

WAVELET-BASED FILTERS FOR ACCURATE COMPUTATION OF DERIVATIVES

MAURICE HASSON

ABSTRACT. Let $f(x)$ be a smooth function whose derivative of a given order must be computed. The signal $f(x)$ is affected by two kinds of perturbation. The perturbation caused by the presence of the machine epsilon ϵ_M of the computer may be considered to be an extremely high-frequency noise of very small amplitude. The way to minimize its effect consists of choosing an appropriate value for the step size of the difference quotient.

The second perturbation, caused by the presence of noise, requires first the signal to be treated in some way. It is the purpose of this work to construct a wavelet-based band-pass filter that deals with the two cited perturbations simultaneously. In effect our wavelet acts like a “smoothed difference quotient” whose stepsize is of the same order as that of the usual difference quotient. Moreover the wavelet effectively removes the noise and computes the derivative with an accuracy equal to the one obtained by the corresponding difference quotient in the absence of noise.

1. INTRODUCTION

Let $f(x)$ be a smooth function whose derivative of a given order must be computed. The signal $f(x)$ is affected by two kinds of perturbation. The perturbation caused by the presence of the machine epsilon ϵ_M of the computer may be considered to be an extremely high-frequency noise of very small amplitude. The way to minimize its effect consists of choosing an appropriate value for the step size of the difference quotient, as recalled below. The second perturbation, caused by the presence of noise, low pass or band-pass filter. requires first the signal to be treated in some way. One technique consists of processing the signal through a low pass or band-pass filter. Another technique consists of the use of regularization. Polynomial interpolation is also a well-spread method. We refer the reader to [1], [11], [14], [17], [18], [19], [22], [23], [25], [27], [28] among others and the references therein. There is one fortuitous case when differentiation reduces to integration: when the function $f(z)$ is analytic so that differentiation is obtained through the Cauchy formula. See [5], [20].

Numerical differentiation using wavelets has been investigated in [6] and [10] in the setting of finite element approximation.

It is the purpose of this work to construct a wavelet-based band-pass filter that performs the two cited operations simultaneously. In effect our wavelet acts like a

Received by the editor July 27, 2004.

2000 *Mathematics Subject Classification.* Primary 41A40, 42A85, 65D25, 65T60.

Key words and phrases. Wavelets, band-pass filters, high-frequency noise.

Supported by a VIGRE Postdoctoral Fellowship at the University of Arizona.

©2005 American Mathematical Society
Reverts to public domain 28 years from publication

“smoothed difference quotient” whose stepsize is of the same order as that of the usual difference quotient. Moreover our wavelet effectively removes the noise and computes the derivative with an accuracy equal to (in fact slightly better than) the one obtained by the corresponding difference quotient in the absence of noise.

It has to be noticed that, no matter how small ϵ_M is, the problem of finding the optimal stepsize as a function of ϵ_M is important because, as we will see, significant deviation from this optimal stepsize produces high error in the computation of the derivative.

The noise that we consider in the numerical computations of Section 5 is of the form $\cos(10000x) + \sin(10000x)$. However numerical experiments show that our wavelet effectively filters out noise of frequency greater than 100 and computes the derivative with an accuracy similar to the accuracy obtained in Subsection 5.2, namely 10^{-11} .

In addition, our method is cost effective because the quadratures involved in the approximation process are found at very low cost, as explained in Subsection 4.1.

To be specific let us recall the classical problem of approximating $f'(x)$ by the difference quotient $\frac{f(x+h)-f(x)}{h}$. It is a well-known fact (that we will review in Section 3), see [2], that

$$(1.1) \quad \left| \frac{f(x+h) - f(x)}{h} - f'(x) \right| \leq \frac{2|xf'(x)|\epsilon_M}{|h|} + \frac{|hf''(x)|}{2} + \mathcal{O}(h^2).$$

The term $\frac{2|xf'(x)|\epsilon_M}{|h|}$ is, of course, not present in infinite precision arithmetic so that the smaller h is the more accurate is the approximation. The situation is markedly different in the presence of a (positive) machine epsilon and in the case of the approximation $\frac{f(x+h)-f(x)}{h}$ the optimal h that we find is $\sqrt{c_1\epsilon_M}$, in which case the minimized error is $\sqrt{d_1\epsilon_M}$, where c_1 and d_1 are constants.

The situation, again in the absence of noise, improves if we replace the previous difference quotient by $\frac{f(x+h)-f(x-h)}{2h}$, in which case we obtain

$$(1.2) \quad \left| \frac{f(x+h) - f(x-h)}{2h} - f'(x) \right| \leq \frac{|xf'(x)|\epsilon_M}{|h|} + \frac{h^2|f^{(3)}(x)|}{6} + \mathcal{O}(h^3).$$

Hence in the case of the approximation $\frac{f(x+h)-f(x-h)}{2h}$ the optimal h is $\sqrt[3]{c_2\epsilon_M}$, in which case the minimized error becomes $\sqrt[3]{d_2\epsilon_M^2}$, where c_2 and d_2 are constants.

In fact by reiterating k times the above process (which, properly interpreted, is a form of Richardson extrapolation) it can be shown that, theoretically, the minimized error in approximating $f'(x)$ will be $c_k\epsilon_M^{\frac{k}{k+1}}$ for an optimal value of $h = d_k\epsilon_M^{\frac{1}{k+1}}$.

However in the presence of high-frequency noise the previous difference quotients, as well as their iterated counterparts, are essentially useless to compute $f'(x)$.

We present in this work a technique based on the construction of an appropriate wavelet, with *many vanishing moments* which, when being convolved in a precise manner with the function, acts as a band-pass filter and, *at the same time*, as a difference quotient. The reason for imposing the vanishing of moments will appear in Section 2. We will in fact compute $f''(x)$ and indicate the needed modification to find derivatives of other orders. The second derivative amplifies even more than the first one the effect of the noise and of the machine epsilon. Hence we choose the second derivative to demonstrate the efficiency of our method. As will be demonstrated in Section 4.1 we compute $f''(x)$ in the presence of high-frequency

noise with an error of the order of $\mathcal{O}\left(\sqrt[4]{\epsilon_M^3}\right)$. This is a theoretical bound as well as a bound on the error obtained through numerical experimentation. Hence with $\epsilon_M = 2^{-52}$ and assuming that the constant implied in \mathcal{O} above is ≈ 1 , the error is theoretically as well as experimentally $\leq 1.818989 \times 10^{-12}$.

The wavelet that we will build and that we call $\Psi_3(x)$ (because it is based on two previously built wavelets that we call $\Psi_2(x)$ and $\Psi_1(x)$) has specific values for its first eight moments. More precisely these moments satisfy (2.1) and (2.2) below.

Once $\Psi_3(x)$ is built $f''(x)$ will be approximated as follows:

$$(1.3) \quad f''(x) \approx \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt.$$

More precisely we will see that

$$(1.4) \quad \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - f''(x) = C_1 \frac{\epsilon_M}{h^2} + C_2 h^6 + \mathcal{O}(h^8)$$

from which the optimal error estimate $\mathcal{O}\left(\sqrt[4]{\epsilon_M^3}\right)$ stated above follows.

In addition $\Psi_3(x)$ will have the desirable band-pass filtering properties needed to deal with the noise present in the signal $f(x)$. It has to be noted that the error estimate in (1.4) is valid, both theoretically as well as experimentally, *with or without* the presence of noise.

Of course this technique, to be effective, requires that the above integrals be computed with high accuracy and at low cost. We will see that, indeed, this is the case.

Relation (1.3), in the context where $\Psi_3(x)$ is a specific moving average, was already anticipated by Shapiro (see [24]) in his treatment of saturation theory. We will see that saturation is already present in our context (Theorem 2.3) in the sense that no matter how smooth $f(x)$ is, the error estimate in (1.4) does not improve (unless $f(x)$ belongs to a very specific saturation class).

Throughout this paper we use the following notation for the Fourier transform $\widehat{f}(\xi)$ of a function $f(x)$:

$$\widehat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\xi x} dx.$$

The inversion formula then takes the form

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \widehat{f}(\xi) e^{i\xi x} d\xi.$$

We will restrict our attention to functions $f(x)$ belonging to the Schwartz class \mathcal{S} so that $\widehat{f}(\xi) \in \mathcal{S}$ also.

This paper is organized as follows. In the next section we build the wavelet fulfilling the required properties (2.1) and (2.2). In addition we prove the necessary error estimates. In Section 3 we derive the discrete analogue of the wavelet $\Psi_3(x)$ and establish the corresponding error estimates. It is this discrete analogue that helps us to establish the estimate (4.1). Section 4 is devoted to estimating the error, in the continuous case, in the presence of ϵ_M . Section 5 produces numerical experiments pertaining to our methods. Finally Section 6 ends with concluding remarks.

2. CONSTRUCTION OF THE WAVELET
AND ITS APPROXIMATION PROPERTIES

This section is devoted to building the wavelet $\Psi_3(x)$ satisfying properties (2.1) and (2.2) below:

$$(2.1) \quad \int_{-\infty}^{\infty} t^k \Psi_3(t) dt = 0, \quad k = 0, 1, 3, 4, 5, 6, 7$$

and

$$(2.2) \quad \int_{-\infty}^{\infty} t^2 \Psi_3(t) dt = 2.$$

The reasons for imposing these moment conditions will appear clearly below. (See Remark 2.4.)

Once $\Psi_3(x)$ is built we will analyse the rate at which it approximates the derivative of a given function.

2.1. Construction of the wavelet. The construction of the required wavelet will be achieved by imposing that for the Fourier transform $\widehat{\Psi}_3(\xi)$ of $\Psi_3(x)$ we have

$$\widehat{\Psi}_3^{(k)}(0) = 0, \quad k = 0, 1, 3, 4, 5, 6, 7$$

and

$$\widehat{\Psi}_3''(0) = -\sqrt{\frac{2}{\pi}}.$$

We will see that $\widehat{\Psi}_3^{(k)}(0) = 0$ if k is odd so that the first condition above reduces to $\widehat{\Psi}_3^{(k)}(0) = 0$, $k = 0, 4, 6$. The required wavelet is built by following an adaptation of the classical Richardson extrapolation method.

We begin with the standard Mexican hat function

$$\Psi(x) = -(1 - x^2)e^{-\frac{x^2}{2}}$$

whose Fourier transform is

$$\widehat{\Psi}(\xi) = -\xi^2 e^{-\frac{\xi^2}{2}}.$$

We refer the reader to [7], [15], [16] and [21] for deep analysis of this wavelet and its applications. Let

$$\widehat{\Psi}_1(\xi) = \frac{\widehat{\Psi}(\xi)}{\sqrt{2\pi}}.$$

By the expansion (2.3) we have

$$\widehat{\Psi}_1^{(k)}(0) = 0, \quad \text{for } k = 0, 1, 3$$

and

$$\widehat{\Psi}_1''(0) = -\sqrt{\frac{2}{\pi}}.$$

Hence we have

$$\int_{-\infty}^{\infty} \left(\frac{x}{h}\right)^2 \frac{1}{h} \Psi_1\left(\frac{x}{h}\right) dx = \int_{-\infty}^{\infty} x^2 \Psi_1(x) dx = 2.$$

We have

$$(2.3) \quad \widehat{\Psi}_1(\xi) = -\frac{\xi^2 e^{-\frac{\xi^2}{2}}}{\sqrt{2\pi}} = -\frac{1}{\sqrt{2\pi}} \xi^2 \left(1 - \frac{\xi^2}{2} + \frac{\xi^4}{8} - \frac{\xi^6}{48} + \dots\right)$$

and

$$\widehat{\Psi}_1\left(\frac{\xi}{2}\right) = -\frac{\xi^2 e^{-\frac{\xi^2}{8}}}{\sqrt{2\pi}} = -\frac{1}{\sqrt{2\pi}} \frac{\xi^2}{4} \left(1 - \frac{\xi^2}{8} + \frac{\xi^4}{128} - \frac{\xi^6}{3072} + \dots\right).$$

We see that for

$$\widehat{\Psi}_2(\xi) := \frac{\widehat{\Psi}_1(\xi) - 16\widehat{\Psi}_1\left(\frac{\xi}{2}\right)}{-3}$$

we have

$$\widehat{\Psi}_2^{(k)}(0) = 0, \text{ for } k = 0, 1, 3, 4, 5,$$

and

$$\widehat{\Psi}_2''(0) = -\sqrt{\frac{2}{\pi}}.$$

Hence

$$\Psi_2(x) = \frac{\Psi_1(x) - 32\Psi_1(2x)}{-3}.$$

We remark that

$$\int_{-\infty}^{\infty} \left(\frac{x}{h}\right)^2 \frac{2}{h} \Psi_2\left(\frac{x}{h}\right) dx = \int_{-\infty}^{\infty} x^2 2\Psi_2(x) dx = 2.$$

By reiterating the Richardson extrapolation technique on

$$\widehat{\Psi}_2(\xi) = \frac{\widehat{\Psi}_1(\xi) - 16\widehat{\Psi}_1\left(\frac{\xi}{2}\right)}{-3},$$

that is to say by considering linear combinations of

$$\widehat{\Psi}_2(\xi) \text{ and } \widehat{\Psi}_2\left(\frac{\xi}{2}\right),$$

and by using as above the Taylor expansion for $\widehat{\Psi}_2(\xi)$, we find that for

$$(2.4) \quad \widehat{\Psi}_3(\xi) = \frac{\widehat{\Psi}_1(\xi) - 80\widehat{\Psi}_1\left(\frac{\xi}{2}\right) + 1024\widehat{\Psi}_1\left(\frac{\xi}{4}\right)}{45}$$

we have

$$\widehat{\Psi}_3^{(k)}(0) = 0, \text{ } k = 0, 1, 3, 4, 5, 6, 7$$

and

$$\widehat{\Psi}_3''(0) = -\sqrt{\frac{2}{\pi}}.$$

Hence

$$\Psi_3(x) = \frac{\Psi_1(x) - 160\Psi_1(2x) + 4096\Psi_1(4x)}{45}$$

satisfies equations (2.1) and (2.2).

Let us summarize the above construction in the form of

Theorem 2.1. *For the wavelet $\Psi_3(x)$ defined by*

$$(2.5) \quad \Psi_3(x) = \frac{1}{45} \frac{(1-x^2)e^{(-1/2x^2)}}{\sqrt{2\pi}} - \frac{32}{9} \frac{(1-4x^2)e^{(-2x^2)}}{\sqrt{2\pi}} + \frac{4096}{45} \frac{(1-16x^2)e^{(-8x^2)}}{\sqrt{2\pi}},$$

equations (2.1) and (2.2) hold.

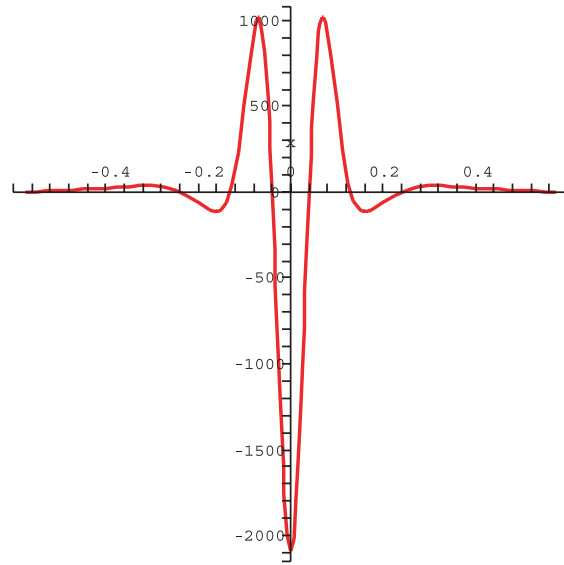


FIGURE 1. The wavelet $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$, $h = \frac{1}{10}$.

We display, in Figure 1, $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$, $h = \frac{1}{10}$.

We also display in Figure 2 the Fourier transform $\widehat{\Psi}_3(\xi)$ of $\Psi_3(x)$ illustrating its band-pass characteristics.

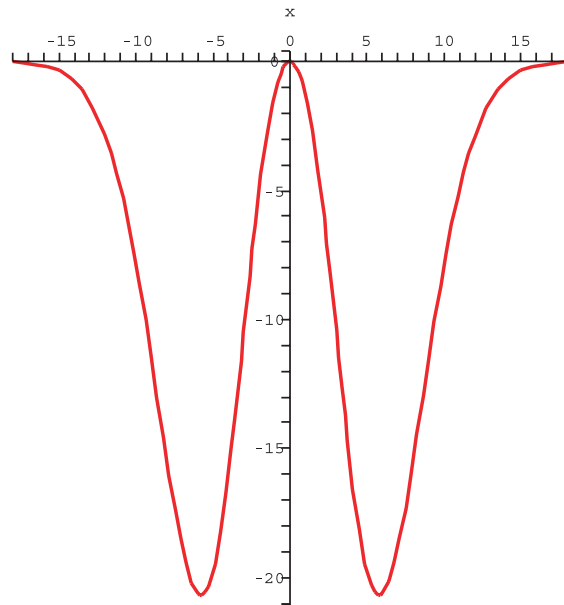


FIGURE 2. $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$, $h = \frac{1}{10}$ (above). The Fourier transform $\widehat{\Psi}_3(\xi)$ of $\Psi_3(x)$ illustrating its band-pass filter characteristics.

Our wavelet $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$ (more precisely its negative) has some superficial resemblance with the classical linear spline Battle-Lemarié wavelet [7], [15] and [21]. However it differs from it in two aspects. $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$ is smooth and is not orthogonal in $L_2(\mathbb{R})$. The smoothness of our wavelet is crucial in our work, whereas orthogonality in $L_2(\mathbb{R})$ is not needed. Without adequate smoothness its Fourier transform would decay too slowly at ∞ and, consequently, the wavelet would not have the needed filtering properties. More importantly the Battle-Lemarié wavelet, regardless of its order, cannot fulfill simultaneously the required moment conditions (2.1) and (2.2).

Our next task consists of analysing the approximation properties of $\Psi_3(x)$.

2.2. Properties of the wavelet $\Psi_3(x)$. Theorem 2.2 below will be the first step of the analysis of the error estimate in the presence of a (nonzero) machine epsilon ϵ_M .

Theorem 2.2. *Let $f(x)$ be a smooth function. Then, in infinite precision arithmetic,*

$$(2.6) \quad \begin{aligned} & \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - f''(x) \\ &= C_1 f^{(8)}(x) h^6 + C_2 f^{(10)}(x) h^8 + \mathcal{O}(h^{10}), \end{aligned}$$

where $\Psi_3(x)$ is given by (2.5). Here

$$(2.7) \quad C_1 = \frac{1}{8!} \int_{-\infty}^{\infty} t^8 \Psi_3(t) dt = 0.00032552$$

and

$$(2.8) \quad C_2 = \frac{1}{10!} \int_{-\infty}^{\infty} t^{10} \Psi_3(t) dt = 0.000053408.$$

More precisely, the error $E = E(x, h)$, expressed in terms of an asymptotic series, is

$$E = \sum_{k=4}^{\infty} \frac{1}{(2k)!} h^{2k-2} f^{(2k)}(x) \int_{-\infty}^{\infty} t^{2k} \Psi_3(t) dt.$$

Proof.

$$(2.9) \quad \begin{aligned} \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt &= \frac{1}{h^2} \int_{-\infty}^{\infty} f(x-th) \Psi_3(t) dt \\ &= \frac{1}{h^2} \int_{-\infty}^{\infty} \left(f(x) - thf'(x) + \frac{t^2 h^2}{2} f''(x) \right) \Psi_3(t) dt \\ &+ \frac{1}{h^2} \mathcal{O}(h^3). \end{aligned}$$

Using now (2.1) with $k = 0, 1$ and (2.2) we see from (2.9) that

$$(2.10) \quad \lim_{h \rightarrow 0} \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt = f''(x).$$

Now

$$\begin{aligned} & \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - f''(x) \\ &= \frac{1}{h^2} \int_{-\infty}^{\infty} \left(\frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - 2t^2 \frac{1}{h} \Psi_3\left(\frac{t}{h}\right) \frac{f''(x)}{2} \right) dt. \end{aligned}$$

Here we again used (2.2).

Hence

$$\begin{aligned} (2.11) \quad & \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - f''(x) \\ &= \frac{1}{h^2} \int_{-\infty}^{\infty} \left(f(x-th) - \frac{t^2 h^2}{2} f''(x) \right) \Psi_3(t) dt. \end{aligned}$$

Now use

$$f(x-th) - \sum_{k=0}^9 (-1)^k \frac{f^{(k)}(x)}{k!} t^k h^k + \mathcal{O}(t^{10} h^{10})$$

together with (2.1) and (2.11) to conclude that

$$\begin{aligned} (2.12) \quad \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - f''(x) &= \frac{1}{h^2} \int_{-\infty}^{\infty} \left(\frac{t^8 h^8}{8!} f^{(8)}(x) \right) \Psi_3(t) dt \\ &+ \frac{1}{h^2} \int_{-\infty}^{\infty} \left(\frac{t^{10} h^{10}}{10!} f^{(10)}(x) \right) \Psi_3(t) dt \\ &+ \frac{1}{h^2} \mathcal{O}(h^{12}). \end{aligned}$$

Here we used

$$\int_{-\infty}^{\infty} t^9 \Psi_3(t) dt = 0.$$

The quantities $C_1 = \frac{1}{8!} \int_{-\infty}^{\infty} t^8 \Psi_3(t) dt$ and $C_2 = \frac{1}{10!} \int_{-\infty}^{\infty} t^{10} \Psi_3(t) dt$ can be found analytically (and of course numerically) and their values are given by (2.7) and (2.8). Hence equations (2.12) together with (2.7) and (2.8) prove (2.6). The proof of the asymptotic formula for the error, which will not be used in the sequel, and which follows the same lines as those of the above proof, is omitted. We refer the reader to [4] for an exhaustive treatment of asymptotic series.

We indicate here how to find C_1 , C_2 being found in an analogous manner. We have

$$\int_{-\infty}^{\infty} x^8 \Psi_3(x) dx = \sqrt{2\pi} \widehat{\Psi}_3^{(8)}(0).$$

Expand $\widehat{\Psi}_3(\xi)$ using its defining relation (2.4) together with (2.3) and find

$$\int_{-\infty}^{\infty} t^8 \Psi_3(t) dt = 13.125.$$

In a similar manner we find

$$\int_{-\infty}^{\infty} t^{10} \Psi_3(t) dt = 193.806.$$

The proof of Theorem 2.2 is complete. \square

Remark 2.1. Fourier analysis techniques can also be used to show that

$$\lim_{h \rightarrow 0} \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt = f''(x).$$

However the error estimate in (2.6) could not be obtained by such techniques.

Remark 2.2. Examination of the above proof, together with the Lebesgue dominated convergence theorem, shows that if we were to assume only that

$$f(x+t) = a_0 + a_1 t + \frac{a_2}{2} t^2 + o(t^2) \text{ near } x,$$

then in that case we would have

$$\lim_{h \rightarrow 0} \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt = a_2.$$

It has to be noted that $f''(x)$ need not exist in this context. However, it does exist and its value is a_2 .

Remark 2.3. Let $P_n(x)$ be a polynomial of degree (at most) n . Then we have, regardless of the value of h ,

$$(2.13) \quad \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} P_n(x-t) \Psi_3\left(\frac{t}{h}\right) dt = P_n''(x)$$

for $n = 0, 1, \dots, 7$.

This follows from equations (2.1) and (2.2).

Remark 2.4. We see now the importance of using wavelets with many vanishing moments. Indeed the same procedure using wavelets $\Psi_1(x)$ and $\Psi_2(x)$ would give us

$$(2.14) \quad \begin{aligned} & \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_1\left(\frac{t}{h}\right) dt - f''(x) \\ &= C_1 f^{(4)}(x) h^2 + C_2 f^{(6)}(x) h^4 + \mathcal{O}(h^6) \end{aligned}$$

and

$$(2.15) \quad \begin{aligned} & \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_2\left(\frac{t}{h}\right) dt - f''(x) \\ &= C_1 f^{(6)}(x) h^4 + C_2 f^{(8)}(x) h^6 + \mathcal{O}(h^8), \end{aligned}$$

respectively.

In infinite precision arithmetic (2.14) and (2.15) show that the derivative can be approximated with arbitrary precision by taking h small enough. However it is in the presence of the machine epsilon that the advantage of $\Psi_3(x)$ over the wavelets $\Psi_1(x)$ and $\Psi_2(x)$ will appear clearly, as will be seen in the sequel. We already observed in (1.1) and (1.2) of the Introduction that in the presence of the machine epsilon h cannot be taken too small. In fact in the case of (2.14) the optimal h is $\mathcal{O}(\sqrt[4]{\epsilon_M})$ and in the case of (2.15) the optimal h is $\mathcal{O}(\sqrt[6]{\epsilon_M})$. It looks as if the situation is even worse for the wavelet $\Psi_3(x)$ in the sense that, now, the optimal h is $\mathcal{O}(\sqrt[8]{\epsilon_M})$. What counts, of course, are the optimal errors which are for the wavelets $\Psi_1(x)$, $\Psi_2(x)$ and $\Psi_3(x)$ of the order $\mathcal{O}(\sqrt{\epsilon_M})$, $\mathcal{O}(\sqrt[3]{\epsilon_M^2})$ and $\mathcal{O}(\sqrt[4]{\epsilon_M^3})$, respectively.

More importantly the wavelets $\Psi_1(x)$ and $\Psi_2(x)$ do not have the appropriate filtering characteristics.

Examination of Theorem 2.2 yields the following *saturation* result. The wavelet $\Psi_3(x)$ is saturated in the following sense: the error in (2.6) cannot be $o(h^6)$ unless $f(x)$ belongs to a specific class.

Theorem 2.3. *If*

$$\frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - f''(x) = o(h^6),$$

then, in fact,

$$\frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - f''(x) = \mathcal{O}(h^8),$$

and this is possible if and only if $f^{(8)}(x) = 0$. That is to say, the saturation class of the wavelet $\Psi_3(x)$, for a given point x , is the class \mathcal{S}_x of those smooth functions for which $f^{(8)}(x) = 0$. More generally if, for some $k = 0, 1, 2, \dots$,

$$\frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - f''(x) = o(h^{6+2k}),$$

then, in fact,

$$\frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt - f''(x) = \mathcal{O}(h^{8+2k})$$

for the same value of k and this is possible if and only if

$$f^{(8+2i)}(x) = 0, \quad i = 0, 1, 2, \dots, k.$$

The corresponding saturation class of the wavelet $\Psi_3(x)$, for a given point x , is the class \mathcal{S}_x^k of those smooth functions for which $f^{(8+2i)}(x) = 0$, $i = 0, 1, 2, \dots, k$.

It has to be noted that computer experiments very clearly confirm the above saturation results (for small values of k , of course).

2.3. The need for many—but not too many—vanishing moments. The proof of Theorem 2.2 shows the importance of equations (2.1) and (2.2) in order to obtain the error estimate (2.6), which will be important to approximate accurately the derivative in the presence of ϵ_M as evidenced by Theorem 4.1 and the remark following its proof.

A question arises naturally at this point: Why not reiterate the wavelet-based Richardson method and consider a wavelet of the form

$$\Psi_4(x) = a_1 \Psi_1(x) + a_2 \Psi_1(2x) + a_3 \Psi_1(4x) + a_4 \Psi_1(8x)$$

where, as before, $\Psi_1(x) = (1-x^2)e^{-\frac{x^2}{2}} \frac{1}{2\pi}$. (2.1) would then be replaced by

$$(2.16) \quad \int_{-\infty}^{\infty} t^k \Psi_4(t) dt = 0, \quad k = 0, 1, 3, 4, 5, 6, 7, 8, 9,$$

and, as before,

$$(2.17) \quad \int_{-\infty}^{\infty} t^2 \Psi_4(t) dt = 2.$$

Indeed the analysis of such a wavelet shows that we would obtain an approximation with an error of the order $\epsilon_M^{\frac{1}{5}}$. order $\frac{1}{8}$. That is to say, we would have an error close to the order of the machine accuracy. The reason for the inadequacy of the wavelet $\Psi_4(x)$ is transparent once we find its Fourier transform $\widehat{\Psi}_4(\xi)$, which contains a

term of the form $\widehat{\Psi}_1\left(\frac{\xi}{8}\right)$. Hence the Fourier transform of $\frac{1}{h}\Psi_4\left(\frac{x}{h}\right)$ contains a term of the form $\widehat{\Psi}_1\left(\frac{h}{8}\xi\right)$. That is to say $\Psi_4(x)$ now has the characteristics of a *high*-pass filter. This wavelet would no longer filter out moderately high-frequency noise, and the theoretical error of the order $\epsilon_m^{\frac{4}{5}}$, which would occur only in the presence of extremely high-frequency noise (or, of course, no noise at all), will not be achieved in the presence of moderately high-frequency noise. The wavelet $\Psi_3(x)$, in contrast, is a band-pass filter and is well suited for our purposes.

3. THE DISCRETE ANALOGUE OF THE WAVELET $\Psi_3(x)$
AND THE EFFECT OF THE MACHINE EPSILON

Our next result deals with the “discrete analogue”, which we call $T_3(x, h)$, of the wavelet $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$. $T_3(x, h)$ will be a distribution consisting of a sum of delta functions centered at appropriately determined points. In this section we will need properties of distributions as well as of their Fourier transforms and we refer the reader to [26] for an excellent treatment of the subject. The purpose of building $T_3(x, h)$ is twofold: first, to analyse the optimal value of h that minimizes the error $\frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t)\Psi_3\left(\frac{t}{h}\right) dt - f''(x)$ for a given value of the machine epsilon of the computer; second, to compare the filtering features of $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$ with those of difference quotients.

The process of building $T_3(x, h)$ is performed as follows. Let

$$T_1(x, h) = \delta(x - h) + \delta(x + h) - 2\delta(x).$$

Here $\delta(x)$ is the usual Dirac mass at 0. Then

$$\int_{-\infty}^{\infty} f(x-t)T_1(t, h)dt = f(x+h) + f(x-h) - 2f(x).$$

Here $\int_{-\infty}^{\infty}$ stands for the usual distributional pairing. Hence

$$\lim_{h \rightarrow 0} \int_{-\infty}^{\infty} f(x-t)\frac{1}{h^2}T_1(t, h)dt = f''(x).$$

We also have

$$\int_{-\infty}^{\infty} t^k T_1(t, 1)dt = 0, \quad k = 0, 1, 3$$

and

$$\int_{-\infty}^{\infty} t^2 T_1(t, 1)dt = 2.$$

It is in that sense that $T_1(x, h)$ is the discrete analogue of the wavelet $\frac{1}{h}\Psi_1\left(\frac{x}{h}\right)$. Recall now that

$$\Psi_3(x) = \frac{\Psi_1(x) - 160\Psi_1(2x) + 4096\Psi_1(4x)}{45}.$$

Now it follows from

$$\delta(ax) = \frac{1}{|a|}\delta(x)$$

that

$$T_1(2x, h) = \frac{1}{2}\delta\left(x - \frac{h}{2}\right) + \frac{1}{2}\delta\left(x + \frac{h}{2}\right) - \delta(x)$$

and

$$T_1(4x, h) = \frac{1}{4}\delta\left(x - \frac{h}{4}\right) + \frac{1}{4}\delta\left(x + \frac{h}{4}\right) - \frac{1}{2}\delta(x).$$

That is to say, the distribution $T_3(x, h)$ defined by

$$\begin{aligned} T_3(x) &= \frac{1}{45}(\delta(x-h) + \delta(x+h) - 2\delta(x)) \\ &\quad - \frac{160}{45}\left(\frac{1}{2}\delta\left(x - \frac{h}{2}\right) + \frac{1}{2}\delta\left(x + \frac{h}{2}\right) - \delta(x)\right) \\ &\quad + \frac{4096}{45}\left(\frac{1}{4}\delta\left(x - \frac{h}{4}\right) + \frac{1}{4}\delta\left(x + \frac{h}{4}\right) - \frac{1}{2}\delta(x)\right) \end{aligned}$$

is the discrete analogue of the wavelet

$$\frac{1}{h}\Psi_3\left(\frac{x}{h}\right).$$

It can be shown, either by direct Taylor expansion, or by using the analogous results for $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$ that we also have

$$\int_{-\infty}^{\infty} t^k T_3(t, 1) dt = 0, \quad k = 0, 1, 3, 4, 5, 6, 7$$

and

$$\int_{-\infty}^{\infty} t^2 T_3(t, 1) dt = 2.$$

Moreover

$$\lim_{h \rightarrow 0} \int_{-\infty}^{\infty} f(x-t) \frac{1}{h^2} T_3(t, h) dt = f''(x).$$

The precise meaning of this statement is

Theorem 3.1.

$$(3.1) \quad \lim_{h \rightarrow 0} \frac{1}{h^2} T_3(x, h) = \lim_{h \rightarrow 0} \frac{1}{h^2} \left(\frac{1}{h} \Psi_3\left(\frac{x}{h}\right) \right) = \delta''(x)$$

in the sense of distributions. Here $\delta''(x)$ is the distribution

$$\langle \delta''(x), \rho(x) \rangle = \rho''(0)$$

for $\rho \in \mathcal{S}$. Moreover, as in the case of $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$,

$$(3.2) \quad \int_{-\infty}^{\infty} \frac{1}{h^2} f(x-t) T_3(t, h) dt - f''(x) = C_1 f^{(8)}(x) h^6 + C_2 f^{(10)}(x) h^8 + \mathcal{O}(h^{10}),$$

where now

$$(3.3) \quad C_1 = \frac{1}{8!} \int_{-\infty}^{\infty} t^8 T_3(t, 1) dt$$

and

$$(3.4) \quad C_2 = \frac{1}{10!} \int_{-\infty}^{\infty} t^{10} T_3(t, 1) dt.$$

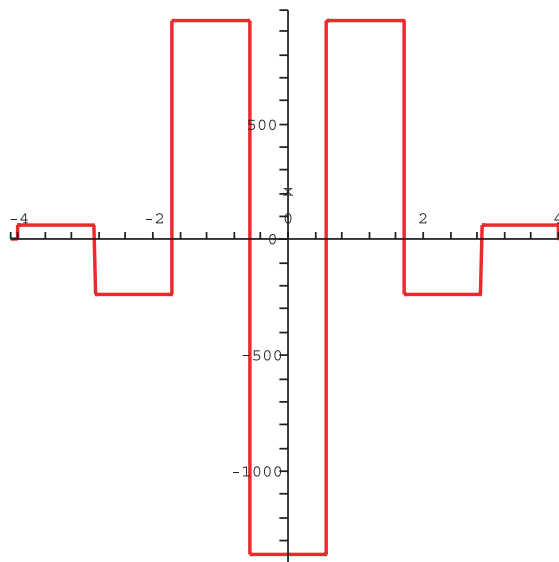


FIGURE 3. $T_3(x, 1)$. (Where $\delta(x) \approx n\chi_{[-\frac{1}{2n}, \frac{1}{2n}]}(x)$.)

This error estimate can be found by an adaptation of the analysis of the error estimate of the continuous case or by direct Taylor expansion.

In other words, $f''(x)$ is approximated with $f_{pp}(x, h)$ where

$$\begin{aligned}
 f_{pp}(x, h) &= \frac{w_1 f_1(x, h) + w_2 f_1(x, \frac{h}{2}) + w_3 f_1(x, \frac{h}{4})}{h^2}, \\
 w_1 &= \frac{1}{45}, \\
 w_2 &= -\frac{16}{9}, \\
 w_3 &= \frac{1024}{45}, \\
 (3.5) \quad f_1(x, h) &= f(x - h) + f(x + h) - 2f(x).
 \end{aligned}$$

We display in Figure 3 the shape of $T_3(x, 1)$ (where $\delta(x) \approx n\chi_{[-\frac{1}{2n}, \frac{1}{2n}]}(x)$).

As in the case of the wavelet $\Psi_3(x)$ we have

Remark 3.1. Let $P_n(x)$ be a polynomial of degree (at most) n . Then we have, regardless of the value of h ,

$$P_{npp}(x, h) = P_n''(x)$$

for $n = 0, 1, \dots, 7$.

We address the problem of finding the optimal value of h , h_{opt} , in order to minimize the discrepancy $\frac{1}{\sqrt{2\pi}} \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x - t) \Psi_3(\frac{t}{h}) dt - f''(x)$. Recall that the machine ϵ , ϵ_M , of a computer is the smallest positive floating-point number, in base 2, such that

$$1 + \epsilon_M > 1.$$

It is then seen that the absolute error which occurs when a number x is entered is of the order $x\epsilon_M$. (In fact the error is $x\frac{\epsilon_M}{2}$ if the machine stores by rounding instead of truncating.) The mean value theorem then shows that the error occurring in $f(x)$ is

$$|e| \approx |xf'(x)|\epsilon_M.$$

Because h is small we assume that this is also a bound for the error in $f(x+h)$. An error occurs also when we enter h . However,

$$\frac{1}{h(1+\epsilon_M)} = \frac{1}{h} (1 - \epsilon_M + \mathcal{O}(\epsilon_M^2)) \approx \frac{1}{h}.$$

Hence the absolute value of the round-off error, E_{round} , in calculating the difference quotient

$$\frac{f(x+h) - f(x)}{h}$$

is then bounded by

$$|E_{\text{round}}| \leq \frac{2|e|}{|h|} \approx \frac{2|xf'(x)|\epsilon_M}{|h|}.$$

The truncation error E_{trunc} is

$$E_{\text{trunc}} = \frac{hf''(\xi)}{2}$$

where $x < \xi < x+h$ (if $h > 0$).

Because h is small we assume that $\xi = x$. For the total error, $E_{\text{tot}} = E_{\text{round}} + E_{\text{trunc}}$, we then have

$$|E_{\text{tot}}| \leq |E_{\text{round}}| + |E_{\text{trunc}}|.$$

We conclude that the optimal step size h can be obtained by minimizing

$$|E_{\text{round}}| + |E_{\text{trunc}}| \leq \frac{2|xf'(x)|\epsilon_M}{|h|} + \frac{|hf''(x)|}{2}.$$

A similar analysis can be performed when we consider the approximation of $f'(x)$ by

$$\frac{f(x+h) - f(x-h)}{2h}$$

where now

$$|E_{\text{round}}| \leq \frac{|xf'(x)|\epsilon_M}{|h|},$$

whereas the round-off error is

$$|E_{\text{round}}| \leq \frac{|f^{(3)}(x)|h^2}{6}.$$

As before the optimal h can be found by minimizing

$$|E_{\text{round}}| + |E_{\text{trunc}}| \leq \frac{|xf'(x)|\epsilon_m}{|h|} + \frac{h^2|f^{(3)}(x)|}{6}.$$

Hence in the case of the approximation $\frac{f(x+h)-f(x)}{h}$ the optimal h that we find is $\sqrt{c_1\epsilon_M}$ and e_{opt} is $\sqrt{d_1\epsilon_M}$, whereas in the case $\frac{f(x+h)-f(x-h)}{2h}$ the optimal h is $\sqrt[3]{c_2\epsilon_M}$ and e_{opt} is $\sqrt[3]{d_2\epsilon_M^2}$ where c_1, c_2, d_1 and d_2 are constants.

3.1. **Proof of Theorem 3.1.** As in the cases of the difference quotients

$$\frac{f(x+h) - f(x)}{h}$$

and

$$\frac{f(x+h) - f(x-h)}{2h}$$

we assume, because h is small, that

$$|e| \approx |xf'(x)|\epsilon_M,$$

besides being a bound for the error in $f(x)$, is also a bound for the error in $f(x+h)$, $f(x-h)$, $f(x+\frac{h}{2})$, $f(x-\frac{h}{2})$, $f(x+\frac{h}{4})$ and $f(x-\frac{h}{4})$. It follows that the absolute value of the round-off error, E_{round} , in calculating the difference quotient $f_{pp}(x, h)$ given by (3.5) is bounded by

$$|E_{\text{round}}| \leq \frac{2|e|}{h^2} \approx \frac{1 + 80 + 1024}{15} \frac{|xf'(x)| \epsilon_m}{h^2}.$$

Of course h is also entered with an error of the order ϵh . But

$$\frac{1}{(h + \epsilon h)^2} = \frac{1}{h^2} (1 + \mathcal{O}(\epsilon)) \approx \frac{1}{h^2}.$$

The value of the constant C_1 given in (3.3) is now needed. It is easily found to be

$$(3.6) \quad C_1 = \frac{1}{8!} \frac{1}{45} \left(2 - 80 \frac{1}{2^8} 2 + \frac{1024}{4^8} 2 \right).$$

Hence for the truncation error we have, in view of Theorem 3.1 and (3.3),

$$|E_{\text{trunc}}| = \frac{1}{8!} \frac{1}{45} \left(2 - 80 \frac{1}{2^8} 2 + \frac{1024}{4^8} 2 \right) h^6 |f^{(8)}(x)|.$$

For the total error, $E_{\text{tot}} = E_{\text{round}} + E_{\text{trunc}}$, we then have

$$|E_{\text{tot}}| \leq |E_{\text{round}}| + |E_{\text{trunc}}|.$$

The optimal step size h can then be obtained by minimizing (for h)

$$\begin{aligned} |E_{\text{round}}| + |E_{\text{trunc}}| &= \frac{1 + 80 + 1024}{15} \frac{|xf'(x)|}{h^2} \epsilon_m \\ &+ \frac{1}{8!} \frac{1}{45} \left(2 - 80 \frac{1}{2^8} 2 + \frac{1024}{4^8} 2 \right) h^6 |f^{(8)}(x)| \end{aligned}$$

or

$$\begin{aligned} |E_{\text{round}}| + |E_{\text{trunc}}| &= 368.33333 \frac{|xf'(x)|}{h^2} \epsilon_m \\ &+ 7.7507526 \times 10^{-7} h^6 |f^{(8)}(x)|. \end{aligned}$$

We obtain

$$h_{\text{opt}} = \mathcal{O}(\sqrt[8]{\epsilon_m}).$$

More precisely

$$h_{\text{opt}} = 10.591854 \sqrt[8]{\frac{|xf'(x)|}{3|f^{(8)}(x)|}} \sqrt[8]{\epsilon_m}.$$

Assuming that $\sqrt[8]{\frac{|xf'(x)|}{3|f^{(8)}(x)|}} \approx 1$, we see that, with $\epsilon_m = 2^{-52}$, we have $h_{\text{opt}} = 0.1170245602199820$. With the same assumption on $\sqrt[8]{\frac{|xf'(x)|}{3|f^{(8)}(x)|}}$ and with the same

value of ϵ_m , we find that $|E_{\text{round}}| + |E_{\text{trunc}}| \approx 10^{-12}$. Numerical experiments, under the above assumptions, very precisely confirm these values for h_{opt} and for the error.

The optimal value of the error, E_{opt} , obtained by using $h = h_{\text{opt}}$, is then

$$E_{\text{opt}} = \left(3.283231 \frac{|xf'(x)|}{\sqrt[4]{\frac{|xf'(x)|}{3|f^{(8)}(x)|}}} + 1.094372 |f^{(8)}(x)| \sqrt[4]{\left| \frac{xf'(x)}{3f^{(8)}(x)} \right|^3} \right) \times \epsilon_m^{\frac{3}{4}}.$$

We omit the details of this lengthy but, otherwise, elementary computation. Hence

$$E_{\text{opt}} = \mathcal{O} \left(\sqrt[4]{\epsilon_m^3} \right).$$

Assuming that the constant implied in \mathcal{O} above is close to 1, namely assuming that x , $f'(x)$ and $f^{(8)}(x)$ are “reasonable,” and with $\epsilon_m = 2^{-52}$, we see that

$$E_{\text{opt}} = 1.81989 \times 10^{-12}.$$

Computer experiments show that for the given value of $h = h_{\text{opt}} = 0.11702$, the error in approximating $f''(x)$ by $\frac{1}{h^2}T(x) * f(x)$ is of the order of (in fact slightly smaller than) $E_{\text{opt}} = 1.81989 \times 10^{-12}$, namely $E = 1.22381 \times 10^{-12}$. Maybe the most remarkable aspect of these experiments is that the error is precisely minimized for $h = 0.11376$, a value very close to $h = h_{\text{opt}} = 0.11702$.

3.2. The filtering characteristics of $T_3(x, h)$ compared to those of $\frac{1}{h}\Psi_3\left(\frac{t}{h}\right)$.

At this point the following question may come to mind. By equation (3.2) the error in approximating $f''(x)$ by $\int_{-\infty}^{\infty} \frac{1}{h^2}f(x-t)T_3(t, h)dt$ gives us the same error estimate as the one using the wavelet $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$. Then what is the advantage of the wavelet approach? Why not approximate $f''(x)$ using simply the quantity f_{pp} given by equation (3.5)? We do not because the distribution $T_3(x, h)$ has no filtering characteristics. Indeed its Fourier transform

$$\widehat{T}(\xi) = \sum_{k=0}^2 a_k \left(e^{i2^{-k}h\xi} + e^{-i2^{-k}h\xi} \right) + C$$

has no decay. The wavelet $\frac{1}{h}\Psi_3\left(\frac{x}{h}\right)$, on the other hand, is a very effective band-pass filter (when an integration scheme is properly used) so that it can find the value of $f''(x)$ very accurately even in the presence of high-frequency noise.

4. THE ERROR ESTIMATE $\frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h}f(x-t)\Psi_3\left(\frac{t}{h}\right) dt - f''(x)$ IN THE PRESENCE OF ϵ_M

We have the following

Theorem 4.1.

$$(4.1) \quad \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h}f(x-t)\Psi_3\left(\frac{t}{h}\right) dt - f''(x) = C_1 \frac{\epsilon_M}{h^2} + C_2 h^6 + \mathcal{O}(h^8).$$

Lemma 4.1. *Let $\epsilon, \delta > 0$. Then there exists η such that*

$$|h| < \eta$$

implies

$$(4.2) \quad \int_{|t|>\delta} \frac{1}{h} \left| \Psi_3\left(\frac{t}{h}\right) \right| < \epsilon.$$

Proof. This lemma is well known for $\frac{1}{h}e^{-\frac{(\frac{t}{h})^2}{2}}$ as well as for its derivatives [12]; hence it holds true for $\frac{1}{h}\Psi_3\left(\frac{t}{h}\right)$ in view of (2.5). \square

Proof of Theorem 4.1. We assume that, as a function of t and for fixed x , we have

$$(4.3) \quad |f'(x-t)(x-t)| \leq M.$$

This is certainly the case if f belongs to the class \mathcal{S} . As already analysed above, the round-off error due to the presence of ϵ_M , in computing $f(x-t)$, is bounded by $|f'(x-t)(x-t)|\epsilon_M$. Hence the round-off error E_{round} , in computing $\int_{-\infty}^{\infty} \frac{1}{h}f(x-t)\Psi_3\left(\frac{t}{h}\right) dt$ is bounded by $\int_{-\infty}^{\infty} \frac{1}{h}|f'(x-t)(x-t)\epsilon_M\Psi_3\left(\frac{t}{h}\right)| dt$. For $\epsilon, \delta > 0$, let $\eta > 0$ be as given by Lemma 4.1. Hence for $|h| < \eta$ it follows that for the error E_{round} we have, in view of (4.2),

$$(4.4) \quad E_{\text{round}} \leq \int_{|t| \leq \delta} \frac{1}{h} \left| f'(x-t)(x-t)\epsilon_M\Psi_3\left(\frac{t}{h}\right) \right| dt + \epsilon M.$$

Because $|t| \leq \delta$ we assume that $f'(x-t)(x-t) \approx f'(x)x$. Hence

$$(4.5) \quad E_{\text{round}} \leq |f'(x)x|\epsilon_M \times \int_{|t| \leq \delta} \frac{1}{h} \left| \Psi_3\left(\frac{t}{h}\right) \right| dt + \epsilon M.$$

Thus, because ϵ can be chosen as small as we please, we have

$$(4.6) \quad E_{\text{round}} \leq |f'(x)x|\epsilon_M \times \int_{-\infty}^{\infty} |\psi_3(t)| dt,$$

in view of

$$\int_{|t| \leq \delta} \frac{1}{h} \left| \Psi_3\left(\frac{t}{h}\right) \right| dt < \int_{-\infty}^{\infty} \frac{1}{h} \left| \Psi_3\left(\frac{t}{h}\right) \right| dt = \int_{-\infty}^{\infty} |\psi_3(t)| dt.$$

Hence the round-off error E , due to the presence of ϵ_M , in computing

$$\frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t)\Psi_3\left(\frac{t}{h}\right) dt$$

is

$$(4.7) \quad E \leq \frac{|f'(x)x|\epsilon_M}{h^2} \times \int_{-\infty}^{\infty} |\psi_3(t)| dt.$$

We know, from Theorem 2.2, that the truncation error in

$$\frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t)\Psi_3\left(\frac{t}{h}\right) dt - f''(x)$$

is

$$(4.8) \quad E_{\text{trunc}} = C_1 f^{(8)}(x)h^6 + C_2 f^{(10)}(x)h^8 + \mathcal{O}(h^{10}).$$

Hence Theorem 4.1 follows from (4.7) and (4.8). \square

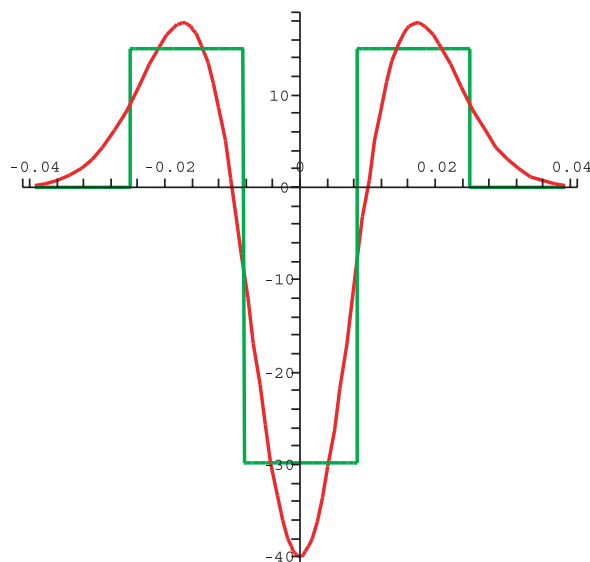


FIGURE 4. The wavelet $\frac{1}{h}\Psi_1\left(\frac{x}{h}\right)$ and its approximation by $T_1(x, h)$.

Theorem 4.1, in addition, gives an optimal error estimate $\mathcal{O}\left(\sqrt[4]{\epsilon_M^3}\right)$ using $h = h_{\text{opt}} = \mathcal{O}\left(\sqrt[8]{\epsilon_m}\right)$.

The errors in the cases $\frac{1}{h}\Psi_3\left(\frac{t}{h}\right)$ and $T_3(x, h)$ are qualitatively similar. This was to be expected because, as noticed earlier, $T_3(x, h)$, when properly interpreted, is an approximation of $\frac{1}{h}\Psi_3\left(\frac{t}{h}\right)$. We illustrate this idea in the case of $T_1(x, h)$ and $\frac{1}{h}\Psi_1\left(\frac{t}{h}\right)$. (If we try to superimpose $T_3(x, h)$ and $\frac{1}{h}\Psi_3\left(\frac{t}{h}\right)$ we get a less clear picture because of the high oscillatory nature of these wavelets.)

Figure 4 (restricted to the case $\Psi_1(x)$ and $T_1(x, 1)$) illustrates the above ideas.

4.1. Numerical considerations. As noticed above, our method, to be effective, requires that the integrals involved in (4.9)

$$(4.9) \quad \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt$$

be computed with high accuracy and at low cost. One of the remarkable aspects of the Mexican hat function and its modifications is the possibility of computing with almost machine accuracy and at very low cost the convolutions involving these wavelets by a quadrature rule as simple as the trapezoidal rule or Simpson's rule. As the reader may guess this is made possible by the appropriate use of the Euler-McLaurin summation formula. It is common knowledge that the Euler-McLaurin summation formula used in conjunction with the trapezoidal rule allows high accuracy computation of integrals of smooth periodic functions (integrated over a period). Not only the same holds true for Simpson's rule, but there is no need for the function to be periodic. It suffices to assume that the derivatives of odd orders are equal at the endpoints of the interval of integration. In that respect see [30]. And this is precisely what (almost) happens when convolving a smooth function with $\frac{1}{h}\Psi_3\left(\frac{t}{h}\right)$. Moreover, because of the fast decay of these

wavelets, numerical experiments show that the quadratures in (4.9) are computed with machine accuracy by replacing them by

$$\frac{1}{h^2} \int_{-9h}^{9h} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt.$$

The wavelets needed to compute derivatives of other orders are built, similarly, by Richardson-like linear combinations of an appropriate derivative of the Gaussian function. These results will be the object of a subsequent paper.

5. COMPARISON OF THE ACCURACIES WITH AND WITHOUT WAVELET FILTERING

Recall that we considered two approximation processes to $f''(x)$: (1.3) that we reproduce here for convenience

$$(5.1) \quad f''(x) \approx \frac{1}{h^2} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi_3\left(\frac{t}{h}\right) dt$$

and the difference quotient

$$(5.2) \quad f''(x) \approx \frac{1}{h^2} \int_{-\infty}^{\infty} f(x-t) T_3(t, h) dt,$$

which amounts to approximating $f''(x)$ by $f_{pp}(x, h)$ (see (3.5)) given by

$$(5.3) \quad f_{pp}(x, h) = \frac{\frac{1}{45}f_1(x, h) - \frac{16}{9}f_1\left(x, \frac{h}{2}\right) + \frac{1024}{45}f_1\left(x, \frac{h}{4}\right)}{h^2},$$

$$f_1(x, h) = f(x-h) + f(x+h) - 2f(x).$$

We now present four numerical results pertaining to our methods.

5.1. Approximation in the absence of noise (and with the presence of ϵ_M). We present in Table 1 two experiments where no noise is present, and with the value of $\epsilon_M = 2^{-52}$.

These experiments pertain to the approximation of $f''(x)$, where $f(x) = \cos(x)$, with the processes (5.1) and (5.3), respectively. We witness results of the same qualitative nature (with a slight improvement in the case of (5.1)). We also observe that the error, in the case of the processes (5.1), is minimized for a rather wide

TABLE 1. Comparison between the two approximation processes in the absence of noise. $f(x) = \cos(x)$.

h	$f''(0)$ computed with (5.1)	$f''(0)$ computed with (5.3)
1	-0.99972265810015	-0.99999923616273
$\frac{1}{5}$	-0.9999997930282	-0.9999999995037
$\frac{1}{10}$	-0.9999999967499	-0.9999999997905
$\frac{1}{15}$	-0.9999999997147	-0.999999999896
$\frac{1}{20}$	-0.999999999512	-0.999999999910
$\frac{1}{25}$	-0.999999999886	-1.0000000000095
$\frac{1}{30}$	-0.999999999994	-0.999999999891
$\frac{1}{100}$	-0.999999999997	-0.999999999902
$\frac{1}{200}$	-0.999999999995	-0.9999999980702
$\frac{1}{400}$	-0.999999999988	-0.9999998939477

range of values of h , namely if $h \in [\frac{1}{400}, \frac{1}{30}]$. Deterioration occurs once $h \leq \frac{1}{450}$. In the case of the processes (5.3) the range of optimal values of h is more restricted: $h \in [\frac{1}{100}, \frac{1}{30}]$. In both cases we see an optimal choice for h is $h = \sqrt[3]{\epsilon_m} \approx \frac{1}{90}$, as predicted by the theory.

5.2. Approximation in the presence of noise (and with the presence of ϵ_M). We present here, in Table 2, two experiments where, now, high-frequency noise is present as indicated, and with the same value of $\epsilon_M = 2^{-52}$.

These experiments pertain to the approximation of $f''(x)$, where

$$f(x) = \cos(x) + \cos(10000x) + \sin(10000x),$$

with the processes (5.1) and (5.3), respectively. We also observe that, again, the error in the case of the processes (5.1) is minimized for a rather wide range of values of h , namely if $h \in [\frac{1}{400}, \frac{1}{30}]$. Deterioration occurs once $h \leq \frac{1}{450}$. In the case of the processes (5.3), clearly, because of the absence of filtering, no approximation is obtained.

TABLE 2. Comparison between the two approximation processes in the presence of noise. $f(x) = \cos(x) + \cos(10000x) + \sin(10000x)$.

h	$f''(0)$ computed with (5.1)	$f''(0)$ computed with (5.3)
1	-0.99972265810015	-9.0117641983302
$\frac{1}{5}$	-0.99999997930242	-210.70216018290
$\frac{1}{10}$	-0.99999997930242	-278.74784641648
$\frac{1}{15}$	-0.9999999997170.	-2030.6779530997
$\frac{1}{20}$	-0.9999999999561	-281.95467246321
$\frac{1}{25}$	-0.9999999999402	-282.00495865726
$\frac{1}{30}$	-0.9999999999651	-3792.4478059765
$\frac{1}{100}$	-0.9999999999563	-4560.3573899134
$\frac{1}{200}$	-0.9999999999493	-2380.5832900073
$\frac{1}{400}$	-0.9999999999301	-5439.2695566901

6. CONCLUDING REMARKS

As noticed in Subsection 4.1 the wavelets needed to compute derivatives of other orders are built, similarly, by Richardson-like linear combinations of an appropriate derivative of the Gaussian function. Hence $f^{(3)}(x)$, say, could be approximated by the use of a wavelet $\Psi(x)$ satisfying

$$(6.1) \quad \int_{-\infty}^{\infty} t^k \Psi(t) dt = 0, \quad k = 0, 1, 2, 4, 5, 6, 7$$

and

$$(6.2) \quad \int_{-\infty}^{\infty} t^3 \Psi(t) dt = 6.$$

$f^{(3)}(x)$ would then be approximated as follows:

$$(6.3) \quad f^{(3)}(x) \approx \frac{1}{h^3} \int_{-\infty}^{\infty} \frac{1}{h} f(x-t) \Psi\left(\frac{t}{h}\right) dt.$$

Here, as in the case of $f''(x)$, it is important to find the optimal value of h , as a function of ϵ_M , that minimizes the error. The analysis is similar to that of Theorem 4.1.

However, as expected, the accuracy of the approximation decreases as the order of the derivative increases. Indeed, our experiments show that we cannot obtain more than four digits of accuracy once the order of the derivative is greater than five. Hence different filters will have to be built to restore high accuracy for the computation of high-order derivatives. However our efforts to build such filters have been unsuccessful.

As noticed by several authors [3], [28], numerical differentiation can be effectively used to detect jumps in the function or in its derivatives of low order (not necessarily of the same order of the approximated derivative). This is a problem of importance in many aspects of signal processing [8], [9] as well as in statistics [29]. We intend, similarly, to use our wavelet $\Psi_3(x)$ to detect singularities in the derivatives of low order of the signal. In this context it is the standard Littlewood-Paley theory, as presented in [13] and [16], that will help us determine the theoretical error estimates.

We plan to report these results in a paper presently in preparation.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Professor Ingrid Daubechies for fruitful discussions, to Professor Dick Gundy for advice and encouragement, and to Professor Juan Restrepo for editorial comments.

REFERENCES

1. Robert S. Anderssen and Markus Hegland, *For numerical differentiation, dimensionality can be a blessing!*, Math. Comp. **227** (1999), 1121–1141. MR1620207 (99i:65019)
2. E. Atkinson, *An introduction to numerical analysis*, Wiley, New York, 1989. MR1007135 (90m:65001)
3. Michèle Basseville and Igor Nikiforov, *Detection of abrupt changes: Theory and application*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1993. MR1210954 (95g:62153)
4. Carl M. Bender and Steven I. Orszag, *Advanced mathematical methods for scientists and engineers. I. Asymptotic methods and perturbation theory*, Springer-Verlag, New York, 1999. MR1721985 (2000m:34116)
5. Franca Calio, Marco Frontini, and Gradimir Milovanovic, *Numerical differentiation of analytic functions using quadratures on the semicircle*, Comput. Math. Appl. **10** (1991), 99–106. MR1136179 (92i:65047)
6. François Chaplais and Sylvain Faure, *Wavelets and differentiation*, Internal Report, Centre Automatique et Systèmes, École Nationale des Mines de Paris (2000), 1–14.
7. Ingrid Daubechies, *Ten lectures on wavelets.*, SIAM, Philadelphia, 1992. MR1162107 (93e:42045)
8. David L. Donoho, *Private communication*, 2003.
9. Dan E. Dudgeon and Russell M. Mersereau, *Multidimensional digital signal processing*, Prentice-Hall, Englewood Cliffs, N.J., 1984.
10. Sylvain Faure, *Implémentation sous simulux de la dérivation des signaux à partir des coefficients d'ondelettes*, Internal Report, Centre Automatique et Systèmes, École Nationale des Mines de Paris (1999), 1–12.
11. Michael Floater, *Error formulas for divided difference expansions and numerical differentiation*, J. Approximation Theory **122** (2003), 1–9. MR1976120 (2004b:65027)
12. Gerald B. Folland, *Introduction to partial differential equations. second edition*, Princeton, NJ, Princeton University Press, 1995. MR1357411 (96h:35001)
13. Michael Frazier, Bjorn Jawerth, and Guido Weiss, *Littlewood-Paley theory and the study of function spaces*, CBMS regional conference series in mathematics, 79, American Mathematical Society, Providence, RI, 1991. MR1107300 (92m:42021)

14. Martin Hanke and Otmar Scherzer, *Inverse problems light: numerical differentiation*, Amer. Math. Monthly **6** (2001), 512–521. MR1840657 (2002e:65089)
15. Wolfgang Hardle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov, *Wavelets, approximation, and statistical applications*. Lecture Notes in Statistics, 129., Springer-Verlag, New York, 1998. MR1618204 (99f:42065)
16. Eugenio Hernandez and Guido Weiss, *A first course on wavelets*, CRC Press, Boca Raton, FL, 1996. MR1408902 (97i:42015)
17. Peter Hoffman and K. C. Reddy, *Numerical differentiation by high order interpolation.*, SIAM J. Sci. Statist. Comput. **6** (1987), 979–987. MR0911068 (89a:65032)
18. Ian Knowles and Robert Wallace, *A variational method for numerical differentiation*, Numer. Math. **1** (1995), 91–110. MR1320703 (96h:65031)
19. Alexandru Lupac and Detlef H. Mache, *On the numerical differentiation*, Rev. Anal. Numer. Theor. Approx. **26** (1997), 109–115. MR1703928 (2000d:65039)
20. Malcon T. McGregor, *Numerical differentiation of analytic functions*, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. **2** (1991), 11–17. MR1162961 (93k:65020)
21. Yves Meyer, *Ondelettes et opérateurs*, Hermann, Paris, 1990. MR1085487 (93i:42002)
22. D. A. Murio, C. E. Meja, and S. Zhan, *Discrete mollification and automatic numerical differentiation*, Comput. Math. Appl. **5** (1998), 1–16. MR1612285 (99d:65066)
23. Alexander G. Ramm and Alexandra Smirnova, *On stable numerical differentiation*, Math. Comp. **70** (2001), 1131–1153. MR1826578 (2002a:65046)
24. Harold S. Shapiro, *Smoothing and approximation of functions*, Van Nostrand Reinhold Co., New York, 1969. MR0412669 (54:791)
25. M. R. Skrzipek, *Generalized associated polynomials and their application in numerical differentiation and quadrature*, Calcolo **40** (2003), 131–147. MR2025599
26. Robert S. Strichartz, *A guide to distribution theory and Fourier transforms*, CRC Press, Boca Raton, FL, 1994. MR1276724 (95f:42001)
27. O. L. Vinogradov and V. V. Zhuk, *Sharp estimates for errors of numerical differentiation type formulas on trigonometric polynomials. Function theory and partial differential equations*, J. Math. Sci. (New York) **105** (2003), 2347–2376. MR1855438 (2002f:65032)
28. Y. B. Wang, X. Z. Jia, and J. Cheng, *A numerical differentiation method and its application to reconstruction of discontinuity*, Inverse Problems **6** (2002), 1461–1476. MR1955897 (2004b:65086)
29. Yazhen Wang, *Jump and sharp cusp detection by wavelets*, Biometrika **82** (1995), 385–397. MR1354236 (96k:62120)
30. J. A. C. Weideman, *Numerical integration of periodic functions: a few examples*, Amer. Math. Monthly **109** (2002), 21–36. MR1903510 (2003c:65017)

PROGRAM IN APPLIED MATHEMATICS, THE UNIVERSITY OF ARIZONA, TUCSON, ARIZONA 85721-0089

E-mail address: hasson@math.arizona.edu