Spectrally Optimized Derivative Formulae

Robert Schaback

April 10, 2008

Abstract

For various numerical methods, it is necessary to have good differentiation formulae for special classes of functions that are defined by their spectral properties. This note shows how to calculate optimal formulae for derivatives from scattered multivariate data under three different but similar assumptions on the spectrum of the functions to be differentiated. Optimal formulae turn out to be necessarily based on certain kernels arising as reproducing kernels for the spaces in question,, and in case of function spaces with Euclidean spectral invariance, like in Sobolev spaces, the optimal formulae are necessarily provided via radial basis functions. Special emphasis is given to the widely neglected issue of scaling, and some examples are provided.

1 Introduction

We consider functions on $\mathbb{R}^d$ whose spectral composition is known, i.e. via the inverse Fourier transform we have

$$f(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{f}(\omega) \exp(i\omega \cdot x) d\omega \text{ for all } x \in \mathbb{R}^d.$$

We want to derive approximate differentiation formulae of the *stencil* form

$$D^\alpha f(x) \approx \sum_{z \in Z} c(\alpha, x, z) f(z), \tag{1}$$

i.e. they should evaluate a multivariate derivative of f at some fixed point $x \in \mathbb{R}^d$ using evaluations of f on points z of a fixed finite set $Z \subset \mathbb{R}^d$ of scattered data locations. Note that regularly structured and univariate data are admitted, but the focus is on the general multivariate scattered case. The resulting spectrally optimized formulae, however, will in a certain specific sense be slightly superior to classical ones.

The error of a formula (1) has the representation

$$\begin{array}{rcl} & & D^\alpha f(x) - \sum_{z\in Z} c(\alpha, x, z) f(z) \\ & = & (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{f}(\omega) \left((i\omega)^\alpha e^{i\omega\cdot x} - \sum_{z\in Z} c(\alpha, x, z) e^{i\omega\cdot z} \right) d\omega \end{array} \tag{2}$$

if the function f is smooth enough to make the integral feasible. Introducing the error functional

$$\begin{array}{rcl} \lambda(f) & := & D^\alpha f(x) - \sum_{z\in Z} c(\alpha, x, z) f(z) \\ & = & \left(\delta_x^\alpha - \sum_{z\in Z} c(\alpha, x, z)\delta_z \right) f \end{array}$$

we see that the overall error in (2) is a weighted mean over all the errors of the functions $\exp(i\omega \cdot x)$, since

$$\begin{array}{rcl} & & D^\alpha f(x) - \sum_{z\in Z} c(\alpha, x, z) f(z) \\ & = & (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{f}(\omega) \lambda^x (e^{i\omega\cdot x}) d\omega \end{array} \tag{3}$$

where the upper index at λ^x indicates that λ acts with respect to the variable x. The function

$$h_\lambda(\omega) := \lambda^x(e^{i\omega\cdot x}) =: \hat{\lambda}(\omega)$$

describes the spectral distribution of the error (it is its distributional Fourier transform). It can be plotted for each differentiation formula, and users can draw conclusions from these plots. But without additional clearly defined criteria, graphical inspection of h_λ is more or less speculative. Thus we have to formulate criteria.

In many cases, users want a small error for functions whose spectrum lies close to the origin. In extreme cases, one might thus want that h_λ and several derivatives $D^\beta h_\lambda$ vanish at zero. By distribution theory, this is the same as requesting that the formula is exact for the polynomials $f(x) = x^\beta$ for the β considered above. Any extended Fourier–based analysis is not necessary, if the Fourier transform at zero plays such a dominant part. Furthermore, in the multivariate case it is not clear which polynomials should be reproduced. Thus we drop polynomial reproduction as a criterion for assessing the quality of formulae for multivariate derivatives, but we will have to come back to it later.

For now, we strongly emphasize that looking at the function h_λ does not mean anything unless one specifies clearly to which class of functions the differentiation formula should be applied.

2 Optimality Criteria

One should ask for formulae which are optimal for all functions f from a certain class of functions with a prescribed spectral behavior. Unfortunately, there are several ways to formalize this, and some cases do not lead to results that can be calculated or even characterized.

Since solutions of partial differential equations usually lie in Sobolev spaces, it makes perfect sense to use the latter. These spaces take the form

$$W_2^k(\mathbb{R}^d) := \left\{ f \;:\; \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 (1 + \|\omega\|_2^2)^k d\omega \le \infty \right\}$$

and are Hilbert spaces under the inner product

$$(f, g)_k := \int_{\mathbb{R}^d} \hat{f}(\omega)\overline{\hat{g}(\omega)}(1 + \|\omega\|_2^2)^k d\omega$$

which penalizes high frequencies. Users who want to find derivative formulae for functions solving partial differential equations are strongly encouraged to use these function classes, because they often host the solution.

The error functional λ of any stencil should be a continuous functional on the Sobolev space. Then it has a norm $\|\lambda\|_k$ in the dual space satisfying

$$|\lambda(f)| \le \|\lambda\|_k \|f\|_k \text{ for all } f \in W_2^k(\mathbb{R}^d),$$

which is an error bound, and finding an optimal stencil amounts to finding a stencil with minimal norm of its error functional $\|\lambda\|_k$.

Since $W_2^k(\mathbb{R}^d)$ is a Hilbert space, the above technique generalizes to any Hilbert space H of functions. If the error functional is continuous, it has a dual norm, and an optimal stencil is one with minimal norm of its error functional. Specific Hilbert spaces arise when the weight function $(1 + \|\omega\|_2^2)^k$ of Sobolev space is replaced by another weight function $W(\omega)$ introducing another penalty for certain frequencies. For instance, the choice

$$W(\omega) = \left\{ \begin{array}{ll} 1 & \|\omega\| \leq C \\ \infty & \|\omega\| > C \end{array} \right.$$

leads to band-limited functions.

Other users may argue that they want to focus on functions f with spectral behavior

$$|\hat{f}(\omega)| \leq T(\omega) \text{ for all } \omega \in \mathbb{R}^d$$

bounded by a nonnegative threshold function T. Then one has the bound

$$|\lambda(f)| \leq (2\pi)^{-d/2} \int_{\mathbb{R}^d} T(\omega) |\lambda^x(e^{i\omega\cdot x})| d\omega$$

and the problem is to find λ such that the right–hand side is minimized.

This is not easy to handle, and is left as an open problem. Things are easier if we either minimize

$$\int_{\mathbb{R}^d} T(\omega) |\lambda^x(e^{i\omega\cdot x})|^2 d\omega$$

or try to push as much of T into the λ part as we can, introducing a weight function W such that we can bound

$$|\lambda(f)|^2 \leq (2\pi)^{-d} \left(\int_{\mathbb{R}^d} \frac{T^2(\omega)}{W(\omega)} d\omega \right) \left(\int_{\mathbb{R}^d} W(\omega) |\lambda^x(e^{i\omega\cdot x})|^2 d\omega \right)$$

such that the first integral is barely bounded, e.g. by choosing

$$W(\omega) = T^2(\omega) \|\omega\|_2^{d+\epsilon}$$

for a small positive ϵ.

Note that all the three cases boil down to minimizing

$$\|\lambda\|_{\mathcal{W}}^2 := \int_{\mathbb{R}^d} W(\omega) |\lambda^x(e^{i\omega\cdot x})|^2 d\omega \tag{4}$$

for a suitable nonnegative weight function W on $\mathbb{R}^d$. Minimizing (4) means finding a formula that is optimal in the space

$$\mathcal{W} := \left\{ f \;:\; \|f\|_{\mathcal{W}}^2 := \int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{W(\omega)} d\omega < \infty \right\} \tag{5}$$

in the sense that the error bound

$$|\lambda(f)| \leq \|\lambda\|_{\mathcal{W}} \|f\|_{\mathcal{W}} \text{ for all } f \in \mathcal{W}$$

is optimized over all differentiation formulae of the same form. This follows easily from introducing W into (3), as we will show below. Note how $1/W$ is penalizing the spectral behavior of the functions to be differentiated, while W penalizes the Fourier transform of the error functional.

3 Optimal Derivative Formulae

We now turn to the minimization of (4). It is an easy exercise in the context of kernel techniques (see H. Wendland's comprehensive book [10] which we use as a convenient reference throughout the paper) that under very weak assumptions on W the space (5) is a with a dual containing λ such that (4) can be written as $\|\lambda\|^2_{\mathcal{W}}$. The duality relation can be read off the simple bound

$$|\mu(f)|^2 \leq \left(\int_{\mathbb{R}^d} \frac{|\hat{f}(\omega)|^2}{W(\omega)} d\omega\right) \left(\int_{\mathbb{R}^d} W(\omega) |\mu^x(e^{i\omega\cdot x})|^2 d\omega\right) = \|\mu\|^2_{\mathcal{W}} \|f\|^2_{\mathcal{W}}$$

for each continuous functional μ. The essential assumption on W is that W should be nonnegative and integrable such that it has a continuous and real–valued inverse Fourier transform $W^\vee =: \Phi$. Then it is convenient to introduce the symmetric continuous translation–invariant *kernel*

$$K(x, y) := \Phi(x - y) \text{ for all } x, y \in \mathbb{R}^d$$

which is reproducing in $\mathcal{W}$, i.e.

$$\mu(f) = (f, \mu^x K(x, \cdot))_{\mathcal{W}} \text{ for all } f \in \mathcal{W},\ \mu \in \mathcal{W}^*.$$

If W is radially invariant, so is Φ, and the kernel K will then be a radial basis function. To avoid certain technical complications, we shall also assume that the weight function W is positive on a set of measure zero around the origin, which makes the kernel positive definite [10].

For the special case of Sobolev space $W_2^k(\mathbb{R}^d)$ with $k > d/2$ we get [10]

$$\Phi(z) = K_{k-d/2}(\|z\|_2) \|z\|_2^{k-d/2}$$

with the modified Bessel function $K_{k-d/2}$. Another easy exercise for reproducing kernel Hilbert spaces yields

$$\|\mu\|^2_{\mathcal{W}} = (\mu^x K(x, \cdot), \mu^y K(y, \cdot))_{\mathcal{W}} = \mu^x \mu^y K(x, y)$$

for all continuous functionals μ, and this implies

$$\begin{array}{rl} & \|\lambda\|^2_{\mathcal{W}} \\ = & \left(\delta^\alpha_x - \sum_{z\in Z} c(\alpha,x,z)\delta_z, \delta^\alpha_x - \sum_{y\in Z} c(\alpha,x,y)\delta_y\right)_{\mathcal{W}} \\ = & (\delta^\alpha_x, \delta^\alpha_x)_{\mathcal{W}} - 2\left(\delta^\alpha_x, \sum_{z\in Z} c(\alpha,x,z)\delta_z\right)_{\mathcal{W}} \\ & +\displaystyle\sum_{z\in Z}\sum_{y\in Z} c(\alpha,x,z)c(\alpha,x,y)\,(\delta_z,\delta_y)_{\mathcal{W}} \\ = & D^{\alpha,u}_x D^{\alpha,v}_x K(u,v) \\ & -2\sum_{z\in Z} c(\alpha,x,z) D^{\alpha,u}_x K(u,z) \\ & +\displaystyle\sum_{z\in Z}\sum_{y\in Z} c(\alpha,x,z)c(\alpha,x,y)K(z,y) \end{array} \qquad (6)$$

for our error functional λ. The above expression holds for any formula, good or bad. It is a quadratic form in the coefficients $c(\alpha,x,z)$ and it can be evaluated easily just by evaluating the kernel and its derivatives. We shall use this later to calculate the norms of various error functionals explicitly.

But since the above expression is a quadratic form, one can use standard arguments of Linear Algebra to prove that the optimal solution is given by solving the linear system

$$\sum_{y\in Z} c(\alpha,x,y)K(z,y) = D^{\alpha,u}_x K(u,z) \text{ for all } z\in Z \qquad (7)$$

which [10] necessarily has a positive definite coefficient matrix if W is positive on a set of measure zero. We summarize:

Theorem 3.1. *For spaces of continuous multivariate functions with specified spectral behavior in one of the three ways described above, optimal differentiation formulae are necessarily provided by kernel techniques. The optimal formulae can be explicitly calculated and turn out to be translation–invariant. In case of radial invariance of the spectral behavior, the optimal formulae are furnished by radial basis functions and have rotational invariance. They are exact on the span of the functions $K(\cdot,z)$ for $z\in Z$, and they can be obtained by taking the derivative of the interpolant based on that span.* □

But readers should note that optimal formulae will not have *scale* invariance in general. We shall come back to this later.

The above result is not new, but it is not widely known among users who need multivariate differentiation formulae. It follows the lines of

Optimal Recovery as surveyed by C.A. Micchelli and Th. Rivlin in [6], while the special application to kernel–based spaces is folklore for people working on reproducing kernel Hilbert spaces or radial basis functions. The article [4] by Y.C. Hon and T. Wei contains more information and also adds regularization strategies, while S. Jakobsson's paper [5] features specific examples and applications. Since the optimal formulae arise as derivatives of kernel–based interpolants, there are error bounds [10], but these hold only asymptotically when there are "enough and close–by" data.

4 Examples: Fixed Data

Univariate cases are not of interest here, because it is well–known that spline interpolants generate optimal differentiation formulae for Sobolev–type spaces.

Let us go into 2D and focus on an example where standard techniques are easily available for comparison. Consider data on an equilateral triangle spanned by the three roots of unity

$$z_1 := (1,0), z_2 := (-0.5, \sqrt{3}/4), z_3 := (-0.5, -\sqrt{3}/4)$$

around zero with distance 1 to zero. We want to predict the values $f(0,0)$, $f_x(0,0)$, $f_y(0,0)$ of a function f from these three values, and we deliberately keep the points fixed because we consider questions of scaling later. Clearly, by looking at affine–linear functions one expects something like

$$\begin{array}{rcrcrcr} f(0,0) & \approx & \frac{1}{3}f(z_1) & + & \frac{1}{3}f(z_2) & + & \frac{1}{3}f(z_3) \\ f_x(0,0) & \approx & \frac{2}{3}f(z_1) & - & \frac{1}{3}f(z_2) & - & \frac{1}{3}f(z_3) \\ f_y(0,0) & \approx & & & \frac{1}{\sqrt{3}}f(z_2) & - & \frac{1}{\sqrt{3}}f(z_3). \end{array}$$

If we remove the denominators row–wise for simplicity, the above stencils can be written in matrix form as

$$\begin{array}{rrr} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 0 & 1 & -1. \end{array} \tag{8}$$

Let us now see what happens if we take function spaces with fixed prescribed spectral behavior and calculate optimal formulae for these points. After row–wise rescaling as above, we get

1.0037	1.0037	1.0037
-2.3229	1.1615	1.1615
0.0000	-1.1615	1.1615

for the Gaussian kernel $\exp(-\|x-y\|_2^2)$, while the Sobolev/Matern kernel $K_2(\|x-y\|_2)\|x-y\|_2^2$ which is optimal for the space $W_2^3(\mathbb{R}^2)$ yields

1.1228	1.1228	1.1228
-2.1776	1.0888	1.0888
0.0000	-1.0888	1.0888

These numbers follow by solving the 3×3 system (7) for 3 different right–hand sides. Note that pointwise differentiation in $2D$ is not a continuous operation in $W_2^2(\mathbb{R}^2)$, forcing us to go into smaller Sobolev spaces.

As is to be expected, the results differ, but they are undoubtedly the optimal choices for the weighted function spaces (5). In order to get more insight into what happens here, we have to go somewhat further.

5 Scaling the Points

Users will often want to use differentiation formulae in a scale–invariant way. For instance, if points $z \in Z$ are scattered around the origin and should be used for approximation of $D^\alpha f(0)$, users will expect that all formulas of the form

$$D^\alpha f(0) \approx h^{-|\alpha|} \sum_{z \in Z} a(z) f(hz),$$

which have the same geometric form and the same (but scaled) coefficients work well for all small h and get even better for $h \to 0$.

This is not true in general, and the optimal formulae derived above will not necessarily be of this form. It is elementary to see that good behavior for $h \to 0$ for such formulae is equivalent to asking for exact differentiation of polynomials, leading to moment conditions for the coefficients. *If a user thinks of a “stencil” this way, there is not much leeway for Fourier techniques taking the spectrum into account.* However, since multivariate polynomial interpolation is a nontrivial issue, it is not straightforward how to define multivariate differentiation formulae which have a prescribed behavior on polynomials. One good choice would be the de Boor–Ron [2] multivariate polynomial interpolation. Another multivariate method that works unconditionally is called Schaback interpolation by Carl de Boor in [1] and goes back to [7].

Strangely enough, one can prove via [7, 8] for the Gaussian kernel that the associated optimal differentiation formulas will tend for $h \to 0$ (and after rescaling) to a formula which is exact on certain polynomials. It is a scale–invariant stencil coinciding with the one provided via de Boor–Ron [2] multivariate polynomial interpolation, and it is a special instance of a finite–difference formula with a certain polynomial exactness depending mysteriously on the point geometry.

6 Scaling the Spaces

But scaling the points has a counterpart within the function spaces. The basic connection is via the scaling relations between the Fourier transform and its inverse, and it is well–known from the standard Shannon–Kotelnikov sampling theorem. If we define $f_c(x) := f(x/c)$ for some $c > 0$, then

$$\hat{f}_c(\omega) = c^d \hat{f}(\omega \cdot c) \text{ for all } \omega \in \mathbb{R}^d.$$

By this transformation, functions f with support in the unit ball will go into functions f_c with support in a ball of radius c, i.e. they will get "spiky" for small c and "flat" for large c with unchanged smoothness properties. Looking at the spectrum, we see that if f is in Sobolev space $W_2^k(\mathbb{R}^d)$ with weight function $W(\omega) = (1 + \|\omega\|_2^2)^k$, the scaled function f_c is in a space with weight $c^{-d}(1 + \|c\,\omega\|_2^2)^k$ with the same numeric value of the norm. Note that for large c there will be much less penalty on high frequencies, as expected. The corresponding Sobolev spaces still consist of functions with a given smoothness, but the functions will have a much wilder local variation, like in turbulence for large Reynolds numbers.

The consequence is that for applications it does not suffice to pick a fixed Sobolev space of the form $W_2^k(\mathbb{R}^d)$, because there is a hidden assumption made on the scaling. One should rather fix an additional scaling c wisely.

Apart from that the functions and kernels in the space get much more "spiky" for small c, there is an interesting geostatistical interpretation (see e.g. the book [9] by M.L. Stein). Kernels model covariances of random fields. If each point $x \in \mathbb{R}^d$ is seen as a random variable, $K(x, y)$ gives the covariance between x and y. This means that spiky kernels for small c model cases where only nearby data are correlated.

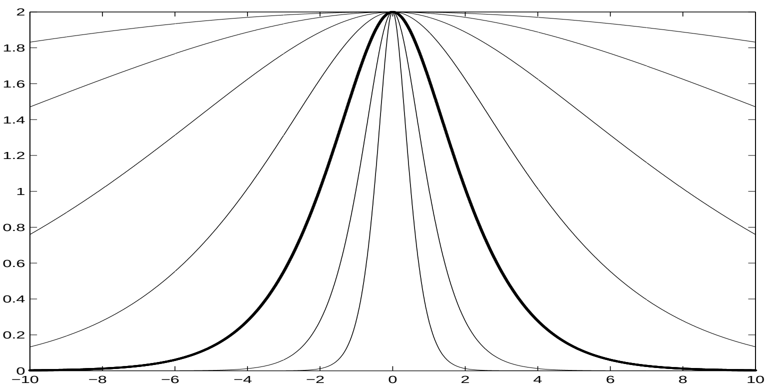

Figure 1: Sobolev kernel $K_2(r/c)(r/c)^2$ for $W_2^3(\mathbb{R}^2)$ at various scales c

Roughly speaking, the variable c indicates the distance up to which one can "rely" on "nearby" data. Again, it is useful to view function spaces with small c like turbulence.

Figure 1 shows scalings of the radial kernel $K_2(r)r^2$ generating $W_2^3(\mathbb{R}^2)$. The thick line is for $c = 1$, i.e. for the "normal" Sobolev space $W_2^3(\mathbb{R}^2)$, while the others have $c = 1/4, 1/2, 2, 4, 8, 16$. One can see that each scale defines a typical range where point values are "correlated", while the exponential decay of the kernel implies that far–apart data points cannot be used favourably for joint estimation of derivatives. This gives us some insight into the behavior of optimal differentiation formulae. The case above with fixed data distance h and fixed c was just a single snapshot. If we keep c fixed and take cases with $h \to 0$, the data are more and more "trustworthy". Similarly, if the data positions are fixed, taking large c means that we focus on spaces with more and more importance of low frequencies. These two situations will not lead to unexpected results.

However, if "*c is too small for h*" or "*h is too large for c*", the formulae still are optimal, but they are forced to work in a "turbulent" situation where it is near to impossible to make a reasonable guess for a derivative. In such cases one can observe that the optimal formulae have a bias towards predicting zero, which is the best they can do under these circumstances.

From the above discussion, it should be clear (and is easy to prove) that the h–scaling of data locations and the c–scaling of frequency spaces are in a 1–1 correspondence, like in the Shannon sampling theory.

We have to point out that there is a special case where optimal differentiation formulae are scale–invariant in the above sense. For this to happen, the underlying function space should have a scale–invariant spectral structure. This is the case for *polyharmonics*, where the kernel has a *generalized* Fourier transform which is a negative power of $\|\omega\|_2$. In particular, the radial basis functions $\phi(r) = r^\beta$ for $\beta \notin 2\mathbb{Z}$ enjoy this property together with their complements, the *thin–plate splines* $\phi(r) = r^\beta \log r$ for $\beta \in 2\mathbb{Z}$. One has to focus on stencils which are exact for low–order polynomials depending on β, but then the argument works as above. Details can easily be filled in using the book [10] by H. Wendland.

Choosing a good scaling for different situations is a well–discussed but unsolved problem in all kernel–based techniques. The recent paper [3] of G. Fasshauer and J.G. Zhang features a fairly general technique for scale estimation, while G.B. Wright and B. Fornberg in [11] deal extensively with scaling of derivative formulae.

7 Scaled Examples

Let us go back to our previous example and introduce scalings h and c on the data and frequency side, respectively. We take

$$z_1 := h(1,0), z_2 := h(-0.5, \sqrt{3}/4), z_3 := h(-0.5, -\sqrt{3}/4)$$

with distance h to zero and predict the values $f(0,0)$, $f_x(0,0)$, $f_y(0,0)$ of a function f from these three values. In what follows, we again remove the denominators from what normally is

$$\begin{array}{rcccccc}
f(0,0) & \approx & \frac{1}{3}f(z_1) & + & \frac{1}{3}f(z_2) & + & \frac{1}{3}f(z_3) \\
f_x(0,0) & \approx & \frac{2}{3h}f(z_1) & - & \frac{1}{3h}f(z_2) & - & \frac{1}{3h}f(z_3) \\
f_y(0,0) & \approx & & & \frac{1}{h\sqrt{3}}f(z_2) & - & \frac{1}{h\sqrt{3}}f(z_3).
\end{array}$$

The first case looks at smaller h for fixed $c = 1$. For $h = 0.1$ and the Gauss kernel with $c = 1$ we find

```
 1.0099    1.0099    1.0099
-2.0099    1.0050    1.0050
 0.0000   -1.0050    1.0050
```

while the Sobolev case for $W_2^3(\mathbb{R}^2)$ has

```
 1.0024    1.0024    1.0024
-2.0104    1.0052    1.0052
-0.0000   -1.0052    1.0052
```

This will be exactly the same as going for $c = 10$ with $h = 1$, and thus we do not need to repeat the results. Smaller h let the optimal stencils converge for $h \to 0$ to the standard one in (8).

If we keep $h = 1$ but go to $c = 0.1$, the prediction has to work under unreliable circumstances. The Sobolev case has

```
 0.0032    0.0032    0.0032
-0.0280    0.0140    0.0140
 0.0000   -0.0140    0.0140
```

while the Gaussian case produces values that cannot be distinguished from zero. If we use the scale–invariant polyharmonic splines on this example, they produce the fixed standard stencil, because they have to obey exactness on affine–linear polynomials.

Since we have (6) to evaluate the norm of *any* error functional on a space with weighted spectrum, we can easily compare the standard scaled stencil with the optimal one. If this criterion is applied, it turns out (see Figures 2 and 3) that the optimal one is only by a small factor better the standard one, and this occurs for all function spaces tested.

The sharp turn at a specific h stands for the fact that the frequency behavior of the fixed function space makes formulae for h larger than a critical value h_0 useless because the data are too far apart to be reliable. In standard terminology, such derivative formulae should have an $\mathcal{O}(h)$ error behavior, but this can be observed only for h being "small enough". Figures 2 and 3 show exactly where the asymptotics start to hold, and they do it uniformly for a function space, not for a specific function.

If errors are plotted as functions of c for a fixed configuration, very similar plots result, showing that errors decrease with a fixed order for increasing c, but also with a typical sharp turn at a "critical" c where the formula starts to be useful for larger c only.

8 Unstructured Examples

Though theory proves that kernel–based derivative formulae are optimal in a special sense for spaces with prescribed spectral behavior, the above examples indicate that if standard formulae of scalable stencil form are available, they can often be used without losing too much. Thus this paper does not encourage users to dump what they have used before.

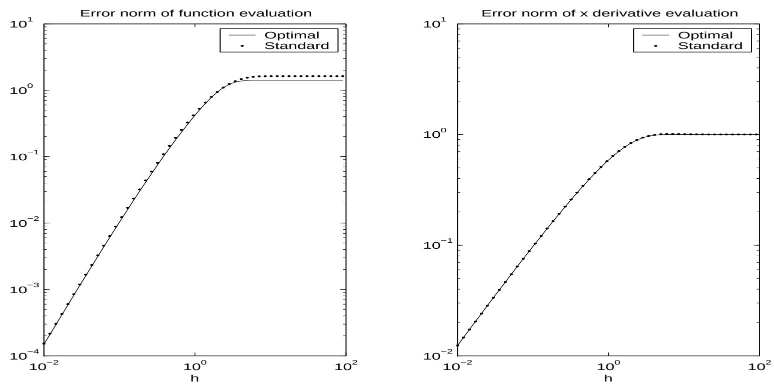

Figure 2: Norms of errors as functions of h, Sobolev kernel for $W_2^3(\mathbb{R}^2)$

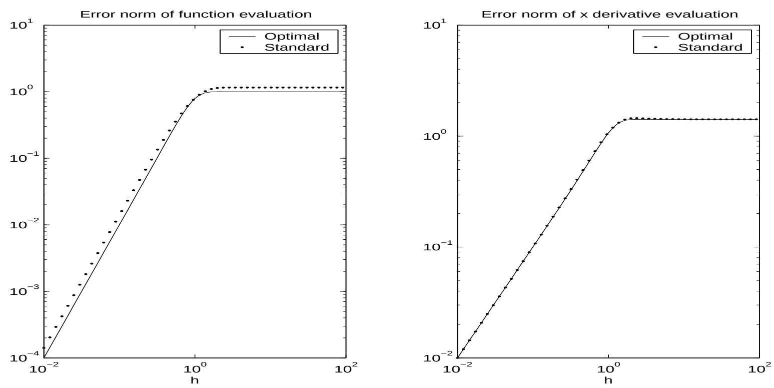

Figure 3: Norms of errors as functions of h, Gaussian kernel

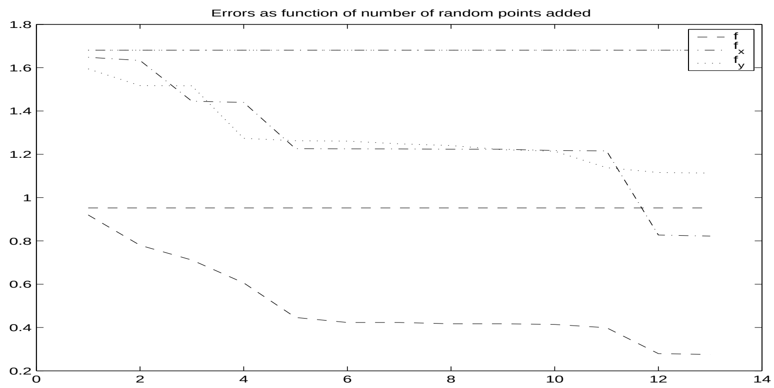

Figure 4: Errors as function of points added

However, for multivariate scattered data there is no generally applicable and widely accepted method for deriving good discretization formulae for derivatives. Optimal methods along the above lines are always applicable at low cost, but they may be bad when other criteria are applied. For instance, they always make optimal use of the available information, and this means that they will use nonzero weights on all the data offered, except when symmetry arguments apply. On the other hand, they can be used for high–order derivatives as well, provided that there are enough data available. Since the optimization principle implies that the errors get smaller if more points are involved, users can look at the tradeoff between better error versus higher complexity for larger data sets.

In our example we can add random data points to the ones we had before, calculate new optimal formulae for each case and then evaluate the errors. Figure 4 shows how the errors decrease. The constant lines are for the errors of the standard 3–point stencil without points added. We chose the kernel for Sobolev space $W_2^3(\mathbb{R}^2)$ with a somewhat too small scaling $c = 0.5$ in order to let the curves start at points where the standard stencil is still competitive. The final point configuration is given in Figure 5. The ©MATLAB programs running the examples are available via
`www.num.math.uni-goettingen.de/schaback/research`

Acknowledgements

Special thanks go to Elisabeth Larsson for a number of helpful comments.

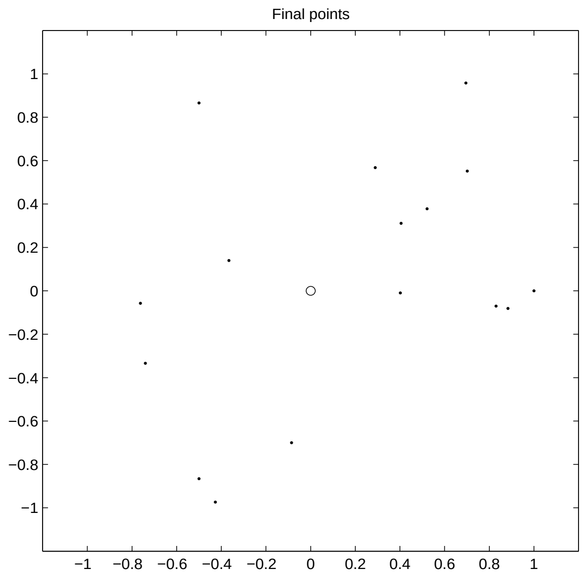

Figure 5: Final points added

References

[1] C. de Boor. On interpolation by radial polynomials. *Advances in Computational Mathematics*, pages 143–153, 2006.

[2] C. de Boor and A. Ron. On multivariate polynomial interpolation. *Constructive Approximation*, 6:287–302, 1990.

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

[4] Y.C. Hon and T. Wei. Numerical differentiation by radial basis functions approximation. *Advances in Computational Mathematics*, 27:247–272, 2007.

[5] S. Jakobsson. Frequency optimized computation methods. *J. Sci. Comput.*, 26(3):329–362, 2006.

[6] C.A. Micchelli and T.J. Rivlin. Lectures on optimal recovery. In P.R. Turner, editor, *Numerical Analysis, Lancaster 1984*, pages 12–93. Springer Verlag, Lecture Notes in Mathematics 1129, 1984.

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

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

[9] M.L. Stein. *Interpolation of spatial data*. Springer Series in Statistics. Springer, 1999.

[10] H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 2005.

[11] G.B. Wright and B. Fornberg. Scattered node compact finite difference-type formulas generated from radial basis functions. *J. Comput. Phys.*, 212(1):99–123, 2006.