AAEC/E398

AAEC/E398

AUSTRALIAN ATOMIC ENERGY COMMISSION

RESEARCH ESTABLISHMENT

LUCAS HEIGHTS

RESOLUTION UNFOLDING WITH LIMITS IMPOSED BY STATISTICAL EXPERIMENTAL ERRORS

by

D.W. LANG

February 1977

ISBN 0 642 99776 4

AUSTRALIAN ATOMIC ENERGY COMMISSION
RESEARCH ESTABLISHMENT
LUCAS HEIGHTS

RESOLUTION UNFOLDING WITH LIMITS IMPOSED BY STATISTICAL EXPERIMENTAL ERRORS

by

D.W. LANG

ABSTRACT

A typical form of the resolution equation is derived by considering the physical measurement of an energy dependent spectrum. It is shown that the information contained in a data set may be expressed by writing the spectrum as a linear combination of a set of resolution functions. Introduction of other functions to describe the spectrum involves extra physical information.

An iterative conjugate gradient technique to obtain a spectrum consistent with the data is described. At each iteration the residual discrepancy between the currently predicted yield and the measured data is used to generate the form and magnitude of the next term to be added to the spectrum. Other unfolding techniques are described and analysed, some faster than the conjugate gradient technique in special cases, but restricted in usefulness by implicit assumptions about the resolution functions.

The nature of residual errors is considered. The variations of independently measured data sets are discussed, and hence, the variations of the sequence of terms appearing in a consequent conjugate gradient analysis. An approximate measure is obtained for the expected variation of independently obtained spectra. Refinements are briefly considered

which apply to a resolution function that is not known precisely or which make use of a requirement that the spectrum be positive throughout its range.

It is concluded that a conjugate gradient technique is best if sufficient computer facilities are available, and that, of the less demanding techniques, the best is one that is essentially a more slowly convergent version of a conjugate gradient method.

National Library of Australia card number and ISBN 0 642 99776 4

CONTENTS

(continued)

CONTENTS (Continued)

1. INTRODUCTION

There are some simple physical experiments in which the concept of a resolution function is not needed. For example in an apparatus to investigate gas laws, where a sample of gas is taken and readings made of its pressure, temperature and volume, uncertainties of pressure, say, are usually errors of reading only and are large compared with the variations of pressure at different points in the volume. This fortunate situation is however the desirable exception.

We consider now a case at almost the other extreme. A person at the bank of a swift muddy stream wishes to choose a safe path across. The depth of water at a given point influences the shape of the surface above that point. It is possible, to some extent, to infer the shape of the bed of the stream from the shape of its surface but there are limitations. For example, the surface shape over a deep hole is much the same as it is over a very deep hole in the bed. Any such difference may be impossible to measure because of random surges of current or other disturbance of the surface. In almost all cases the surface is smoother than the bottom. Uncertainties in the measured surface are multiplied in inferences about the bottom.

Also there is a systematic effect. Structure on the bed is often manifest downstream, but not far upstream. Such features are common wherever resolution limits information. There is a final universal point from this example. A more sensitive experiment in a smaller region is an important check on the information available by analysis of the given data from any experiment with poor resolution. The wise traveller, after choosing a path, takes a stick along and probes the bottom in following the chosen route.

Problems of treatment of resolution functions have been considered for over a century[1]. Rayleigh[2] noted that lines in a spectrum could be considered resolved as separate if the central line of their combined diffraction pattern had a perceptible minimum. The astronomical problem of discovering structure in distant objects has been attacked using progressively more refined instrumentation[1,3,4]. While waiting for better instruments, progress has been made by measuring profiles and resolution functions to enable unfolding, i.e. recovery of available information. The calculation involved, which has become steadily more feasible with the growth of electronic computation, is the subject of this report.

An investigation of the information that can be recovered from the data in cases where the resolution function cannot be ignored must also cover the methods by which the unfolding can be effected. Section 2 transforms the problem that is posed by experiment into its mathematical equivalent. Section 3 is a discussion of the information that exists in the experimental data. Section 4 demonstrates the conjugate gradient technique that has been found effective in extracting the available information, and in Section 5 some techniques are described which, when applicable, can be faster. The procedures used must, at all stages, bear a sensitive relation to the experimental errors, and the general theme of errors is considered separately in Section 6. Some special examples are considered in Section 7, and Section 8 is a brief examination of the effect of constraints on the form of the quantity to be recovered from the data. A number of results needed in the text are treated more fully in the appendices.

2. TYPICAL RESOLUTION PROBLEM

Consider a spectrum $s(E)$ of particles ranging in energy from E_L to E_u so that the number per unit time in the range E to $E + dE$ is proportional to $s(E)dE$. A detecting system is set to a voltage V_j to record those particles incident with, say, some corresponding energy E_j. The actual response of the detector system $R(V_j,E)$ is such that the yield y_j or $y(V_j)$ is given by

$$y_j = K \int_{E_L}^{E_u} R(V_j,E)s(E)dE \qquad (2.1)$$

where the constant K is a warning that there may be considerable complication, e.g. in geometry, to be considered.

In equation 2.1 the assumption is made, and will be made throughout this report, that the problem is linear, i.e. that there is no response which depends on higher powers of $s(E)$. Except in a simple case, covered in Section 7, the resolution function is to be treated as known. All systematic effects are assumed to be incorporated in the resolution functions. It may even be known as a composite of known functions. Thus the recorded response may be composed of a probability $r(V,E)$ that a particle of energy E gives rise to an output voltage V from the detector, and a probability $P(V_j,V)$ that such a voltage V is accepted by the recorder at the setting V_j.

Then

$$R(V_j,E) = \int r(V,E)P(V_j,V)dV \quad . \tag{2.2}$$

The two response functions r and P are 'folded' here into the single response function R and the unobserved variable V usually plays no further part in discussion.

It is convenient experimentally if the resolution function is closely related to a delta function

$$R(V_j,E) \approx \delta(E_o V_j - V_o E) \tag{2.3}$$

so that a range of readings for y_i can be seen to represent s(E) to good approximation. Even for such an almost ideal case, if the resolution function is well enough known, corrections can be applied and a marginally better description obtained for the spectrum. In many cases however the resolution function $R(V_j,E)$ is less tractable, and considerable processing of data is necessary before anything resembling the spectrum s(E) is available.

Equation 2.1 is formally a Fredholm integral equation of the first kind[5,6]. A complete knowledge of y(V) might establish the sort of functional behaviour required[7] for analytical inversion of the equation to obtain s(E). We measure y(V) at a finite set of points and the equation remains formal.

At each point a given measurement m_j cannot be expected to be exactly y_j. In the special case of particle detection m_j is commonly an integral number of pulses and y_j ranges over the real numbers. A set of measurements can be assumed to establish a mean value for m_j and a standard deviation σ_j from that value. For a very large number of measurements the mean is required under this assumption of linearity to approach y_j. Given a particular measurement, it can be assumed to be the actual yield y_j plus a random error ε_j;

$$m_j = y_j + \varepsilon_j \tag{2.4}$$

or

$$m_j = \int_{E_L}^{E_u} R(V_j,E)s(E)dE + \varepsilon_j \quad . \tag{2.5}$$

Equation 2.5 appears explicitly or is implied in unfolding in many investigations[8-83]. Its form is unchanged if V and E are replaced by multidimensional variables. Naturally the computation for unfolding can be expected to take longer if more dimensions lead to more data points.

More dimensions can also require a change of computational method. In Section 5 we note a special case of an unfolding method that works in one dimension but fails if the spectrum is intrinsically two-dimensional.

Equation 2.5 can still apply where the spectrum variable $E^{25,26,30,51,62-67}$ or the measurement variable $V^{32,36,52-54,59,60,70-72,78}$ can have only a discrete set of points. An example of the first occurs if data come from gamma detection as a means of measuring concentration of radiation species, and of the second, where the data are foil activities from which we can recover information about a neutron flux as a function of energy. With suitable care about implicit assumptions, methods of solution usually cope with these cases unchanged.

It is widely known that meaningful consideration of data governed by the equation 2.5 requires simultaneous consideration of the errors. In later sections it is shown that the conjugate gradient technique to arrive at the form of s(E) has considerable advantages over other techniques that have been employed. Special attention is given to the techniques that have been called 'least structure solutions' and 'appropriate solutions', and to cases that can be simplified by Fourier transformation of the equation 2.5. Returning briefly to our special example, we note that a component of the resolution function may be the efficiency $\varepsilon(E)$ with which particles of a given energy are detected. We are at liberty in mathematical techniques to detach such a function when convenient from the resolution function and use a 'detected spectrum' $\varepsilon(E)s(E)$. Such detachments are common in dealing with Fourier transforms of resolution functions.

3. INFORMATION AVAILABLE

From a typical resolution experiment we obtain readings m_j at say, N data points, and also associated standard deviations σ_j. We have

$$m_j = \varepsilon_j + \int_{E_L}^{E_u} R(V_j,E)s(E)dE, \quad j = 1, \ldots N \quad . \qquad (3.1)$$

Obviously, at most N independent parameters can be obtained, which, together with associated information about their errors, describe the spectrum s(E). The aim here is to find a few functions which may be combined to give an adequate description of the data. The usual measure of goodness of fit is chi-squared;

$$\chi^2 = \sum_j^N \frac{(m_j - y_i)^2}{\sigma_j^2} \quad . \tag{3.2}$$

For an iterative process that evaluates the component in the data of one new function at each iteration an acceptable value of chi-squared after n iterations is N - n. In this statement each new function is assumed to be independent of all those before it, as is now defined.

For each member of a set of M trial functions,

$$t_1(E), \ \ldots \ t_k(E), \ \ldots \ t_M(E) \tag{3.3}$$

a set of matrix elements P_{jk} is defined as the set of responses at the data points V_j

$$P_{jk} = \int_{E_L}^{E_u} R(V_j,E)t_k(E)dE \quad . \tag{3.4}$$

Two functions $t_1(E)$ and $t_2(E)$ are defined as conjugate if and only if

$$\sum_j \frac{P_{j1}P_{j2}}{\sigma_j^2} = 0 \quad . \tag{3.5}$$

Given a set of functions we can construct from them a set whose members are conjugate in pairs by the Gram-Schmidt technique;

for if $$\sum_j \frac{P_{j1}P_{j2}}{\sigma_j^2} = A \tag{3.6}$$

and $$\sum_j \frac{P_{j1}^2}{\sigma_j^2} = B \tag{3.7}$$

then $$t_2^1(E) = t_2(E) - (A/B)t_1(E) \tag{3.8}$$

and $t_2^1(E)$ is conjugate to $t_1(E)$. At most N functions of the set can be constructed with at least one of the P_{jk} non-zero for any k, and any pair conjugate. Any reasonably well-behaved function can then be expressed as a linear combination of the set of N functions plus a remainder function which by itself would exhibit no effect at any of the data points. In many areas of interest such a remainder function must be found with a background of detectable functions to make up a physically reasonable spectrum; the cry of a bat is an example of a physical

signal that can be undetectable using the resolution function of a human ear.

To make up an initial set of conjugate functions we could concentrate attention on a set $\{E_k\}$ of discrete points. Hence we can write alternative forms

$$t_k(E) = \delta(E-E_k) \tag{3.3a}$$

or a histogram based on the same set of points

$$t_k(E) = \frac{2}{E_{k+1}-E_{k-1}} \quad \text{for} \quad \frac{E_k+E_{k-1}}{2} < E < \frac{E_k+E_{k+1}}{2}$$

$$= 0 \qquad \text{otherwise} \quad . \tag{3.3b}$$

These two function forms have a considerably different mathematical structure, but are physically indistinguishable if the points are taken close enough together. Theis[75] has shown that well behaved spectrum functions can be approximated either way if the Taylor series for s(E) or for R(V,E) in terms of E, converges rapidly enough. We return to this point in subdividing intervals in Section 7.

In what follows, matrix equations are constructed that implicitly depend on 3.3a. For graphs of results, the form 3.3b or an even smoother form is assumed and any required function in use is defined in terms of the form 3.3a. In whatever way they are represented, the set $\{R(V_j,E)\}$ of actual resolution functions is of fundamental importance[8]. We consider any spectrum function s(E). There is an infinite variety of functions s(E) that give the same predicted yield at the data points. Of these, there is one that has a minimum value of the integral

$$\int_{E_L}^{E_u} s^2(E)dE \tag{3.9}$$

which is a common measure of the magnitude of s(E) considered as a vector, and which, because of the Schwarz lemma, can be considered to be a measure of the complication in s(E). We demonstrate that this smoothest function of the set is a linear combination of the set $\{R(V_j,E)\}$. We can analyse any s(E) by the Gram-Schmidt process as the required linear combination of $\{R(V_j,E)\}$ plus a function f(E) that is conjugate to all members of the set $\{R(V_j,E)\}$. In construction it may emerge that the set $\{R(V_j,E)\}$ is not linearly independent so that there is a relationship of linear dependence among some of the y_j, independent of the form

of the spectrum s(E). We write

$$s(E) = s_0(E) + f(E) \tag{3.10}$$

with $$s_0(E) = \sum_k c_k R'_k(E) \tag{3.11}$$

and the set $\{R'_k(E)\}$ is a mutually conjugate set of linear combinations of the set of $\{R(V_j,E)\}$. We note that s(E) determines the predicted set of yields $\{y_j\}$. In turn these yields determine the values of the set $\{C_k\}$. The form of $s_0(E)$ is thus unambiguous. If the set $\{R(V_j,E)\}$ is linearly dependent and we write

$$s(E) = \sum_k b_j R(V_j,E) + f(E) \quad , \tag{3.12}$$

some of the coefficients b_j may not be well defined, but the functions $s_0(E)$ and hence f(E) are. If the set $R(V_j,E)$ is linearly independent we can obtain a function $s_0(E)$, with the value of the b_j unambiguous, to fit any data set, or set $\{y_j\}$ of predicted yields. For any s(E) we can obtain the equation 3.10 and substitute in the expression 3.9. Then

$$\int s^2(E)dE = \int s_0^2(E)dE + 2\int s_0(E)f(E)dE + \int f^2(E)dE \quad . \tag{3.13}$$

The first term $\int s_0^2(E)dE$ is independent of f(E) and is determined by the given yield prediction. The second term is zero since f(E) is conjugate to every term in $s_0(E)$. The third term can be minimised by making f(E) everywhere zero.

We conclude that the data specify a form of the spectrum s(E) in terms of the resolution functions at the data points. It may be convenient to introduce other functions in the description of the spectrum, but this convenience is not inherent in the data.

An example of a set of response functions that are linearly dependent can be arranged with a high band-width tape recorder and input from a low band-width microphone. An analysis of the recording into its Fourier components involves linear combinations of magnetic intensity at points spread along the tape. Certain linear combinations (components of frequencies outside the microphone band-width) must be zero to within a value corresponding to a noise level of the tape recorder. In Appendix B, an artificial example illustrates the use of linear dependence to check the assigned standard deviations of the data.

For convenience a special set of trial functions is introduced which is closely related to the resolution functions $\{R(V_j,E)\}$. A

function $\nu(E)$ is defined as associated with the effect of error at each point E.

Suppose that for some s(E) a perfect fit to the data has been obtained. The value of χ^2 in equation 3.2 is then zero. We define $\nu(E)$ so that $s(E) + \delta(E-E')\nu(E')$ gives rise to a χ^2 of unity. The value of $\chi^2 = 0$ is obviously a minimum and hence there is no interference term in the new expression for χ^2;

$$\chi^2 = 1 = \nu^2(E') \sum_j \frac{R^2(V_j,E')}{\sigma_j^2} \tag{3.14}$$

$$\nu(E) = \left[\sum_j \frac{R^2(V_j,E)}{\sigma_j^2} \right]^{-\frac{1}{2}} \quad \text{for } \nu(E) \text{ finite} . \tag{3.15}$$

If at some energy E all the $R(V_j,E)$ are zero we have chosen E_L and E_u, the limits of integration, badly and we set $\nu(E) = 0$ there. The set of trial functions

$$\{\nu(E)\ R(V_j,E)\} \tag{3.16}$$

is then useful as a basis for the expansion of s(E). It is, of course, linearly independent if $\{R(V_j,E)\}$ is linearly independent.

These functions of the set 3.16 are largest where the value of χ^2 is least sensitive to the value of s(E), i.e. where a fitting process must make large adjustments at each iteration of the procedure if a good fit is to be obtained in a few iterations.

4. THE USE OF CONJUGATE GRADIENTS

Suppose $\{t_k(E)\}$ is a set of M trial functions and P_{jk} their associated yields. If a linear combination of the trial functions describing the data is desired, we write

$$s(E) = \sum_k a_k\ t_k(E) \quad . \tag{4.1}$$

The set of numbers $\{a_k\}$ can be arranged in order as a row vector $\overleftarrow{a}$ or as a column vector $\vec{a}$ in the M dimensional 'spectrum space'. We may similarly set up a 'yield space' with typical column vectors $\vec{y}$ and $\vec{m}$ having j^{th} elements y_j and m_j respectively.

The elements P_{jk} make up a rectangular N x M matrix P so that

$$y_j = \sum_k P_{jk}\ a_k \tag{4.2}$$

or

$$\vec{y} = P\ \vec{a} \tag{4.3}$$

and

$$\overleftarrow{y} = \overleftarrow{a}\ P^T \tag{4.4}$$

where the superscript T refers to the transpose operator which changes rows into columns and vice versa

$$(P^T)_{jk} = P_{kj}$$

$$\vec{a} = \overleftarrow{a}^T = (\vec{a}^T)^T \quad . \tag{4.5}$$

A number of results are more easily understood if written in 'data significance space' obtained from 'yield space' by dividing each yield element by the standard deviation of the data. We can introduce a diagonal weight matrix w

$$w_{jk} = \delta_{jk}/\sigma_j \quad . \tag{4.6}$$

Weighted data $\vec{g}$ is indicated by

$$g_j = \sigma_j^{-1} m_j \tag{4.7}$$

or

$$\vec{g} = w \vec{m} \quad , \tag{4.8}$$

weighted yields by

$$\vec{x} = w \vec{y} \tag{4.9}$$

and weighted errors by

$$e_j = (\vec{e})_j = w_{jj}\epsilon_j \quad . \tag{4.10}$$

The w and P matrices are combined into the Q matrix

$$Q = w P \tag{4.11}$$

$$Q_{jk} = \sigma_j^{-1} P_{jk} \quad . \tag{4.12}$$

Then

$$\chi^2 = \sum_j \left(\frac{m_j - y_j}{\sigma_j} \right)^2$$

$$= (\overleftarrow{m} - \overleftarrow{y})\, w^2\, (\vec{m} - \vec{y}) \tag{4.13}$$

$$= (\overleftarrow{m} - \overleftarrow{a}\, P^T)\, w^2\, (\vec{m} - P\vec{a}) \tag{4.14}$$

$$= (\overleftarrow{g} - \overleftarrow{a}\, Q^T)\, (\vec{g} - Q\vec{a}) \quad . \tag{4.15}$$

χ^2 is a scalar product or a 1 x 1 matrix, and since our quantities so far are all real, is its own transpose.

We now consider how to reach a minimum of χ^2. Suppose $\vec{a}$ is changed to $\vec{a} + \mu\vec{b}$ where μ is a multiplier whose value is to be determined. The new value of χ^2 is

$$\chi^2 = (\overleftarrow{g} - \overleftarrow{a}\, Q^T)(\vec{g} - Q\vec{a}) - 2\mu\overleftarrow{b}\, Q^T(\vec{g} - Q\vec{a}) + \mu^2 \overleftarrow{b}\, Q^T Q\vec{b} \quad . \tag{4.16}$$

If

$$\overleftarrow{b}\, Q^T Q\vec{a} = 0 \quad , \tag{4.17}$$

a minimum is obtained for χ^2 with respect to μ where

$$\mu = \overleftarrow{b}\, Q^T \vec{g} / \overleftarrow{b}\, Q^T Q\vec{b} \quad . \tag{4.18}$$

If equation 4.17 has not already been satisfied, substituting

$$\vec{b}' = \vec{b} - \vec{a}(\overleftarrow{b}\, Q^T Q\vec{a} / \overleftarrow{a}\, Q^T Q\vec{a}) \tag{4.19}$$

for $\vec{b}$ will do so.

If the expression for χ^2 in equation 4.15 is already the minimum then for all vectors $\vec{b}$

$$\overleftarrow{b}\, Q^T(\vec{g} - Q\vec{a}) = 0 \quad . \tag{4.20}$$

It may be that

$$\vec{g} = Q\vec{a} \tag{4.21}$$

in which case

$$\chi^2 = 0 \quad , \tag{4.22}$$

or quite commonly the lesser condition

$$Q^T(\vec{g} - Q\vec{a}) = 0 \quad . \tag{4.23}$$

Thus in least squares analysis a vector $\vec{a}$ with a small number of components is usually calculated from

$$\vec{a} = (Q^T Q)^{-1} Q^T \vec{g} \quad . \tag{4.24}$$

When M, the number of components in $\vec{a}$, is smaller than N, the inverse $(Q^T Q)^{-1}$ is usually well defined.

As M approaches N the matrix $Q^T Q$ can be expected to become ill-conditioned. If N = M and the matrix is not ill-conditioned then Q^{-1} exists as well as $(Q^T Q)^{-1}$ and the vector $\vec{a}$ is uniquely determined from equation 4.21.

We now examine a typical step of an iterative process. A superscript n in parenthesis indicates the value for some quantity arrived at by n iterations; thus $\vec{a}^{(n)}$ is the spectrum vector after n iterations. To generate the next component vector $\vec{b}^{(n+1)}$ several requirements are imposed. The vector direction is sought giving most rapid improvement of the value of χ^2. Setting $\overleftarrow{b}^{(n+1)}.\vec{b}^{(n+1)}$ to unity the variation of χ^2 with the addition of a small component $\varepsilon\vec{b}^{(n+1)}$ is maximised. If then a minimum of the quantity

$$\varepsilon^{-1}\ (\lambda\varepsilon(\overleftarrow{b}^{(n+1)}.\overrightarrow{b}^{(n+1)}-1) + (\overleftarrow{g}-\overleftarrow{a}^{(n)}Q^T - \varepsilon\overleftarrow{b}^{(n+1)}Q^T).(\overrightarrow{g}-Q\overrightarrow{a}^{(n)} - \varepsilon\overrightarrow{b}^{(n+1)}))$$

(4.25)

is required, with respect to the parameter λ and to the components of $\overrightarrow{b}$ as ε tends to zero, immediately

$$\overrightarrow{b}^{(n+1)} \propto Q^T(\overrightarrow{g}-Q\overrightarrow{a}^{(n)}) \quad . \tag{4.26}$$

If w' is a weighting matrix and the minimum value of $\overleftarrow{b}^{(n+1)}w'^Tw'\overrightarrow{b}^{(n+1)}$ is required the relationship obtained is

$$w'^Tw'\overrightarrow{b}^{(n+1)} \propto Q^T(\overrightarrow{g}-Q\overrightarrow{a}^{(n)}) \quad , \tag{4.27}$$

so that simplicity in isolating $\overrightarrow{b}^{(n+1)}$ indicates we should restrict ourselves to diagonal weighting matrices. Using the definition of $\nu(E)$ (Section 3) enables us to see that if

$$\overrightarrow{b}^{(n+1)} = \nu^2(E)\ Q^T(\overrightarrow{g}-Q\overrightarrow{a}^{(n)}) \tag{4.28}$$

with $\nu^2(E)$ appropriately transformed into a diagonal matrix, components of s(E) are emphasised such that the value of each component has a maximum ratio to the standard deviation associated with that component.

In either case the value of χ^2 is already a minimum with respect to the component of $\overrightarrow{b}^{(n)}$, etc. If the vector $\overrightarrow{b}^{(n+1)}$ is not conjugate to all such vectors then the existing minimum condition may be disturbed. Chi-squared could be minimised with respect to components of all these vectors at each iteration, but it is simpler, before seeking the minimum, to remove the components of these vectors from $\overrightarrow{b}^{(n+1)}$ by the Gram-Schmidt process, to obtain the same result. Appendix A contains a more complete discussion of the technique described here, including a proof that $\overrightarrow{b}^{(n+1)}$ as initially generated (in exact arithmetic) contains a component of $\overrightarrow{b}^{(n)}$ only, and is already conjugate to $\overrightarrow{b}^{(n-1)}$, etc.

Given $\overrightarrow{a}^{(n)}$ and $\overrightarrow{b}^{(n+1)}$ we obtain $\mu^{(n+1)}$ using equation 4.18 and hence

$$\overrightarrow{a}^{(n+1)} = \overrightarrow{a}^{(n)} + \mu^{(n+1)}\overrightarrow{b}^{(n+1)} \quad . \tag{4.29}$$

We write

$$\overrightarrow{x}^{(n)} = Q\ \overrightarrow{a}^{(n)} \tag{4.30}$$

$$\overrightarrow{e}^{(n)} = \overrightarrow{g} - \overrightarrow{x}^{(n)} \tag{4.31}$$

and initially

$$\overrightarrow{b}_1{}^{(n+1)} = Q^T\ \overrightarrow{e}^{(n)} \quad . \tag{4.32}$$

Then

$$\vec{b}_1^{(n+1)} = \vec{b}^{(n+1)} + \alpha^{(n)}\,\vec{b}^{(n)} \tag{4.33}$$

and we find $\alpha^{(n)}$ and remove the component $\alpha^{(n)}\,\vec{b}^{(n)}$ by the step

$$\vec{b}^{(n+1)} = b_1^{(n+1)} - \vec{b}^{(n)}\,(\overleftarrow{b}^{(n)} Q^T Q \vec{b}_1^{(n+1)})/(\overleftarrow{b}^{(n)} Q^T Q\,\vec{b}^{(n)}) \ . \tag{4.34}$$

The sequence of equations from 4.29 to 4.34 which represent a conjugate gradient unfolding[84-88] of the set of equations described by equation 4.2 is obviously well adapted for computer programming. Computer programs based on these sequences have been used in a wide variety of problems outlined in Section 7 and have been found to converge more rapidly if the sequence used incorporates 4.28 rather than the simpler form 4.26.

It is worth noting at this stage that the technique of conjugate gradients involves working on the particular data set. An inverse of the matrix $Q\nu^2Q^T$ can be produced in principle as the method is carried out, but would not be specially useful with a data set generated from a different spectrum. Other authors[89,90] have suggested techniques to find an inverse for the significant portion of the matrix QQ^T in preference to concentrating on the data vector.

The notation is completed by writing

$$Q\mu^{(n)}\vec{b}^{(n)} = c^{(n)}\,\hat{v}^{(n)}, \tag{4.35}$$

where $\hat{v}$ is a unit vector aligned with $\vec{v}$, and

$$Q\vec{b}^{(n)} = \vec{v}^{(n)} \ . \tag{4.36}$$

5. SOME APPROXIMATE RESULTS FOR SPECIAL CASES

5.1 Series Inverse of the Resolution Matrix

From equation 4.21 formally[14]

$$\vec{a} = Q^{-1}\,\vec{g} \tag{5.1}$$

and equally formally

$$Q^{-1} = \mu^{-1}I + \mu^{-1}(I - \mu^{-1}Q) + \mu^{-1}(I - \mu^{-1}Q)^2 \ . \tag{5.2}$$

To make equations 5.1 and 5.2 meaningful in ordinary matrix notation the matrix Q must be square and, for any eigenvalues of the matrix Q

$$|1 - \lambda/\mu| < 1 \ , \tag{5.3}$$

i.e. for real values of λ and μ

$$2 > \lambda/\mu > 0 \ . \tag{5.4}$$

Applications of equation 5.2 are associated with a very large class of physical instruments that give rise to a matrix Q with all elements non-negative. In consequence, as a start towards obtaining the condition 5.4, the eigenvalue of largest magnitude is positive and can be associated with an eigenvector with all its elements non-negative. The second part can be achieved by rewriting equation 5.1

$$\vec{a} = Q^{-2}\, Q\vec{g} \quad . \tag{5.1a}$$

An iterative process is now set up to generate the terms of the series in equation 5.2 and apply them to $\vec{g}$.

Part way through the unfolding, after n iterations, the vector $\vec{a}^{(n)}$ has been obtained and from it the vector $\vec{x}^{(n)}$, given by

$$\vec{x}^{(n)} = Q\vec{a}^{(n)} \quad . \tag{5.5}$$

Hence the remaining discrepancy vector $\vec{e}^{(n)}$ is given by

$$\vec{e}^{(n)} = \vec{g}^{(n)} - \vec{x}^{(n)} \quad . \tag{5.6}$$

We write

$$\vec{a}^{(n+1)} = \vec{a}^{(n)} + \mu^{-1}\, \vec{e}^{(n)} \tag{5.7}$$

so that

$$\vec{x}^{(n+1)} = \vec{x}^{(n)} + \mu^{-1}\, Q\, \vec{e}^{(n)} \tag{5.8}$$

and

$$\vec{e}^{(n+1)} = \vec{e}^{(n)} - \mu^{-1}\, Q\, \vec{e}^{(n)} \tag{5.9}$$

$$= (I - \mu^{-1}\, Q)\vec{e}^{(n)} \quad . \tag{5.10}$$

Consideration of equation 5.10 shows that μ does not have to be fixed through the iterative process, only that the product of the factors

$$\prod_{n} (1 - \mu^{-1(n)}\, Q) \tag{5.11}$$

should converge to zero. Of course the result could apparently be achieved by putting $\mu^{(n)}$ equal to each (non-zero) eigenvalue in turn, but in most resolution situations the eigenvalues are unknown. It is not usually necessary to go beyond very few iterations with reasonably large values of $\mu^{(n)}$. Round-off errors would also cause the eigenvectors associated with the larger eigenvalues to reappear, and any such error present when using a small value of $\mu^{(n)}$ is liable to be multiplied intolerably.

If some of the eigenvalues are negative, μ can be alternated between positive and negative. After a pair of iterations

$$\vec{e}^{(n+2)} = (I - \mu^{-1} Q)(I + \mu^{-1} Q)\, \vec{e}^{(n)} \tag{5.12}$$

$$= (I - \mu^{-2} Q^2)\, \vec{e}^{(n)} \tag{5.13}$$

and the process is equivalent to that suggested by writing equation 5.1a. The expression 5.13 converges provided $|\mu|$ is larger than the magnitude of the largest eigenvalue of Q.

It is unnecessary to force Q to be a square; it follows naturally from equation 4.24 to take

$$\vec{a}^{(o)} = \mu^{-2} Q^T \vec{g} \tag{5.14}$$

leading to

$$\vec{a}^{(n+1)} = a^{(n)} + \mu^{-2} Q^T \vec{e}^{(n)} \tag{5.15}$$

and

$$\vec{x}^{(n+1)} = Q\, \vec{a}^{(n+1)} \tag{5.16}$$

so that

$$\vec{e}^{(n+1)} = (I - \mu^{-2} QQ^T)\, \vec{e}^{(n)} \quad . \tag{5.17}$$

For any vector $\vec{v}$

$$\overleftarrow{v}\, Q\, Q^T \vec{v} = \overleftarrow{u}\, \vec{u} \geqslant 0 \tag{5.18}$$

where $$Q^T \vec{v} = \vec{u} \tag{5.19}$$

so that all eigenvalues of QQ^T are non-negative. For any zero eigenvalue the component of the corresponding eigenvector remains unchanged in $\vec{e}^{(n)}$. We should expect such a component to remain in the discrepancy vector since it could not be generated in the data by any spectrum and must therefore be part of the random noise. If we choose μ^2 close to the largest eigenvalue (λ_m^2) of QQ^T, the eigenvectors are removed preferentially in order of the magnitude of the eigenvalues. After n iterations the k^{th} component in $e^{(n)}$ has been reduced to

$$(1 - (\frac{\lambda_k}{\lambda_m})^2)^n \quad \text{of its initial value.}$$

The method of inversion is useful in sharpening features found on a noisy background. It also lends itself to programming with an analogue computer as noted in Section 7. A similar technique with a matrix in place of μ^2 in equation 5.17 has also been used[31,36] with some success.

5.2 Fourier Unfolding of the Resolution Function

A special class of resolution functions can be written[14,16,28,31,40-42,48,57,68] in the form

$$R(V_j, E) = r_v(V_j)\ r_o(V-E\frac{V_o}{E_o})\ r_e(E) \qquad (5.20)$$

where $r_v(V_j)$ is independent of E,

$r_e(E)$ is independent of V

and $r_o(V-EV_o/E_o)$ depends only on the difference between V and the value EV_o/E_o of 'voltage' that nominally corresponds to E. Both the spectrum variable E and the measurement variable V are usually the same physical type, e.g. readings taken as a function of time often depend on the value of some quantity over a range of earlier times. It has been noticed by many that the two peripheral functions r_v and r_e, may be immediately detached, replacing the data by $m/r_v(V_j)$ and leaving as the final step, division of $r_e(E)\ s(E)$ by $r_e(E)$.

This leaves an equation that can be relabelled to read

$$m(E) = \epsilon(E) + \int r(E-E')\ s(E')\ dE' \ . \qquad (5.21)$$

It is helpful, but not essential, that r(E-E') is usually negligible for large values of $|E-E'|$.

The Fourier transform of a Faltung integral is the product of the Fourier transforms of its components. At this stage we concentrate on predicted yields.

We write

$$Y(k) = \frac{1}{\sqrt{2\pi}} \int y(E)\ \exp\ (-ikE)\ dE \qquad (5.22)$$

$$= \frac{1}{\sqrt{2\pi}} \int r(E-E')\exp\ -ik(E-E')\exp\ (-ikE')s(E')dEdE' \qquad (5.23)$$

$$=\sqrt{2\pi}\ \ R(k)\ S(k) \qquad (5.24)$$

where the script capital letters refer to the Fourier transforms of the indicated functions. Formally equation 5.24 can be divided by $\sqrt{2\ \pi}$ · $R(k)$ and inverse Fourier transforms made. The procedure fails because any well behaved function r(E-E') has a Fourier transform $R(k)$ that tends to zero for large values of (k). This behaviour is then specifically denied in $1/R(k)$. It is possible, however to get useful information by studying the behaviour of $1/R(k)$ for smaller magnitudes of k. We should be able to disregard values of k leading to more than one complete

cycle between adjacent data points. Appendix C contains a discussion of methods[91,92] of expressing data in terms of Fourier components.

We formally define

$$t(E-E') = \frac{1}{\sqrt{2\pi}} \int dk \exp ik(E-E')/R(E) \qquad (5.25)$$

and consider the effect of zeros of $R(k)$, i.e. poles of $1/R(k)$. Poles are ordinarily in pairs;

$$R(-k^*) = R(k)^* \qquad (5.26)$$

so that if

$$R(k_n + iq_n) = 0 \qquad (5.27)$$

it follows that

$$R(-k_n + iq_n) = 0 \qquad (5.28)$$

and $\frac{1}{R(k)}$ contains terms of the form

$$\frac{a_n}{k-k_n-iq_n} - \frac{a_n^*}{k+k_n-iq_n} \quad . \qquad (5.29)$$

For $q_n = 0$ we see that there is a pair of poles at $\pm k_n$ i.e. the instrument cannot detect a spectrum of the form $\alpha \cos(k_n E) + \beta \sin(k_n E)$. The measured data therefore provide no evidence for or against such components.

For a pair of poles with positive q_n the expression 5.29 is the Fourier transform of a function that is zero for E negative, and for E positive is

$$i\sqrt{2\pi}\{a_n \exp(ik_nE-q_nE) - a_n^*\exp(-ik_nE-q_nE)\} \qquad (5.31)$$

$$= 2\sqrt{2\pi}\ \rho_n \exp(-q_nE) \sin(k_nE + \delta_n) \qquad (5.32)$$

where $a_n = \rho_n \exp(i\,\delta_n)$ (5.33)

and E has been substituted for (E-E') to emphasise the important dependences.

For q_n negative, the associated poles give the term

$$2\sqrt{2\pi} \exp(-q_nE) \sin(k_nE + \delta_n)$$

again but only where (E) (i.e. E-E') is negative so that $-q_nE = -(-|q_n|)(-|E|)$ is negative.

We turn aside from the formal Fourier transform to consider the difference equation generated by the matrix formulation of equation 5.21 with data points spread at equal intervals ΔE.

We write

$$r_j = \frac{1}{\Delta E} \int_{E'=j\Delta E-\frac{1}{2}\Delta E}^{E'=j\Delta E+\frac{1}{2}\Delta E} r(E')dE' \tag{5.34}$$

and

$$s_n = \int_{E_o+n\Delta E-\frac{1}{2}\Delta E}^{E_o+n\Delta E+\frac{1}{2}\Delta E} s(E')\ dE' \quad . \tag{5.35}$$

The matrix equivalent of equation 5.21 is

$$Y_n = m_n - \varepsilon_n = \Sigma\ s_k\ r_{n-k} \quad . \tag{5.36}$$

We now consider the reciprocal operator T_j defined by

$$\Sigma\ T_k\ r_{n-k} = \delta_{n\ o} \tag{5.37}$$

where $\delta_{n\ o}$ is the Kronecker delta, i.e. unity if the integer n is zero and zero for all other values of n. Equation 5.37 whose form is clearly recognisable as a linear difference equation, can be rewritten as a linear combination of various order differences of terms of the sequence T_n. It is thus closely related to a linear differential equation, and can be handled in an analogous way. We define $r(\lambda)$ by the equation

$$r(\lambda) = \sum_n \lambda^n\ r_{-n} \quad . \tag{5.38}$$

The general solution for T_n is a sum of the form

$$T_n = \Sigma\ a_q\ \lambda_q^n \tag{5.39}$$

where

$$r(\lambda_q) = 0 \quad . \tag{5.40}$$

A particular solution with different values of a_q in the region of positive and negative n is selected so that T_n tends to zero for $|n|$ large and so that equation 5.37 is satisfied at all values of n including the special case, n = 0.

Note that now we can write

$$\lambda = \exp(-ik\Delta E) \tag{5.41}$$

and see that $r(\lambda) = r(\exp[-ik\Delta E])$ is an approximation to $\sqrt{2\pi}\ R(k)$ as defined by equation 5.24. The values of λ_q satisfying equation 5.40

should be closely related to the zeros of $R(k)$ using equation 5.41. Using the multiplication property of Fourier transforms we see also that as r_{-n} is the coefficient of λ^n in $r(\lambda)$, so T_n is the coefficient of λ^n in a series expansion of $1/r(\lambda)$. To find an inverse, in the one-dimensional case, for the polynomial $r(\lambda)$ we locate its zeros and hence write $\frac{1}{r(\lambda)}$ in partial fractions;

$$\frac{1}{r(\lambda)} = i\lambda^T \sum_q \left(\frac{a_q}{\lambda-\lambda_q} - \frac{a_q^*}{\lambda-\lambda_q^*}\right) \tag{5.42}$$

where the factor λ^T is determined by the coefficient in $r(\lambda)$ representing r_o. The partial fractions in equation 5.42 are expanded in positive or negative powers of λ_q/λ depending on whether $|\lambda_q|$ is less or greater than unity. The expressions 5.31 and 5.42 are both real. Commonsense so dictates, and it is also commonsense to check. Also if $r(\lambda)$ has n roots there are n coefficients T_k given by equation 5.39 for which the expression in positive or negative powers of λ gives the same result. If there are roots λ_q such that $|\lambda_q| = 1$, the expansion of equation 5.42 leads to a non-converging series. As before, there is a component here that would be suppressed in error-free data if it were present in the spectrum. Any component present in the data is therefore ignored or ascribed to noise.

There may be other values of q with $|\lambda_q|$ close to unity which usually have a low signal-to-noise ratio. The form of 5.42 should always be checked to see whether some components should be discarded.

5.3 Triangular Resolution Matrices

There are numerous examples[9,63] where a square matrix Q, which is produced from a resolution equation, obeys one or other of the restrictions

$$Q_{jk} = 0 \quad j < k \text{ and any } k \tag{5.43}$$

or

$$Q'_{jk} = 0 \quad j > k \text{ and any } k \quad . \tag{5.44}$$

The prime on Q'_{jk} of equation 5.44 distinguishes an upper right triangular matrix, which could of course be obtained by reversing the order of both variables in equation 5.43.

Such matrices can be a consequence either of conservation of energy where the resolution function depends on some scattering process, or causality where the resolution involves a time delay. Which form is

obtained depends, in the energy case, on whether the experiment entails, for example, initiating events at a particular energy followed by observation of the total number of reactions, giving rise to equation 5.44, or observing the voltage output of a detector bombarded by a spectrum of particles, giving rise to equation 5.43. (We assume for definition purposes that in both equations, k increases with energy.)

In either case the inverse of the matrix Q has the same triangular form. Thus from equation 5.43

$$(Q^{-1})_{jk} = 0 \quad j < k \text{ and any } k \tag{5.45}$$

and from equation 5.44

$$(Q'^{-1})_{jk} = 0 \quad j > k \text{ and any } k \quad . \tag{5.46}$$

Both cases are equivalent to a matrix that is derived by eliminating variables from simultaneous equations. In either case the simple structure of the determinant involved can be used to investigate the structure of the inverse close to the main diagonal.

We see that

$$(Q^{-1})_{jj} = (Q_{jj})^{-1} \tag{5.47}$$

and

$$(Q^{-1})_{jj-1} = - Q_{jj-1}/(Q_{jj}\, Q_{j-1j-1}) \tag{5.48}$$

and so on.

Concern about errors follows since the forms 5.43 and 5.44 commonly appear with small positive elements on the diagonal and larger positive ones adjacent. From equations 5.47 and 5.48 we then see that there is a threat that an ordinary statistical error in the data can cause a consequential 'ringing' in the spectrum which takes the form of a succession of terms of alternating sign and increasing magnitude as we leave the diagonal of the inverse matrix. Before using the algebraic simplicity of unfolding the triangular matrix form it is therefore wise to check that the elements far away from the diagonal of the inverse do become negligible. This desirable property can sometimes be achieved[9] by recasting the form of the resolution equation. This illustrates, too, the advantage of accepting an unfolding process that does not use a complete inverse. Several problems are discussed in Section 7, where conjugate gradient solutions are indicated rather than inversion of a triangular matrix as has been employed in the past. An extra incentive to abandon the use of a triangular inverse when sufficient computing

facilities are available is often provided by a more detailed examination of the resolution function. For a multitude of experimental reasons the true resolution functions often lead to elements beyond the diagonal in problems superficially described by equation 5.43 or 5.44. Simplicity often allows quick calculation of a good approximation to the spectrum by truncation to a triangular matrix. A better approximation usually demands more detail.

5.4 Perturbations from a Known Inverse

If a useful inverse for some matrix Q has been obtained and we wish to alter some condition slightly, we write

$$\vec{a}_o = Q^{-1} \vec{y}_o \tag{5.49}$$

and wish to obtain

$$\vec{a}_1 = (Q+q)^{-1} \vec{y}_o \quad . \tag{5.50}$$

Noting that

$$Q\, Q^{-1} = I$$

we write

$$(Q^{-1} - r)\ (Q + q) = I \tag{5.51}$$

$$Q^{-1}\, Q - r(Q + q) + Q^{-1}q = Q^{-1}\, Q \quad . \tag{5.52}$$

Multiplying on the right by Q^{-1} and simplifying gives

$$r = Q^{-1}\, q\, Q^{-1} - r\, q\, Q^{-1} \quad . \tag{5.53}$$

Substituting the value of r given by equation 5.53 where r appears in the right-hand side gives

$$r = Q^{-1}qQ^{-1} - Q^{-1}qQ^{-1}qQ^{-1} + rqQ^{-1}qQ^{-1}, \text{ etc.} \tag{5.54}$$

The technique outlined is familiar in obtaining Green's functions where convergence is rapid but it should be used cautiously where the rate of convergence is unknown.

6. ERRORS OF DATA AND SPECTRUM

For a given data point measured values m_j are assumed to be distributed in the neighbourhood of the mean value y_j and $P_j(m_j)dm_j$ is the probability that a measured value of m_j lies between m_j and $m_j + dm_j$. It is usually convenient to use a Gaussian distribution to describe such probabilities. Only random errors are assumed, i.e. all systematic effects are assumed to be included in the resolution functions. Then

$$P(m_j) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp -(m_j - y_j)^2/2\sigma_j^2 \quad . \tag{6.1}$$

The mean value of m_j can be verified to be

$$\int_{-\infty}^{\infty} P(m_j)\ m_j\ dm_j = y_j \tag{6.2}$$

and the mean square deviation from y_j

$$\int_{-\infty}^{\infty} P(m_j)\ (m_j - y_j)^2\ dm_j = \sigma_j^2 \quad . \tag{6.3}$$

The measurements at different data points are assumed to be independent, i.e.

$$P(m_j,m_\ell)\ dm_j\ dm_\ell = P(m_j)dm_j\ P(m_\ell)dm_\ell \quad . \tag{6.4}$$

In nuclear physics and increasingly in other branches of physics, readings are commonly counts of individual events rather than scale settings of meters. For small numbers of counts the distribution of the errors is considerably non-Gaussian. For all m_j and y_j

$$P(m_j) = y_j^{m_j} \exp(-y_j)/(m_j!) \tag{6.5}$$

where the mean y_j need not be an integer, but m_j must be. The mean square deviation is y_j. For large values of y_j the variation of $P(m_j)$ from equation 6.5 is slow from integer to integer near the mean y_j. Using Stirling's approximation we replace equation 6.5 by a form similar to equation 6.1

$$P(m_j) = \frac{1}{\sqrt{2\pi\ y_j}}\ \exp(-(m_j - y_j)^2/2y_j) \quad . \tag{6.6}$$

To combine probabilities over a number of integers both sides of equation 6.6 may be multiplied by the number Δm_j of integers involved. A Gaussian distribution is assumed for the errors unless specifically stated otherwise, but we find it convenient to use

$$m_j \approx y_j \approx \sigma_j^2 \tag{6.7}$$

as a means of evaluating σ_j for large counting rates; the value from equation 6.7 provides a lower limit for σ_j.

The probability of obtaining a particular data vector with components m_1, m_N in the ranges dm_1, dm_N, from a given measurement is then

$$P(m_1 \ldots m_N)\ dm_1 \ldots dm_N = \prod^{N} \frac{1}{\sigma_j\sqrt{2\pi}} \exp\{-(m_j - y_j)^2/2\sigma_j\}\ dm_j \quad . \tag{6.8}$$

The notation of 'data significance space' has already been demonstrated

to have an advantage over that of data space; if two vectors in spectrum space are conjugate, the corresponding vectors in data significance space are orthogonal. Now any pair, or set, of orthogonal vectors in data significance space can be shown to vary independently as written in equation 6.4 and the variance of the component associated with any unit vector in data significance space is the same, i.e. unity. From Section 4

$$g_j = \sigma_j^{-1} m_j = w_j \, m_j \tag{6.9}$$

and

$$\vec{x} = w \, \vec{y} \quad . \tag{6.10}$$

A set of unit data significance vectors $\hat{d}_j$ is obtained from the corresponding set in data space. The data vector corresponding to $\hat{d}_j$ has magnitude σ_j for the g^{th} data point and is zero for all other data points, so that in data significance space

$$(\hat{d}_j)_k = \delta_{jk} \quad . \tag{6.11}$$

Thus

$$\vec{g} = \sum_j g_j \hat{d}_j \quad . \tag{6.12}$$

Also

$$\vec{dg} = w \, \vec{dm} \tag{6.13}$$

so that

$$P(m_1 - m_N) dm_1 \ldots dm_N = P(g_1 \ldots g_N) dg_1 \ldots dg_N \prod_{j=1}^{N} \sigma_j$$

$$= \prod_1^N \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(g_j - x_j)^2\} \, dg_i \quad . \tag{6.14}$$

We can write the data significance vector $\vec{g}$ in terms of any other set of orthonormal vectors $\{\hat{u}_\ell\}$ that span the space. The conjugate gradient technique proceeds by generating members $\hat{v}^{(k)}$ of such a set. If, as indicated by use of the term $\vec{g}_L$ in equation A.6 the set is incomplete it can be completed with any consistent set of unit vectors that span the remainder of the space, giving

$$\vec{x} = \sum_k c_k \hat{u}_k \quad , \tag{6.15}$$

$$\vec{g} = \sum_k b_k \hat{u}_k \quad , \tag{6.16}$$

and

$$\hat{d}_j = \sum_k h_{jk} \hat{u}_k \quad . \tag{6.17}$$

The matrix h_{jk} of coefficients is orthogonal, i.e.

$$h^T = h^{-1} \tag{6.18}$$

and

$$\partial g_j / \partial b_k = h_{jk} \tag{6.19}$$

so that we may rewrite the volume element in equation 6.14

$$\prod_{j=1}^{N} dg_j = \prod_k db_k \left| \frac{\partial \vec{g}}{\partial \vec{b}} \right| \tag{6.20}$$

and the Jacobian determinant $\left| \frac{\partial \vec{g}}{\partial \vec{b}} \right|$ is unity for a proper orthogonal transformation.

Then from equation 6.14

$$\begin{aligned} &\prod_1^N \frac{1}{\sqrt{2\pi}} dg_j \exp - \frac{1}{2} (g_j - x_j)^2 \\ &= \left(\prod_{j=1}^N \frac{1}{\sqrt{2\pi}} dg_j \right) \exp\{ - \sum_j \frac{1}{2} (g_j g_i - 2x_j g_j + x,x_j) \} \\ &= \left(\prod_{k=1}^N \frac{1}{\sqrt{2\pi}} db_k \right) \exp - \frac{1}{2} (\overleftarrow{g}.\overrightarrow{g} - 2\overleftarrow{g}.\overrightarrow{x} + \overleftarrow{x}.\overrightarrow{x}) \\ &= \prod_{k=1}^N \frac{1}{\sqrt{2\pi}} db_k \exp - \frac{1}{2} (b_k - c_k)^2 \quad , \end{aligned} \tag{6.21}$$

since the scalar product of vectors is unchanged by rotation of the coordinate system.

For any particular b_k an individual distribution can be obtained by integration of equation 6.21 over the remainder

$$P(b_k) = \frac{1}{\sqrt{2\pi}} \exp - \frac{1}{2} (b_k - c_k)^2 \quad . \tag{6.22}$$

For any vector in data space there is an associated unit vector in data significance space. The data vector then defines a component magnitude associated with the unit vector. With Gaussian distributions of the independent measurements making up the data set we see that a mean c_k can be defined for such a component that each and every component so defined has the same variance, and that the covariances of orthogonal

components are zero.

This is a general argument but here we are concerned with the series of conjugate vectors in spectrum space produced by the iterative sequence in Section 4 and its associated set of data significance vectors.

If the spectrum were given, the yield could be predicted and compared with a given set of data, enabling a value of χ^2 to be calculated. Its expected value is approximately N, the number of data points. In fitting by conjugate gradient techniques we remove a series of components from the data vector completely, including any associated error of each component. When n components have been removed after n iterations the residual noise component of the data contributes a χ^2 component of approximately N-n. If at this stage

$$\chi^2(x^{(n)}) \approx N-n \quad , \tag{6.23}$$

we can safely attribute the remaining discrepancy between data and predicted yields to the random errors of measurement. There is almost inevitably some admixture of noise in the calculated spectrum.

We could consider the original data to be rather like the grinning Cheshire cat in Alice in Wonderland. A perfect unfolding process removes the cat which is real, leaving behind the errors as a residual grin, not part of the cat at all. As in the original story, we suspect that in general the process of removing the signal from the data involves criteria which do not perfectly define substance so that when we have finished the process the shadow of the cat is preserved in the form of components which are completely absent from the grin. Analysis using conjugate gradient techniques introduces negative correlations between residuals at successive data points. The form of the residual discrepancies is thus dependent both on the real information in the original data and on the process of unfolding those data.

There are three questions about errors that govern the usefulness of a given experiment. The first concerns a particular point of the vector form of the spectrum s(E) and asks how much s(E) can be changed there without doing violence to the data fit.

A similar question was asked and answered by the definition of $\nu(E)$ in Section 3: there we were concerned with the limiting case of variation of s(E) at a point; here we are concerned with variation over a range ΔE. A value of $s_k(E_k)$ defines s(E) over the range ΔE about E_k. We

define ν_k by

$$\nu_k \Delta E = \left(\sum_j R(V_j,E_k)/\sigma_j\right)^{-\frac{1}{2}} \quad . \tag{6.24}$$

Where errors are shown in the spectra in the figures of this report, they are proportional to ν_k. We assume that the residual errors in the data do not interfere with the effect of changing s(E).

A smaller change will do if applied over a wider range ΔE about E. This leads easily to consideration of the effect of changes that are not the same over the whole of ΔE. The second question thus concerns the presence or absence of a particular feature in the spectrum. A very common question involves a Lorentzian interference term in an emission spectrum

$$s(E) = S_o(E) + \frac{a+b(E-E_o)}{(E-E_o)^2 + (\frac{1}{2}\Gamma)^2} \tag{6.25}$$

where $S_o(E)$ is the simplest fit to the data, and often we wish to know if the data support the hypothesis of non-zero values for a and b. Since each makes a linear contribution to the data, a bi-linear form can be obtained for χ^2 in terms of a and b and the previous χ^2. Choosing appropriate values of a and b should lead to a decrease of at least two in χ^2 which is not a significant improvement. Just how large a change is required to be significant depends on how good the previous fit was. If the major component of error was already in the statistical range for χ^2 then introduction of a constant that diminishes χ^2 by three is probably significant. There may be no effect at all on the data fit, which means that this data set or any other data set gathered at the same points reveals nothing about the value of a or of b.

Perhaps the most important question concerns the reproducibility of results. Suppose we have N data points and obtain an acceptable value of χ^2 in n iterations. The same analysis is now applied to obtain s(E) from an entirely new set of data gathered under the same experimental conditions. How much difference is there between the two values of s(E) recovered from the two different data sets?

Appendix A contains the result of following a conjugate gradient process to its end as

$$\vec{g} = \sum_k c^{(k)} \hat{v}^{(k)} + \vec{g}_L \quad , \tag{A.6}$$

stopping only where

$$Q\nu^2 Q^T \vec{g}_L = 0 \quad . \tag{A.9}$$

It appears plausible to assume that usually the quantity

$$\overleftarrow{v}^{(k)} \, Q v^2 \, Q^T \overrightarrow{v}^{(k)} / \overleftarrow{v}^{(k)} . \overrightarrow{v}^{(k)} \tag{6.26}$$

mentioned at A.44 shrinks rapidly as k is increased. Because of the errors that might occur in remeasurement of the data, it is wise to check that the shrinkage does happen for a particular case. A different set of data denoted by the vector $\vec{g}'$, can be written

$$\vec{g}' = \Sigma c^{(k)\prime} \hat{v}^{(k)} + \vec{g}'_L \quad . \tag{6.27}$$

The mean of $c^{(k)}$ is denoted by $\bar{c}^{(k)}$ in a number of trials and the probability of $c^{(k)\prime}$ in the range $c^{(k)}$ to $c^{(k)} + dc^{(k)}$ is

$$P(c^{(k)}) \, dc^{(k)} = \frac{1}{\sqrt{2\pi}} \exp - \frac{1}{2} (c^{(k)} - \bar{c}^{(k)})^2 \quad , \tag{6.28}$$

since from equation 6.22 any vector in 'data significance space' has the same variance.

Similarly

$$P(c^{(k)\prime}) \, dc^{(k)\prime} = \frac{1}{\sqrt{2\pi}} \exp - \frac{1}{2} (c^{(k)\prime} - \bar{c}^{(k)})^2 \quad .$$

Hence

$$P(c^{(k)\prime} - c^{(k)}) \, d(c^{(k)\prime} - dc^{(k)}) =$$

$$\int \frac{d(c^{(k)\prime} + c^{(k)})}{2} P(c^{(k)}) \, P(c^{(k)\prime}) \, d(c^{(k)\prime} - c^{(k)})$$

$$= \frac{1}{\sqrt{2\pi}} d(c^{(k)\prime} - c^{(k)}) \exp - \frac{1}{4} (c^{(k)\prime} - c^{(k)})^2 \quad . \tag{6.29}$$

The first contribution to the variation in the spectrum is $c^{(k)} \pm \sqrt{2}$, where $\sqrt{2}$ is the root mean square variation in the value expected for $c^{(k)}$ at a subsequent measurement. We already have the vector $\vec{b}^{(k)}$ and the k^{th} contribution to error in s is

$$\pm \vec{b}^{(k)} \sqrt{2} [c^{(k)}]^{-1} \quad . \tag{6.30}$$

Allowance should also be made for the different sequence that will appear in the processing of $\vec{g}'$ in equation 6.27. In conjugate gradient iterations on $\vec{g}$ in equation A.6, each component $\vec{v}^{(k)}$ is masked by its predecessors until its turn comes for processing. Although this masking is assumed to be incomplete in the iterations on $\vec{g}'$, it is nevertheless approximated because of the rapid shrinkage in the matrix elements involved. It is likely that at iteration k on $\vec{g}'$ we substantially complete the removal of $c^{(k-1)\prime} \hat{v}^{(k-1)}$, accomplish most of the removal of

$c^{(k)}$' $\hat{v}^{(k)}$ and make a start on the removal of $c^{(k+1)}$' $\hat{v}^{(k+1)}$. It is thus plausible to use $\vec{b}^{(n+1)}$ as a measure of the error involved in stopping after n iterations and of the likely form of error to be found in remeasurement of the same data set, i.e.

$$\vec{a} = \sum_{k=1}^{n} \vec{b}^{(k)} \left(1 \pm \sqrt{2}[c^{(k)}]^{-1}\right) \pm \vec{b}^{(n+1)} \quad . \tag{6.31}$$

Throughout this section we have assumed that iteration will end as soon as an acceptable fit to the measured data is obtained. If not, eventually a spectrum is obtained which fits the data exactly (with the possible exception of a term like $\vec{g}_L$ in equation A.41). Then to analyse errors in the spectrum the complete variance-covariance matrix $(Q^TQ)^{-1}$ must be used, or as much of that inverse as exists. The complete inverse usually contains much larger quantities than those in the portion required by the truncated set of iterations. Small values of the ratio A.44 give large contributions to the inverse matrix. This analysis aims to avoid major uncertainties in the calculated spectrum which are associated with small improvements in the fit to the measured data and which usually produce physical nonsense long before an exact fit is achieved. We emphasise that stopping iterations as soon as an acceptable fit is obtained, increases the probability that the spectra unfolded from different sets of data gathered under the same conditions are acceptably alike.

7. SOME EXAMPLES INVOLVING UNFOLDING

To obtain the energy spectrum of a neutron flux is a continuing problem of nuclear physics. The electric or magnetic fields used for charged particles do not deflect fast neutrons. Here several detector techniques are considered, each associated with a resolution problem.

The first method detects protons knocked on by collisions in hydrogen gas[9-11,13,29,35,43,45,47,55,58,80-82]. Low energy scattering of neutrons by protons is almost completely isotropic in centre of mass coordinates and the near equality of neutron and proton masses then gives a very simple energy distribution. For neutrons of energy E_n the probability that the initial proton energy is in a range E_p to $E_p + dE_p$ is

$$P(E_p)dE_p = dE_p/E \qquad \text{if} \quad E_p < E_n$$
$$\text{and} \qquad = 0 \quad , \qquad \text{if} \quad E_p > E_n \quad . \tag{7.1}$$

If the protons are all being detected and their energies are being measured accurately then the flux $\phi(E_n)dE_n$ of neutrons in the range E_n to $E_n + dE_n$ is formally

$$E_n \sigma(E_n)\phi(E_n) = - \frac{dy(E_p)}{dE_p} \qquad (7.2)$$

where $\sigma(E_n)$ is the scattering cross section for neutrons of energy E_n and $y(E_p)$ is the yield of protons at energy E_p. The uncritical application of equation 7.2 to actual data usually leads to nonsense. In addition to the statistical difficulties in using measurements to approximate a mean yield there are two physical processes to consider, edge effects and position sensitivity in the counter.

Counters to detect the protons produced in the knock-on process always have a finite sensitive volume. The output signal is often proportional to the energy deposited by ionisation in that volume. Some proton tracks are bound to end outside the volume for any external flux of neutrons, and in many counters tracks can also start outside the sensitive volume. Geometric calculations[93] have been made and incorporated in computer programs[9] to allow for the edge effects in special cases. Also collection efficiency appears often to depend significantly on the position in the detector. Equal energy events can thus lead to different voltages output. Beyond the detector comes electronics and some further broadening of resolution.

A program has been written taking into account the edge effects in a spherical proportional counter and assuming that other broadening gives a Gaussian resolution function.

Sample calculations are shown for the measurement of neutron flux from a thick lithium target bombarded with 2.8 MeV protons; the measurements cover the low energy end of the proton recoil spectrum; a user-supplied correction for the high energy neutron flux is included. After the form of the high energy flux is input, the user can either accept a machine calculation of the normalisation or supply a particular value. Figures 1, 2 and 3 are graphs output at the point where the fit to data became acceptable i.e. after 10 iterations. Figures 4 and 5 were produced after 12 iterations with a normalisation supplied deliberately eight per cent low. Figures 6 and 7 illustrate the situation after six iterations of the successful calculation. Graphs in sets of three are produced when convergence has been achieved or at intermediate points specified by the user. The first graph of the trio (Figure 1) which

concerns the data to be fitted, shows the amount subtracted from the original data before fitting, i.e. the contribution of the high energy neutrons to the low energy proton spectrum. (In the approximation represented by equation 7.2 it would be a horizontal straight line.) Four curves are output which are concerned with the residual fitted data. The set of data itself and its reconstruction appear together with curves one standard deviation away from the reconstruction on either side.

The second graph of each trio (Figures 2,4,6) shows the current approximation to the spectrum. The curve of the spectrum is flanked by curves one standard deviation away from it. In all the cases shown the neutron energy groups are three proton channels 1.99 keV wide so that one standard deviation represents a change in the flux of neutrons per MeV over a width of 5.97 keV sufficient to produce an increase of unity in the value of χ^2.

The third graph of each trio (Figures 3,5,7) shows the actual discrepancy in each data channel in multiples of the standard deviation for that channel. The square of the quantity represents the contribution to χ^2 from the channel. The aim of the analysis is to produce pure noise in this graph. Any systematic trend displayed by the errors can be a warning of an unsatisfactory fit.

An important correction that could apply here concerns pile-up of pulses. The number of double counts taking place in a time scale which causes them to be interpreted as single events is proportional to the square of the count rate, i.e. is non-linear. Such corrections are not considered here, but obviously can have considerable bearing on the validity of results.

'Appropriate solutions' of the resolution problem in proton recoil experiments have been produced and reported effective[10]. Such solutions can be useful in proton recoil problems, since the structure in $\phi(E_n)$ is closely related to structure at the same energy in $y(E_p)$.

A computer program has been written which uses actual measured data from monoenergetic beams of neutrons to unfold more complex spectra by the conjugate gradient method, but it is still in the process of development. Using data for a set of monoenergetic neutron beams, this program will unfold neutron pulse height data measured with a plastic scintillation counter.

There is a problem in having the resolution matrix in the form of

experimental data because the value of matrix elements has a smooth dependence on the energy of the neutron or on the strength of the output signal. Measurements of the matrix elements contain statistical errors and may make interpolation difficult. Development of the program includes studies of the effects of interpolation.

The energy of neutrons may also be obtained by measuring their speed. Individual neutrons cannot usually be timed but those produced in pulses within a small volume can. A short, nearly monoenergetic pulse of protons is used to generate neutrons in a target. There are usually also some associated γ-rays which all travel at the speed of light c and arrive with a distribution over time that gives the shape of the original pulse. The neutrons which are slower and vary in speed, arrive later and over a longer time interval. The same detector can be triggered by either gammas or neutrons; standard electronics superimposes results from many pulses and records separately the number of events detected in equal time channels from considerably before the first gammas of each pulse until after the neutron spectrum disappears into the background. If the proton pulse is much shorter in duration than the interval spanned by one channel the gamma spectrum is likely to be restricted to one channel. The channels containing neutron spectra can be assigned a range of time-of-flight and hence of energy. Given the distance d from target to detector, a neutron arriving a time t later than the gamma pulse has a velocity

$$cd/(d + ct) \tag{7.3}$$

and a kinetic energy

$$\frac{1}{2}\, mc^2 d^2/(d + ct)^2 \quad . \tag{7.4}$$

The number of neutrons per unit energy is equal to the number per channel divided by the energy width of a channel. With channels of uniform time width Δt, the energy width is not uniform

$$|\Delta E| \;=\; |\Delta t|\, \frac{mc^3 d^2}{(d+ct)^3} \quad . \tag{7.5}$$

Expressions 7.4 and 7.5 are non-relativistic; for high energies, instead of expression 7.5 we would have

$$|\Delta E| \;=\; |\Delta t|\, m_o\, c^3 \frac{d^2}{\{ct(2d+ct)\}^{3/2}} \quad . \tag{7.6}$$

Usually the original proton pulse is considerably longer than the time width of an individual channel. The detected neutrons no longer have an

unambiguous time-of-flight. For a narrow (unit delta function) pulse the yield would be s(t). With a pulse P(t) that begins to be appreciable at zero time and continues for an interval τ,

$$y(t) = \int_o^\tau P(t')\, s(t-t')dt' \quad . \tag{7.7}$$

If time is divided into steps Δt, then

$$\int_{\ell t-\frac{1}{2}\Delta t}^{\ell t+\frac{1}{2}\Delta t} s(t)dt = s_\ell \tag{7.8}$$

with similar definitions for P_k and y_j. If Δt is made small enough we can obtain any desired approximation to equation 7.7 with the equation

$$y_\ell = \sum_{k=o}^{k_{max}} P_k\, s_{\ell-k} \tag{7.9}$$

where k_{max} is the largest integer less than $\tau/\Delta t + \frac{1}{2}$.

Alternatively equation 7.9 can be written

$$y_k = \sum_{\ell=k-k_{max}}^{\ell=k} P_{k-\ell}\, s_\ell \quad . \tag{7.10}$$

If Δt is chosen as the channel width with which the time-of-flight was measured, then

$$y_k \approx m_k \tag{7.11}$$

where m_k is the number of counts in the appropriate channel. We can also choose to arrange for convenience in calculation by setting

$$\sum_o^{k_{max}} m_k = M \tag{7.12}$$

and

$$P_k = m_k/M \quad . \tag{7.13}$$

In an ideal experiment the values of m_k are very large over some range and zero outside it, i.e. there are no background counts. If also there is a single channel in which gammas are recorded as well as some channels with zero counts between the channel and the neutron channels,

$$P_k = \delta_{k,o} \tag{7.14}$$

and

$$Y_k = s_k \quad . \tag{7.15}$$

More commonly there is a small, but constant background that can be

subtracted with P_k spreading into a number of channels each having a large number of counts. Assuming then that P_k is accurate, equation 7.10 may be unfolded. Note that if it is unfolded for values of $y_\ell \approx m_\ell$ and $\ell=L,\ldots L+L_1$, the values of s_ℓ are usually assumed to be zero except for $\ell=L,\ldots L+L_1 - k_{max} +1$, since otherwise there would be significant information in data channels outside the range chosen. The problem so far outlined can be simply solved with a conjugate gradient technique and results are shown in Figures 8 and 9. For convergence a weighting factor ν^2 must be used in the spectrum space as described in Section 3. The number of iterations to obtain a satisfactory fit to the data is usually less than ten. Other techniques might be useful with the problem, but some that have been tried have led to violently oscillating spectra.

There are modifications to the simplest methods of unfolding, starting from equation 7.10. If the experimental data have an almost constant background count level it is usually subtracted before unfolding. If no subtraction is made there can be unfortunate edge effects because the members of the set $\{y_\ell\}$ of yields outside the set being unfolded are being implicitly, and inconsistently, assumed to be zero and unfolding them produces oscillations of the calculated spectrum at the ends of the range. A procedure similar in principle to subtraction removes edge effects using the data set even in the region where it can be assumed to be solely background at both ends of the meaningful neutron data set. A set of quantities $\{f\}$ is generated,

$$f_\ell = \Sigma P_k f'_{\ell-k} \qquad (7.16)$$

where

$$f'_\ell = 0 \quad \text{if } s_\ell \text{ is assumed zero}$$
$$f'_\ell = 1 \quad \text{if not.} \qquad (7.17)$$

The factor f_ℓ is unity for values of ℓ in the range where the data are meaningful and tail off in the background region at the edges. We now use

$$m'_\ell = m_\ell f_\ell \qquad (7.18)$$

as the data to be fitted and a constant background is fitted accurately, i.e. s_ℓ has the background value right up to the edge.

The minimum value of the dispersion of the data values can be assigned on the basis of Poisson statistics of counting. If the mean

value of the yield at a particular data point is y_k for all data collection runs then the mean of a large number of observations also gives

$$\bar{m}_k = y_k \tag{7.19}$$

and the mean square deviation from y_k is given by

$$\overline{(m_k - y_k)^2} = y_k \quad . \tag{7.20}$$

If however the value of y_k varies because of a random fluctuation independent of the counter system, a mean value $\bar{y}_k$ and a mean square deviation from it must be defined

$$\overline{(y_k - \bar{y}_k)^2} = \overline{y_k^2} - \bar{y}_k^2 \quad . \tag{7.21}$$

We find

$$\bar{m}_k = \bar{y}_k \tag{7.22}$$

as before, but

$$\overline{(m_k - \bar{y}_k)^2} = \bar{y}_k + \overline{y_k^2} - \bar{y}_k^2 \tag{7.23}$$

where we have used the independence hypothesis

$$\overline{(m_k - y_k)(y_k - \bar{y}_k)} = 0 \quad . \tag{7.24}$$

We also note that fluctuations of y_k can be correlated from channel to channel. If a particular run with K channels and no such correlations is repeated

$$\sum_k (m_k - m_k')^2 / m_k \approx 2K \quad . \tag{7.25}$$

We also expect that

$$\sum_k (m_k - m_k')^2 (m_{k+1} - m_{k+1}')^2 / m_k \, m_{k+1} \approx 4K \tag{7.26}$$

where the approximate equality in equation 7.26 if it holds is evidence that there is no correlation of fluctuations in the values of y_k.

We note that the values of P_k in equation 7.13 are subject to error. In addition, equation 7.10 can be considered as an equation to be solved for the best values of P_k or for s_k. It is reasonable to test whether the χ^2 of the fit to all the data, including the P_k, can be markedly improved at any stage by means of a conjugate gradient improvement to the values of P_k. In general the results of such tests were inconclusive. What was done was to test the available improvement without incorporating it. Several more iterations of improvement of the

spectrum were undertaken, and the tests repeated. The improvement available in χ^2 by varying $\{P_k\}$ is generally about the same as that provided by the next iteration on the spectrum. The improvement provided decreases as the number of iterations increases, which casts doubt on the method.

The shape of the functions underlying the data within a single time channel now becomes relevant because of an uncertainty both in the pulse gamma spectrum and in the neutron spectrum arising from such structure. In general the contribution Δy_j to a reading at V_j from an energy range at E_k can be written

$$\Delta y_j = \int_{E_k-\frac{1}{2}\Delta E}^{E_k+\frac{1}{2}\Delta E} R(V_j,E)\ s(E)\ dE \quad . \tag{7.27}$$

Then $R(V_j,E)$ can be written in terms of orthonormal Legendre polynomials $P_\ell(\mu)$ in the range

$$E = E_k - \tfrac{1}{2}\Delta E, \ \ldots, \ E_k + \tfrac{1}{2}\Delta E$$

$$R_{jk} = \sum_\ell r_{\ell k}(V_j)\ P_\ell\ (\frac{E-E_k}{\frac{1}{2}\Delta E}) \tag{7.28}$$

and similarly

$$s(E) = \Sigma\ s_{\ell k}\ P_\ell\ (\frac{E-E_k}{\Delta E}) \tag{7.29}$$

so that

$$y_j = \Sigma\Delta\ y_j = \sum_{k\ell\ell'} r_{\ell k}(V_j)\ s_{\ell' k}\ \delta_{\ell\ell'}\ \Delta E \tag{7.30}$$

where the Kronecker delta $\delta_{\ell\ell'}$ arises from the integration in equation 7.27. When we evaluate y_j, using separate averages over the width ΔE instead of combined integration, we obtain

$$y_j = \sum_k r_{ok}(V_j)\ s_{ok}\ \Delta E \quad . \tag{7.31}$$

The distinction between the expressions 7.30 and 7.31 is unimportant for y_j unless it is likely that both the resolution function and the spectrum have considerable structure on a scale narrower than the interval we are using[75].

Given a set of points in our time-of-flight analysis we may wish to estimate the magnitude of the error introduced. Suppose there are three spectrum points X_-, X_o, X_+ and three pulse profile points P_+, P_o and P_- with a channel width W. To estimate the error in the term $X_o P_o W$ we

assume that both P and X are quadratic over the three channels and obtain the estimate that $X_o P_o W$ should be replaced by

$$X_o P_o W + \frac{1}{3}W(X_+ - X_-)(P_+ - P_-) + \frac{1}{45}W(X_+ + X_- - 2X_o)(P_+ + P_- - 2P_o) \quad . \qquad (7.32)$$

This should be used as an estimate of the error rather than as a basis for a correction.

In the special case of pulsed time-of-flight where there is an uncertainty in both spectrum s or X, and profile P, subdividing the intervals for X and P may improve the fit. If an interval of X is subdivided into two for a form described on the range W about E_k

$$X = a_o P_o + a_1 P_1 \frac{2(E-E_k)}{W} + a_2 P_2 (\frac{2}{W}(E-E_k)) \quad . \qquad (7.33)$$

Then for $X_\pm$, the two expressions with Legendre polynomials defined respectively on the upper and lower ranges of length $\frac{W}{2}$,

$$X_\pm = (a_o \pm a_1 \frac{\sqrt{3}}{2})P_o + (\frac{a_1}{2} \pm \frac{\sqrt{15}}{4} a_2)P_1(\mu) + \frac{a_2}{4}P_2(\mu) \qquad (7.34)$$

where the value of μ and the plus and minus signs refer to the ranges on which the polynomials are defined. Similarly we can subdivide into three ranges, each of length W/3, with the set of polynomials

$$X_\pm = (a_o + (2\sqrt{5}/9)a_2 \pm \frac{2}{3}a_1)P_o + (\frac{1}{3}a_1 \pm 2a_2\sqrt{15}/9)P_1(\mu) + \frac{1}{9}a_2 P_2(\mu) \qquad (7.35)$$

$$X_o = (a_o - 4\sqrt{5}\, a_2/9)P_o + \frac{1}{3}a_1 P_1(\mu) + \frac{1}{9}a_2 P_2(\mu) \quad . \qquad (7.36)$$

The contribution to equation 7.30 from a higher order polynomial is rapidly broken into a summation of smaller components of lower orders by subdivision of the range. The process is uni-directional. A constant is unchanged by subdivision.

The effect of subdividing the range was tried with pulsed time-of-flight. Each pulse channel was split into several subchannels (initially three), the initial pulse in each subchannel of a given channel being put equal to the channel value divided by the number of subchannels. The method of conjugate gradients was then applied to the pulse profile with the provision that changes were made in the subchannel values but not in the total of a channel. Iteration in the spectrum shape continued using all subchannels independently. The convergence of χ^2 was not greatly accelerated. The major change concerned the convergence of the spectrum; previously the spectrum changed rapidly as χ^2 changed rapidly, and after the convergence of χ^2 became slow the spectrum appeared to

start to change pathologically. Such behaviour can result from analysis of residual noise. The assigned errors were perhaps a factor of two or more too low in some cases. With the channels split, the convergence of χ^2 and of the spectrum were rapid initially. As χ^2 appeared to level off, the values of the total in a given spectrum channel also became close to constant, indicating that changes were not greatly meaningful and were producing pathological structure within individual channels.

One technique[32,36,52-54,59,60,71,72,78] used for evaluating a neutron flux as a function of energy is to activate foils and measure activity induced by the flux. For material K the activity A_k is given by

$$A_k = C_k \int_{T_k}^{\infty} \sigma_k(E)\ \phi(E)dE \qquad (7.37)$$

where $\sigma_k(E)$ is the cross section for the reaction producing the activity A_k, with a possible energy threshold T_k. The quantity C_k is a normalisation constant incorporating geometry, etc. The method of conjugate gradients has been programmed for the problem, but has not been applied.

An example of a physical research field hampered by resolution functions is the study of photonuclear cross sections[15,20,22,23,63,74-76]. At present there is no known source of γ-rays which satisfies the three criteria, that it should be high flux, monochromatic beam and simply adjustable in energy. Beams of high flux with a simply adjustable maximum energy are available from bremsstrahlung. When a monoenergetic beam of electrons is incident on a heavy metal target there is a copious beam of gammas, but to first order, the energy flux is a constant for each energy range of gammas, i.e. the number of photons per unit energy is nearly inversely proportional to their energy. Photonuclear reactions are considerably relevant to nuclear physics and therefore some effort has been made to find a better source of gammas and to unfold the cross sections of particular reactions from the bremsstrahlung yield data.

Figures 10 and 11 show that by use of the conjugate gradient technique a nuclear cross section can be unfolded even from such intractable yield curves. In Figure 11 error bars are indicated by the two curves flanking the central spectrum or cross section curve. These lines diverge from the central curve as the energy increases. If the same energy interval is used for points on a cross section curve throughout the energy range considered, this increase of uncertainty is unavoidable

since it is a rigorous consequence of the monotonic decrease of the bremsstrahlung spectrum with increase of gamma energy for a fixed electron energy. The study of photonuclear cross section has included considerable use of least structure solutions[15,22,74,83].

The conjugate gradient technique was applied to investigate the thickness of oxide films on zirconium metal. Fluorine gas can be easily made to penetrate the oxide but not the metal. The nuclear reaction $^{19}F(p,\alpha,\gamma)O^{16}$ has a resonance for (centre of mass) energy 0.878 MeV. A beam of protons of energy 0.924 MeV thus interacts with fluorine at the surface of the oxide layer and the resulting gammas can be detected. High energy protons lose energy as they go into the oxide layer and the resonance occurs for fluorine atoms deeper in the layer. Using the known cross section for the reaction and the fluorine concentration we can predict the yield curve obtained by varying the energy of the proton beam. Given the series of points of a yield curve with their associated errors the conjugate gradient technique obtains the smoothest profile of fluorine concentration consistent with the data. Some results are shown in Figures 12 and 13.

An example of a two-dimensional resolution problem appears in coincidence counting experiments. An important example is the detection of coincident fission fragments[49]. The energies E_1 and E_2 of the two fragments are measured. Using momentum conservation we have

$$m_1E_1 = m_2E_2 \quad . \tag{7.38}$$

Neglecting initially the mass and momentum of any neutrons emitted before detection we have

$$m_1 = ME_2/(E_1+E_2) \quad . \tag{7.39}$$

For each event a pair of mass coordinates can be assigned, producing a contour map with a number of events as the vertical coordinate from the plane of the two mass coordinates. The resolution functions of detectors and electronics result in more gentle contours with lower peaks, i.e. a larger area of the plane occupied. The problem of correcting for resolution is marginally simplified if the processing is done for a contour plot with energies instead of masses, which enables the two resolution functions to be treated separately; obviously storage required on a computer is greatly reduced if one dimension can be unfolded at a time, but neither this case nor the next problem, which

is irreducibly two-dimensional, has been programmed in the study reported here.

Resolution functions are an essential part of the design of optical systems[1-4,14,27,39]. In microscopy the use of electrons has been necessary to illuminate specimens because the detail to be examined came to be comparable with the wavelength of the light used to examine it. Telescopes with mirrors of ever larger diameter have been built. Whatever electromagnetic wavelength is used, for a given optical device there is a limit to the minimum angle subtended by an object if structure is to be seen; the larger the mirror the smaller the angle. As long as there is visual observation it is difficult to make any improvement on the resolution stated for the instruments. With automated processing of data the first result is often a loss of resolution.

A well known example occurs in the photography of other planets from space. The optical system leads to a television camera and the image is sent to earth as an ordered set of readings of brightnesses at a grid of points on the image. The weakest link in the system appears to be the television camera, which allows considerable spread of brightness from adjacent points. The high cost of space missions indicates that a program to unfold the data contained in the transmitted data would be worthwhile even if it achieved a fractional improvement in resolution. For selected regions of data, conjugate gradient techniques would be useful but overall faster methods of approximation to inversion of the resolution matrix by series or Fourier transformation dependent techniques would be applied.

In Appendix E it is shown that there is an intrinsic difference between two-dimensional resolution functions that are the product of two one-dimensional resolution functions (e.g. the fission fragment type) and those that are irreducibly two-dimensional, as would be expected in optics. It is shown that at least one method of one-dimensional solution becomes much more complicated when attempted in two dimensions. Truncated versions of an inverse to the resolution matrix have been used[95] in the past.

One reason for considering the series solution for computer processing of data is that the method lends itself to an analogue technique. In Figure 14 the four boxes represent schematically display television screens showing the received picture, a match to the received picture, a reconstructed (unfolded) image as seen in the space probe and

a difference picture obtained from the matched and received pictures. Each screen is photographed by a system with the same optical and television system as used in the space probe. The signals generated are processed as indicated in the diagram. At each appearance of the picture on the television screen the difference signal generates an addition to the reconstructed signal. The indicated system has not been produced. All computing reported here has been performed on digital machines. The indicated system obviously is dependent on careful technology, such as choice of life-time for the phosphors of the television screens, which is beyond the scope of the present work.

Also beyond the scope here, is consideration of the processing of audio frequency signals. For mechanical recording, the bass frequencies where a large amplitude is associated with little energy are usually attenuated. For reproduction the bass frequencies must be boosted in comparison. To produce noise-free signals, both recording and reproducing apparatus should be faithful to frequencies below the audible range; however to avoid boosting noise output that may be inaudible but can cause discomfort, the frequency range of the reproducing apparatus should be less than that of the recording apparatus.

A number of systems have been produced in computer codes as part of the research reported here. As in experimental data, the Gedunken data and experiments were in general processed better by conjugate gradient techniques than any other available method. The next best technique appears to be the series inversion method. A system chosen for purposes of illustration is reported in Appendix B.

8. CONSTRAINED SOLUTIONS

We have seen that a complete fit to the experimental data is unnecessary and can be disastrous in terms of the credibility of the spectrum obtained. Many attempts have been made to find methods of fitting data adequately while producing a credible spectrum.

A very easy method to understand and program is that of 'appropriate' solutions[34]. The solution is constrained to be positive. The spectrum points s_j are chosen to correspond to the data points $\{m_j\}$ and at each iteration the spectrum is corrected at each point by the equation

$$s_j^{(k+1)} = s_j^{(k)} \; m_j / y_j^{(k)} \tag{8.1}$$

where the superscripts refer to the number of the iteration and $y_j^{(k)}$

is the predicted yield at the j^{th} data point in the k^{th} iteration. It is easy to see that if a fit to the data is achieved one of the standard difficulties is solved. Examples[94] can be constructed where convergence does not occur. If the resolution matrix connecting s and y has a negative eigenvalue, a discrepancy of the component in the spectrum associated with that eigenvalue leads to an opposite discrepancy in the data. Since the process is not linear we cannot be sure that the spectrum discrepancy is increased in all such cases, but it appears likely. In Appendix B we show that convergence can be produced in some cases by alternating iterations.

$$s_j^{(2k+1)} = s_j^{(2k)} \; m_j / y_j^{(2k)} \qquad (8.2)$$

$$s_j^{(2k+2)} = s_j^{(2k+1)} \; y_j^{(2k+1)} / m_j \quad . \qquad (8.3)$$

For spectra with a large constant component the effect of the iterations governed by equation 8.1 is somewhat similar to that of the series inverse of the original matrix described in Section 5, and similarly the paired interations correspond to using the square of the matrix at each iteration of the series inversion.

It is also a difficulty of the 'appropriate' solution technique that the presence or absence of a constant background changes the speed of convergence and the shape of the trial function at each iteration. The form of the spectrum should be limited to protect against negative values without too great a sensitivity to the background.

In general, resolution functions smooth yields; a complicated shape in the spectrum can get lost in the yield, and, with added errors, the residue of structure can be quite overshadowed. Frequently the statistical noise can cause counterfeit structure in the data which is either not present in the yield function or present to a much less marked extent. Apparent structure at the noise level of the data should therefore be treated with caution. There is even more need to restrict complication in the spectrum. The method of 'least structure'[15,22,64,69,73,74,79,83] places a measure on the structure and minimises a combination of the structure and the residual errors.

One commonly used measure of structure in the spectrum s at s_j is the second difference of successive points

$$(\vec{Fs})_j = - s_{j-1} + 2s_j - s_{j+1} \quad , \qquad (8.4)$$

which is a measure of the second degree term required to fit a parabola

through the three points.

A minimum is then sought for the quantity

$$\gamma(\vec{Fs})^T(\vec{Fs}) + (\vec{y}-\vec{m})^T w^2 (\vec{y}-\vec{m})$$

$$= \gamma(\overleftarrow{s}F^T\vec{Fs}) + (\overleftarrow{sQ}^T-\overleftarrow{g})\ (\vec{Qs}-\vec{g}) \quad . \tag{8.5}$$

It is easy to see that the minimum occurs where

$$\gamma F^T\vec{Fs} + Q^T(\vec{Qs}-\vec{g}) = 0 \tag{8.6}$$

$$(\gamma F^T F + Q^T Q)\vec{s} = Q^T\vec{g} \quad . \tag{8.7}$$

We should first note (as proved in Appendix D) that for any positive value of γ there is a solution for $\vec{s}$ which satisfies equation 8.7. The value chosen for γ is in general picked as a compromise. The larger the value of gamma the more of both genuine and noise related structure is suppressed. There are considerable advantages in the method, which has been used extensively in photonuclear unfolding, but it has two related defects. The analysis in Appendix D shows that the process works well if the small eigenvalues of the matrix Q^TQ are related to eigenvectors $\vec{v}$ which are associated with large matrix elements $\overleftarrow{v}F^T\vec{Fv}$, e.g. if elements of $\vec{v}$ alternate in sign and are associated with small components in the vector $Q^T\vec{g}$. The suppression of such components is independent of their magnitude. This can be a defect, but it is plausible to expect small components to be associated with small eigenvalues. The second defect concerns small eigenvalues of Q^TQ with associated eigenvectors $\vec{v}$ that also give small matrix elements $\overleftarrow{v}F^T\vec{Fv}$. The least structure solution does not suppress the corresponding component in the spectrum. 'Slit' resolution functions of the type discussed in Appendix B are usually insensitive to components that vary sinusoidally with period close to the slit width. Any noise in such a component leads to a set of peaks and associated troughs in the calculated spectrum. The least structure fit is likely to be successful as a starting point in most cases, but often an examination of the data is needed to indicate which structure is significant.

A doubly constrained solution has been tried which uses conjugate gradient technique to minimise structure, and some other method to ensure that the spectrum remains positive. One such solution[46] involved taking the gradient with respect to the logarithms of the spectrum values; however the authors found difficulty with convergence.

Another method with a conjugate gradient technique was used to obtain convergence for a resolution problem couched to produce the Hilbert matrix as an approximation to the resolution function. As described in Appendix F, the problem of actual inversion is usually outside the capabilities of single precision arithmetic. However it was possible to isolate a positive spectrum by conjugate gradient iterations leading to a fit of the data at the limits of single precision arithmetic. All spectrum points with negative values were set to zero, and omitted in further fitting. The conjugate gradient technique was repeated but the technique, while successful, is time consuming and would be expected to be of limited value in dealing with actual experimental data.

APPENDIX A

PROPERTIES OF CONJUGATE GRADIENT TECHNIQUES

Given a matrix Q and a data significance vector $\vec{g}$ the conjugate gradient technique is one of many used to seek a vector leading to an acceptable approximation

$$Q\vec{a} = \vec{g} \quad . \tag{A.1}$$

The technique consists of dissecting the data significance vector $\vec{g}$ derived from $\vec{m}$ into orthogonal vectors so that the corresponding spectrum vectors are conjugate with respect to a symmetric matrix of which Q is a factor. The most complicated form used in the body of the paper involves expanding $\vec{a}$ in terms derived from a diagonal weight matrix ν^2 multiplied by Q^T. The symmetric matrix needed for the method is then the N x N product matrix

$$Q\ \nu^2\ Q^T \quad . \tag{A.2}$$

The Lanczos theorem generates a set of orthonormal vectors initiated from $\vec{g}$

$$Q\ \nu^2\ Q^T\ \vec{g} = \vec{v}(1)\ \alpha\ \hat{v}(1) \tag{A.3}$$

where $\hat{v}(1)$ is a unit vector. Then

$$Q\ \nu^2\ Q^T\ \vec{v}(k) = \vec{v}(k+1) + \sum_{r=0}^{k-1} \alpha_r^{(k)}\ \vec{v}(k-r) \quad . \tag{A.4}$$

We construct each vector to be orthogonal to all before it by the Gram-Schmidt process. Equation A.4 using the symmetric product nature of matrix $Q\ \nu^2\ Q^T$, gives

$$\overleftarrow{v}(k+r)\ Q\ \nu^2\ Q^T \vec{v}(k) = \overleftarrow{v}(k)\ Q\ \nu^2\ Q^T \vec{v}(k+r) = 0 \text{ for } |r| > 1 \ . \tag{A.5}$$

The summation in equation A.4 has for exact arithmetic at most two non-zero terms. The number of orthogonal terms to be generated in the sequence $\vec{v}(k)$ is certainly not greater than the number N of data points. When the process is complete (in exact arithmetic) we reach a null vector and expand $\vec{g}$ as far as possible in terms of the set of $\hat{v}(k)$, with a possible residue $\vec{g}_L$ orthogonal to all the set;

$$\vec{g} = \sum_k C^{(k)}\ \hat{v}(k) + \vec{g}_L \tag{A.6}$$

$$Q\nu^2Q^T\vec{g} = \sum_k (c^{(k+1)}\alpha_1^{(k+1)} n_k/n_{k+1} + c^{(k)}\alpha_0^{(k)} + c^{(k-1)} n_k/n_{k-1})\hat{v}_k$$

$$+ Q\nu^2Q^T \vec{g}_L = \vec{v}_k^{(1)} \qquad (A.7)$$

where $$\vec{v}^{(k)} = n_k \hat{v}^{(k)} \quad . \qquad (A.8)$$

First, from the stated orthogonality it follows that $\overleftarrow{g}_L \vec{v}^{(k)}$, $\overleftarrow{g}_L \vec{v}^{(1)}$, $\overleftarrow{g}_L Q\nu^2Q^T\vec{g}$, $\overleftarrow{g}_L Q\nu^2Q^T\vec{g}_L$ and $\nu^2Q^T\vec{g}_L$ are all zero so that

$$Q\nu^2Q^T \vec{g}_L = 0 \quad . \qquad (A.9)$$

The coefficients of $\hat{v}^{(k)}$ except for $\vec{v}^{(1)}$ in equation A.7 must vanish separately. If the first vectors of the set are generated and the associated components from $\vec{g}$ are removed

$$\vec{e}^{(r)} = \sum_{k=r+1} c^{(k)} \hat{v}^{(k)} + \vec{g}_L \qquad (A.10)$$

and

$$Q\nu^2Q^T\vec{e}^{(r)} = c^{(r+1)}\alpha_1^{(r+1)} n_r/n_{r+1} \hat{v}^{(r)} - c^{(r)} n_{r+1}/n_r \hat{v}^{(r+1)} . \qquad (A.11)$$

So that, as indicated at equation 4.32, only a component of the preceding vector must be removed to generate the next data significance vector of the sequence from the residual discrepancy of the fit to the data.

At each step the vector $\vec{b}^{(k+1)}$ can be readily generated so that

$$\nu^2Q^T \vec{e}^{(k)} = \vec{b}_1^{(k+1)} \qquad (A.11)$$

$$Q \vec{b}_1^{(k+1)} = \vec{v}^{(k+1)} c^{(k+1)}/\mu^{(k+1)} + v^{(k)} c^{(k+1)} n_{k+1}/n_k^2 \qquad (A.12)$$

and hence

$$\vec{b}^{(k+1)} = \vec{b}_1^{(k+1)} - \vec{b}^{(k)} \mu^{(k)} c^{(k+1)} n_{k+1}/(n_k^2 c^{(k)}) \quad . \qquad (A.13)$$

The available portion of the inverse at any stage is

$$\sum_{k=1}^{r} \vec{b}^{(k)} \overleftarrow{v}^{(k)}/\overleftarrow{b}^{(k)} Q^TQ \vec{b}^{(k)}$$

so that

$$\vec{a}^{(r)} = \sum_{k=1}^{r} \vec{b}^{(k)} \overleftarrow{v}^{(k)}/\overleftarrow{b}^{(k)} Q^TQ \vec{b}^{(k)} \vec{g} \quad . \qquad (A.14)$$

Equation A.14 depends, but not explicitly, on the weighting matrix ν^2.

The method is a conjugate gradient technique, but not one of steepest descents if the weighting matrix differs from the identity matrix.

We now consider the way in which particular eigenvectors $\hat{u}_k$ of the product matrix $Q\nu^2Q^T$ are removed from $\vec{e}^{(n)}$. As a result of equation A.10

$$\vec{e}^{(o)} = \vec{g} = \vec{g}^{(o)} = \sum_k d_k \hat{u}_k \quad . \tag{A.15}$$

Note that this d_k is not the same as the data significance vector introduced in equation 6.15. We are simply short of letters.

We define $\vec{g}^{(n)}$ by

$$\vec{g}^{(n)} = (Q\nu^2Q^T)^n \vec{g}^{(o)} \tag{A.16}$$

and G_{nm} by

$$G_{nm} = \overleftarrow{g}^{(n)} . \vec{g}^{(m)} \quad . \tag{A.17}$$

We have immediately

$$G_{n+m\ o} = G_{nm} = G_{mn} = \sum_k d_k^2 (\lambda_k)^{m+n}, \tag{A.18}$$

where λ_k is the (non-negative) eigenvalue of the product matrix $(Q\nu^2Q^T)$ associated with the eigenvector $\hat{u}_k$. The eigenvalues are conveniently ordered so that for positive r, $\lambda_{k+r} \geqslant \lambda_k$. We can establish by induction that $\vec{e}^{(n)}$ is a linear combination of $\vec{g}^{(o)}$, $\vec{g}^{(1)}$ ... $\vec{g}^{(n)}$ and that

$$\overleftarrow{g}^{(m)} . \vec{e}^{(n)} = 0 \qquad m=1,\ldots n \quad . \tag{A.19}$$

We could write the linear combination with coefficients f_m, suppressing dependence of f_m upon n, as

$$\vec{e}^{(n)} = \vec{g}^{(o)} - \sum_{m=1}^{n} f_m \vec{g}^{(m)} \tag{A.20}$$

leading to the equation set derived by combining equations A.19 and A.20

$$\sum_m f_m G_{mr} = G_{or} \qquad r=1, \ldots n \quad . \tag{A.21}$$

The members of the set $\{f_m\}$ can then be evaluated by inverting the n x n matrix, consisting of the G_{mr}. First the determinant of G_{mr} is expressed as a sum of determinants in which each row contains powers of only one λ_k. A set of n members is selected from the N λ_k and written as an element of the single determinant, before being summed

$$H_{mr} = d_k^2 \lambda_k^{m+r} \qquad m=1, \ldots n \tag{A.22}$$

where λ_k is the m^{th} member of the chosen set.

Now the determinant of H_{mr} has a factor $d_k^2 \lambda_k^{m+1}$ appropriately for each letter. Any of the forms $\lambda_k(m)$, k_m and λ_{km} can be used to identify the labels. The remainder of the determinant of H_{mr} is the antisymmetric product of the n letters.

$$\prod_{m'>m} (\lambda_{k'}(m') - \lambda_k(m)) \stackrel{def}{=} \text{Prod}(k_1,\ldots k_n) \quad . \tag{A.23}$$

The product A.23 retains its magnitude but changes sign as we exchange any pair. The same set of n letters is used but first all must be made distinct, or the product is zero. We then sum over all the different determinants available by permuting them. We write

$$\text{Prod}(k_1,\ldots k_n) = \varepsilon(k_1,\ldots k_n)\,|\text{Prod}(k_1,\ldots k_n)| \tag{A.24}$$

where

$$\begin{aligned} \varepsilon(k_1,\ldots k_n) &= 1 \quad \text{if } \lambda_{kr} < \lambda_{k'r+1} \text{ for all } r = 1,\ldots n-1 \\ &= 0 \quad \text{if } \lambda_{kr} = \lambda_{k'r+1} \text{ for any } r \end{aligned} \tag{A.25}$$

and changes sign for any interchange $k_r \leftrightarrow k_s$.

Then

$$\begin{aligned} \sum_{\text{set chosen}} \det|H_{mr}| &= |\text{Prod}(k_1,\ldots k_n)| \quad \times \\ \sum_{\text{permutation}} \varepsilon(k_1,\ldots k_n) &\prod_k d_k^2 \lambda_{km}^{m+1} \quad , \end{aligned} \tag{A.26}$$

the summation over permutations is merely the definition of a determinant of a matrix with the element $d_k^2 \lambda_k^{m+1}$ in position k,m and with the set of $\{\lambda_k\}$ written increasing with k. We obtain

$$\sum_{\text{set chosen}} \det|H_{mr}| = (\text{Prod}^2(k_1,\ldots k_n)) \prod_k d_k^2 \lambda_k^2 \quad . \tag{A.27}$$

The determinant of G_{mr} is a sum of the right-hand side of equation A.27 for all choices of the set of n letters out of, at most, the N available.

Next we evaluate the (n-1)x(n-1) cofactors, each associated with an element G_{mr} of the determinant. As before each cofactor can be expressed as a sum over all possible choices of sets of n-1 distinct letters from the set of N, preceded by a sum over the possible permutations of each set for determinants in which each row is concerned with only one eigenvalue of $Q\nu^2Q^T$. To simplify evaluation we recall properties of the elementary symmetric functions for n-1 letters; using the notation

$$S_1^{(n-1)} = \sum_{k=1}^{n-1} \lambda_k \qquad (A.28)$$

$$S_2^{(n-1)} = \sum_{k=1}^{n-2,n-1} \lambda_k \lambda_{k'} \text{ , etc.,} \qquad (A.29)$$

we recall that

$$\prod_{k=1}^{n-1} (\lambda-\lambda_k) = \lambda^{n-1} - S_1^{(n-1)} \lambda^{n-2} + S_2^{(n-1)} \lambda^{n-3} \ldots (-)^{n-1} S_{n-1}^{(n-1)} . \qquad (A.30)$$

The polynomial so formed is zero if $\lambda = \lambda_k$, any member of the set; we can follow with the cofactor determinants the steps of factorisation used above and must now evaluate a determinant whose k^{th} row has elements $1, \lambda_k, \lambda_k^2 \ldots \lambda_k^{n-1}$ with one power of λ_k missing. This power, say r, is the same for all rows.

We make n-1-r column exchanges until a typical row reads

$$1, \lambda_k, \ldots \lambda_k^{n-1}, (-)^{n-1-r} \lambda_k^{n-1}, \lambda_k^{r+1}, \ldots \lambda_k^{n-2} . \qquad (A.31)$$

Equation A.30 is used to rewrite the cofactor as a sum of determinants. All determinants have two columns related by a constant except the single one in which a typical row is

$$1, \lambda_k, \ldots \lambda_k^{r-1}, S_{n-1-r}^{(n-1)} \lambda_k^r, \lambda_k^{r+1}, \ldots \lambda_k^{n-2} . \qquad (A.32)$$

The process leading to equation A.27 can then be parallelled, with two evaluations of determinants of the type just discussed. We obtain the minor in position r,s of the determinant of G_s expressed as a sum over sets of n-1 letters;

$$\text{Minor } (G)_{rs} = \sum_{\substack{\text{sets of} \\ n-1 \text{ letters}}} S_{n-r}^{(n-1)} S_{n-s}^{(n-1)} \text{Prod}^2(k_1, \ldots k_{n-1}) \prod_k (d_k^2 \lambda_k^2) . \qquad (A.33)$$

The coefficient of $\hat{u}_s$ in the reconstructed data $\vec{x}^{(n)}$ is then

$$d_s^{(n)}(x) = d_s \, (\text{Num})_s^{(n)} / \text{Den}^{(n)} \qquad (A.34)$$

$$\text{Den}^{(n)} = \sum_{\substack{\text{sets of} \\ n \text{ letters}}} \text{Prod}^2(k_1, \ldots k_n) \prod_k d_k^2 \lambda_k^2 \qquad (A.35)$$

and the numerator can be obtained by a double application of equation A.30 to the result A.33 so that one letter $\lambda_{k'}$ is singled out,

$$\mathrm{Num}_s^{(n)} = \lambda_s \sum_{k'} d_{k'}^2 \lambda_{k'} \sum_{\substack{\text{sets of} \\ n-1 \text{ letters}}} \mathrm{Prod}^2(k_1,\ldots k_{n-1}) \prod_k d_k^2 \lambda_k^2 (\lambda_k - \lambda_s)(\lambda_k - \lambda_{k'}) \tag{A.36}$$

and we write the coefficient of $\hat{u}_s$ in the discrepancy vector $\vec{e}^{(n)}$

$$e_s^{(n)} = d_s(1 - \mathrm{Num}_s^{(n)}/\mathrm{Den}^{(n)}) \ . \tag{A.37}$$

It then becomes an exercise in symmetric functions and partial fractions to show that

$$e_s^{(n)}\ \mathrm{Den}^{(n)} = d_s \sum_{\substack{\text{sets of} \\ n \text{ letters}}} \mathrm{Prod}^2(k_1,\ldots k_n) \prod_k d_k^2 \lambda_k (\lambda_k - \lambda_s). \tag{A.38}$$

These expressions, while explicit, are not completely transparent. We can however obtain some information by inspection. The factor λ_s in equation A.36 absent from equation A.38, shows that any eigenvector associated with a zero eigenvalue remains solely in the discrepancy between the data and the reconstruction, no matter how many iterations are undertaken. Also if there are degenerate eigenvalues with assigned eigenvectors, any linear combination of these eigenvectors is itself an eigenvector. The component present in the data from any set of degenerate eigenvalues is treated as a single eigenvector. When the number of iterations is equal to the number of discrete non-zero eigenvalues the numerator A.36 and denominator A.38 each contain a single non-zero term. For this value of n

$$\begin{aligned} e_s^{(n)} &= 0 \ , \quad \lambda_s \neq 0 \\ e_s^{(n)} &= d_s \ , \ \lambda_s = 0 \quad . \end{aligned} \tag{A.39}$$

For usual data, meaningful iterations would be expected to terminate long before the stage described by equation A.39. After a number of iterations, the rounding error of any given computer might re-introduce small components of vectors already removed. This is not usually a serious problem; as noted in connection with equation A.5 its orthogonality was only as good as rounding errors permitted. The re-introduced error component should be smaller than the ordinary errors of measurement and, after at most two further iterations, can contribute again, through the discrepancy vector, to its own removal.

The technique that provided equation A.38 can be used to evaluate the coefficient of $\hat{u}_s$ in $\vec{v}^{(n+1)}$.

Then

$$d_s^{(n)}(v)\ \mathrm{Den}^{(n)} =$$

$$d_s\lambda_s \sum_{\substack{\text{sets of}\\ n\ \text{letters}}} \mathrm{Prod}^2(k_1,\ldots k_n) \prod_k \lambda_k^2 d_k^2(\lambda_s-\lambda_k) \quad . \tag{A.40}$$

Again if $\lambda_s = 0$, $\hat{u}_s$ is never a portion of any $\vec{v}^{(k)}$. We may write therefore

$$\vec{g}-\vec{g}_L = \sum_1^{N'} \mu^{(k)}\,\vec{v}^{(k)} = \sum_1^{N'} d_r\,\hat{u}_r \tag{A.41}$$

and

$$\vec{g}_L = \sum_{N'}^{N''} d_s\,\hat{u}_s \tag{A.42}$$

where

$$N' \leqslant N'' \leqslant N,\ \lambda_s = 0 \tag{A.43}$$

for all s in the summation in A.42, and N"-N' is the number of non-zero duplicate values of eigenvalues. For small values of n, the factor λ_s in equation A.40 would appear to indicate that components of eigenvectors of large eigenvalues are removed first. The other factors eventually become important, forcing eigenvectors of lower eigenvalues into consideration. It would therefore be plausible that the ratio

$$\vec{v}^{(n)}\ Q\nu^2Q^T\vec{v}^{(n)} / \vec{v}^{(n)}\cdot\vec{v}^{(n)} \quad , \tag{A.44}$$

which is a weighted mean of the eigenvalues present in $\vec{v}^{(n)}$, should shrink as n increases.

We would also expect that for a positive spectrum and a positive resolution function the data would depend heavily on larger eigenvalues and hence that the contribution to changes in χ^2 from successive vectors $\vec{v}^{(n)}$ would also shrink; there is usually an initial shrinkage followed by oscillation in the changes in χ^2. We now give a counter example to show that both trends are plausible rather than necessary. We set

$$Q\nu^2Q^T = \lambda\begin{pmatrix} p^2 & pq \\ pq & p^2 \end{pmatrix} + \lambda\varepsilon\begin{pmatrix} q^2-pq & \\ -pq & p^2 \end{pmatrix} \tag{A.45}$$

and

$$\vec{g} = d\begin{pmatrix} p \\ q \end{pmatrix} + Md\begin{pmatrix} -q \\ p \end{pmatrix} \tag{A.46}$$

where

$$p^2 + q^2 = 1 \quad . \tag{A.47}$$

By inspection, the matrix is positive definite if λ and ε are greater than zero and ε is less than one. The data set is positive if M, d, p and q are positive and $Mq < p$. The conjugate gradient technique to unfold the data in equation A.46 is complete in at most two iterations. It is easy to check that the first does less to improve the fit than the second does if εM is positive, and is less than the positive member of the pair

$$(M-1)/(M+1) \quad , \quad (M+1)/(1-M) \quad . \tag{A.48}$$

For successive values of the ratio A.44 we have

$$\frac{\overleftarrow{v}^{(1)} Q \nu^2 Q^T \overrightarrow{v}^{(1)}}{\overleftarrow{v}^{(1)} . \overrightarrow{v}^{(1)}} < \frac{\overleftarrow{v}^{(2)} Q \nu^2 Q^T \overrightarrow{v}^{(2)}}{\overleftarrow{v}^{(2)} . \overrightarrow{v}^{(2)}} \tag{A.49}$$

provided that ε is less than unity and

$$\varepsilon M > 1 \quad . \tag{A.50}$$

For $M > 1$ conditions A.48 and A.50 are mutually exclusive.

We do not have a form for the matrix Q, and in fact have a very large number of degrees of freedom for it, since in particular the number of rows is two and the number of columns is only restricted to be at least two. If we make

$$\nu^2 = I \tag{A.51}$$

and

$$Q = Q^T = \lambda^{\frac{1}{2}} \begin{pmatrix} p^2 & pq \\ pq & q^2 \end{pmatrix} + (\varepsilon\lambda)^{\frac{1}{2}} \begin{pmatrix} q^2 & -pq \\ -pq & p^2 \end{pmatrix} \tag{A.52}$$

we find the requirement for a positive spectrum

$$Mq < \varepsilon^{\frac{1}{2}} p \tag{A.53}$$

which is a stronger condition than was imposed by the requirement of positive data. We note that data commonly are expected to have M and ε commensurate, i.e. both larger or both less than unity. In this case neither condition A.48 nor A.50 holds. We expect data and spectrum both to have large components associated with larger eigenvalues. A program to use a conjugate gradient technique should check the behaviour of the fraction A.44 in order to be sure that the error analysis of Section 6 is soundly based.

It is sometimes convenient to describe the conjugate gradient process by means of projection operations. At each step we truncate the space being dealt with by one dimension. All vectors in the truncated space are orthogonal to those removed. We can then formally restart the iterations in the truncated space. It is not usually practicable to carry out the truncation as a computer technique, but some insight into the process involved can be gained. At this 'first' iteration we define

$$\sum_{\ell} d_{\ell}^2 = \overline{d^2} \tag{A.54}$$

$$\sum_{\ell} d_{\ell}^2 \lambda_{\ell} = \overline{\lambda}\ \overline{d^2} \tag{A.55}$$

and

$$\sum_{\ell} d_{\ell}^2 \lambda_{\ell}^2 = \overline{\lambda^2}\ \overline{d^2} \tag{A.56}$$

we then find

$$\mu^{(1)}\ v_{\ell}^{(1)} = x_{\ell}^{(1)} = d_{\ell}\lambda_{\ell}\ \overline{\lambda}\ /\ \overline{\lambda^2} \tag{A.57}$$

so that

$$\overleftarrow{x}^{(1)}\ \overrightarrow{x}^{(1)} = \overline{d^2}\ \overline{\lambda}^2\ /\ \overline{\lambda^2} \tag{A.58}$$

and

$$e_{\ell}^{(1)} = d_{\ell}\ (\overline{\lambda^2} - \lambda_{\ell}\overline{\lambda})/\overline{\lambda^2} \quad . \tag{A.59}$$

For the new value of χ^2

$$\chi^2 = \overline{d^2}\ (1 - \overline{\lambda}^2\ /\ \overline{\lambda^2}) \quad . \tag{A.60}$$

The first iteration is efficient in the fitting process provided that there is a large component of the data associated with each of a few of the largest eigenvalues. At subsequent iterations there is one less eigenvalue in this group and the remaining eigenvalues are moved a little from their positions on the complete space. The process becomes inefficient when in the truncated space

$$\overline{\lambda^2} >> \overline{\lambda}^2 \quad . \tag{A.61}$$

We expect condition A.61 to hold when we have removed most of the components in the larger eigenvalues and are left with much larger values for the new quantities to be associated with small λ_{ℓ} than are still associated with large λ_{ℓ}. It is always our hope that by this stage the value of d_{ℓ} can be considered to be noise for all ℓ.

APPENDIX B

A SIMPLE EXAMPLE

It is convenient to compare techniques of unfolding using a model with simple properties, a model which is simple to understand rather than physically realistic, but somewhat intractable to illustrate the limitations of techniques. Troubles with the end points of the spectrum are avoided by using a spectrum defined on the circumference of a circle. The resolution function is defined to measure the mean value of the spectrum on a range that is close to half the circumference.

In matrix form the spectrum s is assigned a value s_k in channel k and the yield is defined as the average over 2n+1 channels

$$y_k = \frac{1}{2n+1} \sum_{\ell=k-n}^{\ell=k+n} s_\ell \quad . \tag{B.1}$$

The circumference of the circle is then divided into N channels and three special cases can be distinguished

$$N = 4n+1, \tag{B.2}$$

$$= 4n+2 \tag{B.3}$$

and

$$= 4n+3 \quad . \tag{B.4}$$

All subscripts for the model are interpreted modulo N. In this appendix we concentrate attention on the cases B.3 and B.4 and find that iterative techniques provide little distinction between them. In Appendix D we consider least structure calculations for all three cases.

For each case we can write the matrix equation

$$\vec{y} = P \vec{a} \quad , \tag{4.3}$$

where a_k is another name for s_k, and we separate usage by making a_k the value derived from the data. The matrix P is defined by equation B.1 and has an inverse for cases B.2 and B.4 but not for B.3. For case B.3 the matrix P is singular.

The eigenvalues of P for all three cases are given by the expression

$$\lambda_k = (2n+1)^{-1} \sin(k(2n+1)\ \pi/N) \ /\sin(k\pi/N) \quad k = 1,\ldots N-1 \tag{B.5}$$

as well as the special value $\lambda_o = 1$. We thus have for equation B.2, 2n degenerate eigenvalues of zero and, of the remainder, there are n-1 degenerate pairs. If the same standard deviation σ is ascribed independently to each measurement the zero eigenvalues give 2n relations of

the form

$$\sum_k m_k - (2n+1)(m_\ell + m_{\ell+2n+1}) = 0 \pm \sigma(8n^2+4n)^{\frac{1}{2}} \quad . \tag{B.6}$$

We see that the mean value can be established

$$\bar{a} = \bar{m} = (4n+2)^{-1} \sum_k m_k \pm (4n+2)^{-\frac{1}{2}} \sigma \quad . \tag{B.7}$$

Any single pair of opposite channels can be set to have any sum, or even 2n such pairs, with the final pair sum to be determined by equation B.7. The difference of such a pair is given by the data,

$$a_\ell - a_{\ell+n+1} = \tfrac{1}{2}(2n+1)(m_{\ell+n} + m_{\ell-n} - m_{\ell+n+1} - m_{\ell-n-1}) \pm (2n+1)\tilde{\sigma} . \tag{B.8}$$

We may write

$$a_\ell = \bar{a} + \tfrac{1}{4}(2n+1)(m_{\ell+n} + m_{\ell-n} - m_{\ell+n+1} - m_{\ell-n-1}) \quad . \tag{B.9}$$

The form of equation B.9 ignores any non-zero values found in equation B.6 as arising from statistical noise. It is likely, however, that there is still noise incorporated in the form given for the spectrum since the value of χ^2 is slightly less than half the number indicated by the number of independent measurements. In a physical measurement we usually search for structure in the spectrum that is at worst about half the width of our resolution function. In this model unfolding techniques are given the task of recovering much finer detail. The first criterion for success remains that the features present in the original spectrum, and smoothed out considerably in the data, begin to emerge more sharply as any unfolding process is iterated.

Before we consider actual numerical operations we investigate the situation for data gathered under the conditions of equation B.4. The matrix P no longer has any zero eigenvalues, but some vectors in the spectrum space may be expected to have very small effects in data space. The following exact spectrum would produce a given set of data;

$$a_k = \sum_\ell m_\ell - (2n+1)(m_{k+n+1} + m_{k-n-1}) \quad . \tag{B.10}$$

Using an actual value s_k and a value a_k calculated from data

$$a_k = s_k \pm (8n^2 + 4n+1)^{\frac{1}{2}} \sigma \tag{B.11}$$

$$\approx s_k \pm 2.8n\sigma \quad . \tag{B.12}$$

If equation B.10 is used on several independent sets of data the mean value of a_k is expected to be s_k and its root mean square variation

$2.8n\sigma$. We can also find the mean values of some products of channel values in repeated trials

$$\overline{a_k\, a_\ell} = s_k s_\ell + (8n^2 + 4n+1)\sigma^2 \qquad \text{for } k = \ell \text{ as in equation B.11}$$

$$= s_k s_\ell + 4n^2\sigma^2 \qquad \text{for } k = \ell \pm 2(n+1) \qquad \text{(B.13)}$$

and

$$= s_k s_\ell - (4n+1)\sigma^2 \qquad \text{otherwise} \quad . \qquad \text{(B.14)}$$

A small change in the measured data leads to a drastic change in values of the calculated spectrum in two nearly opposite channels, corresponding roughly to the complete liberty in setting the sums of opposite channels when the condition B.3 was used. The χ^2 value of the solution in equation B.12 is the highly improbable number zero and removal of the second half of the noise component has changed the assigned errors. For the fit to data in equation B.9, where the spectrum has a clear restriction on its form, each value a_k has a standard deviation of approximately $n\sigma$. Without the restriction each value in B.12 has more than double this standard deviation.

In an iterative solution, iterations are expected to stop when χ^2 is still double the value reached in equation B.9. We expect to be able to reproduce values of individual a_k considerably better than equation B.9. Any peak in the unfolded spectrum can be expected to be associated with a trough approximately opposite, since the components in the spectrum that lead to major effects in the data are the same under conditions B.3 and B.4. If similar data are used for cases B.3 and B.4 an iterative solution can be expected to reach much the same spectrum.

The standard deviation is defined in terms of the change required in a_k to increase χ^2 by unity and is then $(2n+1)^{\frac{1}{2}}\sigma$; this is a reproducibility criterion rather than a limitation on the possible spectrum.

The computer calculation is set up with n=46, i.e. 2n+1 = 93, and N=186 and 187, and the input spectrum (Figure 15), as

$$s_k = 10^5 + \delta_{k,47}\ 4.65 \times 10^5 \qquad \text{(B.15)}$$

so that

$$y_\ell = 1.05 \times 10^5 \qquad \ell = 1,\ldots 93 \qquad \text{(B.16)}$$

$$= 1.00 \times 10^5 \qquad \ell = 94,\ldots -1 \quad .$$

For the values chosen a matrix inversion routine would not be expected to produce an accurate form of equation B.10, much less the form in

equation B.6. In fact some care is required in programming if accurate results are required from equation B.10 using floating point arithmetic and six significant hexa-decimal digits.

We use the values of y_ℓ alternatively as error free, or (Figure 16) generated as counted values with Poisson statistics, so that σ is approximately 320. Two criteria can be used to define success; the first is the reduction in the discrepancy between the reconstructed and the measured data. If the discrepancies are solely statistical and include all the statistical variations, the sum of their squares is expected to be 1.9×10^7; this is denoted by $S^2 \times 10^7$ and S^2 is tabulated in Table B.1. The second criterion concerns the concentration of variation in the spectrum. In equation B.15 all the variation is in channel 47. In the values for a_k the positive variations from the three channels 46, 47, 48 are added to negative variations in the opposite channels as a measure of the effectiveness of the method. For case B.3 the three channels opposite are used, i.e. 139, 140 and 141, for B.4 the two channels 140 and 141, and in Appendix D with B.2 the two channels, 139 and 140. The target for the concentration for the spectrum is then 4.65×10^5 and we write the value achieved as $C \times 10^5$; C is tabulated in Table B.1.

A glance at the table shows that the series method and the appropriate structure method are both diverging, at least as far as twenty iterations. The difference between them is just emerging. The series solution goes on diverging without limit. The 'appropriate solution' is limited to positive values and eventually converges to a form with a high peak close to channel 140 and over-valued elements for channels 1 to 23 as well as channels 71 to 93. Other elements are forced towards zero because yields there are approximately twice the data values. The 'solution' is stable under iteration but wrong (Figure 20).

The defects of the series method and of the appropriate solution method are associated with negative eigenvalues of P. We rewrite the series method to deal with the square of the matrix, or more generally P^TP, so that the eigenvalues are all positive in the matrix to be inverted. With the matrix and data of this appendix the same result can be achieved for the appropriate solution if we replace the single process of iteration, described by equation 8.1, by a pair of processes for each iteration, as set out in equations 8.2 and 8.3. From Figures 19 and 21, and also from the table, series inversion with a square matrix

TABLE B.1

number of channels		186	186	186	186	187	187	187	187
iteration number		10	10	20	20	10	10	20	20
method of iteration \ data with(+), without(-) statistical error		+	-	+	-	+	-	+	-
conjugate gradient	S^2	3.25	1.98	1.70	0.43	3.88	2.14	1.87	0.73
	C	3.61	3.33	5.16	5.13	3.10	2.88	5.04	4.72
conjugate gradient (positive eigenvalues)	S^2	2.97	1.48	1.54	0.30	3.34	1.68	1.55	0.59
	C	3.63	3.71	5.19	5.25	3.73	3.18	5.33	4.94
series	S^2	390	351	1.76×10^4	1.58×10^4	387	358	1.77×10^4	1.64×10^4
	C	-	-	-	-	-	-	-	-
series (positive eigenvalues)	S^2	17.1	14.6	12.7	10.6	17.6	-	13.3	-
	C	0.54	0.53	0.73	0.71	0.47	-	0.64	-
'appropriate'	S^2	397	355	2.05×10^4	1.84×10^4	392	364	2.08×10^4	1.90×10^4
	C	-	-	-	-	-	-	-	-
'appropriate' (alternating ratios)	S^2	17.1	16.4	12.7	11.2	17.6	16.4	13.3	11.2
	C	0.55	0.54	0.73	0.71	0.47	0.48	0.64	0.62
'appropriate'(structured)	S^2	99.6	95.6	78.6	76.2	91.3	88.1	67.0	65.3
	C	0.48	0.48	0.96	0.93	0.68	0.71	1.28	1.26

resembles the appropriate solution with alternating ratios. The resemblance is played down slightly in the figures. In Figure 19 there are 184 channels for the complete circle so that channels 140 and 141 share the status of being opposite channel 47. In Figure 21 with a total of 186 channels, channel 140 is isolated as the minimum. There is a slight difference that has already emerged in the results after ten iterations. The series solution has no bar against negative values for the spectrum. The spectrum strength gathered up is the same for the two methods, but the appropriate (alternating) solution has a slightly greater concentration in the peak and a smaller concentration in the trough.

By studying the particular problem one might expect to produce a specialised form of the appropriate solution to do better in a given case. Thus

$$s^{(k+1)} = s^{(k)} \; m_j/y_j^{(k)} \tag{8.1}$$

is replaced by

$$s_j^{(k+1)} = s_j^{(k)} \; (m_{j+n} + m_{j-n})/(y_{j+n}^{(k)} + y_{j-n}^{(k)}) \quad . \tag{8.1a}$$

The form 8.1a concentrates attention on the region where a change in s_j leads to large changes in the shape of the data; this method is called the 'appropriate' (structure) solution. In Figure 22 it is seen to be promising, but convergence is slow.

From Figures 17 and 18 we see that the conjugate gradient technique is the most successful of the iterative solutions. Since the matrix is symmetric the simplest form of the conjugate gradient technique works quite well. The results can be improved slightly by using the form suitable to non-symmetric matrices, which also ensures that the eigenvalues are all positive. From Table B.1 we see that twenty iterations are more than sufficient for both methods. The residual errors have moved into the noise range, and the strength associated with the peak and trough is greater than required.

We would prefer to use a conjugate gradient solution if possible, but the series solution with positive eigenvalues is locating important features of the spectrum, as is the modified appropriate solution. We can always set up the position eigenvalue form of the series solution, but special study may be required to find a suitable form of the appropriate solution.

APPENDIX C

FOURIER ANALYSIS OF DATA

In Section 5 it was found that high frequency components of a formal Fourier transform of a resolution function destroyed the initial simplicity of the treatment. It was seen in that section that the important properties of the resolution function depended on lower frequency behaviour. The Fourier analysis of data given in digital form at a finite set of points is described in this appendix.

It may be convenient to write the data in a form with a continuous Fourier spectrum[91], or in terms of a finite set of components[92]. It may be preferable to assign the Fourier spectrum to the lowest possible band, or to select a band not including low frequencies. As usual in data analysis, the possibility of errors of observation should be kept in mind. We limit ourselves to digital data, avoiding analogue devices such as high-fidelity audio equipment.

We usually find it easiest to associate Fourier analysis with frequencies and times. A mnemonic notation is used in which the data are given at times t_r and analysed in frequencies f.

Shannon has shown[91] that a signal containing only frequencies up to some maximum f_m is completely defined if its value Y_n is given at each of the times on the set nT where

$$2 f_m T = 1 \quad . \tag{C.1}$$

Then for any time t the signal Y(t) is given by

$$y(t) = \sum_{r=-\infty}^{\infty} y_r \frac{\sin 2\pi f(t-rT)}{2\pi f(f-rT)} \quad . \tag{C.2}$$

For a particular t in the range

$$nT < t < (n+1)\, T \tag{C.3}$$

the summation would be performed in the order

$$y(t) = \sum_{k=1}^{\infty} (T_k(t) + T_{\ell-k}(t)) \tag{C.4}$$

where

$$T_k = y_{n+k} \frac{\sin 2\pi f(t-(n+k)T)}{2\pi f(t-(n+k)T)} \quad . \tag{C.5}$$

We see that each term in equation C.2 can be written with an infinite product representation of the sine function and that this is formally a Lagrangian interpolation term. Each term is also the Fourier transform

of a signal that has constant modulus in the range of frequencies

$$-f_m < f < f_m \tag{C.6}$$

and is zero outside that range. The phases determine that the r^{th} term has all components at maximum where

$$t = rT \quad . \tag{C.7}$$

We see that the frequency band is economically used, and also that the denominator $2\pi f(t-rT)$ ensures that the influence of any given data point on the values assigned to $y(t)$ between data points shrinks as we move away from the given data point. In an experiment of finite duration $y_N(t)$ could be given in terms of the values of y_r at the points rT for $r=0,1,\ldots N$.

If the apparatus can be reasonably described, $y_N(t)$ would continue to be a good description of the signal strength at t particularly in the interior of the region, i.e. for

$$t \sim \tfrac{1}{2} NT \quad . \tag{C.8}$$

The restriction to a finite data set however gives us the (mathematical) liberty to assign $y(t)$ any value. This is done without changing the frequency range simply by assigning an appropriate value to y_{N+1}. Thus given the value at t_α for N data points we can arrange an increase of G_α with N+1 data points;

$$y_{N+1}(t_\alpha) - y_N(t_\alpha) = G_\alpha \tag{C.9}$$

if y_{N+1} is defined by

$$y_{N+1} = y_{N+1}(t_{N+1}) = G_\alpha \frac{2\pi f(t_\alpha - (N+1)T)}{\sin 2\pi f(t_\alpha - (N+1)T)} . \tag{C.10}$$

The price paid for the frequency condition is obviously unacceptable if the additional term from C.10 leads to large oscillations in the value of $y_{N+1}(t)$ between t_α and $(N+1)T$.

Now consider the data from less regularly spaced points[92]. We can write down immediately a form for $y(t)$ with a finite set of Fourier components and passing through the points of the given data set.

Assume that we have N+1 data pairs (t_r, y_r) in the region

$$0 < t_r < NT \quad . \tag{C.11}$$

We write

$$Z(t) = \exp(2\pi i\, t/NT) \tag{C.12}$$

$$Z_r = Z(t_r) \quad . \tag{C.13}$$

Then

$$y(t) = R_e \sum_r y_r \prod_{s\neq r} \frac{Z-Z_s}{Z_r-Z_s} \quad . \tag{C.14}$$

The form in equation C.14 is preferable to one involving say

$$x_r = \cos(\pi t_r/NT) \tag{C.15}$$

on the assumption that the data are fairly closely and evenly spaced in t. The effect at a considerable distance from the data point of any term in equation C.14 depends on the denominator. To keep all denominators close to equal so that no one piece of data is over-emphasised we try to keep the smallest magnitude factors of the denominator approximately equal. If the data points were equally spaced in terms of the variable x of C.15, a Lagrangian interpolation involving that variable would naturally be preferred.

In the interior of a very long data set evenly spaced in the variable t, small angle approximations apply in the important factors and terms in the summation, so that for well behaved data, equations C.14 and C.5 both give results indistinguishable from those based on the use of equation C.15. At other points there can be a difference. A special case is the repetition of the form in equation C.14 for y(t) at intervals of NT. This repetition can be removed by introducing an extra factor. A strictly limited extra frequency range can be chosen if y_r in equation C.14 is replaced by

$$y_r \frac{\sin k(t-t_r)}{k\, t_r} \tag{C.16}$$

where the factor k has a magnitude such as π/NT depending on how much extra frequency can be tolerated in return for a rapidly diminishing effect. If there is less sensitivity to a sharp cutoff on frequency, a factor like

$$y_r \exp -\left(\frac{t-t_r}{NT}\right)^2 \tag{C.17}$$

can be used.

Equations C.16 and C.17 give rise to a continuous Fourier spectrum rather than sharp components. The spectrum or components can be shifted

into a range with magnitude from F to F+f by rewriting equation C.14

$$y(t) = R_e[\exp(iFt) \sum_r y_r \prod_{s\neq r} \frac{Z-Z_s}{Z_r-Z_s}] \qquad (C.18)$$

and for a similar transformation of equation C.5 we write

$$T_k = y_k \frac{\cos[2\pi\ (F+\frac{f}{2})(t-kT-nT)]\ \sin[2\pi\frac{f}{2}(t-kT-nT)]}{\pi f(t-kT-nT)} \quad . \qquad (C.19)$$

Since it is well known[14] that if a spectrum contains no frequencies above f_m, it can be described completely in terms of the exact signal values at intervals T given by equation C.1, but gains can be made in practice by sampling it more frequently. Given a region NT the form $y_N(t)$ is incomplete without N+1 parameters. It is better to take readings as often as possible and use them in a least square fit for these N+1 parameters.

APPENDIX D

PROPERTIES OF 'LEAST STRUCTURE' SOLUTIONS

First, for any positive value of γ there is a vector $\vec{s}$ satisfying

$$(\gamma F^T F + Q^T Q)\vec{s} = Q^T \vec{g} \quad . \tag{8.7}$$

Both the product matrices $F^T F$ and $Q^T Q$ are positive definite, i.e. the eigenvalues are greater than or equal to zero. If either is non-singular, then so is the sum. Any vector $\vec{g}$ can be written in terms of eigenvectors of QQ^T

$$\vec{g} = \sum_k d_k \hat{v}_k \tag{D.1}$$

where

$$QQ^T \hat{v}_k = \lambda_k \hat{v}_k \tag{D.2}$$

and

$$Q^T \hat{v}_k = \vec{U}_k = \lambda_k^{\frac{1}{2}} \hat{U}_k \tag{D.3}$$

provided that λ_k is not zero.

The unit vector $\hat{U}_k$ is an eigenvector of $Q^T Q$ associated with the eigenvalue λ_k. Then

$$Q^T \vec{g} = \sum_k \lambda_k^{\frac{1}{2}} d_k \hat{U}_k \quad \text{for } \lambda_k \text{ not zero} \quad . \tag{D.4}$$

In the spectrum vector space of all linear combinations of eigenvectors associated with positive eigenvalues of $F^T F$ or of $Q^T Q$, the matrix $\gamma F^T F + Q^T Q$ has positive diagonal matrix elements for all vectors and hence has an inverse. Application of the inverse to $Q^T \vec{g}$, which is also in the space, gives the required vector $\vec{s}$.

Computation is easier and discussion is simpler, but not changed in essential points, if $\gamma F^T F$ and $Q^T Q$ have the same eigenvectors. In the example used in Appendix B the eigenvectors are shared. The eigenvalues given for the matrix P there, are squared to give the eigenvalues of $Q^T Q$ here.

$$\lambda_o = 1$$

$$\lambda_k = (2n+1)^{-1} \sin(k(2n+1)\pi/N) / \sin(k\pi/N) \quad . \tag{B.5}$$

The eigenvalues of $F^T F$ are

$$\mu_o = 0$$

$$\mu_k = 16 \sin^4 (k\pi/N) \quad . \tag{D.5}$$

The only zero eigenvalue of F^TF is associated with the largest eigenvalue of Q^TQ. If

$$d_k = \hat{v}_k . \vec{g}$$

and a_k is the amplitude of $\vec{v}_k$ in S,

$$(\gamma\mu_k + \lambda_k^2) \quad a_k = \lambda_k \, d_k \quad . \tag{D.6}$$

If λ_k is zero, a_k is also zero. If λ_k is small and μ_k is larger, a_k is also small compared to d_k. Such a vector shows rapid variation between adjacent components and can usually be safely ascribed to noise. It is suppressed whether d_k is small or large, i.e. the ground for suppression is not a statistical test. If the magnitude of $\gamma\mu_k$ is too large some vectors with reasonable values of d_k are missing from the spectrum or are much weaker in the spectrum than the data would justify.

If the magnitude of λ_k is small and d_k is also at the noise level, $\gamma\mu_k$ would be expected to be appreciable. This is not always the case. If $\gamma\mu_k$ is much less than λ_k

$$a_k \approx \lambda^{-1} \, d_k \tag{D.7}$$

and a noisy component appears amplified in the spectrum. The least structure solution does not give as much improvement as would be hoped. It does however contain less noise than appears in the limit as gamma tends to zero.

In the example of Appendix B the limits, as gamma tends to zero, are the solutions quoted as B.6 and B.10 in their respective cases. As mentioned in this appendix actual inversion to produce B.10 would not usually be expected of routines involving single precision arithmetic, and, with the magnitudes used, computer round-off error is appreciable in applying equation B.10. Equation D.6 can be used to illustrate the results of inversion of the matrix $(\gamma F^TF + Q^TQ)$ without such round-off problems. The results of examining cases B.2, B.3 and B.4 for data generated with and without statistical error and for various values of the parameter γ are recorded in Table D.1 using the same notation as in Table B.1. In this table the fit to the data is close to adequate for all values of gamma and all three cases. The difference in S^2 made by adding statistical errors is 1.71 or less. There appears to be a

TABLE D.1

Channels in Circumference		185	185	186	186	187	187
Gamma \ Yields		+*	-	+	-	+	-
32.0	S^2	4.16	2.45	4.19	2.41	4.05	2.45
	C	2.19	2.36	2.59	2.51	2.34	2.37
8.0	S^2	3.78	2.10	3.77	2.06	3.67	2.10
	C	2.55	2.71	2.93	2.86	2.68	2.71
2.0	S^2	3.46	1.80	3.43	1.76	3.30	1.80
	C	2.93	3.10	3.32	3.24	3.07	3.10
0.5	S^2	3.20	1.52	3.15	1.47	2.94	1.52
	C	3.31	3.52	3.70	3.66	3.51	3.52

* Statistical (+)
Bare (-)

satisfactory concentration of the spectrum into the crucial set of channels. Figure 23 reinforces this impression for case B.3, i.e. a total of 186 channels. Figure 24 reveals a different facet with 187 channels. The unfolded spectrum contains a peak at channel 47 but is otherwise a very poor guide to the spectrum. As distinct from the case B.3 where zero eigenvalues lead to complete suppression of the associated noise components, in case B.4 there are small but non-zero eigenvalues and, as indicated in the approximation D.7, the associated noise component is amplified.

Note that the success shown in Figure 23 results from simplicity of the model. Because the eigenvalue is zero the component can be discarded. The usual experimental situation would be more likely to produce a small eigenvalue indistinguishable from zero. The component would be nearly suppressed in $Q^T\vec{g}$ and afterwards enormously amplified in the form found for $\vec{s}$. With noise free data, Figure 25 demonstrates that a satisfactory form of the spectrum can be obtained from the methods associated with equation D.6. Figures 26 and 27 show the residual errors associated with the spectra of Figures 23 and 24. The errors do not indicate that one spectrum is a satisfactory representation of the data and the other is not. There is underlying structure in both sets

of errors, associated with the incomplete fit in both cases even with error-free data. The effect of changing gamma is not obvious from a comparison of Figures 23 and 24. Figure 24 has a large, i.e. restrictive, value for gamma and Figure 23 has a smaller and less restrictive value. In both cases the progression from the high value of gamma to a low value of gamma was seen merely to produce more detail in structure that was already present. 'Ledges' on the edges of peaks in the spectrum developed into smaller subsidiary peaks separated from the main peak by shallow valleys. As expected, the graphical representation of results for case B.2 was found to be completely parallel to that for case B.4.

APPENDIX E

DISTINCTION BETWEEN ONE AND TWO DIMENSIONS IN RESOLUTION FUNCTIONS

In Section 5 from equation 5.22 to 5.47 the case was demonstrated of a resolution function that depended on the difference in energy of the spectrum and detector, but was otherwise independent of either. A matrix equation was obtained in finite difference form

$$y_n = \sum_k S_k \, r_{n-k} \tag{5.36}$$

and an inverse T_k was defined by

$$\Sigma \, T_k \, r_{n-k} = \delta_{n,o} \tag{5.37}$$

with $$T_t = \sum_q a_q \lambda_q^t \quad . \tag{5.39}$$

In two dimensions a similar resolution equation can appear. Each element of the yield vector y_{mn} has a double suffix to indicate a position on the plane.

Then

$$y_{mn} = \sum_{t,u} S_{tu} \, r_{m-t,n-u}$$

or equivalently

$$y_{mn} = \sum_{t,u} S_{t+m,u+n} \, r_{-t,-u} \quad . \tag{E.1}$$

T_{tu} is defined by

$$\sum_{t,u} T_{t+m,n+u} \, r_{-t,-u} = \delta_{m,o} \, \delta_{n,o} \quad . \tag{E.2}$$

It is desirable to make the assumption

$$T_{tu} = \sum_{k,\ell} a_{k\ell} \lambda_k^t \, \mu_\ell^u \tag{E.3}$$

where the value of $a_{k\ell}$ is relevant to a particular quadrant, so that T_{tu} tends to zero as $|t|$ or $|u|$ increases, and with a region close to each axis where two expressions of the form E.3 are equal. We now show in constructing an inverse that the form E.3 is an over-simplification. Suppose that an inverse exists with the required asymptotic behaviour and assume that sums can be defined of the form

$$T_u(x) = \sum_{t=-\infty}^{\infty} T_{tu} \, e^{itx} \tag{E.4}$$

and

$$r_{-u}(-x) = \sum_{t} r_{-t,-u}\, e^{-itx} \tag{E.5}$$

so that

$$\sum_{m} e^{imx} \sum_{t,u} T_{m+t,n+u}\, r_{-t,-u} = \sum_{m} e^{imx}\, \delta_{m,o}\, \delta_{n,o} \tag{E.6}$$

$$= \delta_{n,o} \tag{E.7}$$

$$= \sum_{m,t,u} T_{m+t,n+u} \exp(i(m+t)x)\, r_{-t,-u} \exp(-itx) \tag{E.8}$$

$$= \sum_{u} T_{n+u}(x)\; r_{-u}(-x) = \delta_{n,o} \quad . \tag{E.9}$$

Now the equation E.9 has the same form as that of 5.37 so that equation 5.39 can be modified to read, for this case

$$T_{n+u}(x) = \sum_{q} b_q(x)\; \lambda_q^{n+u}(x) \quad . \tag{E.10}$$

If then we obtain the same set of $\{\lambda_q\}$ independent of the value of x it is obvious that we were justified in writing equation E.3. In turn, the assumption is needed that all the members of $r_{-u}(x)$ have a factor depending on u but not on x, and a factor depending on x but not on u, and hence

$$r_{-t,-u} = r'_{-t} r''_{-u} \quad . \tag{E.11}$$

If equation E.11 holds, we are in fact dealing with two one-dimensional resolution problems that have become associated, rather than an intrinsically two-dimensional problem.

More generally, the functions $b_q(x)\; \lambda_q^{n+u}(x)$ can be separated if they are continuously defined for real x. The requirement is violated if, as x varies some one of the $\lambda_q(x)$ increases from less than unity to greater than unity. The quantity $|b_q(x)\; \lambda_q^{n+u}(x)|$ would change discontinuously at that point. At that particular value of x

$$|\lambda_q(x)| = 1 \tag{E.12}$$

which is the condition leading to complete suppression of the yield for some real component that could be present in the spectrum. Thus if equation E.12 is satisfied there can be no inverse at all. If there is no value satisfying equation E.12 for real x, the required inverse can

in principle be constructed. The inverse has a form similar to equation E.3, but with λ_k and μ_ℓ varying slowly with the powers t and u. Since $T_u(x)$ is a sum of terms $b_q(x)\ \lambda_q^u(x)$ each of which is a continuous periodic function of x, it can therefore be described completely by a countable set of Fourier components. Apart from a constant involving π the t^{th} component can be written

$$a_{qq}\ \mu_q^t\ \lambda_q^u \quad . \tag{E.13}$$

If the factorisation stated in equation E.11 exists, the dissection of equation E.13 is worth pursuing further. For our purpose it is enough to satisfy ourselves that we have the essential steps for the construction of an inverse if it exists at all. Given any equation of the type E.1, we can proceed to the form of equation E.10 and from there to the terms given in the form E.13 which are guaranteed to tend to zero as either $|t|$ or $|u|$ grows without limit, though not necessarily completely in the functional form above.

APPENDIX F

HILBERT MATRIX INVERSION

An extreme form of an intractable resolution function might give the resolution equation:

$$y(E) = \int_{E_L}^{E_u} \frac{s(E')\,dE'}{E+E'-\Delta E} \quad . \tag{F.1}$$

Assuming that y(E) is measured at the points

$$E_n = n\,\Delta E,\ n = 1,\ldots N \tag{F.2}$$

and that

$$s(E') = \frac{1}{\Delta E} \sum_{k=1}^{N} s_k\,\delta(E'-k\Delta E) \tag{F.3}$$

$$y_n = \Sigma \frac{s_k}{k+n-1}\ ,\quad k = 1,\ldots N,\ n = 1,\ldots N\ ; \tag{F.4}$$

thus

$$\vec{y} = H\,\vec{s} \tag{F.5}$$

where H is a Hilbert matrix of order N with

$$H_{kn} = \frac{1}{k+n-1} \quad . \tag{F.6}$$

The matrix is non-singular, but extremely ill conditioned. It is easy to evaluate the determinant from the general case

$$\mathcal{H}_{kn} = \frac{1}{X_k+Y_n+\Delta} \tag{F.7}$$

$$\det|\mathcal{H}| = \prod_{k>\ell}(X_k-X_\ell) \prod_{m>n}(Y_m-Y_n) \Big/ \prod_{k,n}(X_k+Y_n+\Delta) \quad . \tag{F.8}$$

Hence

$$\det|H|_N = (1!2!3!,\ldots(N-1)!)^3/(N!(N+1)!\ldots(2N-1)!) \quad . \tag{F.9}$$

The inverse also follows with the element H^{-1}_{kn} given by

$$H^{-1}_{kn} = (-)^{k+n} \frac{(N+k-1)!\ (N+n-1)!}{(k+n-1)\{(k-1)!\ (n-1)!\}^2 (N-k)!\ (N-n)!} \quad . \tag{F.10}$$

It is easy to establish from equation F.6 that for N > 1 the Hilbert matrix has at least one eigenvalue greater than unity, and with a little manipulation it can be discovered that for each eigenvalue of the Hilbert matrix of order N there is one slightly larger in the matrix of order N + 1. The remaining eigenvalue is of the order 2^{-5} times the

smallest eigenvalue for the matrix of order N. Use of the inverse in the form of equation F.10 leads to evaluations of small differences of large numbers. The number of significant figures to be carried to obtain M significant digits in an answer is approximately

$$M + 1.5\ N \quad . \tag{F.11}$$

Restriction of the spectrum and data to positive values is a considerable constraint here. Burrus[18] has considered the use, for N = 20, of data generated by the spectrum

$$s_k = \delta_{k,1} \quad ; \tag{F.12}$$

hence

$$m_\ell = \frac{1}{\ell} \pm \sigma \tag{F.13}$$

and σ was assumed to be the computer round-off, $\pm 10^{-8}$. As already shown a random error of order 10^{-8} almost certainly incorporates a component of the eigenvector that has the largest eigenvalue in the resolution inverse, $\sim 10^{+30}$. A literal inversion then leads to uncertainties in the spectrum of 10^{+22}, more than swamping the genuine spectrum. The spectrum is required to be positive so that

$$s_1 = 1 - e_1$$

$$s_k = e_k, \quad k \neq 1$$

with values of e_k positive and therefore being required to be small, while e_1 is free to be negative, but the data also restrict it to small values.

Consider first the case of e_1 negative:

$$y_k = \frac{1}{k} + \frac{|e_1|}{k} \quad , \tag{F.14}$$

$$\varepsilon_k = m_k - y_k = \pm \sigma - \frac{|e_1|}{k} \quad , \tag{F.15}$$

$$\chi^2 \sim \frac{1}{\sigma^2} \sum_k \sigma^2 + \frac{e_1^2}{\sigma^2} \sum_k \frac{1}{k^2} \quad . \tag{F.16}$$

The value of χ^2 increases by roughly 20 if e_1 satisfies

$$\frac{e_1^2}{\sigma^2} \times 1.596 = 20 \tag{F.17}$$

$$e_1 \sim -3.6\ \sigma \quad . \tag{F.18}$$

We now allow a positive value of e_1 and of some other (single) e_k. As above, the actual effect of any error of order σ that is present is neglected and the maximum value of e_1 consistent with a χ^2 value of order 20 is obtained.

For a given e_1 and the chosen k there is a best available value of e_k given by

$$e_k \sum_{\ell=1}^{N} \frac{1}{(k+\ell-1)^2} = e_1 \sum_{\ell=1}^{N} \frac{1}{\ell} \frac{1}{(k+\ell-1)} \qquad \text{(F.19)}$$

$$= e_1 \frac{N}{k-1} \sum_{\ell=1}^{k-1} \frac{1}{\ell(\ell+N)} \quad . \qquad \text{(F.20)}$$

Selection of k and the optimal e_k leads to a value of χ^2 varying from $0.0805\ e_1^2/\sigma^2$ for k = 2, up to $0.7036\ e_1^2/\sigma^2$ for k = 20. For k = 2, e_1 can range up to about 15.8 σ with a corresponding value for e_k ($=e_2$) of about 25.0 σ. For k = 20, e_1 can range up to only ~ 5.3 σ but e_{20} reaches 31.3 σ.

With particular values for e_1 and for the consequent e_k, to minimise χ^2 an attempt could be made to improve the value of χ^2 by incorporating a further small quantity e_n. The restriction that e_n shall be positive is now important. It is easy to show that e_n is required to be positive in order to improve the fit in case $1 < n < k$. Outside this range, i.e. if $n > k$, e_n must be given a negative value to improve the fit. It is intuitively obvious that the result depends on the slow variation of matrix elements associated with the ill-conditioned nature of the matrix. All the rows (columns) are alike and each row (column) is very similar to an average of the pair, one on either side of it.

The requirement of a positive spectrum is expected to be less restrictive for data generated from (say)

$$s_\ell = \delta_{\ell,10} \quad . \qquad \text{(F.21)}$$

Again the aim is to vary the spectrum by introducing

$$s_{10} = 1 - e_{10} \qquad \text{(F.22)}$$

$$s_k = e_k \quad K \neq 10 \quad .$$

For e_{10} negative, all other e_k are again even more restricted and $|e_{10}|$ is again limited to about 17 σ where the multiplier is larger, simply because the matrix elements are smaller than for the previous case with e_1.

TABLE F.1

k	s_k	m_k	ITERATION 5 $s_k^{(5)}$	ITERATION 5 $y_k^{(5)}$	ITERATION 7 $s_k^{(7)}$	ITERATION 12 $s_k^{(12)}$	ITERATION 22 $s_k^{(22)}$	ITERATION 40 $s_k^{(40)}$
1	0.0	2.750	0.040	2.750	0.040	0.014	0.0	0.0
2	0.0	2.238	-0.788	2.238	-0.783	0.0	0.0	0.0
3	0.0	1.894	3.721	1.894	3.721	1.366	0.197	10.00
4	10.0	1.646	3.774	1.646	3.774	5.267	8.576	0.0
5	0.0	1.458	2.584	1.458	2.584	3.758	3.058	0.0
6	0.0	1.311	1.363	1.311	1.363	1.192	-2.027	0.0
7	0.0	1.192	0.438	1.192	0.438	-1.010	0.0	0.0
8	0.0	1.094	-0.159	1.094	-0.159	0.0	0.0	0.0
9	0.0	1.012	-0.480	1.012	-0.480	0.0	0.0	0.0
10	0.0	0.942	-0.592	0.942	-0.592	0.0	0.0	0.0
11	0.0	0.881	-0.555	0.881	-0.555	0.0	0.0	0.0
12	0.0	0.828	-0.414	0.828	-0.414	0.0	0.0	0.0
13	0.0	0.781	-0.204	0.781	-0.204	0.0	0.0	0.0
14	0.0	0.740	0.048	0.740	0.048	-1.811	0.0	0.0
15	0.0	0.703	0.324	0.703	0.324	-1.016	0.0	0.0
16	0.0	0.669	0.610	0.669	0.610	-1.856	-0.503	0.0
17	0.0	0.639	0.899	0.639	0.899	0.654	+0.339	0.0
18	0.0	0.611	1.182	0.611	1.182	1.485	1.108	-0.827
19	0.0	0.586	1.456	0.586	1.456	2.292	1.804	1.753
20	5.0	0.563	1.719	0.563	1.719	3.069	2.43	4.072

We now use e_{10} and minimise χ^2 by use of e_9 and e_{11}; χ^2 can be multiplied by a factor of 6.58×10^{-5} or e_{10} can be tolerated up to 2.06×10^3 σ. The optimum values of e_9 and e_{11} are respectively $0.433\ e_{10}$ and $0.572\ e_{10}$, i.e. both of order 10^3 σ.

With more than one k such that s_k is positive the complications of correlated errors are bound to occur. The uncertainties can become large if the actual S_K can be used to hide the negative components of an eigenvector corresponding to a small eigenvalue. The effect on the fit to data is small. A program was written to extract a positive spectrum from the data as shown in Table F.1. The program used several iterations of the conjugate gradient method to obtain an acceptably low value for χ^2. Any negative values of the spectrum were then set to zero and iterations restarted.

The positive elements were used as the first vector of a subsequent sequence of conjugate gradient iterations. If an element was negative at two successive checks it was set at zero permanently. The program was also successful in locating positive components if the spectrum points were more closely spaced than the data points.

The Hilbert resolution problem outlined here is considerably more stringent than any expected to arise in experiment. It is therefore mostly useful in testing techniques.

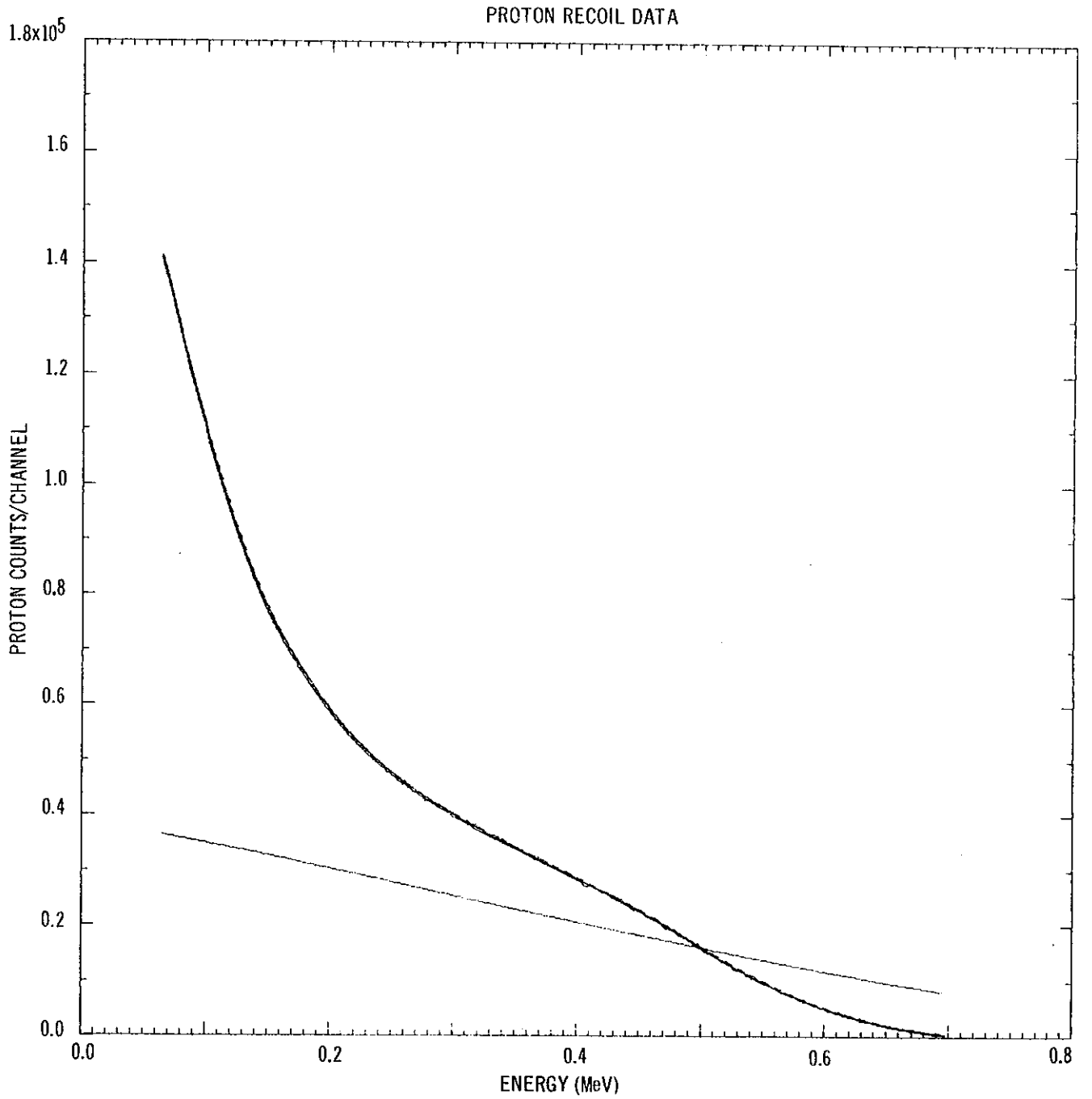

FIGURE 1 DATA FROM A PROTON RECOIL COUNTER MEASURING A NEUTRON FLUX

The spectrum runs to higher energies than the data shown. A correction has been made to remove the effects of the known higher energy neutron spectrum. The amount subtracted at each energy is indicated by the sloping, approximately straight line. The data to be fitted are plotted together with the reconstruction of the data and curves giving the standard deviation from the reconstructed data. The last four curves described are close to superposed.

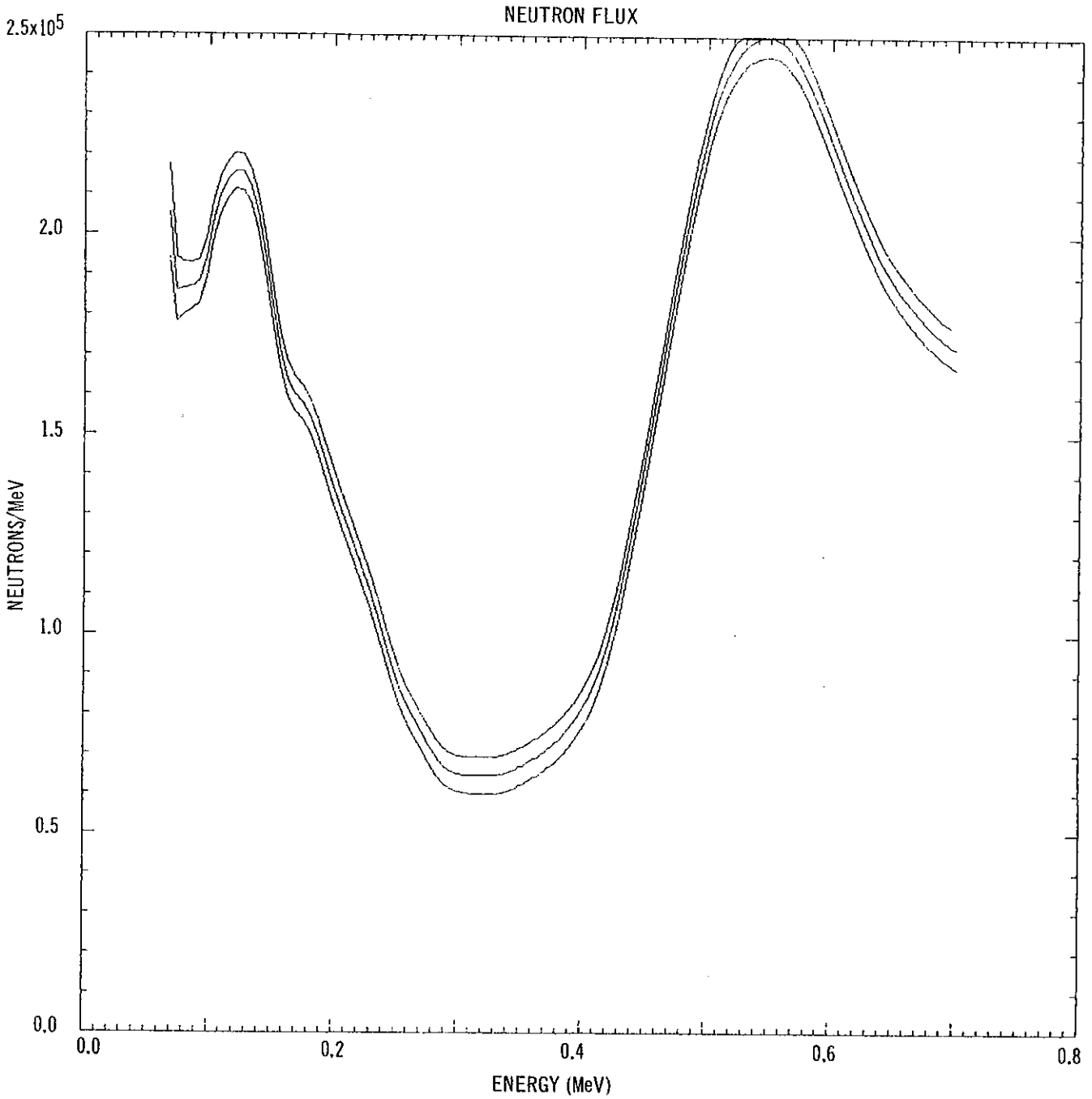

FIGURE 2 THE NEUTRON SPECTRUM DERIVED FROM THE DATA IN FIGURE 1

Also plotted are two curves corresponding to error bars as discussed in Section 6. Since these affect fewer channels for the low energy spectrum they are longer there, i.e. the curves diverge.

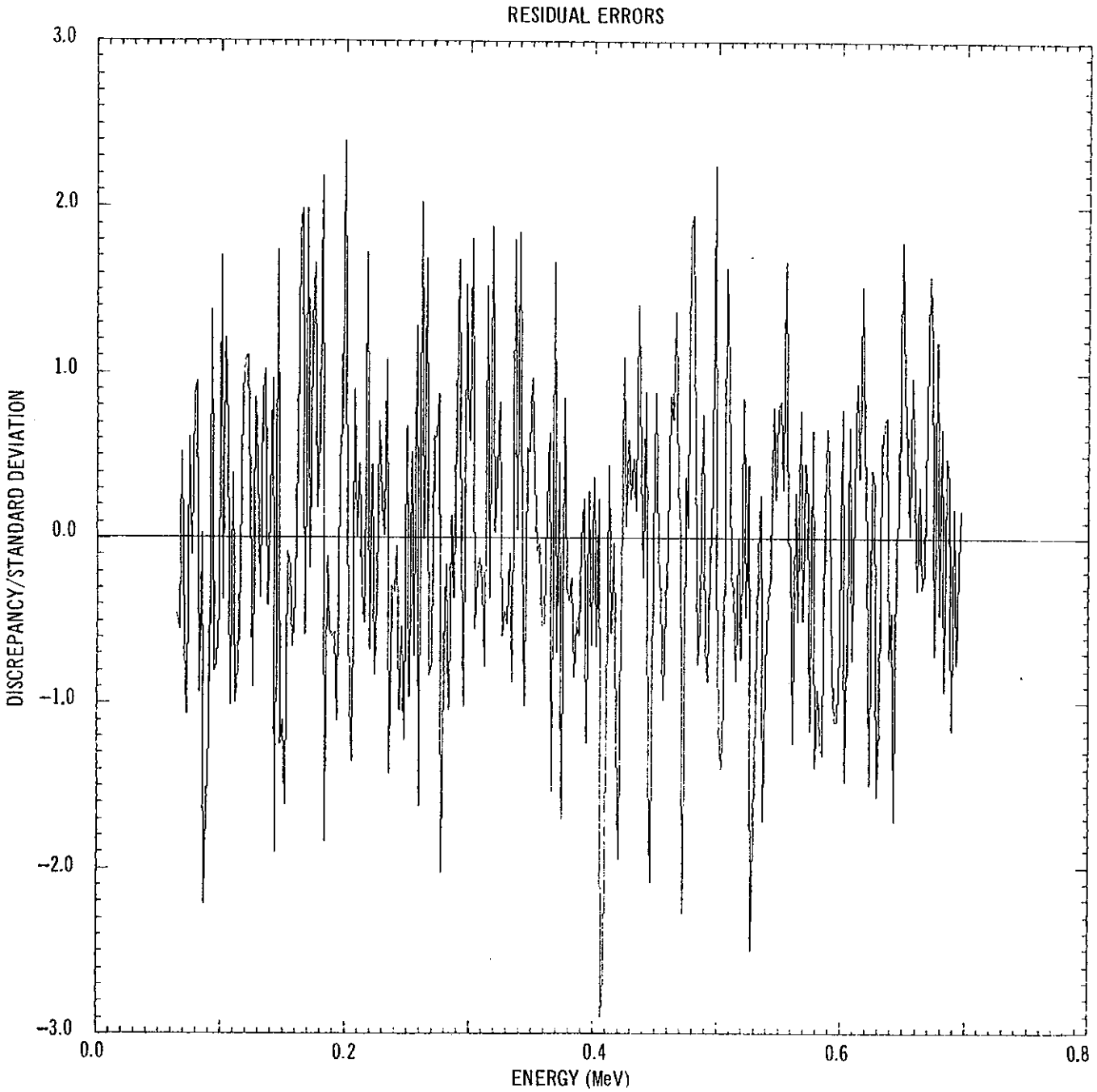

FIGURE 3 RESIDUAL DISCREPANCIES IN THE FIT TO DATA SHOWN IN FIGURE 1

The fit is uniformly good over the whole range of data and the total residual χ^2 (301) is less than the number (309) of degrees of freedom.

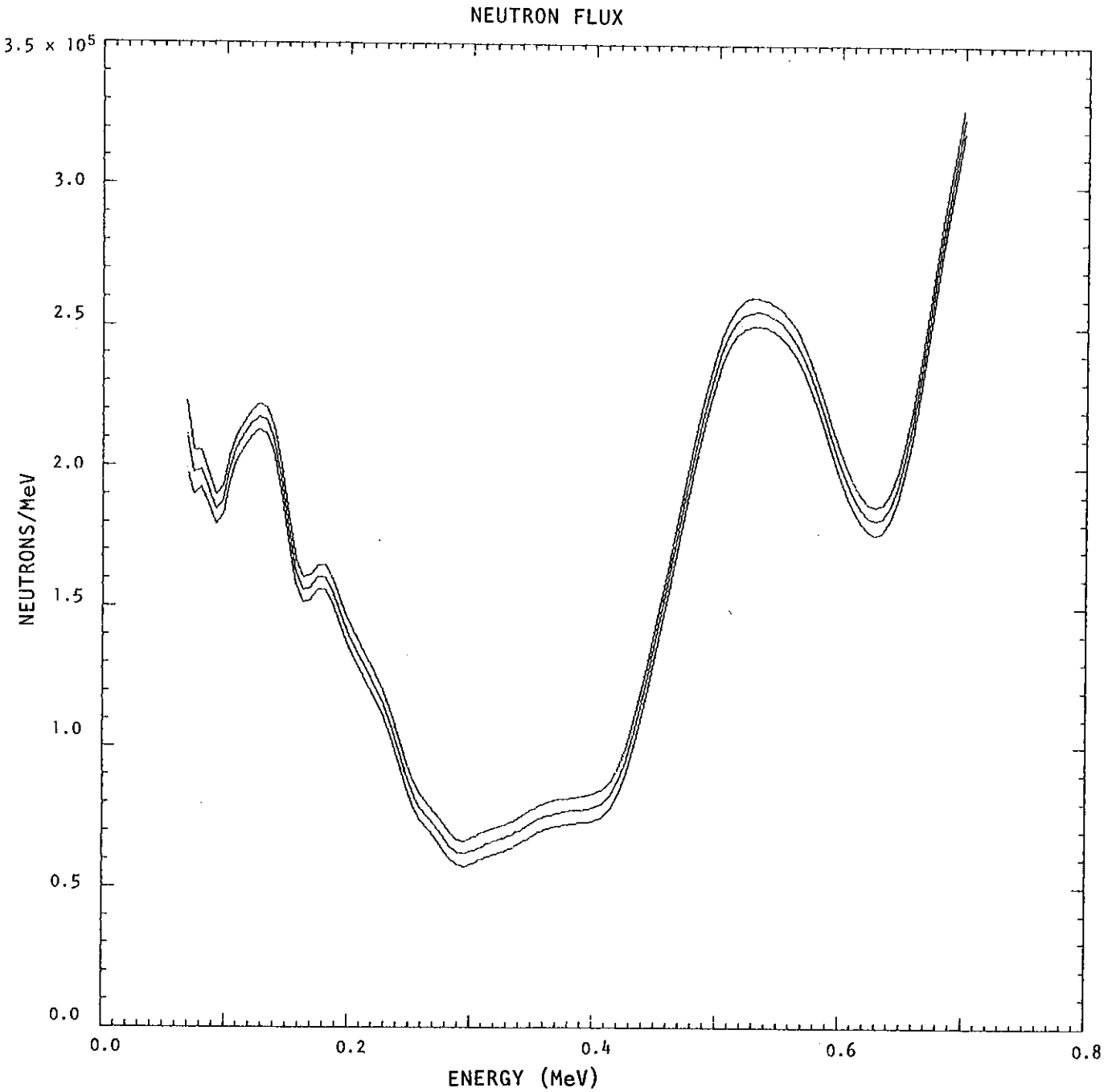

FIGURE 4 SPECTRUM AS IN FIGURE 2 FROM THE SAME DATA AS IN FIGURE 1 BUT WITH AN UNDERESTIMATE OF THE HIGH ENERGY SPECTRUM

Note that the fitting program is giving the same spectrum at low energies, but trying to add at high energies to compensate for the spectrum that was not supplied.

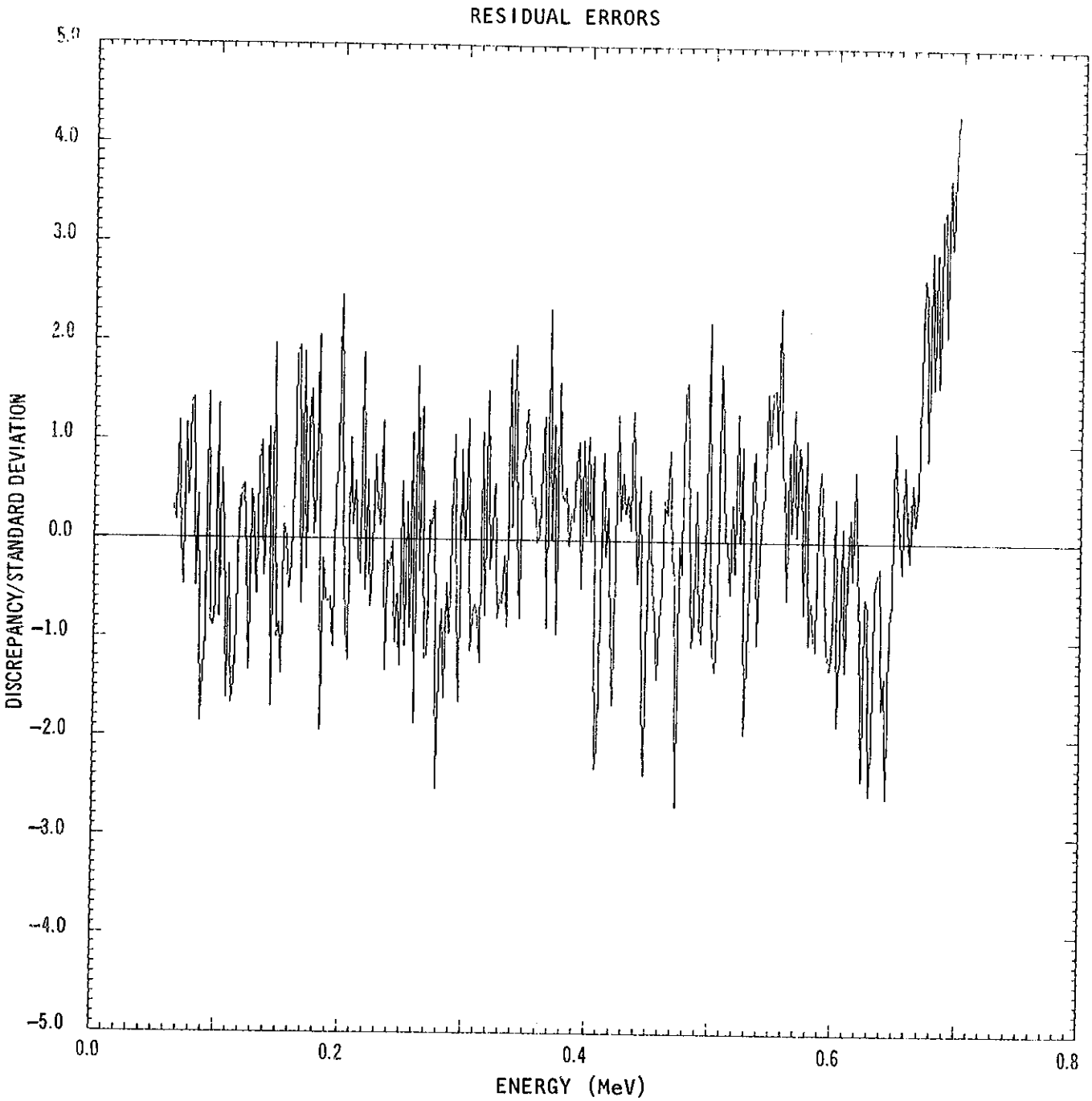

FIGURE 5 RESIDUAL DISCREPANCIES ASSOCIATED WITH THE SPECTRUM IN FIGURE 4

Note that even after 12 iterations, compared with ten for the fit demonstrated in Figure 3, the χ^2 value (439) is too large; and difficulty in establishing a fit at high energies leads to 'ringing' of errors at lower energies.

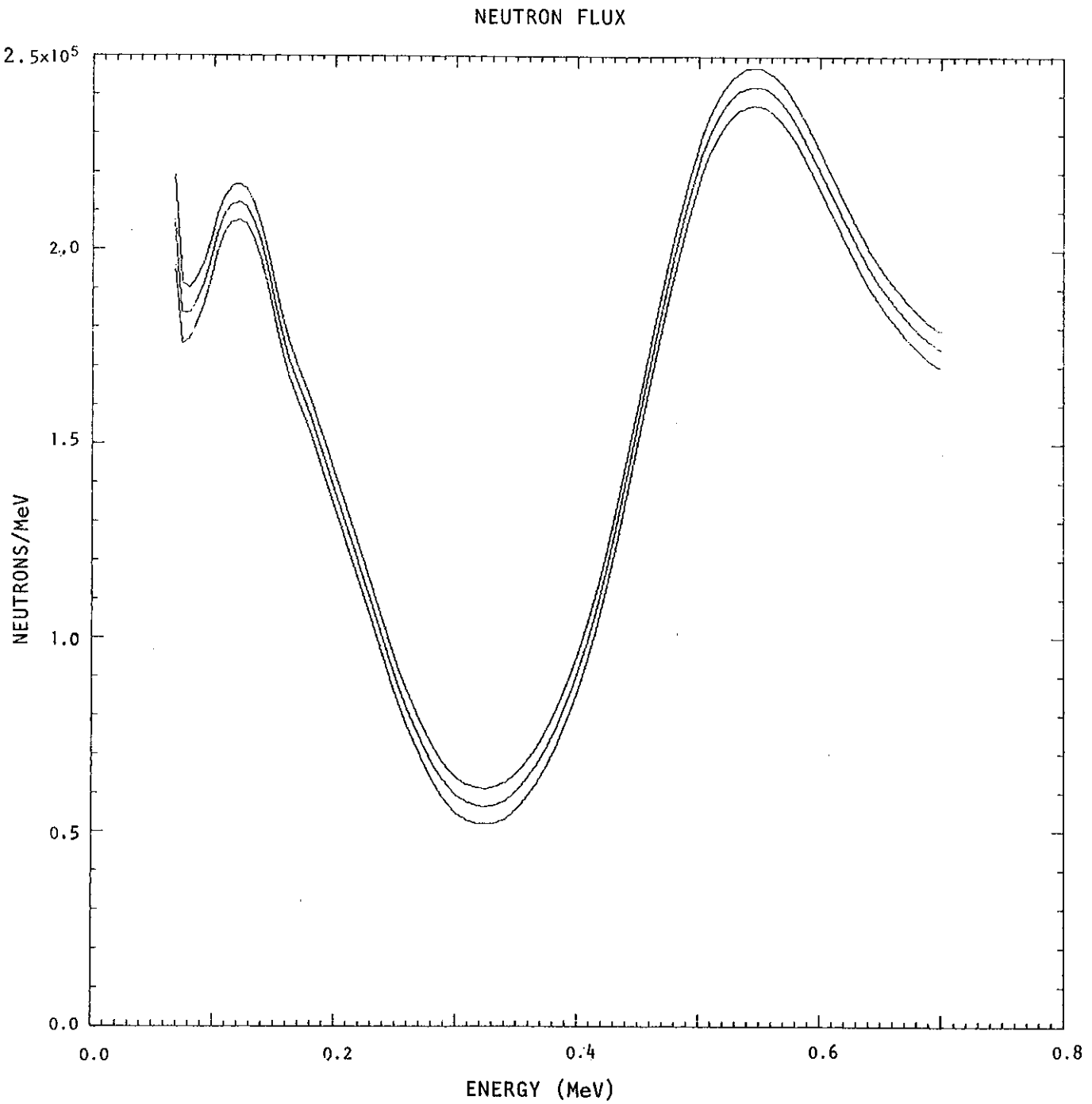

FIGURE 6 THE NEUTRON SPECTRUM OBTAINED AFTER SIX ITERATIONS FOR THE DATA OF FIGURE 1

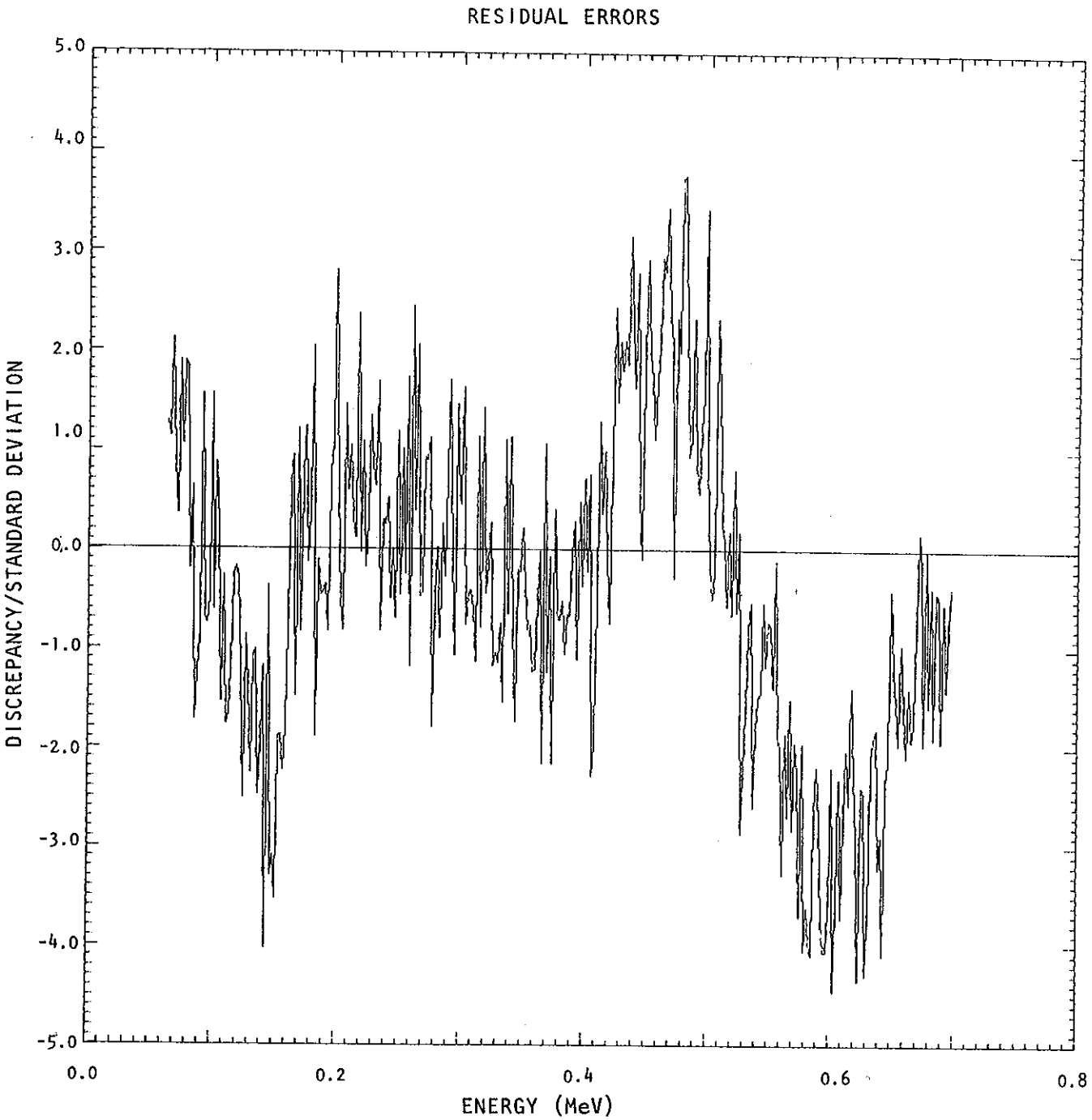

FIGURE 7 THE DISCREPANCIES OCCURRING IN THE FIT TO DATA ASSOCIATED WITH THE SPECTRUM OF FIGURE 6

The χ^2 value is 967. Note that there is 'ringing' as in Figure 5 but more severe; this phenomenon should be investigated wherever it occurs.

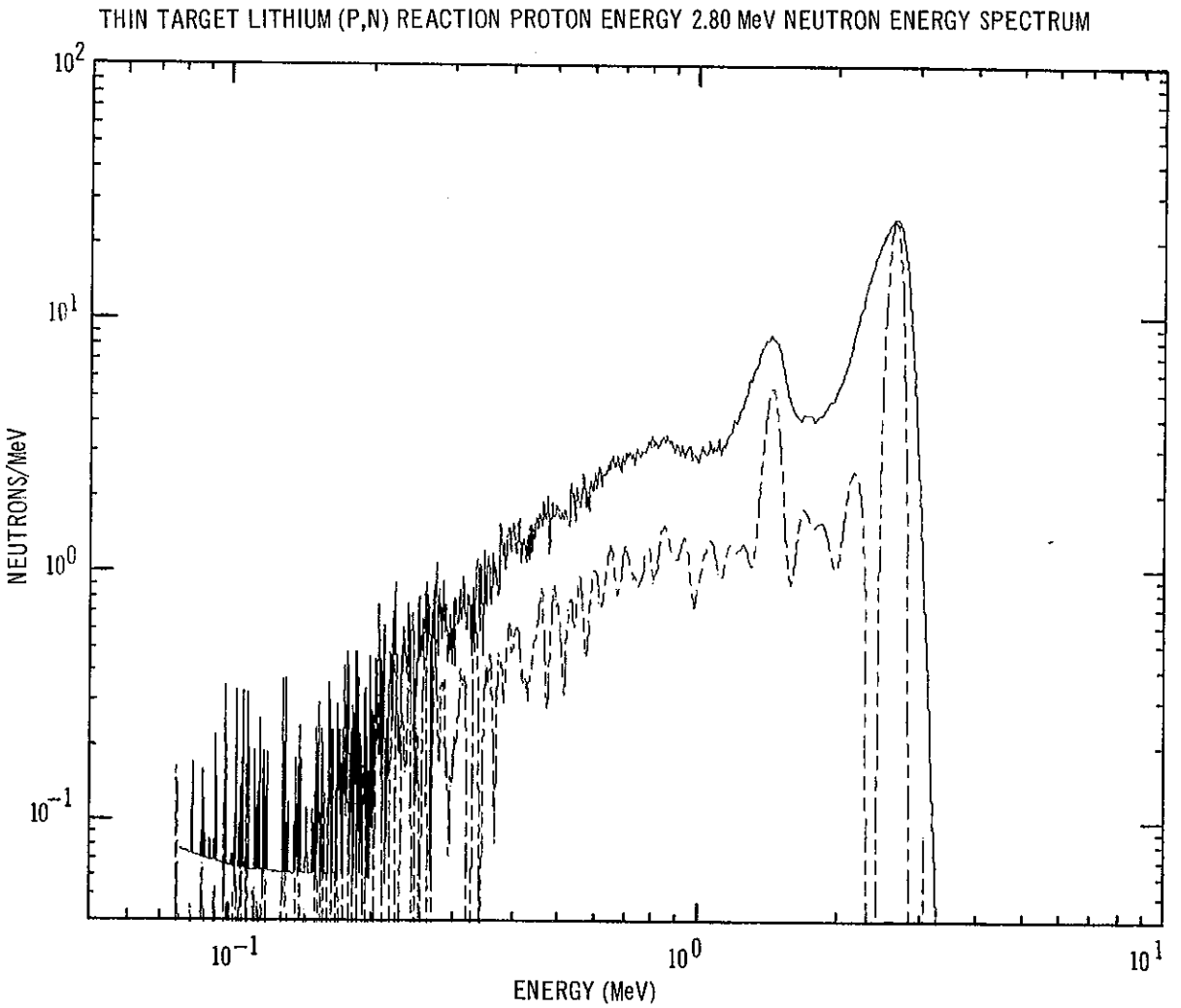

FIGURE 8 NEUTRON ENERGY SPECTRUM OBTAINED WITH PULSED TIME OF FLIGHT WHEN A THIN LITHIUM TARGET WAS BOMBARDED WITH 2.80 MeV PROTONS

The raw data and unfolded spectrum are normalised to have the same maximum. In the shared broad features the spectrum is more sharply defined. The data at lower energy are obviously noise. The strength of a limited unfolding program is that it does not try to make an exact spectrum here. The energy width of time channels shrinks as the energy is decreased. The simplicity of the resolution matrix given by Equation 7.9, in which one set of elements is repeated in many rows, is lost if unfolding is attempted with data expressed as counts/unit energy instead of as counts per channel. As indicated in connection with Equation 5.20, it is worth making sure that the resolution problem has been given its simplest formulation.

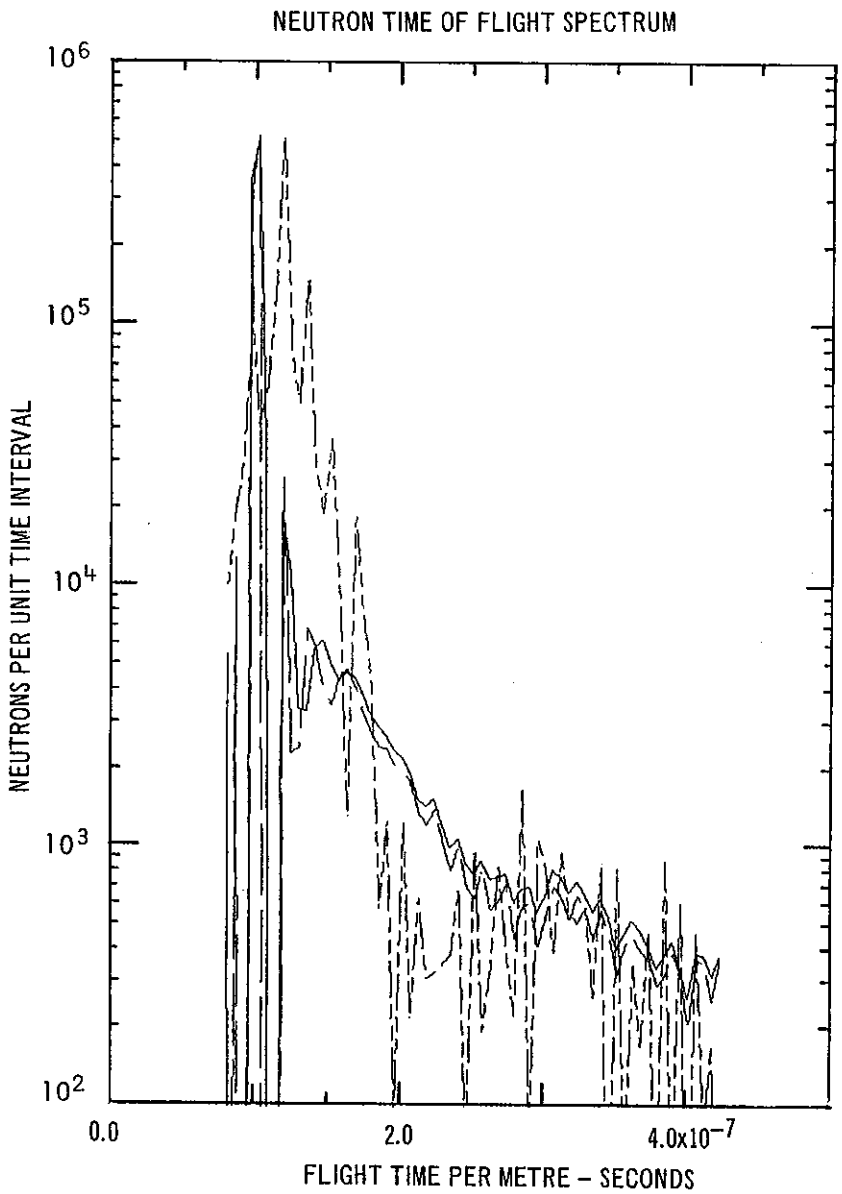

FIGURE 9 DATA AND SPECTRUM FROM A LITHIUM (p,n) EXPERIMENT WITH 2.24 MeV PROTONS

Two spectra are plotted and are close to superposed. The value of χ^2 between them has changed by a factor of two. The change appears to be all concerned with structure narrower than one time-of-flight channel. The spectra are averaged over one time-of-flight channel before plotting (and the data are displaced by five time-of-flight channels).

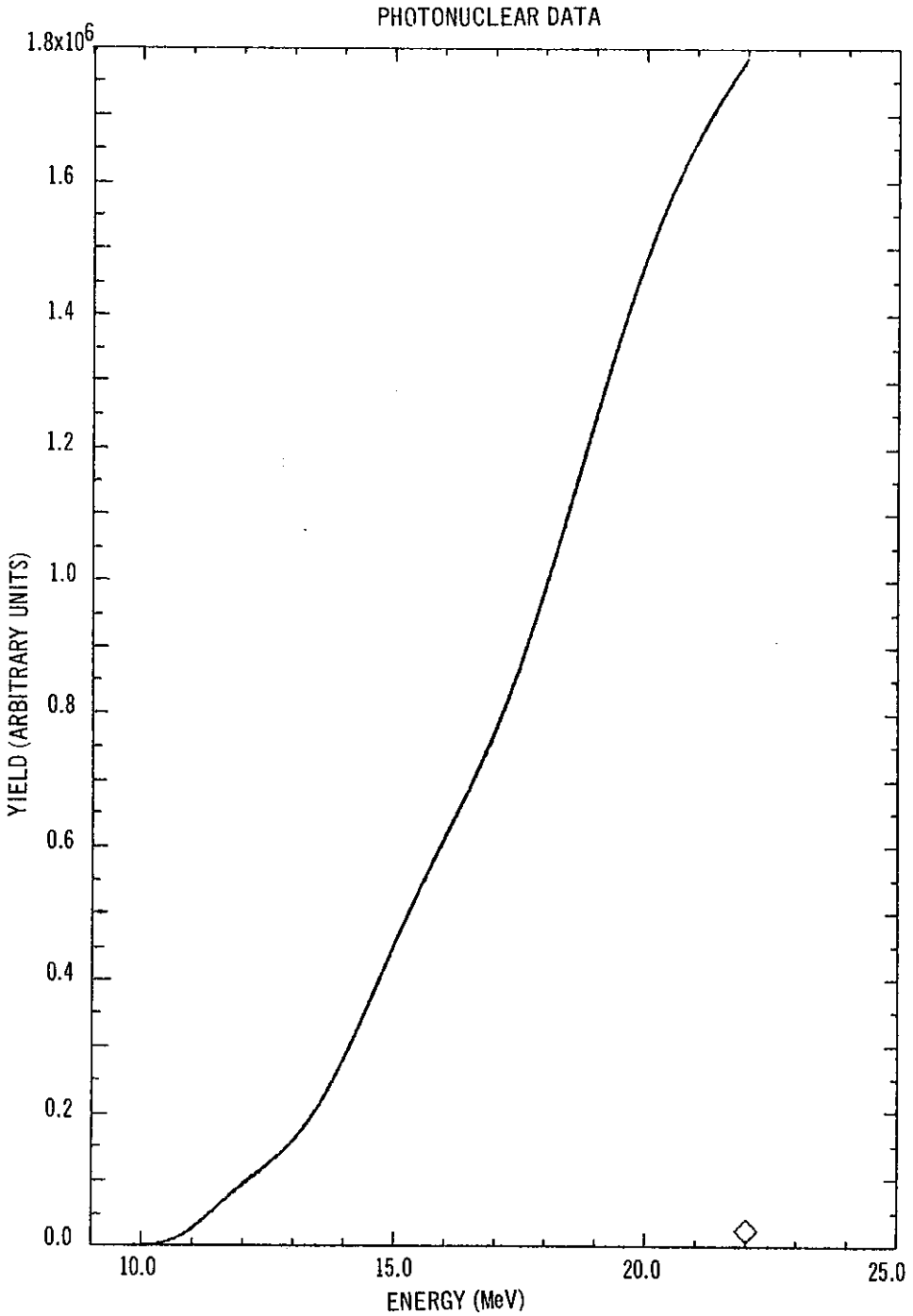

FIGURE 10 GEDUNKEN DATA FROM A PHOTONUCLEAR YIELD EXPERIMENT, SUPPLIED BY DR H.H. THEIS [75]

The reconstructed data are shown on the same scale and appear superposed. The single point at 22.0 MeV is a marker that the fit was achieved with a negative cross section value at this energy.

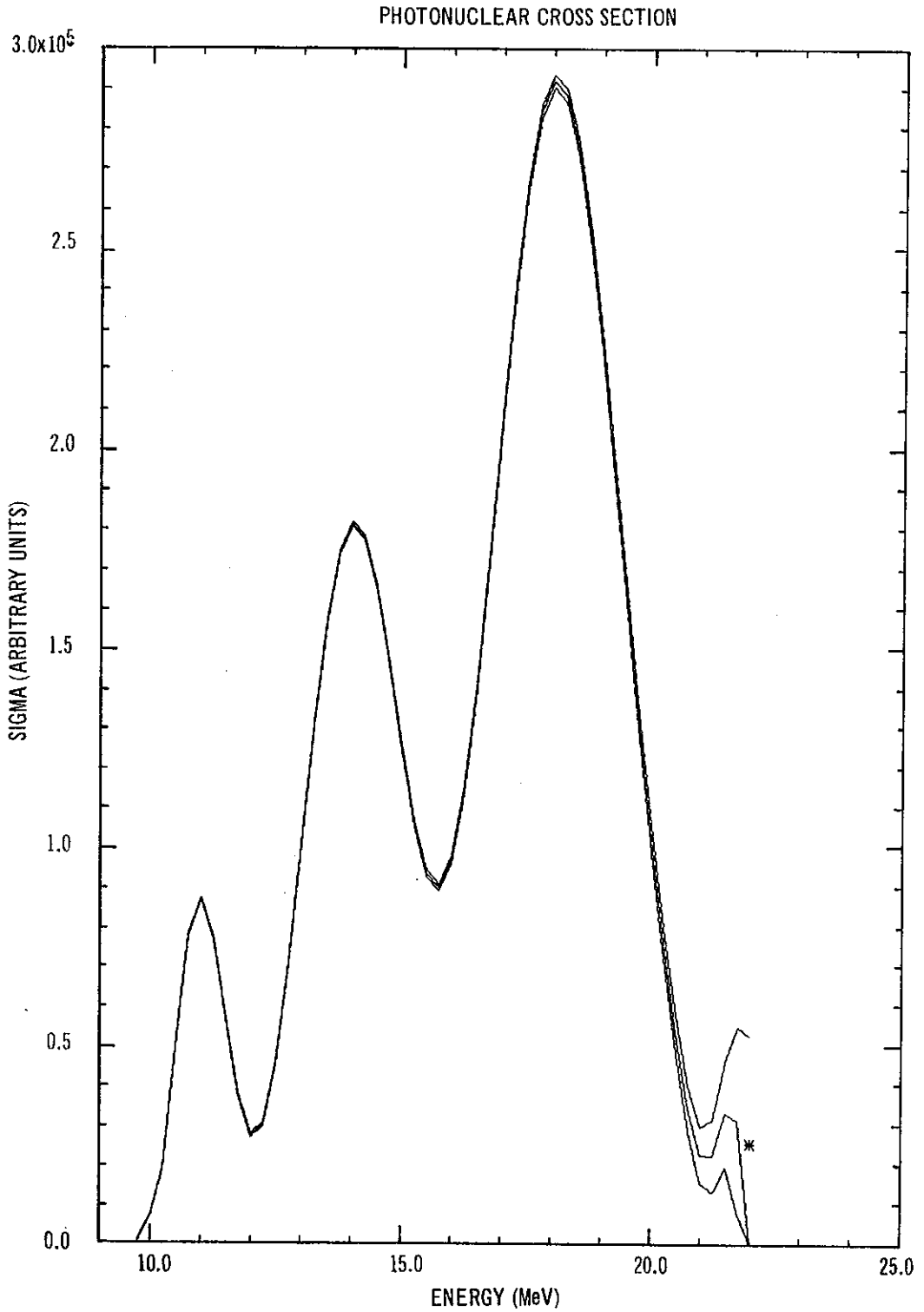

FIGURE 11 THE CROSS SECTION THAT ACHIEVED THE FIT TO THE DATA SHOWN IN FIGURE 10

The bracketing lines are placed one standard deviation away from the spectrum line (for moving one spectrum cross section value). The isolated point at 22.0 MeV is the reflection in the energy axis of the single negative cross section in the spectrum.

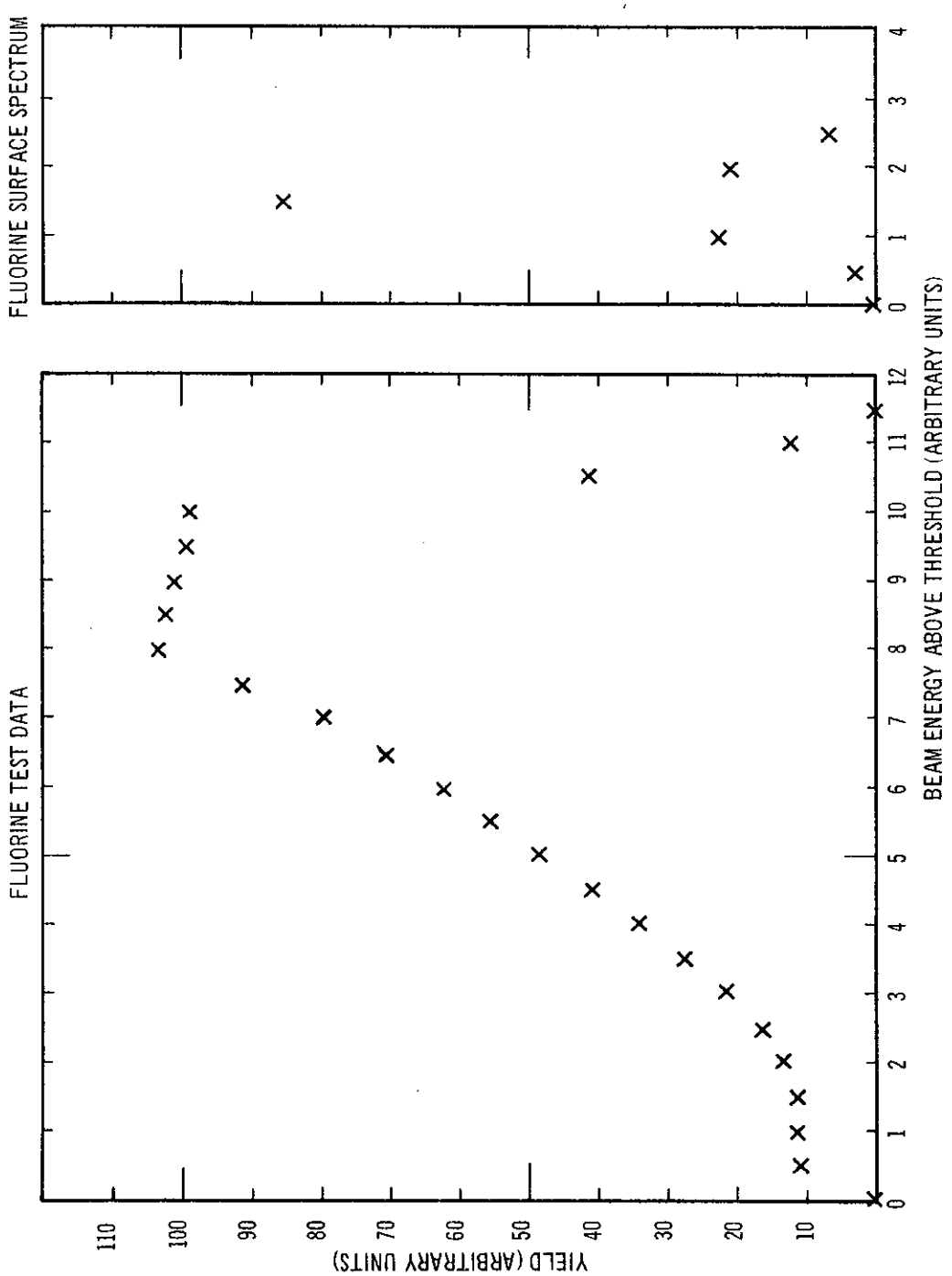

FIGURE 12 (a) TYPICAL SET OF YIELD MEASUREMENTS USED IN TESTING THE FLUORINE PROFILE UNFOLDING METHOD
(b) ASSOCIATED PROFILE FOR SURFACE CROSS SECTION, i.e. THIN FLUORINE TARGET

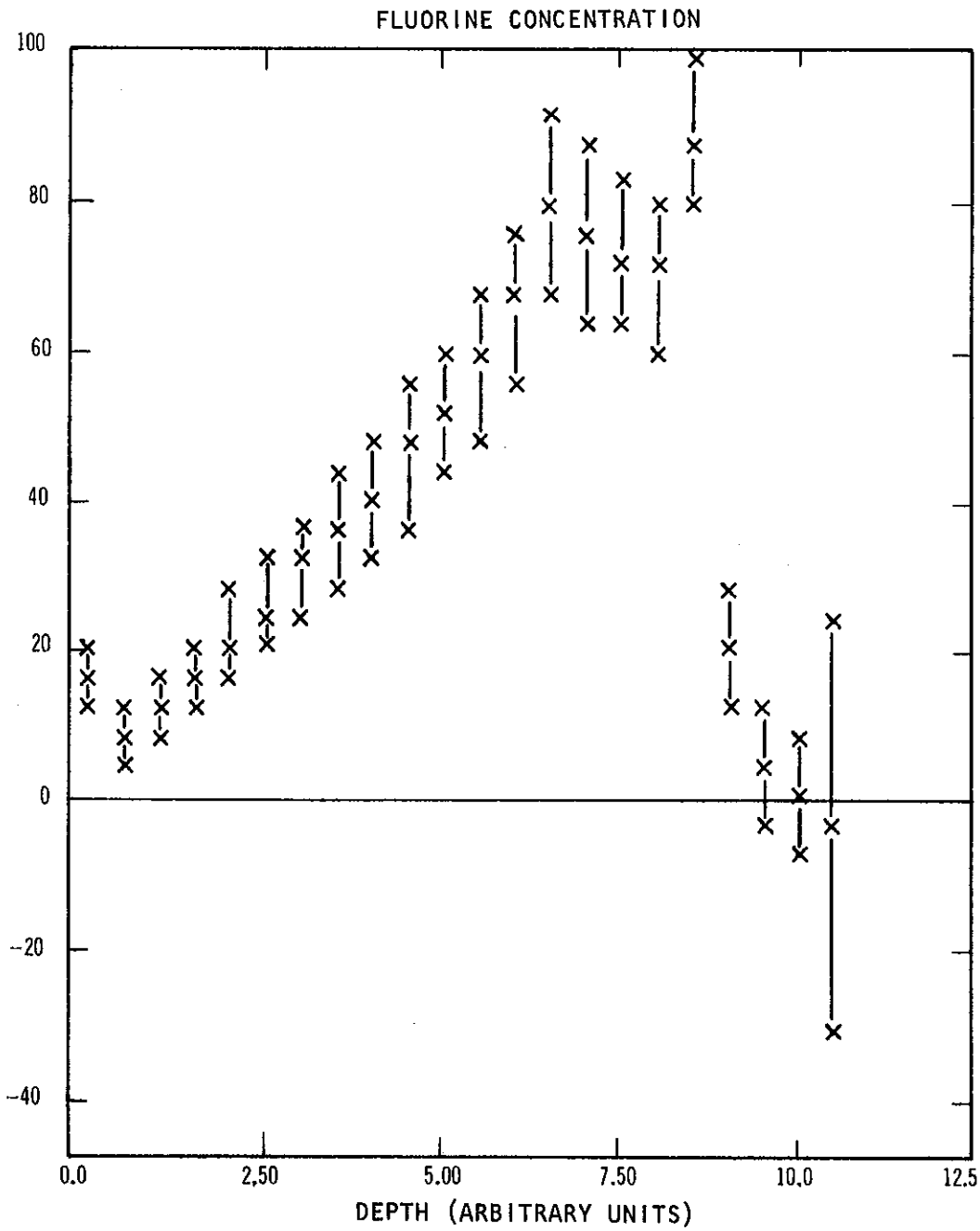

FIGURE 13 THE UNFOLDED FLUORINE PROFILE ASSOCIATED WITH FIGURE 12

The error bars are derived from the counting statistics in Figure 12 and are of the type considered in Section 6.

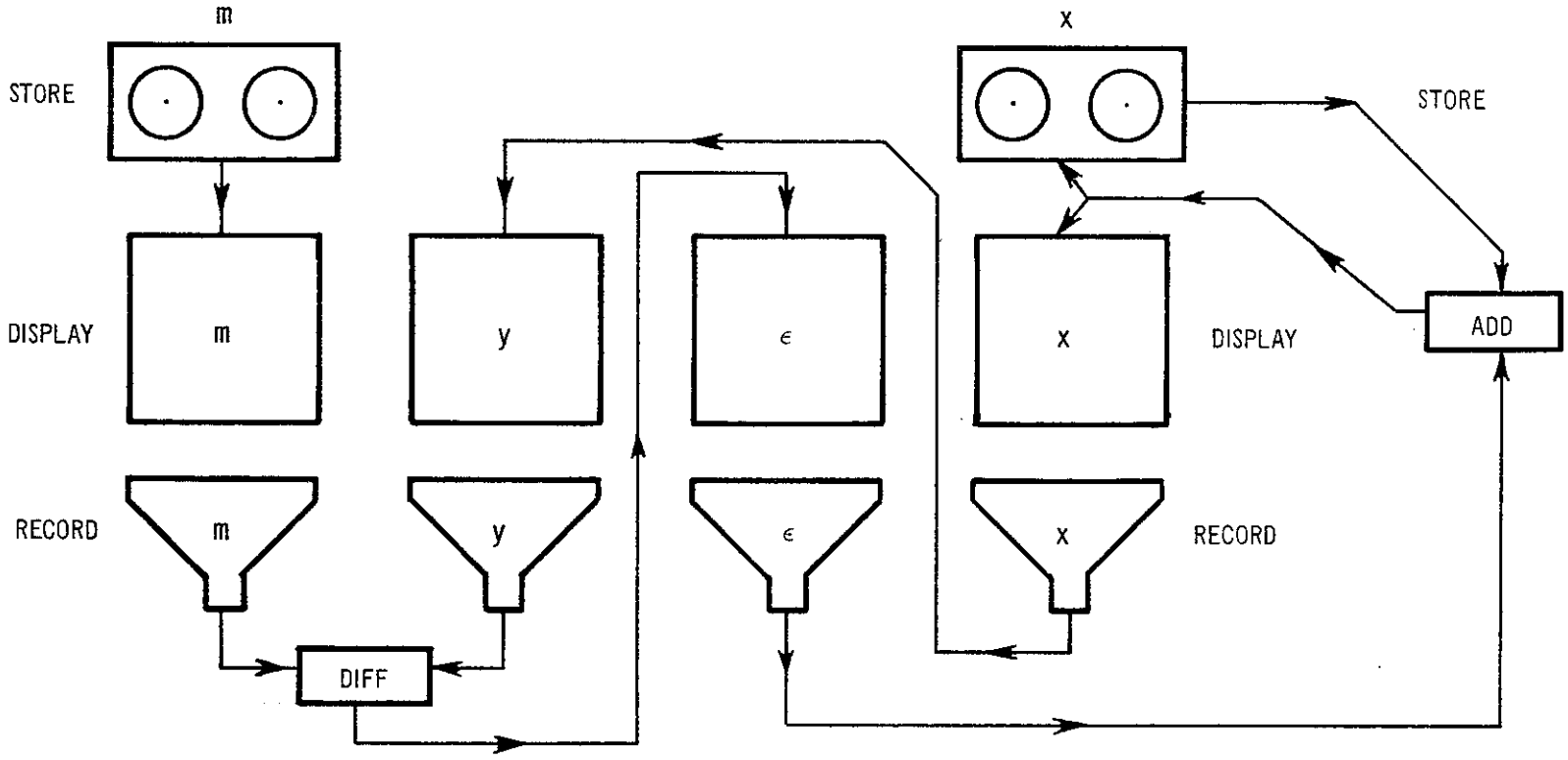

FIGURE 14 A SCHEMATIC OUTLINE OF AN ANALOG SYSTEM TO UNFOLD TELEMETERED TELEVISION DATA BY THE SERIES METHOD

The data are stored as a video signal tape at m and the unfolded signal spectrum at x. The four squares m, y, ε and x represent television screens on which are displayed the data signal, the current approximation to data signal, their difference and the current approximation to the spectrum signal. Each is viewed by a TV camera with the optical system used to generate the original signal data. The signal from x is used to generate y. The difference between m and y is used to generate ε, presumably with a suitable background to make it positive, and the signal from ε without the background is added to the store at x for the next iteration.

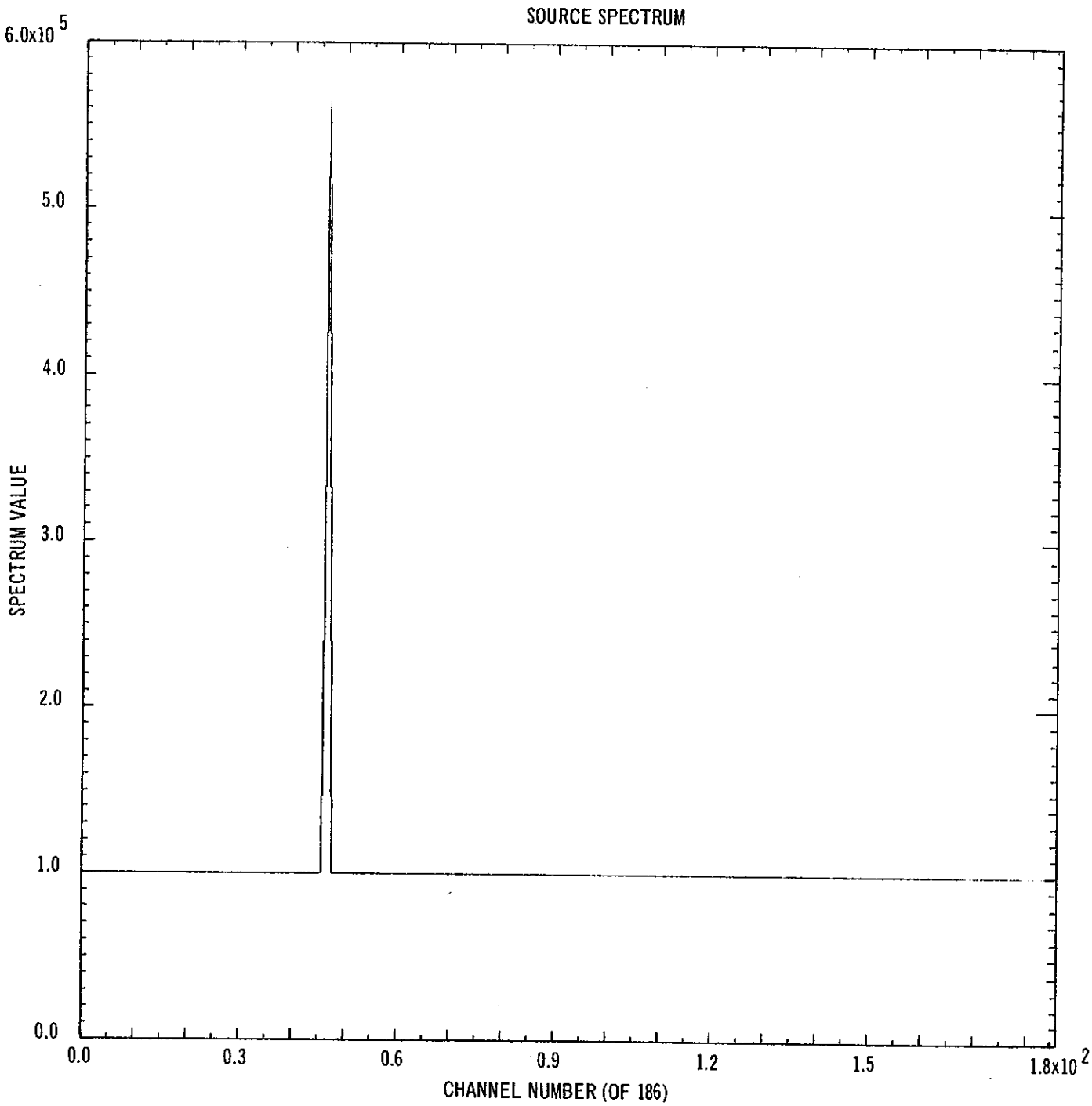

FIGURE 15 THE SOURCE SPECTRUM FOR THE SIMPLE EXAMPLE DEVELOPED IN APPENDIX B

The figure is for case B.3 with 186 channels total. Since the resolution here covers exactly half the circle, the yields would be unchanged if the background is set at 1.05×10^5 instead of 1.00×10^5, and a downward spike of total strength 4.65×10^5 is introduced in channel 140 in place of the upward spike seen here in channel 47.

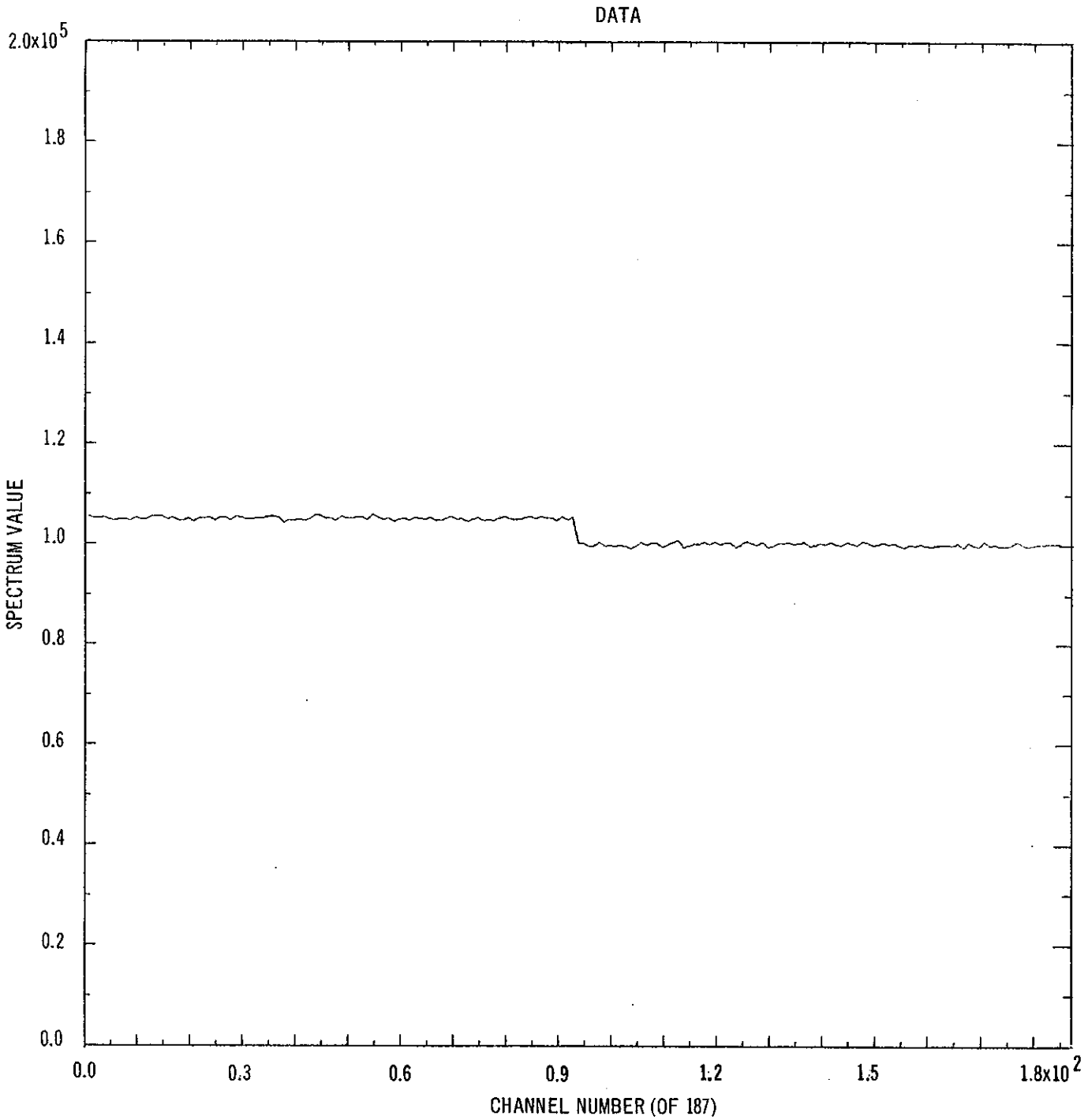

FIGURE 16 THE DATA USED IN THE SIMPLE EXAMPLE OF APPENDIX B IN THE CASE OF B.4 WITH 187 CHANNELS

Each yield point was assumed to result from a counting process, and Poisson statistical counts were obtained. Thus for channels one to 93 the value of σ is 324 and for the remainder of the channels it is 316. With the given number of channels, the measured value of σ for a single data set would not distinguish a significant difference, and for simplicity a single value of σ for all channels is assumed.

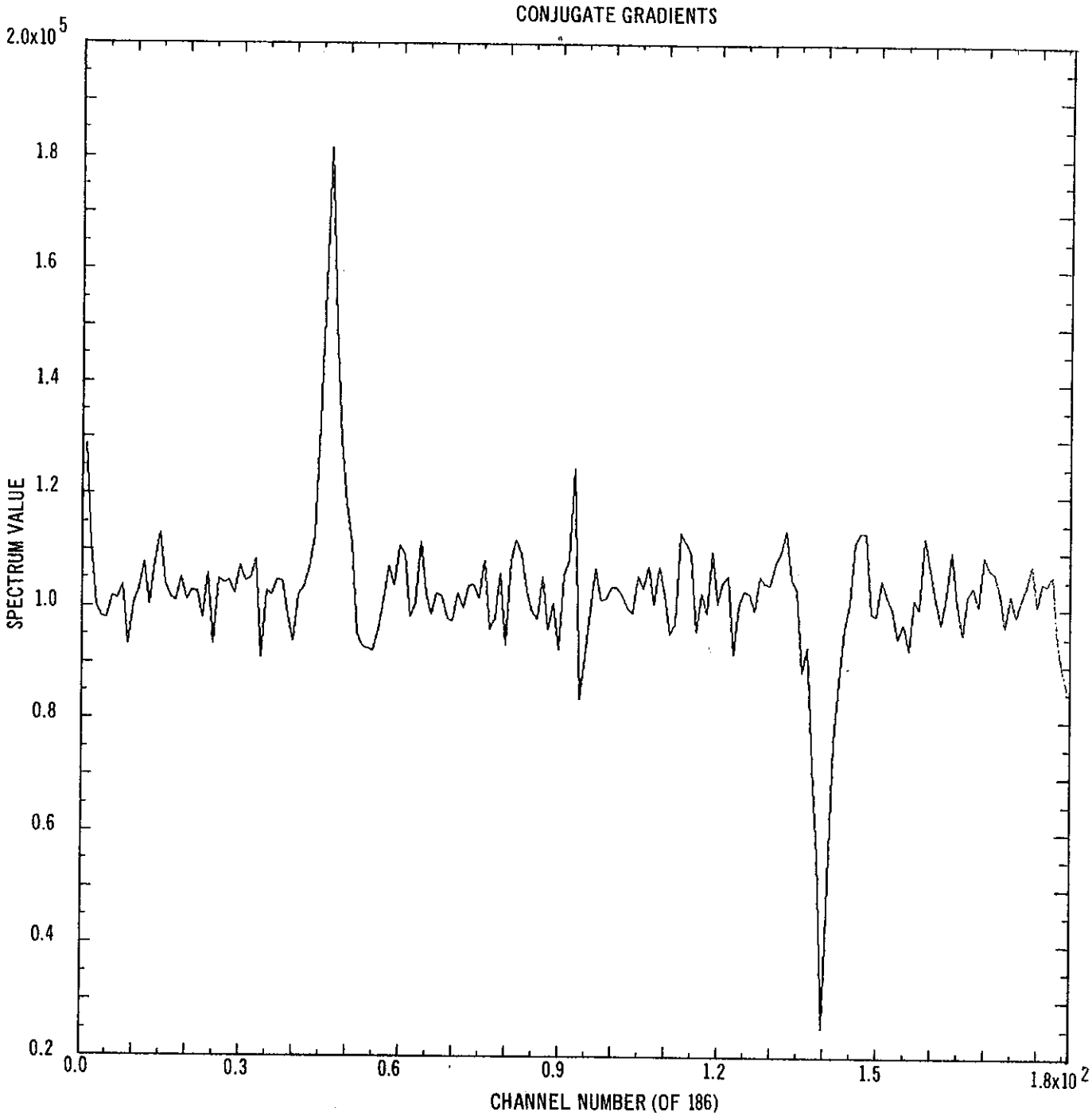

FIGURE 17 THE SPECTRUM OBTAINED AFTER 10 ITERATIONS OF A CONJUGATE GRADIENT TECHNIQUE USING 186 CHANNELS TOTAL, i.e. FOR THE MODEL B.3 WITH RANDOM STATISTICAL ERRORS

At this point S^2 in Table B.1 is 3.25 as compared to 1.48 for error free data. The residual discrepancies therefore include a portion that strongly resists unfolding techniques together with about 1.77 of 'noise'. Again from Table B.1 the amount of strength concentrated, c, is 3.61, as opposed to 3.71 for noise-free data. The statistical errors are not interfering with unfolding the main features of the spectrum.

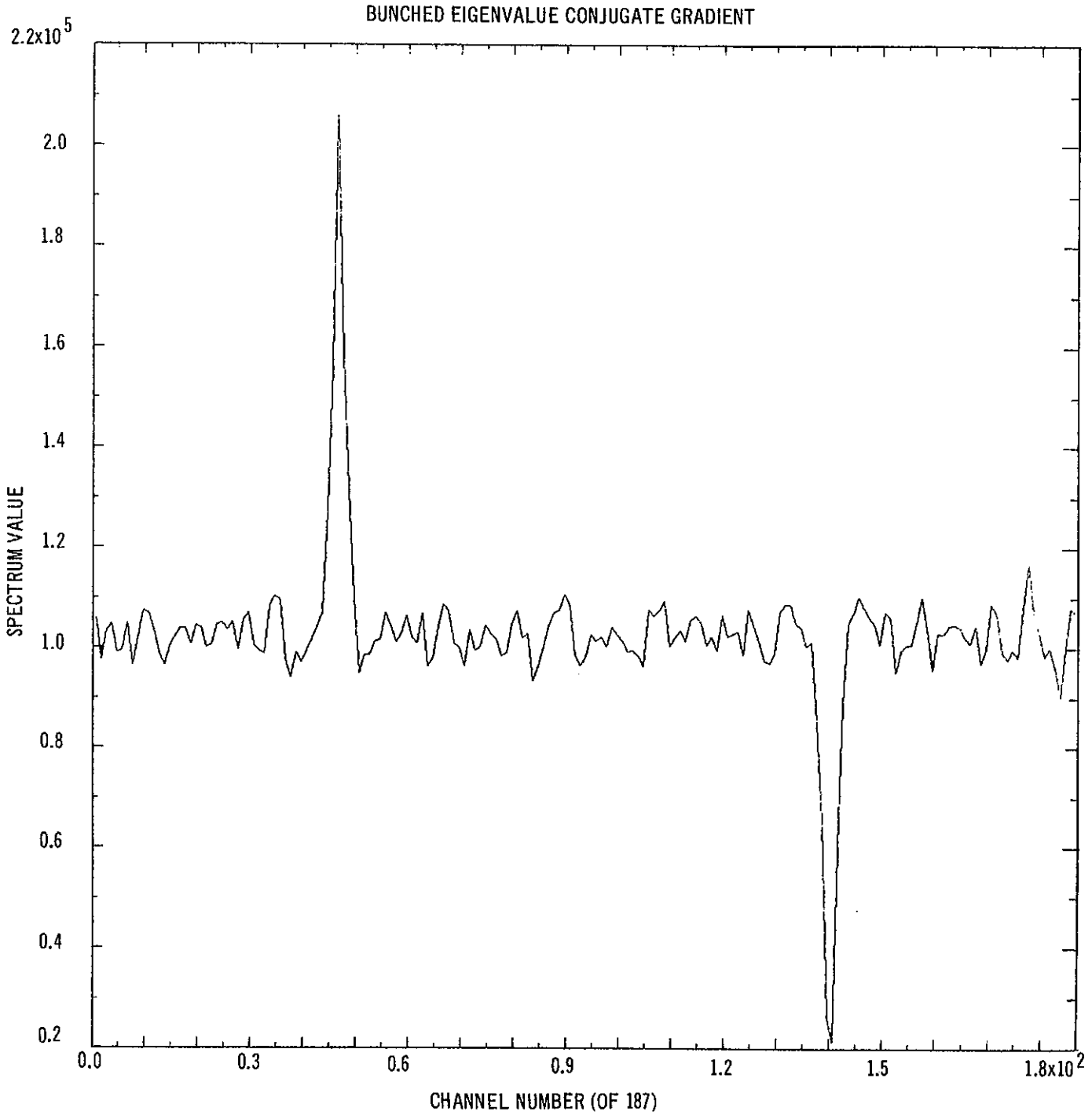

FIGURE 18 THE SPECTRUM OBTAINED AFTER 10 ITERATIONS WITH THE CONJUGATE GRADIENT TECHNIQUE PROGRAMMED AS IF THE RESOLUTION MATRIX WERE NON-SYMMETRIC

The method ensures that the matrix used has all eigenvalues positive. The convergence is slightly more rapid than in the simpler version illustrated in Figure 17. Again the value of S^2 is less than the statistical noise added to the value for S^2 achieved with 10 iterations on noise-free data. The strength of the spike is actually more concentrated than it was with noise-free data.

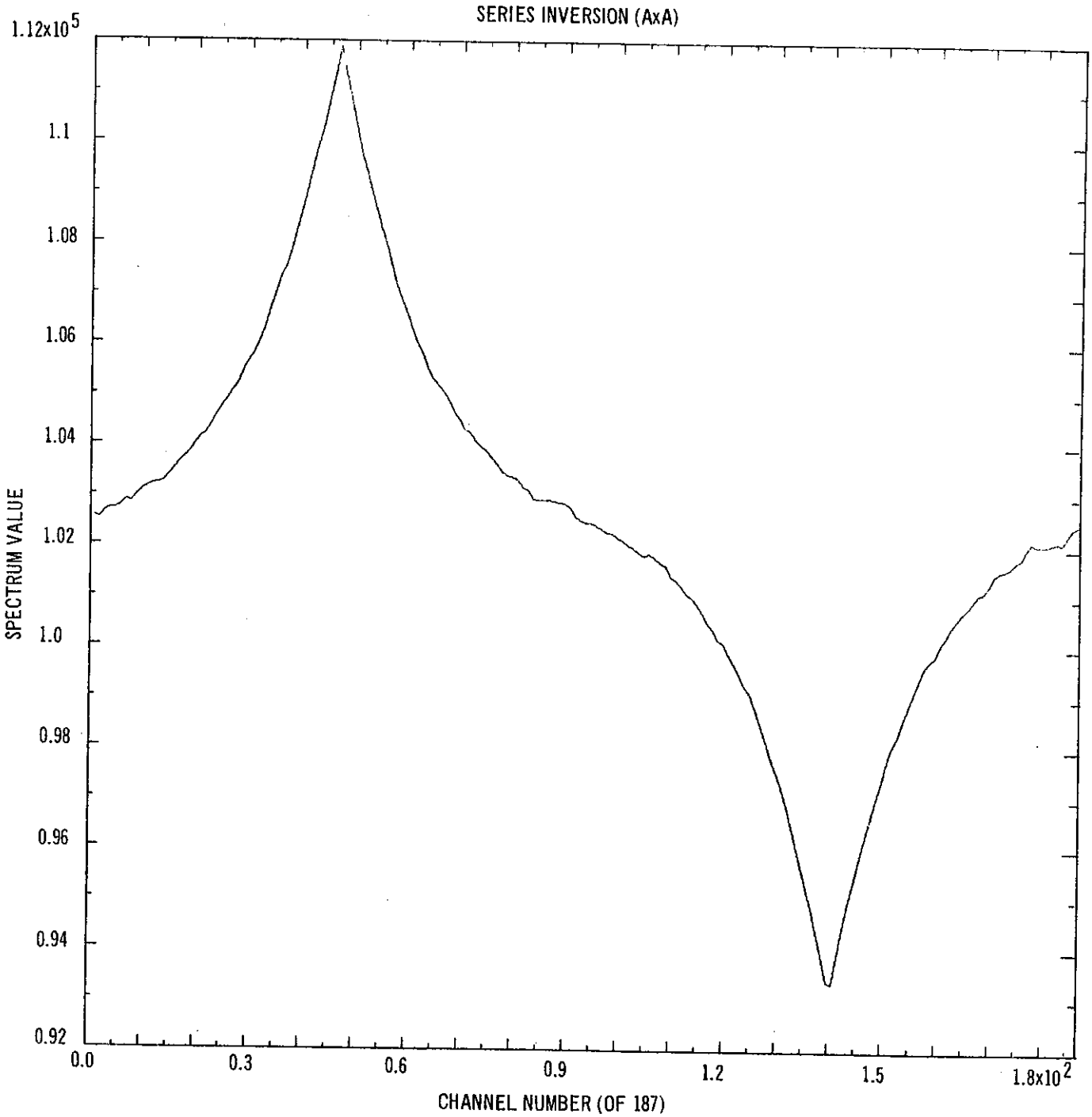

FIGURE 19 UNFOLDING OF CASE B.4 WITH STATISTICAL ERRORS USING THE SERIES METHOD ADAPTED TO ENSURE POSITIVE EIGENVALUES

The fit to the data is still not good but the main feature of the spectrum has been sharpened so that it is concentrated in about a third of the resolution width.

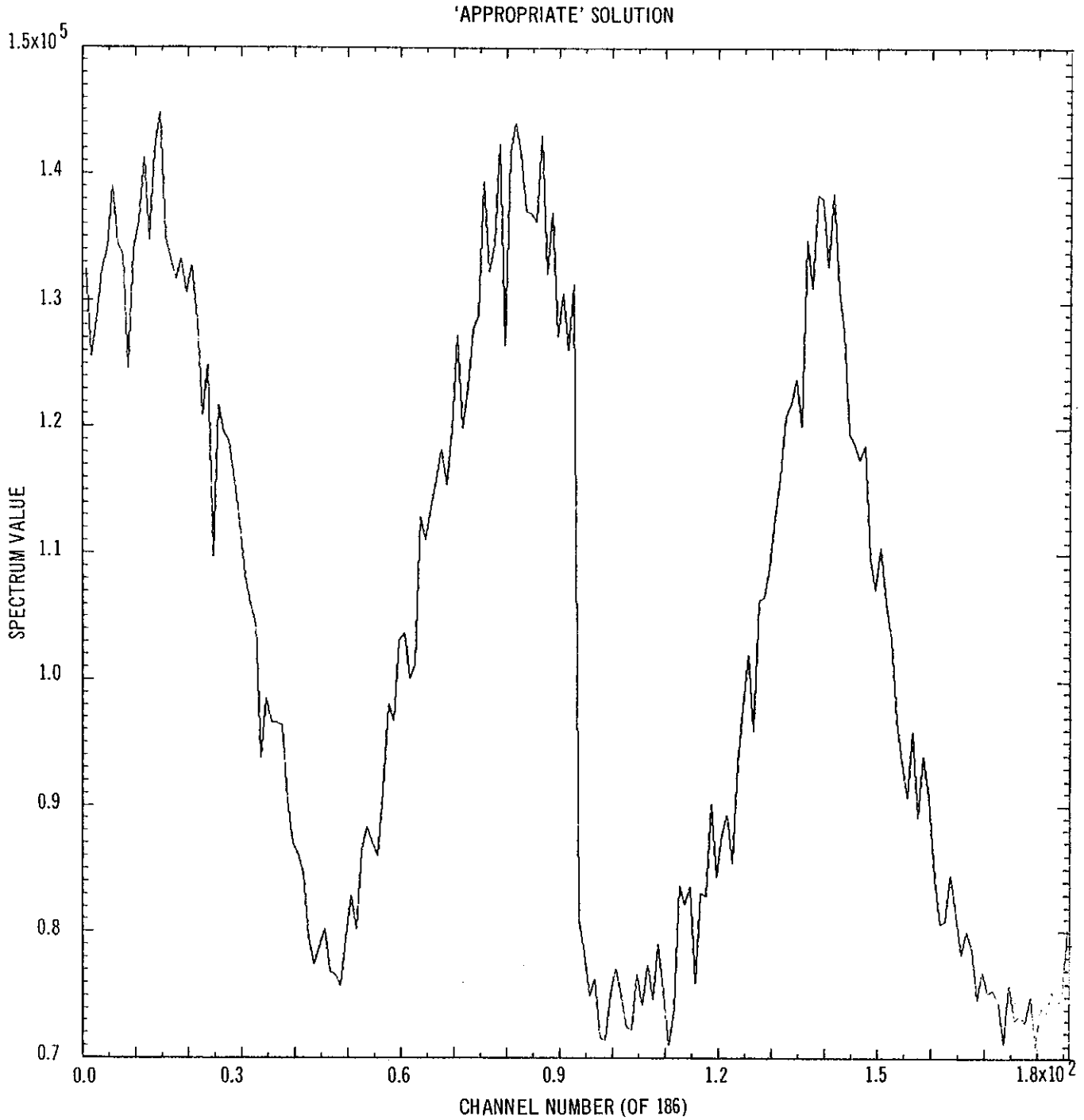

FIGURE 20 AN UNSUCCESSFUL UNFOLDING METHOD

The appropriate solution technique applied to statistical data for the case B.3. The form of the spectrum eventually converges as explained in the text, but the fit to the data gets worse as it does so.

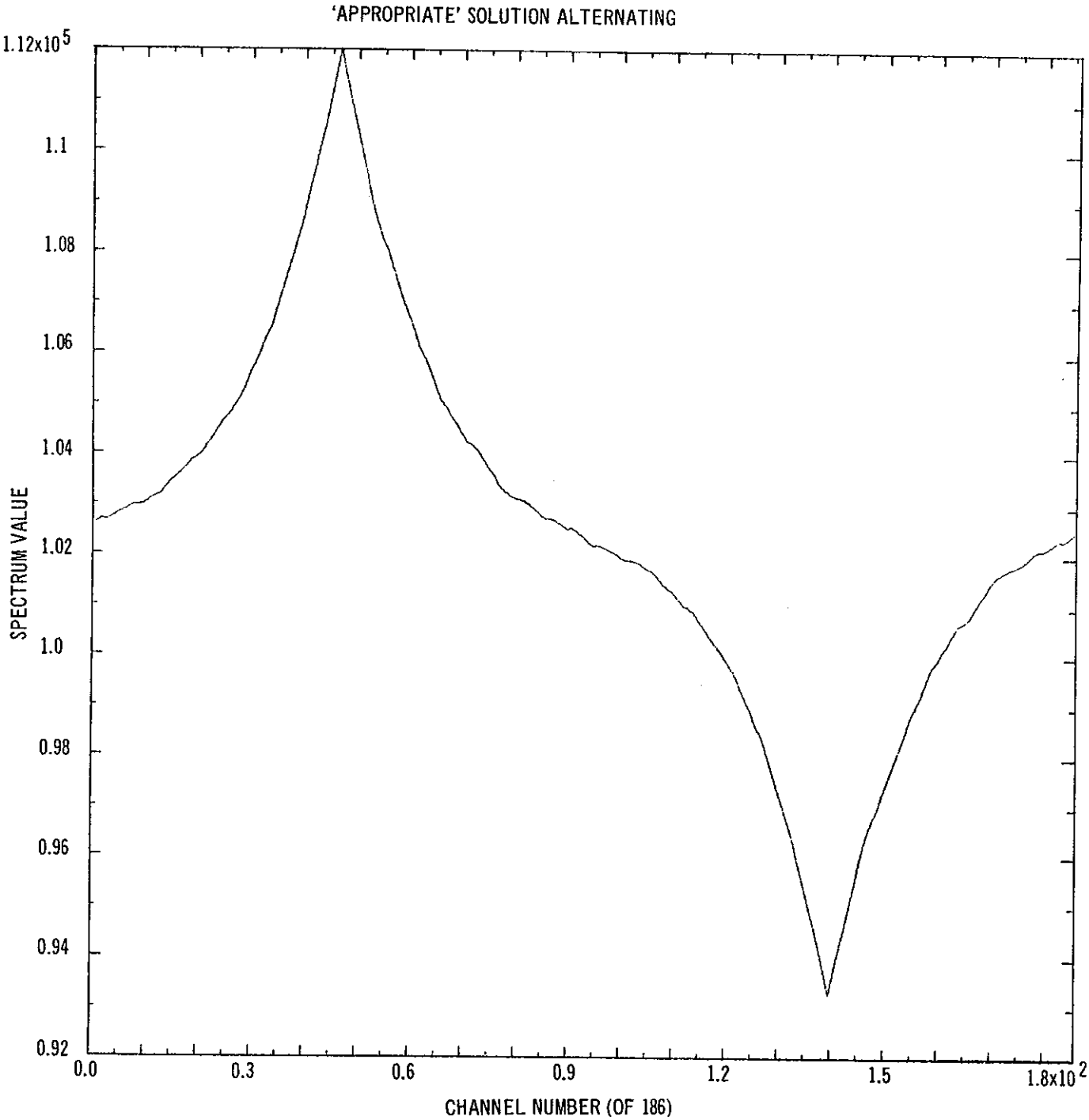

FIGURE 21 UNFOLDING WITH A MODIFIED APPROPRIATE SOLUTION TECHNIQUE

The technique of alternating iterations as explained in the text, produces convergence in a similar fashion to the convergence of the series technique with positive eigenvalues. The strength of the parallel depends here on the large background.

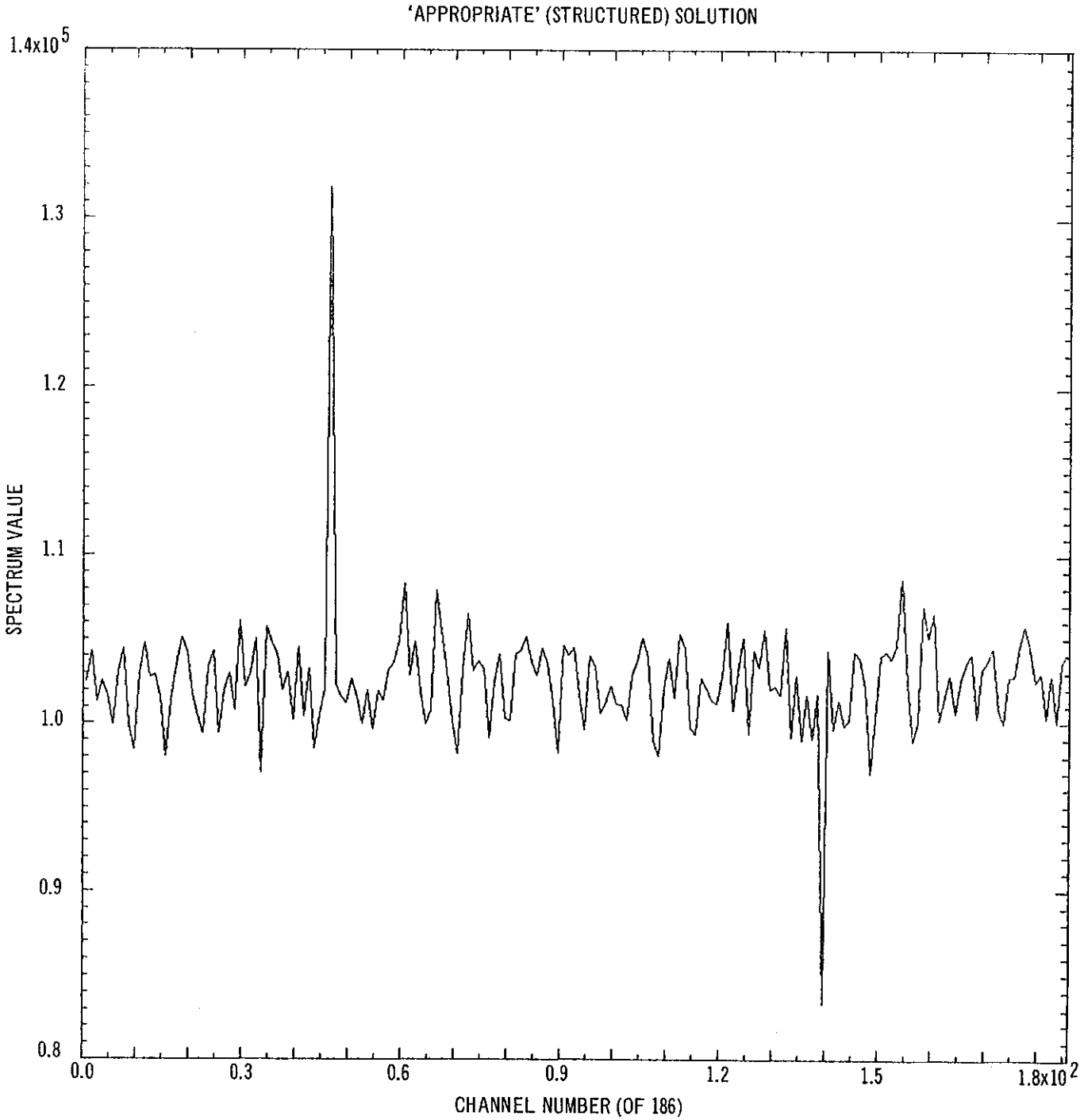

FIGURE 22 A MODIFICATION OF THE APPROPRIATE SOLUTION TECHNIQUE DEPENDING ON THE STRUCTURE OF THE RESOLUTION FUNCTION

This appropriate structure solution is slowly converging here but is selecting the important components of the data. The spectrum shown conceals statistically that there is still a large systematic error in the fit to the data.

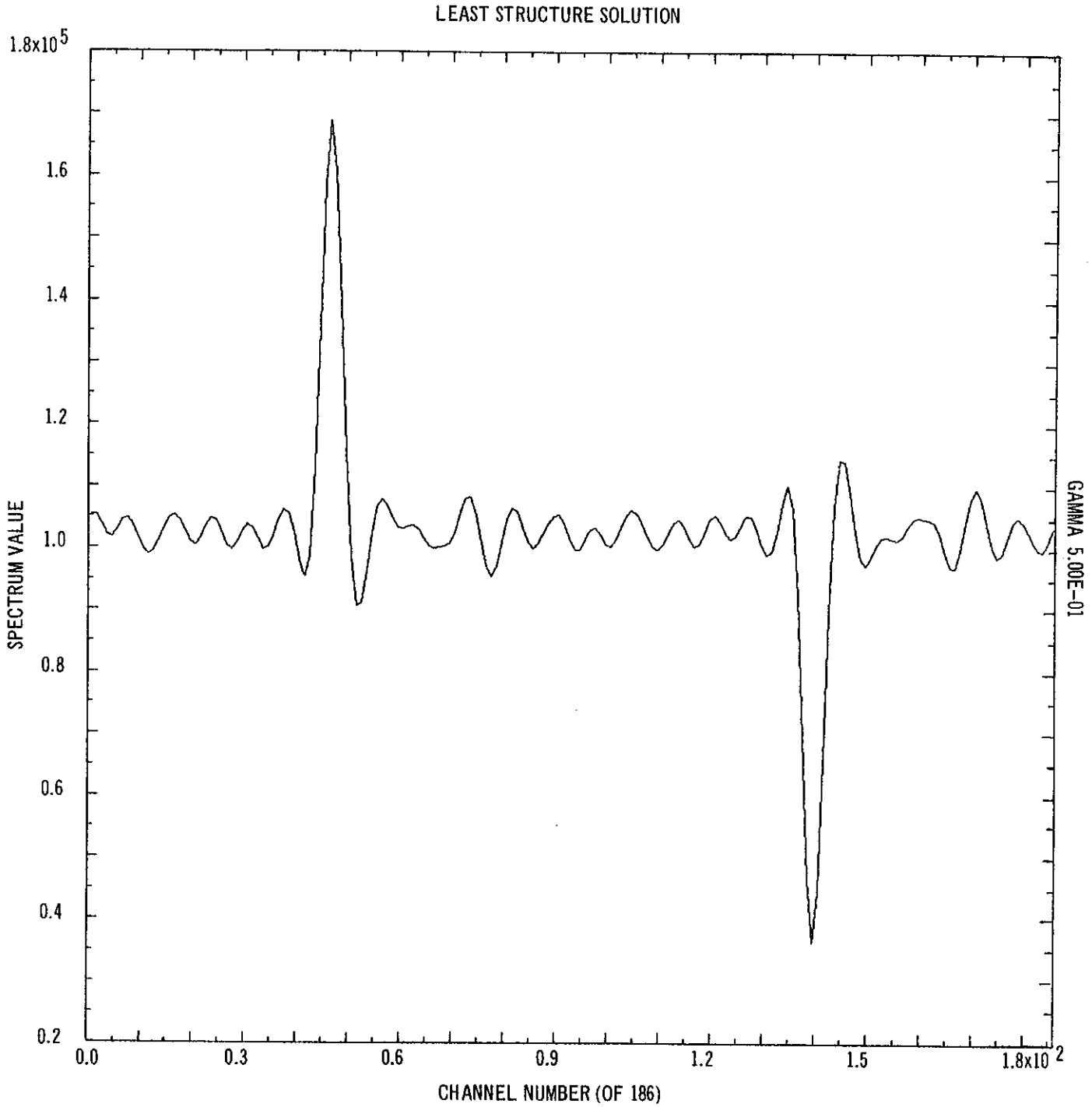

FIGURE 23 LEAST STRUCTURE SOLUTION TO RECOVER THE SPECTRUM OF EQUATION B.15 FROM NOISY DATA UNDER THE CONDITIONS OF B.3

Because we have a simplified inversion technique using equation D.6 instead of a general matrix inversion routine, the results are acceptable.

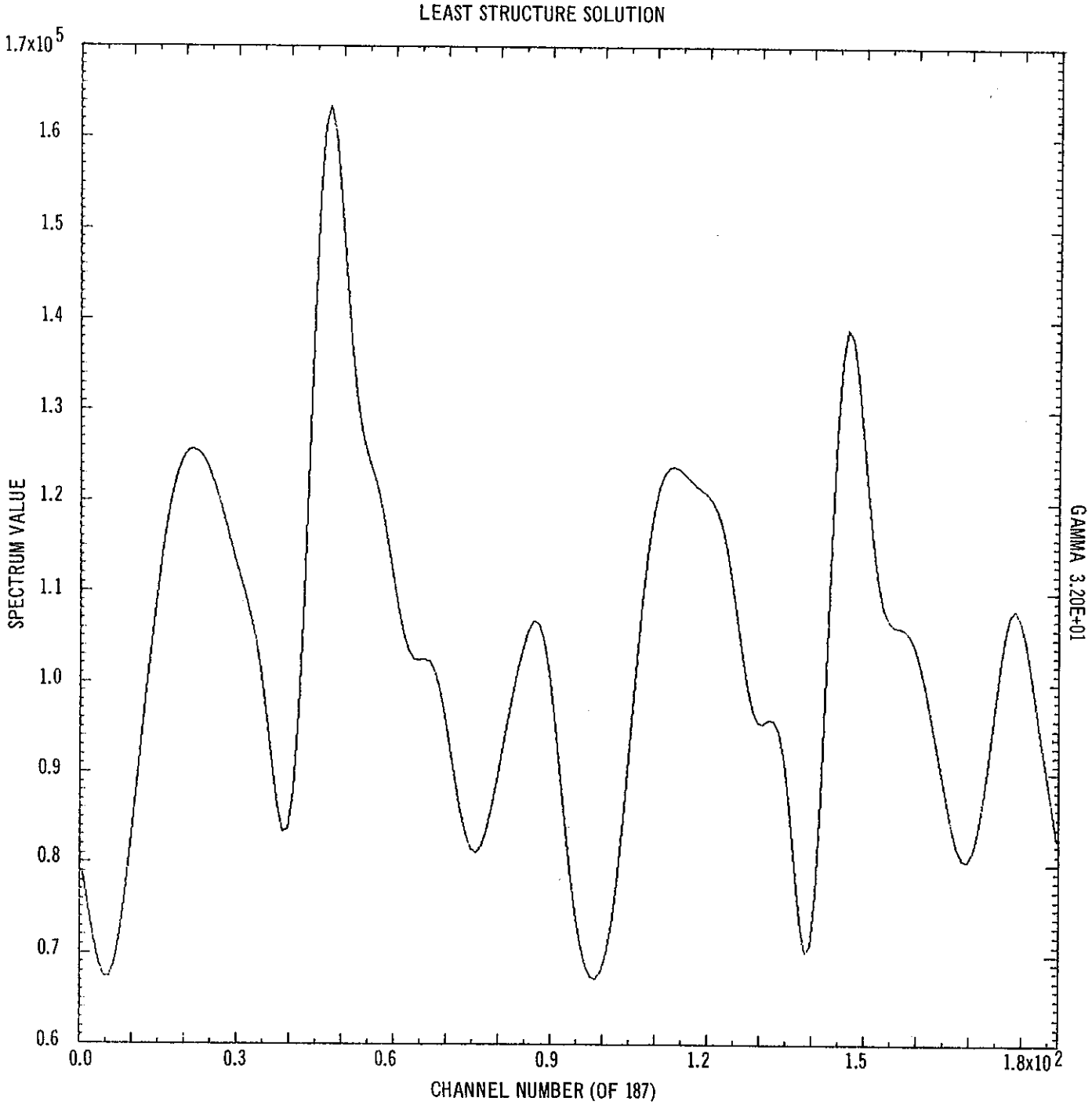

FIGURE 24 THE LEAST STRUCTURE UNFOLDING OF THE DATA OF FIGURE 16

The dominant peak of the original spectrum is obtained but is not seen to be significant because of other features that must be considered artifacts of the method.

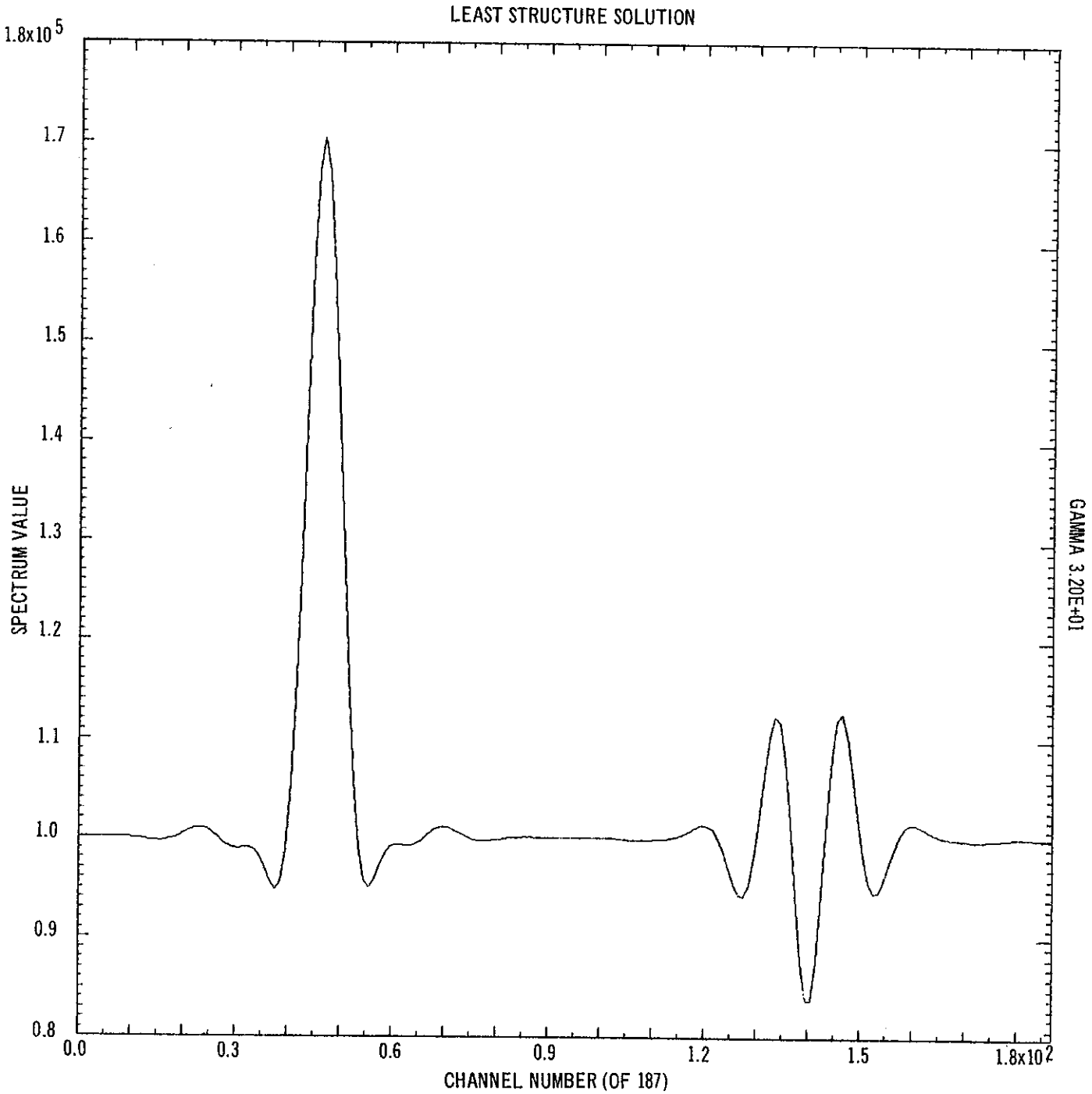

FIGURE 25 THE RESULT OF USING LEAST STRUCTURE UNFOLDING WITH ERROR-FREE DATA

The spectrum recovered appears to be an adequate approximation to that of Figure 15.

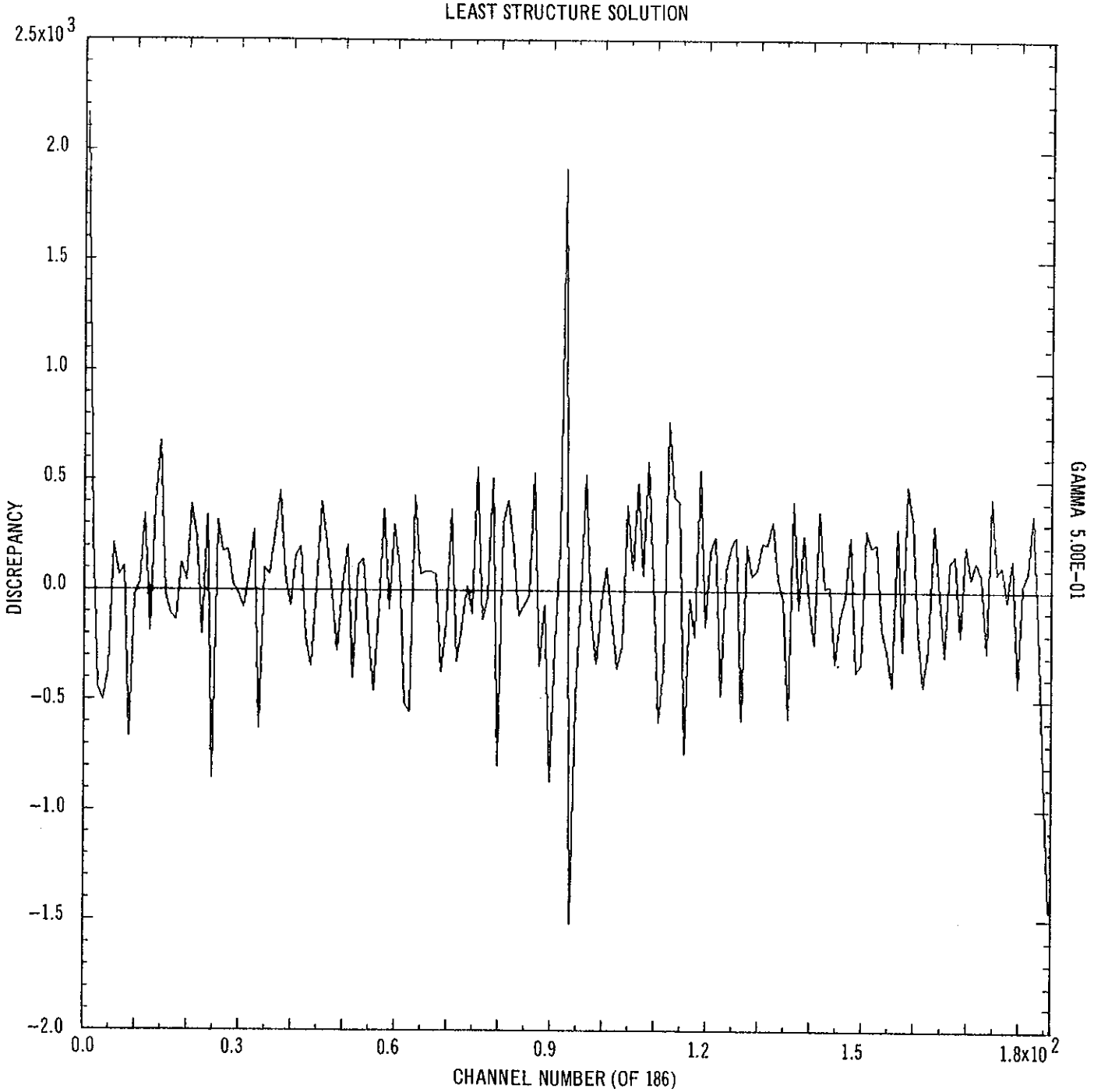

FIGURE 26 RESIDUAL DISCREPANCIES BETWEEN DATA AND YIELDS PREDICTED FROM THE SPECTRUM OF FIGURE 23

There is significant structure indicating a poor fit at channels close to the sudden step in predicted yields. We should be alerted by this structure if we are considering whether there is still unused information in the data. The structure and the noise are characteristic of the data fits achieved with the spectra in Figures 17 and 18 as well as Figure 23. The structure alone shows up in the comparison fits to noise-free data by the associated calculated spectra.

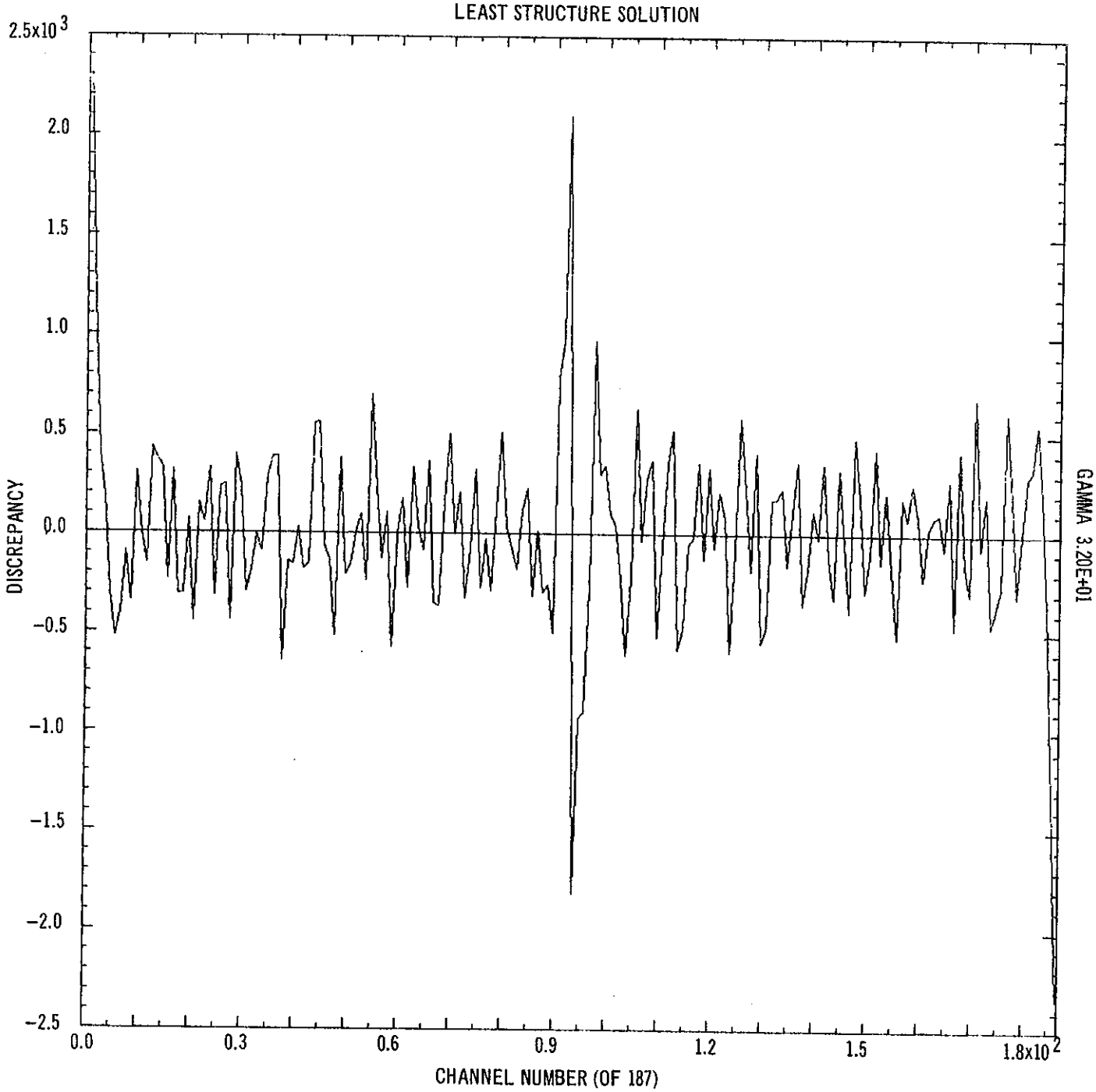

FIGURE 27 THE DISCREPANCIES BETWEEN DATA AND PREDICTED YIELD FOR THE SPECTRUM ILLUSTRATED IN FIGURE 24

The indications of structure are irrelevant to the actual unsatisfactory character of the spectrum in Figure 24.

REFERENCES

1. Fizeau, H. [1868] - C.R. Acad. Sci. Paris, 66 : 934.
2. Lord Rayleigh [1879] - Phil. Mag. 5 (8) 261.
3. Michelson, A.A. [1890] - Phil. Mag. 5 (30) 1.
4. Hanbury-Brown, R., Twiss, R.Q. [1954] - Phil. Mag. 7 (45) 663.
5. Fredholm, Acta Math 27 (1903) 365 and Öfversigt af K. Vetenskaps-Akad. Förhandlinger (Stockholm) 57 (1900) 39.
6. Whittaker, E.T., Watson, G.M. [1927] - A course in modern analysis. Cambridge University Press.
7. Stakgold, Ivar [1967] - Boundary values of mathematical physics. Vol.I, Macmillan, New York.
8. Beers, Richard H. [1971] - Spectrum unfolding using stepwise regression of system response functions. EGG 1183 5000*.
9. Benjamin, P.W., Kemshall, C.D. & Brickstock, A. [1968] - The analysis of recoil proton spectra. AWRE 0-9/68.
10. Bennett, E.F. Gold, R. & Olsen, I.K. [1968] - Analysis and reduction of proton recoil data. ANL 7394.
11. Bennett, E.F. & Yule, T.J. [1971] - Techniques and analyses of fast reactor neutron spectroscopy with proton recoil proportional counters. ANL 7763.
12. Booth, R.S. [1970] - Nucl. Instrum. Methods, 85 : 69.
13. Borgonovi, G.M. & Bromley, G.E. [1971] - Proton recoil data at Gulf Radiation Technology. GULF-RT-10487.
14. Bracewell, R.N. & Roberts, J.A. [1954] - Aust. J. Phys. 7 : 615-640.
15. Bramanis, E., Deague, T.K., Hicks, R.S., Hughes, R.J., Muirhead, E.G., Sambell, R.H. & Stewart, R.J.J. [1972] - Nucl. Instrum. Methods, 100 : 59.
16. Bregman, J.D. & de Mul, F.F.M. [1971] - Nucl. Instrum. Methods, 93 : 109.
17. Bronk, B.V. & Whittem, W.B. [1973] - Nucl. Instrum. Methods, 106 : 319.
18. Burrus, W.R. [1965] - Utilisation of a priori information by means of mathemtical programming in the statistical interpretation of measured distributions. ORNL 3743.
19. Burrus, W.R. & Verbinski, V.V. [1969] - Nucl. Instrum. Methods, 67 : 181-196.

20. Carver, J.H. & Lokan, K.H. [1957] - Aust. J. Phys. 10 : 312.

21. Cerbone, R.J., Harris, L. Jr., Kendrick, H. & Willoughby, D.C. [1971] - Numerical and experimental studies of spectrum unfolding, Vol.I, GULF-RT-10486 VI.

22. Cook, B.C. [1963] - Nucl. Instrum. Methods, 24 : 256-62.

23. Dahmen, H., Dreyer, F., Crawford, D.M. & Thies, H.H. [1973] - Nucl. Instrum. Methods, 107 : 329-32.

24. Dugdale, L.M. [1974] - At. Energy Aust. Vol.17, No.1, pp 11-18.

25. Eckhoff, N.D. [1969] - Nucl. Instrum. Methods, 74 : 77-85.

26. Eckhoff, N.D. [1971] - Nucl. Instrum. Methods, 97 : 263-66.

27. Ekstrom, Michael, P. [1971] - Numerical restoration of random images. (Ph.D. thesis) UCRL 51129.

28. Ekstrom, M.P. & Woods, J.W. [1974] - Application of digital image restoration in X-ray astronomy. UCRL 75895.

29. Estes, C. Michael [1971] - M.S. Thesis, Kansas State University, (unpublished)

30. Filippone, W.L. & Munno, F.J. [1972] - Nucl. Technol. 14 : 200-202.

31. Fioratti, M. Paganini & Piermattei, S. Ricci [1972] - Nucl. Instrum. Methods, 98 : 131-4.

32. Fischer, A. & Turi, L. [1972] - The RFSP program for unfolding neutron spectra from activation data. INDC (HUN)-8/U, Trans. from KFKI-70-39 Rpt.

33. Gold, R. [1964] - Nucl. Sci. Eng., 20 : 493.

34. Gold, R. [1964] - An iterative unfolding method for response matrices. ANL 6984.

35. Gold, R. & Bennett, E.F. [1968] - Nucl. Instrum. Methods, 63 : 285-99.

36. Greer, G.R., Halbeib, J.A. & Walker, J.V. [1967] - A technique for unfolding neutron spectra from activation measurements. SC-RR-67-746.

37. Harris, C., Kendrick, H. & Sperling, S.M. [1970] - An introduction to the principles and use of Ferdor unfolding code. GA9882.

38. Head, J.H. [1972] - Nucl. Instrum. Methods, 98 : 419-28.

39. Horgren, H.M., Villagrana, R.E. & Maher, D.M. [1974] - Computer enhancement of weak-beam images. CALT 767-P-3-32.

40. Hunt, B.R. [1971] - Deconvolution of linear systems by constrained deconvolution regression and its relationship to the Wiener theory. LA-DC-12699.

41. Hunt, B.R. [1970] - Statistical aspects of deconvolution. LA-4556-MS-UC-32.

42. Hunt, B.R. & Andrews, H.C. [1973] - Comparison of different filter structures for restoration of images. LA-DC-73-1333.

43. Iijima, Tsutomu, Hikomukaiyama, Take, & Shirakata, Keisho [1971] - J. Nucl. Sci. Tech., 8 : 192-200.

44. Karayianis, N., Morrison, C.A. & Wortman, D.E. [1970] - Nucl. Sci. Eng., 40 : 38-50.

45. Kemshall, C.D. [1973] - Use of spherical proportional counters for neutron spectrum measurements. AWRE 0-31/73.

46. Kendrick, H., Sperling, S.M., Borgonovi, G.M. & Houston, D.H. [1971] - Numerical and experimental studies of spectral unfolding, Vol. II. GULF-RT-10486 - VII.

47. Kockum, J. [1970] - Nucl. Instrum. Methods, 82 : 285-6.

48. Lang, D.W. [1965] - Nucl. Phys., 72 : 461-74.

49. Leachman, R.B. [1951] - Phys. Rev., 83 : 17-20.

50. Maerker, R.E. & Muckenthaler, F.J. [1969] - Gamma-ray spectra arising from thermal-neutron capture in elements found in soils, concretes and structural material. ORNL-4382.

51. Matthes, W. [1974] - Unfolding of composite spectra by linear regression. EUR 5070e.

52. McElroy, W.N., Berg, S., Crockett, T. & Hawkins, R.G. [1967] - A computer-automated iterative method for neutron flux spectra determination by foil activation. Vol.I, II, III and IV. AFWL-TR-67-71.

53. McElroy, W.N., Berg, S., Crockett, T.B. & Tuttle, R.J. [1969] - Nucl. Sci. Eng., 36 : 15-27.

54. McElroy, W.N., Berg, S. & Gigas, G. [1967] - Nucl. Sci. Eng., 27 : 533.

55. Meyer, W.A. & Mutone, G.A. [1972] - Critical comparison and evaluation of neutron spectra unfolding codes: The Need to Know. Conf. 720901, Book 1, 524.

56. Mijnarends, P.E. [1974] - Unfold Vol. 1: Deconvolution of angular correlations of positron annihilation radiation or Compton line profiles. RCN. 217.

57. Moore, L. [1968] - Brit. J. Appl. Phys. (J. Phys. D) Ser.2, Vol.1, 237-45.

58. Mutone, G.A. & Meyer, W. [1973] - Nucl. Instrum. Methods, 106 : 445-52.

59. Najžed, M., Rant, J. & Šolinc, H. [1972] - 'Set of seven threshold detectors used as a fast neutron spectrometer' and 'Statistical analysis of numerical methods of unfolding spectral data from activation detector'. IAEA-R-536-F.

60. Narum, R.E. [1972] - INSPECT - an interactive computer code for unfolding neutron spectra from activation measurements. ANCR-1035.

61. Parker, J.B., White, P.H. & Webster, R.J. [1963] - Nucl. Instrum. Methods, 23 : 61-8.

62. Peelle, R.W. [1971] - Techniques used at Oak Ridge National Laboratory for unfolding neutron and gamma-ray pulse-height spectra. ORNL-TM-3463.

63. Penfold, A.S. & Leiss, J.E. [1959] - Phys. Rev., 114 1332-37.

64. Phillips, D.L. [1962] - J. Ass. Comp. Mach., 9 : 84-97.

65. Pieroni, Nestor, Rorsch, Detlef, & Wattecamps, Eric, [1974] - Nucl. Instrum. Methods, 115 : 317-23.

66. Rand, R.E. [1962] [1967] - J. Nucl.Inst., 17 : 65-74 & 24 : 127.

67. Rangarnjan, C., Mishra, U.C., Copalakrishnan, Smt. S. & Sadasivan, S. [1973] - Analysis of complex NaI(Tl) gamma spectra from mixtures of nuclides, BARC 686.

68. Rautian, S.G. [1958] - Soviet Physics Uspekhi, 66 (1) 245.

69. Robba, A.A. & Veeser, L.R. [1972] - A least structure unfolding code. LA 5088.

70. Routti, Jorman Tapio, [1969] - Ph.D. Thesis, Lawrence Radiation Laboratory, UCRL 18514.

71. Sanna, R.S. [1971] - A user's guide to SWIFT - a Monte Carlo technique for unfolding neutron spectra. HASL 244.

72. Sanna, R. & O'Brien, K. [1971] - Nucl. Instrum. Methods, 91 : 573-6.

73. Sen, A. & Clarke, J.F. [1971] - Regularization method of unfolding X-ray spectra. ORNL-TM-3582.

74. Tesch, K. [1971] - On the accuracy of the photon-difference method used in high-energy photonuclear experiments, Desy 71/3.

75. Thies, H.H. [1961] - Aust. J. Phys., 14 : 174-87.

76. Thies, H.H., Crawford, D.M., Koch, R. & Thomas, B.W. [1972] - Nucl. Instrum. Methods, 100 : 65.

77. Tominaga, H., Dojyo, M. & Tanaka, M. [1972] - Nucl. Instrum. Methods, 98 : 69-76.

78. Turi, L. & Fischer, A. [1972] - Unfolding of neutron spectra from activation data using the RF 01 and RF 07 codes. INDC (HUN)-7/U. Trans. from KFKI-71-22.

79. Twomey, S. [1963] - J. Ass. Comp. Mach., 10 : 97-101.

80. Verbinski, Victor V. & Giovannini, Raffaele [1974] - Nucl. Instrum. Methods, 114 : 205-31.

81. Verbinski, V.V., Young, J.C., Neill, J.M., Giovannini, R. & Houston, D.H. [1972] - Proton-recoil proportional counter spectrometry for reactor physics measurements. Conf. 720901, Book.2, 959-72.

82. Werle, H. [1972] - Nucl. Instrum. Methods, 99, 295-300.

83. Young, C.S. [1971] - Least structure technique in unfolding. LA-DC-12360.

84. Hestenes, M.R. & Stiefel, E. [1952] - Journal of the National Bureau of Standards, 49, 409-436.

85. Ralston, A. [1965] - A first course in numerical analysis. McGraw-Hill, New York, p.439f.

86. Martin, D.W. & Tee, G.J. [1961] - British Comput. J. (UK), 4 : 242-254.

87. Gastinel, N. [1970] - Linear numerical analysis. Academic Press, New York, p.168f.

88. Reid, J.K. [1970] - A FORTRAN subroutine for the solution of large sparse sets of linear equations by conjugate gradients. AERE-R-6545.

89. Golub, G. [1965] - Numerische mathematik, 7 : 206-216.

90. Businger, P. & Golub, G.H. [1965] - Numerische mathematik, 7 : 269-271.

91. Shannon, C.E. [1949] - Proc. IRE., 37 : 10-21.

92. Patterson, M.R. & Hammerling, F.D. [1971] - Fourier analysis of unequally spaced data. ORNL-TM-3547.

93. Snidow, N.L. & Warren, H.D. [1967] - Nucl. Instrum. Methods, 51 : 109-116.

94. Clements, P.J. [1972] - A discussion of variations on the Scofield gold iterative deconvolution technique. AERE-R-7222.

95. Rosenfeld, Azriel [1969] - Picture processing by computer. Academic Press, New York.

NOTATION

An * indicates that the symbol has a different meaning elsewhere in this report; its context indicates which meaning applies.

Notation	Expression/ Equation No.	Meaning
V	2.1	voltage:setting of recording apparatus
y_j	2.1	yield:population mean of readings with setting V_j
E	2.1	energy:independent variable of spectrum
$s(E)$	2.1	spectrum value at E
E_L, E_u	2.1	particular values of E
$R(V_j,E)$	2.1	resolution or resolving function of apparatus
K	2.1*	normalisation constant for apparatus
$r(V,E)$ $P(V_j,V)$	2.2*	factors of a composite resolution function
$\delta(E_oV_j - V_oE)$	2.3	Dirac delta function of $(E_oV_j - V_oE)$
m_j	2.4	a value obtained for y_j
ϵ_j	2.4	statistically variable discrepancy between m_j and y_j
σ_j	3.1*	root mean square value of ϵ_j
N	3.1	number of data points
χ^2	3.2	chi-squared: measure of agreement with fit to data
$t_k(E)$	3.3	one of set of functions used in linear combination to make up spectrum
P_{jk}	3.4*	resolution matrix element connecting $t_k(E)$ and y_j
A,B	3.6	constants
$t_k'(E)$	3.8	a modified function replacing $t_k(E)$
$\{\ \}$	3.9	set of: $\{R(V_j,E)\}$ set of resolution functions
$s_o(E)$	3.10*	part of s(E) that is a linear combination of $\{R(V_j,E)\}$
$f(E)$	3.10*	the rest of s(E)
$\{R_k'(E)\}$	3.11	set of mutually conjugate linear combinations of $\{R(V_j,E)\}$
c_k	3.11*	coefficients
$\nu(E)$	3.15	weighting function and measure of acceptable variation of s(E)
a_k	4.1*	coefficient of $t_k(E)$ in linear combination making up s(E)
$\vec{a}$	4.3	column vector composed from $\{a_k\}$
$\overleftarrow{a}$	4.4	row vector composed from $\{a_k\}$
P	4.3	matrix with P_{jk} in j^{th} row and k^{th} column
P^T	4.4	transpose of P
w	4.6	weight matrix
δ_{jk}	4.6	Kronecker delta of j,k
$\vec{g}$	4.7	vector of weighted data
$\vec{x}$	4.9	vector of weighted yields
e_j	4.10	weighted error at j^{th} data point
Q	4.11	matrix obtained from P by weighting
$\vec{b}$	4.16	spectrum vector form of term to be added to $\vec{a}$
μ	4.16*	a multiplier
$\vec{a}^{(n)}$	4.25	form of vector $\vec{a}$ associated with n^{th} iteration
$\vec{v}^{(n)}$	4.35	yield vector associated with spectrum vector $b^{(n)}$
$\hat{v}^{(n)}$	4.36	unit vector parallel to $\vec{v}^{(n)}$
$c^{(n)}$	4.35	coefficient of $\hat{v}^{(n)}$
Q^{-1}	5.1	inverse matrix of Q
I	5.2	identity matrix of appropriate dimension
λ	5.3*	an eigenvalue of Q
$\vec{u}$	5.19*	a vector in spectrum space
$r_v(V_j)$ $r_o(V-E\frac{V_o}{E_o})$ $r_e(E)$	5.20	factors of $R(V_j,E)$ in special case depending only on the quantities named
$y(k)$	5.22	Fourier transform of y(E)
$R(k)$	5.24	Fourier transform of $r_o(E-E')$
$S(k)$	5.24	Fourier transform of S(E)
i	5.22	$\sqrt{-1}$
k^*	5.26	complex conjugate of k
$k_n + jq_n$	5.27*	root of R(k)=0

Notation	Expression/ Equation No.	Meaning
$a_n = \rho_n \exp(i\delta_n)$	5.29*	coefficient
ΔE	5.34	spacing (constant) of data points
r_j	5.34	lumped form of r(E) for matrix equation
s_n	5.35	lumped form of spectrum $s(E_n)$
T_k	5.37	element of inverse of special resolution matrix
λ_k	5.39*	root of equation $r(\lambda)=0$
a_k	5.39*	coefficient
q, r	5.51	perturbations (assumed small) of resolution matrix and inverse
$P(m_j)$	6.1*	probability density for measurement m_j
$\{\hat{u}_k\}$	6.13*	a set of orthormal vectors in data significance space
$\{\hat{d}_j\}$	6.13*	one special set of the type $\{\hat{u}_k\}$
$h_{j\ell}$	6.15	element of orthogonal matrix connecting $\hat{d}_j$ and $\hat{u}_\ell$
$\left\|\frac{\partial \bar{c}}{\partial b}\right\|$	6.17	Jacobean determinant of coordinate change
ν_k	6.23	lumped form of ν (E_k)
E_o, Γ, a, b	6.24	nuclear parameters
$\bar{c}^{(k)}$	6.26*	mean of values derived for $c^{(k)}$ from sets of equivalent data
$\sigma(E_n)$	7.2*	nuclear cross section associated with energy E_n
c	7.3*	speed of light
d	7.3*	distance (flight path)
t	7.3*	time (of flight)
m	7.4*	mass (of neutron)
P(t)	7.7*	pulse shape function
P_k	7.9*	lumped pulse shape function
M	7.12*	total counts in pulse profile
k_{max}	7.12*	number of channels in pulse profile
$L, L+L_1$	7.15	boundary channels for time of flight data
f_ℓ, f'_ℓ	7.16*	factors used in correction of edge effects
$P_\ell(\mu)$	7.28*	Legendre polynomial of degree l with argument μ
$X_o, X_\pm$	7.31	three special values of spectrum
$P_o, P_\pm$	7.31*	three special values of pulse profile
a_o, a_1, a_2	7.33	coefficients
W	7.33	a width
$A_k, C_k, \sigma_k(E), \phi(E), T_k$	7.37	standard notation concerning foil activation measurements of neutron flux
m_1, m_2, M, E_1, E_2	7.38*	particular masses and energies
F	8.4	matrix providing measure of structure
γ	8.7	parameter governing suppression of structure
d_k	A.15*	a coefficient
G_{nm}	A.17	matrix: element used in analysis of conjugate gradient process
f_m	A.22*	coefficient
H_{mr}	A.22*	term in element of determinant
$\lambda_k(m), k_m, \lambda_{km}$	A.22	associates the k^{th} eigenvalue with position m in a set of n letters
Prod $(k_1, k_2 \ldots k_n)$	A.23	shortened form of recurring antisymetric product
$\varepsilon(k_1, k_2 \ldots k_n)$	A.24	antisymetric operator with values $\pm 1, 0$.
$\det\|H_{mr}\|$	A.26	determinant with general element H_{mr}
$S_k^{(m)}$	A.28	elementary symmetric function of degree k of a set of m letters
$\text{Num}_s(n)$	A.34	recurring numerator of expression at n^{th} iteration with special eigenvalue λ_s
$\text{Den}^{(n)}$	A.34	corresponding denominator common to all eigenvalues
$\lambda, \varepsilon, p, q$ and M	A.45*	constants in an illustrative example
a_k	B.5*	value calculated from data for s_k
Δ	B.14	a constant
f, f_m	C.1*	frequency (maximum frequency)
$R_e(\ )$	C.18	Real part of expression in parenthesis
μ_k	D.6*	eigenvalue of product matrix F^TF
μ_ℓ, λ_k	E.3*	generalisation of λ_k from eqn (5.39)
H_{kn}	F.6*	element of Hilbert matrix
H_{kn}	F.7	element of generalised Hilbert matrix