Numer. Math. 22, 157—182 (1974)

Numerical Differentiation Procedures for Non-Exact Data*

R. S. Anderssen and P. Bloomfield

Received September 5, 1972

Abstract. The numerical differentiation of data divides naturally into two distinct problems:

(i) the differentiation of exact data, and

(ii) the differentiation of non-exact (experimental) data.

In this paper, we examine the latter. Because methods developed for exact data are based on abstract formalisms which are independent of the structure within the data, they prove, except for the regularization procedure of Cullum, to be unsatisfactory for non-exact data. We therefore adopt the point of view that satisfactory methods for non-exact data must take the structure within the data into account in some natural way, and use the concepts of regression and spectrum analysis as a basis for the development of such methods. The regression procedure is used when either the structure within the non-exact data is known on independent grounds, or the assumptions which underlie the spectrum analysis procedure [viz., stationarity of the (detrended) data] do not apply. In this latter case, the data could be modelled using splines. The spectrum analysis procedure is used when the structure within the non-exact data (or a suitable transformation of it, where the transformation can be differentiated exactly) behaves as if it were generated by a stationary stochastic process. By proving that the regularization procedure of Cullum is equivalent to a certain spectrum analysis procedure, we derive a fast Fourier transform implementation for regularization (based on this equivalence) in which an acceptable value of the regularization parameter is estimated directly from a time series formulation based on this equivalence. Compared with the regularization procedure, which involves $O(n^3)$ operations (where n is the number of data points), the fast Fourier transform implementation only involves $O(n \log n)$.

1. Introduction

The tacit assumptions which underlie most numerical differentiation procedures proposed to date are more or less equivalent to the assumption that the data to be differentiated is exact. This is a direct consequence of the use of abstract concepts as the basis for their formulation. For example, many authors have examined numerical differentiation under the assumption that the step-size (between data points) is large compared with the accuracy of the data. They

* This research was performed with the support of

(a) the Office of Naval Research under contract 0014-67-A-0151-0017

and

(b) the U.S. Army Research Office — Durham, under contract DA-31-124-ARO-D-215.

An earlier version appeared in preprint form as Technical Report 13, Series 2, Department of Statistics, Princeton University, April, 1972.

force this assumption by regarding the data to be exact, and therefore, circumvent the difficulty that the numerical processes defined by these abstract procedures are potentially unstable.

A number of authors have recognized that numerical differentiation procedures do lead to unstable numerical processes and have proposed different procedures for their stabilization. These proposed procedures can be classified as either regularization, optimal step-size or analytic function techniques. However, except for the regularization procedure of Cullum [13], they do not come to grips with the problem posed by real (non-exact) data which arise in experimental situations. Since the explanation lies in the formulation of these methods on the basis of some abstract concept which ignores the source and the nature of the given data, we develop numerical differentiation procedures which take the structure within the data into account. In fact, we develop two statistical procedures. In §3 we sketch briefly the use of regression analysis, while in §4—5 we discuss in detail the use of time series.

One of the main advantages of these two methods is that they allow the curve (signal) from which the derivative has been derived to be constructed. This is then available to the experimenter for comparison with the original data. It represents a direct but not necessarily conclusive check (from the experimentalist's point of view) on the reliability of the numerical derivative obtained.

In §6, we show that the regularization procedure of Cullum [13] is equivalent to a certain spectrum analysis procedure, and use this to

(i) derive a direct estimation procedure for the optimum value of the regularization parameter of the Cullum procedure, and

(ii) construct an automatic computational procedure for numerical differentiation based on a fast Fourier transform implementation of the spectrum analysis equivalence of regularization.

The exploratory nature of the spectrum analysis method is preserved by using a comparison of the data and the reconstructed signal to decide whether the given data can meaningfully support a differentiation. Compared with the regularization procedure, which involves $O(n^3)$ operations (where n is the number of data points), the fast Fourier transform implementation only involves $O(n \log n)$.

A cursory survey of methods based on abstract formalisms, including Cullum's regularization procedure, is given in §2, while the application of the spectrum analysis procedure to synthetic and experimental data is discussed in §8. The order of convergence of the spectral differentiation procedure is established in §7.

The practical importance of differentiating non-exact data is not discussed in this paper. The interested reader is referred to Anderssen and Bloomfield [1].

2. Methods Based on an Abstract Formalism Including Regularization

The abstract formalisms on which numerical differentiation methods have been based can be classified as:

(i) Polynomial Interpolation: see Bickley [3], Brodskii [8], Chakravarti [10], Kranzer [30], Salzer [39, 40] and Schönhage [41].

(ii) Taylor Series and van der Monde Systems: see Björck and Pereyra [5], Galimberti and Pereyra [17], Hunter [27] and Lyness and Moler [34].

(iii) Spline Interpolation: see Birkhoff and de Boor [4] and Secrest [42].

(iv) Contour Integration: see Lyness [31, 32] and Lyness and Moler [33].

(v) Integral Equations and Regularization: As well as illustrating the improperly posed nature of differentiation, the integral equation formulation

$$g(t)=\int_0^t f(s)\,ds, \qquad f(t)=g'(t), \qquad 0\leq s\leq t\leq T, \tag{2.1}$$

can be used to construct finite difference and stabilization procedures for exact and non-exact data, and be used to throw further light on the general nature of differentiation formulas. The general properties of finite difference methods for (2.1) have been examined by de Hoog and Weiss [25, 26], while different regularization procedures have been proposed by Cullum [12, 13], Dolgopolova and Ivanov [14], and Vedeneev and Zhidkov [46].

(vi) Others such as strict estimation in Strom [43], Richardson extrapolation in Engels [16] and random sampling in multidimensional space in Tsuda [44].

We refer to formulas generated with respect to the totality of data points as *global* and reserve the title *local* for formulas generated with respect to a small fixed number M of data points and applied sequentially.

A careful examination and comparison of these methods indicate that:

(i) The most satisfactory method for exact data, when $g(t)$ is known analytically, is the contour integration procedure of Lyness and Moler.

(ii) When the data is exact, but $g(t)$ is not known analytically, the local interpolatory formulas are the most satisfactory (because of their simplicity) as long as the steplength between data points remains large with respect to the rounding error but small with respect to the local variation in $g(t)$. When this is not the case or cannot be guaranteed *a priori*, then a more sophisticated method such as regularization should be used.

(iii) Except for regularization, with an acceptable value for the regularization parameter, none of the above methods is satisfactory for non-exact (experimental) data.

Because it plays a fundamental role in the subsequent analysis, we conclude this section with a brief discussion of the regularization procedure of Cullum [12, 13].

Without loss of generality, we assume that $[0, T]=[0, 1]$ and $g(0)=g(1)=0$. In addition, we introduce the norms

$$\|g\|_2^2=\int_0^1 g^2(x)\,dx,$$

$$\|g\|_4^2=\|g\|_2^2+\|g^{(1)}\|_2^2,$$

and denote the space of twice differentiable functions with second derivatives square summable by W_1^2. Cullum first observes that the Volterra integral equation

formulation (2.1) can be replaced by the Fredholm formulation

$$g(t)=\int_0^1 H(t-s)\,f(s)\,ds=\mathbf{A}f(t), \tag{2.2}$$

with $H(t-s)$ the Heaviside unit step function. Next, by applying regularization in the sense of Tikhonov, the solution of (2.2) is reduced to the following family of optimization problems $P(\alpha)$ $(0<\alpha\leq 1)$: for each α minimize

$$C(\alpha, f)\equiv\|\mathbf{A}f-g\|_2^2+\left(\int_0^1 f(x)\,dx\right)^2+\alpha\,\|f\|_4^2 \tag{2.3}$$

with respect to $f\in W_1^2$. The *regularization parameter* is α. Using a calculus of variations' argument (see Hestenes [24]) the determination of the $f(x)$ which minimizes (2.3) for a fixed α is reduced to the solution of the second kind Fredholm equation

$$\alpha f(t)+\int_0^1 K_0(t, s)\,f(s)\,ds=m(t) \qquad (0\leq t\leq 1) \tag{2.4}$$

in which the kernel $K_0(t, s)$ is computable {see Cullum [13; §4]}. Since Cullum [13] proves that $P(\alpha)$ is well-posed for $0<\alpha\leq 1$, $f(t)$ can be determined from (2.4) by using finite difference methods based on the replacement of the integrals by suitable quadrature formulas.

The only disadvantage of this method is the choice of an acceptable value for the regularization parameter, since the problem of its choice has not previously been formalised.

3 The Regression Procedure

We now turn to the examination of the two independent statistical procedures for the numerical differentiation of non-exact data. The choice of one in favour of the other depends to a certain extent on the nature of the data. From the point of view of the present investigation, we can delineate three (not mutually exclusive) possibilities:

1. A structure for the information in the non-exact data is known from theory or an independent analysis. In this case, we can use standard regression theory to fit a model of the known structure which is then differentiated. If the accuracy of the data is high and only one derivative is required, then this *regression procedure* can be applied directly. If the data is such that it can be transformed to yield a stationary time series {see Box and Jenkins [7] and Hannan [22, 23]} and the accuracy of the data is poor, then use should first be made of a spectrum analysis of the transformed data {see Jenkins and Watts [28] and Grenander and Rosenblatt [19]} to see whether it can meaningfully support a numerical differentiation.

2. An underlying structure for the data is unknown, but it can be transformed to yield a stationary time series. In this case the spectral analysis procedure of §§4, 5 should be used. The advantage of this procedure over regression (when both apply) is that it generates the required n-th derivative $(n\geq 1)$ directly, thus minimizing the enhancement of rounding error effects.

3. An underlying structure for the data is unknown and the data cannot be transformed to yield a stationary time series. In this case, it is necessary to fall back on the use of the regression procedure. Now, however, a set of basis functions such as polynomials or splines, must be chosen for the representation of the structure within the data. Numerically, this raises many difficult questions, and therefore, will not be persued in detail in this paper.

In this section, we examine the regression procedure. A discussion of the spectral procedure is deferred to subsequent sections.

Let

$$\{(g_n, t_n);\, n=0, 1, \ldots, N\}=\{(g_0, t_0), (g_1, t_1), \ldots, (g_N, t_N)\} \tag{3.1}$$

represent the non-exact data which it is necessary to differentiate with

$$0\leq t_0\leq t_1\leq \cdots \leq t_N=T.$$

It is assumed that

(i) $g=g(t)$ is unknown,

(ii) the t_n $(n=0, 1, \ldots, N)$ are known exactly without error,

(iii) the errors in the observations g_n $(n=0, 1, \ldots, N)$ are independent normal random variables with mean zero and variance σ_1^2,

(iv) on the basis of either *a priori* knowledge or the choice of a set of basis functions, it is assumed that $g=g(t)$ can be modelled as the linear form

$$g(t)=a_1\,\phi_1(t)+a_2\,\phi_2(t)+\cdots+a_M\,\phi_M(t)+\tilde{\varepsilon} \tag{3.2}$$

with $M\ll N$, the $\phi_m(t)$ $(m=1, 2, \ldots, M)$ known and linearly independent on $[0, T]$, the $\tilde{\varepsilon}$ a normally distributed stochastic variable with zero mean and variance σ_2^2 and the a_m $(m=1, 2, \ldots, M)$ unknown.

When incorporated in (3.2), the data (3.1) yields the over determined linear system

$$\sum_{m=1}^{M} x_{nm}\,a_m=g_n+\varepsilon_n \qquad (n=0, 1, 2, \ldots, N), \tag{3.3}$$

where $x_{nm}=\phi_m(t_n)$, and the ε_n are random variables with mean zero and variance $\sigma^2=\sigma_1^2+\sigma_2^2$. Using matrix notation, (3.3) can be rewritten as

$$X\boldsymbol{a}=\boldsymbol{g}+\boldsymbol{\epsilon}. \tag{3.4}$$

If, under the above assumptions, it is assumed that the model (3.2) is correct with regard the choice of the ϕ_m and M, then the maximum likelihood estimator $\hat{\boldsymbol{a}}$ for $\boldsymbol{a}$ coincides with the least squares estimator defined by

$$X^T X\hat{\boldsymbol{a}}=X^T\boldsymbol{g}, \tag{3.5}$$

and is normally distributed with mean $\hat{\boldsymbol{a}}$ and variance $\sigma^2(X^TX)^{-1}$.

In addition, the statistic $\hat{\boldsymbol{a}}$ is distributed independently of the defining relationship for the least squares estimator $\hat{\sigma}^2$, of σ^2, viz.

$$N\hat{\sigma}^2=\boldsymbol{g}^T\left(I-X\left(X^TX\right)^{-1}X^T\right)\boldsymbol{g}. \tag{3.6}$$

Since it is known that $\{N\hat{\sigma}^2/\sigma^2\}$ has a X^2_{N-M} distribution {see Mood and Graybill [35; §§13.6–13.7]}, a $(1-\alpha)$ 100% confidence interval for a_m is given by

$$\hat{a}_m \pm t_{N-M}(\alpha)\left\{\frac{N\hat{\sigma}^2}{N-M}[X^TX]^{-1}_{m,m}\right\}^{\frac{1}{2}} \tag{3.7}$$

where $t_{N-M}(\alpha)$ is the two-tail α-point of the t-distribution with $N-M$ degrees of freedom, and $[X^TX]^{-1}_{m,m}$ denotes the m-th diagonal element of $[X^TX]^{-1}$.

It follows from (3.2) that $f(t)=\dot{g}(t)=dg(t)\,dt$ can be modelled as the linear form

$$\dot{g}(t)=a_1\dot{\phi}_1(t)+a_2\dot{\phi}_2(t)+\cdots+a_M\dot{\phi}_M(t). \tag{3.8}$$

When incorporated in (3.8), the unknown data

$$\{(f_n, t_n);\ f_n=\dot{g}(t_n),\ n=0, 1, 2, \ldots, N\} \tag{3.9}$$

yields the overdetermined linear system

$$A\boldsymbol{a}=\boldsymbol{f} \tag{3.10}$$

with $A_{nm}=\phi_m(t_n)$. Thus, the estimation of the numerical derivative $\dot{g}(t_n)$ $(n=0, 1, 2, \ldots, N)$ reduces to a special case of the general statistical problem: *for any matrix* A, *estimate the linear form*

$$\hat{\boldsymbol{f}}=A\hat{\boldsymbol{a}}$$

defined over the regression coefficients of the linear overdetermined system (3.4).

Since, as noted above, $\hat{\boldsymbol{a}}$ is normally distributed it follows that $\hat{\boldsymbol{f}}=A\hat{\boldsymbol{a}}$ must be also. Thus, the substitution of $\hat{\boldsymbol{a}}$ in (3.10) for $\boldsymbol{a}$ is consistent with the situation where the data (3.9) is actually observed, with the errors in $\boldsymbol{f}$ normally distributed, and the $\hat{\boldsymbol{a}}$ is calculated directly as the least squares estimator. In fact, $\hat{\boldsymbol{f}}$ is normally distributed with mean $A\boldsymbol{a}$ and variance $\sigma^2A(X^TX)^{-1}A^T$ and is independent of σ^2. Consequently, a $(1-\alpha)$ 100% confidence interval for the m-th element $\hat{f}_n$ of $\hat{\boldsymbol{f}}$ becomes

$$\hat{f}_n \pm t_{N-M}(\alpha)\left[\hat{\sigma}^2[A(X^TX)^{-1}A^T]_{n,n}\right]^{\frac{1}{2}}. \tag{3.11}$$

This includes the special case $X=A$ when estimates of $\hat{g}_n$ $(n=0, 1, 2, \ldots, N)$ are required.

The actual numerical evaluation of $\hat{\boldsymbol{a}}$, $\hat{\sigma}^2$ and $[X^TX]^{-1}$ should be based on orthogonal factorization or modified Gram-Schmidt orthogonalization procedures in order to minimize the enhancement of rounding error. The reader is referred to Wilkinson and Reinsch [48] as a source reference.

For the given data (3.1), accepting that (ii) and (iii) and consequently the use of the least squares estimator can be justified, the validity of the above rests heavily on the validity of (iv). This depends on the assumption that

(a) a realistic and meaningful model for the data is given by (3.2),
and

(b) only a finite number M of terms of (3.2) are required to define a "safe" model.

Though statistical verification of (a) and (b) cannot be obtained directly, there exist a number of independent techniques which yield indirect evidence as to their possible validity.

A comparison of $\hat{g}(t_n)$ with g_n $(n=0, 1, 2, \ldots, N)$ and the size of the corresponding confidence intervals, defined by (3.11) with $A=X$, yield a first order check on (a). A poor comparison and large confidence intervals imply that (3.11) is inappropriate, if it is clear that they do not arise as a direct consequence of inaccurate data, or the choice of $\phi_m(t)$ $(m=1, 2, \ldots, M)$, which leads to an unstable numerical process for the evaluation of $\hat{\boldsymbol{a}}$, $\hat{\sigma}^2$ or $[X^T X]^{-1}$.

A test that a model with a smaller choice of M than the "safe" choice will adequately describe the situation {i.e. that some of the coefficients a_m in the model (3.2) may be taken as zero} can be used to examine the validity of (b). Such a test has been discussed in Anderssen and Seneta [2; p. 160].

When the choice of the model (3.2) can be predicted with assurance from prior information, the testing of the validity of (b) complements the analysis, and the implementation of the least squares method as a tool for numerical differentiation is straight forward if the numerical process so defined is stable. However, when the choice of the model (3.2) is very much an open question, the testing of (a) and (b) represents a check on the suitability of any choice made.

In such situations, this leaves the general choice of the $\phi_m(t)$ $(m=1, 2, \ldots, M)$ unanswered. In general, one would not choose

$$\phi_m(t)=t^{m-1} \qquad (m=1, 2, \ldots, M)$$

since in this case $X^T X$ coincides with the ill-conditioned Hilbert matrix {see Ralston [38; Chapter 6]}. In fact, the choice should be based on numerical considerations. The potential instability of global and local differentiation formulas based on polynomial interpolation more or less rules out the use of globally defined polynomials, unless some suitable orthogonality condition holds. Since this is a rather specialized assumption which has its own problems, it will not be examined here.

A case for the use of splines can be based on the results of Birkhoff and de Boor [4] and Powell [37]. However, this will not be pursued here. We only pause to note that the implementation of a least squares cubic spline procedure is a non-trivial task which has been discussed in the literature {see, for example, Burchard [9], de Boor and Rice [6] and Powell [37]}.

4 The Time Series Approach

Consider a set of observations, viz. a *time series*,

$$\{g_k; k=0, 1, 2, \ldots, N\} \tag{4.1}$$

of some phenomenon at the evenly spaced time points $t_k=k\Delta$ $(k=1, 2, \ldots, N)$, where Δ denotes a constant steplength. We say that a time series is statistical (as opposed to deterministic), if its future values can be described only in terms of a probability distribution. Further, if a statistical phenomenon evolves in time according to probability laws, then it is referred to as a *stochastic process*. Thus, we can regard a time series as a realization of a stochastic processes, the

properties of which we wish to investigate on the basis of information contained in the time series.

The concept of a *stationary (stochastic) process* is based on the assumption that the process is in a particular state of *statistical equilibrium*. In fact, we say that a stochastic process is *strictly stationary* if its properties are unaffected by a change of time origin. Thus, for a discrete process to be strictly stationary, the joint distribution of any set of $\bar{m}(1 \leq \bar{m} \leq n)$ observations must be unaffected by shifts in the time origin of length $j\Delta$ (j an integer). It follows from this result that the mean and variance of a stationary stochastic process are constant.

Let $p(x)$ denote the *probability density function* of the random variable x, and $P(x)$ the corresponding *probability distribution function*. Since $p(g_k)$ is the same for all k, its shape can be inferred from the histogram of the observations $g_0, g_1, g_2, \ldots, g_N$. This is the fundamental result on which the analysis of a stationary stochastic process hinges: *The analysis of the properties of a stationary stochastic process which generated the given time series can be derived from the statistical properties of this time series.* This has led to the development of a very powerful theory for the analysis of time series. For the breadth and depth of the subject, the reader is referred to Box and Jenkins [7] for an introduction to time series and the autoregressive approach, to Hamming [20, 21] for an introduction to the frequency approach in numerical analysis, to Doob [15] for the basic analysis of stochastic processes, to Hannan [22, 23], Jenkins and Watts [28] and Grenander and Rosenblatt [19] for the spectral approach to time series, and to Tukey [45] for the use of the fast Fourier transform to compute numerical spectrum.

In this section, we show how results from this general theory can be used to define a numerical differentiation procedure which takes the structure within the non-exact data into account. The central assumption is the stationarity of the time series. Though this will not be true for most given time series, it is often easy to introduce some transformation of the data {for example, the removal of a (linear) trend} which reduces it to near stationarity. Any transformation is valid as long as it does not block the possibility of obtaining the derivative of the data or enhances the sensitivity of the method to rounding error.

We assume that the transformed data which is more or less stationary is given by

$$\{v_k;\ k=0, 1, 2, \ldots, N\} \tag{4.2}$$

Suppose that $u(t)$ is the function which underlies the data v_k and the derivative of which we wish to obtain. We assume that $u(t)$ is a stationary stochastic process with the continuous parameter t, that is, a random function. Without loss of generality, we shall also assume that $u(t)$ is observed, or sampled, at the times $t_k = k\Delta$, $k=0, 1, \ldots, N=1/\Delta$, and that the measurement error in the k-th observation is x_k. We assume further that $\{x_k\}$ is a stationary stochastic process with the discrete parameter k, that is, a random sequence. Thus,

$$v_k = u(t_k) + x_k. \tag{4.3}$$

We may clearly assume without further loss of generality that both $u(t)$ and x_k have zero expectation. The Wiener-Khintchine theory of generalised harmonic analysis {see Wiener [47] and Khitchine [29]} now implies that $u(t)$ and x_k may

be represented as

$$u(t) = \int_{-\infty}^{\infty} \exp(i\omega t/\Delta)\, dZ_u(\omega),$$
$$x_k = \int_{-\pi}^{\pi} \exp(i\omega k)\, dZ_x(\omega). \tag{4.4}$$

The term $\exp(i\omega t/\Delta)$ is not standard, but is chosen here since we have a natural time unit, Δ. The relevant feature of the functions Z_u and Z_x is that the variance of an integral

$$\int_I \theta(\omega)\, dZ_u(\omega)$$

is given by

$$\int_I |\theta(\omega)|^2\, dG_u(\omega),$$

where G_u is a nondecreasing function, called the *spectral distribution function*, and similarly for Z_x there exists a spectral distribution function G_x. If neither process contains any purely oscillatory component, as is usually the case, there will exist density functions g_u and g_x, and then the integrals may be written as

$$\int_I |\theta(\omega)|^2 g(\omega)\, d\omega. \tag{4.5}$$

Our present problem is to obtain estimates of $\dot{u}(t_k)$. A more classical problem which is usually discussed within this framework is that of filtering out the signal $u(t_k)$ itself. As we shall see, the solutions to these problems are closely related.

Suppose that we use a filter

$$\sum_{r=-\infty}^{\infty} l_r v_{k-r} \tag{4.3}$$

to estimate $\dot{u}(t_k)$. Now

$$\dot{u}(t) = \int_{-\infty}^{\infty} \exp(i\omega t/\Delta)\, (i\omega/\Delta)\, dZ_u(\omega) \tag{4.6}$$

and thus the error of this estimate is

$$\dot{u}(t_k) - \sum_{r=-\infty}^{\infty} l_r v_{k-r} = \int_{-\infty}^{\infty} \exp(i\omega k)\{i\omega/\Delta - l(\omega)\}\, dZ_u(\omega) - \int_{-\pi}^{\pi} \exp(i\omega k)\, l(\omega)\, dZ_x(\omega), \tag{4.7}$$

where

$$l(\omega) = \sum_{r=-\infty}^{\infty} l_r \exp(-ir\omega).$$

If we assume that error and signal are independent, then the mean squared error, that is the variance of (4.7), is

$$\int_{-\infty}^{\infty} |i\omega/\Delta - l(\omega)|^2 g_u(\omega)\, d\omega + \int_{-\pi}^{\pi} |l(\omega)|^2 g_x(\omega)\, d\omega. \tag{4.8}$$

We now assume that Δ has been chosen to be sufficiently small so that little or no detail of $u(t)$ has been lost by the sampling. This implies that g_u is either small or zero outside the interval $(-\pi, \pi)$ {Hannan [23], Section III.6]}, so we approximate (4.8) by

$$\int_{-\pi}^{\pi} \{|i\omega/\Delta - l(\omega)|^2 g_u(\omega) + |l(\omega)|^2 g_x(\omega)\}\, d\omega. \tag{4.9}$$

From the point of view of the subsequent analysis, it is more appropriate to work with $l_*(\omega) = (\Delta/i\omega)\, l(\omega)$ than $l(\omega)$. Using this notation, (4.9) becomes

$$\int_{-\pi}^{\pi} \omega^2/\Delta^2 \{|1 - l_*(\omega)|^2 g_u(\omega) + |l_*(\omega)|^2 g_x(\omega)\}\, d\omega, \tag{4.10}$$

which we shall denote by $\dot{\mathsf{F}}(l_*)$.

In the more classical problem mentioned above, one estimates $u(t_k)$ by

$$\sum_{r=-\infty}^{\infty} m_r\, v_{k-r},$$

and the mean squared error is

$$\int_{-\pi}^{\pi} \{|1 - m(\omega)|^2 g_u(\omega) + |m(\omega)|^2 g_x(\omega)\}\, d\omega, \tag{4.11}$$

where $m(\omega) = \sum_{r=-\infty}^{\infty} m_r \exp(ir\omega)$. We shall denote (4.11) by $\mathsf{F}(m)$.

Let $\hat{l}(\omega) = g_u(\omega)/\{g_u(\omega) + g_x(\omega)\}$. The central result of this theory is as follows.

Theorem 4.1.

$$\mathsf{F}(m) = \mathsf{F}(\hat{l}) + \|\hat{l} - m\|_\alpha^2, \tag{4.12}$$

$$\dot{\mathsf{F}}(l_*) = \dot{\mathsf{F}}(\hat{l}) + \|\hat{l} - l_*\|_\beta^2, \tag{4.13}$$

where $\|\cdot\|_\alpha$ *and* $\|\cdot\|_\beta$ *denotes the* L_2 *norms with weighting functions* $\{g_u(\omega) + g_x(\omega)\}$ *and* $\frac{\omega^2}{\Delta^2}\{g_u(\omega) + g_x(\omega)\}$, *respectively.*

Proof. We prove only (4.12). Now

$$\begin{aligned}
\mathsf{F}(m) &= \mathsf{F}(\hat{l} + m - \hat{l}) \\
&= \int_{-\pi}^{\pi} \{|1 - \hat{l}(\omega) - m(\omega) + \hat{l}(\omega)|^2 g_u(\omega) + |\hat{l}(\omega) + m(\omega) - \hat{l}(\omega)|^2 g_x(\omega)\}\, d\omega \\
&= \mathsf{F}(\hat{l}) + \|\hat{l} - m\|_\alpha^2 - 2\,\mathrm{Re} \int_{-\pi}^{\pi} [\{1 - \hat{l}(\omega)\}\{m(\omega) - \hat{l}(\omega)\}\, g_u(\omega) \\
&\quad - \hat{l}(\omega)\{m(\omega) - \hat{l}(\omega)\}\, g_x(\omega)]\, d\omega,
\end{aligned}$$

and it is easily verified that this last integral vanishes. The second result is proved similarly.

The connection between our differentiation problem and the classical filtering problem is now clear.

Theorem 4.2. *The minimum mean squared error filters has* $m = \hat{l}$, *and the minimum mean squared error differentiator has* $l_* = \hat{l}$.

Proof. This is an immediate consequence of Theorem 4.1.

It should be noted that the procedure implied by Theorem 4.2 will often not be feasible. Usually, we will require that the differentiation should be carried out by a local formula of the form

$$\sum_{r=-R}^{R} l_r v_{k-r}.$$

Theorem 4.1 then assures us that the best procedure in the sense of mean squared error is obtained by finding that function of the form

$$\hat{l}^{(R)}(\omega) = \sum_{r=-R}^{R} l_r^{(R)} \exp(ir\omega)$$

for which

$$\hat{l}_*^{(R)}(\omega) = (\Delta/i\omega)\, \hat{l}^{(R)}(\omega)$$

best approximates $\hat{l}$ in the β-norm. Clearly, in this constrained problem, the solution will no longer be related to that of the filtering problem, since there the approximation would have to be with respect to the α-norm.

The use of this theory requires knowledge of the spectral density functions g_u and g_x, which will rarely be known *a priori*. In the next section, we discuss estimation of these quantities from the data, and also two ways of establishing differentiation procedures based on the theory contained in this section.

5. The Spectrum Analysis Procedure

In this section, we construct two algorithms based on the time series theory of the previous section where it was shown that the numerical differentiation problem reduces to the determination of

$$\hat{l}(\omega) = g_u(\omega)/\{g_u(\omega) + g_x(\omega)\} \tag{5.1}$$

which minimizes $\dot{\mathsf{F}}(l_*)$. One yields a local differentiation formula while the other yields a global procedure. In both cases, an estimate of $\hat{l}(\omega)$ is required.

A simple step function estimate for $\hat{l}(\omega)$ can be derived if the spectrum of the data consists of regions where either $g_u(\omega) \ll g_x(\omega)$ or $g_u(\omega) \gg g_x(\omega)$; viz.

$$\hat{\hat{l}}(\omega) = \begin{cases} 0 & (g_u(\omega) \ll g_x(\omega)), \\ 1 & (g_u(\omega) \gg g_x(\omega)). \end{cases} \tag{5.2}$$

This proposal is based on the assumption that the spectrum $g_v(\omega)$ of the data (4.2) shows a clear division between signal and noise. If this is not the case, it can be used as a basis for deciding whether the data can meaningfully support a numerical differentiation. This result alone justifies the use of the spectral approach, at least as an exploratory tool.

Though windows of the form (5.2) will yield reliable results, far better results will be obtained when, along with (5.3), the window is tapered (linearly) in regions where $g_u(\omega) \sim g_x(\omega)$. The difficulty is the choice of some "best" estimate for $\hat{l}$ from among the totality of possibilities. In the next section, we see how this can be resolved by showing that the regularization procedure of Cullum is identical to a certain spectrum analysis procedure.

For the estimated $\hat{\hat{l}}(\omega)$, we derive both an explicit and an implicit numerical differentiation algorithm. Both have certain advantages. For example, the implicit method allows the use of quite complex tapering in the shape of $\hat{\hat{l}}(\omega)$. If it is necessary to repeatedly differentiate the same type of data, then it would be advantageous to have a local (explicit) differentiation formula of the form

$$\hat{\hat{u}}(t_k) = \sum_{r=-R}^{R} l_r^{(R)} v_{k-r}. \tag{5.3}$$

In a once off situation, it is usually more appropriate to use an implicit procedure which involves the direct numerical approximation of

$$\begin{aligned} \hat{\hat{u}}(t_k) &= \sum_{r=-\infty}^{\infty} l_r v_{k-r} \\ &= \sum_{r=-\infty}^{\infty} l_r \int_{-\infty}^{\infty} \exp\left(i\omega(k-r)\right) d\left(Z_u(\omega)+Z_x(\omega)\right) \\ &= \int_{-\infty}^{\infty} \exp(i\omega k)\, l(\omega)\, dZ_v(\omega) \\ &= \int_{-\infty}^{\infty} \exp(i\omega k)\,(i\omega/\Delta)\, \hat{l}(\omega)\, dZ_v(\omega). \end{aligned} \tag{5.4}$$

The first algorithm we develope is an *explicit procedure*. If a local differentiator of the form (5.3) is required, then it follows from §4 that the best procedure, in the sense of mean squared error, is obtained by finding

$$\hat{l}^{(R)}(\omega) = \sum_{r=-R}^{R} l_r^{(R)} \exp(ir\omega) \tag{5.5}$$

for which $\hat{l}_*^{(R)}(\omega) = (\Delta/i\omega)\, \hat{l}^{(R)}(\omega)$ best approximates $\hat{l}(\omega)$ in the β-norm. From Theorems 4.1 and 4.2, this reduces to the minimization of

$$\begin{aligned} \|\hat{l}(\omega) - l_*^{(R)}(\omega)\|_\beta &= \int_{-\pi}^{\pi} |\hat{l}(\omega) - l_*^{(R)}(\omega)|^2\, (\omega)^2/\Delta^2)\, g_v(\omega)\, d\omega, \\ &= \int_{-\pi}^{\pi} |(i\omega/\Delta)\, \hat{l}(\omega) - l^{(R)}(\omega)|^2\, g_v(\omega)\, d\omega \end{aligned} \tag{5.6}$$

with respect to the $l_r^{(R)}$ of (5.5). Taking the form of $l_*^{(R)}(\omega)$ into account and using the basic techniques of the calculus of variations, this leads to the following conditions for the evaluation of the $l_k^{(R)}$:

$$\int_{-\pi}^{\pi} \{(i\omega/\Delta)\, \hat{l}(\omega) - l^{(R)}(\omega)\, g_v(\omega) \exp(-i\omega k)\}\, d\omega = 0. \tag{5.7}$$

Using (5.5) and the spectral estimate $\hat{\hat{l}}(\omega)$ for $\hat{l}(\omega)$, (5.7) can be reorganized to yield

$$\sum_{r=-R}^{R} l_r^{(R)} a_{rk} = \gamma_k \qquad (k=-R, \ldots, R), \tag{5.8}$$

with

$$a_{rk} = \int_{-\pi}^{\pi} g_v(\omega) \exp\left(i\omega(r-k)\right) d\omega, \tag{5.9}$$

and

$$\gamma_k = \int_{-\pi}^{\pi} (i\omega/\Delta)\, \hat{\hat{l}}(\omega)\, g_v(\omega) \exp(-i\omega k)\, d\omega. \tag{5.10}$$

By replacing $g_v(\omega)$ in (5.9) and $\hat{\hat{l}}(\omega)\, g_v(\omega)$ in (5.10) by piecewise polynomials on the grid $\{\omega_j = 2\pi j/N;\ j=0, 1, \ldots, N\}$ product integration techniques can be used to evaluate the a_{rk} and γ_k. The fast Fourier transform could also be used to determine the a_{rk} and the γ_k. The $l_r^{(R)}$ are then obtained by inverting the linear system (5.8). The sequential application of (5.3) then yields

$$\hat{\hat{u}}(t_k) \qquad \text{for } k=R, \ldots, N-R.$$

If $\hat{\hat{u}}(t_k)$ is also required at the points

$$k=0, 1, 2, \ldots, R-1, \quad \text{and} \quad k=N-R+1, \ldots, N,$$

then the odd periodic extension of the data (4.2) should be used. We shall show in §6 that the odd periodic extension represents the natural way in which to extend numerical differentiation data.

The second algorithm we examine is *an implicit procedure*. The basis of this procedure is the direct approximation of (5.4) using numerical quadrature. The construction of this algorithm is based on the observation that, in order to calculate $g_v(\omega)$, it is first necessary to evaluate the periodogram of the data (4.2). Hence, it is necessary to use the fast Fourier transform to determine the finite Fourier transform ($\omega_j = 2\pi j/N$)

$$\tilde{v}(\omega_j) = (1/N) \sum_{r=0}^{N-1} v(t_r) \exp(-ir\omega_j) \qquad (j=0, \ldots, N-1) \tag{5.11}$$

with the inverse Fourier transform defined by

$$v(t_r) = \sum_{j=0}^{N-1} \tilde{v}(\omega_j) \exp(i\omega_j r) \qquad (r=0, 1, \ldots, N-1). \tag{5.12}$$

If we write

$$Z_v^{(n)}(\omega) = \sum_{-\pi < \omega_j \leqq \omega} \tilde{v}(\omega_j) \qquad (-\pi < \omega < \pi),$$

where the $\tilde{v}$'s have been extended periodically, it follows that

$$\begin{aligned} v(t_r) &= \sum_{j=0}^{N-1} \exp(i\omega_j r)\, \tilde{v}(\omega_j) \\ &= \int_{-\pi}^{\pi} \exp(i\omega r)\, dZ_v^{(n)}(\omega) \qquad (r=0, \ldots, N-1). \end{aligned}$$

Comparing this with

$$v(t)=\int_{-\pi}^{\pi}\exp(i\omega t/\Delta)\,dZ_v(\omega)\qquad(-\infty<t<\infty),$$

we see that $Z_v^{(n)}$ serves as an approximation to Z_v. Hence it may be used to construct an approximation to (5.4), that is

$$\begin{aligned}\hat{\dot{u}}(t_k)&\cong\int_{-\pi}^{\pi}\exp(i\omega k)\,(i\omega/\Delta)\,\hat{l}(\omega)\,dZ_v^{(n)}(\omega)\\&=\sum_{-\pi<\omega_j\leq\pi}\exp(i\omega_j k)\,(i\omega_j/\Delta)\,\hat{l}(\omega_j)\,\tilde{v}(\omega_j),\end{aligned}\tag{5.13}$$

where the final expression can be regarded as a quadrature formula for $\hat{\dot{u}}(t_k)$. Consequently we estimate the derivative by evaluating (5.13) with $\hat{l}$ replaced by its estimate $\hat{\hat{l}}$.

The results obtained with both these procedures are discussed in §8. We close this section with two comments on the implicit, or global, procedure just derived. The first is that fast Fourier transform algorithms are most efficient when the number of data points being transformed is highly composite, and that an algorithm which works only when this number is a power of two is relatively simple. For this reason, data are often extended by zeroes to make the series length up to a power of two. However, in the next section it emerges that the natural way to extend data for the purposes of numerical differentiation is by the odd periodic extension. This should insure that the biases introduced at the ends of the data by the frequency approach are minimized.

There is one other advantage of the spectral analysis procedure. It also allows the signal $u(t)$ for which the derivative $\dot{u}(t)$ has been derived to be constructed. Thus, the experimentalist can use this as a cross-check on the reliability of the resulting numerical derivative by comparing the reconstructed signal with the data (4.2).

6. A Spectrum Analysis Interpretation for the Regularization Parameter of Cullum's Method

As indicated in §2, of the methods based on an abstract formalism, Cullum's method is the only one which can be applied to the differentiation of non-exact data with reliability, and its successful application depends heavily on the use of the optimum value of the regularization parameter α. Consequently, as is clear from the conclusions in Cullum [13; p. 264], the general applicability of this method depends heavily on finding an interpretation for the optimum α which allows it to be computed accurately.

Here, we show that an interpretation in terms of the implicit spectral analysis method exists. Initially, we observe that, for given data $g(t)$, $0\leq t\leq 1$, with $g(0)=g(1)=0$ and a non-zero and positive α, it follows from Cullum [13; eqn. (4.1)], that the exact regularized solution is defined by

$$\int_t^1\int_0^y f(z)\,dz\,dy+\int_0^1 f(y)\,dy+\alpha f(t)-\alpha f^{(2)}(t)=\int_t^1 g(y)\,dy\tag{6.1}$$

along with the boundary conditions

$$f^{(1)}(0)=f^{(1)}(1)=0. \tag{6.2}$$

Theorem 6.1. *The regularized solution* $f(\alpha;t)$ *of Cullum's method is defined by the differential equation*

$$f(\alpha;t)-\alpha f^{(2)}(\alpha;t)+\alpha f^{(4)}(\alpha;t)=g^{(1)}(t) \tag{6.3}$$

and the boundary conditions (6.2) *and*

$$f^{(3)}(\alpha;0)=f^{(3)}(\alpha;1)=0. \tag{6.4}$$

Proof. We obtain (6.3) by differentiating (6.1) twice with respect to t. The boundary conditions (6.4) are obtained by substituting $t=0$ and $t=1$, respectively, in the first derivative of (6.1)

Corollary 6.1. *If* $g(x)=(k\pi)^{-1}\sin k\pi x$, *then*

$$f(\alpha;x)=\cos k\pi x/\left(1+\alpha(k\pi)^2+\alpha(k\pi)^4\right). \tag{6.5}$$

Proof. Substitution in (6.2), (6.3) and (6.4) confirm that (6.5) is a particular integral of (6.3), while the boundary conditions (6.2) and (6.4) force the complementary function to be zero.

Since we have $g(0)=g(1)=0$, it follows that $g(t)$ may be expanded in a Fourier sine series

$$g(t)=\sum_{r=1}^{\infty}\gamma_r\sin(\pi rt), \tag{6.6}$$

which is continuous at the end points $t=0$ and $t=1$. Hence, from Corollary 6.1, the regularized solution is given by

$$f(\alpha;t)=\sum_{r=1}^{\infty}\gamma_r\pi r\cos(\pi rx)/\{1+\alpha(\pi r)^2+\alpha(\pi r)^4\}. \tag{6.7}$$

The validity of (6.7) hinges on the assumption that the regularized derivative of the Fourier series of g, $g\in L_2$, coincides with the Fourier series of the regularized derivative of g. This is a direct consequence of the well-posedness of the regularization formulation $P(\alpha)$ {see (2.3)} which was established by Cullum [13; Theorem 5.1]. Consequently, in the case when $g(t)$ is given only as data

$$\{g_k=g(t_k);\ t_k=k\Delta,\ \Delta=1/N,\ k=0\ 1,\ \ldots,\ N\} \tag{6.8}$$

we can construct the finite Fourier sine series corresponding to (6.6) by using an odd extension of (6.8). In this way, we obtain (since $g(0)=g(1)=0$)

$$\gamma_r=\frac{2}{N}\sum_{k=1}^{N-1}g_k\sin(\pi krA) \qquad (r=1,2,\ldots,N-1), \tag{6.9}$$

as an approximation to γ_r $(r=1,2,\ldots,N-1)$. The substitution of (6.9) along with $\tilde{\gamma}_r=0$ $(r\geq N)$ in (6.7) then yields

$$f(\alpha;t_k)=\sum_{r=1}^{N-1}\tilde{\gamma}_r\pi r/\{1+\alpha(\pi r)^2+\alpha(\pi r)^4\}\cos(\pi rk\Delta). \tag{6.10}$$

This evidently has a similar form to the implicit procedure of §5, viz.

$$\hat{u}(t_k)=\sum_{j=0}^{N-1}(i\omega_j/\Delta)\,\hat{l}(\omega_j)\,\tilde{g}(\omega_j)\exp(i\omega_j k) \tag{6.11}$$

with

$$\tilde{g}(\omega_j)=(1/N)\sum_{r=0}^{N-1}g_r\exp(-ir\omega_j), \tag{6.12}$$

since, with $\omega_j=2\pi j/N=2\pi j\Delta$, (6.11) and (6.12) can be rewritten to yield

$$\hat{u}(t_k)=\sum_{j=0}^{N-1}(i2\pi j)\,\hat{l}_j\,\tilde{g}_j\exp(i2\pi j\Delta k) \tag{6.13}$$

and

$$\tilde{g}_j=(1/N)\sum_{r=0}^{N-1}g_r\exp(-i2\pi rj\Delta) \tag{6.14}$$

with $\hat{l}_j=\hat{l}(\omega_j)$ and $\tilde{g}_j=\tilde{g}(\omega_j)$.

The difference between the two representations (6.10) and (6.13) relates to the fact that (6.10) has been derived using a Fourier sine series (in order to maintain consistency within the regularization formulation) while (6.13) has been derived using the full Fourier series representation, and therefore, involves fewer frequencies. Consequently, if we apply the spectral procedure to an odd periodic extension of the data (6.8) and use the relevant argument of §5, then we obtain the following formula corresponding to (6.13)

$$\hat{u}(t_k)=\sum_{j=1}^{N-1}\pi j\,\hat{l}_j\tilde{\gamma}_j\cos(\pi jk\Delta). \tag{6.15}$$

Thus, a comparison of (6.15) with (6.10) indicates that the regularization approach can be regarded as a variant of the time series approach in which $\hat{l}(\omega)$ is approximated by a function of the form

$$\lambda_\alpha(\omega)=1/\{1+\alpha(\omega/\Delta)^2+\alpha(\omega/\Delta)^4\}. \tag{6.16}$$

We observe that for a given value of α, an alternative implementation of the regularization procedure of Cullum is obtained by carrying out the implicit spectrum analysis procedure of the previous section with $\hat{l}(\omega)$ replaced by $\lambda_\alpha(\omega)$. This will yield a computationally superior method since this implementation will only involve $O(N\log N)$ operations (when using the Fast Fourier Transform {see Cooley and Tukey [11]}), compared with $O(N^3)$ operations required normally, since the numerical solution of (2.4) reduces to the inversion of a non-sparse matrix of order N.

This equivalence also gives rise to a criterion for comparing values of the regularization parameter and a procedure for estimating the optimal value under this criterion. For suppose the spectrum of the errors, $g_z(\omega)$, were known up to a constant, say

$$g_z(\omega)=b\,h(\omega),$$

where $h(\omega)$ is a known function. This is certainly true if the errors are known to be uncorrelated, when $h(\omega) \equiv 1$. Suppose further that

$$g_u(\omega) = b\,h(\omega)/\{\alpha(\omega/\Delta)^2 + \alpha(\omega/\Delta)^4\}.$$

Then it is easily seen that $\hat{l}(\omega)$ is now precisely $\lambda_\alpha(\omega)$, and that the spectrum of the observed data is given by the two parameter model

$$g_v(\omega; \alpha, b) = b\,h(\omega)\,[1 + 1/\{\alpha(\omega/\Delta)^2 + \alpha(\omega/\Delta)^4\}].$$

Since the probability distribution of the observed data is determined by its spectrum, and hence, by the values of α and b, the statistical likelihood of any suggested values of α and b may be calculated from the observed data. This is discussed more fully in Anderssen and Bloomfield [1], who show how to eliminate b from the problem, and that the optimal vlaue of α under this criterion, that is the maximum likelihood estimate of α, is found by minimising

$$\begin{aligned} N/2 \log\left(\sum_{j=1}^{N/2} I_v(\omega_j)/[h(\omega_j) + h(\omega_j)/\{\alpha(\omega_j/\Delta)^2 + \alpha(\omega_j/\Delta)^4\}]\right) \\ + \sum_{j=1}^{N/2} \log[1 + 1/\{\alpha(\omega_j/\Delta)^2 + \alpha(\omega_j/\Delta)^4\}], \end{aligned} \tag{6.17}$$

where $\omega_j = 2\pi j/N$ and

$$I_v(\omega_j) = \frac{1}{2\pi N}\left|\sum_{t=0}^{N} v_t\, e^{it\omega_j}\right|$$

is the periodogram of the data.

Other consequences of the fact that the regularization method has a well defined spectral interpretation are:

1. The stability and well-posedness of the regularization procedure represents a heuristic proof of the stability and well-posedness of the spectral procedure for numerical differentiation, at least when $\hat{l}(\omega) = \lambda_\alpha(\omega)$ and an odd periodic extension of the data is used.

2. In order to minimize bias due to the ends of the data, it is necessary to

(a) detrend the data so that $g(0) = g(1) = 0$ and

(b) work with an odd periodic extension of this detrended data.

3. Since the time series procedure is designed to cope with errors in the data, it represents justification for Cullum's conclusion (based on numerical experimentation) that the regularization method is stable for noisy data.

7. The Order of Convergence of Spectral Differentiation

Initially, we derive in Theorem 7.1, for smooth exact data $g(t)$, the order of convergence of the corresponding spectral derivative, and then interpret it from the point of view of the differentiation of non-exact data. Assuming that on the interval $[0, 1]$, the data v_k has the structure (4.3) with $x_k = 0$, we prove

Theorem 7.1. *If $u(t)$ satisfies the conditions*

(i) $u \in C^p\,[0, 1]$, $p \geq 3$;

(ii) $u^{(p)}$ *satisfies a Lipschitz condition of order* β, $0 \leq \beta \leq 1$;

(iii) $u(-x) = -u(x)$, $x \in [-1, 1]$;

(iv) $u(0) = u(1) = 0$, $u^{(q)}(0) = u^{(q)}(1)\,(q = 1, 2, \ldots, p)$;

then

$$|\dot{u}(t_k) - \hat{\dot{u}}(t_k)| = O(\alpha) + O(\Delta^{p+\beta-2}) \tag{7.1}$$

if the number of data points $(N+1)$ *is sufficiently large.*

Proof. Let $\dot{u}(\alpha; t)$ denote the regularized derivative of $u(t)$ under the constraint that $\int_0^1 \dot{u}(\alpha; t)\,dt = 0$, then

$$|\dot{u}(t_k) - \hat{\dot{u}}(t_k)| \leq |\dot{u}(t_k) - \dot{u}(\alpha; t_k)| + |\dot{u}(\alpha; t_k) - \hat{\dot{u}}(t_k)|. \tag{7.2}$$

It follows from Theorem 3.1 in Cullum [13] that

$$|\dot{u}(t_k) - \dot{u}(\alpha; t_k)| = O(\alpha^{\frac{1}{2}}).$$

On the other hand, setting $g(t) = u(t)$ in (6.6) and $\dot{u}(\alpha; t) = f(\alpha; t)$ in (6.7), we obtain

$$|\dot{u}(t_k) - \dot{u}(\alpha; t_k)| = O(\alpha).$$

However, this latter estimate does not replace Cullum's since it also applies to the case when $C(\alpha; f)$ of (2.3) is not minimized exactly.

Let the Fourier sine series expansion for $u(t)$ on the interval $[-1, 1]$ be

$$u(t) = \sum_{r=1}^{\infty} \gamma_r \sin(\pi r t), \qquad \gamma_r = 2\int_0^1 u(t) \sin(\pi r t)\,dt,$$

and denote the coefficients of the finite Fourier sine series corresponding to an odd extension of the data $v_k = u(t_k)$ by

$$\tilde{\gamma}_r = \frac{2}{N} \sum_{k=1}^{N} v_k \sin(\pi k r \Delta) \qquad (r = 1, 2, \ldots, N-1).$$

Then, using the equivalence between regularization and spectrum analysis derived in § 6, it follows that

$$\begin{aligned} |\dot{u}(\alpha; t_k) - \hat{\dot{u}}(t_k)| = &\left| \sum_{r=1}^{N-1} (\gamma_r - \tilde{\gamma}_r)\, \pi r\, \lambda_\alpha(\pi r) \cos(\pi r k \Delta) \right| \\ &+ \sum_{j=N}^{\infty} \gamma_r\, \pi r\, \lambda_\alpha(\pi r) \cos(\pi r k \Delta). \end{aligned} \tag{7.4}$$

Now, it is not difficult to establish that

$$|\gamma_r| = O(r^{-p-\beta}), \tag{7.5}$$

and it follows from Hamming [21; Part III] that

$$\tilde{\gamma}_r = \gamma_r + \sum_{q=1}^{\infty} (-\gamma_{2Nq-r} + \gamma_{2Nq+r}). \tag{7.6}$$

Using (7.5) in conjunction with (7.6), it can be shown that

$$|\gamma_r - \tilde{\gamma}_r| = O(N^{-p-\beta}) \quad (r = 1, 2, \ldots, N-1). \tag{7.7}$$

Since $|\cos(\pi r k \Delta)| \leq 1$ and $\lambda_\alpha(\pi r) < 1$, (7.4) yields on applying (7.5) and (7.7) that

$$|\dot{u}(\alpha; t_k) - \hat{\dot{u}}(t_k)| = O(\Delta^{+p+\beta-2}). \tag{7.8}$$

Combining (7.2), (7.3) and (7.8), we obtain the required result (7.1).

Interpretation for Non-Exact Data: In the actual application of spectral differentiation to non-exact data v_k with the structure (4.3), we use the window $\hat{l}(\omega) = \lambda_\alpha(\omega)$ to filter off the noise from the signal. In this way, the Fourier sine coefficients for v_k, $\bar{\gamma}_r$ say, are replaced by $\bar{\gamma}_r \lambda_\alpha(\pi r)$. Let the function with these smoothed coefficients be $g_\alpha(t)$. Then Theorem 7.1 remains valid for this non-exact data, if $g_\alpha(t)$ {instead of $u(t)$}, now satisfies the conditions of Theorem 7.1, since it is the derivative of this function which sprectral differentiation determines.

8. Numerical Results

Numerical experimentation with noisy artificial data as well as real experimental data indicated that:

(i) Though the regression procedure, using fixed knot splines, gave better results than methods based on an abstract formalism (excluding Cullum), the fast Fourier transform implementation of the regularization procedure of Cullum (FFT implementation) invariably yielded better results. The possibility of using a variable knot procedure (see Burchard [9], de Boor and Rice [6]) was not given serious consideration since its computational efficiency would not compare favourably with that of the FFT implementation.

(ii) The estimation procedure for α (viz., the minimization of (6.17)) always yielded values which gave smooth derivatives using the FFT implementation. In addition, a comparison between the given and reconstructed data was favourable as long as the data was not too noisy. The actual minimization was carried out using a quadratic procedure.

(iii) The explicit (local) spectrum procedure gave nearly identical results to the implicit (global) spectrum procedure when the number of points in the explicit formula exceeded 9 (viz., when $R > 4$).

We do not present the results for experimental data in this paper. The interested reader is referred to Anderssen and Bloomfield [1] where the results of applying the FFT implementation to some photoelectric response data of Lüttge (see Pallaghy and Lüttge [36]) is examined at length.

Instead, we apply the FFT implementation to the artifical data

$$\begin{aligned} g(x) &= B x^2 (x-0.1)(x-0.2)(x-0.5)(x-0.75)(x-0.95)(x-1)^2 \\ &\quad + A x^2 + \varepsilon = g_0(x) + \varepsilon, \end{aligned} \tag{8.1}$$

Table 1. Comparison of exact and calculated signal and derivative for data (8.1)
($\Delta = 1/128$, $x_i = i\Delta$, $A = 1.0$, $B = 1.0\,\mathrm{E}+3$, $s = 1.0\,\mathrm{E}-6$)
$\alpha = 3.213\,\mathrm{E}-21$

i	$g(x_i)$	$\hat{g}(x_i)$	$\hat{\dot{g}}(x_i)$	$\hat{g}(x_i)-g_0(x_i)$	$\hat{\dot{g}}(x_i)-\dot{g}_0(x_i)$
1	−2.744 E−6	−4.137 E−6	−1.081 E−2	−4.137 E−6	−1.081 E−2
17	1.996 E−2	1.996 E−2	4.016 E−1	−5.320 E−7	2.244 E−4
33	3.943 E−2	3.943 E−2	−6.682 E−2	−3.348 E−7	7.024 E−5
49	6.937 E−2	6.937 E−2	8.156 E−1	−3.817 E−7	−1.313 E−4
65	2.500 E−1	2.500 E−1	1.844	1.038 E−6	2.083 E−5
81	4.529 E−1	4.529 E−1	1.191	1.270 E−7	−9.619 E−5
97	5.625 E−1	5.625 E−1	8.715 E−1	3.812 E−7	−1.240 E−4
113	7.436 E−1	7.436 E−1	2.049	−2.515 E−7	1.780 E−5
129	1.000	1.000	1.994	7.192 E−7	−5.847 E−3
σ^2				9.336 E−13	1.286 E−6

($\Delta = 1/128$, $x_i = i\Delta$, $A = 1.0$, $B = 1.0\,\mathrm{E}+3$, $s = 1.0\,\mathrm{E}-2$)
$\alpha = 2.087\,\mathrm{E}-12$

i	$g(x_i)$	$\hat{g}(x_i)$	$\hat{\dot{g}}(x_i)$	$\hat{g}(x_i)-g_0(x_i)$	$\hat{\dot{g}}(x_i)-\dot{g}_0(x_i)$
1	−2.744 E−2	−1.289 E−2	2.789 E−1	−1.289 E−2	2.789 E−1
17	1.508 E−2	2.236 E−2	3.011 E−1	2.399 E−3	−1.003 E−1
33	3.699 E−2	3.962 E−2	−6.405 E−2	1.874 E−4	2.848 E−3
49	6.613 E−2	7.186 E−2	8.522 E−1	2.485 E−3	3.652 E−2
65	2.596 E−1	2.519 E−1	1.808	1.874 E−3	−3.567 E−2
81	4.558 E−1	4.522 E−1	1.186	−6.600 E−4	−4.292 E−2
97	5.661 E−1	5.606 E−1	8.915 E−1	−1.924 E−3	1.993 E−2
113	7.422 E−1	7.434 E−1	1.965	−1.975 E−4	−8.482 E−2
129	9.946 E−1	9.970 E−1	2.023	−2.981 E−3	2.286 E−2
σ^2				9.137 E−6	7.881 E−3

defined on the interval [0.1], where ε is a normally distributed random variable with mean zero and standard deviation s. In this way, we know the exact structure of the signal $g_0(x)$, and therefore can compare explicitly the exact and calculated signal and derivative. The results are listed in Tables 1 and 2 where $\hat{g}(x)$ denotes the reconstructured data. In the FFT implementation used, a linear or quadratic trend is removed from the data to force the condition $g(0) = g(1) = 0$. It is clear from Tables 1 and 2 that even for noisy data with the derivative changing rapidly, the FFT implementation yields reliable results.

We illustrate the practical consequences of the complexity results of § 6 by considering the computation time required on the IBM 360/91 computer at Princeton University for 101 data points. The direct implementation of the regularization procedure of Cullum, for a single predetermined α, required 15.75 seconds, while the FFT implementation took only 0.22 seconds, which included the computation of the optimum value of α.

Practical verification of the validity of the equivalence between spectrum analysis and reularization is given in Table 3. The value of the regularized derivative corresponding to the optimal choice of α in § 6 {see column 3 of Table 3}

Table 2. Comparison of exact and calculated signal and derivative for data (8.1)
($\Delta = 1/128$, $x_i = i\Delta$, $A = 1.0$, $B = 1.0\ \mathrm{E}+4$, $s = 0.01$)
$\alpha = 1.101\ \mathrm{E}-14$

i	$g(x_i)$	$\hat{g}(x_i)$	$\hat{\dot{g}}(x_i)$	$\hat{g}(x_i) - g_0(x_i)$	$\hat{\dot{g}}(x_i) - \dot{g}_0(x_i)$
1	$-2.744\ \mathrm{E}-2$	$-1.766\ \mathrm{E}-2$	$2.804\ \mathrm{E}-1$	$1.766\ \mathrm{E}-2$	$2.804\ \mathrm{E}-1$
17	$5.412\ \mathrm{E}-2$	$5.325\ \mathrm{E}-2$	1.941	$-5.743\ \mathrm{E}-3$	$1.771\ \mathrm{E}-1$
33	$-1.707\ \mathrm{E}-1$	$-1.685\ \mathrm{E}-1$	-5.259	$-2.579\ \mathrm{E}-4$	$-9.050\ \mathrm{E}-2$
49	$-5.752\ \mathrm{E}-1$	$-5.676\ \mathrm{E}-1$	1.532	$4.348\ \mathrm{E}-3$	$1.250\ \mathrm{E}-1$
65	$2.596\ \mathrm{E}-1$	$2.527\ \mathrm{E}-1$	9.444	$2.686\ \mathrm{E}-3$	$6.138\ \mathrm{E}-3$
81	1.016	1.011	$6.508\ \mathrm{E}-1$	$-1.753\ \mathrm{E}-3$	$-6.359\ \mathrm{E}-3$
97	$5.661\ \mathrm{E}-1$	$5.565\ \mathrm{E}-1$	-4.897	$-6.019\ \mathrm{E}-3$	$-1.129\ \mathrm{E}-1$
113	$5.441\ \mathrm{E}-1$	$5.438\ \mathrm{E}-1$	4.612	$-1.856\ \mathrm{E}-3$	$-1.320\ \mathrm{E}-1$
129	$9.946\ \mathrm{E}-1$	$1.000\ \mathrm{E}-1$	1.755	$6.227\ \mathrm{E}-5$	$-2.455\ \mathrm{E}-1$
σ^2				$2.065\ \mathrm{E}-5$	$4.793\ \mathrm{E}-2$

($\Delta = 1/128$, $x_i = i\Delta$, $A = 1.0$, $B = 1.0\ \mathrm{E}+4$, $s = 5.0\ \mathrm{E}-2$)
$\alpha = 3.594\ \mathrm{E}-13$

i	$g(x_i)$	$\hat{g}(x_i)$	$\hat{\dot{g}}(x_i)$	$\hat{g}(x_i) - g_0(x_i)$	$\hat{\dot{g}}(x_i) - \dot{g}_0(x_i)$
1	$-1.372\ \mathrm{E}-1$	$-6.644\ \mathrm{E}-2$	1.690	$-6.644\ \mathrm{E}-2$	1.690
17	$3.461\ \mathrm{E}-2$	$5.379\ \mathrm{E}-2$	1.301	$-5.206\ \mathrm{E}-3$	$-4.629\ \mathrm{E}-1$
33	$-1.804\ \mathrm{E}-1$	$-1.620\ \mathrm{E}-1$	-5.498	$6.200\ \mathrm{E}-3$	$-3.292\ \mathrm{E}-1$
49	$-5.881\ \mathrm{E}-1$	$-5.576\ \mathrm{E}-1$	1.731	$1.432\ \mathrm{E}-2$	$3.244\ \mathrm{E}-1$
65	$2.979\ \mathrm{E}-1$	$2.590\ \mathrm{E}-1$	9.273	$9.000\ \mathrm{E}-3$	$-1.641\ \mathrm{E}-1$
81	$1.028\ \mathrm{E}-1$	1.010	$7.588\ \mathrm{E}-1$	$-3.062\ \mathrm{E}-3$	$1.017\ \mathrm{E}-1$
97	$5.804\ \mathrm{E}-1$	$5.473\ \mathrm{E}-1$	-4.846	$-1.522\ \mathrm{E}-2$	$-6.217\ \mathrm{E}-2$
113	$5.383\ \mathrm{E}-1$	$5.480\ \mathrm{E}-1$	4.284	$2.355\ \mathrm{E}-3$	$-4.605\ \mathrm{E}-1$
129	$9.730\ \mathrm{E}-1$	$9.907\ \mathrm{E}-1$	1.979	$-9.267\ \mathrm{E}-3$	$-2.143\ \mathrm{E}-2$
σ^2				$2.578\ \mathrm{E}-4$	$3.371\ \mathrm{E}-1$

is almost identical with the corresponding spectral derivative of Table 1. On the other hand, the regularized derivatives corresponding to non-optimal choice of α differed (often greatly) from the optimal—see remaining columns in Table 3.

9. Generalizations and Conclusions

The development of the fast Fourier transform implementation for the regularization procedure of Cullum described above results from the combination of ideas from time series analysis and numerical analysis. It represents an efficient implementation of a sophisticated numerical analytic procedure, in which use of the fast Fourier transform replaces the numerical inversion of a large nonsparse matrix. Through the equivalence established in § 6, an otherwise uninterpreted parameter in the numerical analytic procedure is interpreted as a parameter in the spectrum of the observed data, thus allowing its estimation from the statistical properties of the data.

Table 3. Comparison of regularized derivatives of the data (8.1)
($\Delta = 1/128$, $x_i = i\Delta$, $A = 1.0$, $B = 1.0\,\mathrm{E}+3$, $s = 0.01$)

	$\alpha = 2.\,\mathrm{E}-10$		$\alpha = 2.\,\mathrm{E}-11$	
i	$\hat{\dot{g}}(x_i)$	$\hat{\dot{g}}(x_i) - \dot{g}_0(x_i)$	$\hat{\dot{g}}(x_i)$	$\hat{\dot{g}}(x_i) - \dot{g}_0(x_i)$
1	2.548 E−1	2.548 E−1	2.549 E−1	2.549 E−1
17	1.714 E−1	−2.300 E−1	2.696 E−1	−1.318 E−1
33	8.108 E−2	1.480 E−1	−1.465 E−2	5.225 E−2
49	8.860 E−1	7.033 E−2	8.348 E−1	1.914 E−2
65	1.642	−2.018 E−1	1.803	−4.102 E−2
81	1.236	4.560 E−2	1.164	−2.712 E−2
97	1.061	1.891 E−1	9.561 E−1	8.456 E−2
113	1.722	3.279 E−1	1.884	−1.656 E−1
129	2.178	1.775 E−1	2.099	9.906 E−2
σ^2		3.251 E−2		1.179 E−2

	$\alpha = 2.087\,\mathrm{E}-12$		$\alpha = 2.\,\mathrm{E}-13$	
i	$\hat{\dot{g}}(x_i)$	$\hat{\dot{g}}(x_i) - \dot{g}_0(x_i)$	$\hat{\dot{g}}(x_i)$	$\hat{\dot{g}}(x_i) - \dot{g}_0(x_i)$
1	2.789 E−1	2.789 E−1	4.088 E−1	4.088 E−1
17	3.011 E−1	−1.003 E−1	3.474 E−1	−5.394 E−2
33	−6.405 E−2	2.848 E−3	−1.538 E−1	−8.691 E−2
49	8.522 E−1	3.652 E−2	8.826 E−1	6.694 E−2
65	1.808	−3.567 E−2	1.821	−2.283 E−2
81	1.186	−4.292 E−2	1.217	2.587 E−2
97	8.915 E−1	1.993 E−2	8.526 E−1	−1.897 E−2
113	1.965	−8.482 E−2	1.962	−8.712 E−2
129	2.023	2.286 E−2	1.967	−3.254 E−2
σ^2		7.881 E−3		1.153 E−2

	$\alpha = 2.\,\mathrm{E}-14$	
i	$\hat{\dot{g}}(x_i)$	$\hat{\dot{g}}(x_i) - \dot{g}_0(x_i)$
1	6.069 E−1	6.069 E−1
17	4.823 E−1	8.097 E−2
33	−1.676 E−1	−1.007 E−1
49	9.163 E−1	1.006 E−1
65	1.850	5.879 E−3
81	1.204	1.327 E−2
97	7.838 E−1	−8.780 E−2
113	1.887	−1.623 E−1
129	1.896	−1.037 E−1
σ^2		2.613 E−2

The nature of the computational advantages is clear from the complexity result discussed in § 6, and the indirect check on the reliability of the computed derivative via the reconstructed signal.

Consequently, it is natural to enquire whether this simple equivalence has generalizations. The extension of the spectral procedure to the situation where the $n-th$ derivative of $g(t)$ is required is straight forward. The analysis and results of §§ 4, 5 carry over with

$$\dot{\mathsf{F}}(l) = \int_{-\pi}^{\pi} \{|i\omega/\Delta - l(\omega)|^2 g_u(\omega) + |l(\omega)|^2 g_x(\omega)\} d\omega$$

replaced by

$$\mathsf{F}^{(n)}(l) = \int_{-\pi}^{\pi} \{|(i\omega/\Delta)^n - l(\omega)|^2 g_u(\omega) + |l(\omega)|^2 g_x(\omega)\} d\omega, \tag{9.1}$$

and with $(i\omega/\Delta)$ and (ω^2/Δ^2) replaced by $(i\omega/\Delta)^n$ and $(\omega^2/\Delta^2)^n$ elsewhere.

However, this leaves the question of an optimal estimate for $\hat{l}(\omega)$ unanswered. At least for the higher derivative case, we could argue on heuristic grounds that the estimate $\lambda_\alpha(\omega)$ of (6.16) was appropriate. The exact form of the equivalence will depend on the way in which the regularization of the integral equation formulation for the n–th derivative, viz.

$$\boldsymbol{A}^n f = \int_0^1 H(t-s)(t-s)^{n-1} f(s)/(n-1)!\, ds = g(t), \qquad 0 \leq t \leq 1, \tag{9.2}$$

where $H(t-s)$ is the Heaviside unit step function, is introduced. For example, when n is odd replace (9.2) by the following family of optimization problems $(0 < \alpha \leq 1)$: for each α minimize

$$C^{(n)}(\alpha, f) = \|\boldsymbol{A}^n f - g\|_2^2 + \left(\int_0^1 \left\{ g_{(x)}^{(1)} + \sum_{i=1}^{n} f^{(2i)}(x) \right\} dx \right)^2 + \alpha \|f\|_4^2 \tag{9.3}$$

for $f \in W_1^2$, under the assumption that

$$g(0) = g^{(1)} = (0), \qquad f_{(0)}^{(2i-1)} = f_{(1)}^{(2i-1)} = 0, \qquad i = 1, 2, \cdots, n. \tag{9.4}$$

Then, using the procedure developed in § 6, we obtain

Theorem 8.1. *The regularized solution $f(\alpha; t)$ of (9.2) is defined by the differential equation*

$$(-1)^n f(\alpha; t) + \alpha f^{(2n)}(\alpha; t) - \alpha f^{(2n+2)}(\alpha, t) = (-1)^n g^{(n)}(t) \tag{9.5}$$

and the boundary conditions (9.4) and

$$f^{(2n+1)}(\alpha; 0) = f^{(2n+1)}(\alpha; 1) = 0.$$

Consequently, repeating the argument of § 6, we are able to conclude that the regularization procedure (9.2) can be regarded as a variant of the time series approach in which $\hat{l}(\omega)$ is approximated by a function of the form

$$\lambda_\alpha^{(n)}(\omega) = 1/[1 + \alpha(\omega/\Delta)^{2n}\{1 + (\omega/\Delta)^2\}]. \tag{9.6}$$

Included within this equivalence are the following special cases:

(i) $n = 1$, the numerical differentiation procedure discussed above, and

(ii) $n = 0$, signal extraction in the Sobolev space W_1^2.

The interesting thing to note about this estimate for $\hat{l}(\omega)$ is its dependence on n. As n increases, $\lambda_\alpha^{(n)}(\omega)$ cuts off more sharply the effect of higher frequency components. This is very interesting since we know from the nature of differentiation that it is these higher frequency components which are responsible for its potential numerical instability. This then represents an intuitive explanation of why regularization defines a stable numerical process for optimal α. It should be borne in mind that the validity of this equivalence depends on the assumed boundary conditions (9.4). Since they are not likely to apply in practice and are difficult to impose artificially, end effects will often be observed.

An important consequence of this result is that derivatives of any order can be computed directly (non-sequentially) from the data, thus avoiding the accumulation of rounding error usually associated with sequential methods.

It is clear that these results extend naturally to integral equations of the form

$$\int_0^t k(t-s)f(s)\,ds = g(t), \qquad 0 < s \leq t \leq 1,$$

where

$$k(t-s) = \sum_{i=0}^{I} a_i\,(t-s)^i,$$

if the problem of appropriate boundary conditions is resolved.

We conclude by noting one further generalization. Nothing in the above analysis restricts one to the use of the Sobolev norm $\|\cdot\|_4$, and therefore, if one requires additional smoothing the above analysis could be reworked for the general Sobolev norm

$$\|g\|_{2m}^2 = \|g\|_2^2 + \|g^{(1)}\|_2^2 + \cdots + \|g^{(m)}\|_2^2.$$

The above regularization-spectrum analysis equivalence for the n-th order derivatives remains valid. However, $\lambda_\alpha^{(n)}(\omega)$ must be replaced by

$$\lambda_\alpha^{(n,m)}(\omega) = 1\Big/\left[1 + \alpha(\omega/\Delta)^{2n}\sum_{j=0}^{m}(\omega/\Delta)^{2j}\right].$$

A comparison of the corresponding theoretical spectrum

$$g_v(\omega;\alpha,b) = b\,h(\omega)\left[1 + 1\Big/\left\{\alpha(\omega/\Delta)^{2n}\sum_{j=0}^{m}(\omega/\Delta)^{2j}\right\}\right]$$

with the periodogram of the data must be used as the basis for the choice of m in a given application {see Anderssen and Bloomfield [1]}.

References

1. Anderssen, R. S., Bloomfield, P.: A time series approach to numerical differentiation. Technometrics (in press)
2. Anderssen, R. S., Seneta, E.: On smoothing techniques for the removal of periodic noise of known period. Math. Geol. **3**, 157–170 (1971)
3. Bickley, W. G.: Formulae for numerical differentiation. Math. Gaz. **25**, 19–27 (1941)
4. Birkhoff, G., de Boor, Carl: Error bounds for spline interpolation. J. Math. and Mech. **13**. 827–835 (1964)

5. Björck, Åke, Pereyra, Victor: Solution of Van der Monde systems of equations. Math. Comp. **24**, 893–903 (1970)
6. de Boor, C., Rice, J. R.: Least squares cubic spline approximation, I — Fixed Knots, II — Variable Knots. Computer Science Department Tech. Report #20 and 21, Purdue University, April, 1968
7. Box, G. E. P., Jenkins, G. M.: Time series analysis, forecasting and control. San Francisco: Holden-Day 1970
8. Brodskii, M. L.: Estimation of the remainder term in formulas of numerical differentiation [Russian], Uspekhi Mat. Nauk **13** (6), 73–77 (1958)
9. Burchard, H. G.: Splines (with optimal knots) are better. Applicable Anal. (1973), to appear
10. Chakravarti, P. C.: Truncation error in interpolation and numerical differentiation, Numer. Math. **3**, 279–284 (1961)
11. Cooley, J. W., Tukey, J. W.: An algorithm for the machine calculation of complex Fourier series. Math. Comp. **19**, 297–301 (1965)
12. Cullum, J.: Regularization and numerical differentiation. IBM Research Report RC2605, Sept. 1969
13. Cullum J.: Numerical differentiation and regularization. SIAM J. Numer. Anal. **8**, 254–265 (1971)
14. Dolgopolova, T. F., Ivanov, V. K.: On numerical differentiation. USSR Comp. Math. and Math. Phys. **6** (4), 223–232 (1966)
15. Doob, J. L.: Stochastic processes. New York: John Wiley 1953
16. Engels, H.: Zur Anwendung der Richardson-Extrapolation auf die numerische Differentiation. Computing **8**, 255–272 (1971)
17. Galimberti, G., Pereyra, V.: Numerical differentiation and the solution of multidimensional Van der Monde systems. Math. Comp. **24**, 357–364 (1970)
18. Gentleman, W. M., Sande, G.: Fast Fourier transforms — for fun and profit. Proc. Fall Joint Computer Conference (1966), 563–578
19. Grenander, U., Rosenblatt, M.: Statistical analysis of stationary time series. New York: John Wiley 1957
20. Hamming, R. W.: Numerical methods for scientists and engineers. New York: McGraw-Hill 1963
21. Hamming, R. W.: The frequency approach to numerical analysis. In: Studies in numerical analysis: Papers presented to Cornelius Lanczos. London: Academic Press 1973
22. Hannan, E. J.: Time series analysis. London: Methuen 1960
23. Hannan, E. J.: Multiple time series. New York: John Wiley 1970
24. Hestenes, M.: Calculus of variations and optimal control theory. New York: John Wiley 1966
25. de Hoog, F., Weiss, R.: High order methods for first kind Volterra equations. SIAM J. Numer. Anal. **10**, 647–664 (1973)
26. de Hoog, F., Weiss, R.: On the solution of Volterra integral equations of the first kind. Numer. Math. **21**, 22–32 (1973)
27. Hunter, D. B.: An iterative method of numerical differentiation. Computer J. **3**, 270–271 (1960)
28. Jenkins, G. M., Watts, D. G.: Spectral analysis and its application. San Francisco: Holden-Day 1968
29. Khintchine, A. Ya.: Korrelationstheorie der stationaren stochastischen Prozesse. Math. Ann. **109**, 604–616 (1934)
30. Kranzer, H. C.: An error formula for numerical differentiation. Numer. Math. **5**, 439–442 (1963)
31. Lyness, J. N.: Numerical approximations based on the theory of complex variable. Proc. 22nd Nat. Conf. ACM Publication P-67, (1967), pp. 125–133
32. Lyness, J. N.: Differentiation formulas for analytic functions. Math. Comp. **22**, 352–362 (1968)
33. Lyness, J. N., Moler, C. B.: Numerical differentiation of analytic functions. SIAM J. Numer. Anal. **4**, 202–210 (1966)

34. Lyness, J. N., Moler, C. B.: Van der Monde systems and numerical differentiation. Numer. Math. **8**, 458–464 (1966)
35. Mood, A. M., Graybill, F. A.: Introduction to the theory of statistics, 2nd edition. New York: McGraw-Hill 1963
36. Pallaghy, C. K., Lüttge, U.: Light-induced and H^+-ion fluxes and bioelectric phenomena in mesophyll cells of Atriplex spongiosa. Zeit. für Pflanz. **62**, 417–425 (1970)
37. Powell, M. J. D.: The local dependence of least squares cubic splines. SIAM J. Numer. Anal. **6**, 398–413 (1969)
38. Ralston, A.: A first course in numerical analysis. New York: McGraw-Hill 1965
39. Salzer, H. E.: Optimal points for numerical differentiation. Numer. Math. **2**, 214–227 (1960)
40. Salzer, H. E.: Equally-weighted formulas for numerical differentiation. Numer. Math. **4**, 381–392 (1963)
41. Schönhage, A.: Optimale Punkte für Differentiation und Integration. Numer. Math. **5**, 303–331 (1963)
42. Secrest, Don: Error bounds for interpolation and differentiation by the use of spline functions. SIAM J. Numer. Anal. **2**, 440–447 (1965)
43. Strom, Torsten: An inclusion formula for derivatives. BIT **11**, 196–198 (1971)
44. Tsuda, T.: Numerical differentiation of functions of very many variables. Numer. Math. **18**, 327–335 (1972)
45. Tukey, J. W.: An introduction to the calculations of numerical spectrum analysis. In: The spectral analysis of time series, ed. B. Harris. New York: John Wiley 1967
46. Vedeneev, E. P., Zhidkov, N. P.: The application of the method of regularization to the differentiation of a function of one variable, given as data [Russian]. Numerical Methods and Programming **3**, 255–261 (1969)
47. Wiener, N.: Generalised harmonic analysis. Acta Math. **55**, 117–258 (1930)
48. Wilkinson, J. H., Reinsch, C.: Linear Algebra, Die Grundlehren der mathematischen Wissenschaften, Band 186. Berlin-Heidelberg-New York: Springer 1971

Dr. R. S. Anderssen
Computer Centre, ANU
P.O. Box 4
Canberra, ACT 2600/Australia

Professor P. Bloomfield
Statistics, Fine Hall,
Princeton NJ 08540 USA