Multistep Approximation Algorithms: Improved Convergence Rates through Postconditioning with Smoothing Kernels

Gregory E. Fasshauer and Joseph W. Jerome

Abstract. We show how certain widely used multistep approximation algorithms can be interpreted as instances of an approximate Newton method. It was shown in an earlier paper by the second author that the convergence rates of approximate Newton methods (in the context of the numerical solution of PDEs) suffer from a "loss of derivatives", and that the subsequent linear rate of convergence can be improved to be superlinear using an adaptation of Nash-Moser iteration for numerical analysis purposes; the essence of the adaptation being a splitting of the inversion and the smoothing into two separate steps. We show how these ideas apply to scattered data approximation as well as the numerical solution of partial differential equations. We investigate the use of several radial kernels for the smoothing operation. In our numerical examples we use radial basis functions also in the inversion step.

1. Introduction

It has been only very recently that the idea of multistep (or multilevel) interpolation/approximation has attracted the interest of a variety of researchers in approximation theory. In [2] the authors use singular integrals on the sphere to perform multilevel approximation, and in [5] locally supported radial basis functions are used to do multilevel interpolation. The paper [2] contains some theoretical results regarding the convergence of the proposed algorithm, whereas [5] is purely computational/experimental in nature. However, the latter paper led to the theoretical investigations of a modified multistep interpolation algorithm in [11]. The paper [5] is also the starting point of our work. We therefore describe the basic iterative algorithm proposed there in more detail in the next section. One of the fundamental ideas in our paper is to interpret the multilevel algorithm as an instance of Newton iteration. This point is explained in detail in Section 2. Once this identification has been made we can draw upon work of the second author (see [7]) which establishes a theory for the superlinear convergence of this Newton iteration if it is coupled with an additional smoothing step. The general framework for these results is formulated in Section 3, and the smoothing operation, which is based on work of Hörmander [6], is carefully analyzed in the fourth section. The remaining sections are dedicated to numerical experiments and concluding remarks.

2. Multilevel Interpolation and Newton Iteration

Since most of our computational results as well as most of the existing work on multilevel methods is based on (locally supported) radial basis functions, we give a brief summary of these functions and some of their properties.

In the context of scattered data interpolation or approximation, radial basis functions are usually defined by starting with a set of nodes (or centers), say $\Xi = \{\xi_1, \ldots, \xi_N\} \subset \mathbb{R}^d$, which, for interpolation, are generally assumed to coincide with the data sites. A radial basis function is then generated by composing an appropriate continuous function $\varphi : [0, \infty) \to \mathbb{R}$ with some norm in $\mathbb{R}^d$ – usually the Euclidean one. A radial basis function, centered at ξ_k, is therefore of the form

$$x \mapsto \varphi(\|x - \xi_k\|_2), \qquad x \in \mathbb{R}^d .$$

In order to have a coherent theory, φ is assumed to be strictly (conditionally) positive definite. With $r = \|x - \cdot\|$, and $c > 0$, the most prominent examples of globally supported radial basis functions are $\varphi(r) = r$ (norm), $\varphi(r) = \sqrt{r^2 + c^2}$ (multiquadrics), $\varphi(r) = 1/\sqrt{r^2 + c^2}$ (reciprocal multiquadrics), $\varphi(r) = e^{-cr^2}$ (Gaussians), and $\varphi(r) = r^2 \log r$ (thin plate splines in $\mathbb{R}^2$). As already mentioned, these functions are globally supported, and therefore (even though they have many very nice properties) their applicability to interpolation or approximation is limited due to the resulting dense system matrices which present severe computational difficulties for large problems. This is what motivated the use of locally supported radial basis functions in a multilevel algorithm in [5]. It seems that the most popular class of these functions is due to Wendland [14]. His local basis functions are the product of a truncated power function and a polynomial. In contrast to the "classical" radial basis functions, these functions are strictly positive definite on $\mathbb{R}^d$ only for d less than or equal to some number d_0, and they possess only finite smoothness, say of order k. However, they lead to sparse matrices. A few examples (with appropriate values of k and d_0) are

$$\begin{aligned}
&\varphi(r) = (1-r)_+ && k=0,\ d_0=1,\\
&\varphi(r) = (1-r)_+^3(3r+1) && k=2,\ d_0=1,\\
&\varphi(r) = (1-r)_+^2 && k=0,\ d_0=3,\\
&\varphi(r) = (1-r)_+^4(4r+1) && k=2,\ d_0=3,\\
&\varphi(r) = (1-r)_+^6(35r^2+18r+3) && k=4,\ d_0=3.
\end{aligned}$$

The multilevel interpolation algorithm of Floater and Iske can be explained as follows. We start with given data sites $\mathcal{X} = \{x_1, \ldots, x_N\} \subset \Omega \subset \mathbb{R}^d$, and associated values $f(x_1), \ldots, f(x_N) \in \mathbb{R}$ which we would like to interpolate. In [5] a thinning algorithm is used to decompose $\mathcal{X}$ into a sequence of nested subsets $\mathcal{X}_1 \subset \mathcal{X}_2 \subset \cdots \subset \mathcal{X}_M = \mathcal{X}$. At each level i in the iteration the interpolant is assumed to be of the form

$$s_i(x) = \sum_{j=1}^{N_i} c_j^{(i)} \varphi_{\delta_i}(\|x - x_j^{(i)}\|) \qquad i \in \{1, \ldots, M\}, \tag{2.1}$$

where $\varphi_{\delta_i}(\cdot) = \varphi(\cdot/\delta_i)$ is a scaled locally supported radial basis function whose scale δ_i matches the data density at level i, and N_i is the number of points in the set $\mathcal{X}_i = \{x_1^{(i)}, \ldots, x_{N_i}^{(i)}\}$.

The algorithm proceeds by interpolating at each level i to the residual of the previous level $i-1$, i.e., s_1 is the interpolant to f on the set $\mathcal{X}_1$, and for $i > 1$, s_i is the interpolant to $f - \sum_{j=1}^{i-1} s_j$ on $\mathcal{X}_i$. This implies that $s_1 + s_2 + \cdots + s_M$ interpolates the entire set of given data.

In order to see the connection of this algorithm to Newton iteration, we define a function $F : L^p(\Omega) \to L^p(\Omega)$ which computes the residual, i.e.,

$$F(u) = u - f.$$

In terms of this function F, we can reformulate the interpolation problem. It now becomes: Find a function u such that $F(u)|_{\mathcal{X}} = 0$. This problem can be solved by Newton iteration,

$$u_i - u_{i-1} = -T_{\mathcal{X}_i} F(u_{i-1}), \tag{2.2}$$

where $T_{\mathcal{X}_i} F(u_{i-1})$ is the *interpolant to the residual* $u_{i-1} - f$ *on* $\mathcal{X}_i$. We can also see that $T_{\mathcal{X}_i}$ is an approximation to F'^{-1} ($=I$, the identity operator here) as in the classical Newton method.

A simple telescoping argument now shows the equivalence of the Floater-Iske algorithm and Newton iteration. Let $u_0 = 0$ and $u_i - u_{i-1} = s_i$, $i = 1, \ldots, M$, and observe:

$$\sum_{j=1}^{i-1} s_j = u_1 + (u_2 - u_1) + \cdots + (u_{i-1} - u_{i-2}) = u_{i-1}.$$

Then, clearly, solving the interpolation problem,

$$s_i|_{\mathcal{X}_i} = (f - \sum_{j=1}^{i-1} s_j)|_{\mathcal{X}_i} \tag{2.3}$$

is equivalent to

$$u_i - u_{i-1}|_{\mathcal{X}_i} = -F(u_{i-1})|_{\mathcal{X}_i},$$

and the final interpolant to the given data on $\mathcal{X}$ is obtained as $s_1 + s_2 + \cdots + s_M = u_M$.

Since we also consider the solution of partial differential equations by this algorithm we give a modification of the Floater-Iske algorithm for linear m-th order partial differential equations

$$L[u](x) = f(x), \qquad x \in \Omega \subset \mathbb{R}^d, \tag{2.4}$$

with homogeneous boundary conditions. Therefore, we define $F : W_p^m(\Omega) \to L^p(\Omega)$, such that

$$F(u) = L[u] - f,$$

for L a linear differential operator of order m, and $u_0 = 0$, $u_i - u_{i-1} = s_i$, $i = 1, 2, \ldots, M$. An approximate solution is obtained by using (2.2) to find a zero of F. Again we can say that $T_{\mathcal{X}_i}$ is an approximation of F'^{-1}, which reduces to L^{-1} in this setting. Therefore, $T_{\mathcal{X}_i} F(u_{i-1})$ corresponds to the *numerical solution of the differential equation* $L[u_{i-1}] - f = 0$ *on the grid* $\mathcal{X}_i$. Written in multistep notation, this becomes

$$\begin{aligned} L[s_1]|_{\mathcal{X}_1} &= f|_{\mathcal{X}_1} \\ L[s_2]|_{\mathcal{X}_2} &= (f - L[s_1])|_{\mathcal{X}_2} \\ &\vdots \\ L[s_M]|_{\mathcal{X}_M} &= (f - \sum_{i=1}^{M-1} L[s_i])|_{\mathcal{X}_M}. \end{aligned} \tag{2.5}$$

Therefore $L[s_1] + L[s_2] + \cdots + L[s_M]|_{\mathcal{X}_M} = f|_{\mathcal{X}_M}$, and so $u_M = \sum_{i=1}^M s_i$ is an approximate solution of the problem (2.4) based on the node set $\mathcal{X}_M$.

3. Adaptive Newton Algorithm with Postconditioning

In [7] Jerome dealt with a more general version of the Newton iteration described in the previous section. The goal of [7] was to establish a convergence theory for the solution of *nonlinear* differential equations also. In this more general setting the operator $T_{\mathcal{X}_i}$ also depends on the intermediate solution u_{i-1}, and therefore, in the following discussion we will replace $T_{\mathcal{X}_i}$ by $T_{\mathcal{X}_i}(u_{i-1})$ and allow nonlinear differential equations also. The results in [7] show that a loss in the convergence rate of the standard Newton iteration can be recovered by adding a smoothing operation in a postconditioning step. More precisely, the iteration now proceeds as

$$u_i - u_{i-1} = -S_{t_i} T_{\mathcal{X}_i}(u_{i-1}) F(u_{i-1}), \tag{3.1}$$

where S_{t_i} is a smoothing operation with smoothing speed t_i. This idea essentially had already been presented by Moser in [10]. However, there the smoothing S

and the numerical inversion T were combined into one single operation. The main idea in [7] was the separation of these two operations, and a thorough theoretical analysis of the individual operations. In this section we present the central points of the Jerome theory. The analysis of the smoothing operation itself (which is based on work of Hörmander [6]) is presented in the next section.

In order to be able to discuss convergence, we take a moment to introduce the function spaces in which we intend to work. Whereas the results in [7], and also in [6], were formulated for functions in Hölder spaces, we intend to present our results in the more widely used Sobolev spaces (with certain Besov or Lipschitz spaces as intermediate spaces). To this end, we recall (see e.g. [13]) that a function u is in the Sobolev space $W_p^k(\Omega)$, $\Omega \subset \mathbb{R}^d$, if the seminorm

$$|u|_{k,p} = \sum_{|\alpha|=k} \|D^\alpha u\|_p$$

is finite. The norm in W_p^k is defined as

$$\|u\|_{k,p} = \|u\|_p + |u|_{k,p},$$

where

$$\|u\|_p = \left(\int_\Omega |u(x)|^p dx \right)^{1/p}$$

is the usual L^p norm. The intermediate Besov spaces $B^s_{p,\infty}(\Omega)$ will be equipped with the norm

$$\|u\|_{s,p,\infty} = \|u\|_p + |u|_{s,p,\infty}$$

where $s = k + r$, $0 < r < 1$, and the seminorm is defined as

$$|u|_{s,p,\infty} = \sum_{|\alpha| \le k} \sup_{h \in \Omega} \frac{\|\Delta_h D^\alpha u\|_p}{|h|^r}.$$

We are now ready to state the main result of [7] as interpreted in the context of Sobolev and Besov spaces. In order to make the formulation of the following theorem more transparent we first state and comment on its three main assumptions:

(A1) Assume m is a nonnegative integer, and set $X_\sigma = B^\sigma_{p,\infty}(\Omega)$ ($W_p^\sigma(\Omega)$ if σ integer), $Y_\sigma = B^{m+\sigma}_{p,\infty}(\Omega)$ ($W_p^{m+\sigma}(\Omega)$ if σ integer), $\sigma \ge 0$. Let $F :\ B_{\epsilon,0} \subset Y_0 \to X_0$ be a differential mapping of order m, from $B_{\epsilon,0} = \{v \in Y_0 :\ \|v - v_0\|_{Y_0} \le \epsilon\}$ to X_0, which satisfies

(1) F is continuously Lipschitz (Fréchet) differentiable on an open set U containing $B_{\epsilon,0}$:

$$\|F'(v) - F'(w)\|_{Y_0,X_0} \le 2M\|v - w\|_{Y_0}, \qquad v, w \in B_{\epsilon,0},$$
$$\|F'(v)\|_{Y_0,X_0} \le M, \qquad v \in B_{\epsilon,0}.$$

(2) There exist numerical inversion operators $T_h(v): X_\gamma \to Y_0$, $\gamma \geq 1$, and a continuous monotone increasing function $\tau: [0,b] \to [0,\infty)$, $\tau(0) = 0$, such that, for all $w \in X_\gamma$,

$$\|[F'(v)T_h(v) - I]w\|_{X_0} \leq \tau(h)\|w\|_{X_\gamma}, \qquad \text{for all } v \in B_{1,\gamma} \subset Y_\gamma. \tag{3.2}$$

(3) $F: B_{\epsilon,0} \cap Y_\sigma \to X_\sigma$ is a well-defined map for $0 \leq \sigma \leq s$ and s sufficiently large.

(4) The maps $\{T_h(v)\}$ are uniformly bounded in h and v from X_σ to $Y_{\sigma-\gamma}$ for $\gamma \leq \sigma \leq s$:

$$\|T_h(v)w\|_{Y_{\sigma-\gamma}} \leq M\|w\|_{X_\sigma}.$$

This first part essentially is a reformulation of the hypotheses of the Newton-Kantorovich theory (see e.g. [8]). The function τ describing the approximation order of the approximate inverse T is usually of the form $\tau(h) = h^\gamma$. Since the parameter γ also enters in the norm of u on the right-hand side of (3.2) the quantity measured by γ is referred to as *loss of derivatives.* In the case of simple approximation, i.e., $m = 0$, the mapping F becomes $F(u) = u - f$, and so $F'(u) = I$. This means that (3.2) is nothing but a standard error estimate of the form

$$\|T_h(v)w - w\|_p \leq Ch^k\|w\|_{k,p}, \qquad \text{for } h \to 0,$$

common in the finite element, spline or radial basis function literature. From this one also sees that for any useful method we should have $\gamma \geq 1$. The parameter s denotes the maximal regularity of the problem.

The next set of assumptions refers to the existence of a smoothing in the sense of Hörmander (see Sect. 4 for more details).

(A2) Assume there exists a smoothing operator S_t, $t \geq 1$, satisfying

$$\begin{array}{lll} \text{(i)} & \|S_t u - u\|_p \to 0 & \text{as } t \to \infty, \\ \text{(ii)} & \|S_t u\|_{\ell,p} \leq Ct^{\ell-k}\|u\|_{k,p}, & 0 \leq k \leq \ell, \\ \text{(iii)} & \|S_t u - u\|_{\ell,p} \leq Ct^{\ell-k}\|u\|_{k,p}, & 0 \leq \ell \leq k, \end{array}$$

with an analogous formulation in Besov spaces in case nonintegral values of k and ℓ are involved.

The third block of assumptions concerns the actual iteration procedure, and tells us how to choose the scaling parameters or *smoothing speeds* for the smoothing operation:

(A3) Assume that the initial iterate $u_0 \in B_{\epsilon,0} \cap Y_s$ satisfies

$$F(u_0) \in X_s, \qquad \|F(u_0)\|_{X_0} \leq \rho^{-\lambda}, \qquad \|F(u_0)\|_{X_s} \leq M\rho^{(s-\lambda)\beta},$$

where $\rho > 0$, $1 < \lambda < s$, and $1 < \beta < 2$ is a *superlinear convergence parameter*. Define u_i by the iteration (3.1) and determine the smoothing speeds t_i using the additional *acceleration parameter* $\theta > 1$ via

$$t_i = \rho^{\theta\beta^i}, \qquad i = 1, 2, \ldots. \tag{3.3}$$

The parameters and exponents are also required to satisfy the following inequalities:

$$1 \le \gamma \le \mu < \lambda < s,$$
$$\lambda \ge \max\{\frac{2\beta\gamma}{2-\beta}, \beta(\gamma + \mu\beta)\} + \frac{\lambda_0}{2-\beta}, \qquad \lambda_0 > 0,$$
$$s \ge \max\{\frac{\theta\gamma}{\theta-1}, \lambda + \frac{\theta\gamma}{\beta-1}\} + s_0 \max\{\frac{1}{\theta-1}, \frac{1}{\beta-1}\}, \qquad s_0 > 0,$$
$$\rho^{\ln\beta} \ge e,$$
$$M^9\rho^{-\lambda_0} \le 1/4,$$
$$M^3\rho^{-s_0} \le 1/2,$$
$$M^4\rho^{-(\lambda-\beta\gamma)} \le \epsilon/2,$$
$$\rho^{\lambda_0 \ln\beta} \ge e.$$

The "meshsize" h_i has to be adaptively selected according to

$$\tau(h_i) \le M^5 \|F(u_{i-1})\|_{X_\gamma}, \qquad i = 1, 2, \ldots. \tag{3.4}$$

Inequality (3.4) is another crucial point in the theory. It shows the correlation of the mesh size to the size of the residual. This inequality is virtually impossible to satisfy precisely in practice since the size of the constant M is not known a priori (it depends on the constants in the Bernstein and Jackson inequalities (ii) and (iii) for the smoothing operation). What can be attempted in test situations for which the solution is known is to estimate M from (3.5) below and then ensure that the mesh size satisfies (3.4). Large variations in M from one iteration to the next indicate trouble.

Theorem 3.1. *If all of the above assumptions (A1)–(A3) are satisfied, then the iterative procedure (3.1) converges to a root u of F in $B_{\epsilon,0} \cap B_{1,\mu}$. The superlinear convergence in Y_0 is described by*

$$\|u - u_i\|_{Y_0} \le M^4 \frac{\rho^{-(\lambda-\beta\gamma)\beta^{i-1}}}{\beta^{i-1}}, \qquad i = 1, 2, \ldots. \tag{3.5}$$

The superlinear convergence in Y_μ is described by

$$\|u - u_i\|_{Y_\mu} \le M^5 \frac{\rho^{-\lambda_0\beta^{i-1}}}{\beta^{i-1}}, \qquad i = 1, 2, \ldots.$$

Furthermore,

$$\|u - u_i\|_{Y_0} = O(\rho_1^{\beta^{i-1}}) \approx \tau(h_i), \qquad \rho_1 = \rho^{-(\lambda-\beta\gamma)}. \tag{3.6}$$

We would like to emphasize the implications of the last statement of the theorem. This shows us that the smoothing recovers the full approximation order of the scheme used for the approximate inversion. However, there is a very delicate dependence on the meshsize h_i which has to be chosen adaptively for (3.6) to hold.

4. Hörmander's Smoothing

In the following we will use convolution with a smoothing kernel as our smoothing operator, i.e., we will consider an operator of the form $S_t u = \phi_t * u$, where ϕ is a suitably chosen kernel.

Below we require a standard lemma from functional analysis whose proof we omit.

Lemma 4.2. *Let $\phi \in L_1(\mathbb{R}^d)$ and $u \in W_p^k(\mathbb{R}^d)$ (or $u \in B_{p,\infty}^k(\mathbb{R}^d)$ for nonintegral values of k), then*

$$\|\phi * u\|_{k,p} \le C\|u\|_{k,p}.$$

The following theorem and its proof are a generalization to Sobolev and Besov spaces of an analogous theorem by Hörmander (see [6]) for Hölder spaces.

Theorem 4.3. *Let $\phi \in L_1(\mathbb{R}^d)$ be a kernel whose Fourier transform Φ is compactly supported and infinitely differentiable in $\mathbb{R}^d$ and $\Phi(\omega) = 1$ in a neighborhood of the origin. Furthermore, let $\phi_t = t^d\phi(t\cdot)$, $t \ge 1$. Then ϕ is a smoothing kernel in the sense of Theorem 3.1, i.e., for $u \in W_p^k(\mathbb{R}^d)$ ($u \in B_{p,\infty}^k(\mathbb{R}^d)$) ϕ satisfies*

$$\lim_{t\to\infty} \|\phi_t * u - u\|_p = 0, \tag{4.1}$$

$$\|\phi_t * u\|_{b,p} \le C t^{b-a}\|u\|_{a,p}, \qquad 0 \le a \le b, \tag{4.2}$$

$$\|\phi_t * u - u\|_{b,p} \le C t^{b-a}\|u\|_{a,p}, \qquad 0 \le b \le a. \tag{4.3}$$

Here the norms $\|\cdot\|_{a,p}$ and $\|\cdot\|_{b,p}$ are used to denote either Sobolev or Besov space norms depending on whether or not a, b are integers.

Proof: The semi-group property (4.1) is a standard result which holds for any Fejér-type kernel, i.e., kernel with the property $\phi_t = t^d\phi(t\cdot)$ (see e.g. [4] for a one-dimensional formulation of this result).

We now turn to the Bernstein inequality (4.2), and note that, for m a positive integer,

$$\|\phi_t * u\|_{a+m,p} = |\sum_{|\alpha|=m} (D^\alpha\phi_t) * u|_{a,p} + \|\phi_t * u\|_p. \tag{4.4}$$

Observe that

$$D^\alpha \phi_t(x) = D^\alpha(t^d\phi(tx)) = t^d t^m (D^\alpha\phi)(tx) = t^m (D^\alpha\phi)_t(x). \tag{4.5}$$

Equations (4.4) and (4.5) imply that

$$\begin{aligned} \|\phi_t * u\|_{a+m,p} &= t^m |\sum_{|\alpha|=m} (D^\alpha\phi)_t * u|_{a,p} + \|\phi_t * u\|_p \\ &\le C t^m \|u\|_{a,p}, \qquad t \ge 1, \end{aligned} \tag{4.6}$$

where the inequality follows from Lemma 4.2 and the triangle inequality since $(D^\alpha\phi)_t \in L^1(\mathbb{R}^d)$. This means (4.2) is true when $b = a + m$.

In order to obtain the validity of (4.2) in the intermediate Besov spaces also, we use the norm interpolation inequality,

$$\|u\|_{(1-\lambda)s_0+\lambda s_1,p,\infty} \le C\|u\|_{s_0,p,\infty}^{1-\lambda} \|u\|_{s_1,p,\infty}^{\lambda}, \tag{4.7}$$

where $0 \le s_0 \le s_1$ and $0 \le \lambda \le 1$ (see [13]).

For our purposes we retain the meaning of m above and let $s = (1-\lambda)a + \lambda(a+m)$. Then (4.7) yields

$$\begin{aligned} \|\phi_t * u\|_{s,p,\infty} &= \|\phi_t * u\|_{(1-\lambda)a+\lambda(a+m),p,\infty} \\ &\le C\|\phi_t * u\|_{a,p}^{1-\lambda} \|\phi_t * u\|_{a+m,p}^{\lambda}. \end{aligned}$$

By Lemma 4.2 and (4.6) the quantity above is no more than

$$C\|u\|_{a,p}^{1-\lambda} t^{m\lambda} \|u\|_{a,p}^{\lambda},$$

which in turn is equal to

$$Ct^{m\lambda}\|u\|_{a,p} = Ct^{s-a}\|u\|_{a,p},$$

since $\lambda m = s - a$ by assumption. Therefore (4.2) is also true for intermediate values $b = s$, $a \le s \le a + m$.

Since (4.3) for $b = a$ follows immediately from an application of the triangle inequality and (4.2), it suffices to prove (4.3) when $b = 0$ (the remaining cases are again filled by norm interpolation).

First we consider (4.3) for $a = s$, $0 \le s \le 1$, i.e., we show for s in this range,

$$\|u - \phi_t * u\|_p \le Ct^{-s}\|u\|_{s,p,\infty}.$$

This follows from the computations below in which we have made use of the definition of the norms involved as well as the assumption that $\Phi(0) = \hat{\phi}(0) = \int_{\mathbb{R}^d} \phi(y)dy = 1$, and the scaling property $\phi_t = t^d\phi(t\cdot)$.

$$
\begin{aligned}
\|u - \phi_t * u\|_p^p &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} [u(x) - u(x-y)]\, \phi_t(y)dy \right|^p dx \\
&\leq \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \frac{|u(x) - u(x-y)|}{|x-(x-y)|^s} |y|^s |\phi_t(y)| dy \right\}^p dx \\
&\leq \int_{\mathbb{R}^d} \sup_{h\in\mathbb{R}^d} \left\{ \frac{|u(x+h) - u(x)|}{|h|^s} \right\}^p dx \left\{ \int_{\mathbb{R}^d} |y|^s |\phi_t(y)| dy \right\}^p \\
&\leq \|u\|_{s,p,\infty}^p \left\{ \int_{\mathbb{R}^d} |y|^s |\phi_t(y)| dy \right\}^p \\
&= \|u\|_{s,p,\infty}^p \left\{ \int_{\mathbb{R}^d} \frac{|y|^s}{t^s} |\phi(y)| dy \right\}^p \\
&\leq \left\{ Ct^{-s} \|u\|_{s,p,\infty} \right\}^p .
\end{aligned}
$$

The special case of (4.3) now follows by taking p-th roots.

The general case $k < s \leq k+1$ can be done similarly, but requires some additional preparations. First we recall the assumptions on the Fourier transform of ϕ, namely $\Phi \in C_0^\infty(\mathbb{R}^d)$ and $\Phi(\omega) = 1$ for $\omega \in B_\epsilon(0)$, a ball with radius ϵ centered at the origin. Next we define a function Ψ by

$$
\Psi(\omega) = \frac{\omega^\alpha (1 - \Phi(\omega))}{\sum_{|\beta|=k} (\omega^\beta)^2}, \qquad |\alpha| = k,
$$

and note that

$$
1 - \Phi(\omega) = \sum_{|\alpha|=k} \omega^\alpha \Psi(\omega). \tag{4.8}
$$

One can easily check that Ψ is zero in a neighborhood of the origin, and that it is homogeneous of degree $-k$ near ∞. This ensures that Ψ is itself a Fourier transform, say $\Psi = \hat{\psi}$. The function ψ is rapidly decreasing at ∞ and has a singularity at most like $|x|^{k-d} \log |x|$ near the origin. The important consequence for us is that ψ is integrable.

We can apply the Fourier transform to $u(x) - \phi_t * u(x)$ and use (4.8) to obtain

$$\begin{aligned}
[u(x) - \phi_t * u(x)]^\wedge &= \hat{u}(\omega) - \hat{\phi}(\frac{\omega}{t})\hat{u}(\omega) \\
&= \hat{u}(\omega)\left[1 - \hat{\phi}(\frac{\omega}{t})\right] \\
&= \hat{u}(\omega)\left[\sum_{|\alpha|=k} (\frac{\omega}{t})^\alpha \Psi(\frac{\omega}{t})\right] \\
&= \sum_{|\alpha|=k} (\frac{\omega}{t})^\alpha \Psi(\frac{\omega}{t})\hat{u}(\omega).
\end{aligned}$$

Next we apply the inverse Fourier transform to our last result. This leads to

$$\begin{aligned}
\left[\sum_{|\alpha|=k} (\frac{\omega}{t})^\alpha \Psi(\frac{\omega}{t})\hat{u}(\omega)\right]^\vee &= \sum_{|\alpha|=k} \left[\Psi(\frac{\omega}{t})\right]^\vee * \left[(\frac{\omega}{t})^\alpha \hat{u}(\omega)\right]^\vee \\
&= \sum_{|\alpha|=k} \left[\Psi(\frac{\omega}{t})\right]^\vee * \left[t^{-|\alpha|}(\omega^\alpha \hat{u}(\omega))\right]^\vee \\
&= \sum_{|\alpha|=k} \psi_t * \left[t^{-k}(-i\partial)^\alpha u\right](x)
\end{aligned}$$

In order for the derivatives of u up to order $k-1$ to vanish at $\pm\infty$ in the integration by parts computation above, we need to assume that $u \in C_0^\infty(\mathbb{R}^d)$. However, since $C_0^\infty(\mathbb{R}^d)$ is dense in both the Sobolev and Besov spaces being used, we can do this without loss of generality.

Therefore, using the abbreviation $D = -i\partial$, we obtain

$$u(x) - \phi_t * u(x) = t^{-k} \sum_{|\alpha|=k} \psi_t * D^\alpha u(x).$$

Since

$$\int_{\mathbb{R}^d} \psi_t(y)dy = \int_{\mathbb{R}^d} t^d \psi(ty)dy = \int_{\mathbb{R}^d} \psi(y)dy = \Psi(0) = 0,$$

we can use the definition of the convolution to rewrite $u(x) - \phi_t * u(x)$ in a form amenable to calculations analogous to those performed in the case $0 \leq s \leq 1$ above, i.e.,

$$u(x) - \phi_t * u(x) = t^{-k} \sum_{|\alpha|=k} \int_{\mathbb{R}^d} \left[D^\alpha u(x-y) - D^\alpha u(x)\right] \psi_t(y)dy.$$

The Jackson inequality now follows by using the norms in higher order intermediate spaces in the same manner as before. ■

Remark. We point out that the transition from the bounded domain Ω in Section 3 to the formulation on $\mathbb{R}^d$ used in Theorem 4.3 is provided by the Calderon extension theorem, which guarantees the existence of a continuous extension operator from the Sobolev (or Besov) space defined on Ω to that defined on $\mathbb{R}^d$.

An example of a kernel satisfying all the hypotheses of Theorem 4.3 was given in [7], and we repeat it here for the reader's convenience. We choose

$$\Phi(\omega) = \begin{cases} 1, & \|\omega\| \leq 1/2, \\ 1 - e^{-\frac{1}{\tan^2 \|\pi\omega/2\| - 1}}, & 1/2 < \|\omega\| < 1, \\ 0, & \|\omega\| \geq 1. \end{cases} \tag{4.9}$$

It is easily checked that $\Phi \in C_0^\infty(\mathbb{R}^d)$, and $\Phi \equiv 1$ in $B_{1/2}(0)$. However, the actual kernel ϕ cannot be computed explicitly for this example.

Remark. Since $S_t u$ is band limited, i.e.,

$$\operatorname{supp}(\widehat{S_t u}) \subset t[-1, 1]^d,$$

one can compute the smoothing in the Fourier domain. We have performed some preliminary tests using this idea. However, considerably more work is needed.

Since the kernel (4.9) is extremely oscillatory, which causes computational difficulties, we have based the numerical examples presented in this paper on the well-known Gauss-Weierstrass kernel, i.e.,

$$\phi(x) = \frac{1}{(4\pi)^{d/2}} e^{-\frac{\|x\|^2}{4}}, \tag{4.10}$$

and the scaled version

$$\phi_t(x) = t^d \phi(tx) = \frac{t^d}{(4\pi)^{d/2}} e^{-t^2 \frac{\|x\|^2}{4}}.$$

We point out that the Fourier transform of the Gauss-Weierstrass kernel satisfies $\Phi(\omega) = 1$ only for $\omega = 0$, and Φ is infinitely differentiable but not compactly supported. Moreover, the Gauss-Weierstrass kernel (as a positive kernel) is known to be saturated, i.e., the Jackson inequality in its form (4.3) does not hold. In fact, we have (see [3], cf. also Korovkin's theorem (e.g., [9]) for positive polynomial operators acting on continuous functions)

$$\|S_t u - u\|_p = o(t^{-2}).$$

Consequently, our numerical experiments do not completely satisfy the theory.

Another class of kernels we have considered are compactly supported on $\mathbb{R}^d$, and also satisfy the condition $\Phi \equiv 1$ in $B_\epsilon(0)$. But they are only of limited regularity. This class of kernels is related to the sinc function. For $d = 1$, one could use

$$\Phi(\omega) = \begin{cases} 1 & |\omega| \leq 1, \\ 0 & \text{otherwise}, \end{cases}$$

which is a C^{-1} function, and leads to the sinc function kernel,

$$\phi(x) = \frac{\sin(x)}{\pi x}.$$

If we desire higher regularity, we can define the continuous function,

$$\Phi(\omega) = \begin{cases} 1 & |\omega| \leq 1/2, \\ 2(1-\omega) & 1/2 < |\omega| \leq 1, \\ 0 & \text{otherwise}, \end{cases}$$

which leads to

$$\phi(x) = 2\frac{\cos(\frac{x}{2}) - \cos(x)}{\pi x^2},$$

or

$$\Phi(\omega) = \begin{cases} 1 & |\omega| \leq 1/2, \\ 4(4\omega^3 - 9\omega^2 + 6\omega - 1) & 1/2 < |\omega| \leq 1, \\ 0 & \text{otherwise}, \end{cases}$$

which is C^1 and leads to

$$\phi(x) = 24\frac{4\cos(\frac{x}{2}) - 4\cos(x) - x\sin(x) - x\sin(\frac{x}{2})}{\pi x^4}.$$

Kernels with more regularity can easily be constructed using Hermite interpolation. However, all of these kernels are also extremely oscillatory and lead to the same kind of numerical problems as the kernel defined in (4.9).

Remark. We note, however, that it is precisely this computationally troublesome oscillatory property of the kernel (which stems from the fact $\Phi \equiv 1$ in $B_\epsilon(0)$) that is theoretically necessary to obtain the required Jackson inequality for the smoothing. As noted above, positive kernels are saturated.

5. Numerical Experiments

5.1. Pure Approximation

We start by giving two examples of (scattered) data approximation. In both of the examples the data is obtained by sampling a given test function. For ease of programming, and also to have a more accurate estimate for convergence rates, we have chosen to sample the test function on a regular set of points: equally spaced points in the unit interval/square for one/two-dimensional approximation, respectively. However, the grid points need not be equally spaced in general. In fact, (3.4) suggests that the mesh size be chosen adaptively.

In both the one-dimensional and the two-dimensional example we chose the locally supported radial basis functions $\varphi(r) = (1-r)_+^4(4r+1)$ of Wendland for the approximate inversion T. (The reader can easily convince himself that the use of globally supported basis functions in a multilevel framework is pointless.) In [15] one can find the following error estimate for these functions:

$$\|T_{\mathcal{X}}(u)f - f\|_\infty \leq Ch^{\frac{3}{2}}\|f\|_{\frac{d+3}{2},2},$$

which suggests that the "loss of derivative" $\gamma = 3/2$ (see (3.2)). The test function for the 1D-case is

$$f(x) = 15e^{\frac{-1}{1-4(x-1/2)^2}}\left[\frac{3}{4}e^{-\frac{(9x-2)^2}{4}} + \frac{3}{4}e^{-\frac{(9x+1)^2}{49}} + \frac{1}{2}e^{-\frac{(9x-7)^2}{4}} - \frac{1}{5}e^{-(9x-4)^2}\right].$$

We have chosen the support of the basis functions in the first step to contain 3 grid points, and made the simplifying assumption that the scales δ_i in (2.1) change linearly with the data separation. This latter assumption is not in complete agreement with the theory. If one assumes $\delta_i = \kappa h_i$, then using an estimate of [12] we get $\tau(h_i) = C(\frac{h_i}{\delta_i})^\gamma$. Since on the other hand $\tau(h_i) \leq M^7\rho^{-(\lambda-\beta\gamma)\beta^{i-1}}$ (which follows from (3.4) and equation (3.1c) of [7]), we see that

$$\tau(h_i) = C(\frac{h_i}{\delta_i})^\gamma = C(\frac{1}{\kappa})^\gamma \leq M^7\rho^{-(\lambda-\beta\gamma)\beta^{i-1}},$$

which might cease to be true for large values of i. This implies that, theoretically, the scaling of the basis functions needs to be performed adaptively, also.

The smoothing operation was done by using a localized discretization of the Gauss-Weierstrass kernel (4.10). Specifically, the kernel was assumed to be zero if its value reached less than 10^{-8}. This means that the number of points on which the discrete convolution is based depends on the meshsize of the grid. In Tables

5.1 and 5.2 we list the ℓ_∞ and ℓ_2 errors along with approximate convergence rates which are computed via

$$\text{rate} = \log\left(\frac{\|F(u_{i-1})\|}{\|F(u_i)\|}\right) / \log 2.$$

The tables contain columns for the case of multistep approximation (actually interpolation) without smoothing, with smoothing using parameters $\beta = 1.1$, $\theta = 1.3$, $\rho = 13$, and using parameters $\beta = 1.2$, $\theta = 1.2$, $\rho = 3$. Recall that these parameters together determine the smoothing speeds t_i via $t_i = \rho^{\theta\beta^i}$ (see (3.3)). Table 5.2 also has two columns in which we have listed the a posteriori estimate for the constant M in Theorem 3.1 obtained via (3.5). When computing the smoothed values it is necessary to extend these values in some way. In both of the following approximation examples we simply extended the test function as a constant outside the domain.

	no smoothing		smoothing 1		smoothing 2	
# pts	ℓ_∞	rate	ℓ_∞	rate	ℓ_∞	rate
3	4.03455e+00		4.01966e+00		4.67686e+00	
5	4.30036e-01	3.230	6.35823e-01	2.660	3.70157e+00	0.337
9	4.33761e-01	-0.012	5.97599e-01	0.089	3.09263e+00	0.259
17	7.48187e-02	2.535	1.94018e-01	1.623	2.44893e+00	0.337
33	2.29188e-02	1.707	5.39304e-02	1.847	1.61769e+00	0.598
65	1.00899e-02	1.184	1.20040e-02	2.168	6.74170e-01	1.263
129	4.19097e-03	1.268	2.15461e-03	2.478	2.34738e-01	1.522
257	1.82722e-03	1.198	8.54868e-04	1.334	4.48747e-02	2.387
513	7.89726e-04	1.210	3.22799e-04	1.405	6.89169e-03	2.703
1025	3.42366e-04	1.206	1.06968e-04	1.593	1.00642e-03	2.776
2049	1.48368e-04	1.206	3.74159e-05	1.515	9.91734e-05	3.343
4097	6.43107e-05	1.206	8.54714e-06	2.130	1.91653e-05	2.371
8193	2.78753e-05	1.206	3.25218e-06	1.394	1.11180e-05	0.786
16385	1.20821e-05	1.206	1.96607e-06	0.726	4.69820e-06	1.243

Table 5.1. ℓ_∞ errors for 1D-approximation.

Although the beneficial effects of the smoothing can be observed in the middle parts of both tables, one can also clearly observe deterioration in the errors in the later stages of the algorithm with smoothing. This phenomenon probably is attributable to the saturation of the Gauss-Weierstrass kernel as mentioned earlier. The large increase for the values of M in the lower part of Table 5.2 also indicate this problem.

	no smoothing		smoothing 1			smoothing 2		
# pts	ℓ_2	rate	ℓ_2	rate	M	ℓ_2	rate	M
3	3.75574e+00		3.72777e+00		144	6.43290e+00		21
5	5.13986e-02	6.191	1.44411e-01	4.690	80	3.39290e+00	0.923	27
9	1.94227e-02	1.404	5.04490e-02	1.517	113	2.20696e+00	0.620	46
17	8.77366e-04	4.468	3.77285e-03	3.741	110	1.16000e+00	0.928	80
33	5.17874e-05	4.083	2.23363e-04	4.078	96	5.67597e-01	1.031	130
65	3.32310e-06	3.962	6.88844e-06	5.019	82	1.28261e-01	2.146	206
129	2.20018e-07	3.917	1.82805e-07	5.236	76	7.49722e-03	4.097	333
257	1.60017e-08	3.781	7.47756e-09	4.612	80	1.63265e-04	5.521	614
513	1.24167e-09	3.688	6.04847e-10	3.628	103	2.45557e-06	6.055	1590
1025	1.02849e-10	3.594	1.34838e-10	2.165	242	4.40398e-08	5.801	6509
2049	8.90893e-12	3.529	7.52027e-11	0.842	576	7.99693e-10	5.783	64627
4097	7.96594e-13	3.483	8.13921e-12	3.208	1084	1.75030e-11	5.514	1664822
8193	7.26505e-14	3.455	2.42981e-12	1.744	2653	1.96715e-12	3.153	61119135
16385	6.70740e-15	3.437	8.24974e-13	1.558	8277	6.19595e-13	1.667	5910162602

Table 5.2. ℓ_2 errors for 1D-approximation.

The difference in the second example is that we choose a bivariate test function,

$$f(x, y) = 1 + \tanh(9x + 9y),$$

and the supports of the basis functions for the approximate inversion initially contain 9 points. The parameters for the smoothing are $\beta = 1.3$, $\theta = 1.2$, and $\rho = 3$. Tables 5.3 and 5.4 list the corresponding errors and rates. Table 5.4 again contains estimates for M.

	no smoothing		smoothing	
grid	ℓ_∞	rate	ℓ_∞	rate
3×3	7.033255e-01		5.422649e-01	
5×5	4.212533e-01	0.740	3.126155e-01	0.795
9×9	2.701228e-01	0.641	2.283325e-01	0.453
17×17	1.348346e-01	1.002	1.085956e-01	1.072
33×33	5.837676e-02	1.208	2.355385e-02	2.205
65×65	2.532966e-02	1.205	9.004403e-03	1.387

Table 5.3. ℓ_∞ errors for 2D-approximation.

In this example we cannot detect the same kind of deterioration as in the previous example. The errors and convergence rates are better once the smoothing is added in the algorithm. Figures 5.1–5.6 support the numerical data for the

	no smoothing		smoothing		
grid	ℓ_2	rate	ℓ_2	rate	M
3×3	1.074078e-01		7.736727e-02		11
5×5	7.756864e-03	3.791	7.340859e-03	3.398	16
9×9	8.123557e-04	3.255	8.700502e-04	3.077	32
17×17	8.655500e-05	3.230	6.689845e-05	3.701	84
33×33	8.204931e-06	3.399	2.853099e-06	4.551	297
65×65	7.334848e-07	3.484	1.353560e-07	4.398	1965

Table 5.4. ℓ_2 errors for 2D-approximation.

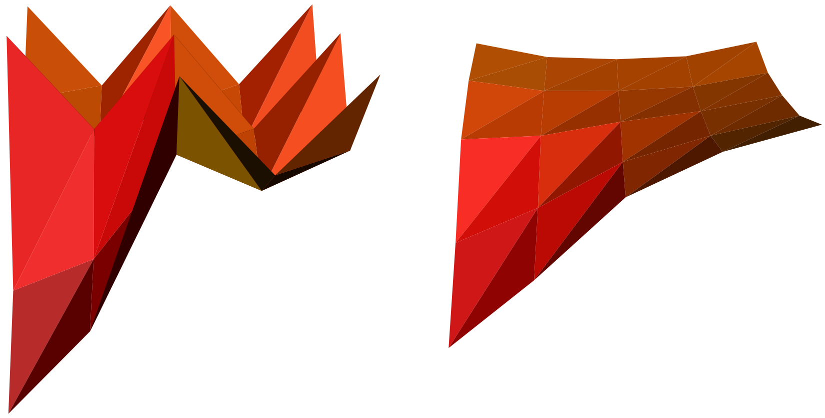

Fig. 5.1. Fit 1 unsmoothed and smoothed.

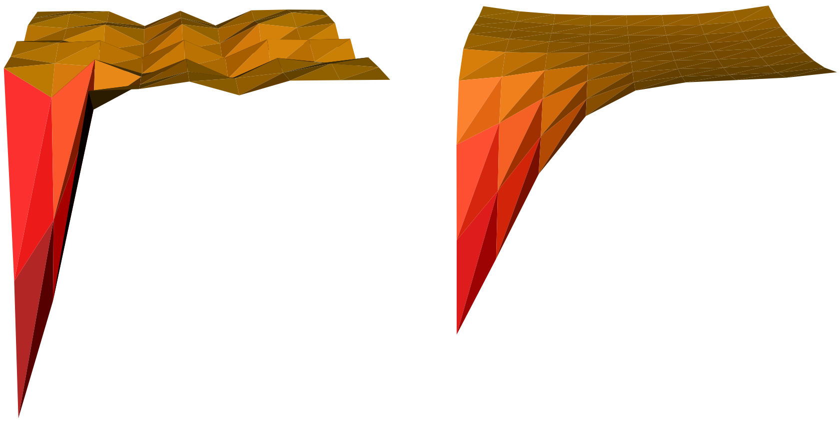

Fig. 5.2. Fit 2 unsmoothed and smoothed.

bivariate example. The six figures correspond to the six grids listed in Tables 5.3 and 5.4. The smoothing effect is clearly visible.

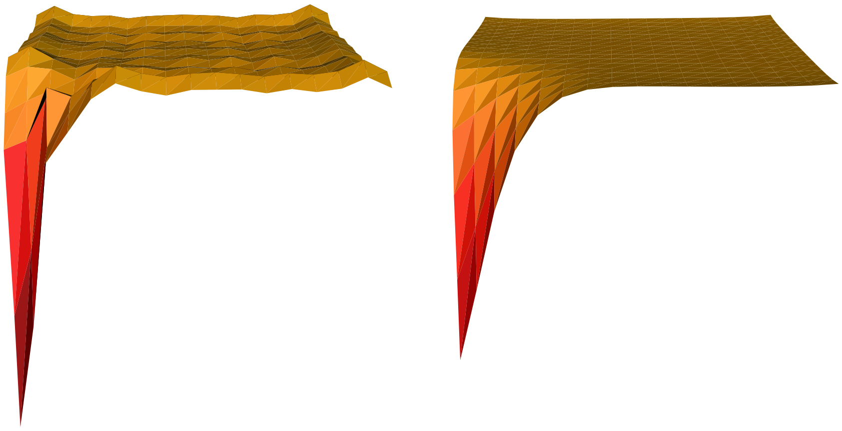

Fig. 5.3. Fit 3 unsmoothed and smoothed.

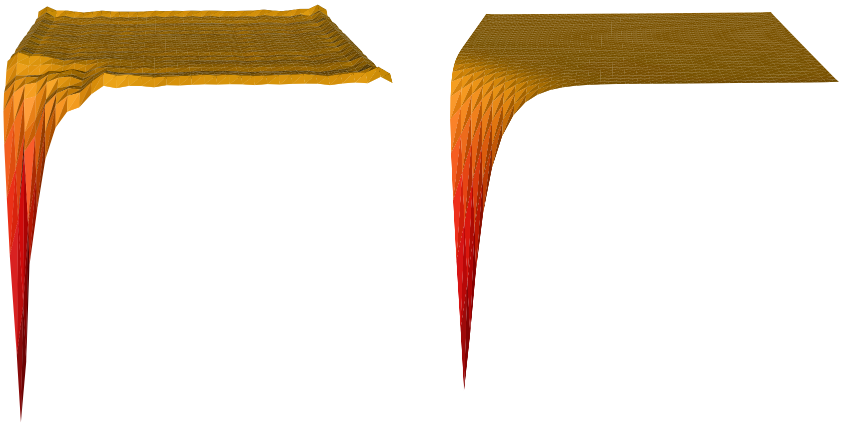

Fig. 5.4. Fit 4 unsmoothed and smoothed.

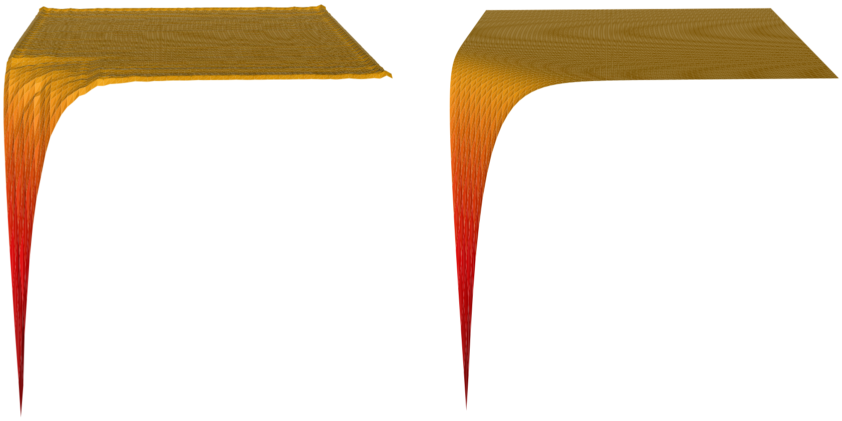

Fig. 5.5. Fit 5 unsmoothed and smoothed.

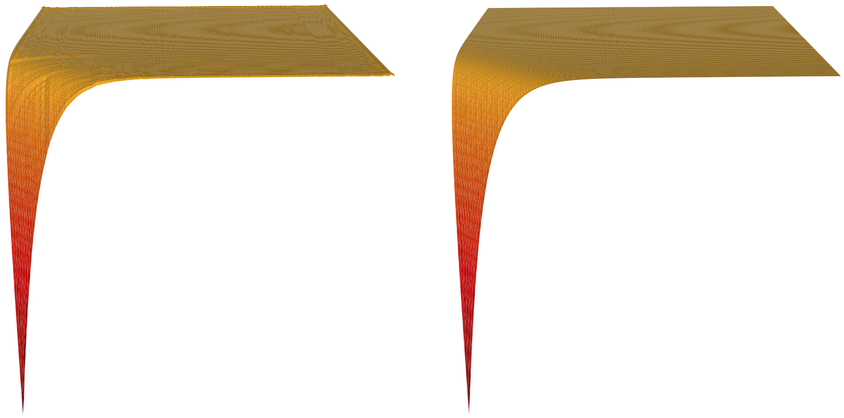

Fig. 5.6. Fit 6 unsmoothed and smoothed.

5.2. Approximate Solution of a Boundary Value Problem

As mentioned in Section 2, we also tested the performance of the multilevel algorithm, with and without smoothing, on a simple two-point boundary value problem. In order to be able to verify our solutions easily, we chose

$$\begin{aligned} -u''(x) + \pi^2 u(x) &= 2\pi^2 \sin \pi x, \qquad x \in (0,1), \\ u(0) = u(1) &= 0, \end{aligned}$$

which has solution $u(x) = \sin \pi x$. As approximate inversion we chose the method of finite differences. In order to pass from one grid to the next finer one, we used linear interpolation of the finite difference values. As can be seen from (2.5), it is now necessary to differentiate the intermediate solutions s_i. In order to do that, we performed an interpolation step using globally supported multiquadrics (hence the low number of points used) to obtain a representation for subsequent differentiation. The smoothing was again done with the Gauss-Weierstrass kernel, and the parameters were $\beta = 1.4$, $\theta = 1.2$, $\rho = 3$. The extension of the data in this example was done periodically, since it is known that the solution will be periodic. The same extension was also used in the interpolation step in order to avoid otherwise very large oscillations at the endpoints of the interval (which have nothing to do with the smoothing, but are solely caused by the interpolation step).

Remark. Tests with locally supported radial basis functions instead of multiquadrics in the interpolation/differentiation step were unsuccessful unless the support was taken very large (=global). This is due to the effects of the chain rule on the scaling of the supports.

Tables 5.5–5.8 list the ℓ_∞ and ℓ_2 errors of the approximate solution, i.e.,

$$\|u - u_i\|_{\infty(2)} \qquad \text{on } \mathcal{X}_{i+1},$$

	no smoothing		smoothing	
# pts	ℓ_∞	rate	ℓ_∞	rate
5	4.827905e-02		6.992085e-02	
9	5.551833e-02	-0.202	1.057068e-02	2.726
17	3.179490e-02	0.804	2.506128e-03	2.077
33	1.740975e-02	0.869	6.591131e-04	1.927
65	8.863612e-03	0.974	1.595434e-04	2.047
129	4.384276e-03	1.016	1.764670e-05	3.176

Table 5.5. ℓ_∞ errors for approximate solution.

	no smoothing		smoothing	
# pts	ℓ_2	rate	ℓ_2	rate
5	7.551016e-04		2.312696e-03	
9	1.177486e-03	-0.641	4.521613e-05	5.677
17	3.437134e-04	1.776	3.103748e-06	3.865
33	7.367566e-05	2.222	1.732496e-07	4.163
65	1.625273e-05	2.181	9.601066e-09	4.174
129	3.793412e-06	2.099	1.096157e-10	6.453

Table 5.6. ℓ_2 errors for approximate solution.

	no smoothing		smoothing	
# pts	ℓ_∞	rate	ℓ_∞	rate
5	1.927824e+01		2.403932e+00	
9	4.236021e+01	-1.136	6.063165e-01	1.987
17	5.552220e+01	-0.390	4.524052e-01	0.422
33	9.432692e+01	-0.765	7.833539e-01	-0.792
65	1.953632e+02	-1.050	8.658790e-01	-0.145
129	3.820763e+02	-0.968	9.984408e-01	-0.206

Table 5.7. ℓ_∞ errors for residuals.

and of the residual

$$\|f - L[u_i]\|_{\infty(2)} = \|F(u_i)\|_{\infty(2)} \qquad \text{on } \mathcal{X}_{i+1}.$$

Figures 5.7–5.12 depict the residuals corresponding to the six iterations listed in Table 5.7. Since the scales on the figures are difficult to read, we included the graph of the original right-hand side function $2\pi^2 \sin \pi x$ in each of the figures as a reference.

	no smoothing		smoothing		
# pts	ℓ_2	rate	ℓ_2	rate	M
5	1.487833e+02		1.422781e+00		5
9	3.341058e+02	-1.167	1.019696e-01	3.803	6
17	6.223095e+02	-0.897	4.925842e-02	1.050	19
33	9.976837e+02	-0.681	1.019520e-01	-1.049	108
65	1.499176e+03	-0.588	9.374004e-02	0.121	1590
129	2.317924e+03	-0.629	1.444180e-01	-0.624	59747

Table 5.8. ℓ_2 errors for residuals.

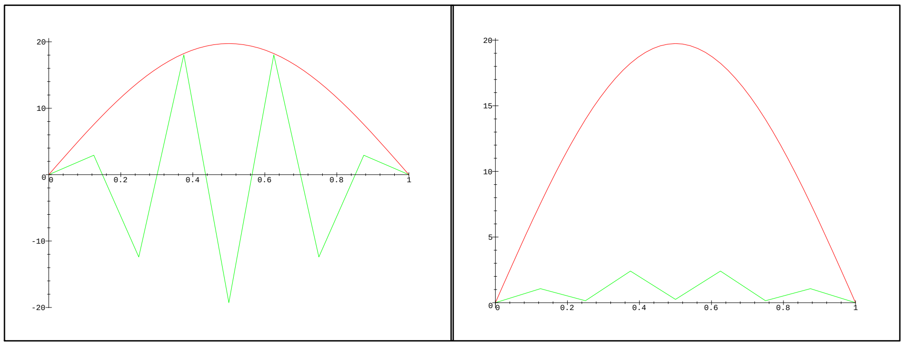

Fig. 5.7. Residual 1 unsmoothed and smoothed.

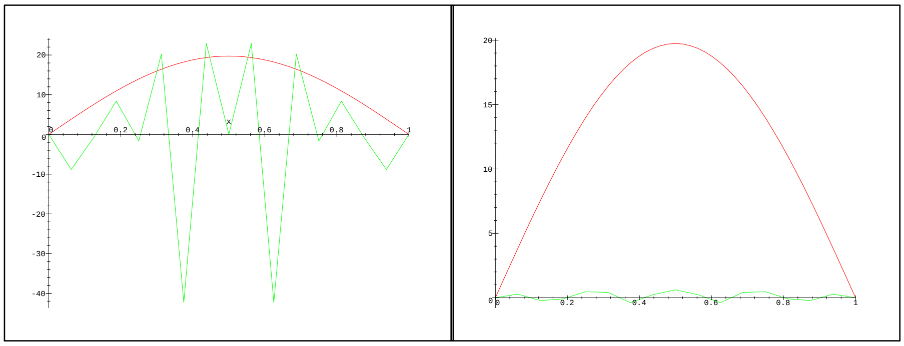

Fig. 5.8. Residual 2 unsmoothed and smoothed.

6. Concluding Remarks

We have generalized results of Hörmander and Jerome to Sobolev and Besov spaces. The theory provides convergence results for certain multilevel approximation algorithms which seem to enjoy a considerable amount of popularity. The established convergence rate is superlinear if a smoothing operation is included in the algorithm as a postconditioner. Our numerical experiments suggest the power

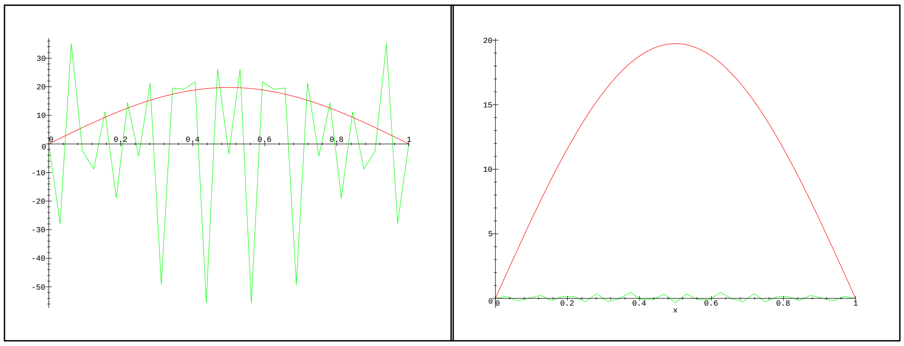

Fig. 5.9. Residual 3 unsmoothed and smoothed.

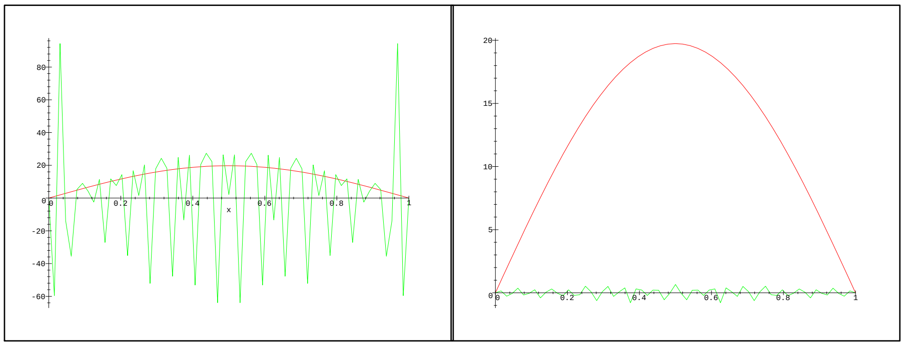

Fig. 5.10. Residual 4 unsmoothed and smoothed.

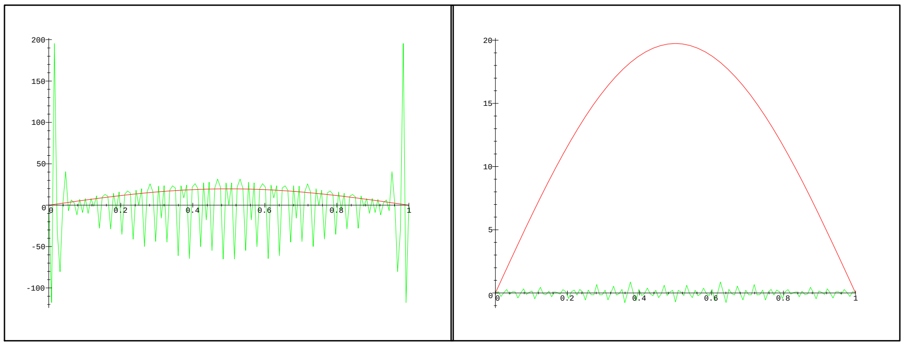

Fig. 5.11. Residual 5 unsmoothed and smoothed.

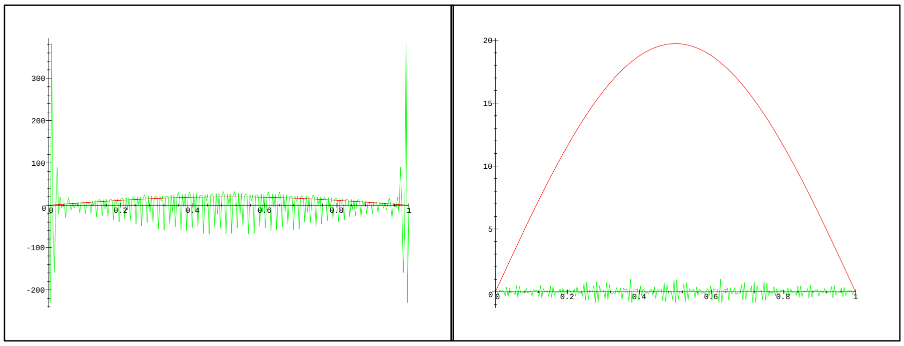

Fig. 5.12. Residual 6 unsmoothed and smoothed.

of the theory, especially for the iterative solution of differential equations. The smoothing was computed using the Gauss-Weierstrass kernel which, as pointed out earlier, does not completely match the requirements of the theory we presented. This is justified, however, by the fact that efficient and accurate computation with oscillatory kernels is highly nontrivial. We plan an efficient implementation of the smoothing based on oscillatory kernels in the Fourier domain in the future. Furthermore, as can be seen from the hypotheses of Theorem 3.1, the theory requires that the solution have a fairly high degree of regularity. We hope to relax these requirements in a future paper. Other work will focus on nonlinear differential equations.

Acknowledgments. We would like to thank S. Hartmann and R. Schaback for pointing out the fact that the scaling of the locally supported radial basis functions used in the approximation examples might have to be done adaptively.

References

1. Adams, R. A., *Sobolev Spaces,* Academic Press, New York, 1975.
2. Brand, R., W. Freeden, and J. Fröhlich, An adaptive hierarchical approximation method on the sphere using axisymmetric locally supported basis functions, Computing **57** (1996), 187–212.
3. Butzer, P. L. and H. Berens, *Semi-Groups of Operators and Approximation, Grundlehren der mathematischen Wissenschaften, Vol.145,* Springer, Heidelberg, 1967.
4. Butzer, P. L. and R. J. Nessel, *Fourier Analysis and Approximation: Vol.I, One-dimensional theory,* Birkhäuser, Basel, 1971.
5. Floater, M. S. and A. Iske, Multistep scattered data interpolation using compactly supported radial basis functions, J. Comput. Applied Math. **73** (1996), 65–78.
6. Hörmander, L., The boundary problems of physical geodesy, Arch. Ration. Mech. Anal. **62** (1976), 1–52.
7. Jerome, J. W., An adaptive Newton algorithm based on numerical inversion: regularization as postconditioner, Numer. Math. **47** (1985), 123–138.
8. Kantorovich, L. and G. Akilov, *Functional Analysis in Normed Spaces,* Pergamon Press, New York, 1964.
9. Lorentz, G. G., *Approximation of Functions,* Holt, Rinehart and Winston, New York, 1966.
10. Moser, J., A rapidly convergent iteration method and nonlinear partial differential equations I., Ann. Scuola Norm. Pisa **XX** (1966), 265–315.

11. Narcowich, F. J., R. Schaback, and J. D. Ward, Multilevel interpolation and approximation, preprint, 1997.

12. Schaback, R., On the efficiency of interpolation by radial basis functions, in *Surface Fitting and Multiresolution Methods,* A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds), Vanderbilt University Press, Nashville TN, 1997, 309–318.

13. Triebel, H., *Spaces of Besov-Hardy-Sobolev Type,* Teubner, Leipzig, 1978.

14. Wendland, H., Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree, Adv. in Comput. Math. **4** (1995), 389–396.

15. Wendland, H., Error estimates for interpolation by compactly supported radial basis functions of minimal degree, Universität Göttingen, 1996.

Gregory E. Fasshauer
Department of Computer Science and Applied Mathematics
Illinois Institute of Technology
Chicago, IL 60616
`fass@amadeus.csam.iit.edu`

Joseph W. Jerome
Department of Mathematics
Northwestern University
Evanston, IL 60208
`jwj@math.nwu.edu`