N)_{L_{\infty}} := \inf_{h \in \Pi_N} \|f - h\|_{L_{\infty}}$$

and assume that there is a constant $a \geq 2$ such that the quantity

$$\epsilon_N := \sup_{\lambda_\ell, \lambda_k \leq N} \text{dist}(\varphi_\ell \varphi_k, \Pi_{aN})_{L_\infty} \quad (16)$$

satisfies $N^m \epsilon_N \rightarrow 0$ as $N \rightarrow \infty$, for all $m > 0$.

Definition 3.3. Let the product assumption hold. For fixed $1 \leq p \leq \infty$, a family $(\mu_N)_{N=1}^\infty$ of positive quadrature measures of order N , respectively, is called a family of *Marcinkiewicz–Zygmund quadrature measures* of order N , respectively, if

$$\|f\|_{|\mu_N|, L_p(\mathbb{X})} \asymp \|f\|_{L_p(\mathbb{X})}, \quad \text{for all } f \in \Pi_N, \quad (17)$$

where $|\mu_N|$ denotes the total variation measure of μ_N and $\|f\|_{|\mu_N|, L_p(\mathbb{X})}$ denotes the L_p -norm of f with respect to $|\mu_N|$.

Under fairly general assumptions, the results in [12, Theorem 3.1] and [13, Theorem 5.8] imply the existence of a family of finitely supported Marcinkiewicz–Zygmund quadrature measures $(\mu_N)_{N=1}^\infty$ of order N , respectively, such that $\#\text{supp}(\mu_N) \asymp N^\alpha$, where α is as in Definition 2.1.

3.3. Widths and characterization of approximation spaces

Fix $1 \leq p \leq \infty$ and suppose that $(\mu_N)_{N=1}^\infty$ is a family of Marcinkiewicz–Zygmund quadrature measures of order N , respectively. According to [12, Inequality (3.13)], we have

$$\|\sigma_N(f, \mu_N)\|_{L_p} \lesssim \|f\|_{|\mu_N|, L_p}, \quad f \in L_p(\mathbb{X}, |\mu_N|), \quad (18)$$

and the generic constant can be chosen independently of f and N . Later, we shall also need that (17) extends to all functions in $X_p(\mathbb{X})$, i.e., that $\|f\|_{|\mu_N|, L_p} \lesssim \|f\|_{L_p(\mathbb{X})}$ holds, for all $f \in X_p(\mathbb{X})$. The latter is fine for $p = \infty$. For $1 \leq p < \infty$, we have not yet found any explicit example except for the measure μ itself. Therefore, we shall simply restrict ourselves to $\mu_N = \mu$, $N = 1, 2, \dots$ in this case.

We can also estimate

$$\sup_{y \in \mathbb{X}} \int_{\mathbb{X}} |K_N(x, y)| d|\mu_N|(x) \lesssim 1, \quad (19)$$

cf. [12, Inequality (3.12)]. Note that (19) is the quadrature version of (14). Next, we recall the characterization of $\mathcal{A}^s(L_p(\mathbb{X}))$ using σ_N , see [12, 25, 27, 28]:

Theorem 3.4. Suppose that $1 \leq p \leq \infty$ and assume that $(\mu_N)_{N=1}^\infty$ is a family of Marcinkiewicz–Zygmund quadrature measures of order N , respectively, if $p = \infty$. For $1 \leq p < \infty$ we choose $\mu_N = \mu$, $N = 1, 2, \dots$. Assume further that H is a low-pass filter. Then, for all $f \in \mathcal{A}^s(L_p(\mathbb{X}))$, we have

$$\|f - \sigma_N(f, \mu_N)\|_{L_p(\mathbb{X})} \lesssim N^{-s} \|f\|_{\mathcal{A}^s(L_p(\mathbb{X}))}, \quad (20)$$

where the generic constant does not depend on N or f . On the other hand, if, for $f \in L_p(\mathbb{X})$, there are generic constants not depending on N such that

$$\|f - \sigma_N(f, \mu_N)\|_{L_p(\mathbb{X})} \lesssim N^{-s},$$

then $f \in \mathcal{A}^s(L_p(\mathbb{X}))$.

Remark 3.5. Let us point out again that we suppose $\mu_N = \mu$, $N = 1, 2, \dots$ for $1 \leq p < \infty$. In this case, the term $\sigma_N(f, \mu_N)$ contains spectral information $\int_{\mathbb{X}} f(x) \varphi_k(x) d\mu(x)$. If $p = \infty$ and μ_N has finite support, then we have an approximation scheme that uses finitely many training data consisting of function evaluations at scattered data points in $\text{supp}(\mu_N)$.

We have already determined the asymptotics of the Kolmogorov metric entropy. Here, we shall determine the n -widths for the global approximation space.

Theorem 3.6. *The continuous nonlinear n -widths of $\overline{\mathcal{A}^s}(X_p(\mathbb{X}))$ in $X_p(\mathbb{X})$ satisfy*

$$d_n(\overline{\mathcal{A}^s}(X_p(\mathbb{X})), X_p(\mathbb{X})) \asymp n^{-s/\alpha}.$$

In order to verify [Theorem 3.6](#) we shall consider two more types of n -widths, cf. [\[33\]](#). The linear n -width of a subset Y in a Banach space Z is

$$\mathcal{L}_n(Y, Z) := \inf_{F_n} \sup_{x \in Y} \|x - F_n(x)\|,$$

where the infimum is taken over all bounded linear operators F_n on Z whose range is of dimension at most n . The Bernstein n -width of Y in Z is

$$\mathcal{B}_n(Y, Z) := \sup_{Z_{n+1}} \sup\{\lambda : \lambda \overline{Z_{n+1}} \subset Y\}$$

where the supremum is taken over all subspaces Z_{n+1} of Z of dimension at least $n+1$ and $\overline{Z_{n+1}}$ denotes the unit ball in Z_{n+1} . The continuous nonlinear n -widths are sandwiched by

$$\mathcal{B}_n(Y, Z) \leq d_n(Y, Z) \leq \mathcal{L}_n(Y, Z), \quad (21)$$

cf. [\[6\]](#). We can now take care of the proof. We point out that we shall derive a lower bound on the Bernstein n -width, which is closely related to the Bernstein estimates in [\[25\]](#) for integer derivatives when combined with K-functionals, cf. [\[6\]](#).

Proof of Theorem 3.6. The operator σ_N is bounded on $L_p(\mathbb{X})$ and $\sigma_N(f)$ is an element in Π_N . Consider N with $N \asymp n^{1/\alpha}$, so that $\dim(\Pi_N) \asymp N^\alpha$ yields

$$\mathcal{L}_n(\overline{\mathcal{A}^s}(X_p(\mathbb{X})), X_p(\mathbb{X})) \lesssim n^{-s/\alpha},$$

cf. [Theorem 3.4](#). Hence, the second inequality in [\(21\)](#) implies $d_n(\overline{\mathcal{A}^s}(X_p(\mathbb{X})), X_p(\mathbb{X})) \lesssim n^{-s/\alpha}$.

To establish the lower bound, we take the subspace Π_{N+1} and aim to derive a generic constant $c > 0$ such that $cN^{-s} \overline{\Pi_{N+1}} \subset \overline{\mathcal{A}^s}(X_p(\mathbb{X}))$, where $\overline{\Pi_{N+1}} = \{f \in \Pi_{N+1} : \|f\|_{L_p(\mathbb{X})} \leq 1\}$. For $f \in \overline{\Pi_{N+1}}$, let $g := N^{-s} f$. We obtain $\|g\|_{L_p(\mathbb{X})} \leq N^{-s}$ and, for $M > N$, we derive $E(g, \Pi_M, L_p(\mathbb{X})) = 0$. The choice $M \leq N$ yields

$$E(g, \Pi_M, L_p(\mathbb{X})) = N^{-s} \|f - \sigma_M(f)\|_{L_p(\mathbb{X})} \lesssim N^{-s},$$

because $\|f\|_{L_p(\mathbb{X})} \leq 1$ and $\|\sigma_M(f)\|_{L_p(\mathbb{X})} \lesssim \|f\|_{L_p(\mathbb{X})}$. Thus,

$$\sup_{M \geq 1} M^s E(g, \Pi_M, L_p(\mathbb{X})) \lesssim 1$$

holds. We have verified that there is a generic constant c such that

$$cN^{-s} \overline{\Pi_{N+1}} \subset \overline{\mathcal{A}^s}(L_p(\mathbb{X})).$$

Therefore, we obtain $\mathcal{B}_n(\overline{\mathcal{A}^s}(X_p(\mathbb{X})), X_p(\mathbb{X})) \gtrsim n^{-s/\alpha}$, so that [\(21\)](#) concludes the proof. \square

It should be mentioned that upper bounds on linear n -widths for compact Riemannian manifolds were already derived in [\[16\]](#), where also the exact asymptotics were obtained for compact homogeneous manifolds.

4. Bit representation in global approximation spaces

This section is dedicated to verify that linear quantization of the approximation scheme $\sigma_N(f, \mu_N)$ enables bit representations matching the optimality bounds stated in [Theorem 2.6](#) up to a logarithmic factor. First, we recall the formula [\(15\)](#),

$$\sigma_N(f, \mu_N) = \sum_{k=0}^{\infty} H\left(\frac{\lambda_k}{N}\right) \int_{\mathbb{X}} f(x) \varphi_k(x) \varphi_k d\mu_N(x),$$

where $(\mu_N)_{N=1}^{\infty}$ is a family of Marcinkiewicz–Zygmund quadrature measures of order N , respectively, if $p = \infty$. Again, if $1 \leq p < \infty$, then we choose $\mu_N = \mu$, $N = 1, 2, \dots$

Since $H(t) = 1$, for $t \in [0, 1/2]$ and $H(t) = 0$, for $t > 1$, we observe that $H(\frac{\lambda_k}{N})H(\frac{\lambda_k}{2N}) = H(\frac{\lambda_k}{N})$. If $(\nu_N)_{N=1}^{\infty}$ is a family of quadrature measures of order $2N$, respectively, then a straightforward calculation yields

$$\sigma_N(f, \mu_N) = \int_{\mathbb{X}} \sigma_N(f, \mu_N, x) \sum_{k=0}^{\infty} H\left(\frac{\lambda_k}{2N}\right) \varphi_k(x) d\nu_N(x) \varphi_k. \quad (22)$$

The representation [\(22\)](#) involves the quadrature measure ν_N and the Marcinkiewicz–Zygmund quadrature measure μ_N . To design the final approximation scheme, we fix some $S > 1$, apply the quantization

$$I_N(f, \mu_N, x) = \lfloor N^S \sigma_N(f, \mu_N, x) \rfloor, \quad (23)$$

and define the actual approximation by

$$\sigma_N^{\circ}(f, \mu_N, \nu_N) := N^{-S} \int_{\mathbb{X}} I_N(f, \mu_N, x) \sum_{k=0}^{\infty} H\left(\frac{\lambda_k}{2N}\right) \varphi_k(x) d\nu_N(x) \varphi_k. \quad (24)$$

In other words, we replace $\sigma_N(f, \mu_N, x)$ in [\(22\)](#) with a number on the grid $\frac{1}{N^S} \mathbb{Z}$.

We have the following result for the ball $\overline{\mathcal{A}^s}(L_p(\mathbb{X}))$ of radius 1 of the global approximation space given by [\(5\)](#). It extends results in [\[11\]](#) from compact Riemannian manifolds and $p = \infty$ to diffusion measure spaces and to the entire range $1 \leq p \leq \infty$:

Theorem 4.1. *For the case $p = \infty$, we assume that $(\mu_N)_{N=1}^{\infty}$ is a family of Marcinkiewicz–Zygmund quadrature measures of order N , respectively. For $1 \leq p < \infty$ we choose $\mu_N = \mu$, $N = 1, 2, \dots$. Assume further that H is a low-pass filter. We suppose that $(\nu_N)_{N=1}^{\infty}$ are Marcinkiewicz–Zygmund quadrature measures of order $2N$ with $\#\text{supp}(\nu_N) \lesssim N^{\alpha}$. For fixed $s > 0$ and $S > \max(1, s)$, we apply the discretizations [\(23\)](#) and [\(24\)](#). Then there is a constant $c > 0$ such that, for all $f \in \overline{\mathcal{A}^s}(L_p, \mathbb{X})$,*

$$\|f - \sigma_N^{\circ}(f, \mu_N, \nu_N)\|_{L_p(\mathbb{X})} \leq cN^{-s} \quad (25)$$

holds. For $cN^{-s} = \varepsilon \leq 1$ and $\varepsilon \leq 1$, the number of bits needed to represent all integers $\{I_N(f, \mu_N, x) : x \in \text{supp}(\nu_N)\}$ does not exceed a positive constant (independent of ε) times

$$(1/\varepsilon)^{\alpha/s} (1 + \log_2(1/\varepsilon)). \quad (26)$$

Proof of Theorem 4.1. The triangle inequality yields

$$\begin{aligned} \|f - \sigma_N^{\circ}(f, \mu_N, \nu_N)\|_{L_p(\mathbb{X})} &\lesssim \|f - \sigma_N(f, \mu_N)\|_{L_p(\mathbb{X})} \\ &\quad + \|\sigma_N(f, \mu_N) - \sigma_N^{\circ}(f, \mu_N, \nu_N)\|_{L_p(\mathbb{X})}. \end{aligned}$$

Since Theorem 3.4 implies $\|f - \sigma_N(f, \mu_N)\|_{L_p(\mathbb{X})} \lesssim N^{-s} \|f\|_{\mathcal{A}^s(L_p(\mathbb{X}))}$, we only need to take care of the term on the far most right. The quantization (23) immediately yields

$$|\sigma_N(f, \mu_N, x) - N^{-S} I_N(f, \mu_N, x)| \leq N^{-S}, \quad \text{for all } x \in \text{supp}(\nu_N), \quad (27)$$

so that (22) and (19) imply

$$\begin{aligned} \|\sigma_N(f, \mu_N) - \sigma_N^\circ(f, \mu_N, \nu_N)\|_{L_p(\mathbb{X})} &= \left\| \int_{\mathbb{X}} \left(\sigma_N(f, \mu_N, y) - N^{-S} I_N(f, \mu_N, y) \right) K_{2N}(\cdot, y) d\nu_N(y) \right\|_{L_p(\mathbb{X})} \\ &\lesssim N^{-S} \leq N^{-s}. \end{aligned}$$

Hence, we have derived (25).

To tackle (26), we observe that the localization property (13) yields $\|g\|_{L_\infty} \lesssim N^\alpha \|g\|_{L_1}$, for all $g \in \Pi_N$, see also [27, Lemma 5.5] for more general Nikolskii inequalities. We apply (27) and then use $\sigma_N(f, \mu_N) \in \Pi_N$ with $L_p \hookrightarrow L_1$, which yields

$$|I_N(f, \mu_N, x)| \lesssim N^S \|\sigma_N(f, \mu_N)\|_{L_\infty(\mathbb{X})} \lesssim N^{S+\alpha} \|\sigma_N(f, \mu_N)\|_{L_p(\mathbb{X})}.$$

According to [27, Theorem 5.1], $\|\sigma_N(f, \mu_N)\|_{L_p(\mathbb{X})} \lesssim \|f\|_{L_p(\mathbb{X})}$ holds. Since f is contained in the ball of radius 1, so that $\|f\|_{L_p(\mathbb{X})} \leq 1$, we see that

$$|I_N(f, \mu_N, x)| \lesssim N^{S+\alpha}.$$

Thus, the number of bits needed to represent each $I_N(f, \mu_N, x)$ is at most $\log_2(c_1 N^{S+\alpha})$, where $c_1 \geq 1$ is a positive constant. Note that we can assume that $c_1 N^{S+\alpha} \geq 1$ because otherwise $I_N(f, \mu_N, x)$ would be zero. Since $\#\text{supp}(\nu_N) \lesssim N^\alpha$, we have $\#\{I_N(f, \mu_N, x) : x \in \text{supp}(\nu_N)\} \lesssim N^\alpha$. Therefore, the total number of bits needed to represent all numbers $\{I_N(f, \mu_N, x) : x \in \text{supp}(\nu_N)\}$ is at most $c_2 N^\alpha \log_2(c_1 N^{S+\alpha})$, where c_2 is a positive constant. By using $cN^{-s} = \varepsilon \leq 1$ and $\varepsilon \leq 1$, we derive that the number of necessary bits does not exceed

$$\begin{aligned} c_2 c^{\alpha/s} (1/\varepsilon)^{\alpha/s} \log_2(c_1 (c/\varepsilon)^{(S+\alpha)/s}) &\lesssim (1/\varepsilon)^{\alpha/s} \log_2((c_1)^{s/(S+\alpha)} c/\varepsilon) \\ &\lesssim (1/\varepsilon)^{\alpha/s} \log_2((c_1/\varepsilon)^{s/(S+\alpha)} c/\varepsilon) \\ &\lesssim (1/\varepsilon)^{\alpha/s} (1 + \log_2(1/\varepsilon)), \end{aligned}$$

which concludes the proof. \square

Theorem 4.1 yields that our bit representation scheme is optimal with respect to the metric entropy as stated in Theorem 2.6 at least up to a logarithmic factor.

5. Bit representation of locally smooth functions

In the previous section we derived a bit-representation scheme for the global approximation space, i.e., $B = \mathbb{X}$. It turns out that the case $B \subsetneq \mathbb{X}$ is more involved because we do not have results that characterize $\mathcal{A}^s(L_p(B))$ by means of σ_N . In fact, σ_N requires functions to be defined globally so that one would be forced to deal with boundary effects. On the other hand, B itself may not be a diffusion measure space satisfying all required assumptions. We circumvent such difficulties by dealing with a modified approximation space, for which we can construct a bit representation scheme.

5.1. Characterization of local smoothness by local approximation rates

Before we can discuss local smoothness, few technical details need to be introduced and we make use of $\mathcal{C}^\infty(\mathbb{X}) := \bigcap_{s>0} \mathcal{A}^s(L_\infty(\mathbb{X}))$:

Definition 5.1. We say that \mathbb{X} satisfies the *smooth cut-off property* if for any $s > 0$ and any two concentric balls B', B with $B' \subsetneq B$ there is $\phi \in \mathcal{C}^\infty(\mathbb{X})$ such that ϕ equals 1 on B' and ϕ vanishes outside of B with $|\phi(x)| \leq 1$, for all $x \in \mathbb{X}$.

Note that any smooth manifold satisfies the smooth cut-off property.

Definition 5.2. For $x \in \mathbb{X}$, the *local approximation space in x* is denoted by $\mathcal{A}^s(X_p(\mathbb{X}), x)$ and defined as the collection of $f \in X_p(\mathbb{X})$ such that there is an open ball B containing x with $f\phi \in \mathcal{A}^s(X_p(\mathbb{X}))$, for all $\phi \in \mathcal{C}^\infty(\mathbb{X})$ with support in B .

It turns out that the approximation rate of $\sigma_N(f, \mu_N)$ characterizes the approximation class of f , at least when switching to dyadic numbers $N = 2^n$, $n = 1, 2, \dots$, cf. [12,25,27,28]:

Theorem 5.3. Let \mathbb{X} satisfy the smooth cut-off property. For $p = \infty$, suppose that $(\mu_{2^n})_{n=1}^\infty$ is a family of Marcinkiewicz–Zygmund quadrature measures of order 2^n , respectively. For $1 \leq p < \infty$, we choose $\mu_{2^n} = \mu$, $n = 1, 2, \dots$. If H is a low-pass filter, then the following points are equivalent:

- (i) $f \in \mathcal{A}^s(X_p(\mathbb{X}), x)$,
- (ii) there is a ball B centered at x such that

$$\|f - \sigma_{2^n}(f, \mu_{2^n})\|_{L_p(B)} \lesssim 2^{-ns}. \quad (28)$$

Note that the generic constant in (28) may depend on x and f . Nonetheless, the above theorem characterizes local approximation by means of σ_{2^n} .

The local approximation space $\mathcal{A}^s(X_p(\mathbb{X}), x)$, for $x \in \mathbb{X}$, is not endowed with any norm. In view of Theorem 5.3, we fix some ball B and introduce a new approximation space in the following.

Definition 5.4. Let B be a nonempty ball in \mathbb{X} . For $p = \infty$, suppose that $(\mu_{2^n})_{n=1}^\infty$ is a family of Marcinkiewicz–Zygmund quadrature measures of order 2^n , respectively. For $1 \leq p < \infty$, we choose $\mu_{2^n} = \mu$, $n = 1, 2, \dots$. If H is a low-pass filter, then we define the *local approximation space in B* by

$$\mathcal{A}^s(X_p(\mathbb{X}), B) := \{f \in X_p(\mathbb{X}) : \|f\|_{\mathcal{A}^s(X_p(\mathbb{X}), B)} < \infty\},$$

where

$$\|f\|_{\mathcal{A}^s(X_p(\mathbb{X}), B)} := \|f\|_{L_p(\mathbb{X})} + \sup_{n \geq 1} 2^{ns} \|f - \sigma_{2^n}(f, \mu_{2^n})\|_{L_p(B)}.$$

Note that if $p = \infty$, then the space $\mathcal{A}^s(X_p(\mathbb{X}), B)$ implicitly depends on the family $(\mu_{2^n})_{n=1}^\infty$ of Marcinkiewicz–Zygmund quadrature measures of order 2^n , respectively. As opposed to $\mathcal{A}^s(X_p(B))$ defined in (5), the space $\mathcal{A}^s(X_p(\mathbb{X}), B)$ consists of functions defined globally that inherit approximation properties locally. By definition, we have

$$\|f - \sigma_{2^n}(f, \mu_{2^n})\|_{L_p(B)} \leq 2^{-ns} \|f\|_{\mathcal{A}^s(X_p(\mathbb{X}), B)}. \quad (29)$$

Since $\sigma_{2^n}(f, \mu_{2^n})$ is a diffusion polynomial, we observe that

$$\mathcal{A}^s(X_p(\mathbb{X}), B)|_B \hookrightarrow \mathcal{A}^s(X_p(B)).$$

However, we cannot claim that the reverse embedding also holds.

It should be mentioned that $\sigma_{2^n}(f, \mu_{2^n})$ in (29) approximates f locally but its definition needs global knowledge of f or at least on $\text{supp}(\mu_{2^n})$ if $p = \infty$. To enable the design of an approximation scheme that involves local information on f exclusively, we define one more approximation space by using some cut-off function:

Definition 5.5. For some fixed $\phi \in \mathcal{C}^\infty(\mathbb{X})$, define

$$\mathcal{A}^s(X_p(\mathbb{X}), \phi) := \{f \in X_p(\mathbb{X}) : f\phi \in \mathcal{A}^s(X_p(\mathbb{X}))\} \quad (30)$$

endowed with the norm $\|f\|_{\mathcal{A}^s(X_p(\mathbb{X}), \phi)} := \|f\|_{L_p(\mathbb{X})} + \sup_{n \geq 1} 2^{ns} E(f\phi, \Pi_{2^n}, L_p(\mathbb{X}))$.

To study local approximation, choose two concentric balls B', B with $B' \subsetneq B$. If \mathbb{X} satisfies the smooth cut-off property, then we can fix some $\phi \in \mathcal{C}^\infty(\mathbb{X})$ that is one on B' and zero outside of B . The Definition (30) yields that $f \in \mathcal{A}^s(X_p(\mathbb{X}), \phi)$ implies

$$\|f - \sigma_{2^n}(f\phi, \mu_{2^n})\|_{L_p(B')} \leq \|f\phi - \sigma_{2^n}(f\phi, \mu_{2^n})\|_{L_p(\mathbb{X})} \lesssim 2^{-ns} \|f\|_{\mathcal{A}^s(X_p(\mathbb{X}), \phi)}.$$

In the subsequent section, we shall consider balls of radius r for both spaces,

$$\overline{\mathcal{A}^s}(X_p(\mathbb{X}), B) := \{f \in X_p(\mathbb{X}) : \|f\|_{\mathcal{A}^s(X_p(\mathbb{X}), B)} \leq 1\}, \quad (31)$$

$$\overline{\mathcal{A}^s}(X_p(\mathbb{X}), \phi) := \{f \in X_p(\mathbb{X}) : \|f\|_{\mathcal{A}^s(X_p(\mathbb{X}), \phi)} \leq 1\}, \quad (32)$$

and aim to develop approximation schemes requiring only few bits.

5.2. Local bit-representation in approximation classes

To design an approximation scheme for the spaces $\overline{\mathcal{A}^s}(X_p(\mathbb{X}), B)$ and $\overline{\mathcal{A}^s}(X_p(\mathbb{X}), \phi)$, let B be a ball in \mathbb{X} and let $B' \subset B$ be another nonempty ball concentric with B and of radius strictly less. It will turn out that the following scheme enables us to approximate f on B' . For some fixed $S > 1$, we apply the quantization $I_{2^n}(f, \mu_{2^n}, x)$ as in (23) but in place of (24) we define the local approximation by

$$\sigma_{2^n}^\circ(f, \mu_{2^n}, \nu_{2^n}, B) := 2^{-nS} \int_B I_{2^n}(f, \mu_{2^n}, x) \sum_{k=0}^{\infty} H\left(\frac{\lambda_k}{2^{n+1}}\right) \varphi_k(x) d\nu_{2^n}(x) \varphi_k. \quad (33)$$

We have the following result for the ball $\overline{\mathcal{A}^s}(X_p(\mathbb{X}), B)$ of the localized approximation space given by (31):

Theorem 5.6. Suppose that \mathbb{X} satisfies the smooth cut-off property and that $(\mu_{2^n})_{n=1}^\infty$ is a family of Marcinkiewicz–Zygmund quadrature measures of order 2^n , respectively, if $p = \infty$. For $1 \leq p < \infty$ we choose $\mu_{2^n} = \mu$, $n = 1, 2, \dots$. Assume further that H is a low-pass filter. Let B, B' be two concentric balls, so that $B' \subsetneq B$. We also suppose that $(\nu_{2^n})_{n=1}^\infty$ are Marcinkiewicz–Zygmund quadrature measures of order 2^{n+1} with $\#\text{supp}(\nu_{2^n}) \lesssim 2^{n\alpha}$. For fixed $s > 0$ and $S > \max(1, s)$, we apply the discretizations (23) and (33). Then there is a constant $c > 0$ such that, for all $f \in \overline{\mathcal{A}^s}(X_p(\mathbb{X}), B)$,

$$\|f - \sigma_{2^n}^\circ(f, \mu_{2^n}, \nu_{2^n}, B)\|_{L_p(B')} \leq c 2^{-ns} \quad (34)$$

holds. For $c2^{-ns} = \varepsilon \leq 1$ and $\varepsilon \leq 1$, the number of bits needed to represent all integers $\{I_{2^n}(f, \mu_{2^n}, x) : x \in \text{supp}(v_{2^n}) \cap B\}$ does not exceed a positive constant (independent of ε) times

$$(1/\varepsilon)^{\alpha/s} (1 + \log_2(1/\varepsilon)). \quad (35)$$

Proof of Theorem 5.6. For $f \in \overline{\mathcal{A}^s}(X_p(\mathbb{X}), B)$, we use the localization property (12) with $r = \alpha + S$ and the embedding $L_p \hookrightarrow L_1$ in the compact case to derive, for $y \in B'$,

$$\begin{aligned} \int_{\mathbb{X} \setminus B} |\sigma_{2^n}(f, \mu_{2^n}, x) K_{2^{n+1}}(x, y)| d|v_{2^n}|(x) &\lesssim 2^{-nS} \|\sigma_{2^n}(f, \mu_{2^n})\|_{|v_{2^n}|, L_p} \\ &\lesssim 2^{-nS} \|\sigma_{2^n}(f, \mu_{2^n})\|_{|\mu_{2^n}|, L_p}. \end{aligned} \quad (36)$$

The latter estimate holds because both $(v_{2^n})_{n=1}^\infty$ and $(\mu_{2^n})_{n=1}^\infty$ are families of Marcinkiewicz–Zygmund measures. The quantization (23) immediately yields

$$|\sigma_{2^n}(f, \mu_{2^n}, x) - 2^{-nS} I_{2^n}(f, \mu_{2^n}, x)| \leq 2^{-nS}, \quad \text{for all } x \in \text{supp}(v_{2^n}). \quad (37)$$

By using (37), (22), and (19), we derive

$$\begin{aligned} \|\sigma_{2^n}(f, \mu_{2^n}) - \sigma_{2^n}^\circ(f, \mu_{2^n}, v_{2^n}, B)\|_{L_p(B')} \\ &= \|\sigma_{2^n}(f, \mu_{2^n}) - \int_B 2^{-nS} I_{2^n}(f, \mu_{2^n}, x) K_{2^{n+1}}(x, \cdot) dv_{2^n}(x)\|_{L_p(B')} \\ &\lesssim \|\sigma_{2^n}(f, \mu_{2^n}) - \int_B \sigma_{2^n}(f, \mu_{2^n}, x) K_{2^{n+1}}(x, \cdot) dv_{2^n}(x)\|_{L_p(B')} \\ &\quad + 2^{-nS}. \end{aligned}$$

Next, we make use of (22) and (36) with (18) to obtain

$$\begin{aligned} \|\sigma_{2^n}(f, \mu_{2^n}) - \sigma_{2^n}^\circ(f, \mu_{2^n}, B)\|_{L_p(B')} &\lesssim \left\| \int_{\mathbb{X} \setminus B} \sigma_{2^n}(f, \mu_{2^n}, x) K_{2^{n+1}}(x, \cdot) dv_{2^n}(x) \right\|_{L_p(B')} \\ &\quad + 2^{-nS} \\ &\lesssim 2^{-nS} \|f\|_{|\mu_{2^n}|, L_p} + 2^{-nS} \lesssim 2^{-nS}. \end{aligned}$$

Here, it is important that we assume $\mu_{2^n} = \mu$, for $1 \leq p < \infty$, so that $\|f\|_{|\mu_{2^n}|, L_p} \lesssim \|f\|_{L_p(\mathbb{X})}$ holds for the entire range $1 \leq p \leq \infty$. Note that $\|f\|_{L_p(\mathbb{X})} \leq 1$ follows from $f \in \overline{\mathcal{A}^s}(X_p(\mathbb{X}), B)$. The triangle inequality with (29) and the above estimate yield

$$\begin{aligned} \|f - \sigma_{2^n}^\circ(f, \mu_{2^n}, B)\|_{L_p(B')} &\lesssim \|f - \sigma_{2^n}(f, \mu_{2^n})\|_{L_p(B')} \\ &\quad + \|\sigma_{2^n}(f, \mu_{2^n}) - \sigma_{2^n}^\circ(f, \mu_{2^n}, v_{2^n}, B)\|_{L_p(B')} \\ &\lesssim 2^{-ns} + 2^{-nS} \lesssim 2^{-ns}, \end{aligned}$$

which verifies (34).

For the remaining part, we can follow the lines of the proof of Theorem 4.1. \square

Note that Theorem 5.6 still requires global knowledge of f because we need to build $\sigma_{2^n}(f, \mu_{2^n})$. For the sake of completeness, we use a cut-off function to feed in local information only:

Theorem 5.7. Under the same assumption as in Theorem 4.1, let $\phi \in \mathcal{C}^\infty(\mathbb{X})$ be one on B' and zero outside of B . Then there is a constant $c > 0$ such that, for $f \in \overline{\mathcal{A}^s}(X_p(\mathbb{X}), \phi)$,

$$\|f - \sigma_{2^n}^\circ(f\phi, \mu_{2^n}, v_{2^n})\|_{L_p(B')} \leq c2^{-ns} \quad (38)$$

holds. For $c2^{-ns} = \varepsilon \leq 1$, the number of bits needed to represent all integers $\{I_{2^n}(f\phi, \mu_{2^n}, x) : x \in \text{supp}(\nu_{2^n})\}$ does not exceed a positive constant (independent of ε) times

$$(1/\varepsilon)^{\alpha/s}(1 + \log_2(1/\varepsilon)). \quad (39)$$

Proof. The smooth cut-off property of ϕ yields that $f \in \overline{\mathcal{A}^s}(X_p(\mathbb{X}), \phi)$ leads to $f\phi \in \overline{\mathcal{A}^s}(X_p(\mathbb{X}))$, due to $|\phi(x)| \leq 1$, for all $x \in \mathbb{X}$. Thus, Theorem 4.1 applied to $f\phi$ implies (38) and (39). \square

Acknowledgments

M.E. has been funded by the Vienna Science and Technology Fund (WWTF) through project VRG12-009. The authors thank H.N. Mhaskar for many fruitful discussions.

Appendix. Summary of the technical assumptions

The asymptotic bounds on the metric entropy in Theorem 2.6 hold for any diffusion measure space \mathbb{X} as introduced in Definition 2.1 provided that the restrictions $\{\varphi_{k|B}\}_{k=0}^\infty$ are linearly independent on the ball $B \subset \mathbb{X}$. Our computational scheme that matches this bound up to a logarithmic factor needs few additional technical assumptions that are distributed within the present paper. Here, we list all the required assumptions for the sake of completeness:

- (I) \mathbb{X} is a diffusion measure space (Definition 2.1),
- (II) ϵ_N defined in (16) satisfies $N^m \epsilon_N \rightarrow 0$ as $N \rightarrow \infty$, for all $m > 0$,
- (III) there is a family $(\nu_N)_{N=1}^\infty$ of Marcinkiewicz–Zygmund quadrature measures of order $2N$, respectively, satisfying $\#\text{supp}(\nu_N) \lesssim N^\alpha$ (Definitions 3.2, 3.3 and (22)),
- (IV) the smooth cut-off property holds (Definition 5.1).

Condition (I) is the general framework, and the additional conditions are further technical details. Note that (IV) is only needed in Section 5, and it is well-known that this holds for smooth manifolds. All of the above conditions hold for compact homogeneous manifolds, e.g., the sphere and the Grassmann manifold, if the function system $\{\varphi_k\}_{k=0}^\infty$ are eigenfunctions of the Laplace operator, cf. [15]. Moreover, it was pointed out in [3] that the conditions are also satisfied for smooth compact Riemannian manifolds with nonnegative Ricci curvature, see also Remark 2.2. Families of Marcinkiewicz–Zygmund quadrature measures $(\mu_N)_{N=1}^\infty$ were then constructed in [13], such that $\#\text{supp}(\mu_N) \asymp N^\alpha$.

References

- [1] B. Carl, I. Stephani, *Entropy, Compactness and the Approximation of Operators*, Cambridge University Press, 1990.
- [2] C.K. Chui, F. Filbir, H.N. Mhaskar, Representation of functions on big data: Graphs and trees, *Appl. Comput. Harmon. Anal.* 38 (2015) 489–509.
- [3] C.K. Chui, H.N. Mhaskar, Smooth function extension based on high dimensional unstructured data, *Math. Comp.* 83 (290) (2014) 2865–2891.
- [4] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F.J. Warner, S.W. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data part. i: Diffusion maps, *Proc. Natl. Acad. Sci.* 102 (2005) 7426–7431.
- [5] P. de la Harpe, C. Pache, Cubature formulas, geometrical designs, reproducing kernels, and Markov operators, in: *Infinite Groups: Geometric, Combinatorial and Dynamical Aspects*, Vol. 248, Birkhäuser, Basel, 2005, pp. 219–267.
- [6] R.A. DeVore, R. Howard, C. Micchelli, Optimal nonlinear approximation, *Manuscripta Math.* 63 (4) (1989), 469–478.

- [7] R.A. DeVore, G.G. Lorentz, *Constructive Approximation*, Springer-Verlag, 1993.
- [8] D. Dung, Non-linear approximations using sets of finite cardinality or finite pseudo-dimension, *J. Complexity* 17 (2) (2001) 467–492.
- [9] D. Dung, T. Ullrich, n -widths and ε -dimensions for high-dimensional approximations, *Found. Comput. Math.* 13 (2013) 965–1003.
- [10] D.E. Edmunds, H. Triebel, *Function Spaces, Entropy Numbers, Differential Operators*, Cambridge University Press, 1996.
- [11] M. Ehler, F. Filbir, ε -coverings of Hölder-Zygmund type spaces on data-defined manifolds, in: *Abstract and Applied Analysis*, 2014.
- [12] F. Filbir, H.N. Mhaskar, A quadrature formula for diffusion polynomials corresponding to a generalized heat kernel, *J. Fourier Anal. Appl.* 16 (5) (2010) 629–657.
- [13] F. Filbir, H.N. Mhaskar, Marcinkiewicz–Zygmund measures on manifolds, *J. Complexity* 27 (6) (2011) 568–596.
- [14] S. Foucart, A. Pajor, H. Rauhut, T. Ullrich, The Gelfand widths of l_p -balls for $0 < p \leq 1$, *J. Complexity* 26 (2010) 629–640.
- [15] D. Geller, I.Z. Pesenson, Band-limited localized Parseval frames and Besov spaces on compact homogeneous manifolds, *J. Geom. Anal.* 21 (2) (2011) 334–371.
- [16] D. Geller, I.Z. Pesenson, n -widths and approximation theory on compact Riemannian manifolds, in: A. Mayeli, P.E.T. Jorgensen, G. Ólafsson (Eds.), *Commutative and Noncommutative Harmonic Analysis and Applications*, in: *Contemporary Mathematics*, vol. 603, 2013.
- [17] D. Geller, I.Z. Pesenson, Kolmogorov and linear widths of balls in Sobolev spaces on compact manifolds, *Math. Scand.* 115 (1) (2014) 96–122.
- [18] A. Grigor'yan, Estimates of heat kernels on riemannian manifolds, in: B. Davies, Yu. Safarov (Eds.), *Spectral Theory and Geometry*. ICMS Instructional Conference, Edinburgh, 1998, in: *London Math. Soc. Lecture Notes*, vol. 273, Cambridge Univ. Press, 1999, pp. 140–225.
- [19] D.D. Haroske, H. Triebel, *Distributions, Sobolev Spaces, Elliptic Equations*, EMS Publishing House, 2007.
- [20] A.N. Kolmogorov, Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse, *Ann. of Math.* 37 (1936) 107–111.
- [21] A.N. Kolmogorov, V.M. Tihomirov, ε -entropy and ε -capacity of sets in function spaces, *Uspekhi Mat. Nauk* 14 (1959) 3–86 English transl. in *Amer. Math. Soc. Transl.* 17, 277–364 MR0112032 (22:2890).
- [22] S.G. Krantz, H.R. Parks, *A Primer of Real Analytic Functions*, Birkhäuser, 1992.
- [23] G.G. Lorentz, Metric entropy and approximation, *Bull. Amer. Math. Soc.* 72 (6) (1966) 903–937.
- [24] G.G. Lorentz, M.v. Golitschek, Y. Makovoz, *Constructive Approximation, Advanced Problems*, Springer Verlag, New York, 1996.
- [25] M. Maggioni, H.N. Mhaskar, Diffusion polynomial frames on metric measure spaces, *Appl. Comput. Harmon. Anal.* 24 (3) (2008) 329–353.
- [26] H.N. Mhaskar, On the representation of smooth functions on the sphere using finitely many bits, *Appl. Comput. Harmon. Anal.* 18 (3) (2005) 215–233.
- [27] H.N. Mhaskar, Eignets for function approximation on manifolds, *Appl. Comput. Harmon. Anal.* 29 (2010) 63–87.
- [28] H.N. Mhaskar, A generalized diffusion frame for parsimonious representation of functions on data defined manifolds, *Neural Netw.* 24 (2011) 345–359.
- [29] C.A. Micchelli, T.J. Rivlin, A survey of optimal recovery, in: *Optimal Estimation in Approximation Theory*, Plenum Press, 1977, pp. 1–54 (Chapter).
- [30] B. Nadler, M. Galun, Fundamental limitations of spectral clustering, *Neural Inf. Process. Syst.* 19 (2007), 1017–1024.
- [31] E. Novak, Optimal recovery and n -widths for convex classes of functions, *J. Approx. Theory* 80 (3) (1995) 390–408.
- [32] A. Pinkus, *n -Widths in Approximation Theory*, Springer-Verlag, New York, 1985.
- [33] A. Pinkus, n -widths of Sobolev spaces in L^p , *Constr. Approx.* 1 (1985) 15–62.
- [34] T.J. Rivlin, *The Chebyshev Polynomials*, John Wiley and Sons, 1974.
- [35] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 290 (5500) (2000) 2323–2326.
- [36] B. Schölkopf, A. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.
- [37] V.N. Temlyakov, Estimates for the asymptotic characteristics of classes of functions with bounded mixed derivative or difference, *Proc. Steklov Inst. Math.* 189 (161–197) (1990).

- [38] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science* 290 (5500) (2000) 2319–2323.
- [39] J.F. Traub, H. Wozniakowski, *A General Theory of Optimal Algorithms*, Academic Press, New York, 1980.
- [40] A.G. Vitushkin, *Theory of the Transmission and Processing of Information*, Pergamon Press, 1961.