

SMOOTH FUNCTION EXTENSION BASED ON HIGH DIMENSIONAL UNSTRUCTURED DATA

CHARLES K. CHUI AND H. N. MHASKAR

ABSTRACT. Many applications, including the image search engine, image inpainting, hyperspectral image dimensionality reduction, pattern recognition, and time series prediction, can be facilitated by considering the given discrete data-set as a point-cloud \mathcal{P} in some high dimensional Euclidean space \mathbb{R}^s . Then the problem is to extend a desirable objective function f from a certain relatively smaller training subset $\mathcal{C} \subset \mathcal{P}$ to some continuous manifold $\mathbb{X} \subset \mathbb{R}^s$ that contains \mathcal{P} , at least approximately. More precisely, when the point cloud \mathcal{P} of the given data-set is modeled in the abstract by some unknown compact manifold embedded in the ambient Euclidean space \mathbb{R}^s , the extension problem can be considered as the interpolation problem of seeking the objective function on the manifold \mathbb{X} that agrees with f on \mathcal{C} under certain desirable specifications. For instance, by considering groups of cardinality s of data values as points in a point-cloud in \mathbb{R}^s , such groups that are far apart in the original spatial data domain in \mathbb{R}^1 or \mathbb{R}^2 , but have similar geometric properties, can be arranged to be close neighbors on the manifold. The objective of this paper is to incorporate the consideration of data geometry and spatial approximation, with immediate implications to the various directions of application areas. Our main result is a point-cloud interpolation formula that provides a near-optimal degree of approximation to the target objective function on the unknown manifold.

1. INTRODUCTION

With the recent rapid technological advance, significantly lower manufacturing cost, and high demand in such areas as: image sensor and capture, satellite and medical imaging, powerful search engines, convenient internet access, as well as low-cost wireless communication, memory devices, and computing power, the tremendously huge amount of data information to be processed and understood is overwhelming. One of the most popular current approaches is to consider a collection of all pieces of information as a point-cloud in some (unknown) manifold, of relatively low dimension, embedded in a high dimensional Euclidean space, with each point in the point-cloud representing a piece of the information data-set. For example, each point in the point-cloud could be an image thumbnail in application to an image search, where the dimension of the Euclidean space under consideration is the maximum resolution of the image thumbnail collection. Other important

Received by the editor February 17, 2012 and, in revised form, February 23, 2013.

2010 *Mathematics Subject Classification*. Primary 41A25, 42C15, 68Q32.

The first author's research was supported by ARO Grants # W911NF-07-1-0525 and # W911NF-11-1-0426.

The second author's research was supported, in part, by grant DMS-0908037 from the National Science Foundation and grant W911NF-09-1-0465 from the U.S. Army Research Office.

applications include document library sorting, hyperspectral image analysis, and feature pattern recognition.

In order to encode the similarities as well as dissimilarities among the data points, and more generally, local geometry of the data-set, it is convenient to assume that the point-cloud \mathcal{P} (at least approximately) lies on some continuous manifold $\mathbb{X} \subset \mathbb{R}^s$. This encoding is accomplished by inheriting some q -dimensional parametric representation of the manifold \mathbb{X} , where q denotes the (unknown) manifold dimension. Here, the manifold parametric representation could be determined by minimizing certain deviation from the data-set and the parameterization can be derived from some coordinate functions that are eigenvectors of the corresponding heat diffusion operator on the manifold \mathbb{X} (see [9]). Or equivalently, the manifold parametric representation could be determined by maximizing a certain variance of the data-set, and the parameterization can be derived from certain coordinate functions that reside in the nullspace of some Laplace-Beltrami operator on \mathbb{X} . Hence, there are two (dual) approaches to analyze and understand the point-cloud \mathcal{P} ; namely, by considering the (integral) kernels \mathcal{K} of these two types of operators, both governed by the given point-cloud \mathcal{P} . The interested reader is referred to our Handbook article [7] for a discussion and comparison of these approaches and corresponding nonlinear methods.

As a motivation to the following discussion, let us consider the widely popular ad hoc bilateral filter for image noise reduction with minimal image blurring, introduced by Tomasi and Manduchi [35] and extended by Elad [16], by incorporating a multiplicative radiometric filter component with the Gaussian convolution filter. While the Gaussian component of the bilateral filter has the capability of noise reduction by image smoothing, the radiometric component compares pixel similarities at pixel locations that could be spatially far apart to preserve image features, and hence avoiding excessive image blur. Observe that by considering the entire \mathbb{R}^2 as the image domain, then replacing the variance of the Gaussian function with a positive constant c multiple of some parameter t , the Gaussian convolution filter becomes the heat kernel of isotropic diffusion with time variable t and conductivity constant c . In our work [6], iterative Laplace-Beltrami operators for bilateral filtering are formulated when the radiometric parameter is replaced by some positive power of t ; and thus showing that bilateral filtering indeed leads to a certain anisotropic diffusion process.

For instance, to adopt the manifold approach to incorporate the similarities and dissimilarities of the data geometry, let us consider the example of lifting a data-set defined in \mathbb{R}^1 to some point-cloud in a higher dimensional space \mathbb{R}^s . A somewhat naive and yet effective way is to consider blocks of size s of consecutive data points. These blocks should be overlapping, and all data are supposed to belong to at least one block. In other words, each block is a point in the point cloud \mathcal{P} in \mathbb{R}^s , with coordinates given by the data values of the data points in the block. Hence, points corresponding to blocks that have similar geometry, except for some scaling factor (which is taken care of by the manifold parametric representation), are close neighbors. If ranking of similarities is desired to apply the theory and methods of approximation and interpolation of functions on manifolds to be introduced in the present paper, the anisotropic transform (AT) discussed in our recent work [8] can be applied. The AT is formulated in terms of certain decomposition of the matrix version of the (integral) kernels \mathcal{K} corresponding to the heat diffusion or

Laplace-Beltrami operators, as discussed in some detail in [8]. For fast computation, randomized projection is integrated with the AT (called RAT algorithms), and the trade-off between computational speed and possible inaccuracy (with arbitrarily small probability) is discussed in [8] and illustrated with experimental results for the method of Iso-map, introduced in [34].

Many practical applications such as pattern recognition, time series prediction, and missing data recovery, involve the estimation of an objective function defined on the point-cloud as discussed above. For example, in application to missing data recovery and repairing damaged data, the objective function could be defined at each point $x \in \mathcal{P}$ by the value of the data point at the center of the corresponding block x that constitutes the point in the point cloud. However, if some data is missing or damaged, this objective function is known only on a subset $\mathcal{C} \subset \mathcal{P}$. The question of missing/damaged data recovery is then the question of extending the objective function from \mathcal{C} to \mathcal{P} . This is parallel to the applications of manifold learning, where the number of points in the point cloud \mathcal{P} is too large to handle, and the standard approach is to extend the objective function defined only on some relatively smaller training subset $\mathcal{C} \subset \mathcal{P}$ to all of \mathcal{P} .

The important problem of extending a real-valued smooth function from a given set in a Euclidean space to the whole space so as to minimize a certain Sobolev-type norm has been investigated extensively by Fefferman and his collaborators [17–21]. In our study of contextual data completion in [5], we have also explored the question of extension of functions from the exterior of a sub-domain of a known manifold to the interior of this sub-domain. Let us classify such problem areas as smooth function extension from structured subsets to the entire domains, including manifolds, in order to distinguish our study in the present paper from the results in this vast literature. To the best of our knowledge, we do not know of any existing literature on the extension of functions from an unstructured finite set in a high dimensional Euclidean space to an (unknown) manifold of relatively lower dimension. Our point of view is that the study of smooth function extension based on unstructured data would require some interpolation scheme governed by certain approximation specifications. The main objective of this paper is to initiate a study of this problem by deriving an interpolation formula that possesses the approximation order determined by the desired smoothness function class in which the objective function lives.

As to the relevant existing literature, let us first mention that the well developed classical theory of approximation of functions is inadequate to capture the similarity and dissimilarities of data behavior that could be spatially far apart. For example, in application to time series prediction, similarity of data information of recurrent events taken in the past cannot be directly used effectively for current prediction. Similarly, in order to repair damaged music recording, it is definitely more effective if the recurring themes of the music are taken into consideration, as opposed to solely applying local approximation methods. Secondly, we would also like to mention that the existing literature of approximation theory on manifolds is not directly applicable to the study of point-clouds, since our consideration requires approximation on **unknown** manifolds. In particular, techniques based on special functions of some parametric representation of a known manifold are not applicable to point-cloud considerations. The referee of our paper has kindly pointed out the

papers [11, 14, 15] among many applications of diffusion geometry techniques, which do not require that the manifold in question be known beforehand.

The foundation of approximation theory on finite sets \mathcal{C} with the approximants defined on some unknown manifold \mathbb{X} that contains \mathcal{C} was only studied recently by the second author and his collaborators in [29, 31, 32], where the objective was to preserve the smoothness class to which the target functions are assumed to belong. However, since \mathbb{X} is not known, the classical definition of smoothness classes cannot be adopted directly. Nevertheless, the eigenfunctions (and eigenvalues) of the diffusion operator can be computed approximately. Therefore, along the line of classical approximation theory, the smoothness of a function f on \mathbb{X} at a point $\mathbf{x}_0 \in \mathbb{X}$ was defined in [29, 32] in terms of the degree of approximation of f from spaces of “diffusion polynomials” (see Section 2.3 for precise definitions), so that the behavior of the terms of some wavelet-like expansion of f in the neighborhood of a given point \mathbf{x}_0 completely determines the smoothness of f at \mathbf{x}_0 .

If the manifold is known and the smooth function class has been determined, a typical approach for solving this type of extension problems is to define the desired smooth function extension as the solution of some optimization problem, such as the minimization of some penalty functional or regularization pre-conditioning. For unknown manifolds, we choose the route of constructing certain linear operators which can be implemented as matrix vector multiplications by using pre-computed matrices to solve this ill-posed function extension problem. Our constructions are based either on the spectral information on the target function f or on the values of f on a subset $\mathcal{C} \subset \mathcal{P}$. We have proved the so-called “converse theorems” of approximation to show that our constructions are asymptotically optimal for every individual function in terms of the amount of information used to achieve a given degree of approximation. Of course, we may treat the unknown solution of the optimization problem as the target function f . In the context of approximation by diffusion polynomials, our constructions then provide a near-optimal solution to a suitably defined optimization problem, and we will illustrate the optimality nature of our extension with an example to be discussed in Section 4.

To apply the methods of diffusion maps, it is well known that the bottleneck is in the computation of eigenvalues and eigenfunctions of the diffusion operator. In [10], Coifman and Maggioni avoid this problem by pointing out that iterates of the diffusion operator can be used effectively to generate an approximate span of the eigenfunctions. This idea is explored by the second author in [32], by constructing a wavelet-like decomposition based solely on linear combinations of iterates of the diffusion operator, thereby obviating the need to compute eigenfunctions and eigenvalues. In the present paper, we will further develop the approach in [32] and introduce approximation methods to extend a $f : \mathcal{C} \rightarrow \mathbb{R}$ as a smooth function on \mathbb{X} . By assuming that the functions on the training set \mathcal{C} are values of some unknown smooth function f on \mathbb{X} , then our interpolant $\mathcal{I}(f)$ defined in (4.14) provides an optimal degree of approximation to this unknown function. The first stage of our construction of $\mathcal{I}(f)$ will be a discretized version of a linear combination of iterates of a suitable operator \mathcal{T} as in [32]. While the constructions in [32] require integration with respect to the volume measure of the unknown manifold \mathbb{X} , the method of construction in this present paper is based solely on the training data.

The idea behind our construction in this paper can be outlined as follows. Motivated by potential extensions of our theory to harmonic analysis on graphs, we

follow [10, 29, 32] and develop our theory in the more general set up of quasi-metric measure spaces. We first discretize the operators defined in [32] by applying the quadrature formulas developed in [22, 23]. This yields an approximation Q to the unknown function f . We will then use the operators from [32] again to interpolate the data $\{(x_k, y_k - Q(x_k))\}_{k=1}^M$ and add the resulting interpolant as a correction term to Q . The localization properties of the operators given in [32] ensure not only that such an interpolatory construction is possible, but also that the matrices involved are well conditioned, and the interpolation operator is bounded. Together with the approximation results in [31, 32], this leads to desired approximation bounds to be described in Theorem 3.1.

To summarize the main results in this paper, we will describe in Theorem 3.1, a recipe for converting bounds on the degrees of approximation from subsets of a normed linear space into a wavelet-like decomposition. This implies that for developing such a decomposition in various settings, we need only establish the bounds as required in Theorem 3.1. We will then obtain such bounds in four different settings. Theorem 4.1 is devoted to the case of quasi-interpolatory diffusion polynomial approximation operators based on information in the form of integrals with respect to a very general measure satisfying certain technical conditions described there. We would like to mention that, in particular, this is the analogue of the corresponding statement in [29, Eqn. (2.15)], where the information was in the form of a distinguished measure on the quasi-metric space. The other main results in Section 4 are Theorems 4.2, 4.3 and 4.4. Theorem 4.3 is the analogue of Theorem 4.1 where the operators involved are based on iterates of one member of a fixed family of operators. Theorems 4.2 and 4.4 are analogues of Theorems 4.1 and 4.3, respectively, where the approximation operators are also required to interpolate at the training data set. These theorems also assure that the collocation matrices involved in the interpolation are well conditioned.

2. PRELIMINARIES

To facilitate our presentation, we describe in this section the notations and basic concepts required throughout our paper. In Section 2.1, we summarize some well-known definitions about quasi-metric measure spaces, and review the Stieltjes integral notation for representing weighted sums. The basic assumptions on the quasi-metric spaces and measures that are commonly referred to throughout the paper are listed in Section 2.2. One of the main objectives of the paper is to develop wavelet-like decompositions of functions and to characterize smoothness (function) classes by the wavelet components. These smoothness classes will be introduced in Section 2.3. The wavelet-like decompositions are to be defined in terms of iterates of operators by using certain kernels. In this regard, we will need a variety of other kernels in the proofs of our main results. All of these will be defined in Section 2.4. In Section 2.5, we formulate some further terminology on the data-sets and associated measures to be used in the rest of this paper. A summary of the basic properties of the measures associated with the data-sets will also be discussed in Section 2.5.

2.1. Quasi-metric and measures. Let \mathbb{X} be a nonempty set. A *quasi-metric* on \mathbb{X} is a function $\rho : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ that satisfies the following properties: For all $x, y, z \in \mathbb{X}$,

- (1) $\rho(x, y) \geq 0$,
- (2) $\rho(x, y) = 0$ if and only if $x = y$,
- (3) $\rho(x, y) = \rho(y, x)$,
- (4) there exists a constant $\kappa > 0$ such that

$$(2.1) \quad \rho(x, y) \leq \kappa\{\rho(x, z) + \rho(z, y)\}, \quad x, y, z \in \mathbb{X}.$$

For example, the geodesic distance on a Riemannian manifold \mathbb{X} is a quasi-metric.

The quasi-metric ρ gives rise to a topology on \mathbb{X} , with

$$\{y \in \mathbb{X} : \rho(x, y) < r\}, \quad x \in \mathbb{X}, \quad r > 0.$$

being a basis for the topology. In the sequel, we will write

$$\mathbb{B}(x, r) = \{y \in \mathbb{X} : \rho(x, y) \leq r\}, \quad \Delta(x, r) = \mathbb{X} \setminus \mathbb{B}(x, r), \quad x \in \mathbb{X}, \quad r > 0.$$

With the topology in place, the notion of Borel measures is well defined. If ν is a (signed) Borel measure on \mathbb{X} , its total variation measure $|\nu|$ is defined by

$$|\nu|(B) = \sup_{\mathcal{U}} \sum_{U \in \mathcal{U}} |\nu(U)|,$$

where the supremum is taken over all partitions \mathcal{U} of B that consist of Borel sets. Here, the quantity $|\nu|(\mathbb{X})$ is called the total variation of ν . If ν is a signed measure, then its total variation is always finite. If ν is a positive measure, it is said to be of bounded variation if its total variation is finite. The measure ν is said to be complete if for any Borel set B with $\nu(B) = 0$ and any subset $A \subseteq B$, the so-called outer ν -measure of A is 0. Since any measure can be extended to a complete measure by suitably enlarging the underlying sigma-algebra, we will assume in the sequel that all the measures to be introduced in this paper are complete.

From the point of view of applications, we will be interested in a fixed probability measure μ on \mathbb{X} , such that the point-cloud may be thought of as a sample of \mathbb{X} with respect to μ . We will also be interested in measures of the following form. If $\mathcal{C} \subseteq \mathbb{X}$ is a finite set, the measure τ that associates with each $x \in \mathcal{C}$ the mass w_x , is defined by

$$\tau(B) = \sum_{x \in B} w_x$$

for the Borel subsets of \mathbb{X} . Obviously, the total variation of the measure τ is given by

$$|\tau|(B) = \sum_{x \in B} |w_x|.$$

If $f : \mathcal{C} \rightarrow \mathbb{C}$, then

$$\int_B f d\tau = \sum_{x \in \mathcal{C} \cap B} w_x f(x).$$

In theoretical analysis, it is necessary to consider a sequence of sets

$$\mathcal{C}_n = \{x_{n,1}, \dots, x_{n,M_n}\},$$

and measures τ_n associating a mass $w_{n,k}$ with each $x_{n,k} \in \mathcal{C}_n$. In this case, we have

$$\int_B f d\tau_n = \sum_{\substack{1 \leq k \leq M_n \\ x_{n,k} \in B}} w_{n,k} f(x_{n,k}).$$

The precise locations of the points of \mathcal{C}_n , or the number $w_{n,k}$, or even the numbers M_n will play no role in our theoretical development. The notation $\int_B f d\tau_n$

(or $\int_B f(x)d\tau_n(x)$) is a convenient way to denote the far more cumbersome sum expression on the right-hand side of the above formula. The measure notation will also help us to keep track of the various constants, as well as isolate the conditions on the point-set system $\{\mathcal{C}_n\}$ which are significantly germane to our analysis.

2.2. Assumptions. In the sequel, let μ be a fixed probability measure on \mathbb{X} . We fix a nondecreasing sequence $\{\ell_k\}_{k=0}^\infty$ of nonnegative numbers such that $\ell_0 = 0$, and $\ell_k \uparrow \infty$ as $k \rightarrow \infty$. Also, we fix a system of continuous and bounded functions $\{\phi_k\}_{k=0}^\infty$, orthonormal with respect to μ ; namely, for all nonnegative integers j, k ,

$$(2.2) \quad \int_{\mathbb{X}} \phi_k(x) \overline{\phi_j(x)} d\mu(x) = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise.} \end{cases}$$

We will assume in the sequel that $\phi_0(x) = 1$ for all $x \in \mathbb{X}$. For example, in the case of a point-cloud on a compact Riemannian manifold \mathbb{X} , we may consider the point-cloud to be a sample from \mathbb{X} taken according to a probability measure μ . In the case when this measure is the volume measure on \mathbb{X} , we may take ϕ_k 's to be the eigenfunctions of the Laplace–Beltrami operator on \mathbb{X} , corresponding to the eigenvalues $-\ell_k^2$. If the point-cloud is not sampled according to the volume measure, then we may need to consider differential operators other than the Laplace–Beltrami operator. Also, it is sometimes not necessary to use the exact eigenvalues of such operators. For example, in the case when \mathbb{X} is the unit sphere embedded in \mathbb{R}^3 , the eigenvalues of the Laplace–Beltrami operator are given by $-\sqrt{k(k+1)}$. The analysis is easier if we use $-k$ instead. In general, while the exact eigenvalues might be hard to compute, an asymptotic expression is often available. While these considerations motivate our definitions, we observe that we are considering a very general scenario with quasi-metric measure spaces, where differential operators are not defined.

In the theory of diffusion maps, the heat kernel plays a central role (cf. [1–3, 9, 10]). In particular, while the metric space \mathbb{X} , the measure μ , etc., are only models for the point-cloud, the heat kernel can be approximated well in terms of the data-set itself [9]. Accordingly, we will formulate our assumptions in terms of a formal heat kernel. The *heat kernel* on \mathbb{X} is defined formally by

$$(2.3) \quad K_t(x, y) = \sum_{k=0}^{\infty} \exp(-\ell_k^2 t) \phi_k(x) \phi_k(y), \quad x, y \in \mathbb{X}, t > 0.$$

Although K_t satisfies the semigroup property, and in light of the fact that $\ell_0 = 0$, $\phi_0(x) \equiv 1$, we have

$$(2.4) \quad \int_{\mathbb{X}} K_t(x, y) d\mu(y) = 1, \quad x \in \mathbb{X},$$

yet K_t may not be the heat kernel in the classical sense. In particular, we need not assume K_t to be nonnegative.

Definition 2.1. The set \mathbb{X} (more precisely, the quintuplet $(\mathbb{X}, \rho, \mu, \{\ell_k\}_{k=0}^\infty, \{\phi_k\}_{k=0}^\infty)$) is called a diffusion measure space with a quasi-metric (DIMSWM) if the following conditions are satisfied.

- (1) For each $x \in \mathbb{X}$ and $r > 0$, the ball $\mathbb{B}(x, r)$ is compact.
- (2) There exists $\alpha > 0$ and $\kappa_2 > 0$ such that the following power growth bound condition holds:

$$(2.5) \quad \mu(\mathbb{B}(x, r)) = \mu(\{y \in \mathbb{X} : \rho(x, y) < r\}) \leq \kappa_2 r^\alpha, \quad x \in \mathbb{X}, r > 0.$$

(3) With α as above, there exist $\kappa_3, \kappa_4 > 0$ such that the following Gaussian upper bound holds:

$$(2.6) \quad |K_t(x, y)| \leq \kappa_3 t^{-\alpha/2} \exp\left(-\kappa_4 \frac{\rho(x, y)^2}{t}\right), \quad x, y \in \mathbb{X}, \quad 0 < t \leq 1.$$

(4) With α as above, there exists $\kappa_5 > 0$ such that the following on–diagonal lower bound holds:

$$(2.7) \quad K_t(x, x) \geq \kappa_5 t^{-\alpha}, \quad x \in \mathbb{X}, \quad 0 < t < 1.$$

There is a great deal of discussion in the literature on the validity of the conditions in the above definition and their relationship with many other objects related to the quasi–metric space in question, (cf., for example, [12, 25–27]). In particular, it is shown in [12, Section 5.5] that any complete, connected Riemannian manifold with nonnegative Ricci curvature is a DIMSWM. It is shown in [28] that both of our assumptions on the heat kernel are valid in the case when \mathbb{X} is a complete Riemannian manifold with bounded geometry, and $\{\ell_j^2\}$, respectively $\{\phi_j\}$, are eigenvalues, respectively eigenfunctions, for a uniformly elliptic second order differential operator satisfying certain technical conditions.

We have shown in [22, Proposition 4.1] and [31, Lemma 5.2] that (2.6) with $y = x$ and (2.7) are equivalent to the following *spectral bounds* on the so–called Christoffel (or spectral) function (defined by the middle expression of (2.8) below):

$$(2.8) \quad c_1 N^\alpha \leq \sum_{\ell_j \leq N} |\phi_j(x)|^2 \leq c_2 N^\alpha, \quad x \in \mathbb{X}, \quad N > 0,$$

where c_1, c_2 are positive constants depending only on κ_2 . We have also proved in [22, Theorem 4.1] that (2.6) with $y \neq x$ is equivalent to the so–called finite speed of wave propagation. We refer to [22, 29] for a more detailed explanation of this property. Here, we wish only to remark that our assumptions (2.5) and (2.6) above are equivalent to the three assumptions made in [29].

The bounds on the heat kernel are closely connected with the measures of the balls $B(x, r)$. For example, it is proved in [27] that the conditions (2.4), (2.5), and (2.6) imply that there exists a constant $c > 0$ such that

$$(2.9) \quad \mu(B(x, r)) \geq cr^\alpha, \quad 0 < r \leq 1, \quad x \in \mathbb{X}.$$

In view of (2.5), this shows that μ satisfies the homogeneity condition

$$(2.10) \quad \mu(B(x, R)) \leq c_1 (R/r)^\alpha \mu(B(x, r)), \quad x \in \mathbb{X}, \quad r \in (0, 1], \quad R > 0,$$

where $c_1 > 0$ is a suitable constant. In many of the examples cited above, the kernel K_t also satisfies a lower bound to match the upper bound in (2.6). In this case, Grigoryán [27] has also shown that (2.5) is satisfied for $0 < r < 1$.

In the sequel, we assume that \mathbb{X} is a DIMSWM, and make the following convention.

Constant convention: *In the sequel, the symbols c, c_1, \dots will denote positive constants depending only on $\mathbb{X}, \rho, \mu, \kappa_1, \dots, \kappa_5$, and other similar fixed quantities. They will not depend upon the systems $\{\phi_k\}, \{\ell_k\}$ by themselves, except through the quantities mentioned above. On occasions when we need to have the constants depend upon additional variables, these will be listed explicitly. Their values may be different at different occurrences, even within a single formula. The notation $A \sim B$ will mean $c_1 A \leq B \leq c_2 A$.*

2.3. Function spaces and smoothness classes. For any Borel measure ν on \mathbb{X} , a Borel measurable subset $B \subseteq \mathbb{X}$, and a Borel measurable $f : \mathbb{B} \rightarrow \mathbb{C}$, we define

$$\|f\|_{\nu;B,p} := \begin{cases} \left\{ \int_B |f(x)|^p d|\nu|(x) \right\}^{1/p}, & \text{if } 0 < p < \infty, \\ |\nu| - \operatorname{ess\,sup}_{x \in B} |f(x)|, & \text{if } p = \infty. \end{cases}$$

With the convention that two functions are considered to be equal if they are equal $|\nu|$ -almost everywhere, the set of all ν -measurable functions f for which $\|f\|_{\nu;B,p} < \infty$ is denoted by $L^p(\nu;B)$. We will also write $L^p(\nu) = L^p(\nu;B)$ and $\|f\|_{\nu;p} = \|f\|_{\nu;\mathbb{X},p}$, etc. If $1 < p < \infty$, we define its conjugate by $p' = p/(p - 1)$ and the notation is extended, as usual, by defining $1' = \infty$, $\infty' = 1$. The space of all uniformly continuous, bounded functions $f : \mathbb{X} \rightarrow \mathbb{C}$, equipped with the norm $\|\cdot\|_\infty$, will be denoted by $C(\mathbb{X})$.

Using the sequences $\{\ell_k\}_{k=0}^\infty$ and $\{\phi_k\}_{k=0}^\infty$ given in Section 2.2, we define as in [22, 29],

$$\Pi_N := \operatorname{span}\{\phi_j : \ell_j \leq N\}, \quad N \geq 0,$$

and refer to an element of Π_N as a *diffusion polynomial* of degree at most N . We note that diffusion polynomials are not necessarily algebraic or trigonometric polynomials, and noninteger values of N are allowed. If $0 < p \leq \infty$ and $f \in L^p(\mu)$, we define the *degree of approximation* of f by

$$(2.11) \quad E_{N,p}(f) := \inf\{\|f - P\|_{\mu;p} : P \in \Pi_N\}.$$

For any $f \in L^\infty(\mu)$, the mapping $N \mapsto E_{N,p}(f)$ is a nonincreasing function of $N \in [0, \infty)$. The class of all functions f for which $E_{N,p}(f) \rightarrow 0$ as $N \rightarrow \infty$ will be denoted by X^p . Necessarily, $X^p \subseteq L^p(\mu)$. In most cases of interest, $X^p = L^p(\mu)$ if $1 \leq p < \infty$ and $X^\infty = C(\mathbb{X})$. The notation X^p is used to avoid the need of either to make this assumption tacitly (which leads to some problems, for example, in using the Riesz–Thorin interpolation theorem) or to state it over and over again. Also, since we are not assuming \mathbb{X} to be a compact manifold (without boundary), the spectrum of the Laplace–Beltrami operator may not be discrete. Hence, if we choose a countable subset of the spectrum as $\{\phi_k\}$, it will be not be true that $X^p = L^p(\mu)$, even for $p = 2$. Our notation avoids all these technical problems.

In the approximation theory literature, there are plenty of examples where the rate at which $E_{N,p}(f) \rightarrow 0$ as $N \rightarrow \infty$ characterizes the smoothness of the function f , measured in terms of specially designed “moduli of smoothness” (see, for example, [13, Chapters 6, 7, 8, 12]). In order to define the function spaces which we are interested in, we first define a sequence space. For a sequence $\mathbf{s} = \{s_n\}_{n=0}^\infty$ of complex numbers, and $a, \gamma > 0$, we write

$$(2.12) \quad \|\mathbf{s}\|_{\gamma,a} := \begin{cases} \left\{ \sum_{n=0}^\infty (2^{na}|s_n|)^\gamma \right\}^{1/\gamma}, & \text{if } 0 < \gamma < \infty, \\ \sup_{n \in \mathbb{Z}_+} 2^{na}|s_n|, & \text{if } \gamma = \infty, \end{cases}$$

where \mathbb{Z}_+ denotes the set of all nonnegative integers. The sequence \mathbf{s} is said to be in the space $\mathbf{b}_{\gamma,a}$ if $\|\mathbf{s}\|_{\gamma,a} < \infty$. It is easy to verify that the quantity $\|\cdot\|_{\gamma,a}$ is a quasi-norm on $\mathbf{b}_{\gamma,a}$ if $\gamma > 0$ and a norm if $\gamma \geq 1$. We define the *Besov space* $B_{p,\gamma}^a$ to be the class of all $f \in X^p$ such that $\{E_{2^n,p}(f)\}_{n=0}^\infty \in \mathbf{b}_{\gamma,a}$. In the general context of functions on a quasi-metric space, we do not have a notion of differentiable

functions. For convenience, in light of the vast literature in approximation theory in various special settings, we think intuitively of a function in $B_{p,\gamma}^a$ as being a -times differentiable.

2.4. Kernels and operators. Our paper will deal with a variety of different operators. All of these will have the form

$$f \mapsto \int_{\mathbb{X}} f(t)\Psi(\circ, t)d\nu(t)$$

for some suitable kernels Ψ and measures ν . Accordingly, we will often abuse the terminology, by referring to kernels Ψ and the corresponding operators interchangeably.

First, we define kernels which play the same role in our theory as the classical de la Vallée Poussin operators in the theory of Fourier series. If $H : [0, \infty) \rightarrow \mathbb{R}$ and $N > 0$, we define formally

$$(2.13) \quad \Phi_N(H, x, y) := \sum_{k=0}^{\infty} H\left(\frac{\ell_k}{N}\right) \phi_k(x)\phi_k(y).$$

If $f \in L^1(\mu)$, we define

$$(2.14) \quad \hat{f}(k) := \int_{\mathbb{X}} f(y)\phi_k(y)d\mu(y), \quad k = 0, 1, \dots$$

Corresponding to the formal kernel in (2.13), we have the formal operator for any measure ν on \mathbb{X} :

$$(2.15) \quad \sigma_N(\nu; H, f, x) := \int_{\mathbb{X}} \Phi_N(H, x, y)f(y)d\nu(y), \quad f \in L^1(\nu), N > 0, x \in \mathbb{X}.$$

It is convenient to define $\Phi_0(H, x, y) = \sum_{j:\ell_j=0} \phi_j(x)\phi_j(y)$, and

$$(2.16) \quad \sigma_0(\nu; H, f, x) := \int_{\mathbb{X}} \Phi_0(H, x, y)f(y)d\nu(y), \quad f \in L^1(\nu), x \in \mathbb{X}.$$

Many of our constructions here will be based on iterates of a family of kernels:

$$(2.17) \quad G_N(x, y) = \sum_{j=0}^{\infty} b\left(\frac{\ell_j}{N}\right) \phi_j(x)\phi_j(y), \quad x, y \in \mathbb{X},$$

where $b : [0, \infty) \rightarrow \mathbb{R}$, and the corresponding operator, defined formally by

$$(2.18) \quad \mathcal{T}_{G_N}(f, x) = \sum_{j=0}^{\infty} b\left(\frac{\ell_j}{N}\right) \hat{f}(j)\phi_j(x) = \int_{\mathbb{X}} f(y)G_N(x, y)d\mu(y), \quad f \in L^1(\mu), x \in \mathbb{X}.$$

For example, if $b(u) = \exp(-u^2t)$, the kernel G_N is the heat kernel at time t/N^2 .

Using (2.13) and (2.15) with b in place of H , we note that if $N > 0$, then for $x, y \in \mathbb{X}$ and $f \in L^1(\mu)$,

$$(2.19) \quad G_N(x, y) = \Phi_N(b, x, y), \quad \mathcal{T}_{G_N}(f, x) = \sigma_N(\mu; b, f, x).$$

Nevertheless, the notations G_N and \mathcal{T}_{G_N} are more natural for our context, rather than using the notations as in (2.19) for this purpose.

The technical conditions on the function b are given below.

Definition 2.2. Let $r \geq 1$ be an integer, and $b : \mathbb{R} \rightarrow [0, \infty)$. Then $b \in \mathcal{B}_r$ if the following conditions are satisfied.

- (1) The function b satisfies $b(u) = b(-u)$, $u \in \mathbb{R}$, $b(0) = 1$.
- (2) The function b is r -times continuously differentiable on \mathbb{R} .
- (3) There exists an even function $B : \mathbb{R} \rightarrow [0, \infty)$ such that B is nonincreasing on $[0, \infty)$, and each of the following conditions is satisfied (with α as in Definition 2.1):

$$(2.20) \quad |(1 + u)^j b^{(j)}(u)| \leq B(u), \quad u \in [0, \infty), \quad j = 0, \dots, r,$$

$$(2.21) \quad \lim_{u \rightarrow \infty} B(u)u^\alpha = 0,$$

and

$$(2.22) \quad \int_0^\infty B(u)u^{\alpha-1} du < \infty.$$

The function B will be called a \mathcal{B}_r majorant of b , with the notation

$$(2.23) \quad \|B\| := \left(\sup_{u \in \mathbb{R}} B(u) \right) + \int_0^\infty v^{\alpha-1} B(v) dv.$$

For example, $b(u) = (1 + u^2)^{-\beta/2}$, $\beta > \alpha$, satisfies these conditions for every $r \geq 1$ with $B(u) = c(1 + u^2)^{-\beta/2}$, where $c > 0$ is a constant depending only on β and r . With this choice of b , G_N is the Green’s function of a pseudo-differential operator defined for functions on \mathbb{X} . Similarly, if $b(u) = \exp(-tu^2)$ for some $t > 0$, then $b^{(j)}(u)$ has the form $P(u) \exp(-tu^2)$ for some polynomial P of degree j . Hence, b satisfies all these conditions for every $r \geq 1$ with $B(u) = c \exp(-tu^2/2)$ for some constant $c > 0$ depending on r and t . With this choice of b , G_N is the heat kernel K_t to be defined in Section 2.2. It is obvious that any compactly supported function satisfying the properties 1, 2 above is in \mathcal{B}_r . We observe that \mathcal{B}_r is a vector space for each $r \geq 1$.

We have proved in [32, Proposition 4.1] that if $b \in \mathcal{B}_1$, then the series in (2.17) converges absolutely and uniformly for all $x, y \in \mathbb{X}$. Hence, for each N , G_N is a uniformly continuous and bounded function in each of its variables.

If $g : [0, \infty) \rightarrow \mathbb{R}$, then the operator $g(\mathcal{T}_{G_N})$ is defined formally by

$$g(\mathcal{T}_{G_N})(f, x) = \sum_{j=0}^\infty g\left(b\left(\frac{\ell_j}{N}\right)\right) \hat{f}(j) \phi_j(x), \quad f \in L^1(\mu), \quad x \in \mathbb{X}.$$

The formal kernel corresponding to $g(\mathcal{T}_{G_N})$ will be denoted by $\Psi(g, G_N)$:

$$\Psi(g, G_N, x, y) = \sum_{j=0}^\infty g\left(b\left(\frac{\ell_j}{N}\right)\right) \phi_j(x) \phi_j(y), \quad x, y \in \mathbb{X}.$$

We will be particularly interested in the case when g is a polynomial.

2.5. Data-sets and associated measures. If $\mathcal{C} \subseteq \mathbb{X}$ is a finite set, its density content (i.e., mesh norm or fill distance) $\delta(\mathcal{C})$ and minimal separation $\eta(\mathcal{C})$ are defined by

$$(2.24) \quad \delta(\mathcal{C}) = \sup_{x \in \mathbb{X}} \rho(x, \mathcal{C}), \quad \eta(\mathcal{C}) = \min_{x, y \in \mathcal{C}, x \neq y} \rho(x, y).$$

It is clear that $\eta(\mathcal{C}) \leq 2\delta(\mathcal{C})$. The following proposition [31, Proposition 2.1] reveals the important connections among sequences of given data-sets and constructions of their subsequences with $\delta(\mathcal{C}) \sim \eta(\mathcal{C})$.

Proposition 2.1. (a) *If $\mathcal{C} \subset \mathbb{X}$ is a finite set and $\epsilon > 0$, there exists $\tilde{\mathcal{C}} \subseteq \mathcal{C}$ such that*

$$\delta(\mathcal{C}) \leq \delta(\tilde{\mathcal{C}}) \leq 2\delta(\mathcal{C}) \leq 2\eta(\tilde{\mathcal{C}}).$$

(b) *If $\mathcal{C}_0 \subseteq \mathcal{C}_1 \subset \mathbb{X}$ are finite subsets with $\delta(\mathcal{C}_1) \leq (1/2)\delta(\mathcal{C}_0) \leq \eta(\mathcal{C}_0)$, then there exists \mathcal{C}_1^* , with $\mathcal{C}_0 \subseteq \mathcal{C}_1^* \subseteq \mathcal{C}_1$, such that $\delta(\mathcal{C}_1) \leq \delta(\mathcal{C}_1^*) \leq 2\delta(\mathcal{C}_1) \leq 2\eta(\mathcal{C}_1^*)$.*

(c) *Let $\{\mathcal{C}_m\}$ be a sequence of finite subsets of \mathbb{X} , with $\delta(\mathcal{C}_m) \sim 1/m$, and $\mathcal{C}_m \subseteq \mathcal{C}_{m+1}$, $m = 1, 2, \dots$. Then there exists a sequence of subsets $\{\tilde{\mathcal{C}}_m \subseteq \mathcal{C}_m\}$, where, for $m = 1, 2, \dots$, $\delta(\tilde{\mathcal{C}}_m) \sim 1/m$, $\tilde{\mathcal{C}}_m \subseteq \tilde{\mathcal{C}}_{m+1}$, $\delta(\tilde{\mathcal{C}}_m) \leq 2\eta(\tilde{\mathcal{C}}_m)$.*

Next, we discuss measures associated with the data-sets.

Definition 2.3. (a) Let $d > 0$. A (signed or positive) Borel measure ν on \mathbb{X} is said to be d -regular if

$$(2.25) \quad \nu(B(x, d)) \leq cd^\alpha, \quad x \in \mathbb{X}.$$

The infimum of all constants c in (2.25) will be denoted by $\|\nu\|_{R,d}$.

(b) Let $n > 0$. A (signed or positive) Borel measure ν on \mathbb{X} is called an approximate quadrature measure of order n , if for every $R > 0$, and $P_1, P_2 \in \Pi_n$,

$$(2.26) \quad \left| \int_{\mathbb{X}} P_1 P_2 d\mu - \int_{\mathbb{X}} P_1 P_2 d\nu \right| \leq \frac{c(R)}{n^R} \|P_1\|_{\mu;1} \|P_2\|_{\mu;1}.$$

The properties of d -regular measures have been studied extensively in [23]. In particular, we have reconciled the various definitions of the term “ d -regular” in our different papers. For the convenience of the reader, we state various equivalent formulations, and a couple of other interesting facts about regular measures in the following proposition.

Proposition 2.2. (a) *If ν is d -regular, then for each $r > 0$ and $x \in \mathbb{X}$,*

$$(2.27) \quad |\nu|(B(x, r)) \leq c\|\nu\|_{R,d} \mu(B(x, r+d)) \leq c_1\|\nu\|_{R,d}(r+d)^\alpha.$$

Conversely, if for some $A > 0$, $|\nu|(B(x, r)) \leq A(r+d)^\alpha$ or each $r > 0$ and $x \in \mathbb{X}$, then ν is d -regular, and $\|\nu\|_{R,d} \leq 2^\alpha A$.

(b) *For each $\gamma > 1$,*

$$(2.28) \quad \|\nu\|_{R,\gamma d} \leq c_1(\gamma+1)^\alpha \|\nu\|_{R,d} \leq c_1(\gamma+1)^\alpha \gamma^\alpha \|\nu\|_{R,\gamma d},$$

where c_1 is the constant appearing in (2.27).

(c) *Let $\mathcal{C} \subset \mathbb{X}$ be a finite set, $\eta = \eta(\mathcal{C}) \leq 1$, and τ be a measure that associates the mass 1 with each $x \in \mathcal{C}$. Then τ is a η -regular measure, and $\|\nu\|_{R,\eta} \leq c\eta^{-\alpha}$, the constant being independent of η .*

The results in [23] are stated and proved for compact Riemannian manifolds where an additional condition on the gradient of the heat kernel is assumed. However, the statements listed above are valid in the generality in which we have stated them. For the convenience of the reader, we include a proof just before the end of Section 5. Finally, we observe that under additional conditions, we have established in [22, 31] the existence of $1/N$ -regular approximate quadrature measures of order N ; indeed, even the existence of positive measures satisfying these conditions.

3. BACKGROUND

In this paper, we are interested in developing a wavelet-like decomposition (also known as atomic decomposition or Littlewood–Paley expansion) of functions in X^p , especially in X^∞ , using diffusion polynomials, iterates of the operators \mathcal{T}_{G_N} , and their interpolatory analogues, so that the behavior of the terms in this expansion characterizes the smoothness class to which the function belongs. In Section 3.1, we develop a general theorem that reduces this development to estimates on the degrees of approximation. In the case of diffusion polynomials, the basis of these constructions is a family of localized kernels of the form $\Phi_n(h, x, y)$. The required localization estimates are given in Theorem 3.2 in Section 3.2. In contrast to the constructions of Coifman and Maggioni in [10], the results about degree of approximation are obtained using judicious linear combinations of the iterates of the operators \mathcal{T}_{G_N} ; i.e., operators of the form $Q_{n,r}(\mathcal{T}_{G_N})$ for properly selected (algebraic) polynomials $Q_{n,r}$ with appropriate parameters n and r (cf. (3.12) and (3.13) below). The choice of these parameters and construction of the polynomials is designed so that the kernels of the operators $Q_{n,r}(\mathcal{T}_{G_N})$ approximate the kernels $\Phi_n(h, x, y)$ sufficiently well, preserving the localization properties. The general definition of the polynomials $Q_{n,r}$ and the corresponding kernels is also given in Section 3.2.

3.1. Wavelet-like decomposition. Since our smoothness classes are defined in terms of degrees of approximation, the characterization of smoothness classes by a wavelet-like expansion can be done in great generality. We review the ideas in this subsection in this generality (cf. [30]).

Let \mathcal{X} be a normed linear space, $\mathcal{V}_n \subseteq \mathcal{V}_{n+1}$, $n = 0, 1, \dots$ be an increasing sequence of closed subsets of \mathcal{X} , such that $\bigcup_{n=0}^\infty \mathcal{V}_n$ is dense in \mathcal{X} . For $f \in \mathcal{X}$, we define

$$\text{dist}(f, \mathcal{V}_n) = \inf\{\|f - P\|_{\mathcal{X}} : P \in \mathcal{V}_n\}.$$

Theorem 3.1. *Let $m^* \geq 0$ be an integer, $0 < \gamma \leq \infty$, $0 < a < S$. Let $\{\mathcal{U}_n\}_{n=m^*}^\infty$ be a sequence of linear operators on \mathcal{X} such that for each $f \in \mathcal{X}$, and for $n = m^*, m^* + 1, \dots$,*

$$(3.1) \quad \begin{aligned} \text{dist}(f, \mathcal{V}_n) &\leq c \{ \|f - \mathcal{U}_n(f)\|_{\mathcal{X}} + 2^{-nS} \|f\|_{\mathcal{X}} \} \\ &\leq c_1 \{ \text{dist}(f, \mathcal{V}_{n-m^*}) + 2^{-nS} \|f\|_{\mathcal{X}} \}, \end{aligned}$$

for some constants $c, c_1 > 0$ independent of f or n . Then it follows that

$$(3.2) \quad f = \mathcal{U}_{m^*}(f) + \sum_{n=m^*}^\infty (\mathcal{U}_{n+1}(f) - \mathcal{U}_n(f)),$$

with convergence in the sense of \mathcal{X} . Moreover,

$$(3.3) \quad \{\text{dist}(f, \mathcal{V}_n)\}_{n=0}^\infty \in \mathbf{b}_{\gamma, a}$$

if and only if

$$(3.4) \quad \{\|\mathcal{U}_{n+1}(f) - \mathcal{U}_n(f)\|_{\mathcal{X}}\}_{n=m^*}^\infty \in \mathbf{b}_{\gamma, a}.$$

Proof. In view of our assumption that $\bigcup_{n=0}^\infty \mathcal{V}_n$ is dense in \mathcal{X} , the second estimate in (3.1) implies that

$$\limsup_{n \rightarrow \infty} \|f - \mathcal{U}_n(f)\|_{\mathcal{X}} \leq c \lim_{n \rightarrow \infty} \{ \text{dist}(f, \mathcal{V}_{n-m^*}) + 2^{-nS} \|f\|_{\mathcal{X}} \} = 0.$$

The expansion (3.2) is a restatement of this fact.

In view of the second estimate in (3.1), we have for integer $n \geq m^*$,

$$\begin{aligned} \|\mathcal{U}_{n+1}(f) - \mathcal{U}_n(f)\|_{\mathcal{X}} &\leq \|f - \mathcal{U}_{n+1}(f)\|_{\mathcal{X}} + \|f - \mathcal{U}_n(f)\|_{\mathcal{X}} \\ &\leq 2c \{ \text{dist}(f, \mathcal{V}_{n-m^*}) + 2^{-nS} \|f\|_{\mathcal{X}} \}. \end{aligned}$$

Hence, (3.3) implies (3.4).

Finally, let (3.4) hold. Using the first estimate in (3.1), we deduce that for integer $N \geq m^* + 1$,

$$\begin{aligned} \text{dist}(f, \mathcal{V}_N) &\leq c \{ \|f - \mathcal{U}_N(f)\|_{\mathcal{X}} + 2^{-NS} \|f\|_{\mathcal{X}} \} \\ &= c \left\{ \left\| \sum_{n=N}^{\infty} (\mathcal{U}_{n+1}(f) - \mathcal{U}_n(f)) \right\|_{\mathcal{X}} + 2^{-NS} \|f\|_{\mathcal{X}} \right\} \\ (3.5) \quad &\leq c \left\{ \sum_{n=N}^{\infty} \|\mathcal{U}_{n+1}(f) - \mathcal{U}_n(f)\|_{\mathcal{X}} + 2^{-NS} \|f\|_{\mathcal{X}} \right\}. \end{aligned}$$

We now recall that the discrete Hardy inequality [13, Lemma 3.4, p. 27] implies that if $\{a_j\}_{j=0}^{\infty} \in \mathbf{b}_{\gamma,a}$, then $\{\sum_{j=k}^{\infty} a_j\}_{k=0}^{\infty} \in \mathbf{b}_{\gamma,a}$ as well. Since

$$\{\|\mathcal{U}_{n+1}(f) - \mathcal{U}_n(f)\|_{\mathcal{X}}\}_{n=m^*}^{\infty} \in \mathbf{b}_{\gamma,a},$$

the estimate (3.5) leads to (3.3). □

Thus, in the sequel, in order to obtain a wavelet-like decomposition together with the characterization of smoothness classes for various operators, we need only to establish an estimate of the form (3.1) as in Theorem 3.1.

3.2. Approximation by certain operators. In the remainder of this paper, we assume that \mathbb{X} is a DIMSWM. The starting point in our analysis is the following theorem [32, Theorem 5.1]:

Theorem 3.2. *Let $r > \alpha + 3$ be an integer, $H \in \mathcal{B}_r$ and let \mathbb{H} be a \mathcal{B}_r majorant of H . Then for $N > 0$,*

$$(3.6) \quad |\Phi_N(H, x, y)| \leq c(r) \|\mathbb{H}\| \frac{N^\alpha}{\max(1, (N\rho(x, y))^{r-3})}, \quad x, y \in \mathbb{X},$$

where $\|\mathbb{H}\|$ is defined as in (2.23).

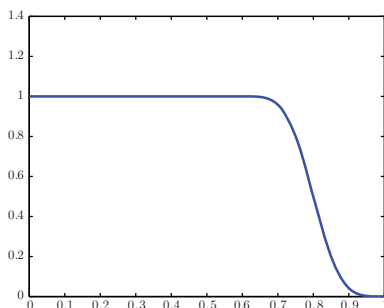
Throughout the rest of the paper, we will use the even function h in $C^\infty(\mathbb{R})$ as our low-pass window function, defined by

$$(3.7) \quad h(t) := \begin{cases} 1, & \text{if } 0 \leq t \leq 1/2, \\ \exp\left(-\frac{\exp(2/(1-2t))}{1-t}\right), & \text{if } 1/2 < t < 1, \\ 0, & \text{if } t \geq 1. \end{cases}$$

We note that h is nonincreasing on $(0, \infty)$ (cf. Figure 1).

Consequently, $h \in \mathcal{B}_{S+3}$ for all $S > 0$. Therefore, it follows from the above theorem that

$$(3.8) \quad |\Phi_N(h, x, y)| \leq c(S) \frac{N^\alpha}{\max(1, (N\rho(x, y))^S)}, \quad x, y \in \mathbb{X}, \quad N > 0.$$

FIGURE 1. The function h in (3.7) on $[0, 1]$.

Using this localization, we proved in [29, Theorem 4.2] that for every p with $1 \leq p \leq \infty$ and $f \in L^p(\mu)$,

$$(3.9) \quad \|\sigma_n(\mu; h, f)\|_{\mu; p} \leq c\|f\|_{\mu; p}, \quad n > 0.$$

The fact that $h(t) = 1$ if $|t| \leq 1/2$ can be used to deduce that $\sigma_n(\mu; h, P) = P$ for all $P \in \Pi_{n/2}$. Also, since $h(t) = 0$ if $|t| \geq 1$, $\sigma_n(\mu; h, f) \in \Pi_n$ for all $f \in L^1(\mu)$. Together with (3.9), these two facts lead to

$$(3.10) \quad \begin{aligned} E_{n,p}(f) &\leq \|f - \sigma_n(\mu; h, f)\|_p \leq cE_{n/2,p}(f), \\ f &\in L^p(\mu), \quad 1 \leq p \leq \infty, \quad n > 0. \end{aligned}$$

In particular, a wavelet-like decomposition can be constructed by using σ_{2^n} in place of \mathcal{U}_n in Theorem 3.1.

While the construction of the operator σ_n requires the knowledge of ℓ_k and ϕ_k , we obtained in [32] an approximation to this operator using iterates of the operator \mathcal{T}_{G_N} . To describe this approximation, we introduce some further notation.

For $r > -1$, and $n = 0, 1, \dots$, let $p_n^{(r,r)}$ denote the orthonormalized Jacobi (ultraspherical) polynomial of degree n such that for integers $n, m \geq 0$,

$$\int_{-1}^1 p_n^{(r,r)}(t)p_m^{(r,r)}(t)(1-t^2)^r dt = \begin{cases} 0, & \text{if } n \neq m, \\ 1, & \text{if } n = m. \end{cases}$$

For $n = 1, 2, \dots$, let $x_{n,n} < x_{n-1,n} < \dots < x_{1,n}$ be the zeros of $p_n^{(r,r)}$, and let the corresponding Cotes' numbers be defined by

$$w_{k,n} = \left\{ \sum_{j=0}^n p_j^{(r,r)}(x_{k,n})^2 \right\}^{-1}.$$

For $f \in C[-1, 1]$, $n = 1, 2, \dots$, and $0 \leq j \leq n$, let

$$(3.11) \quad a_{j,n,r}(f) = \sum_{k=1}^n w_{k,n} f(x_{k,n}) p_j^{(r,r)}(x_{k,n}),$$

and define the function \tilde{h} by $\tilde{h}(2b(u) - 1) = h(u)$. For $r > 0$, we define a sequence of polynomials $Q_{n,r}^*$ by

$$\begin{aligned}
 (3.12) \quad Q_{n,r}^*(t) = & t^{r+1} \left\{ \frac{(2r+1)!}{r!} \tilde{h}(1) \sum_{j=0}^r \frac{(-1)^j}{j!(r-j)!(j+r+1)} t^j \right. \\
 & + \sum_{k=1}^{n-1} (-1)^k h\left(\frac{k}{n}\right) a_{k,n;r}(\tilde{h}'/(1-\circ)^2)^r \\
 & \left. \times \left\{ \frac{2^{2r+1}(2k+2r+1)(k+2r)!}{k!} \right\}^{1/2} \sum_{j=0}^{k+r} \frac{(-1)^j (k+r+j)!}{j!(k+r-j)!(r+j+1)!} t^j \right\}.
 \end{aligned}$$

Finally, we set

$$(3.13) \quad Q_{n,r}(u) := Q_{n,r}^*(b(u)), \quad u \in \mathbb{R},$$

and observe that for $f \in L^1(\mu)$,

$$Q_{n,r}^*(\mathcal{T}_{G_N})(f) = \sigma_N(\mu; Q_{n,r}, f).$$

The following theorem (cf. [32, Theorem 6.1]) describes the approximation of the kernels $\Phi_N(h)$ and the corresponding operators by $\Phi_N(Q_{n,r})$ and the corresponding operators $Q_{n,r}(\mathcal{T}_{G_N})$.

Theorem 3.3. *Let $S > \alpha$, $R > r \geq S + 3$, let $n \geq 1$ be integers, $b \in \mathcal{B}_{R+1}$, let B be its \mathcal{B}_{R+1} majorant, and let b be strictly decreasing on $[0, \infty)$, and $x, y \in \mathbb{X}$. Then the quantity $\Phi_N(Q_{n,r}, x, y)$ is well defined. Moreover,*

$$\begin{aligned}
 (3.14) \quad & |\Phi_N(h, x, y) - \Phi_N(Q_{n,r}, x, y)| \\
 & \leq \frac{c(R) \|B^{r+1}\|}{n^{R-r}} \frac{N^\alpha}{\max(1, (N\rho(x, y))^S)}, \quad N \geq 1
 \end{aligned}$$

and

$$(3.15) \quad \sup_{x \in \mathbb{X}, N \geq 1} \int_{\mathbb{X}} |\Phi_N(h, x, y) - \Phi_N(Q_{n,r}, x, y)| d\mu(y) \leq \frac{c \|B^{r+1}\|}{n^{R-r}}.$$

Consequently, for $1 \leq p \leq \infty$,

$$(3.16) \quad \|\sigma_N(\mu; h, f) - Q_{n,r}^*(\mathcal{T}_{G_N})(f)\|_{\mu;p} \leq \frac{c_1 \|B^{r+1}\|}{n^{R-r}} \|f\|_p, \quad f \in L^p(\mu), \quad N \geq 1.$$

We observe the following corollary, which is immediate from (3.16) and (3.10). For notational simplicity, we will write

$$\begin{aligned}
 (3.17) \quad \sigma_{2^n}^*(\mu; f) &= Q_{2^n, S+3}^*(\mathcal{T}_{G_{2^n}})(f) \\
 &= \int_{\mathbb{X}} f(y) \Phi_{2^n}(Q_{2^n, S+3}, \circ, y) d\mu(y), \quad f \in L^1(\mu), \quad n \geq 0.
 \end{aligned}$$

More generally, if ν is any measure, we define

$$(3.18) \quad \sigma_{2^n}^*(\nu; f) = \int_{\mathbb{X}} f(y) \Phi_{2^n}(Q_{2^n, S+3}, \circ, y) d\nu(y), \quad f \in L^1(\nu), \quad n \geq 0.$$

Corollary 3.1. *Let $1 \leq p \leq \infty$, $f \in L^p(\mu)$, $n \geq 0$ be an integer. Then*

$$(3.19) \quad E_{2^n, p}(f) \leq c \{ \|f - \sigma_{2^n}^*(\mu; f)\|_{\mu;p} + 2^{-nS} \|f\|_{\mu;p} \} \leq c E_{2^{n-1}, p}(f).$$

In particular, a wavelet-like decomposition can be constructed by using $\sigma_{2^n}^*$ in place of \mathcal{U}_n in Theorem 3.1. By considering the operators $Q_{M2^n, S+3}^*(\mathcal{T}_{G_{2^n}})$ for a sufficiently large M , we may obtain frame properties as well, but this is not the focus for the present paper.

4. MAIN RESULTS

In this section, we state our main results, first for approximation and interpolation by diffusion polynomials, and then in the case when the various operators involved are replaced by their analogues based on the iterates of operators.

We remind the reader that \mathbb{X} is assumed to be a DIMSWM throughout the paper. Our main objective in this paper is to obtain analogues of Corollary 3.1 in the case when we have values of the target function rather than its Fourier- $\{\phi_k\}$ coefficients $\{\hat{f}(k)\}$. In this case, we would also like to achieve an interpolatory approximation; i.e., an approximation where the approximating operator reproduces the training data. As explained earlier, it is convenient to express these results in terms of integrals with respect to measures rather than weighted sums of the training data.

Theorem 4.1. *Let \mathbb{X} be a DIMSWM, $1 \leq p \leq \infty$, and $\{\nu_n\}$ be a sequence of Borel measures on \mathbb{X} such that each ν_n is a 2^{-n} -regular, approximate quadrature measure of order 2^n in the sense of Definition 2.3. Furthermore, assume that $\|\nu_n\|_{R, 2^{-n}} \leq c$, and $\|f\|_{\nu; p} \leq c\|f\|_{\mu; p}$ for all $f \in X^p(\mu)$. Then for $f \in X^p(\mu)$,*

$$(4.1) \quad E_{2^n, p}(f) \leq \|f - \sigma_{2^n}(\nu_n; f)\|_{\mu; p} \leq c\{E_{2^{n-1}, p}(f) + 2^{-nS}\|f\|_{\mu; p}\}.$$

Observe that we do not require the measures ν_n in the above theorem to be supported on a finite set. Nevertheless, the most interesting case for the purpose of this paper is when these sets are finite. In this case, we remind the reader that the operators $\sigma_{2^n}(\nu_n; f)$ take on the form

$$(4.2) \quad \sigma_{2^n}(\nu_n; f, x) = \sum_{m=1}^{M_n} w_{m,n} f(y_{m,n}) \Phi_n(h, x, y_{m,n}),$$

where $\mathcal{C}_n = \{y_{1,n}, \dots, y_{M_n,n}\} \subset \mathbb{X}$ and $w_{m,n}$'s are weights so that for the measure ν_n that associates the mass $w_{m,n}$ with each $y_{m,n}$ satisfies the properties listed in the above theorem. In particular, these operators can be evaluated as a matrix vector multiplication with pre-computed matrices. Of course, since function evaluations are not defined for the spaces $L^p(\mu)$, $1 \leq p < \infty$, Theorem 4.1 in this case is valid only for $f \in X^\infty \subseteq C(\mathbb{X})$.

In the case when the ν_n 's are supported on finite sets, we are able to achieve interpolatory approximation. Towards this end, let \mathcal{C}_n be the finite support of ν_n . Continuing the notation of the above paragraph, we may then construct the operator $\sigma_{2^n}(\nu_n; f)$ as in (4.2). These do not interpolate the data $(y_{m,n}, f(y_{m,n}))$; i.e., they do not have the property that $\sigma_{2^n}(\nu_n; f, y_{m,n}) = f(y_{m,n})$, $m = 1, \dots, M_n$, $n = 1, 2, \dots$. So, we add a correction term which interpolates the data $(y_{m,n}, f(y_{m,n}) - \sigma_{2^n}(\nu_n; f, y_{m,n}))$, and add this correction to $\sigma_{2^n}(\nu_n; f)$. The resulting operator, called *blending operator*, will interpolate as well as provide a near best approximation as required by Theorem 3.1, provided the interpolatory correction is obtained carefully. In particular, we need the collocation matrix to be not just invertible, but well conditioned. We remark that the interpolation and quasi-approximation operators are not commutative in the formulation of the blending operators, as shown

in [4] for the construction of local multivariate spline interpolants. The content of the following Theorem 4.2 is that our blending operators can be constructed by using diffusion polynomials in $\Pi_{2^{n+m^*}}$ for a constant m^* independent of f and n . We define a family of matrices $\mathbf{A}_{m,n}$ by

$$(4.3) \quad \mathbf{A}_{m,n}(x, y) = \Phi_{2^m}(h, x, y), \quad x, y \in \mathcal{C}_n,$$

and a sequence of column vectors

$$(4.4) \quad \mathbf{e}_n^d(f, x) = f(x) - \sigma_{2^n}(\nu_n; f, x), \quad x \in \mathcal{C}_n.$$

Theorem 4.2. *Let \mathbb{X} and $\{\nu_n\}$ be as in Theorem 4.1 (with $p = \infty$). In addition, let $\mathcal{C}_n = \text{supp}(\nu_n)$ be a finite set, satisfying*

$$(4.5) \quad B_1 2^{-n} \leq \eta(\mathcal{C}_n) \leq 2\delta(\mathcal{C}_n) \leq 4\eta(\mathcal{C}_n) \leq B_2 2^{-n}, \quad n \geq 1,$$

where B_1, B_2 are positive constants independent of n . There exists $c^* > 0$ such that each of the following statements hold.

(a) *Each of the matrices $\mathbf{A}_{n+c^*,n}$ is invertible, and its condition number (with respect to ℓ^∞ norms) satisfies*

$$(4.6) \quad \text{cond}(\mathbf{A}_{n+c^*,n}) \leq c.$$

(b) *Let $f \in X^\infty(\mu)$, and for $x \in \mathbb{X}$,*

$$(4.7) \quad \mathcal{I}_n^d(f, x) = \sigma_{2^n}(\nu_n; f, x) + \sum_{y \in \mathcal{C}_n} (\mathbf{A}_{n+c^*,n}^{-1} \mathbf{e}_n^d(f, y)) \mathbf{A}_{n+c^*,n}(x, y).$$

Then

$$(4.8) \quad \mathcal{I}_n^d(f, x) = f(x), \quad x \in \mathcal{C}_n$$

and

$$(4.9) \quad E_{2^{n+c^*},\infty} \leq \|f - \mathcal{I}_n^d(f)\|_\infty \leq c\{E_{2^{n-1},\infty}(f) + 2^{-nS}\|f\|_{\mu;\infty}\}.$$

In particular, we may obtain a wavelet-like decomposition based on the interpolatory operators $\{\mathcal{I}_n^d\}$ by applying Theorem 3.1. We note that the operators $\mathcal{I}_n^d(f)$ can also be evaluated as a linear combination of the values of f , by matrix-vector multiplication with pre-computed matrices.

We would like to make a remark here regarding an optimality property of the operators $\{\mathcal{I}_n^d\}$. In this paragraph only, let a pseudo-differential operator $\Delta^r(f)$ be defined formally by

$$\widehat{\Delta^r(f)}(k) = (1 + \ell_k^2)^{r/2} \hat{f}(k), \quad f \in L^1(\mu), \quad k = 0, 1, \dots, \quad r \in \mathbb{R}.$$

Following the proof of [31, Proposition 5.3(b)], it is easy to show that (4.9) implies that

$$\|\Delta^r(f) - \Delta^r(\mathcal{I}_n^d(f))\|_\infty \leq c\{E_{2^{n-1},\infty}(\Delta^r(f)) + 2^{-nS}\|\Delta^r(f)\|_{\mu;\infty}\}.$$

In particular,

$$(4.10) \quad \|\Delta^r(\mathcal{I}_n^d(f))\|_\infty \leq c\|\Delta^r(f)\|_{\mu;\infty}.$$

Thus, writing $z = f(x)$, if one considers as usual the minimization problem

$$g = \arg \min\{\|\Delta^r(g)\|_{\mu;\infty} : g(x) = z, \quad x \in \mathcal{C}_n\},$$

then $\mathcal{I}_n^d(f) = \mathcal{I}_n^d(g)$ (since both these expressions depend only the data $\{(x, z)\}_{x \in \mathcal{C}_n}$), and (4.10) shows that the extension $\mathcal{I}_n^d(f)$, **obtained without having to solve any optimization problem**, still satisfies

$$\|\Delta^r(\mathcal{I}_n^d(f))\|_\infty \leq c \min\{\|\Delta^r(g)\|_{\mu; \infty} : g(x) = z, x \in \mathcal{C}_n\}.$$

The analogues of the above two theorems in the case when the approximation is based on iterates of operators rather than diffusion polynomials can be obtained by applying Theorem 3.3. To state these results, we define the quantities \mathbf{e}_n , and $\mathbf{B}_{m,n}$ as follows:

$$(4.11) \quad \begin{aligned} \mathbf{e}_n(f, x) &= f(x) - \sigma_{2^n}^*(\nu_n; f, x), & x \in \mathcal{C}_n, \\ \mathbf{B}_{m,n} &= \Phi_{2^m}(Q_{2^n, S+3}, x, y), & x, y \in \mathcal{C}_n. \end{aligned}$$

As before, in the definition (3.18) of $\sigma_{2^n}^*(\nu_n; f)$, we need not assume that ν_n is finitely supported. However, in (4.11), ν_n is assumed to be finitely supported, and \mathcal{C}_n is assumed to be the finite support of ν_n .

Theorem 4.3. *Let $b \in \mathcal{B}_{2S+3}$, B be its \mathcal{B}_{2S+3} majorant, and let b be strictly decreasing on $[0, \infty)$. Under the same assumptions as stated in Theorem 4.1, for $f \in X^p(\mu)$,*

$$(4.12) \quad \begin{aligned} E_{2^n, p}(f) &\leq c \{ \|f - \sigma_{2^n}^*(\nu_n; f)\|_{\mu; p} + 2^{-nS} \|f\|_{\mu; p} \} \\ &\leq c \{ E_{2^{n-1}, p}(f) + 2^{-nS} \|f\|_{\mu; p} \}. \end{aligned}$$

Theorem 4.4. *Let $b \in \mathcal{B}_{2S+3}$, B be its \mathcal{B}_{2S+3} majorant, and b be strictly decreasing on $[0, \infty)$. Under the same assumptions as stated in Theorem 4.2, there exist constants $c^*, c_1^* > 0$ such that the following statements hold for $n \geq c_1^*$.*

(a) *Each of the matrices $\mathbf{B}_{n+c^*, n}$ is invertible, and its condition number (with respect to ℓ^∞ norms) satisfies*

$$(4.13) \quad \text{cond}(\mathbf{B}_{n+c^*, n}) \leq c.$$

(b) *Let $f \in X^\infty(\mu)$, and for $x \in \mathbb{X}$,*

$$(4.14) \quad \mathcal{I}_n(f, x) = \sigma_{2^n}^*(\nu_n; f, x) + \sum_{y \in \mathcal{C}_n} (\mathbf{B}_{n+c^*, n}^{-1} \mathbf{e}_n(f, y)) \mathbf{B}_{n+c^*, n}(x, y).$$

Then

$$(4.15) \quad \mathcal{I}_n(f, x) = f(x), \quad x \in \mathcal{C}_n,$$

and

$$(4.16) \quad \begin{aligned} E_{2^n+c^*, \infty} &\leq c \{ \|f - \mathcal{I}_n(f)\|_\infty + 2^{-nS} \|f\|_{\mu; \infty} \} \\ &\leq c \{ E_{2^{n-1}, \infty}(f) + 2^{-nS} \|f\|_{\mu; \infty} \}. \end{aligned}$$

5. PROOFS OF THE MAIN RESULTS

The starting point of our proofs is the following proposition (cf. [31, Proposition 5.1(a)]).

Proposition 5.1. *Let $d > 0$ and ν be a d -regular measure. If $g_1 : [0, \infty) \rightarrow [0, \infty)$ is a nonincreasing function, then for any $N > 0, r > 0, x \in \mathbb{X}$,*

$$(5.1) \quad N^\alpha \int_{\Delta(x, r)} g_1(N\rho(x, y)) d|\nu|(y) \leq \frac{2^\alpha (c + (d/r)^\alpha) \alpha}{1 - 2^{-\alpha}} \|\nu\|_{R, d} \int_{rN/2}^\infty g_1(u) u^{\alpha-1} du.$$

In particular,

$$(5.2) \quad N^\alpha \int_{\Delta(x,r)} \frac{d|\nu|(y)}{\max(1, (N\rho(x,y))^S)} \leq \frac{c_1(c + (d/r)^\alpha)\alpha}{\max(1, (rN/2)^{S-\alpha})} \|\nu\|_{R,d}$$

and

$$(5.3) \quad N^\alpha \int_{\mathbb{X}} \frac{d|\nu|(y)}{\max(1, (N\rho(x,y))^S)} \leq c_1(c + (Nd)^\alpha) \|\nu\|_{R,d}.$$

We recall that if $P \in \Pi_{2^{n-1}}$, then $\sigma_{2^n}(\mu; h, P) = P$. If ν is only an approximate quadrature measure, this may not be true if we replace μ by ν . Our next objective is to approximate $P \in \Pi_{2^{n-1}}$ by $\sigma_{2^n}(\nu; h, P)$ and $\sigma_{2^n}(\nu; Q_{2^n, S+3}, P)$.

Lemma 5.1. *Let $b \in \mathcal{B}_{2S+3}$, B be its \mathcal{B}_{2S+3} majorant, and let b be strictly decreasing on $[0, \infty)$. Under the same assumptions as stated in Theorem 4.1, the following statements hold.*

(a) For $f \in X^p(\mu)$ and $n \geq 0$,

$$(5.4) \quad \|\sigma_{2^n}(\nu_n; h, f)\|_{\mu;p} \leq c\|f\|_{\nu_n;p} \leq c\|f\|_{\mu;p}$$

and

$$(5.5) \quad \|\sigma_{2^n}(\nu_n; h, f) - \sigma_{2^n}^*(\nu_n; f)\|_{\mu;p} \leq \frac{c\|B^{S+4}\|}{2^{nS}}\|f\|_{\nu_n;p} \leq \frac{c\|B^{S+4}\|}{2^{nS}}\|f\|_{\mu;p}.$$

(b) For $n \geq 1$ and $P \in \Pi_{2^{n-1}}$,

$$(5.6) \quad \|P - \sigma_{2^n}(\nu_n; h, P)\|_{\mu;\infty} \leq c2^{-nS}\|P\|_{\mu;1}$$

and

$$(5.7) \quad \|P - \sigma_{2^n}^*(\nu_n; P)\|_{\mu;p} \leq c2^{-nS}(1 + \|B^{S+4}\|)\|P\|_{\mu;p}.$$

Proof. We use the localization estimate (3.8) to obtain

$$(5.8) \quad |\Phi_{2^n}(h, x, y)| \leq \frac{c2^{n\alpha}}{\max(1, (2^n\rho(x, y))^S)}, \quad x, y \in \mathbb{X}.$$

Since ν_n is 2^{-n} -regular, and $\|\nu_n\|_{R,2^{-n}} \leq c$, we may use (5.3) with $N = 2^n$, $d = 2^{-n}$ to conclude from (5.8) that

$$(5.9) \quad \int_{\mathbb{X}} |\Phi_{2^n}(h, x, y)| d|\nu_n|(y) \leq c.$$

Similarly, since μ is 2^{-n} -regular, and $\|\mu\|_{R,2^{-n}} \leq c$ for every n , we have

$$(5.10) \quad \int_{\mathbb{X}} |\Phi_{2^n}(h, x, y)| d\mu(y) \leq c.$$

The estimate (5.4) follows from (5.9), (5.10), and Young’s inequality (cf. [32, Lemma 5.3]).

Next, we use (3.14) with $r = S + 3$, $R = 2S + 3$, $N = 2^n$, and 2^n in place of n to conclude that

$$(5.11) \quad \begin{aligned} & |\Phi_{2^n}(h, x, y) - \Phi_{2^n}(Q_{2^n, S+3}, x, y)| \\ & \leq \frac{c\|B^{S+4}\|}{2^{nS}} \frac{2^{n\alpha}}{\max(1, (2^n\rho(x, y))^S)}, \quad x, y \in \mathbb{X}. \end{aligned}$$

This leads to (5.5) in exactly the same way as (5.8) leads to (5.4). This completes the proof of part (a).

To prove part (b), let $P \in \Pi_{2^n-1}$, and $x \in \mathbb{X}$. Then $\sigma_{2^n}(\mu; h, P) = P$. Since ν_n is an approximate quadrature measure of order 2^n , we may use the definition (2.26) with P in place of P_1 , $\Phi_n(h, x, \circ) \in \Pi_{2^n}$, and S in place of R to obtain

$$\begin{aligned} & |P(x) - \sigma_{2^n}(\nu_n; h, P, x)| \\ (5.12) \quad &= \left| \int_{\mathbb{X}} P(y) \Phi_{2^n}(h, x, y) d\mu(y) - \int_{\mathbb{X}} P(y) \Phi_{2^n}(h, x, y) d\nu_n(y) \right| \\ &\leq \frac{c}{2^n S} \|P\|_{\mu;1} \|\Phi_{2^n}(h, x, \circ)\|_{\mu;1}. \end{aligned}$$

Together with (5.10), this implies (5.6). Since μ is a probability measure, (5.6) and (5.5) together lead to (5.7). \square

In the sequel, we consider $b \in \mathcal{B}_{2S+3}$ and B to be fixed, and the constants may depend upon these.

Proof of Theorem 4.1. Since $\sigma_{2^n}(\nu_n; h, f) \in \Pi_{2^n}$, the first inequality in (4.1) is clear. To prove the second inequality, let $P \in \Pi_{2^n-1}$ satisfy

$$\|f - P\|_{\mu;p} \leq 2E_{2^n-1,p}(f).$$

Then using (5.4) and (5.6), we deduce that

$$\begin{aligned} (5.13) \quad & \|f - \sigma_{2^n}(\nu_n; h, f)\|_{\mu;p} = \|f - P - \sigma_{2^n}(\nu_n; h, f - P) + P - \sigma_{2^n}(\nu_n; h, P)\|_{\mu;p} \\ &\leq \|f - P\|_{\mu;p} + \|\sigma_{2^n}(\nu_n; h, f - P)\|_{\mu;p} + \|P - \sigma_{2^n}(\nu_n; h, P)\|_{\mu;p} \\ &\leq c\|f - P\|_{\mu;p} + \frac{c_1}{2^n S} \|P\|_{\mu;p} \\ &\leq c \left\{ E_{2^n-1,p}(f) + \frac{1}{2^n S} \|f\|_{\mu;p} \right\}. \end{aligned}$$

This completes the proof of the second inequality in (4.1). \square

Since this preparation is sufficient to prove Theorem 4.3, we prove it now, postponing the proof of Theorem 4.2, which requires some further preparation.

Proof of Theorem 4.3. Using (4.1) and (5.5), we obtain that

$$\begin{aligned} E_{2^n;p}(f) &\leq \|f - \sigma_{2^n}(\nu_n; h, f)\|_{\mu;p} \\ &\leq \|f - \sigma_{2^n}^*(\nu_n; f)\|_{\mu;p} + \|\sigma_{2^n}^*(\nu_n; f) - \sigma_{2^n}(\nu_n; h, f)\|_{\mu;p} \\ &\leq \|f - \sigma_{2^n}^*(\nu_n; f)\|_{\mu;p} + \frac{c}{2^n S} \|f\|_{\mu;p}. \end{aligned}$$

This proves the first inequality in (4.12). The second inequality is proved easily by applying (4.1) and (5.5) :

$$\begin{aligned} (5.14) \quad & \|f - \sigma_{2^n}^*(\nu_n; f)\|_{\mu;p} \\ &\leq \|f - \sigma_{2^n}(\nu_n; h, f)\|_{\mu;p} + \|\sigma_{2^n}(\nu_n; h, f) - \sigma_{2^n}^*(\nu_n; f)\|_{\mu;p} \\ &\leq c \left\{ E_{2^n-1,p}(f) + \frac{1}{2^n S} \|f\|_{\mu;p} \right\}. \quad \square \end{aligned}$$

In order to prove the theorems concerning interpolatory approximation, we recall a couple of well-known results from linear algebra. The first of these is an analogue of the Gershgorin theorem (cf. [31, Proposition 6.1]). We will show that the matrices $\mathbf{A}_{m,n}$ satisfy the conditions of Proposition 5.2 if $m \geq n + c^*$.

Proposition 5.2. *Let $M \geq 1$ be an integer and \mathbf{A} an $M \times M$ matrix whose (i, j) -th entry is $A_{i,j}$. $1 \leq p \leq \infty$, and $\gamma \in [0, 1)$. If*

$$(5.15) \quad \sum_{\substack{i=1 \\ i \neq j}}^M |A_{j,i}| \leq \gamma |A_{j,j}|, \quad \sum_{\substack{i=1 \\ i \neq j}}^M |A_{i,j}| \leq \gamma |A_{j,j}|, \quad j = 1, \dots, M,$$

and $\lambda = \max_{1 \leq i \leq M} |A_{i,i}| > 0$, then \mathbf{A} is invertible, and

$$(5.16) \quad \|\mathbf{A}^{-1}\mathbf{z}\|_{\ell^p} \leq ((1 - \gamma)\lambda)^{-1} \|\mathbf{z}\|_{\ell^p}, \quad \mathbf{z} \in \mathbb{C}^M.$$

The second fact is a restatement of [24, Theorem 2.3.4], which will help us to carry out a perturbation argument to derive results about $\mathbf{B}_{m,n}$ from those about $\mathbf{A}_{m,n}$.

Proposition 5.3. *Let \mathbf{A} and \mathbf{B} be matrices such that \mathbf{A} is invertible, and for some $r \in (0, 1)$,*

$$(5.17) \quad \|\mathbf{A} - \mathbf{B}\| \leq \frac{r}{\|\mathbf{A}^{-1}\|}.$$

Then \mathbf{B} is invertible, and

$$(5.18) \quad \|\mathbf{A}^{-1} - \mathbf{B}^{-1}\| \leq \frac{r}{1 - r} \|\mathbf{A}^{-1}\|.$$

Proof. This is a simple consequence of [24, Theorem 2.3.4]. To apply this theorem, we note that (5.17) leads to

$$\|\mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\| \leq r < 1. \quad \square$$

It is convenient to formulate certain calculations needed in deriving estimates on the matrices in a separate lemma, as follows.

Lemma 5.2. *Let $\mathcal{C} \subset \mathbb{X}$ be a finite set, and $\eta = \eta(\mathcal{C})$. Then for $N \geq (\eta/4)^{-1}$ and $x \in \mathbb{X}$,*

$$(5.19) \quad \sum_{\substack{y \in \mathcal{C} \\ \rho(x,y) \geq \eta/2}} \frac{N^\alpha}{\max(1, (N\rho(x,y))^S)} \leq c(\eta N)^{\alpha-S} \eta^{-\alpha}.$$

In particular, for $x \in \mathcal{C}$,

$$(5.20) \quad \sum_{\substack{y \in \mathcal{C} \\ x \neq y}} \frac{N^\alpha}{\max(1, (N\rho(x,y))^S)} \leq c(\eta N)^{\alpha-S} \eta^{-\alpha}.$$

Furthermore, for $x \in \mathbb{X}$,

$$(5.21) \quad \sum_{y \in \mathcal{C}} \frac{N^\alpha}{\max(1, (N\rho(x,y))^S)} \leq c(c_1 + (N\eta)^\alpha) \eta^{-\alpha}.$$

Proof. Let τ be the measure that associates the mass 1 with each element of \mathcal{C} . In view of Proposition 2.2(c), $\|\tau\|_{R,\eta} \leq c\eta^{-\alpha}$. Since $N\eta/2 \geq 2$, we may apply the estimate (5.2) in Proposition 5.1 with $r = \eta/2$ and $d = \eta$ to obtain for $x \in \mathbb{X}$,

$$(5.22) \quad \sum_{\substack{y \in \mathcal{C} \\ \rho(x,y) \geq \eta/2}} \frac{N^\alpha}{\max(1, (N\rho(x,y))^S)} = \int_{\Delta(x,\eta/2)} \frac{N^\alpha}{\max(1, (N\rho(x,y))^S)} d\tau(y) \leq c(\eta N)^{\alpha-S} \eta^{-\alpha}.$$

This proves (5.19). If $x \in \mathcal{C}$, the definition of $\eta(\mathcal{C})$ implies that $\{y \in \mathcal{C} : y \neq x\} \subseteq \{y \in \mathcal{C} : \rho(x, y) \geq \eta/2\}$. Hence, (5.19) implies (5.20). The estimate (5.21) follows directly from (5.3) used with τ in place of ν . \square

Lemma 5.3. *We assume the set up as in Theorem 4.2. Then, for $2^m \geq (\eta/4)^{-1}$,*

$$(5.23) \quad \sum_{\substack{y \in \mathcal{C} \\ \rho(x,y) \geq \eta(\mathcal{C}_n)/2}} |\mathbf{A}_{m,n}(x, y)| \leq c \frac{2^{m\alpha}}{2^{(m-n)S}}, \quad x \in \mathbb{X},$$

$$(5.24) \quad |\mathbf{A}_{m,n}(x, x)| \sim 2^{m\alpha}, \quad x \in \mathbb{X}.$$

Moreover, for $m \geq n$, and $2^m \geq (\eta/4)^{-1}$,

$$(5.25) \quad \sum_{y \in \mathcal{C}_n} |\mathbf{A}_{m,n}(x, y) - \mathbf{B}_{m,n}(x, y)| \leq c_1 \frac{2^{m\alpha}}{2^{nS}}, \quad x \in \mathbb{X}.$$

In particular, for $x \in \mathcal{C}_n$, (5.23) implies that

$$(5.26) \quad \sum_{\substack{y \in \mathcal{C} \\ y \neq x}} |\mathbf{A}_{m,n}(x, y)| \leq c \frac{2^{m\alpha}}{2^{(m-n)S}},$$

Proof. The localization estimate (5.8), (5.19) in Lemma 5.2 and our assumption (4.5) regarding the minimal separation of set \mathcal{C}_n imply that for all $x \in \mathcal{C}_n$,

$$(5.27) \quad \begin{aligned} \sum_{\substack{y \in \mathcal{C}_n \\ \rho(x,y) \geq \eta(\mathcal{C}_n)/2}} |\Phi_{2^m}(h, x, y)| &\leq c_2 \sum_{\substack{y \in \mathcal{C}_n \\ \rho(x,y) \geq \eta(\mathcal{C}_n)/2}} \frac{2^{m\alpha}}{\max(1, (2^m \rho(x, y))^S)} \\ &\leq c_3 (\eta(\mathcal{C}_n) 2^m)^{\alpha-S} \eta(\mathcal{C}_n)^{-\alpha} \\ &\leq c_4 \frac{2^{n\alpha}}{2^{(m-n)S-\alpha}} = c_4 2^{-(m-n)S} 2^{m\alpha}. \end{aligned}$$

This proves (5.23). The estimate (5.26) follows from (5.20) in the same way.

Since $h(t) \geq 0$ for all t and $h(t) = 1$ if $t \in [0, 1/2]$, (2.8) shows that for all $x \in \mathbb{X}$,

$$(5.28) \quad |\Phi_{2^m}(h, x, x)| = \sum_j h\left(\frac{\ell_j}{2^m}\right) |\phi_j(x)|^2 \geq \sum_{\ell_j \leq 2^{m-1}} |\phi_j(x)|^2 \geq c_5 2^{m\alpha}.$$

Using (2.8) and the facts that $0 \leq h(t) \leq 1$ for all $t \in \mathbb{R}$, and $h(t) = 0$ if $|t| \geq 1$, we deduce that

$$|\Phi_{2^m}(h, x, x)| \leq c_6 2^{m\alpha}, \quad x \in \mathbb{X}.$$

This proves (5.24).

Next, we use (3.14) with $N = 2^m$, $R = 2S + 3$, and 2^n in place of n to obtain

$$(5.29) \quad |\Phi_{2^m}(h, x, y) - \Phi_{2^m}(Q_{2^n, S+3}, x, y)| \leq \frac{c_5}{2^{nS}} \frac{2^{m\alpha}}{\max(1, (2^m \rho(x, y))^S)}, \quad x, y \in \mathbb{X}.$$

We deduce (5.25) using (5.21), our assumption (4.5) regarding the minimal separation of set \mathcal{C}_n , and the fact that $m \geq n$. \square

Proof of Theorem 4.2. We recall that the norms of matrices are taken with respect to the ℓ^∞ norms on the underlying Euclidean spaces. Let $m \geq n$ and $2^m \geq (\eta/4)^{-1}$. In view of (5.23) and (5.24), we see that

$$(5.30) \quad \sum_{y \in \mathcal{C}_n} |\mathbf{A}_{m,n}(x, y)| \leq c 2^{m\alpha}, \quad x \in \mathbb{X}.$$

In particular,

$$(5.31) \quad \|\mathbf{A}_{m,n}\| = \max_{x \in \mathcal{C}_n} \sum_{y \in \mathcal{C}_n} |\mathbf{A}_{m,n}(x, y)| \leq c2^{m\alpha}.$$

Also, it is clear from (5.26) and (5.24) that for sufficiently large c^* , we have

$$(5.32) \quad \sum_{\substack{y \in \mathcal{C} \\ y \neq x}} |\mathbf{A}_{m,n}(x, y)| \leq (1/4) \min_{z \in \mathcal{C}_n} |\mathbf{A}_{m,n}(z, z)|, \quad x \in \mathcal{C}_n, \quad m \geq n + c^*.$$

Therefore, we may apply Proposition 5.2 with $\gamma = 1/4$, and in view of (5.24), with $\lambda \sim 2^{m\alpha}$ to conclude that if $m \geq n + c^*$ then $\mathbf{A}_{m,n}$ is invertible, and

$$(5.33) \quad \|\mathbf{A}_{m,n}^{-1}\| \leq c2^{-m\alpha}.$$

The estimate (4.6) follows from (5.31) and (5.33).

In the remainder of this proof, we take $m = n + c^*$. The equation (4.8) is clear from the definitions. The first estimate in (4.9) follows from the definition (4.7) and the fact that $\Phi_{2^{n+c^*}}(h, x, \circ) \in \Pi_{2^{n+c^*}}$. In view of (4.1),

$$(5.34) \quad \|\mathbf{e}_n^d(f)\|_{\ell^\infty} \leq c\{E_{2^{n-1}, \infty}(f) + 2^{-nS}\|f\|_{\mu; \infty}\} =: \epsilon_n.$$

In view of (5.33), this yields

$$\|\mathbf{A}_{n+c^*, n}^{-1} \mathbf{e}_n^d(f)\|_{\ell^\infty} \leq c2^{-m\alpha} \epsilon_n.$$

Therefore, (5.30) implies that

$$\left| \sum_{y \in \mathcal{C}_n} (\mathbf{A}_{n+c^*, n}^{-1} \mathbf{e}_n^d(f, y)) \Phi_{2^{n+c^*}}(h, x, y) \right| \leq c\epsilon_n.$$

Together with (4.1), this implies the second inequality in (4.9). □

Theorem 4.4 can be obtained from Theorem 4.2 by a perturbation argument.

Proof of Theorem 4.4. In this proof only, let $\mathbf{A} = \mathbf{A}_{n+c^*, n}$, $\mathbf{B} = \mathbf{B}_{n+c^*, n}$, and \mathbf{D} be the diagonal matrix of \mathbf{A} . Then (5.24) shows that

$$(5.35) \quad \|\mathbf{D}^{-1}\| \sim 2^{-n\alpha},$$

and (5.32) can be rewritten in the form

$$\|\mathbf{D} - \mathbf{A}\| \leq (1/4)\|\mathbf{D}^{-1}\|^{-1}.$$

Hence, Proposition 5.3 implies that

$$\|\mathbf{D}^{-1} - \mathbf{A}^{-1}\| \leq (1/3)\|\mathbf{D}^{-1}\|.$$

Thus,

$$\|\mathbf{A}^{-1}\| \sim \|\mathbf{D}^{-1}\| \sim 2^{-n\alpha}.$$

The estimate (5.25) used with $m = n + c^* > n$ can now be rewritten in the form

$$\|\mathbf{A} - \mathbf{B}\| \leq \frac{c}{2^{nS}} \|\mathbf{A}^{-1}\|^{-1}.$$

Hence,

$$(5.36) \quad \|\mathbf{B}\| \leq c2^{n\alpha}.$$

Further, if c_1^* is chosen so that $2^{nS} > c/4$ above for $n \geq c_1^*$, then Proposition 5.3 can be applied again to deduce that

$$(5.37) \quad \|\mathbf{A}^{-1} - \mathbf{B}^{-1}\| \leq \frac{c_1}{2^{nS}} \|\mathbf{A}^{-1}\| \sim 2^{n(-\alpha-S)}, \quad \|\mathbf{B}^{-1}\| \sim 2^{-n\alpha}.$$

Together with (5.36), this establishes the bound on the condition number of \mathbf{B} as in (4.13).

Next, using (5.5) with $p = \infty$, we deduce that

$$(5.38) \quad \|\mathbf{e}_n(f) - \mathbf{e}_n^d(f)\|_{\ell^\infty} \leq \|\sigma_{2^n}(\nu_n; h, f) - \sigma_{2^n}^*(\nu_n; f)\|_{\mu; \infty} \leq c2^{-nS} \|f\|_{\mu; \infty}.$$

In view of (5.34) (including the notation there) and (5.37), we may now conclude that

$$(5.39) \quad \begin{aligned} & \|\mathbf{A}^{-1}\mathbf{e}_n^d(f) - \mathbf{B}^{-1}\mathbf{e}_n(f)\|_{\ell^\infty} \\ &= \|(\mathbf{A}^{-1} - \mathbf{B}^{-1})\mathbf{e}_n^d(f) + \mathbf{B}^{-1}(\mathbf{e}_n^d(f) - \mathbf{e}_n(f))\|_{\ell^\infty} \\ &\leq \{2^{-n(\alpha+S)}\epsilon_n + 2^{-n(\alpha+S)}\|f\|_{\mu; \infty}\} \leq c2^{-n(\alpha+S)}\|f\|_{\mu; \infty}. \end{aligned}$$

Using (5.25) with $m = n + c^*$, it is not difficult to deduce that

$$\begin{aligned} & \|\mathcal{I}_n^d(f) - \mathcal{I}_n(f)\|_{\mu; \infty} \leq \|\sigma_{2^n}(\nu_n; h, f) - \sigma_{2^n}^*(\nu_n; f)\|_{\mu; \infty} \\ &+ \left\| \sum_{y \in \mathcal{C}_n} ((\mathbf{A}^{-1}\mathbf{e}_n^d(f))\Phi_{2^m}(h, \circ, y) - (\mathbf{B}^{-1}\mathbf{e}_n(f))\Phi_{2^m}(Q_{2^n, S+3}, \circ, y)) \right\|_{\mu; \infty} \\ &\leq c2^{-nS}\|f\|_{\mu; \infty}. \end{aligned}$$

Therefore, (4.16) follows from the corresponding estimate (4.9) for \mathcal{I}_n^d . □

Finally, we prove Proposition 2.2. Towards this goal, we first make the following observation [23]. If $K \subseteq \mathbb{X}$ is a compact subset and $\epsilon > 0$, we will say that a subset $\mathcal{C} \subseteq K$ is ϵ -separated if $\rho(x, y) \geq \epsilon$ for every $x, y \in \mathcal{C}$, $x \neq y$. Since K is compact, there exists a finite, maximal ϵ -separated subset $\{x_1, \dots, x_M\}$ of K . If $x \in K \setminus \bigcup_{k=1}^M B(x_k, \epsilon)$, then $\{x, x_1, \dots, x_M\}$ is a strictly larger ϵ -separated subset of K . So, $K \subseteq \bigcup_{k=1}^M B(x_k, \epsilon)$. Moreover, the balls $B(x_k, \epsilon/2)$ are mutually disjoint.

Proof of Proposition 2.2. The proofs of parts (a) and (b) are reproduced verbatim from [23], except to confirm to the notation of this paper. The proof of part (c) is as given in [31], except for a change of notation.

In the proof of part (a) only, let $\lambda > \|\nu\|_{R,d}$, $r > 0$, $x \in \mathbb{X}$, and let $\{y_1, \dots, y_N\}$ be a maximal $2d/3$ -separated subset of $\mathbb{B}(x, r+2d/3)$. Then $\mathbb{B}(x, r) \subseteq \mathbb{B}(x, r+2d/3) \subseteq \bigcup_{j=1}^N \mathbb{B}(y_j, 2d/3)$. So,

$$|\nu|(\mathbb{B}(x, r)) \leq |\nu|(\mathbb{B}(x, r+2d/3)) \leq \sum_{j=1}^N |\nu|(\mathbb{B}(y_j, 2d/3)) \leq \sum_{j=1}^N |\nu|(\mathbb{B}(y_j, d)) \leq \lambda N d^\alpha.$$

The balls $\mathbb{B}(y_j, d/3)$ are mutually disjoint, and $\bigcup_{j=1}^N \mathbb{B}(y_j, d/3) \subseteq \mathbb{B}(x, r+d)$. In view of (2.9), $d^\alpha \leq c\mu(\mathbb{B}(y_j, d/3))$ for each j . So,

$$\begin{aligned} |\nu|(\mathbb{B}(x, r)) &\leq \lambda N d^\alpha \leq c\lambda \sum_{j=1}^N \mu(\mathbb{B}(y_j, d/3)) \\ &= c\lambda \mu\left(\bigcup_{j=1}^N \mathbb{B}(y_j, d/3)\right) \leq c\lambda \mu(\mathbb{B}(x, r+d)). \end{aligned}$$

Since $\lambda > \|\nu\|_{R,d}$ was arbitrary, this leads to the first inequality in (2.27). The second inequality follows from (2.5). The converse statement is obvious. This completes the proof of part (a).

The second estimate in (2.28) is clear from the definitions. The first estimate in (2.28) follows by applying (2.27) with $r = \gamma d$. This completes the proof of part (b).

In order to prove part (c), we use the characterization given in part (a). Let $x_0 \in \mathbb{X}$, $r > 0$ be arbitrary, and in this proof only, let $\mathbb{B}(x_0, r) \cap \mathcal{C} = \{z_1, \dots, z_J\}$, then the balls $\mathbb{B}(z_j, \eta/2)$ are disjoint, and $\bigcup_{j=1}^J \mathbb{B}(z_j, \eta/2) \subset \mathbb{B}(x_0, r + \eta/2)$. Using the fact that $\tau(\mathbb{B}(x_0, r)) = J$, and recalling (2.9), we obtain

$$\begin{aligned} \mu(\mathbb{B}(x_0, r + \eta/2)) &\geq \mu\left(\bigcup_{j=1}^J \mathbb{B}(z_j, \eta/2)\right) \\ &= \sum_{j=1}^J \mu(\mathbb{B}(z_j, \eta/2)) \geq cJ\eta^\alpha = c\tau(\mathbb{B}(x_0, r))\eta^\alpha. \end{aligned}$$

Thus part (a) of this proposition implies that τ is η -regular, and $\|\tau\|_{R,\eta} \leq c\eta^{-\alpha}$. \square

ACKNOWLEDGEMENT

The author's thank the referee for his/her many useful suggestions for the improvement of the first draft of this paper.

REFERENCES

- [1] M. Belkin and P. Niyogi, *Semi-supervised learning on Riemannian manifolds*, Machine Learning Journal **56** (2004), 209–239.
- [2] Mikhail Belkin and Partha Niyogi, *Towards a theoretical foundation for Laplacian-based manifold methods*, J. Comput. System Sci. **74** (2008), no. 8, 1289–1308, DOI 10.1016/j.jcss.2007.08.006. MR2460286 (2009m:68207)
- [3] M. Belkin, P. Niyogi, *Convergence of Laplacian Eigenmaps*, Manuscript (<http://www.cse.ohio-state.edu/~mbelkin/papers/CLEM.08.pdf>).
- [4] Charles K. Chui and Harvey Diamond, *A general framework for local interpolation*, Numer. Math. **58** (1991), no. 6, 569–581, DOI 10.1007/BF01385640. MR1083520 (92b:41005)
- [5] Charles K. Chui and H. N. Mhaskar, *MRA contextual-recovery extension of smooth functions on manifolds*, Appl. Comput. Harmon. Anal. **28** (2010), no. 1, 104–113, DOI 10.1016/j.acha.2009.04.004. MR2563262 (2011j:94010)
- [6] Charles K. Chui and Jianzhong Wang, *PDE models associated with the bilateral filter*, Adv. Comput. Math. **31** (2009), no. 1-3, 131–156, DOI 10.1007/s10444-008-9095-2. MR2511577 (2010e:94028)
- [7] C.K. Chui and J.Z. Wang, *Dimensionality Reduction of Hyper-spectral Imagery Data for Feature Classification*, in “Handbook of Geomathematics”, pp. 1005–1049, Freeden, W., Nashed, Z., and Sonar, T. (eds.), Springer, 2010.
- [8] Charles K. Chui and Jianzhong Wang, *Randomized anisotropic transform for nonlinear dimensionality reduction*, GEM Int. J. Geomath. **1** (2010), no. 1, 23–50, DOI 10.1007/s13137-010-0004-8. MR2747641 (2012b:62012)
- [9] Ronald R. Coifman and Stéphane Lafon, *Diffusion maps*, Appl. Comput. Harmon. Anal. **21** (2006), no. 1, 5–30, DOI 10.1016/j.acha.2006.04.006. MR2238665 (2008a:60210)
- [10] Ronald R. Coifman and Mauro Maggioni, *Diffusion wavelets*, Appl. Comput. Harmon. Anal. **21** (2006), no. 1, 53–94, DOI 10.1016/j.acha.2006.04.004. MR2238667 (2007d:42067)
- [11] W. Czaja and M. Ehler, *Schrödinger eigenmaps for the analysis of bio-medical data*, IEEE Trans. Pattern Anal. Mach. Intelligence, DOI:10.1109/TPAMI.2012.270.
- [12] E. B. Davies, *L^p spectral theory of higher-order elliptic differential operators*, Bull. London Math. Soc. **29** (1997), no. 5, 513–546, DOI 10.1112/S002460939700324X. MR1458713 (98d:35164)
- [13] Ronald A. DeVore and George G. Lorentz, *Constructive Approximation*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 303, Springer-Verlag, Berlin, 1993. MR1261635 (95f:41001)
- [14] M. Ehler, *The multiresolution structure of pairs of dual wavelet frames for a pair of Sobolev spaces*, Jaen J. Approx., **2** (2) (2010), 193–214.
- [15] M. Ehler, F. Filbir, and H. N. Mhaskar, *Locally learning biomedical data using diffusion frames*, J. Comput. Biol. **19** (2012), no. 11, 1251–1264, DOI 10.1089/cmb.2012.0187. MR2994881

- [16] Michael Elad, *On the origin of the bilateral filter and ways to improve it*, IEEE Trans. Image Process. **11** (2002), no. 10, 1141–1151, DOI 10.1109/TIP.2002.801126. MR1966043 (2004c:94042)
- [17] Charles Fefferman, *Whitney’s extension problem for C^m* , Ann. of Math. (2) **164** (2006), no. 1, 313–359, DOI 10.4007/annals.2006.164.313. MR2233850 (2007g:58013)
- [18] Charles Fefferman, *The structure of linear extension operators for C^m* , Rev. Mat. Iberoam. **23** (2007), no. 1, 269–280, DOI 10.4171/RMI/495. MR2351135 (2009c:41003)
- [19] Charles Fefferman and Bo’az Klartag, *Fitting a C^m -smooth function to data. I*, Ann. of Math. (2) **169** (2009), no. 1, 315–346, DOI 10.4007/annals.2009.169.315. MR2480607 (2011g:58011)
- [20] Charles Fefferman and Bo’az Klartag, *Fitting a C^m -smooth function to data. II*, Rev. Mat. Iberoam. **25** (2009), no. 1, 49–273, DOI 10.4171/RMI/569. MR2514338 (2011f:58020)
- [21] Charles Fefferman, *Fitting a C^m -smooth function to data. III*, Ann. of Math. (2) **170** (2009), no. 1, 427–441, DOI 10.4007/annals.2009.170.427. MR2521121 (2011e:58014)
- [22] F. Filbir and H. N. Mhaskar, *A quadrature formula for diffusion polynomials corresponding to a generalized heat kernel*, J. Fourier Anal. Appl. **16** (2010), no. 5, 629–657, DOI 10.1007/s00041-010-9119-4. MR2673702 (2011j:41007)
- [23] F. Filbir and H. N. Mhaskar, *Marcinkiewicz-Zygmund measures on manifolds*, J. Complexity **27** (2011), no. 6, 568–596, DOI 10.1016/j.jco.2011.03.002. MR2846706
- [24] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1996. MR1417720 (97g:65006)
- [25] Alexander Grigor’yan, *Estimates of heat kernels on Riemannian manifolds*, Spectral theory and geometry (Edinburgh, 1998), London Math. Soc. Lecture Note Ser., vol. 273, Cambridge Univ. Press, Cambridge, 1999, pp. 140–225, DOI 10.1017/CBO9780511566165.008. MR1736868 (2001b:58040)
- [26] Alexander Grigor’yan, *Heat kernels on weighted manifolds and applications*, The ubiquitous heat kernel, Contemp. Math., vol. 398, Amer. Math. Soc., Providence, RI, 2006, pp. 93–191, DOI 10.1090/conm/398/07486. MR2218016 (2007a:58028)
- [27] Alexander Grigor’yan, *Heat kernels on metric measure spaces with regular volume growth*, Handbook of geometric analysis, No. 2, Adv. Lect. Math. (ALM), vol. 13, Int. Press, Somerville, MA, 2010, pp. 1–60. MR2743439 (2012d:58037)
- [28] Yu. A. Kordyukov, *L^p -theory of elliptic differential operators on manifolds of bounded geometry*, Acta Appl. Math. **23** (1991), no. 3, 223–260. MR1120831 (92f:58164)
- [29] M. Maggioni and H. N. Mhaskar, *Diffusion polynomial frames on metric measure spaces*, Appl. Comput. Harmon. Anal. **24** (2008), no. 3, 329–353, DOI 10.1016/j.acha.2007.07.001. MR2407008 (2010d:42052)
- [30] H. N. Mhaskar, *Polynomial operators and local smoothness classes on the unit interval*, J. Approx. Theory **131** (2004), no. 2, 243–267, DOI 10.1016/j.jat.2004.10.002. MR2106540 (2005k:41011)
- [31] H. N. Mhaskar, *Eignets for function approximation on manifolds*, Appl. Comput. Harmon. Anal. **29** (2010), no. 1, 63–87, DOI 10.1016/j.acha.2009.08.006. MR2647012 (2012c:41031)
- [32] H.N. Mhaskar, *A generalized diffusion frame for parsimonious representation of functions on data defined manifolds*, Neural Networks **24** (2011), 345–359.
- [33] Naoki Saito, *Data analysis and representation on a general domain using eigenfunctions of Laplacian*, Appl. Comput. Harmon. Anal. **25** (2008), no. 1, 68–97, DOI 10.1016/j.acha.2007.09.005. MR2419705 (2009d:35038)
- [34] J.B. Tenenbaum, V. de Silva, and J.C. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science **290** (2000), 2319–2323.
- [35] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*, in “Proc. 6th Int. Conf. Computer Vision”, New Delhi, India, 1998, pp. 839–846.

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, STANFORD, CALIFORNIA 94305
E-mail address: ckchui@stanford.edu

DEPARTMENT OF MATHEMATICS, CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA 91125 — AND — INSTITUTE OF MATHEMATICAL SCIENCES, CLAREMONT GRADUATE UNIVERSITY, CLAREMONT, CALIFORNIA 91711
E-mail address: hmhaska@gmail.com