On Local RBF Approximation *[†]

T.Hangelbroek[‡]

August 28, 2018

Abstract

The purpose of this paper is to investigate RBF approximation with highly nonuniform centers. Recently, DeVore and Ron have developed a notion of the local density of a set of centers – a notion that permits precise pointwise error estimates for surface spline approximation. We give an equivalent, alternative characterization of local density, one that allows effective placement of centers at different resolutions. We compare, also, the pointwise results of DeVore–Ron to previously works of Wu and Schaback and of Duchon.

1 Introduction

This brief article concerns local approximation results for radial basis function (RBF) approximation with the goal of effectively placing centers at varying resolutions. We consider RBF approximants of the form $x \mapsto \sum_{\xi \in \Xi} \phi(x - \xi)$, where the arrangement of centers, Ξ , may be highly nonuniform. A motivation for this set up is that centers may be placed strategically to treat defects in the target function. This becomes very important in high dimensions, where conventional quasi-uniform placement of centers is extremely costly; error estimates assuming a (small) fill distance h require a placement of $O(h^{-d})$ centers; obtaining a comparable error with fewer centers is clearly desirable. To this end, we seek a method by which Ξ can be chosen to achieve a pointwise error that reflects the local arrangement of Ξ . In turn, this requires finding a useful measure of the *local density* of Ξ .

In [4], DeVore and Ron establish powerful local error estimates for kernel based approximation; along the way they develop a satisfactory notion of local density – the *majorant* – expressed as a function over a domain containing the centers (see (3) below). This function gives, roughly, the distance to the nearest unisolvent subset of Ξ . However, it also satisfies an extra condition of global

^{*2000} Mathematics Subject Classification: 41A17, 26B35, 41A63, 42C15

 $^{^{\}dagger}Key \ words:$ error estimates, nonlinear approximation, optimal approximation, radial basis functions, scattered data, thin-plate splines, surface splines, approximation order

[‡]Texas A&M University College Station, TX 77843, USA. Research supported by NSF Postdoctoral Research Fellowship.

compatibility: it may not grow or shrink too rapidly. This condition is not stated explicitly, rather it is contained in the definition, but it plays an essential role in their local error estimates. In this article we investigate this compatibility condition and give an equivalent condition that lends itself to effective placement of highly nonuniform centers.

Error estimates in [4] show that kernel approximation bears a strong similarity to univariate spline approximation, which exhibits local convergence in the sense that error decays rapidly over intervals where breakpoints are tightly spaced. Indeed, such results have long been known for spline quasiinterpolation, at least since [2]. If Q_T is the quasi-interpolation operator of order r associated with knots $T = (t_j)$, then [3, (4.18)] in conjunction with a theorem of Whitney [3, Theorem 4.2] tells us that for $x \in [t_j, t_{j+1}]$,

$$|f(x) - Q_T(f)(x)| \le C_r \omega_{r+1}(f, t_{j+r} - t_{j-r+1}).$$

We note that the error from spline approximation at x depends only on the distribution of knots near x (the r^{th} nearest neighbors) and the smoothness of the target function in a neighborhood of x. This result can be attributed to the local nature of the B-splines basis (and its associated dual functionals).

Because kernels are often globally supported, and because a truly local basis similar to the B-splines seems to be out of the question, the distance to the nearest neighbors may be unsuitable for measuring local density. Kernel approximants exhibit *far field* effects, meaning that each kernel has a global influence, and the majorant of DeVore - Ron, via the global compatibility condition, penalizes remote, sparse density in an effort to mitigate such far field effects.

Remarkably, it is sometimes the case that RBF interpolation is local in a stronger sense than considered by DeVore – Ron. Specifically, rates of convergence for interpolation of certain target functions (those coming from the *native space*) with certain RBFs can be measured in terms of an expression that takes into account only the local distribution of centers. However, the drawback is that such results are presently only known to hold for target functions from a specific class.

In the following section, we present an alternative, equivalent characterization of the majorant of Devore and Ron, one that involves a global compatibility condition, and we give a self-contained development of their pointwise error estimates that explicitly uses this compatibility condition. In Section 3, we demonstrate how the compatibility condition may be used to place centers nonuniformly. Section 4 is a discussion of (previously known) local error estimates in the native space context.

2 Local Estimates

In [4], DeVore and Ron construct a local approximation scheme using a simple measure of the local density of centers. This initial notion of density (given in Definition 1) is not suitable to capture far field effects, and the analysis of the scheme's convergence eventually relies on a more refined notion of local density: the initial density's *majorant*. In this section, we recast the DeVore – Ron result with a different but equivalent local density parameter, one that lends itself to efficient placement

of centers. As in [4], we begin by giving the initial definition of local density, a function on \mathbb{R}^d which at each $\alpha \in \mathbb{R}^d$ indicates a radius sufficient to capture a K-stable local polynomial reproduction of order ℓ :

Definition 1 (Local Density). Given a set of centers $\Xi \subset \mathbb{R}^d$, a local density (LD) $\rho : \mathbb{R}^d \to \mathbb{R}_+$ is a function with an associated local polynomial reproduction of precision ℓ . That is to say, there is a kernel $a : \Xi \times \mathbb{R}^d \to \mathbb{R} : (\xi, \alpha) \mapsto a(\xi, \alpha)$ so that the following hold:

(Support) For $|\xi - \alpha| > \rho(\alpha)$, $a(\xi, \alpha) = 0$.

(**Precision**) For all $p \in \Pi_{\ell}$ we have $\sum_{\xi \in \Xi} a(\xi, \alpha) p(\xi) = p(\alpha)$.

(Stability) There is K > 0 such that $\sum_{\xi \in \Xi} |a(\xi, \alpha)| < K$ for all α .

We note that this definition is given, in more or less the same form, by Wu and Schaback in [8, Lemma 2] (we discuss their local results in Section 4.2).

This construction allows the surface spline, ϕ

$$\phi(x) := \phi_k(x) := C_{k,d} \begin{cases} |x|^{2k-d} \log |x| & \text{for even } d\\ |x|^{2k-d} & \text{for odd } d \end{cases}$$
(1)

(also known as polyharmonic splines because they are the fundamental solution of the k-fold Laplacian Δ^k), to be approximated by a linear combination of nearby shifts. This is accomplished with a small, local error:

$$|\phi(x-\alpha) - \sum a(\xi,\alpha)\phi(x-\xi)| \le C\,\rho(\alpha)^{2k-d} \left(1 + \frac{|x-\alpha|}{\rho(\alpha)}\right)^{-\nu}, \quad \nu := \ell + d - 2k$$

and one can generate the approximant:

$$T_{\Xi}f(x) := \int_{\mathbb{R}^d} \Delta^k f(\alpha) \sum_{\xi \in \Xi} a(\xi, \alpha) \phi(x - \xi) \, \mathrm{d}\alpha.$$

This leads to convenient pointwise error estimates:

$$|f(x) - T_{\Xi}f(x)| \le C \int_{\mathbb{R}^d} |\Delta^k f(\alpha)| \,\rho(\alpha)^{2k-d} \left(1 + \frac{|x-\alpha|}{\rho(\alpha)}\right)^{-\nu} \,\mathrm{d}\alpha.$$
(2)

Observe that the construction of the approximation operator T_{Ξ} depends only on a local polynomial reproduction $(\xi, \alpha) \mapsto a(\xi, \alpha)$. The estimate (2) holds for any LD with local polynomial reproduction a. In particular, it holds for any function ρ' with $\rho' \ge \rho$. With this in mind, it may seem tempting to use an optimally small ρ : the LD so that $\rho(\alpha)$ is the minimal radius around α that captures a K-stable polynomial reproduction of order ℓ . Sadly, because of the global nature of the kernels, this is unsuitable for producing precise estimates.

In [4] the LD is used to create a "majorant":

$$H(x) = \sup_{y \in \mathbb{R}^d} \rho(y) \left(1 + \frac{|x - y|}{\rho(y)} \right)^{-r}.$$
 (3)

In short, the coefficient kernel associated to the LD is used to create the approximant, but, in order to attack the estimate (2), an expression involving the majorant controls the error. The initial LD only plays an ancillary role: to construct the approximant via the coefficient kernel and to generate the majorant.

To be sure, the entire exercise could be repeated using only H and entirely without ρ . It is a simple task to show that H is itself an LD (since $H(x) \ge \rho(x)$ and the local polynomial reproduction $(\xi, x) \mapsto a(\xi, x)$ is a local polynomial reproduction for H as well). Moreover, the majorant of H is simply a constant multiple of H (a constant depending only on r). It follows that by replacing the initial LD ρ with its majorant H, one would obtain the same results.

Alternately, from the beginning one may impose the assumption that the LD is equivalent to its majorant (i.e., it is *self-majorizing*). This is a change of perspective: from the point of view that centers have been given outside of our control (and with the goal of remaining faithful to the local distribution of data by having error estimates reflecting the local density) to the setting where the spacing of centers is chosen to reflect regions of interest or to attack defects in the target function. We will choose centers having an LD that is correct from the start, having an extra condition designed to handle far field effects. The condition on the LD will be different from that of [4]; it is a slightly more easily verified property: *slow growth*. To proceed, we formalize both extra assumptions (slow growth and self-majorization) under the heading of global compatibility and take a moment to discuss their equivalence.

Definition 2 (Global Compatibility). If there is a constant $C_{sg} > 0$ so that for every x and α , we have

$$\rho(\alpha) \le C_{sg}\rho(x) \left(1 + \frac{|x - \alpha|}{\rho(x)}\right)^{1 - \epsilon},\tag{4}$$

we say ρ exhibits $1 - \epsilon$ slow growth.

The function $\rho : \Omega \to \mathbb{R}_+$ exhibits self-majorization of order r if there is a constant $C_{sm} > 0$ so that for every x and y, we have

$$\rho(y) \ge C_{sm}\rho(x) \left(1 + \frac{|x-y|}{\rho(x)}\right)^{-r} \tag{5}$$

The equivalence of these two assumptions can be expressed formally:

Lemma 3. If ρ satisfies the property of self-majorization (5) then it satisfies the property of slow growth (4) with $\epsilon = \frac{1}{r+1}$ and constant C_{sg} depending only on r and C_{sm} . Likewise, if ρ satisfies (4) then it satisfies (5) with $r = \frac{1-\epsilon}{\epsilon}$ and constant C_{sm} depending only on ϵ and C_{sg} . *Proof.* When $|x - \alpha| \ge \rho(x)$, (4) implies that $\rho(\alpha) \le 2^{1-\epsilon} C_{sg} \rho(x)^{\epsilon} |x - \alpha|^{1-\epsilon}$, so

$$\frac{2^{\epsilon-1}}{C_{sg}}\rho(\alpha)^{\epsilon} \left(\frac{|x-\alpha|}{\rho(\alpha)}\right)^{\epsilon-1} \le \rho(x)^{\epsilon}.$$

On the other hand, when $|x - \alpha| < \rho(x)$, $\rho(\alpha) \le 2^{1-\epsilon} C_{sg} \rho(x)$, so ρ satisfies self-majorization with $C_{sm} = \min\left(\frac{2^{\epsilon-1}}{C_{sg}}, \left(\frac{2^{\epsilon-1}}{C_{sg}}\right)^{1/\epsilon}\right)$.

When $|x - \alpha| \ge \rho(x)$, (5) implies that $\rho(\alpha) \ge 2^{-r} C_{sm} \rho(x)^{1+r} |x - \alpha|^{-r}$, so

$$\frac{2^r}{C_{sm}}\rho(\alpha)^{1+r}\left(\frac{|x-\alpha|}{\rho(\alpha)}\right)^r \ge \rho(x)^{1+r}.$$

On the other hand, when $|x - \alpha| < \rho(x)$, $\rho(\alpha) \leq 2^{-r}C_{sm}\rho(x)$, so ρ satisfies slow growth with $C_{sg} = \max\left(\frac{2^r}{C_{sm}}, \left(\frac{2^r}{C_{sm}}\right)^{1/(1+r)}\right)$.

Either of these extra assumptions on ρ are sufficient to obtain the error estimate in [4]:

Theorem 4 (DeVore Ron I). Let $\ell > 2k - d + 1$. Suppose that ρ satisfies Slow Growth with $\epsilon > \frac{2k}{\ell}$. There is a constant C so that for $f \in C^{2k}(\mathbb{R}^d)$ having compact support,

$$|f(x) - T_{\Xi}f(x)| \le C\rho(x)^{2k} \|\Delta^k f\|_{\infty}$$

Proof. This follows by applying the growth assumption to (2) and writing $\gamma = 1 - \epsilon$ to obtain:

$$\begin{aligned} |f(x) - T_{\Xi}f(x)| &\leq C \int_{\mathbb{R}^d} |\Delta^k f(\alpha)| \rho(x)^{2k-d} \left(1 + \frac{|x-\alpha|}{\rho(x)}\right)^{\gamma(2k-d)} \left(1 + \frac{\frac{|x-\alpha|}{\rho(x)}}{\left(1 + \frac{|x-\alpha|}{\rho(x)}\right)^{\gamma}}\right)^{-\nu} d\alpha \\ &\leq C\rho(x)^{2k-d} \|\Delta^k f(\alpha)\|_{\infty} \int_{\mathbb{R}^d} \left(1 + \frac{|x-\alpha|}{\rho(x)}\right)^{2k-d-\ell+\ell\gamma} d\alpha \\ &\leq C\rho(x)^{2k} \|\Delta^k f(\alpha)\|_{\infty} \int_0^\infty (1+R)^{2k-d-\ell+\ell\gamma} R^{d-1} dR. \end{aligned}$$

The second inequality follows by writing $(2k - d)\gamma - \nu(1 - \gamma) = (2k - d)\gamma - (\ell - 2k + d)(1 - \gamma)$. The convergence of the last integral is a consequence of the assumption $\gamma < 1 - \frac{2k}{\ell}$.

A further result from [4], is that functions of lower smoothness can also be treated with local error estimates. The operator T_{Ξ} is instrumental in obtaining low smoothness results, albeit indirectly. This is the point of [4, Theorem 5.3], which, for completeness, we rephrase in a simplified form as Theorem 5 in terms of the slow growth assumption. The lower order result is technically more complicated than that of Theorem 4. It is a common technique to use interpolation theory to obtain direct approximation results for functions of lower smoothness. DeVore and Ron use an argument of this type, that splits f into a rough but benign part, b, and a smooth part, g. This is done in a way that is not entirely straightforward, by controlling the size of b (and the smoothness of g) in a precise way to match the local density.

To discuss lower smoothness estimates, we first introduce fractional smoothness spaces. These can be expressed in numerous different ways: as Besov spaces, Triebel-Lizorkin spaces or (more familiarly) Hölder-Zygmund spaces. In the setting we consider, these are the same spaces. That is, we consider $F^{\sigma}_{\infty,\infty}(\mathbb{R}^d) = B^{\sigma}_{\infty,\infty}(\mathbb{R}^d) = C^{(\sigma)}(\mathbb{R}^d)$. However, the exact smoothness norm we employ is the $B^{\sigma}_{\infty,\infty}$ norm, defined in terms of wavelet coefficients. Smooth functions can be expanded as $f = \sum_{i=0}^{\infty} c_j \psi_j$ and the smoothness seminorm is expressed in terms of coefficients c_j .

A totally standard construction, used also in [4], indexes wavelets by gendered, dyadic cubes: $\nu \in \mathcal{D}$, where each $\nu = (e_{\nu}, I_{\nu})$ is a pair comprising:

- a gender $e = e_{\nu} \in \{0, 1\}^d \setminus \{\mathbf{0}\}$
- and a dyadic cube $I := I_{\nu} = 2^{-j} (k + [0, 1]^d).$

For a general dyadic cube of this form we denote the corner by $c(I) := 2^{-j}k$ and the side-length by $\ell(I) := 2^{-j}$. These definitions extend for gendered cubes: $c(\nu) := c(I_{\nu})$ and $\ell(\nu) := \ell(I_{\nu})$. Under this indexing, each gendered cube ν has exactly one parent ν' , where $I_{\nu} \subset I_{\nu'}$, $e_{\nu} = e_{\nu'}$ and $\ell(\nu') = 2\ell(\nu)$.

The wavelet system we employ is a family of C^r , compactly supported functions, with r > 2k. Each wavelet is related to one of $2^d - 1$ prototypes by affine changes of variable: $\psi_{\nu}(x) = \Psi_{e_{\nu}}\left(\frac{x-c(\nu)}{\ell(\nu)}\right)$. In other words, each wavelet is a translated, rescaled copy of one of the $2^d - 1$ functions $\Psi_e \in C^r(\mathbb{R}^d)$. Consequently the supports of wavelets are obtained by affine transformations, and each is contained in a ball with radius proportional to the side-length and centered at the corner of the cube $I(\nu)$. I.e., there is $\Gamma > 0$ so that for all $\nu \in \mathcal{D}$

$$\overline{I_{\nu}} := \operatorname{supp}(\psi_{\nu}) = c(\nu) + \ell(\nu) \times \operatorname{supp}(\Psi_{e_{\nu}}) \subset B(c(\nu), \Gamma\ell(\nu))$$

For orthogonal wavelet systems, compactly supported continuous functions have the unique expansion $f = \sum_{\nu \in \mathcal{D}} f_{\nu} \psi_{\nu}$. The smoothness seminorm of f is

$$|f|_{B^{\sigma}_{\infty,\infty}} := \sup_{\nu \in \mathcal{D}} \left(\ell(\nu)^{-\sigma} |f_{\nu}| \right).$$

Theorem 5 (DeVore Ron II). Let $\ell > 2k - d + 1$. Suppose that ρ satisfies Slow Growth with $\epsilon > \frac{2k}{\ell}$. There is C > 0 so that for all compactly supported $f \in B^{\sigma}_{\infty,\infty}$, $\sigma < 2k$, there is $s_{f,\Xi} \in \text{span}(\phi, \Xi)$ so that

$$|f(x) - s_{f,\Xi}(x)| \le C\rho(x)^{\sigma} ||f||_{B^{\sigma}_{\infty,\infty}}.$$

Proof. We split f = g + b where $|b(x)| \leq \rho(x)^{\sigma} |f|_{B_{\infty,\infty}^{\sigma}}$ and $|\Delta^k g(x)| \leq \rho(x)^{\sigma-2k} |f|_{B_{\infty,\infty}^{\sigma}}$. A consequence of this and Theorem 4 is that $|f(x) - T_{\Xi}g(x)| \leq \rho(x)^{\sigma} ||f||_{B_{\infty,\infty}^{\sigma}}$ and the theorem follows with $s_{f,\Xi} = T_{\Xi}g$.

To obtain the split, partition $\mathcal{D} = \mathcal{D}_q \cup \mathcal{D}_b$ by selecting cubes ν according to the density over \overline{I}_{ν} :

$$\nu \in \mathcal{D}_g \quad \text{iff} \quad \ell(\nu) \ge \rho(\nu) := \max_{y \in \overline{I}_{\nu}} \rho(y).$$

Define $g := \sum_{\nu \in \mathcal{D}_g} f_{\nu} \psi_{\nu} = \sum_{\ell(\nu) \ge \rho(\nu)} f_{\nu} \psi_{\nu}$. Estimating the iterated Laplacian of a wavelet is straightforward: for $x \in \overline{I}_{\nu}$, $|\Delta^k \psi_{\nu}(x)| \le C\ell(\nu)^{-2k}$. Consequently,

$$|\Delta^k g(x)| \le C \sum_{\substack{\nu \in \mathfrak{D}_g \\ x \in \overline{I_\nu}}} f_\nu \ell(\nu)^{-2k} \le C \, |f|_{B^{\sigma}_{\infty,\infty}} \sum_{\substack{\nu \in \mathfrak{D}_g \\ x \in \overline{I_\nu}}} \ell(\nu)^{\sigma-2k} \le C' |f|_{B^{\sigma}_{\infty,\infty}} (\rho(x))^{\sigma-2k} \le C' |f|_{B^{\sigma}_{\infty,\infty}} (\rho(x))$$

The final estimate deserves some explanation. Note that $\nu \in \mathcal{D}_g$ and $x \in \overline{I_\nu}$ imply that $\ell(\nu) \ge \rho(x)$. Finding $j \in \mathbb{Z}$ so that $2^j \ge \rho(x) > 2^{j-1}$, the number of wavelets with $\ell(\nu) = 2^j$ that have x in their support is bounded, $\#\{\nu : \ell(\nu) = 2^j, x \in \overline{I_\nu}\} \le N$, with a constant independent of x and j. We may rewrite the sum in the next to last expression in the chain of inequalities as

$$\sum_{\substack{\nu \in \underline{\mathcal{D}}_g \\ x \in \overline{I_\nu}}} \ell(\nu)^{\sigma - 2k} \le \sum_{\substack{\ell(\nu) = 2^j \\ x \in \overline{I_\nu}}} \sum_{i=0}^{\infty} \left(2^{j+i}\right)^{\sigma - 2k} \le N 2^{j(\sigma - 2k)} \sum_{i=0}^{\infty} \left(2^i\right)^{\sigma - 2k} \le C(\rho(x))^{\sigma - 2k}.$$

Estimating the size of $b(x) := \sum_{\nu \in \mathcal{D}_b} f_{\nu} \psi_{\nu}$, we write $|b(x)| \leq \sum_{\nu \in \mathcal{D}_b, x \in \overline{I}_{\nu}} |f_{\nu}|$, which is bounded by $|f|_{B^{\sigma}_{\infty,\infty}} \sum_{\nu \in \mathcal{D}_b, x \in \overline{I}_{\nu}} (\ell(\nu))^{\sigma}$. If x and y are in \overline{I}_{ν} and if $\rho(y) > \ell(\nu)$ then

$$\rho(x) \ge C_{sm}\rho(y) \left(1 + \frac{|x-y|}{\rho(y)}\right)^{-r} \ge \ell(\nu)C_{sm}(1+2\Gamma)^{-r} \quad \Rightarrow \quad \ell(\nu) \le C\rho(x)$$

As in the case of g, it follows that the series $|b(x)| \leq |f|_{B^{\sigma}_{\infty,\infty}} \sum_{\ell(\nu) \leq C\rho(x), x \in \overline{I}_{\nu}} (\ell(\nu))^{\sigma}$ can be rewritten as a sum of geometric series, to obtain $|b(x)| \leq C |f|_{B^{\sigma}_{\infty,\infty}} (\rho(x))^{\sigma}$.

3 Placing Centers

We now turn to a discussion of how condition (4) may be directly implemented to produce effective global approximants with local error estimates. In this section we present an algorithm for generating a set of centers with a fixed spacing, having a more refined spacing on a particular subset. This can be done in such a way that the pointwise error from surface spline approximation using ϕ_k (the surface spline of order k > d/2), reflects the local arrangement of centers.

We begin with a compact set Ω (e.g., a finite set or some lower dimensional manifold) in which we wish to place centers with increased density (say a spacing of h^s , s > 1). Furthermore, we wish to have an "ambient" density of h outside of Ω . Without loss, we assume $h = 2^{-j}$ and $h^s = 2^{-js}$.

Invoking a $1 - \epsilon$ slow growth condition from Definition 2 (a range of valid ϵ 's will be determined momentarily), with $\rho(\alpha) = 2^{-sj}$ and $\rho(x) = 2^{-j}$, we see that extra centers must be placed in

a region $\tilde{\Omega}$ with $\Omega \subset \tilde{\Omega} \subset \{x \mid \operatorname{dist}(x, \Omega) \leq 2^{-j(\frac{1-s\epsilon}{1-\epsilon})}\}$. This imposes a certain condition on the relationship between s, the slow growth parameter ϵ , and ultimately the polynomial precision ℓ (by the conditions of Theorem 4).

To force the set $\tilde{\Omega} \setminus \Omega$ to shrink with $h, \epsilon \times s$ should be less than 1; the smaller this product is, the smaller the region of extra centers will be. In turn, Theorem 4 forces $\ell > 2k/\epsilon$.

We place extra, gridded centers with dyadic spacing in annular regions around Ω . That is, we identify a sequence of annular regions $\Omega_0, \ldots, \Omega_{j_0}$ (with $j_0 = sj - j - 1$) around Ω . In each region Ω_J we place centers Ξ_J having constant spacing, and this spacing diminishes the farther Ω_J is from Ω . Specifically, the *J*th region, Ω_J , has centers with spacing 2^{-js+J} . The union of these sets $\bigcup_{J=0}^{sj-j-1} \Omega_J$ is $\widetilde{\Omega}$.

Initial step: The first such set, Ω_0 , contains Ω and has centers with spacing 2^{-sj} . We make it slightly larger, in order to ensure that sufficiently many centers are present to satisfy Definition 1. Thus we set

$$\Omega_0 = \{x \mid \operatorname{dist}(x, \Omega) \le \ell\} \quad \text{and} \quad \Xi_0 = 2^{-sj} \mathbb{Z}^d \cap \Omega_0$$

Inside Ω_0 we have placed a set of gridded centers with spacing 2^{-sj} .

From this, we have $\rho(\alpha) = \ell 2^{-sj}$ for $\alpha \in \Omega$. Indeed, for $\alpha \in \Omega$, there is a simplex S_{α} containing α and contained in Ω_0 , with side-length $\ell 2^{-sj}$ and corners in Ξ_0 . The points $S_{\alpha} \cap \Xi_0$ are in general position for interpolation by Π_{ℓ} , and the associated Lagrange functions for polynomial interpolation $l_{\xi}(\alpha)$ give the required local polynomial reproduction $a(\xi, \alpha) = l_{\xi}(\alpha)$. The stability constant K is none other than the Lebesgue constant for this interpolation problem, which is bounded by $\binom{2\ell-1}{\ell}$, as demonstrated in [1, Theorem 2.2].

General step: As we did before, the width of Ω_J can be determined from the slow growth condition. Setting

$$\Omega_J := \left\{ x \mid \operatorname{dist}(x,\Omega) \le \ell \, 2^{\left(\frac{J}{1-\epsilon}-js\right)} \right\} \setminus \bigcup_{\nu=0}^{J-1} \Omega_{\nu} \quad \text{and} \quad \Xi_J := \Omega_J \cap 2^{J-sj} \mathbb{Z}^d$$

guarantees that $\rho(\alpha) = \ell 2^{J-sj}$ in Ω_{J-1} , since $\alpha \in \Omega_{J-1}$ is at the center of a ball of radius $\ell 2^{J-sj}$ contained in $\bigcup_{\nu=0}^{J} \Omega_{\nu}$. As before, there is a simplex S_{α} containing α contained in this ball, and the corresponding Lagrange functions give the required coefficient kernel.

Verifying the slow growth condition: It follows that for $x \in \Omega$ and $\alpha \in \Omega_J$ the distance $\operatorname{dist}(x, \alpha) \geq \ell 2^{\frac{J-1}{1-\epsilon}-js}$, and

$$2^{J-sj} = 2^{-sj+1} \left(\frac{2^{\frac{J-1}{1-\epsilon}-sj}}{2^{-sj}} \right)^{1-\epsilon} \qquad \Longrightarrow \qquad \rho(\alpha) \le 2\rho(x) \left(1 + \frac{\operatorname{dist}(x,\alpha)}{\rho(x)} \right)^{1-\epsilon}$$

Likewise, for $\alpha \in \Omega_{J'}$ and $x \in \Omega_J$ with J = J' + m and $m \ge 1$, we can bound the distance by

$$\operatorname{dist}(x,\alpha) \ge \ell 2^{\frac{J'+m}{1-\epsilon}-js} - \ell 2^{\frac{J'}{1-\epsilon}-js} \ge \ell 2^{-js} 2^{\frac{J'}{1-\epsilon}} \left(2^{\frac{m}{1-\epsilon}}-1\right) \ge \frac{1}{2} 2^{-js} 2^{J'} 2^{\frac{m}{1-\epsilon}}.$$

This allows us to bound $\frac{\operatorname{dist}(x,\alpha)}{2^{J'+1-js}}$ from below by $\frac{1}{4}2^{\frac{m}{1-\epsilon}}$; in turn, $2^m \leq 4^{1-\epsilon} \left(\frac{\operatorname{dist}(x,\alpha)}{2^{J'+1-js}}\right)^{1-\epsilon}$. Thus,

$$\rho(x) = 2^{J'+m+1-js} \le 4^{1-\epsilon} 2^{J'+1-js} \left(1 + \frac{\operatorname{dist}(x,\alpha)}{2^{J'+1-js}}\right)^{1-\epsilon} = 4^{1-\epsilon} \rho(\alpha) \left(1 + \frac{\operatorname{dist}(x,\alpha)}{\rho(\alpha)}\right)^{1-\epsilon}.$$

It follows that Theorems 4 and 5 hold for approximation by ϕ_k and for this set of centers with local density ρ .

Example. In this example the global spacing is $h = 2^{-j}$, while the spacing near the origin will be $h^2 = 2^{-2j}$. We choose to impose a slow growth condition with $\epsilon = 1/3$. By Theorem 4 we observe that the LD must have precision $\ell > 2k/\epsilon$, so we choose $\ell = 7k$.



Figure 1: A configuration of centers for thin plate spline approximation (k = 2, as in the first remark) where the density ranges from $\rho(0) \sim h^2 = 2^{-6}$ to the coarsest density is $\rho(x) \sim h = 2^{-3}$.

The spacing of centers immediately around the origin should be 2^{-2j} , but there is an intermediate region where the spacing grows slowly. We decompose this in j regions of increasing width:

$$\Omega_J := \{ x : |x| \le 7k \times 2^{\frac{3J}{2} - 2j} \} \setminus \bigcup_{0 \le \ell < J} \Omega_\ell, \quad J = 1, \dots, j$$

in which we place gridded centers with spacing $h_J = 2^{J-2j}$. Thus, $\rho(x) \leq 4^{2/3}\rho(y)(1 + \frac{|x|}{\rho(y)})^{2/3}$ holds throughout \mathbb{R}^d , and by Theorem 5 there is a constant C so that for any f with smoothness $0 < s \leq 2k$ we have a good approximant with dramatically increased accuracy at the origin:

 $||f - s_f||_{\infty} \le Ch^s ||f||_s$ and $|f(0) - s_f(0)| \le Ch^{2s} ||f||_s$.

In fact, for $x \in \Omega_J$, one has $|f(x) - s_f(x)| \le C2^{s(J-2j)} ||f||_s$.

We note finally that the cardinality of centers in the ball $B(0, 7k \times 2^{-j/2})$ is less than

$$\sum_{J=0}^{j} (7k)^d 2^{d(\frac{3J}{2}-2j)} 2^{-d(J-2j)} \le (7k)^d \sum_{J=0}^{j} 2^{d(\frac{J}{2})} \le C_d(7k)^d 2^{d(\frac{j}{2})}.$$

Since the same ball filled with centers having uniform spacing $h = 2^{-j}$ holds roughly $(7k)^d 2^{d(j/2)}$ centers, the increased precision comes at a cost of adding only a multiple of the original centers.

4 Local Error Estimates for RBF Interpolation

Many RBF interpolation results provide (implicitly or explicitly) local error estimates similar to those of Theorem 4. In this section we discuss two well known techniques for providing error estimates for RBF interpolation which happen to give estimates that are local.

In sharp contrast to the results of DeVore and Ron, only Definition 1 is needed for these estimates, and no extra global compatibility assumptions such as Definition 2 are needed for these (in other words, the estimates depend only on the nearest neighbors to point where the error is being measured). The drawback comes from the fact that the estimates hold for only particular classes of target functions (i.e., for functions in the native space), and, although the error being measured is pointwise, the smoothness assumption on the target function is generally not measured in L_{∞} .

This section is organized in two parts: first we discuss the localness inherent in Duchon's seminal result [5] concerning surface spline interpolation, and then we discuss the more general (in that it treats more kernels, and also addresses error of the derivative) "power function" method.

4.1 Local Error Estimates for Surface Spline Interpolation

In this case we consider interpolation of a function f, defined initially on a bounded domain Ω with Lipschitz boundary, at a finite set of points $\Xi \subset \Omega$ using surface splines defined in (1). That is, for k > d/2 we consider the unique interpolant to f at Ξ of the form

$$I_{\Xi}f(x) = \sum_{\xi \in \Xi} A_{\xi}\phi(x-\xi) + p(x), \quad \text{with } p \in \Pi_{k-1} \text{ and } \sum_{\xi \in \Xi} A_{\xi}q(\xi) = 0 \text{ for all } q \in \Pi_{k-1}.$$

This is equivalent to finding the interpolant $s_{f,\Xi}$ in the Beppo-Levi space (also known as the homogeneous Sobolev space) $D^{-k}L_2(\mathbb{R}^d) := \{f \in C(\mathbb{R}^d) \mid D^{\alpha}f \in L_2(\mathbb{R}^d) \forall |\alpha| = k\}$ to f at Ξ that minimizes the Sobolev seminorm on \mathbb{R}^d : $|u|_k := |u|_{W_2^k(\mathbb{R}^d)}$, where the Sobolev seminorm defined on a measurable set U is defined as

$$|u|_{W_2^k(U)}^2 := \int_U \sum_{|\alpha|=k} {k \choose \alpha} |D^{\alpha}u(x)|^2 \,\mathrm{d}x.$$

Duchon's error estimates result from two key observations:

- 1. a "zeros lemma" stating that any function defined on a ball B, in the Sobolev class $W_2^k(B)$ (with k > d/2), having many zeros in B is necessarily small in $L_{\infty}(B)$, and
- 2. an observation that the interpolation projector is bounded on $D^{-k}L_2$, indeed, it is orthogonal with respect to the $D^{-k}L_2$ semi-inner product, with the space of functions vanishing on Ξ as its nullspace.

1. Zeros Lemma It is not difficult, in this context, to derive a simplified, pointwise version of Duchon's zeros lemma [5, Proposition 2]. Namely, for k > d/2 and $u \in W_2^k$ vanishing on a set Ξ satisfying the conditions of Definition 1 with precision $\ell \ge k - 1$, we have

$$|u(\alpha)| \le C(\rho(\alpha))^{m-d/2} |u|_{W_2^k(B(\alpha,\rho(\alpha)))}.$$
(6)

To demonstrate this, we begin by observing that $|u(\alpha)| \leq (1+K) ||p-u||_{L_{\infty}(B(\alpha,\rho(\alpha)))}$. This follows directly from the fact that $|u(\alpha)| \leq |p(\alpha)| + |u(\alpha) - p(\alpha)|$, for all $p \in \Pi_{\ell}$, and from Definition 1:

$$|p(\alpha)| \le K \sup_{\xi \in \Xi \cap B(\alpha, \rho(\alpha))} |p(\xi)| \le K \sup_{\xi \in \Xi \cap B(\alpha, \rho(\alpha))} |p(\xi) - u(\xi)| \le K ||p - u||_{L_{\infty}(B(\alpha, \rho(\alpha)))}$$

Estimate (6) follows by dilating and translating a well-known result from polynomial approximation: $\inf_{p\in\Pi_{\ell}} \|p-u\|_{L_{\infty}\left(B(0,1)\right)} \leq \inf_{p\in\Pi_{k-1}} \|p-u\|_{L_{\infty}\left(B(0,1)\right)} \leq C|u|_{W_{2}^{k}\left(B(0,1)\right)}.$

2. Orthogonal Projection By [5, Lemma 3.2], the interpolation operator

$$I_{\Xi}: D^{-k}L_2(\mathbb{R}^d) \mapsto \left\{ \sum_{\xi \in \Xi} A_{\Xi} \phi(\cdot - \xi) + p \mid p \in \Pi_{m-1} \text{ and } \sum_{\xi \in \Xi} A_{\xi} q(\xi) = 0 \text{ for all } q \in \Pi_{k-1} \right\}$$

satisfies the orthogonality relationship

$$|I_{\Xi}f|_{k}^{2} + |f - I_{\Xi}f|_{k}^{2} = |f|_{k}^{2}$$

for all $f \in D^{-k}L_2(\mathbb{R}^d)$.

Local Error Estimate At this point, we can apply the argument of [5, Proposition 3]. We have

$$|f(\alpha) - I_{\Xi}f(\alpha)| \leq C(\rho(\alpha))^{k-d/2} |f - I_{\Xi}f|_{W_2^k(B(\alpha,\rho(\alpha)))}$$

$$\leq C(\rho(\alpha))^{k-d/2} |f - I_{\Xi}f|_k$$

$$\leq C(\rho(\alpha))^{k-d/2} |f|_k.$$

We note that no extra conditions are necessary beyond Definition 1, but this comes with the drawback that the order is known only for functions in $W_2^k(\mathbb{R}^d)$, and that this is an unconventional smoothness space (for this problem), in that it measures smoothness in L_2 rather than L_{∞} . In contrast, Theorems 4 and 5 seem to require Definition 2, but they give orders in the range $s \in [0, 2k]$ for functions from C^s .

4.2 Local error estimates via the power function method

An alternative method for measuring error, due originally to Wu and Schaback, estimates the pointwise error in terms of the 'power function' associated with a conditionally positive definite kernel ϕ ,

$$\mathcal{P}_{\Xi}(x) := \left(\phi(0) - 2\sum_{\xi \in \Xi} \chi_{\xi}(x)\phi(x-\xi) + \sum_{\xi,\zeta \in \Xi} \chi_{\zeta}(x)\chi_{\xi}(x)\phi(\zeta-\xi)\right)^{1/2}.$$

At each point x, it measures, roughly, the norm of the representer of the interpolation error at x: $\delta_x(\mathrm{Id} - I_{\Xi})$ in a certain reproducing kernel semi-Hilbert space, the *native space*. Our discussion is follows its development in the article of Wu and Schaback, [8], where the local RBF error estimates we now discuss were first presented, and Wendland's text [7].

Consider RBF interpolation by a radial function that is conditionally positive definite of order m: a radial function ϕ which, for any finite $\Xi \subset \mathbb{R}^d$, has a collocation matrix $\mathcal{C}_{\Xi} := (\phi(\xi - \zeta))_{\xi,\zeta}$ that is positive definite on vectors $\mathbf{A} \in \mathbb{R}^{\Xi}$ satisfying

$$\sum_{\xi \in \Xi} A_{\xi} p(\xi) = 0 \quad \text{for all} \quad p \in \Pi_{m-1}$$
(7)

(we assume $\Pi_{-1} = \{0\}$ when m = 0, in this case the kernel is simply positive definite). For any finite, unisolvent $\Xi \in \mathbb{R}^d$ (with respect to Π_{m-1}), the interpolation problem possesses a unique solution $I_{\Xi}f$ in $S_m(\Xi) := \{\sum_{\xi \in \Xi} A_{\xi}\phi(\cdot - \xi) + p \mid p \in \Pi_{m-1}, \mathbf{A} \text{ satisfying (7)} \}.$

The associated native space is a semi-Hilbert space \mathcal{N} with a semi-inner product $\langle \cdot, \cdot \rangle_{\mathcal{N}}$ determined by ϕ . See [6] or [7, Chapter 8] for a detailed construction. The semi-inner product has Π_{m-1} as its nullspace and there is a reproducing kernel, in the following sense:

$$f(x) = (\operatorname{Proj}_{m-1}f)(x) + \langle f, G(\cdot, x) \rangle_{\mathcal{N}} \quad \text{for all } f \in \mathcal{N}.$$
(8)

The operator $\operatorname{Proj}_{m-1}$ is a projector onto \prod_{m-1} with nullspace determined by a fixed set of points X poised for interpolation by Π_{m-1} . I.e., X is a fixed, unisolvent set of points (with respect to Π_{m-1}) with $\#X = \dim \Pi_{m-1}$. The nullspace of $\operatorname{Proj}_{m-1}$ is simply $\left(\operatorname{span}_{x_j \in X} \delta_{x_j}\right) \perp$, the joint kernel of the functionals δ_{x_i} . For each x, the function $G(\cdot, x)$ (which is uniquely determined by the projector $\operatorname{Proj}_{m-1}$ and, in turn, by the fixed set X) is in \mathcal{N} , and reproduces the functional $\delta_{(x)} := \delta_x - \delta_x \operatorname{Proj}_{m-1}.$

By expressing the interpolant in terms of the Lagrange basis, $I_{\Xi}f = \sum_{\xi \in \Xi} f(\xi)\chi_{\xi}$, the interpolation error can be expressed, with the help of (8) as

I

$$|f(x) - I_{\Xi}f(x)| = \left| \langle f, G(\cdot, x) \rangle_{\mathbb{N}} + \operatorname{Proj}_{m-1}f(x) - \sum_{\xi \in \Xi} \left(\langle f, G(\cdot, \xi) \rangle_{\mathbb{N}} + \operatorname{Proj}_{m-1}f(\xi) \right) \chi_{\xi}(x) \right|.$$

Since $\Pi_{m-1} \subset S_m(\Xi)$, $p \in \Pi_{m-1}$ can be written as $p = \sum_{\xi \in \Xi} p(\xi) \chi_{\xi}$, and we can simplify the above expression:

$$|f(x) - I_{\Xi}f(x)| = \left| \left\langle f, G(\cdot, x) - \sum_{\xi \in \Xi} \chi_{\xi}(x) G(\cdot, \xi) \right\rangle_{\mathcal{N}} \right| \le |f|_{\mathcal{N}} \left| G(\cdot, x) - \sum_{\xi \in \Xi} \chi_{\xi}(x) G(\cdot, \xi) \right|_{\mathcal{N}}$$

The quadratic form $Q_x(\mathbf{u}) := \phi(0) - 2\mathbf{u}^{\mathbf{T}} R_{\Xi}(x) + \mathbf{u}^{T} \mathfrak{C}_{\Xi} \mathbf{u}$, with $R_{\Xi}(x) = (\phi(x-\xi))_{\xi \in \Xi}$, is defined for $\mathbf{u} \in \mathbb{R}^{\Xi}$, and one easily sees that $\mathcal{P}_{\Xi}(x)^2 = Q_x(\mathbf{u}^*)$ with $\mathbf{u}^* = (\chi_{\xi}(x))_{\xi \in \Xi}$. By [7, Lemma 11.3], one has for certain admissible vectors \mathbf{u} – namely for $\mathbf{u} \in \mathbb{M}_x := \{\mathbf{u} \in \mathbb{R}^{\Xi} \mid \sum_{\xi \in \Xi} u_{\xi} p(\xi) = p(x)\}$ that the quadratic form is related to the function G by $Q_x(\mathbf{u}) = |G(\cdot, x) - \sum_{\xi \in \Xi} u_\xi G(\cdot, \xi)|_{\mathcal{N}}^2$. Therefore, it follows that $|f(x) - I_{\Xi}f(x)| \leq |f|_{\mathcal{N}} \mathcal{P}_{\Xi}(x)$, which is [8, Theorem 4], and was alluded to in the first paragraph of this subsection.

We can say more, however, since the minimum of $Q_x(\mathbf{u})$ over \mathbb{M}_x is $\mathcal{P}_{\Xi}(x)^2$, [8, Theorem 1]. It follows that one can estimate the power function at x using $Q_x(\mathbf{u})$, for any other $\mathbf{u} \in \mathbb{M}_x$. We choose **u** determined by the coefficient kernel $u_{\xi} = a(\xi, x)$ obtained from Definition 1 (with precision $\ell \geq m-1$). In other words,

$$\left(\mathcal{P}_{\Xi}(x)\right)^2 \le \phi(0) - 2\sum_{\xi \in \Xi} a(\xi, x)\phi(x-\xi) + \sum_{\xi,\zeta \in \Xi} a(\zeta, x)a(\xi, x)\phi(\zeta-\xi)$$

Assume now that the RBF ϕ is in $C^s(\mathbb{R}^d)$, $s \in (0, \infty)$ and that Definition 1 holds with polynomial precision ℓ where $\ell = \max([s], m) - 1$. Polynomial reproduction then gives, for any $p \in \Pi_{\ell}$:

$$(\mathfrak{P}_{\Xi}(x))^{2} \leq \phi(x-x) - p(x-x) - \sum_{\xi \in \Xi} a(\xi, x) (\phi(x-\xi) - p(x-\xi)) - \sum_{\xi \in \Xi} a(\xi, x) \left((\phi(x-\xi) - p(x-\xi)) - \sum_{\zeta \in \Xi} a(\zeta, x) (\phi(\zeta-\xi) - p(\zeta-\xi)) \right) \leq (1+K) \|\phi - p\|_{L_{\infty}(0,\rho(x))} + (K+K^{2}) \|\phi - p\|_{L_{\infty}(0,2\rho(x))} \leq C(1+K)^{2} (\rho(x))^{s}.$$

It follows that for $f \in \mathbb{N}$ the interpolation error satisfies the pointwise bound:

$$|f(x) - I_{\Xi}f(x)| \le C(1+K)(\rho(x))^{s/2} |f|_{\mathbb{N}}.$$

In this case, we note that, again, no extra conditions are necessary beyond Definition 1, but this comes with the drawback that the order is known only for functions in \mathcal{N} , which typically measures smoothness in L_2 rather than L_{∞} . In this case, however, the power function approach has an extra advantage, which we have not discussed (treated in [8] and [7, Chapter 11]) in that it gives estimates for derivatives of the error, as well.

References

- L. P. Bos, Bounding the Lebesgue function for Lagrange interpolation in a simplex, J. Approx. Theory, 38 (1983), pp. 43–59.
- [2] C. DE BOOR AND G. J. FIX, Spline approximation by quasiinterpolants, J. Approximation Theory, 8 (1973), pp. 19–45. Collection of articles dedicated to Isaac Jacob Schoenberg on his 70th birthday, I.
- [3] R. A. DEVORE AND G. G. LORENTZ, Constructive approximation, vol. 303 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1993.
- [4] R. A. DEVORE AND A. RON, Approximation using scattered shifts of a multivariate function. Unpublished manuscript.
- [5] J. DUCHON, Sur l'erreur d'interpolation des fonctions de plusieurs variables par les D^m-splines, RAIRO Anal. Numér., 12 (1978), pp. 325–334, vi.
- [6] R. SCHABACK, Native Hilbert spaces for radial basis functions. I, in New developments in approximation theory (Dortmund, 1998), vol. 132 of Internat. Ser. Numer. Math., Birkhäuser, Basel, 1999, pp. 255–282.
- [7] H. WENDLAND, *Scattered data approximation*, vol. 17 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2005.
- [8] Z. M. WU AND R. SCHABACK, Local error estimates for radial basis function interpolation of scattered data, IMA J. Numer. Anal., 13 (1993), pp. 13–27.