

Green's Functions: Taking Another Look at Kernel Approximation, Radial Basis Functions and Splines

Gregory E. Fasshauer

Abstract The theories for radial basis functions (RBFs) as well as piecewise polynomial splines have reached a stage of relative maturity as is demonstrated by the recent publication of a number of monographs in either field. However, there remain a number of issues that deserve to be investigated further. For instance, it is well known that both splines and radial basis functions yield “optimal” interpolants, which in the case of radial basis functions are discussed within the so-called native space setting. It is also known that the theory of reproducing kernels provides a common framework for the interpretation of both RBFs and splines. However, the associated reproducing kernel Hilbert spaces (or native spaces) are often not that well understood — especially in the case of radial basis functions. By linking (conditionally) positive definite kernels to Green's functions of differential operators we obtain new insights that enable us to better understand the nature of the native space as a generalized Sobolev space. An additional feature that appears when viewing things from this perspective is the notion of scale built into the definition of these function spaces. Furthermore, the eigenfunction expansion of a positive definite kernel via Mercer's theorem provides a tool for making progress on such important questions as stable computation with flat radial basis functions and dimension independent error bounds.

1 Introduction

A number of monographs and survey papers dealing with splines, radial basis functions and, more generally, reproducing kernels, have appeared in recent years. The following list is representative, but certainly far from complete: [1, 3, 8, 18, 20, 29, 34, 39, 40, 41, 43]. Even though (or precisely because) there is regrettably little interaction between different mathematical communities, we have included references

Gregory E. Fasshauer
Illinois Institute of Technology, Chicago, IL 60616, U.S.A., e-mail: fasshauer@iit.edu

from approximation theory as well as probability/statistics and machine learning since reproducing kernels play a central — and perhaps increasing — role in all of these communities.

In this paper we will address several questions that — to our knowledge — have not been addressed sufficiently in the existing literature. The first few questions center around the notion of an RBF *native space* (to be defined below). We will recall existing interpretations and claim that most of them are not very “intuitive”. This has, in fact, been a point of criticism of RBF methods. What are these native spaces, and what kind of functions do they contain? How do they relate to classical function spaces such as Sobolev spaces? We try to shed some light on this topic by discussing recent work of [12, 44] in Section 2.

Another set of questions is related to the role of *scale*. RBF practitioners have known for a long time that the proper scaling of the basis functions plays a very important role. It might affect the accuracy of an approximation, its numerical stability and its efficiency. Should a notion of scale be included in the definition of the native space? Our framework of Section 2 does indeed provide a natural way of doing this.

An appropriate scaling of the kernel has been used to establish a connection between infinitely smooth RBFs and polynomial interpolants in the literature (see Section 3 and the references listed there). If the kernels are “flat”, we get convergence of RBF interpolants to polynomial interpolants. We will report on a recent investigation [38] that reveals a similar connection between RBFs of limited smoothness and piecewise polynomial splines.

Even though researchers have struggled for many years with the ill-conditioning of RBF systems, relatively little progress has been made in this direction. For univariate piecewise polynomial splines it is well known that moving from the basis of truncated power functions to the B-spline basis provides well-conditioned, and even banded, matrices. Aside from some scattered work on preconditioning of RBF systems, only Bengt Fornberg together with his co-workers has tackled this problem with some success. We are especially motivated by their RBF-QR idea [14, 15] and will provide some of our own thoughts on this approach in Section 4.

Finally, many papers on rates of convergence of the RBF approximation method exist in the literature. However, none of these papers address the question of dimension-dependence of these bounds. In fact, it is quite obvious that all of the existing bounds suffer from the *curse of dimensionality*. In Section 5 we review recent work [9] on *dimension-independent* error bounds for RBF approximation methods.

It turns out that a unifying theme underlying all of these questions is the notion of *Green’s functions* and *eigenfunction expansions*. Therefore, these topics will be reviewed in the next section. Connections between either splines and Green’s functions or radial basis functions and Green’s functions have repeatedly been made over the past decades (see, e.g., [4, 7, 19, 25, 27, 28, 37, 41]). However, many of the connections presented in the following seem to go beyond the discussion in the existing literature. Throughout the paper we will use (simple) examples to illustrate the various topics.

2 Toward an Intuitive Interpretation of Native Spaces

2.1 What is the current situation?

Even though piecewise polynomial splines and radial basis functions are conceptually very similar (some people do not even distinguish between the two and use the term *spline* to refer to either method), there are relatively few intersections in the literature on these two approximation methods. Perhaps the most prominent common feature of the two methods is given by the fact that they both yield *minimum norm interpolants* (see, e.g., [2, 8, 37, 43]). In fact, it is precisely this property that led Schoenberg to refer to piecewise polynomial univariate approximating functions as *splines* [36].

To begin with a specific example, we recall that the natural spline $s_{f,2m}$ of order $2m$ provides the smoothest interpolant to data sampled from any function f in the Sobolev space $H^m(a,b)$ of functions whose m^{th} derivative is square integrable on $[a,b]$ and whose derivatives of orders m through $2m-2$ vanish at the endpoints of the interval $[a,b]$, i.e.,

$$s_{f,2m} = \operatorname{argmin}_{f \in H^m(a,b)} \left\{ \int_a^b [f^{(m)}(x)]^2 dx \mid f(x_i) = y_i, i = 1, \dots, N, \right. \\ \left. f^{(\ell)}(a) = f^{(\ell)}(b) = 0, \ell = m, \dots, 2m-2 \right\}. \quad (1)$$

Now let us consider the corresponding minimum norm property as it is commonly found for radial basis functions, or more generally reproducing kernel interpolants (see, e.g., [43]). The reproducing kernel interpolant $s_{f,K}$ is optimal in the sense that it is the minimum norm interpolant to data sampled from any function f in $\mathcal{H}(K, \Omega)$, the reproducing kernel Hilbert space (or *native space*) associated with K . This can be stated as

$$s_{f,K} = \operatorname{argmin}_{f \in \mathcal{H}(K, \Omega)} \{ \|f\|_{\mathcal{H}(K, \Omega)} \mid s_{f,K}(\mathbf{x}_i) = f(\mathbf{x}_i), i = 1, \dots, N \}. \quad (2)$$

While the function space $H^m(a,b)$ that appears in (1) can be rather easily understood in terms of the smoothness and boundary conditions imposed, the native space $\mathcal{H}(K, \Omega)$ in (2) looks a bit more cryptic. What is this mysterious native space and how is its norm defined?

For a general positive definite kernel K and domain $\Omega \subseteq \mathbb{R}^d$ the native space is commonly defined as

$$\mathcal{H}(K, \Omega) = \operatorname{span} \{ K(\cdot, \mathbf{z}) \mid \mathbf{z} \in \Omega \},$$

i.e., the native space is given by all linear combinations of — often infinitely many — “shifts” of the kernel K . This is certainly a valid definition, but what sort of func-

tions does $\mathcal{H}(K, \Omega)$ contain? The literature is more specific for the case in which we use translation invariant (in the statistics literature also referred to as *stationary*) kernels on $\Omega = \mathbb{R}^d$, i.e., if the kernel is really a function of one variable, namely the difference of two points, or $\tilde{K}(\mathbf{x} - \mathbf{z}) = K(\mathbf{x}, \mathbf{z})$. In this case, if $\tilde{K} \in C(\mathbb{R}^d) \cap L_1(\mathbb{R}^d)$, then

$$\mathcal{H}(\tilde{K}, \mathbb{R}^d) = \left\{ f \in L_2(\mathbb{R}^d) \cap C(\mathbb{R}^d) \mid \frac{\mathcal{F}f}{\sqrt{\mathcal{F}\tilde{K}}} \in L_2(\mathbb{R}^d) \right\},$$

i.e., a function f belongs to the native space $\mathcal{H}(\tilde{K}, \mathbb{R}^d)$ of the kernel \tilde{K} if the decay of its Fourier transform $\mathcal{F}f$ relative to that of the Fourier transform $\mathcal{F}\tilde{K}$ of the kernel is rapid enough. This characterization certainly encodes some kind of smoothness information, but it is not very intuitive. The previous material is covered in much more detail in [43].

As mentioned above, we are not only interested in understanding the type of functions contained in the native space, but also the norm this space is equipped with. Since both the spline and kernel spaces are Hilbert spaces it is natural to look at their inner products. In the natural spline case this is the standard Sobolev inner product whose induced norm appears in (1). What does the native space inner product look like?

For a general positive definite kernel K on a general domain Ω we take functions $f, g \in \mathcal{H}(K, \Omega)$ and use the notation $N_K = \dim(\mathcal{H}(K, \Omega))$ for the dimension of the native space (note that $N_K = \infty$ is common). Then

$$\langle f, g \rangle_{\mathcal{H}(K, \Omega)} = \left\langle \sum_{j=1}^{N_K} c_j K(\cdot, \mathbf{x}_j), \sum_{k=1}^{N_K} d_k K(\cdot, \mathbf{z}_k) \right\rangle_{\mathcal{H}(K, \Omega)} = \sum_{j=1}^{N_K} \sum_{k=1}^{N_K} c_j d_k K(\mathbf{x}_j, \mathbf{z}_k).$$

Once again, one might wonder how to interpret this. As before, for translation invariant kernels on $\Omega = \mathbb{R}^d$, i.e., $\tilde{K}(\mathbf{x} - \mathbf{z}) = K(\mathbf{x}, \mathbf{z})$, we can employ Fourier transforms. Then we have

$$\langle f, g \rangle_{\mathcal{H}(\tilde{K}, \mathbb{R}^d)} = \frac{1}{\sqrt{(2\pi)^d}} \left\langle \frac{\mathcal{F}f}{\sqrt{\mathcal{F}\tilde{K}}}, \frac{\mathcal{F}g}{\sqrt{\mathcal{F}\tilde{K}}} \right\rangle_{L_2(\mathbb{R}^d)}$$

provided $\tilde{K} \in C(\mathbb{R}^d) \cap L_1(\mathbb{R}^d)$ and $f, g \in \mathcal{H}(\tilde{K}, \mathbb{R}^d)$.

Before we begin our discussion relating kernel methods to Green's functions — and thereby providing an interpretation of native spaces as *generalized Sobolev spaces* — we mention a few examples of kernels whose native spaces already are known to be Sobolev spaces. Since all of these kernels are radial (or *isotropic*) kernels we introduce the notation $\kappa(\|\mathbf{x} - \mathbf{z}\|) = K(\mathbf{x}, \mathbf{z})$. This also helps us avoid confusion between a kernel K and the modified Bessel function of the second kind $K_{m-d/2}$ that appears below.

Matérn kernels (sometimes also called *Sobolev splines*, see, e.g., [8]) are of the form

$$\kappa(r) \doteq K_{m-d/2}(r) r^{m-d/2}, \quad m > \frac{d}{2},$$

where we have used the notation \doteq to indicate that equality holds up to a multiplicative constant. It is quite natural to use the term Sobolev splines to refer to these functions since their native space is given by a classical Sobolev space, i.e., $\mathcal{H}(\kappa, \mathbb{R}^d) = H^m(\mathbb{R}^d)$.

A second example is given by the entire family of *Wendland's compactly supported radial basis functions* (see, e.g., [8, 43]). A popular member of this family is of the form

$$\kappa(r) \doteq (1-r)_+^4(4r+1),$$

and its native space $\mathcal{H}(\kappa, \mathbb{R}^3)$ is norm-equivalent to the classical Sobolev space $H^3(\mathbb{R}^3)$ (see [42]).

The family of *polyharmonic splines* is another famous (albeit only conditionally positive definite) example that fits this list. These functions are of the form

$$\kappa(r) \doteq \begin{cases} r^{2m-d}, & d \text{ odd,} \\ r^{2m-d} \log r, & d \text{ even,} \end{cases}$$

and the native space $\mathcal{H}(\kappa, \mathbb{R}^d)$ is a *Beppo-Levi space of order m*

$$\text{BL}_m(\mathbb{R}^d) = \left\{ f \in C(\mathbb{R}^d) \mid D^\alpha f \in L_2(\mathbb{R}^d) \text{ for all } |\alpha| = m \right\}.$$

We may consider this space as a homogeneous Sobolev space of order m (see [8]).

This latter example is also featured in the recent paper [4], and spherical versions of the latter two examples are discussed in [25]. We became aware of both of these papers only after the initial submission of our own paper and it is interesting to note that they both use the connection between reproducing kernels and Green's functions as an essential ingredient to obtain their L_p approximation results.

2.2 Mercer's theorem and eigenvalue problems

We will limit most of our discussion to positive definite kernels. A perspective on positive definite kernels that appears quite frequently in the literature on statistical learning (but not so much in approximation theory) is their characterization via an eigenfunction expansion. This fact goes back many years to the early work of James Mercer [24]. We quote here a version of this result from [29].

Theorem 1 (Mercer's theorem). *Let (Ω, μ) be a finite measure space and $K \in L_\infty(\Omega^2, \mu^2)$ be a kernel such that the integral operator $T_K : L_2(\Omega, \mu) \rightarrow L_2(\Omega, \mu)$ defined by*

$$(T_K f)(\mathbf{x}) = \int_{\Omega} K(\mathbf{x}, \mathbf{z}) f(\mathbf{z}) d\mu(\mathbf{z})$$

is positive definite. Let $\varphi_n \in L_2(\Omega, \mu)$ be the normalized eigenfunctions of T_K associated with the eigenvalues $\lambda_n > 0$. Then

1. the eigenvalues $\{\lambda_n\}_{n=1}^{\infty}$ are absolutely summable,
2. $K(\mathbf{x}, \mathbf{z}) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(\mathbf{x}) \varphi_n(\mathbf{z})$ holds μ^2 almost everywhere, and the series converges absolutely and uniformly μ^2 almost everywhere.

More generally, Hilbert-Schmidt theory ensures the existence of L_2 -convergent eigenfunction expansions of compact, self-adjoint operators.

We now consider a kernel $K : \Omega \times \Omega \rightarrow \mathbb{R}$ on a general domain Ω and define an inner product with positive weight function σ (instead of using the measure theoretic notation of the theorem) as

$$\langle f, g \rangle = \int_{\Omega} f(\mathbf{x})g(\mathbf{x})\sigma(\mathbf{x})d\mathbf{x}.$$

The eigenvalue problem for the integral operator $T_K : f \mapsto \int_{\Omega} K(\cdot, \mathbf{z})f(\mathbf{z})\sigma(\mathbf{z})d\mathbf{z}$ consists of finding solutions λ and φ of

$$\int_{\Omega} K(\mathbf{x}, \mathbf{z})\varphi(\mathbf{z})\sigma(\mathbf{z})d\mathbf{z} = \lambda\varphi(\mathbf{x}). \quad (3)$$

This represents a homogeneous Fredholm integral equation of the 2nd kind and it is therefore not obvious how we should go about finding the eigenvalues and eigenfunctions of T_K . The idea we will pursue here is to relate the integral equation to a differential equation which may be easier to solve.

2.3 Green's functions and eigenfunction expansions

Green's functions play a central role in the solution of differential equations. We now consider the nonhomogeneous linear (ordinary or partial) differential equation

$$(Lu)(\mathbf{x}) = f(\mathbf{x}) \quad \text{on } \Omega \subset \mathbb{R}^d$$

with a linear and elliptic operator L and some appropriate homogeneous boundary conditions. We will be more specific about the boundary conditions later.

The solution of this differential equation can be written in terms of a *Green's function* G as

$$u(\mathbf{x}) = \int_{\Omega} f(\mathbf{z})G(\mathbf{x}, \mathbf{z})d\mathbf{z}, \quad (4)$$

where the Green's function satisfies the differential equation

$$(LG)(\mathbf{x}, \mathbf{z}) = \delta(\mathbf{x} - \mathbf{z}).$$

Here δ denotes the standard delta function(al), and the point \mathbf{z} denotes a fixed (and arbitrary) "source". The boundary conditions are the same as above.

We now establish a connection between the integral operator eigenvalue problem that is needed for Mercer's series representation of a positive definite kernel discussed above and a related eigenvalue problem for a differential operator. For simplicity we assume that K is a free space Green's function for the differential operator L , i.e., $(LK)(\mathbf{x}, \mathbf{z}) = \delta(\mathbf{x} - \mathbf{z})$ ¹.

We apply the differential operator L to the integral equation (3), interchange integration and differentiation and use the definition of the Green's function to obtain

$$L \int_{\Omega} K(\mathbf{x}, \mathbf{z}) \varphi(\mathbf{z}) \sigma(\mathbf{z}) d\mathbf{z} = L\lambda \varphi(\mathbf{x}) \iff \int_{\Omega} \delta(\mathbf{x} - \mathbf{z}) \varphi(\mathbf{z}) \sigma(\mathbf{z}) d\mathbf{z} = \lambda L\varphi(\mathbf{x}).$$

Using the definition of the delta function this gives us

$$L\varphi(\mathbf{x}) = \frac{1}{\lambda} \sigma(\mathbf{x}) \varphi(\mathbf{x}),$$

which shows that the eigenvalues of the integral operator correspond to reciprocals of eigenvalues of the differential operator, while the corresponding eigenfunctions are the same.

We now present a simple and well-known example (on a bounded interval). The kernel in this example is sometimes referred to as the *Brownian bridge kernel* (see, e.g., [1]) since it is the covariance kernel of a Brownian motion with zero boundary conditions at both ends of the interval, also known as a Brownian bridge.

Example 1 (Brownian bridge kernel). Consider the domain $\Omega = [0, 1]$, and let

$$K(x, z) = \min(x, z) - xz = \begin{cases} x(1-z), & x \leq z, \\ z(1-x), & x > z. \end{cases}$$

This kernel may be obtained by integrating

$$-\frac{d^2}{dx^2} K(x, z) = \frac{1}{2} \delta(x - z)$$

twice using the boundary conditions

$$K(0, z) = K(1, z) = 0.$$

In other words, K is the Green's function (up to a factor 2) of the differential operator $L = -\frac{d^2}{dx^2}$ with corresponding boundary conditions.

We now consider the integral operator eigenvalue problem

$$\int_{\Omega} K(x, z) \varphi(z) \sigma(z) dz = \lambda \varphi(x)$$

¹ The problem is considerably more difficult on bounded domains, i.e., for differential equations including boundary conditions, and we do not discuss that case here.

with $\sigma(x) \equiv 1$ and K and Ω as above. Using the piecewise definition of K this corresponds to

$$\int_0^x z\varphi(z)dz + \int_x^1 x\varphi(z)dz - \int_0^1 xz\varphi(z)dz = \lambda\varphi(x).$$

If we apply the differential operator $L = -\frac{d^2}{dx^2}$ to this integral equation and use two elementary differentiation steps we obtain

$$\begin{aligned} \frac{d}{dx} \left\{ x\varphi(x) - \int_1^x \varphi(z)dz - x\varphi(x) - \int_0^1 z\varphi(z)dz \right\} &= \lambda\varphi''(x) \\ \iff -\varphi''(x) &= \frac{1}{\lambda}\varphi(x), \end{aligned}$$

which again illustrates that the eigenvalues of the integral operator are the reciprocals of the eigenvalues of the differential operator. We will continue this example below.

The second piece of the puzzle is to express the Green's function of the differential equation in terms of the eigenvalues and eigenfunctions of a related Sturm-Liouville eigenvalue problem. To this end we start with the generic ordinary differential equation

$$(LG)(x, z) = \delta(x - z)$$

with regular Sturm-Liouville boundary conditions. The so-called *Sturm-Liouville eigenvalue problem* is then given by

$$(L\varphi)(x) = \frac{1}{\lambda}\sigma(x)\varphi(x), \quad (5)$$

where we need to add the same set of regular Sturm-Liouville boundary conditions. Here σ is a weight function whose choice is basically free, but of course determines the specific form of the eigenfunctions and eigenvalues by defining different inner products. For a fixed choice of σ we can represent the Green's function G via an eigenfunction expansion of the form

$$G(x, z) = \sum_{n=1}^{\infty} c_n(z)\varphi_n(x). \quad (6)$$

We again consider z as an arbitrary, but fixed, source point and therefore the expansion coefficients (generalized Fourier coefficients) c_n will depend on z . In order to determine these coefficients we apply the differential operator L to (6) and use linearity along with the definitions of the Green's function and the Sturm-Liouville eigenvalue problem to arrive at

$$\delta(x - z) = (LG)(x, z) = \sum_{n=1}^{\infty} c_n(z)(L\varphi_n)(x) = \sum_{n=1}^{\infty} \frac{c_n(z)\sigma(x)\varphi_n(x)}{\lambda_n}.$$

Multiplication of this identity by $\varphi_m(x)$ and integration from a to b yields

$$c_n(z) = \frac{\lambda_n \varphi_n(z)}{\int_a^b \varphi_n^2(x) \sigma(x) dx},$$

where we have used the orthogonality of the eigenfunctions φ_n . If we identify the Green's function G with the kernel K in (3) then we see that the coefficients $c_n(z)$ are nothing but the generalized Fourier coefficients of G , i.e., the appropriately normalized inner product of $G(\cdot, z)$ with φ_n . In particular, if the eigenfunctions are orthonormal with respect to σ then

$$G(x, z) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x) \varphi_n(z).$$

This argument works analogously in higher space dimensions.

Example 2 (More Brownian bridge). A simple exercise in standard Sturm-Liouville theory tells us that the boundary value problem

$$-\varphi''(x) = \frac{1}{\lambda} \varphi(x), \quad \varphi(0) = \varphi(1) = 0,$$

has eigenvalues and eigenfunctions

$$\lambda_n = \frac{1}{(n\pi)^2}, \quad \varphi_n(x) = \sin n\pi x, \quad n = 1, 2, 3, \dots,$$

and we can verify

$$G(x, z) = \min(x, z) - xz = \sum_{n=1}^{\infty} c_n(z) \sin n\pi x$$

with

$$c_n(z) = \int_0^1 (\min(x, z) - xz) \sin n\pi x dx = \frac{\sin n\pi z}{(n\pi)^2} = \lambda_n \varphi_n(z).$$

2.4 Generalized Sobolev spaces

We now briefly discuss how we interpret the native space of a kernel K in terms of an associated differential operator L . Many more details are given in [44] which also appears in this volume, and in the more general paper [12]. A rigorous theoretical framework supporting this interpretation is provided in [44] for vector differential operators \mathbf{P} , and in [12] for more general vector distributional operators. While these two papers contain the theory for generalized Sobolev spaces on the unbounded domain \mathbb{R}^d , we will illustrate the framework here with some examples on bounded

domains. A theoretical framework for that (more complicated) case is the subject of [13].

Our approach depends on the ability to identify the differential operator L that corresponds to a given Green's kernel K (or vice versa). Given such an L , we then decompose it into

$$L = \mathbf{P}^{*T} \mathbf{P} = \sum_{j=1}^J P_j^* P_j,$$

where the P_j are themselves differential operators and P_j^* is an appropriately defined adjoint. For example, for the Brownian bridge kernel $K(x, z) = \min(x, z) - xz$ discussed in previous examples we have $L = -\frac{d}{dx^2}$, $\mathbf{P} = P = \frac{d}{dx}$, and $\mathbf{P}^* = P^* = -\frac{d}{dx}$, i.e., $J = 1$. We point out that the theory in [12, 44] is not limited to finite J . In particular, [44] contains an example for the Gaussian kernel whose corresponding vector differential operator \mathbf{P} is infinite-dimensional.

We then define the *generalized Sobolev space* $H_{\mathbf{P}}(\mathbb{R}^d)$ as the set of slowly increasing locally integrable functions f for which $\mathbf{P}f \in L_2(\mathbb{R}^d)$, i.e.,

$$H_{\mathbf{P}}(\mathbb{R}^d) = \left\{ f \in L_1^{loc}(\mathbb{R}^d) \cap SI \mid P_j f \in L_2(\mathbb{R}^d), j = 1, \dots, J \right\}.$$

The (semi-)inner product for this space is also defined in terms of \mathbf{P} . Namely,

$$\langle f, g \rangle_{H_{\mathbf{P}}(\mathbb{R}^d)} = \sum_{j=1}^J \int_{\mathbb{R}^d} P_j f(\mathbf{x}) \overline{P_j g(\mathbf{x})} d\mathbf{x}.$$

For our running Brownian bridge example the reproducing kernel Hilbert space is the standard Sobolev space

$$H_{0,1}^1(0,1) = H_{\mathbf{P}}(0,1) = \{ Pf = f' \in L_2(0,1) : f(0) = f(1) = 0 \}$$

with inner product

$$\langle f, g \rangle_{H_{\mathbf{P}}(0,1)} = \int_0^1 Pf(x)Pg(x)dx = \int_0^1 f'(x)g'(x)dx.$$

We then have that $K(x, z) = \min(x, z) - xz$ is the reproducing kernel of $H_{\mathbf{P}}(0, 1)$.

The left graph in Figure 1 shows multiple copies of the piecewise linear Brownian bridge kernel centered at equally spaced points in the interval $[0, 1]$. Note how this kernel is neither radial (isotropic), nor translation invariant (stationary).

We provide two more examples that are obtained from the previous one by a simple shift of the eigenvalues.

Example 3 (Tension spline kernel). We begin with the Sturm-Liouville ODE eigenvalue problem

$$\varphi''(x) + (\lambda - \varepsilon^2)\varphi(x) = 0, \quad \varphi(0) = \varphi(1) = 0,$$

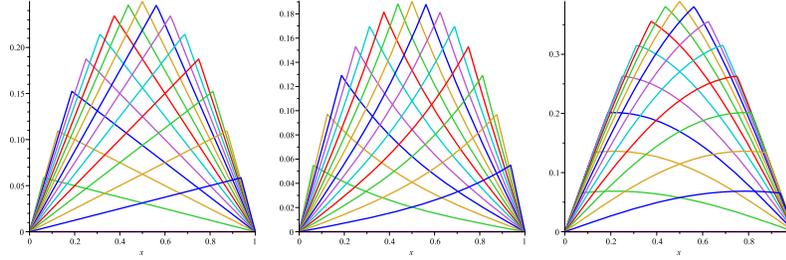


Fig. 1 Copies of the Brownian bridge (left), tension spline (middle) and relaxation spline (right) kernels $K(x, z)$ for 15 equally spaced values of z in $[0, 1]$.

where ε is an additional parameter, often referred to as *shape parameter* or *tension parameter*. This means that $L = -\frac{d}{dx^2} + \varepsilon^2 I$ and $\mathbf{P} = \left[\frac{d}{dx}, \varepsilon I\right]^T$. Note that we now indeed have a vector differential operator \mathbf{P} . The eigenvalues and eigenfunctions can easily be found to be

$$\lambda_n = n^2 \pi^2 + \varepsilon^2, \quad \varphi_n(x) = \sin n\pi x, \quad n = 1, 2, \dots$$

The kernel (i.e., Green's function) is given by

$$K(x, z) = \begin{cases} \frac{\sinh(\varepsilon x) \sinh \varepsilon(1-z)}{\varepsilon \sinh \varepsilon}, & x < z, \\ \frac{\sinh(\varepsilon z) \sinh \varepsilon(1-x)}{\varepsilon \sinh \varepsilon}, & x > z. \end{cases}$$

For this example, the reproducing kernel Hilbert space is again a standard Sobolev space, namely

$$H_{\mathbf{P}}(0, 1) = \{f, f' \in L_2(0, 1) : f(0) = f(1) = 0\}.$$

However, the inner product is now given by

$$\langle f, g \rangle_{H_{\mathbf{P}}(0,1)} = \sum_{j=1}^2 \int_0^1 P_j f(x) \overline{P_j g(x)} dx = \int_0^1 f'(x) g'(x) dx + \varepsilon \int_0^1 f(x) g(x) dx.$$

One of the most notable points here is that the so-called *shape parameter* ε of the kernel is intimately related to the inner-product and therefore the norm of the function space. Through this feature, which is usually completely ignored in the discussion of function spaces used in approximation theory, we are able to introduce a more refined notion of a function space for a certain approximation problem at hand by our ability to capture a certain length scale represented in the data. This length scale is defined by the relative importance of function values and derivatives. Therefore, we might want to denote this Sobolev space by $H_{\varepsilon}^1(0, 1)$. As a consequence, the definition of a generalized Sobolev space encodes both smoothness and “peakiness” information of the functions it contains.

The middle graph in Figure 1 shows multiple copies of the tension spline kernel centered at equally spaced points in the interval $[0, 1]$. Note how this kernel has a certain tension specified by the choice of the shape parameter ε . For the graphs in Figure 1 the value $\varepsilon = 2$ was used.

If we use the same differential operator, i.e., $L = -\frac{d}{dx^2} + \varepsilon^2 I$, but eliminate the boundary conditions, then we obtain the related radial kernel

$$\kappa(r) = \frac{1}{2\varepsilon} e^{-\varepsilon r}, \quad r = |x - z|,$$

(see [12]). This kernel is well-known in the literature as one of the members of the *Matérn* family, or as the *Ornstein-Uhlenbeck* kernel. Its reproducing kernel Hilbert space is the classical Sobolev space $H^1(\mathbb{R})$ (see also [1]).

Example 4 (Relaxation spline kernel). By adding the shift ε^2 in the Sturm-Liouville equation of the previous example instead of subtracting we obtain a different set of eigenvalues and eigenfunctions, namely

$$\lambda_n = n^2 \pi^2 - \varepsilon^2, \quad \varphi_n(x) = \sin n\pi x, \quad n = 1, 2, \dots$$

The kernel in this example is given by

$$K(x, z) = \begin{cases} \frac{\sin(\varepsilon x) \sin \varepsilon(1-z)}{\varepsilon \sin \varepsilon}, & x < z, \\ \frac{\sin(\varepsilon z) \sin \varepsilon(1-x)}{\varepsilon \sin \varepsilon}, & x > z, \end{cases}$$

and the generalized Sobolev space and inner product are defined analogously. The right graph in Figure 1 shows different copies of this kernel. Since the effects of the shape parameter here amount to a relaxation instead of a tension we chose to call this kernel a *relaxation spline* kernel.

More examples of reproducing kernels, their associated differential operators, as well as eigenvalues and eigenfunctions — also in higher space dimensions — are presented below. It should also be noted that a connection between piecewise polynomial splines and Green's functions has been mentioned in the literature before (see, e.g., [37, 41]). However, in both instances the splines were related to Green's functions of *initial value problems* whereas our framework uses Green's functions of boundary value problems.

3 Flat Limits

In this section we will take a closer look at the effect of the shape parameter ε present in the definition of some of our kernels. In particular, we are interested in understanding the behavior of *radial* kernel interpolants for the limiting case of $\varepsilon \rightarrow 0$, i.e., flat kernels. A radial kernel is of the form $\kappa(\|\mathbf{x} - \mathbf{z}\|) = K(\mathbf{x}, \mathbf{z})$, i.e., it is

invariant under both translation and rotation. In the statistics literature such a kernel is called stationary and isotropic.

The results in this section specifically address the scattered data interpolation problem. In other words, we are given data sites $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset \mathbb{R}^d$ with associated data values $\{f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)\}$ sampled from some function f and wish to reconstruct f by a function of the form

$$s_{f,\kappa}^\varepsilon(\mathbf{x}) = \sum_{j=1}^N c_j \kappa(\varepsilon \|\mathbf{x} - \mathbf{x}_j\|), \quad \mathbf{x} \in \mathbb{R}^d,$$

where the coefficients c_j are determined by satisfying the interpolation conditions

$$s_{f,\kappa}^\varepsilon(\mathbf{x}_i) = f(\mathbf{x}_i), \quad i = 1, \dots, N.$$

3.1 Infinitely smooth RBFs

In recent years so-called *flat radial basis functions* (RBFs) have received much attention in the case when the kernels are infinitely smooth (see, e.g., [5, 16, 21, 22, 23, 33]). We begin by summarizing the essential insight gained in these papers, and then present some recent results from [38] that deal with radial kernels of finite smoothness in the next subsection.

Theorem 2. *Assume the positive definite radial kernel κ has an expansion of the form*

$$\kappa(r) = \sum_{n=0}^{\infty} a_n r^{2n}$$

into even powers of r (i.e., κ is infinitely smooth), and that the data \mathcal{X} are unisolvent with respect to any set of N linearly independent polynomials of degree at most m . Then

$$\lim_{\varepsilon \rightarrow 0} s_{f,\kappa}^\varepsilon(\mathbf{x}) = p_{m,f}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d,$$

where $p_{m,f}$ is determined as follows:

- *If interpolation with polynomials of degree at most m is unique, then $p_{m,f}$ is that unique polynomial interpolant.*
- *If interpolation with polynomials of degree at most m is not unique, then $p_{m,f}$ is a polynomial interpolant whose form depends on the choice of RBF.*

This theorem applies to kernels such as

$$\kappa(\varepsilon r) = \frac{1}{1 + \varepsilon^2 r^2} = 1 - (\varepsilon r)^2 + (\varepsilon r)^4 - (\varepsilon r)^6 + (\varepsilon r)^8 + \dots \quad (\text{IQ}),$$

$$\kappa(\varepsilon r) = e^{-\varepsilon^2 r^2} = 1 - (\varepsilon r)^2 + \frac{1}{2}(\varepsilon r)^4 - \frac{1}{6}(\varepsilon r)^6 + \frac{1}{24}(\varepsilon r)^8 + \dots \quad (\text{Gaussian}),$$

$$\kappa(\varepsilon r) = \frac{1}{\sqrt{1 + \varepsilon^2 r^2}} = 1 - \frac{1}{2}(\varepsilon r)^2 + \frac{3}{8}(\varepsilon r)^4 - \frac{5}{16}(\varepsilon r)^6 + \frac{35}{128}(\varepsilon r)^8 + \dots \quad (\text{IMQ}).$$

The implications of this theorem are quite deep since it essentially establishes radial basis functions as generalizations of polynomial spectral methods. As a consequence, this opens the door to the design of algorithms for function approximation as well as the numerical solution of partial differential equations that are more accurate than the standard polynomial spectral methods. Moreover, the scattered data setting in which radial basis functions are used allows for more flexibility with respect to geometry and adaptivity.

We will come back to the Gaussian kernel in the next two sections of this paper where we address two important issues: computational stability and rates of convergence.

3.2 Finitely smooth RBFs

To our knowledge, the flat limit of RBFs with finite smoothness was not studied until the recent paper [38] in which interpolation on \mathbb{R}^d was investigated.

Before we explain the results obtained in [38], we look at a few finitely smooth radial kernels as full space Green's functions as discussed in the earlier sections.

Example 5 (Radial kernels with finite smoothness).

1. We have already mentioned the univariate C^0 Matérn kernel $K(x, z) \doteq e^{-\varepsilon|x-z|}$. For this first example we remember that the differential operator L associated with this full-space Green's function was given by

$$L = -\frac{d^2}{dx^2} + \varepsilon^2 I.$$

On the other hand, it is well-known that univariate C^0 piecewise linear splines may be expressed in terms of kernels of the form $K(x, z) \doteq |x - z|$. The corresponding differential operator in this case is

$$L = -\frac{d^2}{dx^2}.$$

Note that the differential operator associated with the Matérn kernel “converges” to that of the piecewise linear splines as $\varepsilon \rightarrow 0$. We also remark that the piecewise linear Brownian bridge kernel does not fit into this discussion since it is associated with a boundary value problem, i.e., it is not a full-space Green's function.

2. The univariate C^2 tension spline kernel [30] $K(x, z) \doteq e^{-\varepsilon|x-z|} + \varepsilon|x-z|$ is the Green's kernel of

$$L = -\frac{d^4}{dx^4} + \varepsilon^2 \frac{d^2}{dx^2},$$

while the univariate C^2 cubic spline kernel $K(x, z) \doteq |x-z|^3$ corresponds to

$$L = -\frac{d^4}{dx^4}.$$

Again, the differential operator associated with the tension spline “converges” to that of the cubic spline as $\varepsilon \rightarrow 0$.

3. In [1] we find a so-called univariate *Sobolev kernel* of the form $K(x, z) \doteq e^{-\varepsilon|x-z|} \sin\left(\varepsilon|x-z| + \frac{\pi}{4}\right)$ which is associated with

$$L = -\frac{d^4}{dx^4} - \varepsilon^2 I.$$

The operator for this kernel also “converges” to the cubic spline kernel, but the effect of the scale parameter is analogous to that of the relaxation spline of Example 4.

4. The general *multivariate Matérn kernels* are of the form

$$K(\mathbf{x}, \mathbf{z}) \doteq K_{m-d/2}(\varepsilon\|\mathbf{x}-\mathbf{z}\|)(\varepsilon\|\mathbf{x}-\mathbf{z}\|)^{m-d/2}, \quad m > \frac{d}{2},$$

and can be obtained as Green's kernels of (see [44])

$$L = (\varepsilon^2 I - \Delta)^m, \quad m > \frac{d}{2}.$$

We contrast this with the *polyharmonic spline kernels*

$$K(\mathbf{x}, \mathbf{z}) \doteq \begin{cases} \|\mathbf{x}-\mathbf{z}\|^{2m-d}, & d \text{ odd,} \\ \|\mathbf{x}-\mathbf{z}\|^{2m-d} \log \|\mathbf{x}-\mathbf{z}\|, & d \text{ even,} \end{cases}$$

and

$$L = (-1)^m \Delta^m, \quad m > \frac{d}{2}.$$

In summary, all of these examples show that the differential operators associated with finitely smooth RBF kernels “converge” to those of a piecewise polynomial or polyharmonic spline kernel as $\varepsilon \rightarrow 0$. This motivates us to ask whether RBF interpolants based on finitely smooth kernels converge to (polyharmonic) spline interpolants for $\varepsilon \rightarrow 0$ mimicking the relation between infinitely smooth radial kernels and polynomials. As the following theorem shows, this is indeed true.

As mentioned in Theorem 2, infinitely smooth radial kernels can be expanded into an infinite series of even powers of r . Finitely smooth radial kernels can also be expanded into an infinite series of powers of r . However, in this case there al-

ways exists some minimal odd power of r with nonzero coefficient indicating the smoothness of the kernel. For example, for univariate C^0 , C^2 and C^4 Matérn kernels, respectively, we have

$$\begin{aligned}\kappa(\varepsilon r) &\doteq e^{-\varepsilon r} = 1 - \varepsilon r + \frac{1}{2}(\varepsilon r)^2 - \frac{1}{6}(\varepsilon r)^3 + \dots, \\ \kappa(\varepsilon r) &\doteq (1 + \varepsilon r)e^{-\varepsilon r} = 1 - \frac{1}{2}(\varepsilon r)^2 + \frac{1}{3}(\varepsilon r)^3 - \frac{1}{8}(\varepsilon r)^4 + \dots, \\ \kappa(\varepsilon r) &\doteq (3 + 3\varepsilon r + (\varepsilon r)^2)e^{-\varepsilon r} = 3 - \frac{1}{2}(\varepsilon r)^2 + \frac{1}{8}(\varepsilon r)^4 - \frac{1}{15}(\varepsilon r)^5 + \frac{1}{48}(\varepsilon r)^6 + \dots.\end{aligned}$$

Theorem 3 ([38]). *Suppose κ is conditionally positive definite of order $m \leq n$ with an expansion of the form*

$$\kappa(r) = a_0 + a_2 r^2 + \dots + a_{2n} r^{2n} + a_{2n+1} r^{2n+1} + a_{2n+2} r^{2n+2} + \dots,$$

where $2n+1$ denotes the smallest odd power of r present in the expansion (i.e., κ is finitely smooth). Also assume that the data \mathcal{X} contain a unisolvent set with respect to the space $\pi_{2n}(\mathbb{R}^d)$ of d -variate polynomials of degree less than $2n$. Then

$$\lim_{\varepsilon \rightarrow 0} s_{f,\kappa}^\varepsilon(\mathbf{x}) = \sum_{j=1}^N c_j \|\mathbf{x} - \mathbf{x}_j\|^{2n+1} + \sum_{k=1}^M d_k p_k(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d,$$

where $\{p_k \mid k = 1, \dots, M\}$ denotes a basis of $\pi_n(\mathbb{R}^d)$.

In other words, the “flat” limit of a piecewise smooth RBF interpolant is nothing but a polyharmonic spline interpolant. Therefore, just as infinitely smooth RBFs can be interpreted as generalizations of polynomials, we can view finitely smooth RBFs as generalizations of piecewise polynomial (or more generally polyharmonic) splines.

We point out that Theorem 3 does not cover Matérn kernels with odd-order smoothness. However, all other examples listed above are covered by the theorem.

Figure 2 illustrates the convergence of univariate C^0 and C^2 Matérn interpolants to piecewise linear and piecewise cubic spline interpolants, respectively.

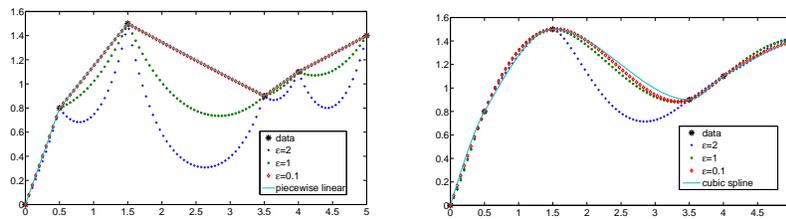


Fig. 2 Convergence of C^0 (left) and C^2 (right) Matérn interpolants to piecewise linear (left) and cubic (right) spline interpolants.

4 Stable Computation

We now look at some practical consequences of working with “flat” RBF kernels. It is well-known that interpolation with “flat” Gaussian kernels leads to a notoriously ill-conditioned interpolation matrix K . This is due to the fact that the standard set of basis functions $\{e^{-\varepsilon^2(x-x_j)^2}, j = 1, \dots, N\}$ becomes numerically linearly dependent. It needs to be emphasized that the resulting numerical instabilities are due only to this “bad” choice of basis and not to the choice of function space itself. In fact, we will discuss in the next section how well one can approximate with linear combinations of Gaussians.

Even though Gaussian kernels are rather popular — especially in the machine learning community, it has been widely accepted that working with Gaussians is an ill-conditioned problem. As a result, the literature contains many references to a so-called *uncertainty* or *trade-off principle* (see, e.g., [31] or the more recent paper [35]). This uncertainty principle, however, is tied directly to the use of the standard (“bad”) basis, and we demonstrate below how it can be circumvented by choosing a better — orthonormal — basis. The following discussion is motivated by the recent work of Bengt Fornberg and his collaborators [14, 15] in which they have proposed a so-called RBF-QR algorithm which allows for stable RBF computations. In addition to this QR-based approach they have also proposed other stable algorithms such as the Contour-Padé algorithm [16]. The guiding principle in this work is always the fact that the RBF-direct algorithm (based on the use of the “bad” standard basis) is ill-conditioned, but the RBF interpolation problem itself is not.

4.1 An eigenfunction expansion for Gaussians

In [29] (and already [45], albeit with incorrect normalization) one can find the following general eigenfunction expansion

$$e^{-b(x-z)^2} = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x) \varphi_n(z), \quad (7)$$

where the eigenfunctions φ_n are orthonormal in $L_2(\mathbb{R}, \rho)$ with weight function

$$\rho(x) = \sqrt{\frac{2a}{\pi}} e^{-2ax^2}.$$

Here a and b are arbitrary positive numbers. If we let $c = \sqrt{a^2 + 2ab}$, then the eigenfunctions φ_n turn out to be

$$\varphi_n(x) = \frac{1}{\sqrt{2^{n-1}(n-1)!} \sqrt{\frac{a}{c}}} e^{-(c-a)x^2} H_{n-1}(\sqrt{2c}x), \quad n = 1, 2, \dots,$$

with H_n the classical Hermite polynomials of degree n , i.e.,

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \quad \text{for all } x \in \mathbb{R}, \quad n = 0, 1, 2, \dots$$

so that

$$\int_{\mathbb{R}} H_n^2(x) e^{-x^2} dx = \sqrt{\pi} 2^n n! \quad \text{for } n = 0, 1, 2, \dots$$

The corresponding eigenvalues are

$$\lambda_n = \sqrt{\frac{2a}{a+b+c}} \left(\frac{b}{a+b+c} \right)^{n-1}, \quad n = 1, 2, \dots$$

In particular, we will want to use the Gaussian kernel in its usual form with shape parameter ε as

$$K(x, z) = e^{-\varepsilon^2(x-z)^2}$$

so that $b = \varepsilon^2$. Moreover, we take $a = \frac{1}{2}$ and therefore $c = \frac{1}{2}\sqrt{1+4\varepsilon^2}$ (see also [9]). This leads to eigenvalues

$$\begin{aligned} \lambda_n &= \frac{1}{\sqrt{\frac{1}{2}(1+\sqrt{1+4\varepsilon^2}) + \varepsilon^2}} \left(\frac{\varepsilon^2}{\frac{1}{2}(1+\sqrt{1+4\varepsilon^2}) + \varepsilon^2} \right)^{n-1} \\ &= \frac{\varepsilon^{2(n-1)}}{\left(\frac{1}{2}(1+\sqrt{1+4\varepsilon^2}) + \varepsilon^2 \right)^{n-\frac{1}{2}}}, \quad n = 1, 2, \dots \end{aligned} \quad (8)$$

and eigenfunctions

$$\varphi_n(x) = \sqrt{\frac{(1+4\varepsilon^2)^{1/4}}{2^{n-1}(n-1)!}} \exp\left(-\frac{\varepsilon^2 x^2}{\frac{1}{2}(1+\sqrt{1+4\varepsilon^2})}\right) H_{n-1}\left((1+4\varepsilon^2)^{1/4} x\right). \quad (9)$$

4.2 The RBF-QR algorithm

The starting point for the bivariate Gaussian RBF-QR algorithm in [14] was an expansion of the form

$$e^{-\varepsilon^2(x-z)^2} = \sum_{n=0}^{\infty} \frac{(2\varepsilon^2)^n}{n!} x^n e^{-\varepsilon^2 x^2} z^n e^{-\varepsilon^2 z^2}, \quad x, z \in \mathbb{R}. \quad (10)$$

However, the authors claimed that this series is not ideal since, coupled with the RBF-QR strategy described below, it does not provide an effective reduction of the conditioning of the RBF interpolation algorithm. Most likely, the poor conditioning of the new basis that results from this expansion is due to the fact that the func-

tions $x \mapsto x^n e^{-\varepsilon^2 x^2}$ are not orthogonal in $L_2(\mathbb{R})$. Indeed, for $\varepsilon \rightarrow 0$ these functions converge to the standard monomial basis which is known to be ill-conditioned (cf. Vandermonde matrices). Therefore, the authors followed up their initial expansion with a transformation to polar coordinates and an expansion in terms of Chebyshev polynomials. This leads to an RBF-QR algorithm for Gaussians that is indeed stable, but limited to problems in \mathbb{R}^2 .

The following discussion based on the eigenfunction expansion (7) of the Gaussian kernel will be applicable in any space dimension. Due to the product nature of the kernel we describe only the 1D version here. A comment at the end of this section indicates how to approach the general multivariate setting.

We will now show that if we use an expansion of the kernel in terms of orthonormal (eigen-)functions, then the source of ill-conditioning of the Gaussian basis is moved entirely into its eigenvalues. Since the eigenvalues of the Gaussian kernel decay very quickly we are now able to directly follow the QR-based strategy suggested in [14] — without the need for any additional transformation to Chebyshev polynomials.

In particular, we use the eigenvalues (8) and eigenfunctions (9) of the Gaussian kernel as discussed above.

The QR-based algorithm of [14] corresponds to the following. Starting with an expansion of the basis functions centered at x_j , $j = 1, \dots, N$, of the form

$$K(x, x_j) = \sum_{n=1}^{\infty} \varepsilon^{2(n-1)} b_n(x_j) \varphi_n(x), \quad b_n(x_j) := \varepsilon^{-2(n-1)} \lambda_n \varphi_n(x_j),$$

i.e., a generalized Fourier expansion with x_j -dependent Fourier coefficients, we obtain

$$\begin{bmatrix} K(x, x_1) \\ K(x, x_2) \\ \vdots \\ \vdots \\ K(x, x_N) \end{bmatrix} = \begin{bmatrix} \cdots & \cdots \\ \cdots & \cdots \\ \cdots & \mathbf{B} & \cdots \\ \cdots & \cdots \\ \cdots & \cdots \end{bmatrix} \begin{bmatrix} \varepsilon^0 & & & & \\ & \varepsilon^2 & & & \\ & & \ddots & & \\ & & & \varepsilon^{2n} & \\ & & & & \ddots \end{bmatrix} \begin{bmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \vdots \\ \varphi_n(x) \\ \vdots \end{bmatrix}.$$

Using more compact matrix-vector notation we can denote this by

$$\mathbf{k}(x) = \mathbf{B}\mathbf{E}\phi(x), \quad (11)$$

where $\mathbf{k}(x) = (K(x, x_j))_{j=1}^N$ and $\phi(x) = (\varphi_n(x))_{n=1}^{\infty}$ are the vectors of standard basis functions and eigenfunctions, respectively, evaluated at x , $\mathbf{B} = (b_n(x_j))_{j=1, n=1}^{N, \infty}$, and $\mathbf{E} = \text{diag}(\varepsilon^0, \varepsilon^2, \dots, \varepsilon^{2n}, \dots)$ is the diagonal matrix of increasing even powers of ε . Note that \mathbf{E} and ϕ are infinite and need to be appropriately truncated for practical applications. The matrix \mathbf{B} has N rows, but infinitely many columns. However, since we are working with an eigenfunction expansion, truncating the representation at M terms will provide the best (in the L_2 -sense) M -term approximation to the full series. Note that since

$$\begin{aligned} B_{jn} &= \left(\frac{2}{1 + \sqrt{1 + 4\varepsilon^2} + 2\varepsilon^2} \right)^{n-\frac{1}{2}} \sqrt{\frac{(1 + 4\varepsilon^2)^{1/4}}{2^{n-1}(n-1)!}} e^{-\frac{2\varepsilon^2 x_j^2}{1 + \sqrt{1 + 4\varepsilon^2}}} H_{n-1} \left((1 + 4\varepsilon^2)^{1/4} x_j \right) \\ &\rightarrow \frac{1}{\sqrt{2^{n-1}(n-1)!}} H_{n-1}(x_j) \quad \text{as } \varepsilon \rightarrow 0 \end{aligned}$$

the matrix \mathbf{B} remains “nice” as $\varepsilon \rightarrow 0$. Moreover, this limiting relation is another indication of the polynomial limit of Gaussian kernel interpolation as discussed in the previous section.

The QR idea now consists in first computing the QR-decomposition of \mathbf{B} , i.e.,

$$\mathbf{B} = \mathbf{Q}\mathbf{R}$$

with unitary matrix \mathbf{Q} and upper triangular matrix \mathbf{R} . Next, we multiply the relation $\mathbf{k}(x) = \mathbf{B}\mathbf{E}\phi(x)$ on both sides by the non-singular matrix $\mathbf{E}^{-1}\mathbf{Q}^*$. The crucial observation here is that this does not change the function space spanned by the (poorly conditioned) standard basis functions $K(\cdot, x_1), \dots, K(\cdot, x_N)$.

As a result, using (11), the QR-decomposition of \mathbf{B} and the fact that \mathbf{Q} is unitary we obtain a new basis for the Gaussian approximation space, namely

$$\boldsymbol{\psi}(x) = \mathbf{E}^{-1}\mathbf{Q}^*\mathbf{k}(x) = \mathbf{E}^{-1}\mathbf{Q}^*\mathbf{Q}\mathbf{R}\mathbf{E}\phi(x) = \mathbf{E}^{-1}\mathbf{R}\mathbf{E}\phi(x),$$

where $\boldsymbol{\psi}(x) = (\psi_n(x))_{n=1}^\infty$. Note that the matrix $\mathbf{E}^{-1}\mathbf{R}\mathbf{E}$ is upper triangular and due to the scaling from the left and right should be relatively well-conditioned.

We are currently in the process of implementing this algorithm [10], and preliminary tests indicate that it is now possible to compute Gaussian RBF interpolants with this new eigenfunction basis stably also in the “flat” limit as $\varepsilon \rightarrow 0$. Incidentally, this is precisely the approach taken in [15] for stable radial basis function approximation on the sphere. It is interesting to note that traditionally there has been a much closer connection between (zonal) kernels used on the sphere and spherical harmonics, i.e., the eigenfunctions of the Laplace-Beltrami operator on the sphere (see, e.g., [11]). Furthermore, the RBF-QR approach should be successfully applicable whenever an eigenfunction expansion of the kernel is available.

As mentioned above, for the d -variate case we can use the fact that the Gaussian is a tensor product kernel:

$$K(\mathbf{x}, \mathbf{z}) = e^{-\varepsilon_1^2(x_1-z_1)^2 - \dots - \varepsilon_d^2(x_d-z_d)^2} = \sum_{\mathbf{n} \in \mathbb{N}^d} \lambda_{\mathbf{n}} \varphi_{\mathbf{n}}(\mathbf{x}) \varphi_{\mathbf{n}}(\mathbf{z}),$$

so that the multivariate eigenvalues and eigenvectors are simply the products of the one-dimensional ones, i.e.,

$$\lambda_{\mathbf{n}} = \prod_{\ell=1}^d \lambda_{n_\ell} \quad \text{and} \quad \varphi_{\mathbf{n}}(\mathbf{x}) = \prod_{\ell=1}^d \varphi_{n_\ell}(x_\ell).$$

One of the advantages — both practical and theoretical — of this product approach is that we can take different shape parameters ε_ℓ for different dimensions, i.e., we can employ an *anisotropic* kernel K . Of course, the isotropic (or radial) case can still be recovered if we choose $\varepsilon_\ell = \varepsilon$, $\ell = 1, \dots, d$. We will exploit this ability to generalize to anisotropic Gaussian kernels in the next section on convergence rates.

5 Dimension Independent Error Bounds

In the last section of this paper we mention some new results (see [9] for much more details) on the rates of convergence of Gaussian kernel approximation. To be more specific, we will address weighted L_2 approximation when the data is specified either by function values of an unknown function f (from the native space of the kernel) or with the help of arbitrary linear functionals. Our convergence results pay special attention to the dependence of the estimates on the space dimension d . We will see that the use of anisotropic Gaussian kernels instead of isotropic ones provides improved convergence rates. It should also be mentioned that the work in [9] deals with linear approximation algorithms, while the recent paper [17] addresses nonlinear Gaussian approximation.

5.1 The current situation

A good resource for standard RBF scattered data approximation results up to the year 2005 is [43]. There we can find two different L_∞ error bounds for isotropic Gaussian interpolation to data sampled from a function f in the native space $\mathcal{H}(K, \Omega)$ of the Gaussian. Both of these results are formulated in terms of the *fill distance*

$$h_{\mathcal{X}, \Omega} = \sup_{\mathbf{x} \in \Omega} \min_{1 \leq j \leq N} \|\mathbf{x} - \mathbf{x}_j\|,$$

where $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ denotes the set of data sites as before. Since the results we mention below are in terms of N , the number of data, we will restate the error bounds from [43] also in terms of N using the fact that for quasi-uniformly distributed data sites we have $h_{\mathcal{X}, \Omega} = \mathcal{O}(N^{-1/d})$.

If f has derivatives up to total order p and $s_{f,K}$ is the interpolant based on the Gaussian kernel $K(\mathbf{x}, \mathbf{z}) = e^{-\varepsilon^2 \|\mathbf{x} - \mathbf{z}\|^2}$, i.e.,

$$s_{f,K}(\mathbf{x}) = \sum_{j=1}^N c_j K(\mathbf{x}, \mathbf{x}_j) \quad \text{such that} \quad s_{f,K}(\mathbf{x}_i) = f(\mathbf{x}_i), \quad i = 1, \dots, N,$$

then the first error bound is of the form

$$\|f - s_{f,K}\|_\infty \leq C_d N^{-p/d} \|f\|_{\mathcal{H}(K, \Omega)}$$

with some possibly dimension-dependent constant C_d . Therefore, infinitely smooth functions can be approximated with order $p = \infty$. With some extra effort one can also obtain the spectral estimate

$$\|f - s_{f,K}\|_\infty \leq e^{-\frac{c}{d}N^{1/d} \log N} \|f\|_{\mathcal{H}(K,\Omega)}.$$

It is apparent from both of these bounds that the rate of convergence deteriorates as d increases. Moreover, the dependence of the constants on d is not clear. Therefore, these kinds of error bounds — and in fact almost all error bounds in the RBF literature — suffer from the *curse of dimensionality*. We will now present some results from [9] on *dimension-independent* convergence rates for Gaussian kernel approximation.

5.2 New results on (minimal) worst-case weighted L_2 error

As already indicated above, we will make several assumptions in order to be able to obtain dimension-independent error bounds.

We define the worst-case weighted $L_{2,\rho}$ error as

$$\text{err}_{2,\rho}^{\text{wc}} = \sup_{\|f\|_{\mathcal{H}(K,\mathbb{R}^d)} \leq 1} \|f - s_{f,K}\|_{2,\rho},$$

where $s_{f,K}$ is our kernel (minimum norm) approximation calculated in the usual way. Therefore

$$\|f - s_{f,K}\|_{2,\rho} \leq \text{err}_{2,\rho}^{\text{wc}} \|f\|_{\mathcal{H}(K,\mathbb{R}^d)} \quad \text{for all } f \in \mathcal{H}(K,\mathbb{R}^d).$$

The N^{th} *minimal worst case error* $\text{err}_{2,\rho}^{\text{wc}}(N)$ refers to the worst case error that can be achieved with an optimal design, i.e., data generated by N optimally chosen linear functionals. For function approximation this means that the data sites have to be chosen in an optimal way. The results in [9] are non-constructive, i.e., no such optimal design is specified. However, a Smolyak or sparse grid algorithm is a natural candidate for such a design. If we are allowed to choose arbitrary linear functionals, then the optimal choice for weighted L_2 approximation is known. In this case we use generalized Fourier coefficients, i.e., the optimal linear functionals are $L_j = \langle \cdot, \varphi_j \rangle_{\mathcal{H}(K,\mathbb{R}^d)}$ and we obtain the truncated generalized Fourier series approximation

$$s_{f,K}(\mathbf{x}) = \sum_{n=1}^N \langle f, \varphi_n \rangle_{\mathcal{H}(K,\mathbb{R}^d)} \varphi_n(\mathbf{x}) \quad \text{for all } f \in \mathcal{H}(K,\mathbb{R}^d),$$

where

$$K(\mathbf{x}, \mathbf{z}) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(\mathbf{x}) \varphi_n(\mathbf{z}), \quad \int_{\Omega} K(\mathbf{x}, \mathbf{z}) \varphi_n(\mathbf{z}) \rho(\mathbf{z}) d\mathbf{z} = \lambda_n \varphi_n(\mathbf{x}).$$

It is then known [26] that

$$\text{err}_{2,\rho}^{\text{wc}}(N) = \sqrt{\lambda_{N+1}},$$

the $(N+1)^{\text{st}}$ largest eigenvalue, which is easy to identify in the univariate case, but takes some care to specify in the multivariate setting.

In [9] it is then proved that in the isotropic case, i.e., with a truly radial Gaussian kernel of the form

$$K(\mathbf{x}, \mathbf{z}) = e^{-\varepsilon^2 \|\mathbf{x}-\mathbf{z}\|^2}$$

one can approximate

- function data with an N^{th} minimal error of the order $\mathcal{O}(n^{-1/4+\delta})$, and
- Fourier data (i.e., arbitrary linear functional data) with an N^{th} minimal error of the order $\mathcal{O}(n^{-1/2+\delta})$.

Here the constants in the \mathcal{O} -notation do not depend on the dimension d and δ is arbitrarily small.

With anisotropic kernels, i.e.,

$$K(\mathbf{x}, \mathbf{z}) = e^{-\varepsilon_1^2(x_1-z_1)^2 - \dots - \varepsilon_d^2(x_d-z_d)^2}$$

one can do much better. In this case, if the shape parameters decay like $\varepsilon_\ell = \ell^{-\alpha}$, then one can approximate

- function data with an N^{th} minimal error of the order $\mathcal{O}(n^{-\max(\frac{\alpha^2}{2+\alpha}, 1/4)+\delta})$, and
- Fourier data (i.e., arbitrary linear functional data) with an N^{th} minimal error of the order $\mathcal{O}(n^{-\max(\alpha, 1/2)+\delta})$.

Again, the constants in the \mathcal{O} -notation do not depend on the dimension d .

In order to prove the above results it was essential to have the eigenvalues (cf. (8))

$$\lambda_n = \frac{\varepsilon^{2(n-1)}}{\left(\frac{1}{2}(1 + \sqrt{1 + 4\varepsilon^2}) + \varepsilon^2\right)^{n-\frac{1}{2}}}, \quad n = 1, 2, \dots,$$

and eigenfunctions (cf. (9))

$$\varphi_n(x) = \sqrt{\frac{(1 + 4\varepsilon^2)^{1/4}}{2^{n-1}(n-1)!}} \exp\left(-\frac{\varepsilon^2 x^2}{\frac{1}{2}(1 + \sqrt{1 + 4\varepsilon^2})}\right) H_{n-1}\left(\left(1 + 4\varepsilon^2\right)^{1/4} x\right)$$

of the univariate Gaussian kernel $K(x, z) = e^{-\varepsilon^2(x-z)^2}$. As mentioned in the previous section, the multivariate (and anisotropic) case can be handled using products of univariate eigenvalues and eigenfunctions.

Even if we do not have an eigenfunction expansion of a specific kernel available, the work of [9] shows that for any radial (isotropic) kernel one has a dimension-independent Monte-Carlo type convergence rate of $\mathcal{O}(n^{-1/2+\delta})$ provided arbitrary linear functionals are allowed to generate the data. For translation-invariant (stationary) kernels the situation is similar. However, the constant in the \mathcal{O} -notation depends

— in any case — on the sum of the eigenvalues of the kernel. For the radial case this sum is simply $\kappa(0)$ (independent of d), while for general translation invariant kernels it is $\tilde{K}(\mathbf{0})$, which may depend on d .

These results show that — even though RBF methods are often advertised as being “dimension-blind” — their rates of convergence are only excellent (i.e., spectral for infinitely smooth kernels) if the dimension d is small. For large dimensions the constants in the \mathcal{O} -notation take over. If one, however, permits an anisotropic scaling of the kernel (i.e., elliptical symmetry instead of strict radial symmetry) and if those scale parameters decay rapidly with increasing dimension, then excellent convergence rates for approximation of smooth functions can be maintained independent of d .

6 Summary

In this paper we have attempted to shed some new light on the connections between piecewise polynomial splines and approximation methods based on reproducing kernels and radial basis functions in particular. Using Mercer’s theorem and the resulting eigenfunction expansions of positive definite kernels along with an interpretation of these kernels as Green’s functions of appropriate differential operators we provided a new interpretation of RBF native spaces as generalized Sobolev spaces (cf. [12, 13, 44]). As a result we have a more intuitive interpretation of RBF native spaces in terms of the smoothness of the functions they contain. Moreover, special attention is paid to the native space norm and how it encodes information of the inherent scale of the functions it contains.

Extreme scaling of kernels, i.e., “flat” limits are investigated and they provide a new connection between finitely smooth RBF kernels and piecewise polynomial or polyharmonic splines (see [38]). We also use the eigenfunction expansions to move Fornberg’s RBF-QR algorithm onto a more standard theoretical foundation which provides at the same time an algorithm for Gaussians that is applicable in any space dimension.

Finally, we discussed some of the results of [9] on dimension-independent convergence rates for Gaussians. The main insight obtained from these results is that one needs to allow for the use of an anisotropic scaling of the kernel with rapidly decaying scale parameters in order to be able to guarantee high rates of convergence in high space dimensions.

There is still much work to be done in the future. The theoretical framework for Green’s functions on bounded domains needs to be completed, the new RBF-QR algorithm for Gaussians needs to be implemented, and the hunt for kernels with readily available or relatively easily computable eigenfunction expansions is on. Any such kernel benefits from the ideas laid out for stable computation and dimension-independent error bounds. There is also room to generalize the results on flat limits of piecewise smooth RBF kernels. Finally, it is expected that the eigenfunction expansions discussed here can be exploited to obtain fast multipole-type algorithms.

Acknowledgements The work presented in this paper was supported by NSF Grant DMS–0713848. The author is also indebted to Fred Hickernell, Mike McCourt, John Riddle, Guohui Song, Henryk Woźniakowski, and Qi Ye, all of whom played important roles in various parts of this paper.

References

1. Berline, A., Thomas-Agnan, C.: *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer, Dordrecht (2004)
2. de Boor, C.: *A Practical Guide to Splines*. Springer, New York (1978, revised edition 2001)
3. Buhmann, M.D.: *Radial Basis Functions: Theory and Implementations*. Cambridge University Press, Cambridge (2003)
4. DeVore, R., Ron, A.: Approximation using scattered shifts of a multivariate function. *Trans. Amer. Math. Soc.*, DOI 10.1090/S0002-9947-2010-05070-6 (2010)
5. Driscoll, T.A., Fornberg, B.: Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.* **43**, 413–422 (2002)
6. Duffy, D.G.: *Green's Functions with Applications*. Chapman & Hall/CRC, Boca Raton (2001)
7. Dyn, N., Levin, D., Rippa, S.: Numerical procedures for surface fitting of scattered data by radial functions. *SIAM J. Sci. Statist. Comput.* **7**, 639–659 (1986)
8. Fasshauer, G.E.: *Meshfree Approximation Methods with MATLAB*. World Scientific Publishers, Singapore (2007)
9. Fasshauer, G.E., Hickernell, F.J., Woźniakowski, H.: Rate of convergence and tractability of the radial function approximation problem. Submitted
10. Fasshauer, G.E., McCourt, M.J.: Stable evaluation of Gaussian RBF interpolants. In preparation
11. Fasshauer, G.E., Schumaker, L.L.: Scattered data fitting on the sphere. In: Dæhlen, M., Lyche, T., Schumaker, L.L. (eds.) *Mathematical Methods for Curves and Surfaces II*, pp. 117–166. Vanderbilt University Press, Nashville (1998)
12. Fasshauer, G.E., Ye, Q.: Reproducing kernels of generalized Sobolev spaces via a Green function approach with distributional operators. Submitted
13. Fasshauer, G.E., Ye, Q.: Reproducing kernels of Sobolev spaces in bounded domains via a Green's kernel approach. In preparation
14. Fornberg, B., Larsson, E., Flyer, N.: Stable computations with Gaussian radial basis functions in 2-D. Technical Report 2009-020, Uppsala University, Department of Information Technology.
15. Fornberg, B., Piret, C.: A stable algorithm for flat radial basis functions on a sphere. *SIAM J. Sci. Comp.* **30**, 60–80 (2007)
16. Fornberg, B., Wright, G.: Stable computation of multiquadric interpolants for all values of the shape parameter. *Comput. Math. Appl.* **47**, 497–523 (2004)
17. Hangelbroek, T., Ron, A.: Nonlinear approximation using Gaussian kernels. *J. Funct. Anal.* **259** no. 1, 203–219 (2010)
18. Iske, A.: *Multiresolution Methods in Scattered Data Modelling*. Lecture Notes in Computational Science and Engineering 37, Springer Verlag, Berlin (2004)
19. Kybic, J., Blu, T., Unser, M.: Generalized sampling: A variational approach — Part I: Theory. *IEEE Trans. Signal Proc.* **50**, 1965–1976 (2002)
20. Lai, M.J., Schumaker, L.L.: *Spline Functions on Triangulations*. Cambridge University Press, Cambridge (2007)
21. Larsson, E., Fornberg, B.: A numerical study of some radial basis function based solution methods for elliptic PDEs. *Comput. Math. Appl.* **46**, 891–902 (2003)
22. Larsson, E., Fornberg, B.: Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. *Comput. Math. Appl.* **49**, 103–130 (2005)

23. Lee, Y.J., Yoon, G.J., Yoon, J.: Convergence of increasingly flat radial basis interpolants to polynomial interpolants. *SIAM J. Math. Anal.* **39**, 537–553 (2007)
24. Mercer, J.: Functions of positive and negative type, and their connection with the theory of integral equations. *Phil. Trans. Royal Soc. London Series A* **209**, 415–446 (1909)
25. Mhaskar, H.N., Narcowich, F.J., Prestin, J., Ward, J.D.: L^p Bernstein estimates and approximation by spherical basis functions. *Math. Comp.* **79**, 1647–1679 (2010)
26. Novak, E., Woźniakowski, H.: *Tractability of Multivariate Problems, Volume 1: Linear Information*. EMS Tracts in Mathematics, no. 6, European Mathematical Society (2008)
27. Parzen, E.: Statistical inference on time series by RKHS methods. In: Pyke, R. (ed.) *Proc. 12th Biennial Seminar*, pp. 1–37. Canadian Mathematical Congress, Montreal, Canada (1970)
28. Pesenson, I.: Variational splines on Riemannian manifolds with applications to integral geometry. *Adv. Appl. Math.* **33**, 548–572 (2004)
29. Rasmussen, C.E., Williams, C.: *Gaussian Processes for Machine Learning*. MIT Press, 2006 (online version at <http://www.gaussianprocess.org/gpml/>)
30. Renka, R.J.: Interpolatory tension splines with automatic selection of tension factors. *SIAM J. Sci. Stat. Comput.* **8**, 393–415 (1987)
31. Schaback, R.: Error estimates and condition numbers for radial basis function interpolation. *Adv. in Comput. Math.* **3**, 251–264 (1995)
32. Schaback, R.: Multivariate interpolation by polynomials and radial basis functions. *Constr. Approx.* **21** 293–317 (2005)
33. Schaback, R.: Limit problems for interpolation by analytic radial basis functions. *J. Comp. Appl. Math.* **212** 127–149 (2008)
34. Schaback, R., Wendland, H.: Kernel techniques: From machine learning to meshless methods. *Acta Numerica* **15** 543–639 (2006)
35. Schmid, D.: A trade-off principle in connection with the approximation by positive definite kernels. In: Neamtu, M., Schumaker, L.L. (eds.) *Approximation Theory XII: San Antonio 2007*, pp. 348–359. Nashboro Press, Brentwood, TN (2008)
36. Schoenberg, I.J.: Contributions to the problem of approximation of equidistant data by analytic functions, Parts A & B. *Quart. Appl. Math.* **4**, 45–99 & 112–141 (1946)
37. Schumaker, L.L.: *Spline Functions: Basic Theory*. John Wiley & Sons, New York (1981, reprinted by Krieger Publishing 1993)
38. Song, G., Riddle, J., Fasshauer, G.E., Hickernell, F.J.: Multivariate interpolation with increasingly flat radial basis functions of finite smoothness. Submitted
39. Stein, M.L.: *Interpolation of Spatial Data. Some theory for Kriging*. Springer Series in Statistics, Springer-Verlag, New York (1999)
40. Steinwart, I., Christmann, A.: *Support Vector Machines*. Springer Verlag, Berlin (2008)
41. Wahba, G.: *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics 59, SIAM, Philadelphia (1990)
42. Wendland, H.: Sobolev-type error estimates for interpolation by radial basis functions. In: Le Méhauté, A., Rabut, C., Schumaker, L.L. (eds.) *Surface Fitting and Multiresolution Methods*, pp. 337–344. Vanderbilt University Press, Nashville (1997)
43. Wendland, H.: *Scattered Data Approximation*. Cambridge University Press, Cambridge (2005)
44. Ye, Q.: Reproducing kernels of generalized Sobolev spaces via a Green function approach with differential operators. Submitted to these proceedings
45. Zhu, H., Williams, C.K., Rohwer, R.J., Morciniec, M.: Gaussian regression and optimal finite dimensional linear models. In: Bishop, C.M. (ed.) *Neural Networks and Machine Learning*. Springer, Berlin (1998)