Scaling of Radial Basis Functions

E. Larsson[1, 3] and R. Schaback[2, 3]

October 12, 2022

Abstract

This paper studies the influence of scaling on the behavior of Radial Basis Function interpolation. It focuses on certain central aspects, but does not try to be exhaustive. The most important questions are: How does the error of a kernel-based interpolant vary with the scale of the kernel chosen? How does the standard error bound vary? And since fixed functions may be in spaces that allow scalings, like global Sobolev spaces, is there a scale of the space that matches the function best? The last question is answered in the affirmative for Sobolev spaces, but the required scale may be hard to estimate. Scalability of functions turns out to be restricted for spaces generated by analytic kernels, unless the functions are band-limited. In contrast to other papers, polynomials and polyharmonics are included as *flat limits* when checking scales experimentally, with an independent computation. The numerical results show that the hunt for near-flat scales is questionable, if users include the flat limit cases right from the start. When there are not enough data to evaluate errors directly, the scale of the standard error bound can be varied, up to replacing the norm of the unknown function by the norm of the interpolant. This follows the behavior of the actual error qualitatively well, but is only of limited value for estimating error-optimal scales. For kernels and functions with unlimited smoothness, the given interpolation data are proven to be insufficient for determining useful scales.

1 Introduction

Throughout the paper it is assumed that readers are familiar with the basics of kernel-based methods, e.g. from books by M.D. Buhmann [4], H.

[1]Scientific Computing, Department of Information Technology, Uppsala University, Box 337, SE-751 05 Uppsala, Sweden

[2]Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Lotzestraße 16–18, D–37083 Göttingen, Germany

[3]Authors in alphabetical order

Wendland [23], and G. Fasshauer/M. McCourt [8].

It is a well-known fact that interpolation and approximation with radial basis functions is crucially dependent on *scaling*. There are two ways to introduce scaling into a radial basis function. For instance, the classical multiquadrics are often written as kernels

$$\Phi(\|x-y\|_2) := (c^2+\|x-y\|_2^2)^\beta = c^{2\beta}(1+\|x-y\|_2^2/c^2)^\beta \text{ for all } x,y \in \mathbb{R}^d,$$

and compactly supported Wendland functions are scaled like in

$$\Phi(\|x-y\|_2) := (1-\|x-y\|_2/c)_+^m \, p_{d,k}(\|x-y\|_2/c) \text{ for all } x,y \in \mathbb{R}^d$$

to let the support radius be $c > 0$. The PDE–oriented literature often calls c a *shape parameter* (see e.g. [14]). In another context started by B. Fornberg and his collaborators [7] one works with $\varepsilon = 1/c$ instead and calls $\varepsilon \to 0$ the *flat limit*. For details on flat limits, see e.g. [7, 13, 15, 18, 19, 21] with a good summary in [8]. Here, we shall stick to the latter notation and and consider scaled radial basis functions

$$\Phi_\varepsilon(\|x-y\|_2) := \Phi(\varepsilon\|x-y\|_2)$$

for $x, y \in \mathbb{R}^d$ with $\Phi : [0,\infty) \to \mathbb{R}$. We ignore multiplication of the kernel by a scalar here, because the scalar cancels out for interpolation and approximation. However, it turns up as a *process variance* in methods that apply probabilistic estimation after rewriting kernel methods via Gaussian processes. See [8, 20] for the connection to the deterministic situation.

If data $f(x_1),\dots,f(x_n) \in \mathbb{R}$ of a function f are to be interpolated in scattered locations $x_1,\dots,x_n \in \mathbb{R}^d$, they determine coefficients α_j by the linear system

$$f(x_k) = \sum_{j=1}^n \alpha_j \Phi_\varepsilon(x_k,x_j),\ 1 \le k \le n \tag{1}$$

with a positive definite *kernel matrix* A^{X,Φ_ε} with entries $\Phi_\varepsilon(x_j,x_k)$, $1 \le j,k \le n$. Then the interpolant is written as

$$s_{f,X,\Phi_\varepsilon}(x) = \sum_{j=1}^n \alpha_j \Phi_\varepsilon(x,x_j),$$

or as a linear combination

$$s_{f,X,\Phi_\varepsilon}(x) = \sum_{j=1}^n f(x_j) u_{x_j}^{X,\Phi_\varepsilon}(x)$$

for *Lagrangians* or *cardinal interpolants* with $u_{x_j}^{X,\Phi_\varepsilon}(x_i) = \delta_{ji}$ given by

$$u_{x_j}^{X,\Phi_\varepsilon}(x) = \sum_{k=1}^n \beta_{jk}^{X,\Phi_\varepsilon} \Phi_\varepsilon(x,x_k)$$

via the coefficients $\beta_{jk}^{X,\Phi_\varepsilon}$, $1 \le j,k \le n$ from the inverse of the kernel matrix A^{X,Φ_ε} because of

$$u_{x_j}^{X,\Phi_\varepsilon}(x_i) = \delta_{ji} = \sum_k = 1^n \beta_{jk}^{X,\Phi_\varepsilon} \Phi_\varepsilon(x_i,x_k), 1 \le j,i \le n.$$

Theoretically, one can use arbitrary positive definite radial basis functions with arbitrary scales. The result can be seen as independent of f, if one only looks at the Lagrangians. This will be the topic of Section 2.3.

Fixing a norm for f and the interpolant, one can ask for an *error-optimal* scale that minimizes $\|f - s_{f,X,\Phi_\varepsilon}\|$ over all manageable scales. This is an interesting problem in theory and practice, and still open in both respects. There is a vast literature on numerical methods for estimating optimal scales in practice, but we shall not compare these here.

If users minimize errors or error bounds for a specific X, the distribution and density of points may have a strong influence on the result, leading to different suggested scales for different X. Users should be cautious when seeing estimates of optimal scales behaving like $1/h(X,\Omega)$ for the *fill distance*

$$h = h(X,\Omega) = \sup_{y\in\Omega} \min_{x\in X} \|x-y\|_2,$$

since it is well-known [2, 1] that there is no convergence in such a *stationary* situation except for conditionally positive kernels.

A possible workaround to minimizing the error directly is to ask for an *error-bound-optimal* scale that minimizes some specified norm of the error. The standard pointwise L_∞ error bound is

$$|f(x) - s_{f,X,\Phi_\varepsilon}(x)| \le P_{X,\Phi_\varepsilon}(x)\|f\|_{\Omega,\Phi_\varepsilon} \tag{2}$$

using the domain-independent *Power Function*. It holds for all functions f from the local *native space* $\mathscr{H}_{\Omega,\Phi}$ and all finite sets $X \subset \Omega$ of centers. By standard extension arguments [23, Section 10.7] one has

$$\|f\|_{\Omega,\Phi_\varepsilon} \le \|E_\Omega(f)\|_{\mathbb{R}^d,\Phi_\varepsilon}$$

for the canonical extension $E_\Omega(f)$ of f to the whole space. If norms are taken, one gets two variations. The local bound is

$$\|f - s_{f,X,\Phi_\varepsilon}\|_{\Omega,\infty} \le \|P_{X,\Phi_\varepsilon}\|_{\Omega,\infty}\|f\|_{\Omega,\Phi_\varepsilon} \le \|P_{X,\Phi_\varepsilon}\|_{\Omega,\infty}\|E_\Omega f\|_{\mathbb{R}^d,\Phi_\varepsilon} \tag{3}$$

holding on the local native space $\mathscr{H}_{\Omega,\Phi}$, while a global one is

$$\|f - s_{f,X,\Phi_\varepsilon}\|_{\Omega,\infty} \le \|P_{X,\Phi_\varepsilon}\|_{\Omega,\infty}\|f\|_{\mathbb{R}^d,\Phi_\varepsilon} \tag{4}$$

holding on the global native space $\mathscr{H}_{\mathbb{R}^d,\Phi}$. This is the version we shall analyze in Section 3. Note that (4) turns into (3) if f is replaced by $E_\Omega(f)$. Both factors in the right-hand side of (4) will depend on ε. The bound splits into an f-dependent part $\|f\|_{\mathbb{R}^d,\Phi_\varepsilon}$ and an X-dependent part $\|P_{X,\Phi_\varepsilon}\|_{\Omega,\infty}$ whatever the scaling is. This raises the question whether there is a scale ε that minimizes $\|f\|_{\mathbb{R}^d,\Phi_\varepsilon}$ without caring for the data.

Furthermore, convergence rates for interpolation often do not depend on scaling, but the factor in front of the rates does. This means that there must be an X-independent part in the whole problem.

All of this raises the question whether functions have a "natural scaling" that can be recovered approximately somehow, being independent of the interpolation problem. And since $\|f\|_{\Phi_\varepsilon}$ is a factor in the error bound (4), there is some hope that a minimal $\|f\|_{\Phi_\varepsilon}$ also cares for small factors in front of convergence rates. Surprisingly, functions in global Sobolev spaces have such a natural scale, as proven in Section 4.

But there are practical limits to scaling. It is well–known [17] that the condition of the kernel matrices A^{X,Φ_ε} increases dramatically for $\varepsilon \to 0$, while the interpolants still exist for all $\varepsilon > 0$. There are many workarounds for this, e.g. Contour-Padé [12], RBF-QR [10], and RBF-GA [11] by the group around Bengt Fornberg, and Hilbert-Schmidt-SVD by Fasshauer/McCourt [8, Chapter 13]. But we shall only focus on the flat limit here, not on methods to reach it via small scales. It requires to distinguish between two kinds of radial kernels: the *analytic* ones have convergent expansions into powers of $\|x-y\|_2^2$, like the Gaussian or inverse multiquadrics $\Phi(x-y) = (1+\|x-y\|_2^2)^{-m}$ for positive m. These have infinite smoothness, and their Fourier transform decays exponentially towards infinity. The other *non-analytic* kind has limited smoothness, and their Fourier transform decays algebraically, i.e. as a negative finite power towards infinity. These include Matérn-Sobolev and compactly supported Wendland [22] or Buhmann [3] kernels, among others. The flat limit for the analytic kernels is a polynomial [7] except for certain degenerations [15, 16, 19] depending on the point sets, while the flat limit for the others [21] is a polyharmonic interpolant based on the conditionally positive definite kernels

$$K(x,y) = \begin{cases} \|x-y\|^{2m-d} & d \text{ odd} \\ \|x-y\|^{2m} \log \|x-y\|_2 & d \text{ even} \end{cases}$$

with generalized Fourier transforms $\|\omega\|_2^{-2m}$ on $\mathbb{R}^d$. These kernels are scale-invariant, because the scale comes to the front as a scalar factor in the Fourier transform. They are hard to beat if it comes to handle stencils for approximation of derivatives [5], and they offer a convenient bypass around all scaling

problems. Situations with very small ε and kernels with finite smoothness will be explained perfectly by interpolation with polyharmonic kernels that should be used right from the start in such cases. The discussion of flat limits should therefore be confined to analytic kernels.

Summarizing, the above discussion shows that the strong dependence of radial basis function techniques on scaling is a feature, not a bug. The functions supplying the data already have a hidden natural scale, independent of how the reconstruction by interpolation or approximation is done, and good recovery methods should therefore not be scale–independent. This paper tries to clarify the scaling effects to some extent.

It starts by collecting some basic facts on scaling in Section 2 for the convenience of readers. These include some useful invariance relations and describe the dependence of Power Functions and Lagrangians on scaling. Optimal scales of functions result from studying norms $\|f\|_{\Phi_\varepsilon}$ for fixed f as functions of ε in section 2.1.

Finally, we look at the limit $\varepsilon \to 0$ for analytic kernels Φ in section 6, ignoring possible degenerations. It is known that the limit interpolant is a polynomial, but here we study the behavior of the norm of the interpolation error as a function of ε. Experimentally, it is a function of ε that can have sharp local minima at seemingly unexpected scales, but there are also many cases where the flat limit has an optimal error norm. Certain criteria for these two cases are provided, but they are hard to handle in practice. However, it is proven that any fixed set of interpolation data does not determine whether the flat limit is optimal or not. Users need additional data for error evaluation when they search for optimal scales.

2 Basic Facts on Scaling and Spaces

Throughout the paper, functions will be real-valued, defined on $\mathbb{R}^d$, and Fourier transformable. We do not treat localized versions of scaling here. Furthermore, we focus on functions with continuous point evaluation, and therefore we work on Hilbert spaces $\mathscr{H}_\Phi$ with reproducing kernels Φ. These kernels will be translation-invariant and Fourier transformable, i.e. there is a reproduction property

$$f(x) = (f, \Phi(x-\cdot))_{\mathscr{H}_\Phi} \text{ for all } f \in \mathscr{H}_\Phi,\ x \in \mathbb{R}^d$$

and an inner product

$$(f,g)_{\mathscr{H}_\Phi} = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\hat{f}(\omega)\overline{\hat{g}(\omega)}}{\hat{\Phi}(\omega)} d\omega \text{ for all } f \in \mathscr{H}_\Phi$$

with the usual property

$$\Phi(x-y) = (\Phi(x-\cdot), \Phi(\cdot - y))_{\mathscr{H}_\Phi} \text{ for all } x, y \in \mathbb{R}^d. \tag{5}$$

A particularly interesting case is Sobolev space $W_2^m(\mathbb{R}^d)$ for $m > d/2$ with the Whittle-Matérn kernel

$$\Phi(x) := \frac{2^{1-m}}{\Gamma(m)} \|x\|_2^{m-d/2} K_{m-d/2}(\|x\|_2) \tag{6}$$

using the modified Bessel function of second kind. It has the d-variate Fourier transform

$$\hat{\Phi}(\omega) = (1 + \|\omega\|_2^2)^{-m} \text{ for all } \omega \in \mathbb{R}^d \tag{7}$$

to make it compatible with the inner product above.

Definition 1. *A function* $f : \mathbb{R}^d \to \mathbb{R}$ *will be* scaled *by*

$$f_\varepsilon(x) := f(\varepsilon x) \text{ for all } x \in \mathbb{R}^d, \varepsilon > 0. \tag{8}$$

Note that scaling is "inverted" in the Fourier domain by

$$\widehat{f_\varepsilon}(\omega) = \varepsilon^{-d} \hat{f}(\omega/\varepsilon) = \varepsilon^{-d} \hat{f}_{1/\varepsilon}(\omega) \text{ for all } \omega \in \mathbb{R}^d, \varepsilon > 0. \tag{9}$$

Thus it is equivalent up to a factor to consider scaling of a function or its Fourier transform. If one would define scaling differently, namely by $f_\varepsilon(x) := \varepsilon^{d/2} f(\varepsilon x)$, there would be Fourier transform symmetry. We avoid this, because we want to keep the fact that $\varepsilon \to 0$ implies that f_ε does not go to zero, while the Fourier transform of f_ε goes to zero if it decays fast enough at infinity. This is the "flat limit" situation, and this paper will keep an eye on flat limits throughout.

The scaling law (8) will scale compact supports properly, keeping $f_\varepsilon(0) = f(0)$ invariant, but this will not be true in frequency space. There, the integral over frequency space is invariant.

When we scale the kernel K of the Hilbert space $\mathscr{H}_\Phi$, we shall denote the scaled inner product by $(\cdot,\cdot)_{\Phi_\varepsilon}$ belonging to the scaled kernel K_ε of the Hilbert space $\mathscr{H}_{\Phi_\varepsilon}$ Then the dual version

$$(\delta_x, \delta_y)_\Phi = (\Phi(x-\cdot), \Phi(\cdot - y))_{\mathscr{H}_\Phi} = \Phi(x-y) \text{ for all } x, y \in \mathbb{R}^d$$

of (5) yields

$$(\delta_x, \delta_y)_{\Phi_\varepsilon} = \Phi_\varepsilon(x-y) = \Phi(\varepsilon x - \varepsilon y) = (\delta_{\varepsilon x}, \delta_{\varepsilon y})_\Phi \tag{10}$$

for all $x, y \in \mathbb{R}^d$, $\varepsilon > 0$. This is a *scaling law* for point evaluation functionals.

When treating fixed functions $f \in \mathscr{H}_\Phi$ with scaled kernels K_{Φ_ε}, it is not clear whether all scaled native spaces $\mathscr{H}_{\Phi_\varepsilon}$ contain f, and whether the native spaces are nested or norm-equivalent. We postpone this to the study of norms $\|f\|_{\Phi_\varepsilon}$ for $f \in \mathscr{H}_\Phi$, and it will turn out that analytic kernels will cause problems.

2.1 Scaling of Kernel-Based Norms

We consider functions $f \in \mathscr{H}_\Phi$ and start with a simple scaling law

$$\|f_\varepsilon\|_{\Phi_\varepsilon} = \|f\|_\Phi \tag{11}$$

that follows from

$$\begin{aligned}
\|f_\varepsilon\|^2_{\Phi_\varepsilon} &= \int_{\mathbb{R}^d} \frac{|\hat{f}_\varepsilon(\omega)|^2}{\hat{\Phi}_\varepsilon(\omega)} d\omega \\
&= \varepsilon^{-2d} \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega/\varepsilon)|^2}{\hat{\Phi}_\varepsilon(\omega)} d\omega \\
&= \varepsilon^{-d} \int_{\mathbb{R}^d} \frac{|\hat{f}(\eta)|^2}{\hat{\Phi}_\varepsilon(\eta\varepsilon)} d\eta \\
&= \int_{\mathbb{R}^d} \frac{|\hat{f}(\eta)|^2}{\hat{\Phi}(\eta)} d\eta \\
&= \|f\|^2_\Phi.
\end{aligned}$$

In the form

$$\|f_\varepsilon\|_\Phi = \|f\|_{\Phi_{1/\varepsilon}}$$

this proves that scaling a function or a kernel is the same thing for calculating native space norms, as long as one of the sides exists.

Now we fix functions $f \in \mathscr{H}_\Phi$ and check for which kernel scales we have $f \in \mathscr{H}_{\Phi_\varepsilon}$. A scale ε is called Φ-*admissible for* $f \in \mathscr{H}_\Phi$, if $\|f\|_{\Phi_\varepsilon}$ is finite.

A simple restriction on admissible scales is

$$\begin{aligned}
\|f\|^2_{\Phi_\varepsilon} &= \varepsilon^d \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{\Phi}(\omega/\varepsilon)} d\omega \\
&\geq \varepsilon^d \|f\|_{L_2} \|\hat{\Phi}\|^{-1}_\infty
\end{aligned} \tag{12}$$

with three implications: for large ε the norm $\|f\|^2_{\Phi_\varepsilon}$ must be large, for small ε it can not be smaller than $\mathscr{O}(\varepsilon^d)$, and admissible scales ε with $\|f\|^2_{\Phi_\varepsilon} \leq \|f\|^2_\Phi$ are bounded above. Altogether, the case of large ε is not interesting.

Theorem 1. *For kernels with finite smoothness, arbitrary scales are admissible. In particular, if the Fourier transform of* Φ *behaves like* $\hat{\Phi}(\omega) = \Theta(\|\omega\|^{-\beta})$ *with* $\beta > d$ *near infinity, then*

$$\|f\|^2_{\Phi_\varepsilon} = \Theta(\varepsilon^d \max(1, \varepsilon^{-\beta}))\|f\|^2_\Phi$$

for all $\varepsilon > 0$ *and all* $f \in \mathscr{H}_\Phi$. *The spaces* $\mathscr{H}_{\Phi_\varepsilon}$ *are identical as sets and norm-equivalent, the equivalence constants behaving like* $\varepsilon^d \max(1, \varepsilon^{-\beta})$

For Sobolev space $W_2^m(\mathbb{R}^d)$ with integer m, this holds for $\beta = 2m$, but we can get the explicit formula

$$\|f\|_{\Phi_\varepsilon}^2 = \varepsilon^d \sum_{j=0}^{m} \binom{m}{j} \varepsilon^{-2j} |f|_{W_2^j(\mathbb{R}^d)}^2 \text{ for all } f \in W_2^m(\mathbb{R}^d).$$

Since we are in global Sobolev space, the seminorms are norms, and thus none of the $|f|_{W_2^j(\mathbb{R}^d)}$ can vanish.

Proof. We consider

$$\begin{aligned} \|f\|_{\Phi_\varepsilon}^2 &= \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{\Phi}_\varepsilon(\omega)} d\omega \\ &= \varepsilon^d \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{\Phi}(\omega/\varepsilon)} \\ &= \varepsilon^d \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{\hat{\Phi}(\omega)} \frac{\hat{\Phi}(\omega)}{\hat{\Phi}(\omega/\varepsilon)} d\omega \end{aligned} \tag{13}$$

and get the two-sided bound

$$\inf_{\omega \in \mathbb{R}^d} \frac{\hat{\Phi}(\omega)}{\hat{\Phi}(\omega/\varepsilon)} \le \frac{\|f\|_{\Phi_\varepsilon}^2}{\varepsilon^d \|f\|_\Phi^2} \le \sup_{\omega \in \mathbb{R}^d} \frac{\hat{\Phi}(\omega)}{\hat{\Phi}(\omega/\varepsilon)}.$$

To show that both bounds behave like $\max(1, \varepsilon^{-\beta})$ for the finite smoothness case, we assume

$$\begin{array}{rcccccl} 0 & < & C_0 & \le & \hat{\Phi}(\omega) & \le & C_1 \quad \text{for } \|\omega\|_2 \le c \\ 0 & < & c_0 \|\omega\|_2^{-\beta} & \le & \hat{\Phi}(\omega) & \le & c_1 \|\omega\|_2^{-\beta} \quad \text{for } \|\omega\|_2 \ge c \end{array}$$

and bound the quotient by

$$\begin{array}{rcccl} \frac{C_0}{C_1} & \le & \frac{\hat{\Phi}(\omega)}{\hat{\Phi}(\omega/\varepsilon)} & \le & \frac{C_1}{C_0} \quad \text{for } \|\omega\|_2 \le c, \|\omega\|_2/\varepsilon \le c, \\ \frac{C_0 c^\beta}{c_1} & \le & \frac{\hat{\Phi}(\omega)}{\hat{\Phi}(\omega/\varepsilon)} & \le & \varepsilon^{-\beta} \frac{C_1 c^\beta}{c_0} \quad \text{for } \|\omega\|_2 \le c, \|\omega\|_2/\varepsilon \ge c, \\ \varepsilon^{-\beta} \frac{c_0 c^\beta}{C_1} & \le & \frac{\hat{\Phi}(\omega)}{\hat{\Phi}(\omega/\varepsilon)} & \le & \frac{c_1 c^{-\beta}}{C_0} \quad \text{for } \|\omega\|_2 \ge c, \|\omega\|_2/\varepsilon \le c, \\ \varepsilon^{-\beta} \frac{c_0}{c_1} & \le & \frac{\hat{\Phi}(\omega)}{\hat{\Phi}(\omega/\varepsilon)} & \le & \varepsilon^{-\beta} \frac{c_1}{c_0} \quad \text{for } \|\omega\|_2 \ge c, \|\omega\|_2/\varepsilon \ge c. \end{array}$$

If m is an integer, we can continue from (13) to

$$\begin{aligned} \|f\|_{\Phi_\varepsilon}^2 &= (2\pi)^{-d/2} \varepsilon^d \sum_{j=0}^{m} \binom{m}{j} \varepsilon^{-2j} \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 \|\omega\|_2^{2j} d\omega \\ &= \varepsilon^d \sum_{j=0}^{m} \binom{m}{j} \varepsilon^{-2j} |f|_{W_2^j(\mathbb{R}^d)}^2. \end{aligned} \tag{14}$$

□

Theorem 2. *For Gaussians and other kernels with exponential decay of the Fourier transform at infinity, the inclusion $\mathscr{H}_\Phi \subseteq \mathscr{H}_{\Phi_\varepsilon}$ is true only for $\varepsilon \geq 1$. For single functions, kernel scales $\varepsilon < 1$ may be admissible, but this depends on the function. For bandlimited functions and all kernels, all kernel scales are admissible.*

Proof. If we assume an exponential law

$$\hat{\Phi}(\omega) = c\exp(-\gamma\|\omega\|_2)$$

near infinity, we get

$$\frac{\hat{\Phi}(\omega)}{\hat{\Phi}(\omega/\varepsilon)} \quad = \quad \exp(-\gamma\|\omega\|_2)\exp(\gamma\|\omega\|_2/\varepsilon).$$

Now $\|f\|^2_{\Phi_\varepsilon}$ gets unbounded for $\varepsilon < 1$ provided that the integrable function $|\hat{f}(\omega)|^2/\Phi(\omega)$ is bounded below near infinity by an arbitrarily large negative power of $\|\omega\|_2$. The same argument works for the Gaussian, and shows that admissible scales are strongly f-dependent.

For bandlimited functions, all scales are admissible by the above argumentation. If the spectrum of f is limited by $\|\omega\|_2 \leq B$, and if we define

$$\delta^-_\Phi(K) := \inf_{\|\omega\|_2 \leq K} \hat{\Phi}(\omega),\ \delta^+_\Phi(K) := \sup_{\|\omega\|_2 \leq K} \hat{\Phi}(\omega),$$

equation (13) yields

$$\frac{\delta^-_\Phi(B)}{\delta^+_\Phi(B/\varepsilon)} \quad \leq \quad \frac{\|f\|^2_{\Phi_\varepsilon}}{\varepsilon^d\|f\|^2_\Phi} \quad \leq \quad \frac{\delta^+_\Phi(B)}{\delta^-_\Phi(B/\varepsilon)} \tag{15}$$

to show admissibility of all scales for all kernels. □

Because the native space norm does not exist in the limit, Theorem 2 forbids to go to flat limits for analytic kernels if $\|f\|^2_{\Phi_\varepsilon}$ occurs, e.g. for standard error bounds in terms of the Power Function. The flat limit of Lagrangians in the left-hand side of (17) exists, but the right-hand side cannot be formed for $\varepsilon < 1$.

Here, we ignored domain-dependent native spaces and their norms. These are always equivalent to the global norms on $\mathbb{R}^d$, but the norm equivalence constants will vary when kernels are scaled. Since Section 3 will be confined to global norms as well, we ignore the localized norms.

2.2 Scaling the Power Function

For interpolation in $X = \{x_1, \ldots, x_n\}$, the power function is

$$P^2_{X,\Phi}(x) := \inf_{a \in \mathbb{R}^n} \left\| \delta_x - \sum_{j=1}^{n} a_j \delta_{x_j} \right\|^2_{\Phi}.$$

and if we apply the scaling law (10), then

$$P_{X,\Phi_\varepsilon}(x) = P_{\varepsilon X,\Phi}(\varepsilon x) \text{ for all } x \in \mathbb{R}^d.$$

For Power Functions, scaling points and scaling kernels is the same, and the invariance relation is

$$P_{X,\Phi}(x) = P_{\varepsilon X,\Phi_{1/\varepsilon}}(\varepsilon x).$$

The Power Function $P_{X,\Phi_\varepsilon}(x) = P_{\varepsilon X,\Phi}(\varepsilon x)$ can be bounded via the standard arguments in [23]. If the usual geometric properties on the domain Ω and the point set X are satisfied, there is a Φ- and geometry-dependent function F_Φ such that

$$P_{X,\Phi}(x) \le F_\Phi(h) \text{ for all } x \in \Omega$$

for the *fill distance*

$$h := \sup_{y \in \Omega} \min_{x \in X} \|y - x\|_2.$$

Scaling X and x to εX and εx leaves the geometry invariant, and therefore

$$P_{\varepsilon X,\Phi}(\varepsilon x) \le F_\Phi(\varepsilon h)$$

for $\varepsilon \le 1$. The function $F_\Phi(h)$ can have algebraic or exponential decay towards zero for $h \to 0$. A similar argument is in [5] for convergence of stencils, though restricted to Sobolev spaces. In the case $W_2^m(\Omega)$ for $\Omega \subset \mathbb{R}^d$, the above function is $F_\Phi(h) = ch^{m-d/2}$, and for kernels satisfying Theorem1 it is $F_\Phi(h) = ch^{\beta/2-d/2}$.

Note that [17] provides a two-sided bound with the same F_Φ in case of asymptotically uniformly distributed points and kernels with finite smoothness.

2.3 Scaling of Lagrangians

For Lagrange basis functions $u_j^{X,\Phi}$ we get

$$\begin{aligned} \delta_{j,k} = u_j^{X,\Phi_\varepsilon}(x_k) &= \sum_{x_i \in X} a_{ji}(X,\Phi_\varepsilon)\Phi(\varepsilon x_k - \varepsilon x_i), \\ u_j^{X,\Phi_\varepsilon}(x) &= \sum_{x_i \in X} a_{ji}(\varepsilon X,\Phi)\Phi(\varepsilon x - \varepsilon x_i), \\ &= u_j^{\varepsilon X,\Phi}(\varepsilon x). \end{aligned}$$

This proves that scaling points or kernels is the same thing for calculating Lagrangians, and there is an invariance relation

$$u_j^{X,\Phi}(x) = u_j^{\varepsilon X,\Phi_{1/\varepsilon}}(\varepsilon x) \text{ or } u_j^{X,\Phi_\varepsilon}(x) = u_j^{\varepsilon X,\Phi}(\varepsilon x).$$

The paper [6] proves that Lagrangians for kernels with limited smoothness are uniformly bounded if points are asymptotically uniformly distributed. Checking the proof under consideration of scaling reveals that it works independent of scaling, because the scale factor cancels out. The geometric assumptions are satisfied, because the form $u_j^{\varepsilon X,\Phi}(\varepsilon x)$ just scales the geometry by ε. But the proof uses a bump function argument, and therefore fails for analytic kernels like Gaussians or inverse multiquadrics. This is in line with the fact that Lagrangians diverge in the flat limit for analytic kernels if they do not converge to polynomials [13] in degenerate cases, and it is also in line with the convergence towards polyharmonic interpolants [21] for non-analytic kernels.

2.4 L_∞ Error Under Scaling

If a user just has a single interpolation problem for a fixed point set X and wants to use the best possible kernel and scaling, the error

$$f(x) - s_{f,X,\Phi_\varepsilon}(x) = f(x) - \sum_{x_j \in X} f(x_j) u_{x_j}^{X,\Phi_\varepsilon}$$

just depends on the behavior of the Lagrangians $u_{x_j}^{X,\Phi_\varepsilon}$. For large ε, these converge to delta functions $\delta_{x_j}(x)$ for all analytic kernels, and for $\varepsilon \to 0$ they will converge to a limit Lagrangian, except for degenerate situations in case of analytic kernels, where they converge to $\pm\infty$. The limit Lagrangians are polynomials for analytic kernels and polyharmonics for kernels with finite smoothness.

When evaluating the error as a function of ε, it may be that the minimum is at $\varepsilon = 0$, i.e. the flat limit. In particular, this is to be expected when f is close to a polynomial and analytic kernels are used, or if f is close to a polyharmonic otherwise. But plenty of easy experiments show that the error above may have a minimal L_∞ norm for some positive ε, in a specific situation determined by f, X, and Φ. Observing such cases does not exhibit a clear pattern because of the strong dependence on f, and it is hard to attribute to one of the three ingredients.

One can get rid of the influence of f by focusing on the Lagrangians, but this will not be what single-case users want. They might not even be interested at all in function spaces or smoothness properties of f, using only the values $f(X)$. To go on from here, additional properties of Lagrangians are needed, e.g. moment conditions, but this leads to open problems.

3 Scaling of Error Bounds

The standard bound

$$|f(x) - s_{f,X,\Phi}(x)| \le P_{X,\Phi}(x)\|f\|_\Phi \tag{16}$$

can be scaled in different ways. Scaling only Φ results in

$$\left| f(x) - \sum_{x_j \in X} f(x_j) u_j^{X,\Phi_\varepsilon}(x) \right| \le P_{X,\Phi_\varepsilon}(x)\|f\|_{\Phi_\varepsilon}, \tag{17}$$

playing the left-hand side back to the Lagrangians and Section 2.3. To see whether the bound gives more information, both factors in the right-hand side need further consideration. Scaling only f and scaling only the points are not very useful and therefore omitted here.

The inequality is tight for some ε-dependent function, but it is not clear how the ingredients vary with ε for a fixed f and whether the two sides stay close. Recall that Theorem 2 excludes scales $\varepsilon < 1$ for arbitrary f in the case of analytic kernels.

The factors in the right-hand side were dealt with in sections 2.1 and 2.2, and we only have to look at the product. The error bound could be improved if the native space norm were taken in the domain-dependent native space. But this localized norm is bounded above by the global one [23], and this is the way the error bound is used in practice. We keep the global norm here.

Theorem 3. *For kernels satisfying Theorem 1 and for asymptotically uniformly placed points, the right-hand side of the standard error bound (16) behaves like*

$$\begin{array}{ll} \mathcal{O}(1) & \text{for } \varepsilon \to 0, \\ \mathcal{O}(\varepsilon^\beta) & \text{for } \varepsilon \to \infty. \end{array}$$

In Sobolev spaces $W_2^m(\mathbb{R}^d)$ with integer m, the square of the bound behaves like

$$\varepsilon^{2m} \sum_{j=0}^{m} \binom{m}{j} \varepsilon^{-2j} |f|^2_{W_2^j(\mathbb{R}^d)} \text{ for } \varepsilon \to 0.$$

□

Proof. From the previous section, we get that the error bound is $\mathcal{O}(1)$ for $\varepsilon \to 0$, caring for the flat limit. For large ε, the behavior of both factors leads to the $\mathcal{O}(\varepsilon^\beta)$ asymptotics of the bound. □

So far, this result is disappointing, because it is not clear if there is a minimum with respect to ε. Some experimental observations are in Section 5.4 showing typical cases in Figure 4.

4 Natural Scales of Functions

To proceed towards finding a "natural" scale of a global function $f \in \mathscr{H}_\Phi$ with respect to the kernel chosen, we consider the minimum of $\|f\|_{\Phi_\varepsilon}$ over all scales ε that are Φ-admissible for f. By (12), the search for optimal scales can be limited to

$$\varepsilon^d \leq \frac{\|f\|_\Phi^2 \|\hat{\Phi}\|_\infty}{\|f\|_2^2}. \tag{18}$$

For kernels with finite smoothness, Theorem 1 guarantees such a minimum, because $\|f\|_{\Phi_\varepsilon}$ goes to infinity like ε^d for $\varepsilon \to \infty$, and like $\varepsilon^{d-\beta}$ for $\varepsilon \to 0$ with $\beta > d$.

For general functions and analytic kernels, Theorem 2 limits scales to $\varepsilon \geq 1$, while (18) gives an upper limit in all cases. Therefore there must be a minimum of $\|f\|_{\Phi_\varepsilon}$ in this restricted domain. But things are strongly f-dependent, and if f is bandlimited, we can look at the limit $\varepsilon \to 0$ even for analytic kernels. Then (15) shows that the left bound gets constant for $\varepsilon \to 0$, but the right one will blow up. There may be a minimum at $\varepsilon = 0$, i.e. for the flat limit case.

By the scaling law $\|f_\varepsilon\|_\Phi^2 = \|f\|_{\Phi_{1/\varepsilon}}^2$, the minimization of $\|f_\varepsilon\|_\Phi^2$ is equivalent to minimization of $\|f\|_{\Phi_{1/\varepsilon}}^2$, and is therefore also covered by the above argumentation, except that ε must be replaced by $1/\varepsilon$.

Summarizing, all functions in native spaces of a kernel with finite smoothness have a natural scale as an ε that minimizes $\|f\|_{\Phi_\varepsilon}^2$. For analytic kernels, existence of a minimum of $\|f\|_{\Phi_\varepsilon}$ will be crucially dependent on f, but it may arise in the flat limit $\varepsilon = 0$. Even if it could be estimated efficiently and accurately, it would not necessarily lead to a minimal error in (17) for specific cases. But it may have a small factor in front of the convergence rate associated with the unscaled kernel. Some experimental results are in Section 5.3, but they replace $\|f\|_{\Phi_\varepsilon}$ by $\|s_{f,X,\Phi_\varepsilon}\|_{\Phi_\varepsilon}$ and the resulting scales are not particularly useful.

5 Examples

A comparison of existing scale estimation algorithms is not attempted here. The examples will be strictly confined to illustrations of the theoretical arguments of this paper. Sophisticated algorithms are avoided in favor of what scientists do when they want quick results. Throughout, we work in the unit square in $\mathbb{R}^2$, and we limit all hazardous scales in a primitive way by stopping when MATLAB's `condest` exceeds 10^{14}. In contrast to many other papers, we add the flat limit situation, not as a limit, but by adding a polynomial and a polyharmonic solution.

5.1 Remarks on Kernels and Functions

To keep the number of final figures small, only four kernels and functions were selected as typical cases.

From the class of analytic kernels, we only consider the Gaussian and the inverse multiquadric $\Phi(x,y) = (1+\|x-y\|_2)^{-1/2}$. For analytic functions, we only consider the scaled class $f_\varepsilon(x) = (1+\varepsilon^2\|x\|_2^2)^{-1}$ that models the Runge phenomenon. Their complex extensions have singularities at distance $1/\varepsilon$ from the real axis, and we use $\varepsilon = 1/4$ for a good and $\varepsilon = 25$ for a bad case. Both are polynomials up to machine accuracy, but the bad case needs a very high degree for this. One can expect that the polynomial flat limit works fine for the good case and fails for the bad one, even though the function is analytic on the real numbers.

The non-analytic case has to compromise on the kernel side. The intention is to use kernels with comparable $W_2^3(\mathbb{R}^2)$ smoothness. The matching 2D polyharmonics are $\|x-y\|_2^4 \log\|x-y\|_2$, and the matching radial Matérn-Sobolev kernel is $K_2(r)r^2$. But standard Wendland kernels $\phi_{3,k}$ on $\mathbb{R}^2$ work in $W_2^{k+3/2}(\mathbb{R}^2)$ and cannot match $W_2^3(\mathbb{R}^2)$. Picking $k=2$, i.e. the radial kernel $(1-r)_+^6(35r^2+18r+3)$ leads to $W_2^{3.5}(\mathbb{R}^2)$ smoothness, and therefore we also use the polyharmonic kernel $\|x-y\|_2^6 \log\|x-y\|_2$ working in $W_2^4(\mathbb{R}^2)$.

In legends of plots, kernels are abbreviated by `g`, `mq`, `ms3`, `w3.5`, and for the flat limit we use `p`, `ph3`, and `ph4` for polynomials and polyharmonics. The trailing number indicates Sobolev smoothness, not the usual RBF parameters. To match the basic kernel scale to the domain $[-1,+1]^2$ and regular point sets of 121 or 441 points, the Gaussian, the inverse multiquadric and the Wendland function were pre-scaled by $\varepsilon = 10, 10$, and 0.2, respectively. Note that for small ε the Wendland functions have large supports, and the matrices are not sparse. Error evaluation is made in cell midpoints only.

Polynomials were fitted to the data by brute force, multiplying data values by the pseudoinverse of a high-order Vandermonde matrix. This is a hazardous thing by itself, but we wanted to use primitive techniques only. Since the result deteriorates numerically when large polynomial degrees are taken, we picked a degree that yields a minimal L_∞ error. Of course, there are much better techniques for polynomial interpolation.

For non-analytic functions we picked $\|x\|_2^3$ for a case with a single interior point of non-smoothness, and $\|x\|_\infty$ for a piecewise linear continuous case with derivative discontinuities along lines. The chosen kernels and functions are the result of many test runs suppressed here.

5.2 L_∞ Errors as Functions of Scales

Figure 1 shows observed L_∞ errors as functions of scales, for the four functions and kernels. The point set X consisted of 121 regular points on $[-1,+1]^2$, while Figure 2 is for 441 points, just to demonstrate the dependence on X. The errors for polynomials and polyharmonics are given as horizontal lines with markers.

Here is a list of observations.

1. The top two functions are analytic, but only the left one is close to a low-degree polynomial. There, the flat limit for analytic kernels is optimal, but the non-analytic kernels do not perform badly, as well as the polyharmonics.
2. The top right case shows the Runge effect. Well-chosen scales of analytic kernels may give quite some improvement over the others, but the scale estimation will be hazardous. Non-analytic kernels and polyharmonics work fine when data sets get dense.
3. The bottom left case is a mildly nonsmooth non-polynomial function with the non-smoothness located just in the center of the domain. Polyharmonics are the best choice.
4. The bottom right case is continuous, but nondifferentiable along lines. Except for lucky scales, analytic kernels should be avoided. The others work well.
5. The hunt for error-minimizing scales may pay off for analytic kernels, but is a risky thing that often is not much better than what kernels with low smoothness or flat-limit polyharmonics can do.
6. If one calculates a polynomial and a polyharmonic solution to pick the better one, there is not much to gain by bridging the gap between the flat limit and the smallest ε that works within a crude `condest` limit. Continuing the curves to the left will not give something much better.
7. Motivated by the result that small errors imply large conditions of kernel matrices [17], users were tempted to take the smallest ε they could handle. The right-hand plots show that this is misleading for analytic kernels. For non-analytic kernels it works fine, but the polyharmonic flat limit will be a better and faster solution, realizing the smallest ε ever.
8. The vast literature on algorithms for scale estimation has no clear winner yet. A possible reason is that the above plots show widely varying situations to care for. It is suggested to include $\varepsilon = 0$ into testing of scales, because the flat limit solution, polynomial or polyharmonic, may be hard to beat by any other scaling.

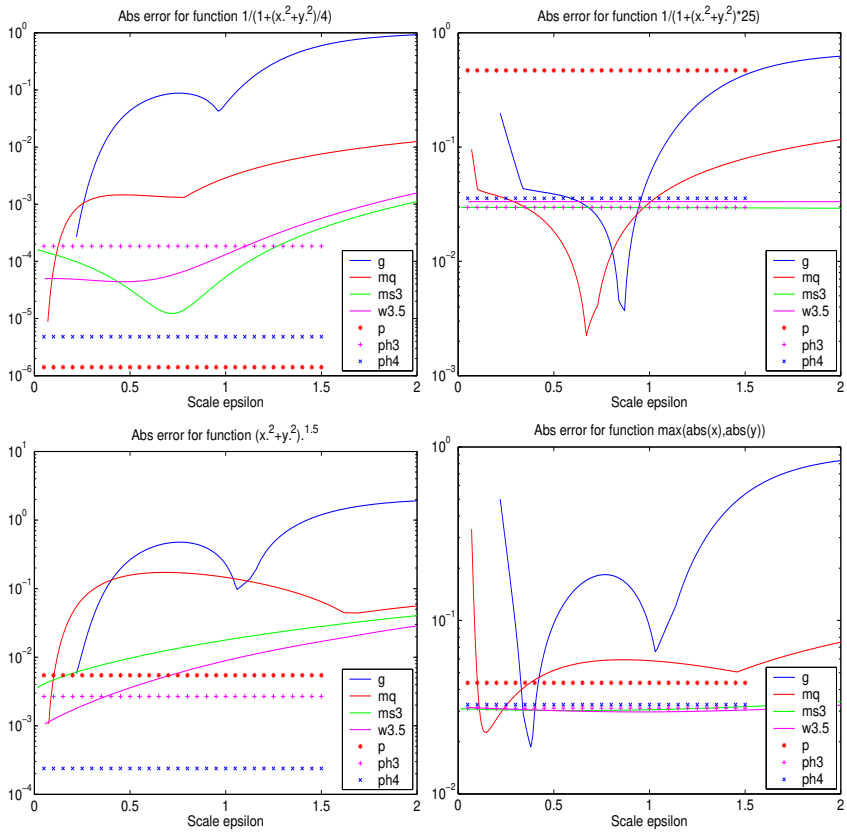

Figure 1: ℓ_∞ errors on midpoints as functions of scale for 121 regular data locations on $[-1,+1]^2$.

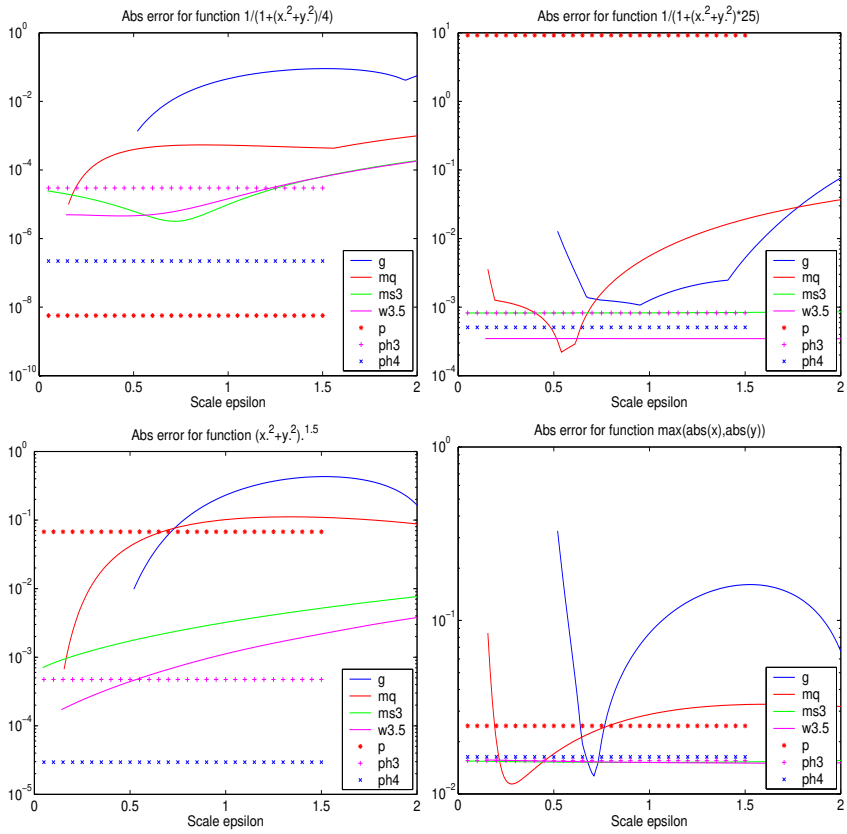

Figure 2: ℓ_∞ errors on midpoints as functions of scale for 441 regular data locations on $[-1,+1]^2$.

5.3 Native Space Norms as Functions of Scales

Section 4 calls for plots of $\|f\|_{\Phi_\varepsilon}$ as functions of ε with clear minima. This would be another estimation method for scales, if it were cheaply and accurately available. But only native space norms $\|s_{f,X,\Phi_\varepsilon}\|_{\Phi_\varepsilon} \le \|f\|_{\Phi_\varepsilon}$ of interpolants on point sets X are available, and it is not clear how close they are to $\|f\|_{\Phi_\varepsilon}$ and whether they have the same behavior for varying ε as predicted for $\|f\|_{\Phi_\varepsilon}$ by Section 4, in particular Theorem 1. There, as functions of $c = 1/\varepsilon$, the quantity $c^d\|f\|_{\Phi_{1/c}}$ should be a polynomial in c of degree m, if f is in $W_2^m(\mathbb{R}^d)$ and Φ is the kernel for that space. At least, one should look at $c^d\|s_{f,X,\Phi_{1/c}}\|_{\Phi_{1/c}}$ as a function of large c for various kernels. This is still open, and maybe it yields an estimate for a good smoothness parameter.

Figure 3 shows $\|s_{f,X,\Phi_\varepsilon}\|_{\Phi_\varepsilon}$ as a function of ε for the four selected cases in the 441 point situation. The scale-independent native space norms of the polyharmonics are given for comparison as horizontal lines with markers. There is no transition to the flat limit, because $\|s_{f,X,\Phi_\varepsilon}\|_{\Phi_\varepsilon}$ seems to behave similarly to $\|f\|_{\Phi_\varepsilon}$, i.e. going to infinity in the flat limit, exponentially for analytic kernels and algebraically for the others. Minima of the native space norm, if there are any, do not appear to be useful as criteria for good scales.

5.4 Error Bounds as Functions of Scales

Figure 4 shows the accessible part of the right-hand side of the standard error bound 4. The replacement of $\|f\|_{\Phi_\varepsilon}$ by $\|s_{f,X,\Phi_\varepsilon}\|_{\Phi_\varepsilon}$ deletes the reliability as an upper bound, but it is interesting to see whether it indicates good scales when compared to Figure 2 for the error. Note that the curves in Figure 4 can be calculated without having any additional points for error evaluation. The qualitative behavior is unexpectedly similar, and the observations made after seeing the error plots could have been made without evaluating errors.

Figures 6 and 5 combine Figures 2 and 4 for easier comparison of actual errors (top row) to native space norms (center row) and error bounds (bottom row). Figure 5 has two moderately well-behaving cases, while Figure 6 shows results for the nonsmooth function $\max(|x|,|y|)$ to the right and $1/(1+25(x^2+y^2))$ with an expectable Runge phenomenon on the left.

In the flat limit, the factors in the error bound 4 tend towards $0 \cdot \infty$ with unknown behavior. The flat limit marks in Figure 4 are the error bounds for the polyharmonic case, and not attained as the limits of the curves, unlike Figure 2. The conclusion is that the error bounds are useless when traced to the flat limit.

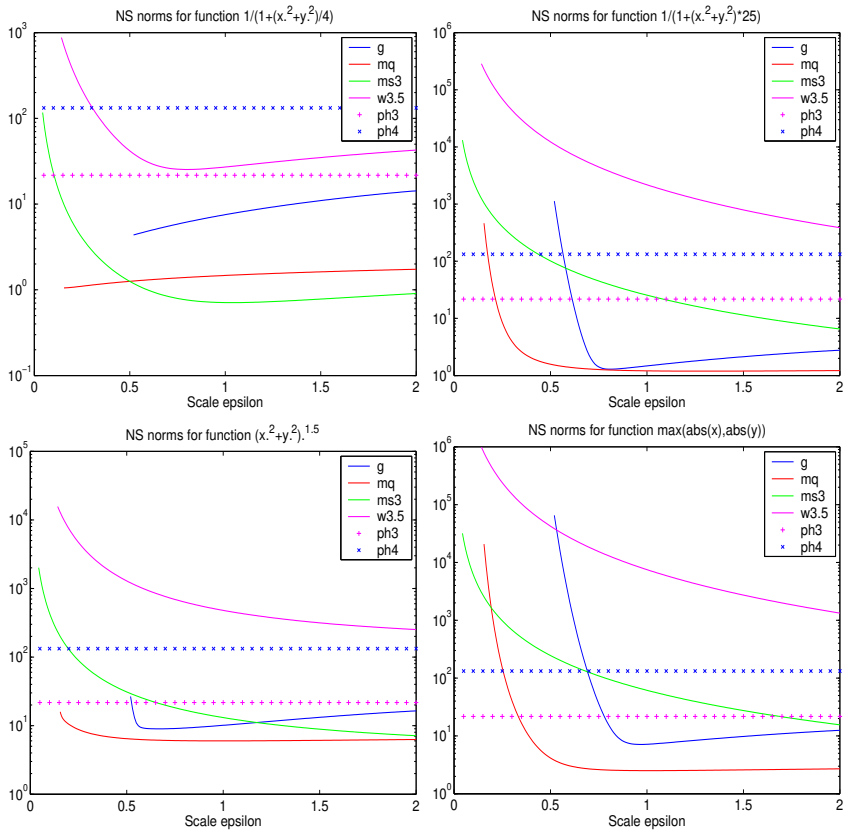

Figure 3: Native space norms for varying scale.

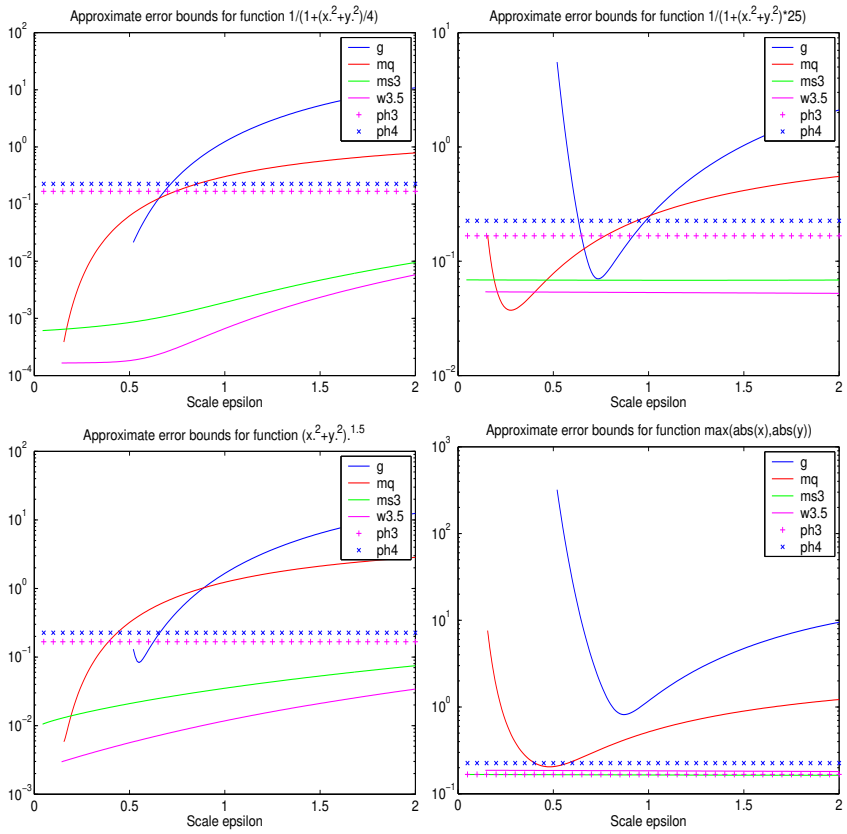

Figure 4: Approximate error bounds for varying scale.

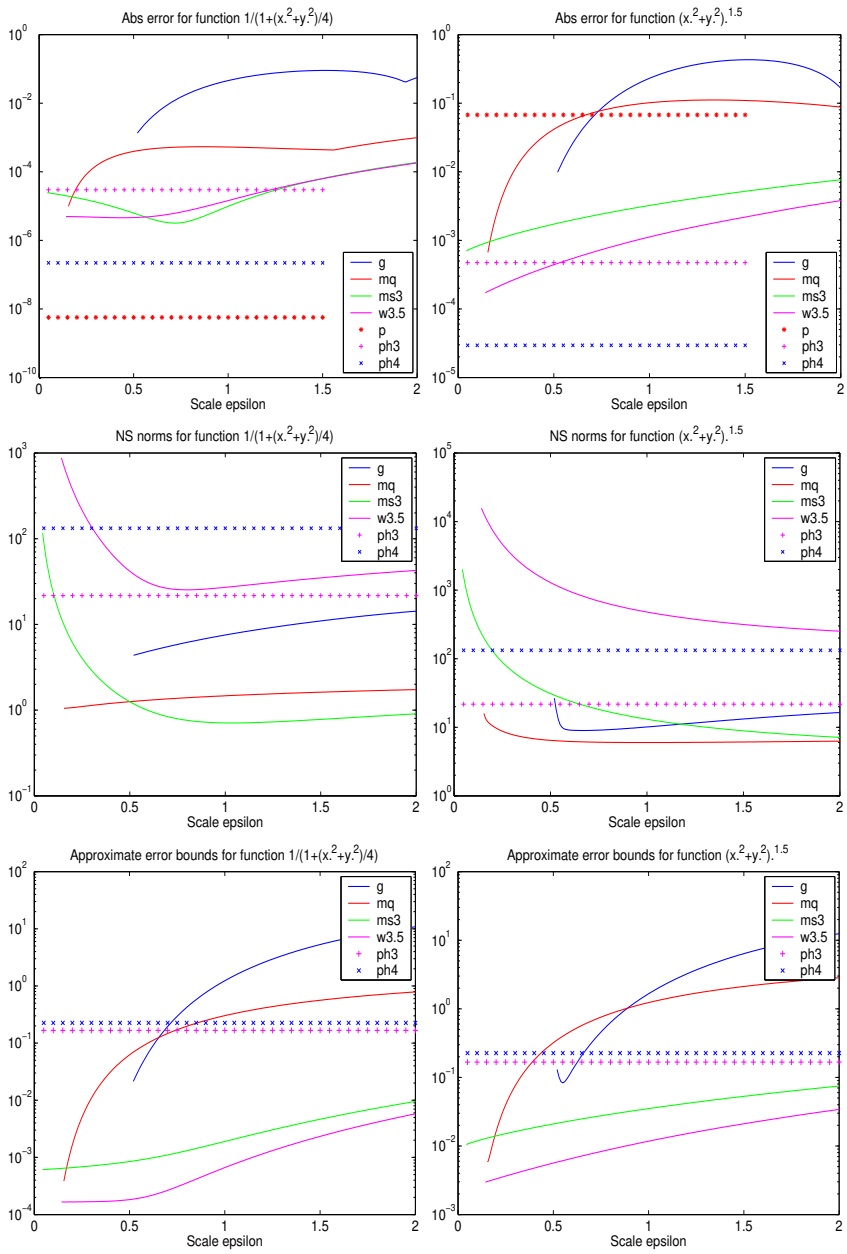

Figure 5: Errors, native space norms, and approximate error bounds for two selected functions.

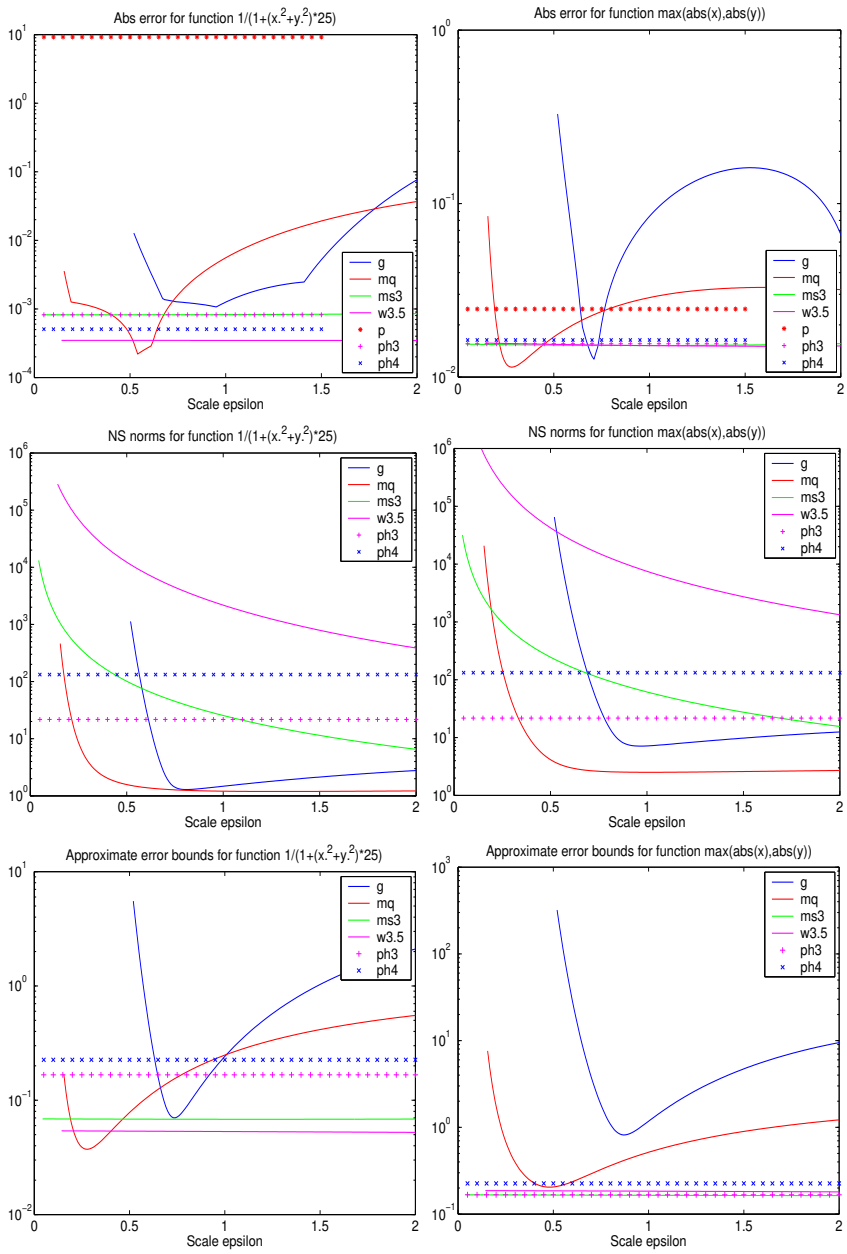

Figure 6: Errors, native space norms, and approximate error bounds for two selected functions

6 Error Expansions

We now look at cases with infinite smoothness. There are cases where a positive scale ε gives smaller errors than the flat limit, and we want to see under which conditions this works. To this end, we have to expand the interpolation error into a power series in ε and see how it behaves.

Consider the interpolation problem on a set X of n points $x_1, \ldots, x_n$. as in Section 1. For the Gaussian kernel [18] and other analytic radial basis functions [15], the flat limit for $\varepsilon \to 0$ exists and is a polynomial, if certain nondegeneracy conditions hold [15, I-III, p. 110]. In addition, then there is an expansion

$$s_{f,X,\Phi_\varepsilon}(x) = p_{0,f}(x) + \varepsilon^2 p_{2,f}(x) + \varepsilon^4 p_{4,f}(x) + \cdots, \tag{19}$$

of the interpolant, where $p_{0,f}$ is the polynomial interpolant in the flat limit, and where the other $p_{2j,f}$ are polynomials vanishing on all data sites. They are linear in the data of f on X, i.e. they have the form

$$p_{2j,f} = \sum_{i=1}^{n} f(x_i) p_{2j,i} \tag{20}$$

with polynomials $p_{2j,i}$ that are only dependent on X and the kernel Φ. If $p_{0,f}$ has degree K, then $p_{2j,f}$ has degree $K+2j$ [15].

When going towards scales that yield a smaller error than the flat limit for given f and X, consider

$$\sigma_f(x) = \operatorname{sgn}\left(p_{0,f}(x) - f(x)\right)$$

and write the error as a perturbation of the flat limit error as

$$\sigma_f(x)\left(s_{f,X,\Phi_\varepsilon}(x) - f(x)\right) = |p_{0,f}(x) - f(x)| + \sigma_f(x) \sum_{j=k}^{\infty} \varepsilon^{2j} \sum_{i=1}^{n} f(x_i) p_{2j,i}(x).$$

Some $\varepsilon > 0$ can only be an improvement over the flat limit, if

$$\sigma_f(x) \sum_{j=k}^{\infty} \varepsilon^{2j} \sum_{i=1}^{n} f(x_i) p_{2j,i}(x) < 0 \tag{21}$$

on all extremal points x of $f - p_{0,f}$. This might be true even for values of ε that are not extremely small. If we focus on very small ε and take the smallest k such that $p_{2k,f}$ is nonzero, we need

$$\sigma_f(x) \sum_{i=1}^{n} f(x_i) p_{2k,i}(x) < 0.$$

Both criteria are rather hard to check and analyze.

Users may think that the data points X and the function values $y_j = f(x_j)$ already determine whether the flat limit is optimal or not, and that this suffices to estimate a good scale. But this is not true. Even if the interpolation data are fixed, there are functions that can lead to any possible outcome.

Theorem 4. *For all choices of point sets $X = \{x_1, \ldots, x_n\}$ and function values $y_1, \ldots, y_n$ it cannot be decided whether the flat limit is error-optimal or not. More precisely, there are functions f that attain the values $y_j = f(x_j)$, $1 \le j \le n$ and have an optimal flat limit or not.*

Proof. Let the point set X and the kernel Φ be fixed. Furthermore, we consider fixed data values y_j at the $x_j \in X$ and let the functions f with $f(x_j) = y_j$, $x_j \in X$ vary. For each such function, the polynomials $p_{2k,f}$ are the same, and we denote them now by $p_{2k,X,Y}$ to separate the set X of data locations from the set Y of function values. Now let g be any function that vanishes on X and consider the function $f_g(x) = p_{0,X,Y}(x) - g(x)$. Then

$$\begin{aligned} s_{f_g,X,\Phi_\varepsilon}(x) - f_g(x) &= s_{p_{0,X,Y},X,\Phi_\varepsilon}(x) - p_{0,X,Y}(x) + g(x) \\ &= g(x) + \sum_{j=k}^{\infty} \varepsilon^{2j} p_{2j,X,Y}(x) \end{aligned}$$

shows that all functions g that vanish on X can arise as error functions in the flat limit. Let $k(X,Y,g)$ be the smallest $k \ge 1$ such that $p_{2k,X,Y}$ does not vanish on the extremal points of g. Then

$$\operatorname{sgn} p_{2k(X,Y,g),X,Y}(x) = -\operatorname{sgn} g(x)$$

should hold on the extremal points of g if the flat limit is not optimal for small ε.

This is the situation when X, Y, and g are prescribed. But one may take $k(X,Y)$ to be the smallest $k \ge 1$ such that $p_{2k,X,Y}$ does not vanish, and then fix $g = -p_{2k,X,Y}$ to arrive at the same situation, while the choice $g = +p_{2k,X,Y}$ leads to local optimality of the flat limit. □

The consequence is that users must spare data points for error evaluation, if they want to get anywhere with a numerical estimation of scales. This is part of the Leave-one-out-cross-validation (LOOCV) technique, see [9].

7 Conclusions and Open Problems

Choosing "good" scales for kernel-based interpolation remains a challenge, both in theory and practice. A few things could be contributed here, but there are many open issues.

Sobolev spaces on $\mathbb{R}^d$ form a scale of spaces, and each function belonging to one of these spaces belongs to all of them. When minimizing the norm of a given function over all scales of spaces, there is a unique minimum, defining a "natural" scale of the function. But this scale is hard to estimate, and it is only of limited practical value for finding error-optimal scales, because it minimizes a function norm, not an interpolation error norm. However, the "natural scale" may be useful elsewhere.

If the error of an interpolant cannot be evaluated because there are no additional data, one might look at the minimum of the standard error bound, because it does not need additional data. But, in contrast to the natural scale, there is no theoretical guarantee for a nonzero scale that minimizes the error. In particular, the flat limit may be the zero scale that yields the optimal error. There is some experimental support for the strategy to calculate the flat limit case in parallel to the standard scale-dependent solution. In a variety of cases, the flat limit error comes out small enough to let the unstable hunt for an error-optimal scale be not worth while. Criteria for this case would be helpful.

In the case of analytic kernels like the Gaussian or the multiquadrics, changes of scales imply nontrivial changes of the underlying Reproducing Kernel Hilbert space, making scale changes a hazardous issue even in theory. But this is no restriction if one focuses on weak error norms. In particular, one can expand the pointwise error into a series wrt. the scale, but the result does not easily yield useful criteria for "good" scales. These criteria should be improved. However, the expansions can be used for showing that for fixed interpolation data, the unknown function on the other points can be chosen to yield any possible error function. In addition, the flat limit may be optimal or not, just by changing the function outside the given interpolation data.

Readers are encouraged to extend the above situations, and this paper hopefully serves as a starting point.

References

[1] B. Baxter and R. Brummelhuis. Convergence Estimates for Stationary Radial Basis Function Interpolation and for Semi-discrete Collocation-Schemes. *J. Fourier Anal. Appl.* 28, 53 (2022)

[2] M.D. Buhmann. *Multivariable interpolation using radial basis functions*. PhD thesis, University of Cambridge, 1989.

[3] M.D. Buhmann. Radial functions on compact support. *Proceedings of the Edinburgh Mathematical Society*, 41:33–46, 1998.

[4] M.D. Buhmann. *Radial Basis Functions, Theory and Implementations*. Cambridge University Press, Cambridge,UK, 2003.

[5] O. Davydov and R. Schaback. Optimal stencils in Sobolev spaces. *IMA Journal of Numerical Analysis*, 39:398–422, 2019.

[6] St. De Marchi and R. Schaback. Stability of kernel-based interpolation. *Adv. in Comp. Math.*, 32:155–161, 2010.

[7] T.A. Driscoll and B. Fornberg. Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.*, 43:413–422, 2002.

[8] G. Fasshauer and M. McCourt. *Kernel-based Approximation Methods using MATLAB*, volume 19 of *Interdisciplinary Mathematical Sciences*. World Scientific, Singapore, 2015.

[9] G. Fasshauer and J.G. Zhang. On choosing optimal shape parameters for RBF approximation. *Numerical Algorithms*, 45:345–368, 2007.

[10] B. Fornberg, E. Larsson, and N. Flyer. Stable computations with Gaussian radial basis functions. *SIAM J. Sci. Comput.*, 33(2):869–892, 2011.

[11] B. Fornberg, E. Lehto, and C. Powell. Stable calculation of Gaussian-based RBF-FD stencils. *Computers and Mathematics with Applications*, 65:627–637, 2013.

[12] B. Fornberg and G. Wright. Stable computation of multiquadric interpolants for all values of the shape parameter. *Comput. Math. Appl.*, 48(5-6):853–867, 2004.

[13] B. Fornberg, G. Wright, and E. Larsson. Some observations regarding interpolants in the limit of flat radial basis functions. *Computers & Mathematics with Applications*, 47:37–55, 2004. doi:10.1016/S0898-1221(04)90004-1.

[14] E. J. Kansa and R.E. Carlson. Improved accuracy of multiquadric interpolation using variable shape parameters. *Comput. Math. Appl.*, 24:99–120, 1992.

[15] Elisabeth Larsson and Bengt Fornberg. Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. *Comput. Math. Appl.*, 49:103–130, 2005.

[16] Y.J. Lee, G.J. Yoon, and J. Yoon. Convergence property of increasingly flat radial basis function interpolation to polynomial interpolation. preprint, 2005.

[17] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3:251–264, 1995.

[18] R. Schaback. Multivariate interpolation by polynomials and radial basis functions. *Constructive Approximation*, 21:293–317, 2005.

[19] R. Schaback. Limit problems for interpolation by analytic radial basis functions. *J. Comp. Appl. Math.*, 212:127–149, 2008.

[20] M. Scheuerer, M. Schlather, and R. Schaback. Interpolation of spatial data - a stochastic or a deterministic problem? *European Journal of Applied Mathematics*, 24:601–629, 2013.

[21] G. Song, J. Riddle, G.E. Fasshauer, and F.J. Hickernell. Multivariate interpolation with increasingly flat radial basis functions of finite smoothness. *Adv. Comp. Math.*, 36:485–501, 2012.

[22] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396, 1995.

[23] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, Cambridge,UK, 2005.