J. Phys. E: Sci. Instrum., Vol. 16, 1983. Printed in Great Britain

Optimal smoothing of 'noisy' data by fast Fourier transform

E L Kosarev† and E Pantos‡
† Institute for Physical Problems, USSR Academy of Sciences, Moscow 117334, USSR
‡ Science and Engineering Research Council, Daresbury Laboratory, Warrington WA4 4AD, UK

Received 3 March 1982, in final form 5 January 1983

Abstract. The theory of optimal filtering and smoothing of noisy data is presented. Implementation of this theory is made on the AS/7000 computer at Daresbury Laboratory. The Fortran code and examples of application on 'typical' spectroscopic data are given. The routine size is 35 kbytes and the CPU time 0.11 s for 1024 points.

1. Introduction

Almost any type of analysis of experimental data requires that some kind of data smoothing is carried out. For instance, this may involve a least square fit of the data to some function, a specific one in the fortunate case where the result of an experiment is expected to have a form that can be expressed analytically, e.g. an exponential decay, or to an expression that can be expanded in terms of a basis set of functions. The choice of functions may be based on physical intuition or the experimental conditions. This type of problem is more appropriately described as function fitting or function minimisation prodecure. Several methods are described in the literature (see for instance Wolfe, 1978) that make use of a computer program for finding the best set of functions and parameters. Some of them are available from computer libraries.

The obvious limitation is that in cases where the experimental results (typically a function $\hat{y}=\hat{y}(x)$ where the experimental error $\delta\hat{y}$ is considered to be only in the measurement of the ordinate values $\hat{y}$) can not be approximated by a reasonable analytical expression or where physical intuition is not enough in determining a correct and physically meaningful choice of basis functions such an approach would be quite unsatisfactory.

There are numerous examples from spectroscopy, throughout the range of the electromagnetic spectrum where the true signal may consist of a nonconstant background with superimposed lines of various intensities and degree of overlap. The level of measurement errors (noise) may complicate the appearance of such data even further so that it becomes impossible to extract even simple, but often critical, information like peak positions and intensities and band widths with any measurable certainty.

One approach in these cases is to use a spline interpolation or approximation (Spath 1978), where sections of the experimentally measured function $\hat{y}_i=f(x_i)$ are fitted to a polynomial, usually a cubic. The data is swept from beginning to end fitting a section at a time, the main problem being how to smoothly join (spline) these sections in order to produce a smoothly changing final curve. Again, there are algorithms of varying complexity and efficiency based on the same idea.

The main disadvantage of this approach is that any related statistical information that may exist, either *a priori* from the experimental design and data recording conditions, or *a posteriori* from the noise content of the data, is not utilised in the smoothing procedure. It is thus possible to 'oversmooth' or 'undersmooth'. Oversmoothing is almost certain to take place when the signal has a rather high frequency content while undersmoothing occurs in the presence of 'outliers', i.e. points that are far out from the general trend of neighbouring data points (blunders, in statistical terminology).

Thus, it is necessary to use a procedure which utilises all the information available about our data including the noise statistics. For example, when a Fourier transform technique is used all the information contained in the original data is represented by the amplitudes of the sine and/or cosine functions. All that is needed is a criterion for separating the noise component from the total signal, leaving us with the 'true' or smoothed data. An algorithm that would do this could then be described as a low pass digital filter, since in most situations the noise is represented by the high frequency components. Such an algorithm is described by Inouye *et al* (1969). The implication is that statistical information that exists in some way or another can be used in deciding which frequencies to keep and which to reject. Kaiser and Reed (1977) have described a procedure for finding an appropriate filter function. However, they do not offer a specific algorithm that selects a filter which matches the properties of the signal and the noise.

Since the filtering process is usually carried out at the initial stage of data analysis, possibly on-line if the experiment is under computer control, it is essential that a fast Fourier transform (FFT) algorithm is employed. The Fourier expansion may be done either in terms of cosine or sine functions or both. The aim is to obtain a Fourier spectrum where the coefficients corresponding to the signal decrease and reach the noise level as quickly as possible. The choice between the first two options should be based on the symmetry and analytical properties of the measured function.

This implies that a Fourier filtering program offering these options might have to be interactive, whereupon the experimenter should choose the type of Fourier transform and maybe some kind of data adjustment procedure after displaying the original measured function. FFT algorithms that expand only in terms of sine or cosine functions do exist (Christiansen and Hockney 1971) but we have not used them in the present work. Instead we have made use of the NAG library FFT routine CO6AAF (NAG Library Manual 1975) which expands in both sine and cosine based on Singleton's (1968) FFT procedure.

The most serious problem of using a Fourier filtering program is in the way the truncation in the frequency space, the Fourier spectrum, is to take place. Anyone familiar with Fourier analysis and reconstruction will be aware of the 'Gibbs phenomenon' whereby oscillations can be introduced in the reconstructed data if the high frequency rejection is abrupt, i.e. the Fourier coefficients are simply set to zero from a given frequency k onwards.

This problem is usually treated by multiplying the Fourier coefficients with an appropriate, smoothly changing at the cut-off frequency, edge function. We have experimented with several such edge functions including the Gaussian tails of Inouye *et al* (1969) but quickly came to the conclusion that none could be either general or optimal. In the next two sections we described how optimal filtering may be achieved on a variety of experimental data using an optimal procedure.

We also compare our results with those obtained by a spline approximation algorithm which uses the Harwell library subroutines VC03A and VB06A (Hopper 1978) written by Powell (1967). It will be shown that our approach gives significantly better results. Execution is up to three times faster

and the main storage required is nearly half of that required by the spline program.

Two other important aspects of the use of Fourier filtering for data smoothing, mainly integration within an analysis package and compression of data information with corresponding economies in data storage will be discussed in the final section.

2. Theory of optimal filtering

Most of the results of this theory are applicable not only for Fourier transforms but also for cases of transforms with any other basis functions. Our discussion however will mainly be concerned with Fourier transforms.

The theory is equally applicable if the data is given as a continuous segment $x_1 \leqslant x \leqslant x_N$ or as a discrete set of points $x = x_1, x_2, \ldots, x_N$. Since we are dealing with digitally recorded data the second case will concern us most.

Let us consider the measured values $\hat{f}_i$ of a function $f(x)$ at discrete points $x_1, x_2, \ldots, x_N$ with corresponding measurement error (noise) ε_i

$$\hat{f}_i = f(x_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, N \tag{1}$$

and let us consider the case where there is no systematic error, i.e.

$$\bar{\varepsilon}_i = 0 \tag{2}$$

and that the noise at the different points is uncorrelated, i.e.

$$\overline{\varepsilon_i \varepsilon_j} = \sigma^2 \delta_{ij} \tag{3}$$

where δ_{ij} is Kronecker's delta and the variance of noise σ^2 is the same at all points.

The majority of experiments satisfy the simple requirements (1)–(3). They are important not only for simplicity in the theory but also for efficiency in application, since correlated errors will result in systematic errors in the function $f(x)$.

The function $f(x)$ can be expanded in terms of the basis functions $\{\psi_\alpha(x)\}$

$$f(x) = \sum_\alpha C_\alpha \psi_\alpha(x), \qquad \alpha = 1, 2, \ldots \tag{4}$$

where the functions $\{\psi_\alpha(x)\}$ are orthonormalised

$$(\psi_\alpha, \psi_\beta) = \sum_i \psi_\alpha(x_i)\psi_\beta(x_i) = \delta_{\alpha\beta}. \tag{5}$$

The analytical form of the basis functions is chosen according to the physical problem we encounter, e.g. a set of exponentials if $f(x)$ represents a decay curve, etc.

It is well known (see for example Berezein and Zhidkov (1965) that the minimum approximation error R_M^2

$$\min_{\{C_\alpha\}} R_M^2, \ \alpha = 1, 2, \ldots, M, \ldots \tag{6}$$

where

$$R_M^2 = \sum_i \Big[f(x_i) - \sum_{\alpha=1}^{M} C_\alpha \psi_\alpha(x_i)\Big]^2 \qquad M = 1, 2, \ldots \tag{7}$$

is found when the coefficients C_α

$$C_\alpha = (f, \psi_\alpha) = \sum_i f(x_i)\psi_\alpha(x_i). \tag{8}$$

It is very important to appreciate that $\min R_M^2$ is not the same as $\min \hat{R}_M^2$, where

$$\hat{R}_M^2 = \sum_i \Big[f(x_i) - \sum_{\alpha=1}^{M} \hat{C}_\alpha \psi_\alpha(x_i)\Big]^2, \tag{9}$$

and

$$\hat{C}_\alpha = (\hat{f}, \psi_\alpha) = \sum_i \hat{f}(x_i)\psi_\alpha(x_i). \tag{10}$$

The minimum of $\hat{R}_M^2$ corresponds to the minimum difference between the unknown function $f(x)$ and its approximation. Because we do not know the values of $f(x_i)$ but only the values of $\hat{f}_i$ we cannot use formula (8) but only formula (10).

Because the noise ε_i is random, the values of $\hat{C}_\alpha$, $\hat{f}(x)$ and $\hat{R}_M^2$ will also be random. We will study at first the statistical properties of $\hat{C}_\alpha$. According to the central limit theorem of probability theory (see Brandt 1973) the values of $\hat{C}_\alpha$ have a normal distribution for the broad classes of noise probability distributions and, the mean value of $\hat{C}_\alpha$ is

$$\bar{\hat{C}}_\alpha = \sum_i \bar{\hat{f}}_i \psi_\alpha(x_i) = \sum_i f(x_i)\psi_\alpha(x_i) = C_\alpha \tag{11}$$

i.e. equal to the values of the unknown coefficients C_α. The covariance matrix (or matrix of errors) $D\hat{C}$ of the values of $\hat{C}_\alpha$ is, by definition, equal to

$$(D\hat{C})_{\alpha\beta} = \sigma(\hat{C}_\alpha)\sigma(\hat{C}_\beta)\rho_{\alpha\beta}, \tag{12}$$

i.e. the product of the standard deviations of $\hat{C}_\alpha$ and $\hat{C}_\beta$ and their correlation coefficients $\rho_{\alpha\beta}$. From equation (10) we have

$$(D\hat{C})_{\alpha\beta} = \sigma^2 \delta_{\alpha\beta}. \tag{13}$$

This means that the $\hat{C}_\alpha$'s are uncorrelated. This follows not only from the fact that the noise is uncorrelated (3), but also that the basis functions $\{\psi_\alpha\}$ are orthogonal (5).

Because the $\hat{R}_M^2$'s from (9) are random, strictly speaking, it is necessary to find not $\min \hat{R}_M^2$, but the minimum of the mean value of $\hat{R}_M^2$. For this purpose we introduce in (10) the unknown filter coefficients k_α in order to have a new filtered function

$$\hat{f}_1(x) = \sum_\alpha k_\alpha \hat{C}_\alpha \psi_\alpha(x). \tag{14}$$

Now we have a well-posted mathematical problem: to find coefficients $\{k_\alpha\}$ in order to minimise $\overline{\hat{R}^2}$, where

$$\hat{R}^2 = \sum_i \{f(x_i) - \sum_\alpha k_\alpha \hat{C}_\alpha \psi_\alpha(x_i)\}^2 \tag{15}$$

and

$$\overline{\hat{R}^2} = \sum_\alpha [k_\alpha^2 D(\hat{C}_\alpha) + C_\alpha^2 (1 - k_\alpha)^2], \tag{16}$$

and

$$D(\hat{C}_\alpha) = (D\hat{C})_{\alpha\alpha}. \tag{17}$$

We have

$$\frac{\partial \overline{\hat{R}^2}}{\partial k_\alpha} = 0 \tag{18}$$

and finally

$$k_\alpha = \frac{C_\alpha^2}{C_\alpha^2 + D(\hat{C}_\alpha)}. \tag{19}$$

This important formula corresponds to the well known Wiener filter (WF) in the theory of optimal detection of random signals from random noise when the statistical properties of both signal and noise are known (Wainstein and Zubakov 1962). In our case we have nonrandom signal $f(x_i)$ and only random noise ε_i. In other words we can now state that the Wiener filter (19) gives the minimum $\overline{\hat{R}^2}$ not only for both random signal and random noise, but for nonrandom signal and random noise, too. This is very important for the treatment of physical experiments, because sometimes we can have only nonrandom signal (absorption spectrum of a gas, for example) and random

measurement noise. The use of (19) is a generalisation of the approach previously described by Kosarev (1980).

The minimal value of $\overline{\hat{R}^2}$ is then

$$\overline{\hat{R}^2_{\min}} = \sum_\alpha \frac{C_\alpha^2}{1 + C_\alpha^2 / D(\hat{C}_\alpha)}. \tag{20}$$

Finally, from (14) and (13), we have the expression for the error in the approximation of the function $f_1(x)$

$$\delta \hat{f}_1(x) = 2\sigma \left(\sum_\alpha k_\alpha^2 \psi_\alpha^2(x) \right)^{1/2}. \tag{21}$$

It is seen from (14)–(19) that this expression is approximately equal to the error in the approximation of the unknown function $f(x)$, the one we are interested in. The multiplier of 2 in (21) approximately corresponds to the usually adopted 95% confidence interval (see for example Brandt's (1973) book).

The Wiener filter formula (19) for optimal filtering is only implicit, because it depends on the unknown amplitudes C_α. If we could know the amplitudes C_α we could construct the optimal filter for extracting the signal from the noise. However, this would not be necessary since we would know the amplitudes C_α, hence the signal!

This disadvantage of formula (19) proves to be at the same time its strong advantage. It is at this point in our procedure where we can make use of any *a priori* or *a posteriori* information about the spectrum of the unknown signal.

In this paper we use formula (22) for the WF

$$\hat{k}_\alpha = \frac{\hat{C}_\alpha^2}{\hat{C}_\alpha^2 + D(\hat{C}_\alpha)} \tag{22}$$

if the signal $\hat{C}_\alpha^2$ is larger than the noise,

$$\hat{C}_\alpha \geqslant D(\hat{C}_\alpha) \text{ for } \alpha \leqslant \alpha_0 \tag{23}$$

and the simple straight line extrapolation

$$\ln(\hat{C}_\alpha^2) \approx A_1 \alpha + B_1 \text{ for } \alpha_0 \leqslant \alpha \leqslant \alpha_1 \tag{24}$$

for that part of the spectrum where the signal is less than the noise

$$\hat{C}_\alpha^2 \ll D(\hat{C}_\alpha).$$

This we have found sufficient for most spectroscopic data. There are cases, however, e.g. EXAFS (decaying sinusoid) where the rate of decrease of the Fourier spectrum is best approximated by

$$\ln(\hat{C}_\alpha^2) \approx A_2 \ln \alpha + B_2. \tag{25}$$

This approach gives a smooth decreasing filter function which does not give rise to Gibb's oscillations which would arise if the filter function has an abrupt cut-off.

3. Application to Fourier transform

We use Fourier transform for two reasons. Firstly there are everal standard fast Fourier transform (FFT) algorithms in various computer libraries, and secondly, Fourier transform is a very powerful tool for solving various deconvolution problems which are very often met in experimental situations.

There are, however, two restrictions imposed by standard FFT subroutines:

(a) The data has to be known at equidistant points

$$x_i = \Delta x \cdot i, \; i = 1, 2, \ldots, N \tag{26}$$

and

(b) the number of data N should be equal to a power of two

$$N = 2^m. \tag{27}$$

The first limitation could be avoided by using an interpolation routine before the FFT, but in that case we would spend additional computer time and, what is more important, we would have *correlated* measurement errors. The undesirability of this was explained in § 2. The best way of avoiding interpolation is to make measurements of equidistant points. This is not unduly cumbersome if the data is collected under computer control. For instance, an EXAFS spectrum, usually recorded at equidistant wavelength units, could be recorded at equidistant inverse wavelength units (k-space) if full advantage is to be taken of the FFT as a noise filter as well as an analysis tool.

The second restriction may be remedied by adding zeros, as many as needed, to make up to a power of two†.

If the support interval for the function $f(x)$ is

$$x_1 = a \leqslant x \leqslant f = x_N \tag{28}$$

and the support interval for the new function $f^{(1)}$ is

$$a \leqslant x \leqslant c, \text{ where } c \geqslant bf \tag{29}$$

and

$$f^{(1)}(x) = \begin{cases} f(x) & \text{for } a \leqslant x \leqslant b \\ 0 & \text{for } b < x \leqslant c \end{cases} \tag{30}$$

it is obvious that nothing is lost of the original information.

Standard analysis by Fourier series in the interval (a, b) uses cosine and sine basis function with periods

$$\Delta_k = (b-a)/k, \; k = 1, 2, 3, \ldots \tag{31}$$

and these functions are orthogonal in this interval (a, b). When we use the larger interval (a, c) and the new cosine and sine functions we have periods

$$\Delta_k^{(1)} = (c-a)/k. \tag{32}$$

Hence, only those functions from the new basis functions will be simultaneously basis functions for both support intervals that have indices

$$k^{(1)} = k \frac{c-a}{b-a}, \; k = 1, 2, 3, \ldots, \tag{33}$$

and only these will be orthogonal in the old support interval. We proved in § 2 that only the amplitudes of these Fourier harmonics will be uncorrelated. In fact additional harmonics which arise when we use the new support interval (a, c) are necessary only to represent the additional zeros outside the old interval (a, b). Because of this, in our program we take into account the correlation properties of the new Fourier coefficients. So, from now on we consider that $N = 2^m$ and we assign a new variable n equal to half the number of data points

$$n = N/2. \tag{34}$$

The standard formula for the discrete Fourier transform of the function

$$\hat{y}_j = \hat{f}_{j+1} \tag{35}$$

is

$$\hat{y}_j = \frac{\hat{A}_0}{2} + \sum_{k=1}^{n-1} \left[\hat{A}_k \cos\left(\pi \frac{k_j}{n}\right) + \hat{B}_k \sin\left(\pi \frac{k_j}{n}\right) \right] + \frac{\hat{A}_n}{2} \cos(\pi j) \tag{36}$$

$$j = 0, 1, 2, \ldots, N-1,$$

† NAG MK8 routine now available for any N where prime factor $\leqslant 19$.

where the cosine and sine amplitudes are

$$\hat{A}_k=\frac{1}{n}\sum_{j=0}^{N-1}\hat{y}_j\cos\left(\pi\frac{k_j}{n}\right)$$
$$\hat{B}_k=\frac{1}{n}\sum_{j=0}^{N-1}\hat{y}_j\sin\left(\pi\frac{k_j}{n}\right) \quad (37)$$
$$k=0,1,2,\ldots,n.$$

For $N=2n$ data points we have $(n+1)$ amplitudes for the cosine harmonics and only $(n-1)$ amplitudes for the sine harmonics (because always $\hat{B}_0=\hat{B}_n=0$), and the total number of harmonics equals $2n=N$ precisely.

In order to increase the rate of decrease to zero of the Fourier amplitudes A_k and B_k we subtract from the original data $\hat{y}_l$ the straight baseline

$$\hat{y}'_j=\hat{y}_j-\left(\frac{\hat{y}_N-\hat{y}_1}{x_N-x_1}(x_j-x_1)+\hat{y}_1\right) \quad (38)$$

which is added to the smoothed data for final restoration.

The covariance matrix for the Fourier amplitudes is

$$\mathrm{DF}=\frac{\sigma^2}{n}\begin{pmatrix}2 & & 0\\ & 1 & \\ & \ddots & \\ 0 & & 1\end{pmatrix} \quad (39)$$

instead of (13). For convenience we define two new variables, the noise level

$$\mathrm{NOISE}=2\sigma^2/n \quad (40)$$

since the noise fluctuation level of the sum of the square of the Fourier amplitudes $\hat{A}_k^2+\hat{B}_k^2$ will be approximately equal to $2\sigma^2/n$, and the observed intensity of the Fourier spectrum

$$\hat{S}_k=\hat{A}_k^2+\hat{B}_k^2. \quad (41)$$

The main assumption in this work which results from (3) is that the Fourier spectrum of noise is constant over all frequencies. This assumption was valid for most of the real experimental data on which we applied our algorithm.

The strategy of our algorithm is as follows

(i) Subtract the baseline (38).

(ii) Compute Fourier transform of data $\hat{y}'_j$ using the NAG Library subroutine C06AAF (NAG Library Manual, 1975)

(iii) Calculate σ^2 and NOISE from the high frequency part (the second half) of the Fourier spectrum where the signal amplitude is very much less than that of the noise. It is very important to take into acount here only uncorrelated harmonics in accordance with (33).

(iv) Calculate the Wiener filter in accordance with (22)–(24) from Fourier spectrum of signal and noise together.

(v) Multiply the Fourier amplitudes by the WF coefficients $\hat{k}_\alpha$ (22) and then compute the inverse Fourier transform using C06AAF.

(vi) Restore ordinates by adding the baseline.

The complete listing of the program, based on (i)–(vi) above is published in the Daresbury Laboratory preprint DL/CSE/P10, February 1982 and it is available from the authors. This listing includes only the calling references to the NAG Library subroutine C06AAF and some comments on its use. Instead of that NAG subroutine one can possibly use any FFT subroutines e.g. published in the book 'Programs for Digital Signal Processing', IEEE Press, NY, 1979.

The result will be optimally filtered (smoothed) data. Calculation of the error of approximation can be done by (21), or, more roughly but faster by

$$\delta f_1(x)\sim 2\sigma\sqrt{\frac{M}{N}}, \quad (42)$$

where M equals the number of uncorrelated harmonics used in the reconstruction.

The crucial step in this procedure is (iv). Let us consider it in more detail.

The problem is how to find the points α_0 and α_1 in (23), (24), i.e. the start and the end of 'tail' of the Fourier spectrum $\hat{S}_k$.

A Fourier spectrum of typical spectroscopic data is shown in figure 1 on semilogarithmic scale.

From this figure we can see that our main assumption about the Fourier spectrum of noise is correct – in this example the noise occupies about 90% of the total frequency space with constant intensity. We can also see that our estimate of the noise level from the high frequency part of the Fourier spectrum is correct, too – the mean value of the noise fluctuation is about zero (in logarithmic scale). This example is not exceptional. The Fourier spectra of most of the data we have studied were like that one of figure 1.

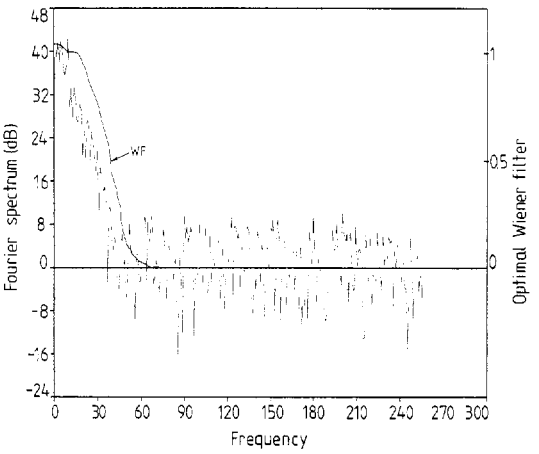

Figure 1. Fourier spectrum $(\hat{A}^2+\hat{B}^2)$/NOISE of typical spectroscopic data and optimal Wiener filter for these data.

In order to find the cut-off between signal and noise it is necessary to study in more detail the statistical properties of $\hat{S}_k$. According to the general theory both $\hat{A}_k$ and $\hat{B}_k$ have normal distributions with mean values

$$\bar{\hat{A}}_k+=A_k \text{ and } \bar{\hat{B}}_k=B_k. \quad (43)$$

If at some k the Fourier amplitudes are zero

$$A_k=B_k=0 \quad (44)$$

then for that frequency there is no signal but only noise. In this case the observed spectral intensity divided by NOISE has a χ^2 distribution with two degrees of freedom

$$\hat{S}_k/\mathrm{NOISE}\sim\chi_2^2 \quad (45)$$

(see Brandt (1973) § 6.5).

The probability density of a random value

$$x=\chi_2^2 \quad (46)$$

equals

$$p(x)=\tfrac{1}{2}\exp(-x/2),\ 0\leqslant x<\infty \quad (47)$$

and the distribution function $P(x)$

$$P(x)=\text{Prob}(\chi^2\leqslant x)=\int_0^x p(x)\,dx=1-\exp(-x/2). \quad (48)$$

Using this formula we compute the values of table 1, i.e. the values of probability that

$$\hat{S}_k/\text{NOISE}\leqslant x. \quad (49)$$

Table 1. Probability that $\chi_2^2\leqslant x$ and for $\chi_2^2>x$.

x	$P(x)$ (%)	$1-P(x)$ (%)	$10\lg x$ (dB)
0.001	0.05	99.95	−30
0.01	0.5	99.5	−20
0.1	4.9	95.1	−10
1.0	39.3	60.7	0
10.0	99.3	0.67	+10
10.6	99.5	0.5	10.25
13.9	99.9	0.1	11.4

We can understand from this table that approximately in 5% of the cases the value of $\hat{S}_k$/NOISE will be less than the −10 dB level and only in less than 0.7% of cases this value would be exceeding the +10 dB level.

All of these predictions are realised in figure 1. One is reminded that this table corresponds to case (44) when there is no signal but noise only.

For the starting point of the tail we choose the point j_0 where for the first time

$$\hat{S}_{j_0}<\text{NOISE} \quad (50)$$

and for end point j_1 we choose the point where the straight line extrapolation (24) decreases below the −20 dB level S/N = 0.01). The parameters A_1 and B_1 in (24) are found by least squares method

$$A_1=\frac{xy-j_0x_My_M}{xx-j_0x_M^2},\ B_1=y_M-A_1x_M, \quad (51)$$

where

$$xy=\sum_{i=1}^{j_0} i\hat{S}_i,\ xx=\sum_{i=1}^{j_0} i^2,\ x_M=(1+j_0)/2,\ y_M=\sum_{i=1}^{j_0}\hat{S}_i/j_0. \quad (52)$$

4. Outliers rejection

The procedure outlined above has been found to work extremely well with spectroscopic data and random uncorrelated noise. In the cases where apart from the random noise we also have the presence of 'outliers' i.e. points far removed from the mean of the normal distribution, it is necessary to identify them and reject them during restoration.

The implemented algorithm includes a test for such points; that is, after the first restoration the residuals are calculated and if any of the residual amplitudes exceeds a threshold value the smoothing procedure is repeated but this time the outlier points have been replaced by the smoothed ones from the previous run. This is repeated until no outliers remain.

The value of the threshold residual amplitude R strictly speaking is an increasing function of the total number of data points N.

In the tables of Pearson and Hartley (1956) this function was tabulated at $1\leqslant N\leqslant 30$. For $N>30$ we compute this function by the formula

$$R=\Psi[(1-\varepsilon/2)^{1/N}] \quad (53)$$

where Ψ is inverse probability integral

$$\Phi[\Psi(p)]\equiv p,\ \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-t^2/2)\,dt \quad (54)$$

and ε is probability to reject random noise points but not outliers (see book of Bolshev and Smirnov (1968)). At $\varepsilon=1\%$ and $64\leqslant N\leqslant 4096$ there is a very simple approximation

$$R=3.8+0.15(\log_2 N-6) \quad (55)$$

which has precision not worse than 0.05, and this approximation we use in the program.

5. Examples

Figures 2 and 3 give a fine example of the usefulness of our

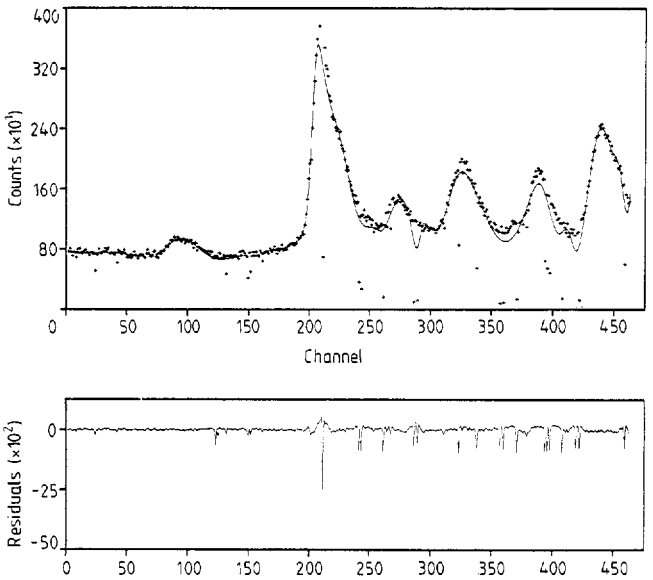

Figure 2. Splined smoothed (line) and original (crosses) data. 476 points, CPU time 1.93 s.

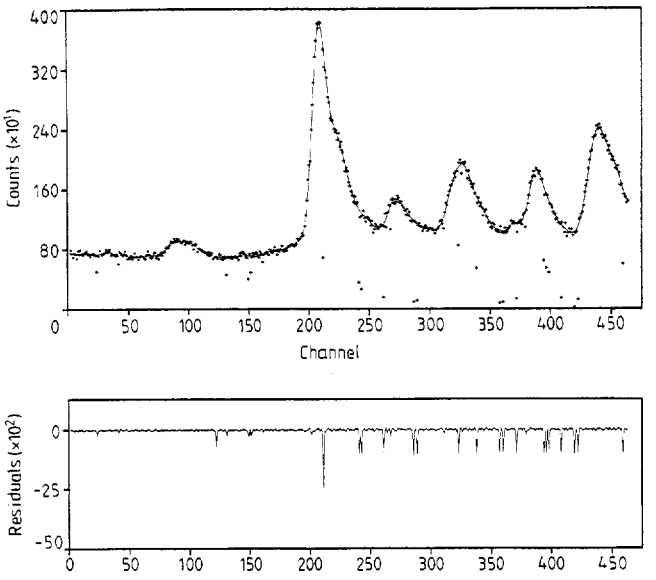

Figure 3. Wiener filtered (line) and original (crosses) data. 26 outliers removed in eight iterations. 476 points, CPU time 0.47 s. The Fourier spectrum is given in figure 1. Frequency $j_0=37$.

approach compared to a conventional spline fitting procedure. The data represent the phosphorescence excitation spectrum of Naphthalene in a rare gas matrix (Hamilton and Najbar (1981)). Apart from the usual statistical noise there are several outlier points cause by a fault in the recording system. The only way a spline method could be employed satisfactorily would be by 'editing out' the outliers or by using appropriate weighting factors for the corresponding points. This would be a laborious and time consuming exercise and always specific to one data set. The approach described above not only results in a smooth curve through the real experimental points but also rejects the outliers automatically. Notice that although eight iterations were needed the CPU time required was less than for the spline. Fourier spectrum of these data and optimal Wiener filter are presented in the figure 1.

Figures 4 and 5 give an example of a dataset with many sharp peaks (hydrogen emission spectrum) and relatively low noise level. The spline method drastically oversmooths the peaks and valleys between peaks whilst the Wiener filter method retains the experimental resolution. Notice also that in this case only one iteration was used with significant time advantages.

Figure 6 tells us there are cases where the assumption that the second half of the Fourier spectrum represents noise only is no longer true. This figure presents the Fourier spectrum of the data from figures 4 and 5 if the noise level is estimated according to point 3 of the algorithm strategy from the second half of the Fourier spectrum. In this case however the noise occupied much less than the whole second half of the Fourier spectrum. The cut-off frequency may be better selected interactively to estimate correctly the noise level.

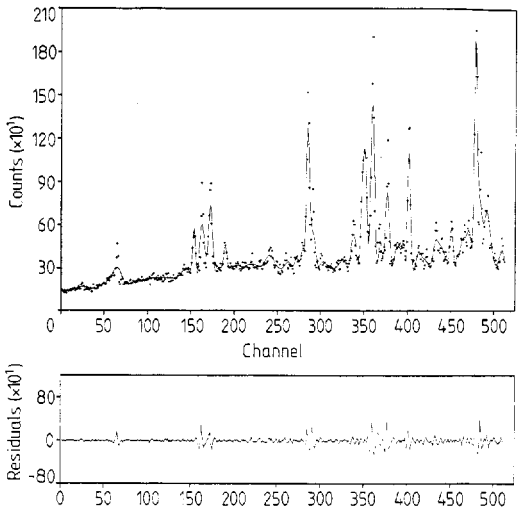

Figure 4. Splined smoothed (line) and original (crosses) data. 512 points, CPU time 5.57 s.

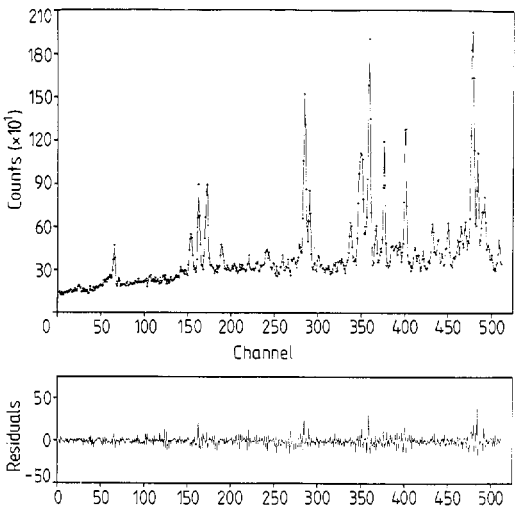

Figure 5. Wiener filtered (line) and original (crosses) data. 512 points, CPU time 0.05 s. Frequency $j_0 = 105$.

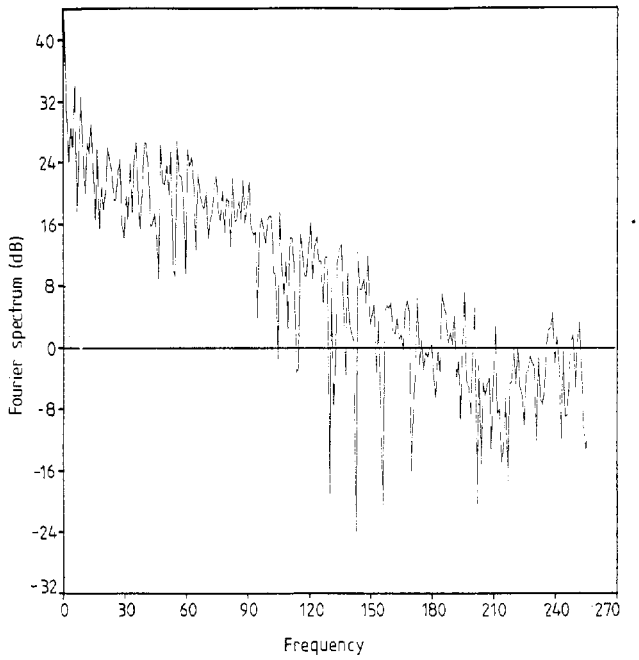

Figure 6. Fourier spectrum $(\hat{A}^2 + \hat{B}^2)$/NOISE of the data presented in the figures 4 and 5.

6. Conclusions

We have presented an optimum and CPU efficient procedure for automatic data smoothing. Optimum of course does not mean the one and only correct and appropriate way to smooth any spectroscopic data. The restrictions imposed by (3), (26) and (27), normally irrelevant in a spline procedure, should be borne in mind.

If made interactive with intermediate graphic displays of the FT, WF and reconstructed data, a physically meaningful selection process for the cut-off frequency is made available in contrast to other more heavy-handed procedures (Cutkosky 1981).

A most important aspect of this method is that in cases of data collected at an experimental station by a minicomputer, hardware or assembler coded fast Fourier transforms may be available. This would increase execution speed even more.

Another spin-off is that since only a small part of the Fourier spectrum is used in the reconstruction (~20%) the smoothed dataset can be stored in the form of its truncated Fourier spectrum.

This data compression technique is used widely in the cases of two dimensional data sets (e.g. digital images) where only a small portion of the transform spectrum is used either in transmission over long distances (satellites) or for mass storage (Landsat photographs).

Finally, in the context of 2D image processing now, the

Weiner filter algorithm we have implemented can be extended in 2D noise filtering and image restoration problems.

The program version used at the Daresbury Laboratory includes interactive graphics (program SMFFT) with a data reading routine that can handle any data produced at experimental stations at the Synchrotron Radiation Source and is part of the SRS Program Library (Pantos 1981).

Acknowledgments

The authors wish to thank the Daresbury Laboratory and the Committee of Synchrotron Radiation of the USSR Academy of Sciences for supporting exchange visits to Daresbury and Moscow, respectively, in connection with this work.

References

Berezin M A and Zhidkov N P 1965 *Computing Methods*, vol. **1** translated from the Russian (Oxford: Pergamon) chap 5

Bolshev L N and Smirnov N V 1968 *Tables of Mathematical Statistics* (In Russian) (Moscow: Nauka)

Brandt S 1973 *Statistical and Computational Methods in Data Analysis* (Amsterdam: North-Holland)

Christiansen J and Hockney R W 1971 FOUR67: a fast Fourier transform package
Comput. Phys. Commun. **2** 127–38

Cutkosky R E 1981 Spline interpolation and smoothing of data
Comput. Phys. Commun. **23** 287–99

Hamilton T D S and Najbar J 1981 Unpublished results

Hopper M J 1978 Harwell Subroutine library. A Catalogue of Subroutines (Harwell: AERE)
Harwell Report AERE R9185

Inouye T, Harper T and Rasmussen N C 1969 Application of Fourier transforms to the analysis of spectral data
Nucl. Instrum. Meth. **67** 125–32

Kaiser J F and Reed W A 1977 Data smoothing using low-pass digital filters
Rev. Sci. Instrum. **48** 1447–57

Kosarev E L 1980 Applications of the first kind integral equations in experimental physics
Comput. Phys. Commun. **20** 69–75

NAG Library Manual 1975 *NAGFLIB 1061/571: Mk. 5*

Pantos E 1981 *The SRS Program Library*, Issue 1, Daresbury Laboratory Preprint DL/SCI/P346E

Pearson E S and Hartley H O (ed.) 1956 *Biometrika Tables for Statisticians*, vol. 1 (Cambridge: Cambridge University Press)

Powell M J D 1967 Curve fitting by cubic splines (Harwell: AERE) *Harwell Publication TP307*

Programs for Digital Signal Processing 1979 (New York: IEEE Press)

Singleton R C 1968 Algorithm 338 Algol procedure for the fast Fourier transform
Commun. ACM **11** 773–6

Spath H 1978 *Spline-Algorithmen zur Konstruktion glatter Furven und Flachen* (Munchen: R. Oldenbourgh Verlag)

Wainstein L A and Zubakov V D 1962 *Extraction of Signals from Noise;* translated from the Russian (Englewood Cliffs, N.J.: Prentice-Hall)

Wolfe M A 1978 *Numerical Methods for Unconstrained Optimization* (New York: Van Nostrand)