BULL. AUSTRAL. MATH. SOC. 65D25
VOL. 58 (1998) [223–232]

A STABLE FINITE DIFFERENCE ANSATZ FOR HIGHER ORDER DIFFERENTIATION OF NON-EXACT DATA

BOB ANDERSSEN, FRANK DE HOOG AND MARKUS HEGLAND

If standard central difference formulas are used to compute second or third order derivatives from measured data even quite precise data can lead to totally unusable results due to the basic instability of the differentiation process. Here an averaging procedure is presented and analysed which allows the stable computation of low order derivatives from measured data. The new method first averages the data, then samples the averages and finally applies standard difference formulas. The size of the averaging set acts like a regularisation parameter and has to be chosen as a function of the grid size h.

1. INTRODUCTION

Let the given (observational or non-exact) data be defined by

$$\mathbf{d} := \left\{ d_j = f(t_j) + \varepsilon_j;\ t_j = jh,\ h = 1/n,\ j = 0, 1, 2, \ldots, n \right\}, \tag{1}$$

where $f(t)$ denotes the underlying, but unknown, signal process and the ε_j denote the (observational or non-exact) errors which are assumed to be identical and independently distributed normal random variables with $E[\varepsilon_j] = 0$ and $E[\varepsilon_j \varepsilon_k] = \sigma^2 \delta_{jk}$, where δ_{jk} denotes the Kronecker delta function and $E[\cdot]$ the expectation operator.

In the numerical differentiation of non-exact data, the goal is to recover, from the given data $\mathbf{d}$, an estimate of some (lower order) derivative

$$f^{(p)}(t) := d^p f(t)/dt^p, \qquad p = 1,\ 2,\ \text{or } 3, \quad \text{(say)},$$

rather than f itself. Often, it is only the first derivative that is required. This situation has been examined in considerable detail in the literature, under the assumption that higher order differentiation is some natural generalisation of the results for first order differentiation. Clearly, even the fractional differentiation of non-exact data (see [2] and [8]) can, on occasions, be given a similar interpretation. However, though not incorrect, this assumption glosses over important practical details which are the focus of this paper.

Received 28th January, 1998

NOTATION AND ASSUMPTIONS. The notation $f_j^{[p]}$ will be used to denote the finite difference value of the p-th derivative evaluated at the grid point jh. The function f will be assumed to have the smoothness required by the formulas presented below in terms of the differentiation they involve.

2. THE FINITE DIFFERENCE ANSATZ FOR FIRST ORDER DIFFERENTIATION

Often, in the past (see [1]), and even today, data are differentiated using a ruler to obtain an estimate of the first derivative of f. It is fast and has a natural intuitive appeal. In fact, for the practitioner, who has just measured or calculated (graphically) some specific data the first derivative of which must be estimated before their interpretation is possible, the ruler approach represents a realistic alternative (see [3]). It is the reason why it is still used today, at least as a quick exploratory tool. However, its greatest drawback is (and was) its lack of objectivity in that the form of the first derivative determined by a particular individual will be influenced by their level of familiarity with the context within which that data have been derived.

Mathematically, numerical differentiation was initially seen as simply a special case of constructing finite difference approximations to derivatives. Consequently, the earliest methods proposed for the numerical differentiation of accurate numerical, though not necessarily exact, data were finite difference formulas. They were derived, in one way or another, through the manipulation of either the definition of a derivative or Taylor's expansion (theorem). They predate the computer (see [7]), and relate closely to the earliest ideas about the numerical approximation of derivatives in differential equations (see [9]). However, the interconnection is not clear cut. On the one hand, forward, backwards and central difference approximations take centre stage in the numerical solution of differential equations, whereas it is only the central (centred) difference formulas (or, equivalently, centred moving-averages)

$$(2) \qquad f_j^{[1]}[m] = \frac{d_{j+m} - d_{j-m}}{2mh}, \quad m = 1, 2, \ldots, \quad j = m, m+1, \ldots, n-m,$$

which are the key to the numerical differentiation of data. Normally, m is chosen to have the value 1, but the possibility of choosing a greater value has been implicitly examined by a number of authors within the context of optimising the choice of h (see [5, Section 7.1]). In part, the advantage of a central difference formula, over the alternatives such as forward and backwards differences, is the associated higher order of convergence, which, in turn, can be explained algebraically and graphically as a practical realisation of the mean value theorem. In fact, the ruler differentiation simply corresponds to an analogue realisation of the mean value theorem applied directly to the data.

Statistically, the approach adopted was quite different. Whereas the finite difference formula approach more or less determines the derivatives at the data points directly and explicitly from the data, the statistical approach is indirect. Here, one first estimates

statistically the parameters β in an assumed parametric estimate $\hat{f}(t;\beta)$ of $f(t)$, and then estimates $f^{(1)}(t)$ as

$$\hat{f}^{[1]}(t;\beta).$$

This statistical approach leads naturally to the following two important generalisations:

(a) NON-PARAMETRIC DIFFERENTIATION. This is simply the non-parametric counterpart to the parametric procedure outlined above, where one replaces the specific choice of a parametric model for $f(t)$ by a non-parametric functional characterising its structure. For example, a popular choice for the non-parametric functional is the least squares smoothing spline criterion (see [**10**])

$$f_\alpha(t) = arg\Big\{ \min_{f\in\mathcal{H}_1} \Big[\sum_{j=0}^{n} \big(f(t_j) - d_j\big)^2 + \alpha \int_{t_0}^{t_n} \big(d^2 f(t)/dt^2\big)^2 dt \Big] \Big\}, \tag{3}$$

where $\mathcal{H}_1$ denotes the Sobolev space of absolutely continuous first derivatives.

(b) FOURIER-WIENER DIFFERENTIATION. If it is assumed that the data has been generated by a stationary stochastic process, then discrete Fourier analysis and Wiener filtering can be applied directly to the data to recover an estimate of the (first) derivative of $f(t)$ (see [**3**]).

Though the major emphasis in [**3**] was on the implementation of numerical differentiation as a Wiener filtering process, they showed how, for given data, the Wiener filtering theory could be used to construct a type of centred moving-average (local differentiator) for performing the differentiation. In essence, these local differentiators are simply central difference formulas. However, the utility of such formulas does not appear to have been pursued in any great detail.

In an independent study, [**4**] analysed the stability properties of multi-point finite difference differentiators of the form

$$f_j^{[1]} = f^{[1]}(t_j) = \sum_{k=-r}^{r} W_k d_{j+k}, \tag{4}$$

where the W_k, $k = -r,\ -r+1,\ \ldots,\ r-1,\ r$, denote appropriately chosen weights. They first observed that, if the weights satisfy

$$W_k = -W_{-k}, \quad k = 0,\ 1,\ 2,\ \ldots,\ r,$$

which implies that $W_0 = 0$, then the multipoint finite difference differentiators (4) are exact for constant data, and can be rewritten as the following sum of the central difference differentiators $f_j^{[1]}[k]$

$$f_j^{[1]} = \sum_{k=1}^{r} w_k f_j^{[1]}[k], \quad W_k = w_k/(2kh), \quad k = 1,\ 2,\ \ldots,\ r, \tag{5}$$

https://doi.org/10.1017/S0004972700032196 Published online by Cambridge University Press

where

$$\sum_{k=1}^{r} w_k = 1. \tag{6}$$

In part, the goal of that paper was to show that the stabilisation of such formulas was controlled by its length r. In fact, it was established that the choice of the length r must be related to the size of the step-length h so that r increases appropriately as h decreases. Consequently, under such circumstances, the differentiator (5) can be given a regularisation interpretation in which the length r and the weights w_j play, respectively, the role of the regularisation parameter and the regularisation. A formal characterisation of this fact was also derived.

In a spline-type context, such results can be formalised using mollification (see [8]).

Pragmatically, this result yields a natural *ansatz* for the construction of finite difference formulas for the stabilised numerical differentiation of one-dimensional observational or non-exact data; namely,

THE FINITE DIFFERENCE ANSATZ FOR FIRST ORDER DIFFERENTIATION.

"Choose as the local differentiator, a weighted sum of the central difference differentiators $f_j^{[p]}[m]$ so that, when it is applied to the data as a centred moving-average, an appropriately smooth estimate of the derivative $f^{(p)}(t)$ of the signal $f(t)$ results."

The purpose of this paper is an examination of the applicability of this *ansatz* to higher order differentiation, when, for a fixed m, the averaging is performed with respect to j. A natural motivation for this approach can be based on the advantages of performing repeated measurements in a statistical analysis. In fact, if, relative to the smoothness of f, the size of h is very small, then, for small r, the values of

$$f_{j+l}^{[1]}[m], \qquad l = -r,\ -r+1, \ldots,\ r-1,\ r,$$

can be viewed as repeated measurements of $f_j^{[1]}[m]$.

3. THE FINITE DIFFERENCE OPERATORS

Let:

(a) D^p denote the differential operator of order p defined by

$$D^p = \frac{d^p}{dt^p}, \qquad p = 1,\ 2,\ \ldots.$$

(b) I correspond to the unit interval $[0,1]$ on IR.

https://doi.org/10.1017/S0004972700032196 Published online by Cambridge University Press

(c) f be a real-valued function defined on I with sufficient regularity such that

$$D^p f(t) = \frac{d^p f}{dt^p}(t),$$

exists.

(d) G_h denote the uniform grid of points

$$t_j = jh, \qquad j = 0,\ 1,\ 2, \ldots,\ n, \qquad h = 1/n.$$

(e) $\mathbf{f} = f_{G_h}$ denote the restriction of f to the grid G_h.

(f) $\Delta^{(p)}_{h,m}$ denote a family, parameterised by m, of difference operators, defined in terms of their action on $\mathbf{f}$, which approximate $(D^p f)(ih)$, $i \in I$, in the sense that

$$\Delta^{(p)}_{h,m}\mathbf{f}(ih) = (D^p f)(ih) + O\big((mh)^2\big), \tag{7}$$

where $\Delta^{(p)}_{h,m}\mathbf{f}(ih)$ only acts on the grid values $f_{i\pm mj}$ for $j = 0,\ 1,\ 2,\ \ldots$.

For the first and second derivatives, the natural examples of such second order finite difference formulas are, respectively,

$$\Delta^{(1)}_{h,m}\mathbf{f}(ih) = \frac{f(ih+mh) - f(ih-mh)}{2mh}, \tag{8}$$

$$\Delta^{(2)}_{h,m}\mathbf{f}(ih) = \frac{f(ih+mh) - 2f(ih) + f(ih-mh)}{(mh)^2}. \tag{9}$$

NOTE. Clearly, for a given j, there will be an upper bound on the value of m which guarantees that $0 \leqslant (j-m)h < (j+m)h \leqslant 1$. However, this is a rather technical matter which can be circumvented by assuming that, with respect to a given choice of j, the value of n is such as to generate a sufficiently fine grid which guarantees the application of any particular formula considered. In other words, it is assumed that one has sufficient data to perform the relevant operations examined and discussed below. Such situations occur naturally in situations where the data is collected by a computer in an on-line monitoring scenario.

3.1 The Averaged Finite Difference Formulas

Let

$$f_i^{[p]}[m] = \Delta^{(p)}_{h,m}\mathbf{f}(ih). \tag{10}$$

Here, we examine the following averaging of these finite difference differentiation formulas

$$\overline{f_i^{[p]}}[m] = \frac{1}{2r+1}\sum_{j=-r}^{r} f_{i+j}^{[p]}[m]. \tag{11}$$

The repeated measurement interpretation follows from the fact that

$$\overline{f_i^{[p]}}[m] = \overline{f}_i^{[p]}[m]; \tag{12}$$

https://doi.org/10.1017/S0004972700032196 Published online by Cambridge University Press

That is, the above averaging of the finite difference formulas corresponds to the application of the chosen finite difference formula (for a fixed m) to the averaging of the data at the grid points $(i+m)h$, ih and $(i-m)h$ by the formula

$$\overline{f}_i h = \frac{1}{2r+1} \sum_{j=-r}^{r} f_{ih+j}, \tag{13}$$

where ih corresponds to $(i+m)h$, im and $(i-m)h$, respectively.

In order to guarantee that the errors generated by the application of the numerical differentiation formulas (10) to the observational data $\{d_j\}$ remain uncorrelated, one must ensure that, with respect to a given r, the value of m is suitably large. The simplest strategy is to replace m in (10) by $kr+1$ to obtain

$$\overline{f}_i^{[p]}[kr+1] = \frac{1}{2r+1} \sum_{j=-r}^{r} f_{i+j}^{[p]}[kr+1], \tag{14}$$

and to constrain k to satisfy $k \geqslant 2$. If $m = 2r+1$, then every consecutive point about the grid point ih is utilised in the evaluation of (13). This clearly represents the most efficient use of the data.

4. Convergence and Stability

The proof of convergence and stability exploits the data-averaging duality of the formula (11). Because the proof for odd order derivatives can be constructed in a similar manner, attention is restricted to the situation where p is an even integer $2q$. In fact, since, for suitably smooth functions f,

$$f_{i\pm j} = f_i \pm jhf_i^{(1)} + \frac{(jh)^2}{2!} f_i^{(2)} + \sum_{l=3}^{p+1} \frac{(\pm 1)^l (jh)^l}{l!} f_i^{(l)} + O\left((jh)^{p+2}\right), \tag{15}$$

it follows that, with $p = 2q$,

$$\overline{f}_i = \frac{1}{2r+1} \sum_{j=-r}^{r} f_{i+j} = f_i + \frac{r(r+1)h^2}{12} f_i^{(2)} + \sum_{l=2}^{q} P_{2l}(r) h^{2l} f_i^{(2l)} + O\left((rh)^{2q+2}\right), \tag{16}$$

where the $P_{2l}(r)$ denote polynomials of degree $2l$ in r.

Therefore, it follows that

$$\begin{aligned} \overline{f_i^{[p]}}[kr+1] &= \overline{f}_i^{[p]}[kr+1] \\ &= \Delta_{h,kr+1}^{(p)} \mathbf{f}(ih) + \frac{r(r+1)h^2}{12} \Delta_{h,kr+1}^{(p)} \mathbf{f}^{(2)}(ih) \\ &\quad + \sum_{l=2}^{q} P_{2l}(r) h^{2l} \Delta_{h,kr+1}^{(p)} \mathbf{f}^{(2l)}(ih) + O\left(\frac{(rh)^{2q+2}}{((kr+1)h)^{2q}} \right). \end{aligned} \tag{17}$$

https://doi.org/10.1017/S0004972700032196 Published online by Cambridge University Press

If one recalls that

$$\Delta^{(p)}_{h,kr+1}\mathbf{f}^{(2l)}(ih) = f_i^{(2l+p)} + K((kr+1)h)^2 f^{(2l+p+2)}(\zeta), \qquad K = \text{constant},$$

where the value of the constant K depends on the explicit form of the finite difference formula used and ζ denotes the appropriate mean value, this last result becomes

$$\begin{aligned}\overline{f}_i^{[p]}[kr+1] &= f_i^{(p)} + K((kr+1)h)^2 f^{(p+2)}(\zeta_2) \\ &\quad + \frac{r(r+1)h^2}{12} f_i^{(p+2)} + \frac{Kr(r+1)h^2(krh)^2}{12} f^{(p+4)}(\zeta_4) \\ &\quad + \sum_{l=2}^{q} P_{2l}(r)h^{2l}\left(f_i^{(p+2l)} + K((kr+1)h)^2 f^{(p+2l+2)}(\zeta_{2l+2})\right) \\ &\quad + O\left(\frac{(rh)^{2q+2}}{((kr+1)h)^{2q}}\right), \end{aligned} \tag{18}$$

where the ζ_2, ζ_4, et cetera denote the appropriate mean values.

On recalling that (see [6, Section 5.8])

$$\frac{1}{2r+1}\sum_{j=-r}^{r} \varepsilon_{i+j} = \frac{1}{\sqrt{2r+1}}\eta_i,$$

where the η_i denote identically distributed Gaussian random variables with zero mean and variance σ^2, it follows that, because $k \geqslant 2$, the $\eta_{i+(kr+1)j}$, for fixed i and $j = 0,\ 1,\ 2,\ \ldots$, are independent and

$$\Delta^{(p)}_{h,kr+1}\mathbf{d}(ih) = \overline{f}_i^{[p]}[kr+1] + \frac{\overline{K}}{\sqrt{(2r+1)}}((kr+1)h)^{-p}\eta_i, \tag{19}$$

where the value of $\overline{K}$ depends on the nature of the finite difference formula chosen.

It follows from (18) and (19) that $\overline{f}_i^{[p]}[kr+1]$ converges to $f_i^{(p)}$, if

(i) $krh \rightarrow 0$,

(ii) $rh \rightarrow 0$,

(iii) $r^{1/2}(krh)^p \sim 1$.

If it is assumed that $krh = h^s$, then, from (iii), one obtains that $r \sim h^{-2ps}$. But, (ii) then implies that h^{1-2ps} must tend to zero, and thus that convergence is guaranteed if

$$2ps < 1, \qquad k \text{ finite}. \tag{20}$$

The size of the perturbation in the actual values obtained is clearly controlled by the value of k. In particular, the larger k the smaller the perturbation errors. Thus, for a very stable solution, one requires, as well as (i), (ii) and (iii), that

https://doi.org/10.1017/S0004972700032196 Published online by Cambridge University Press

(iv) $k \to \infty$.

If it is again assumed that $krh = h^s$, then it now follows that convergence is guaranteed if

$$(2p+1)s < 1. \tag{21}$$

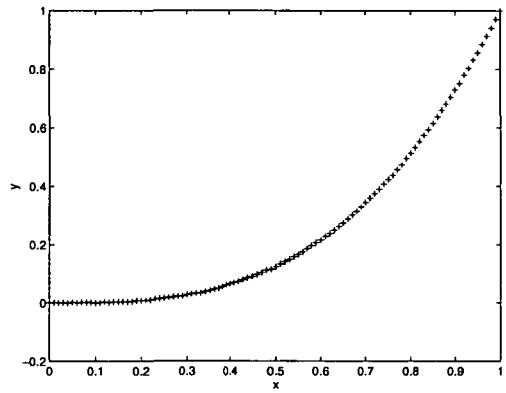

Figure 1: Data $y = x^3 + \varepsilon_x$

5. Implementation and Exemplification

Once, for a given p, the form of the local differentiator has been chosen, implementation reduces to choosing the values of k and r. Clearly, the actual choice of k and r will depend of the nature of the observational data.

The efficiency of the proposed differentiator shall now be demonstrated on synthetic data. Let $f(x) = x^3$ and the standard deviation be $\sigma = 0.001$ such that the data is

$$y_i = (ih)^3 + \varepsilon_i. \tag{22}$$

The data is plotted for $h = 0.01$ in Figure 1.

The second derivative for this example is $f^{(2)}(x) = 6x$. In Figure 2 the values of the second differences are plotted. One sees that the variance is huge such that no trend can be detected even thogh the the synthetic data is quite precise.

Using the new differentiator with $r = 4$ and $k = 2$ (which is minimal) one obtains the approximations for $f^{(2)}$ displayed in Figure 3. There is now an obvious trend and the points are fairly close to the expected points. However, due to the sampling procedure, there are many fewer data points in the derivative. But an interpolation of these points still gives a good result in this case. When the original data has higher curvature one might require more data points to reconstruct the derivative.

https://doi.org/10.1017/S0004972700032196 Published online by Cambridge University Press

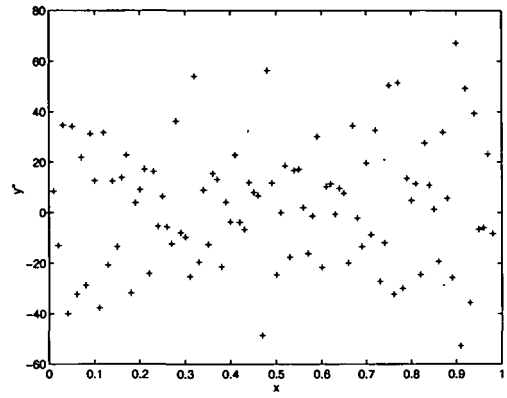

Figure 2: Central difference approximation of 2nd derivative with measurement error

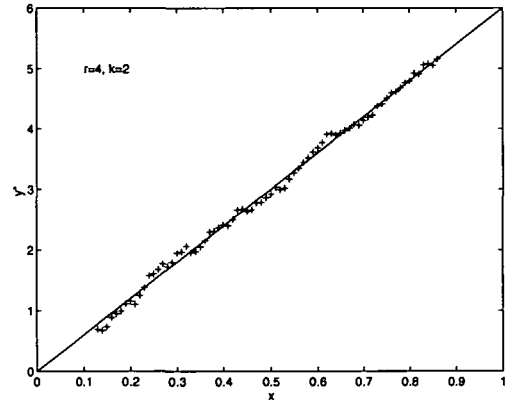

Figure 3: Central difference approximation of the averaged function

References

[1] Chr. Andersen, 'The ruler method - an examination of a method for numerical determination of Fourier coefficients', *Acta Polytech. Scand. Math. and Comput. Mach. Ser.* **8** (1963), 19–73.

[2] R.S. Anderssen, 'Stable procedures for the inversion of Abel's equation', *J. Inst. Math. Appl.* **17** (1976), 329–342.

[3] R.S. Anderssen and P. Bloomfield, 'Numerical differentiation procedures for non-exact data', *Numer. Math.* **22** (1974), 157–182.

[4] R.S. Anderssen and F.R. de Hoog, 'Finite difference methods for the numerical differentiation of non-exact data', *Computing* **33** (1984), 259–267.

[5] S.D. Conte and C. de Boor, *Elementary numerical analysis: an algorithmic approach*, (2nd ed.) (McGraw-Hill, New York, 1972).

[6] D.J. Finney, *Statistics for mathematicians: an introduction* (Oliver and Boyd, Edinburgh, 1968).

[7] D.R. Hartree, *Numerical analysis* (Clarendon Press, Oxford, 1952).

[8] M. Hegland and R.S. Anderssen, *A mollification framework for improperly posed problems*, Mathematics Research Report MRR 085-95 (CMA, Australian National University, Canberra, Australia) (submitted).

[9] L.F. Richardson, 'The approximate arithmetic solution by finite difference of physical problems involving differential equations with an application to the stresses in a masonary dam', *Philos. Trans. Roy. Soc. London Ser. A* **210** (1910), 307–357.

[10] G. Wahba, *Spline models for observational data* (SIAM, Philadelphia, PA, 1990).

Division of Mathematics and Statistics
CSIRO
GPO Box 1965
Canberra ACT 2601
Australia

Division of Mathematics and Statistics
CSIRO
GPO Box 1965
Canberra ACT 2601
Australia

Computer Sciences Laboratory
RSISE
Australian National University
Canberra ACT 0200
Australia