MATHEMATICAL TREATMENT

OF

DATA CONTAINING

INSTRUMENT RESPONSE

MATHEMATICAL TREATMENT OF DIGITIZED DATA CONTAINING INSTRUMENT RESPONSE AND STATISTICAL DEVIATIONS

by

DARIUS DOMAS SLAVINSKAS, B.A.Sc., M.Sc.

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

April 1966

DOCTOR OF PHILOSOPHY (1966) McMASTER UNIVERSITY
Hamilton, Ontario.

TITLE: Mathematical Treatment of Digitized Data Containing Instrument Response and Statistical Deviations.

AUTHOR: Darius Domas Slavinskas, B.A.Sc. (University of Toronto)
M.Sc. (McMaster University)

SUPERVISOR: Dr. T. J. Kennett

NUMBER OF PAGES: vi, 145

SCOPE AND CONTENTS:

Two methods of approach are presented for the analysis of digitized data containing instrument response effects. The first method corrects the data vector by means of multiplication by the inverse response matrix. In order to accomodate large size data fields, a special effort is made to obtain expressions for the inverse matrix elements in closed form. Reduction of statistical uncertainties is accomplished by application of non-negativity conditions. The second approach is based on the method of least squares. Applications to two-dimensional coincidence spectra and nonlinear model functions are discussed in some detail. Although the main emphasis is placed on analysis of nuclear spectra, the techniques presented need not be limited to this application alone.

ACKNOWLEDGEMENTS

I wish to thank the members of my Supervisory Committee, Dr. T. J. Kennett, Dr. M. W. Johns and Dr. G.L. Keech, whose guidance and encouragement have made the completion of this work possible. In particular I am indebted to my Research Director, Dr. T. J. Kennett, whose special insight and a wealth of suggestions have provided avenues of approach to the more difficult problems.

I am also indebted to Dr. B. Banaschewski for his interest and fruitful suggestions regarding some mathematical aspects of this study. Frequent discussions with Dr. W. V. Prestwich and his constructive criticisms have been most helpful, for which I express my sincere gratitude. The patient assistance provided by Dr. G. L. Keech, Dr. D. J. Kenworthy and Dr. W. H.Fleming when computer programming difficulties arose is also greatly appreciated. I wish to thank Dr. N. P. Archer for providing a computer program for generation of γ-ray response shapes in NaI(Tℓ) detectors.

A special vote of thanks is due Mrs. H. Kennelly who bravely tackled the typist's nightmare of reproducing a multitude of formulae. Her speed and accuracy were of great service to me.

Last, but not least, I wish to thank my wife, Gintara, for her moral support and willingness to sacrifice her personal comforts during the years of my study.

Financial support by the Province of Ontario is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

Recent developments in nuclear spectroscopy have flourished to a high degree of sophistication in data acquisition equipment and techniques. Multi-channel analyzers with up to 16,000 channels are becoming increasingly common. As a consequence of the sheer bulk of data generated the experimenter must perforce have a less intimate contact with raw results and take recourse to automatic data processing. Fortunately this approach is made possible by the increasing availability of high speed and large memory capacity computers.

The physically meaningful parameters are in general not directly available from the raw data. Usually it is necessary to create a family of mathematical models for the spectrum and then apply a process of selecting a particular model which will adequately describe the observed data. The required physical quantities are then contained within the parameters of this model. Construction and mathematical methods of processing such models will be the main subject of this thesis.

The nuclear spectra considered here possess two particular characteristics which render their analysis difficult. Firstly, the detector response to mono-energetic radiation is usually a complicated function which may contain more than one peak. Secondly, the number of counts accumulated in each analyzer channel is subject to statistical deviations. As a consequence, the observed spectrum may bear little

similarity to the actual distribution of energies emanating from the nuclide under study. An example of such response is given in Figure 1 which shows the ^{137}Cs γ-ray spectrum taken with a NaI(Tℓ) scintillator. This particular nuclide decays by β^- emission to ^{137}Ba which subsequently de-excites by emitting a 662 keV γ-ray. The main peak shown in channel 157 is produced by the photo-electric effect which usually results in deposition of all incident γ-ray energy inside the scintillator. Structure to the left of this peak is due to partial energy losses. The plateau below channel 110 is caused by Compton scattering events after which energy-degraded photons may escape from the detector. Compton-scattered γ-rays from material outside the scintillator produce the small peak at channel 46. The entire response displays effects of a Gaussian resolution which is caused in part by statistical fluctuations of the number of electrons collected in the photo-multiplier. This effect presents the greatest difficulty when the spectrum contains γ-rays with small energy separation. Most types of response functions considered in this thesis will take the Gaussian resolution into consideration.

It is possible to derive a general mathematical expression for the observed spectrum. Let the detector response function to incident radiation of energy y be given by $R(x,y)$. If we stipulate the condition of normalization

$$\int_0^\infty R(x,y)dx=1 \ , \qquad 1.1$$

then we can think of $R(x,y)dx$ as the probability that an event of energy y in the true spectrum will produce a response in the energy interval x to

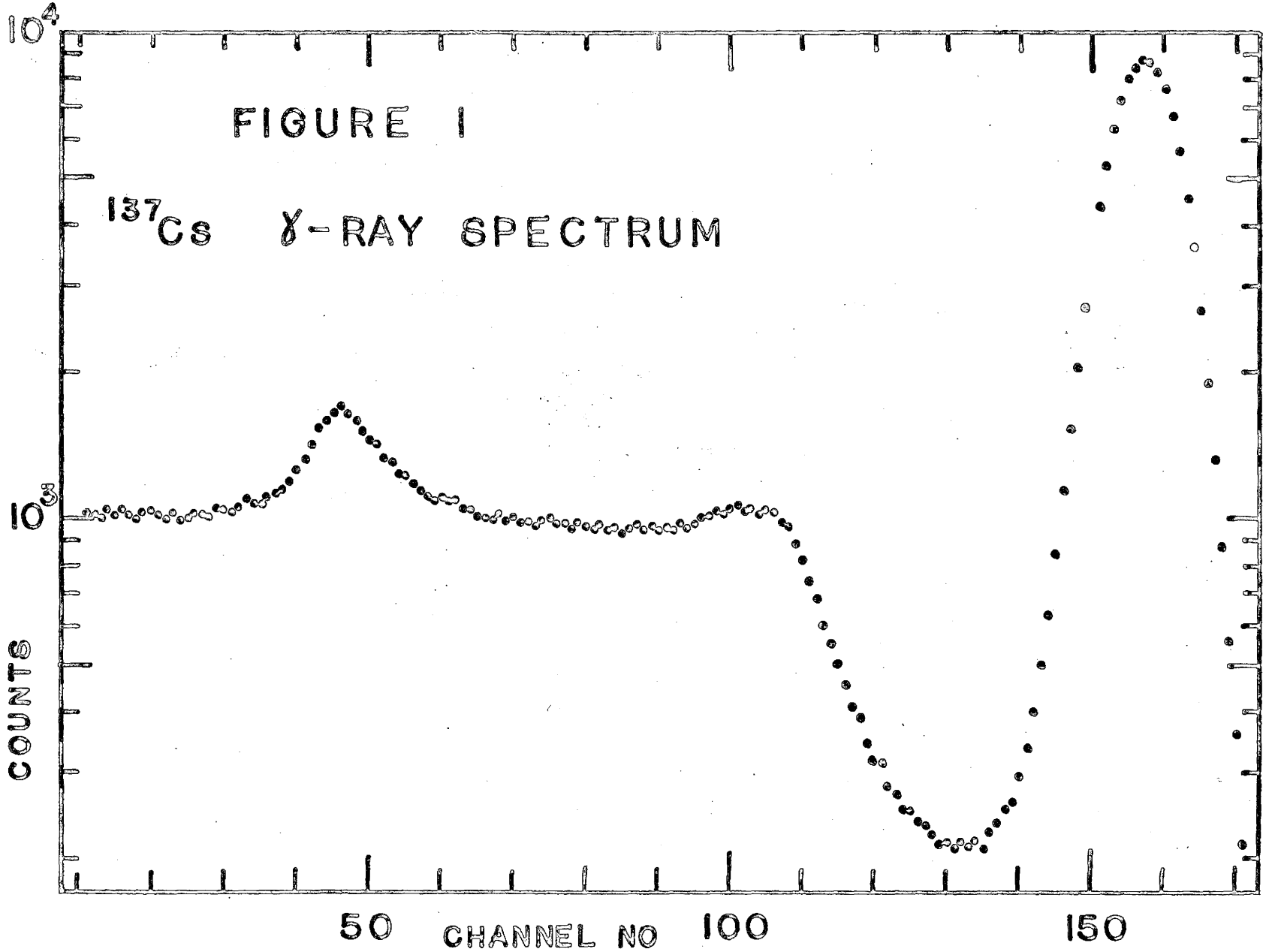


FIGURE 1
^{137}Cs γ-RAY SPECTRUM
10^4
10^3
COUNTS
50
CHANNEL NO
100
150

to x + dx. The concept of the "true" spectrum T(y) as used in this context perhaps needs some clarification. Imagine an experiment performed with a perfect detector having the response function $R(x,y) = \delta(x-y)$. When accumulated into channels the results represent a histogram with statistical uncertainties. If we now let the channel width Δ approach an infinitesimal value dy and take the average of a large number of identical experiments (approaching infinity in number), then the histogram approaches a smooth function given by T(y)dy. In other words, T(y) is proportional to the probability density function t(y) of the sample emitting radiation having energy y. We can therefore write

$$T(y) = \epsilon N_o \, t(y) \quad , \tag{1.2}$$

where ϵ is the detector efficiency and N_o is the average number of emissions radiated over the duration of experiment. Since t(y) must be normalized to unity, we have

$$\int_0^\infty T(y)dy = \epsilon N_o . \tag{1.3}$$

As our next step we can consider the effects of response R(x,y) being other than a delta function. The original function T(y) is modified to another function M(x) according to the convolution integral

$$M(x) = \int_0^\infty R(x,y)\, T(y)dt \tag{1.4}$$

Note that by virtue of eq.(1.1) the area under the function is conserved, i.e.

$$\int_0^\infty M(x)dx = \epsilon N_o . \tag{1.5}$$

If we think of x as the pulse height, then the probability density function of pulse heights entering the analyzer is given by $M(x)/(\mathcal{E}N_o)$. The analyzer sorts these pulses into channels centered at positions x_i. For the moment we ignore any analyzer nonlinearity effects and assume that everywhere we have the same constant difference $x_{i+1} - x_i = \Delta$. If now the i^{th} channel has a pulse acceptance profile $P_i(x)$, then the contents of this channel are on the average given by

$$m_i = \int_o^\infty M(x)\, P_i(x)dx. \qquad 1.6$$

An ideal channel profile would have the shape

$$P_i(x) = \frac{1}{\Delta} \quad \text{for} \quad |x_i - x| < \frac{\Delta}{2}$$

$$= 0 \quad \text{for} \quad |x_i - x| > \frac{\Delta}{2}. \qquad 1.7$$

In practice the profile does not have such sharp cut-off edges[1], however we can use the form of (1.7) without departing too much from reality. Substituting accordingly in eq. (1.6) we obtain

$$m_i = \frac{1}{\Delta} \int_{x_i-\Delta/2}^{x_i+\Delta/2} M(x)dx\,. \qquad 1.8$$

Up to now we talked about expectation values of spectra and ignored statistical effects. In a particular experiment one observes a statistical spectrum which we denote by m_i'. Throughout this thesis primes will be used exclusively for the purpose of indicating statistical quantities. As shown in Appendix I the probability of obtaining a particular value m_i' is given by the Poisson frequency function

$$p(m_i', m_i) = \frac{m_i^{m_i'}}{(m_i')!} e^{-m_i} \quad . \tag{1.9}$$

The same appendix also shows that for large m_i expression (1.9) is closely approximated by a Gaussian frequency function, i.e. we have the asymptotic expression

$$p(m_i', m_i) \sim \frac{1}{\sqrt{2\pi m_i}} e^{-\frac{(m_i' - m_i)^2}{2 m_i}} \tag{1.10}$$

We can thus separate the statistical part of the spectrum and write

$$m_i' \sim m_i + s_i' \quad , \tag{1.11}$$

where

$$s_i' = m_i' - m_i \quad , \tag{1.12}$$

and the statistical part has the frequency function

$$f(s_i', m_i) = \frac{1}{\sqrt{2\pi m_i}} e^{-\frac{(s_i')^2}{2m_i}} \quad . \tag{1.13}$$

Collecting the results of eqs. (1.4), (1.8) and (1.11) we have the final mathematical model of the observed spectrum,

$$m_i' \sim \frac{1}{\Delta} \int_{x_i - \frac{\Delta}{2}}^{x_i + \frac{\Delta}{2}} dx \int_0^{\infty} R(x,y)T(y)dy + s_i' \quad . \tag{1.14}$$

This spectrum can be modified somewhat by addition of pulses which occur within the electronic resolving time. Such events may be due to chance or they may occur on account of two nuclear transitions being in prompt cascade.

Corrections (both experimental and analytical) for these effects have been discussed by various authors [2,3,4,5] and a short summary is given in Appendix II. In the main text we shall assume that, if necessary, the appropriate corrections have been made and that eq.(1.14) represents the experimental spectrum.

Our problem is to determine T(y) as accurately as possible from the given set of observed values m_i'. An exact determination is of course impossible, mainly due to the statistical uncertainty in s_i' and the integration of data into a finite number of channels.

There appear to be two main avenues of approach in determining T(y). The first approach is based on inversion of the response matrix and will be treated in Chapter II. In this approach no assumptions about the particular form of T(y) are necessary; however the condition that T(y) be non-negative can be put to effective use in reducing statistical uncertainties. The second approach discussed in Chapter III requires that T(y) be described by a model function in which the parameters are adjusted to give a best fit to the data. Usually this approach results in a weighted least squares calculation.

Chapter II includes a detailed description of matrix inversion methods applied to resolution corrections of β-ray spectra taken with organic scintillators. Linear Fermi-plots are obtained for almost the entire range of the energy spectrum. In addition this chapter presents a derivation of an inverse matrix in closed form for a certain class of infinite size response matrices. The class is limited to response functions which have shapes essentially independent of energy.

Chapter III includes discussion of two-dimensional time-correlation

experiments. A method of reducing the large scale problem to a simplified solution by parts is presented together with the statistical properties of the parameters thus obtained. A model calculation with γ-ray coincidence spectra is used to illustrate how one may determine γ-ray cascades, branching ratios and absolute transition rates independently of knowledge of detection efficiencies. The effects of various least squares calculation weighting functions are also studied.

The important class of least squares calculations with model functions which are non-linear in their parameters is discussed in some detail in Chapter III. Conditions are given under which the probability density function of estimates of parameters departs significantly from a Gaussian shape. Approximate probability density functions are derived for a certain class of models and compared to the results of Monte Carlo calculations.

Chapter IV summarizes the results and discusses relative merits of the two above-mentioned approaches in determining $T(y)$. It is shown how under certain conditions the two approaches merge to become identical.

CHAPTER II

RESPONSE MATRIX APPROACH

2.1 Matrix Model of System Response

In this chapter we shall treat the detector response problem by methods of matrix algebra. The experimental spectrum m_i' of model equation (1.14) is already of a form which can be expressed by a vertically arranged vector (one column matrix)

$$\underline{M}' \equiv \begin{Vmatrix} m_1' \\ m_2' \\ m_3' \\ \vdots \\ m_N' \end{Vmatrix} . \qquad 2.1.1$$

As indicated by the subscript of the last component, we are limiting ourselves to consideration of a data field with N channels. In a similar way we can define the random vector

$$\underline{S}' \equiv \begin{Vmatrix} s_1' \\ s_2' \\ s_3' \\ \vdots \\ s_N' \end{Vmatrix} . \qquad 2.1.2$$

The response function R(x,y) and true spectrum T(y) have in principle continuous arguments and cannot be exactly represented by matrices with a finite number of elements. It thus becomes necessary to make the approximating assumption that the true spectrum T(y) can be replaced by a digitized version

$$T(y) \longrightarrow \sum_{j=1}^{N} t_j \, \delta(y-y_j) \, , \qquad 2.1.3$$

where

$$t_j = \frac{1}{\Delta} \int_{y_j-\frac{\Delta}{2}}^{y_j+\frac{\Delta}{2}} T(y)dy. \qquad 2.1.4$$

Note that these definitions preserve the original normalization condition:

$$\int_0^{\infty} T(y)dy = \sum_{i=1}^{N} t_j = \epsilon N_o \, . \qquad 2.1.5$$

Substitution of definitions (2.1.3) and (2.1.4) in model equation (1.1 4) yields the result

$$m_i' = \sum_{j=1}^{N} \left[\frac{1}{\Delta} \int_{x_i-\frac{\Delta}{2}}^{x_i+\frac{\Delta}{2}} R(x,y_j)dx \right] t_j + s_i' \, . \qquad 2.1.6$$

We can now define a response matrix

$$\underline{R} = \begin{Vmatrix} r_{11} & r_{12} \cdots r_{1N} \\ r_{21} & r_{22} \cdots r_{2N} \\ \cdots & \cdots \\ \cdots & \cdots \\ r_{N1} & r_{N2} \cdots r_{NN} \end{Vmatrix} \, , \qquad 2.1.7$$

where the elements are given by

$$r_{ij} = \frac{1}{\Delta} \int_{x_i - \frac{\Delta}{2}}^{x_i + \frac{\Delta}{2}} R(x, y_j)dx. \qquad 2.18$$

Using these definitions we can replace eq. (2.1.6) by the simplified expression

$$m_i' = \sum_{j=1}^{N} r_{ij} t_j + s_i' \quad . \qquad 2.19$$

By the rules of matrix multiplication and addition this last statement is just the matrix equation

$$\underline{M}' = \underline{R}\,\underline{T} + \underline{S}' \quad , \qquad 2.1.10$$

where $\underline{T}$ is a vertical vector with components t_j.

Solution for $\underline{T}$ can be attempted by methods of matrix algebra. If one obtains the inverse of response matrix $\underline{R}$, then multiplication of eq. (2.1.10) by $\underline{R}^{-1}$ from the left will produce the result

$$\underline{R}^{-1}\,\underline{M}' = \underline{T} + \underline{R}^{-1}\,\underline{S}' \quad . \qquad 2.1.11$$

Unfortunately the right hand side of this expression contains the undesirable statistical vector $\underline{R}^{-1}\,\underline{S}'$. For inverse matrices $\underline{R}^{-1}$ containing elements much larger than unity this vector may represent a great magnification of the original statistical uncertainties, even to the extent as to overshadow the structure of the true spectrum $\underline{T}$. This is notably true for Gaussian response shapes having widths extending over many channels. However, there are cases where $\underline{R}^{-1}$ does not contain large elements and practical application of matrix inversion

can be made, as is illustrated by the response corrections to β-ray spectra treated in a later section of this chapter. Even in cases where magnification of statistical uncertainties is large the results might be salvaged by applying the condition that the true vector $\underline{T}$ must be non-negative. This procedure may result in considerable reduction of statistical deviations. The statistical effects created by various inverse matrices $\underline{R}^{-1}$ and the application of non-negativity are discussed in the last two sections of the present chapter.

Expression (2.1.3) contains the assumption that the true spectrum consists of delta functions placed in the middle of each channel. Consequently a spectrum T(y) containing entries at intermediate positions has no representation in the matrix model. When T(y) is a slowly varying function of y, this limitation is not serious since the set of "sampling" values t_j adequately defines T(y). However, if the true T(y) contains a delta function between channel centers, the unfolded spectrum $\underline{R}^{-1}\ \underline{M}^{0}$ tends to split the contents between the two adjacent channels. Therefore the matrix model cannot distinguish one delta function placed between channel centers from two delta functions placed at centers of adjacent channels. This effect is illustrated by calculations in section 2.5.

A number of numerical methods of unfolding (correction for response matrix) have been previously reported[6,7,8,9,10]. These methods generally require involved calculations, the difficulty of which increases rapidly with N, the number of channels used. Apparently the simplest to apply are the iterative methods[6,7]. However, their rate of convergence depends considerably on the type of response matrix $\underline{R}$.

In particular, when the response has a Gaussian shape, the convergence can be very slow. The Gaussian resolution correction has been treated analytically by Dixon et al[11]. Their solution for T(y) is given in series form containing increasing orders of derivatives of the observed spectrum M(y). Since high order derivatives are difficult to evaluate the series must perforce be terminated. This has the effect of introducing oscillations in the calculated spectrum. Various methods of numerical matrix inversion are discussed in section 2.3.

Since in practice the inversion of matrices by numerical methods is limited to rank not much larger than, say, 50, the present work is concentrated on attempts to obtain a representation of the inverse matrix in closed form. A great advantage of this approach is the fact that data fields of almost unlimited number of channels can be considered. Computer calculations can be performed quite rapidly without the danger of large rounding-off errors that are often encountered during numerical inversion of large size matrices.

2.2 Correction for Analyzer Nonlinearity

Usually analyzers possess some degree of nonlinearity which results in pulse heights being distributed into bins of unequal width Δ. Although as a rule slight, the effects of this phenomenon are twofold: firstly, the abscissa, or pulse-height scale is distorted and secondly, channels with greater widths receive more than their proper share of counts. For linear pulse-height scales the response matrix can often be derived from physical considerations (see section 2.4 on β-ray spectra). It may thus be desirable to correct the observed spectrum by transferring its contents to a new set of channels which span equal pulse-height intervals.

The required pulse-height vs. channel curve can be obtained by the use of a calibrated pulser. From the set of analyzer addresses x, obtained at various pulse-height settings y, the following curve was derived for a 256 channel analyzer:

$$y = A(B + x + 1.09 \times 10^{-4} x^2) . \qquad 2.2.1$$

The parameters A and B represent the gain and origin shift respectively. Thus for a particular set-up of instruments only these two parameters need be determined for a full calibration.

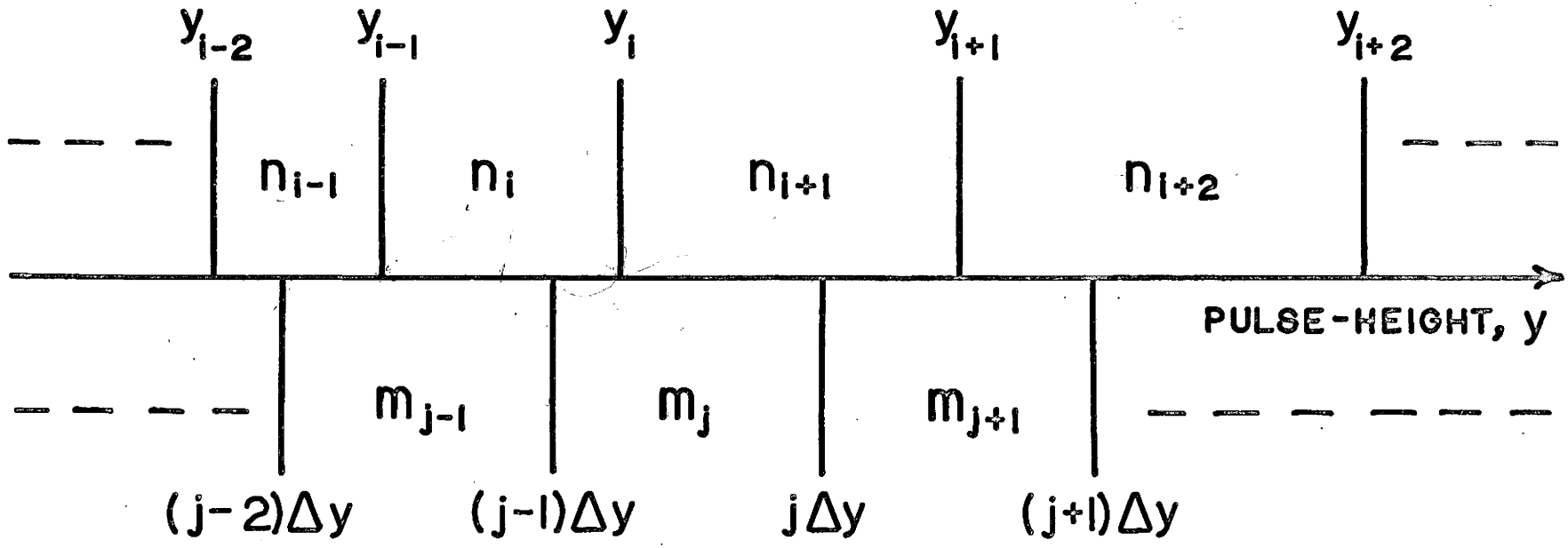
Having obtained the explicit relationship $y = f(x)$ we can proceed with the correction as illustrated in Figure 2. Channel bins which vary in width on the pulse-height scale are shown at the top. The position of the right edge of the i^{th} bin is calculated by substituting $x = i + \frac{1}{2}$ in formula (2.2.1). The new bins of constant width Δy are shown at the bottom. Counts from the channel bins are transferred to the bottom bins in direct proportion to the corresponding overlap. If counts in the original spectrum are given by n_i, then according to Figure 2 the counts in the j^{th} bin of the new spectrum become

$$m_j = \frac{y_i - (j-1)\Delta y}{y_i - y_{i-1}} \, n_i + \frac{j\Delta y - y_i}{y_{i+1} - y_i} \, n_{i+1} . \qquad 2.2.2$$

In this derivation it was assumed that all pulses spanned by the width of an original channel occurred with equal probability. When the spectrum has a pronounced slope in the region considered this assumption is not strictly true since the rate of pulse arrival differs at the two channel sides. However, if this effect is too pronounced,

FIGURE 2

CORRECTION FOR ANALYZER NONLINEARITY



much available information is lost due to channel integration and the need for a new experiment with finer channel mesh would be indicated.

2.3 Numerical Methods of Matrix Inversion

This section is not essential to the main development presented in this thesis since the methods proposed are largely dependent on response matrices which can be inverted in closed form. However, for the sake of completeness, it is deemed advisable to include a short discussion of some numerical inversion methods. A short study of inversion difficulty as a function of matrix size is also provided. This latter consideration will perhaps serve to illustrate the advantages gained by having the inverse matrix in closed form.

In this section it is being assumed that the inverse matrix $\underline{R}^{-1}$ exists. A necessary and sufficient condition for this to be true is that matrix R have rank N (i.e. all N columns of the matrix constitute a set of linearly independent N-dimensional vectors). An equivalent necessary and sufficient condition is that $\det(\underline{R}) \neq 0$. The response matrices encountered here generally satisfy these conditions so that their inverses can be calculated. Section 2.5 will cover a few special types of response matrices which under certain conditions become singular.

The subject of matrix inversion has received attention in numerous volumes (see for example references [12,13,14,15,16]). A number of papers[17,18,19] have discussed the subject from the point of view of electronic computer application.

Numerical methods of matrix inversion can be divided into two broad classes - direct methods and iterative methods. The direct methods

obtain the required solution in a finite number of operations, whereas iterative methods only approach the true solution with each iteration. After a certain number of iterations the remaining corrections may become insignificant and the process is then terminated. For large size matrices the iterative methods are often more economical in effort and may yield more accurate answers.

Perhaps the simplest of direct methods is the Gauss[9] method of systematic elimination. The process is analogous to solution of linear equations by progressive elimination of unknown variables. One wishes to find the inverse matrix $\underline{R}^{-1}$ such that

$$\underline{R}^{-1}\,\underline{R} = \underline{I}\,, \qquad 2.3.1$$

where $\underline{I}$ is an N by N size identity matrix having elements equal to unity along the principal diagonal and zero elsewhere. Suppose that we divide the matrices $\underline{R}^{-1}$ and $\underline{I}$ into N row vectors and denote the resulting vectors by $\underline{R}_i^{-1}$ and $\underline{I}_i$ respectively. Then for each value of index i we have a set of N linear simultaneous equations

$$\underline{R}_i^{-1}\,\underline{R} = \underline{I}_i\,, \qquad 2.3.2$$

which can be written out in the detailed form

$$\begin{aligned} r_{i1}^{-1}\, r_{11} + r_{i2}^{-1}\, r_{21} + \ldots + r_{iN}^{-1}\, r_{N1} &= \delta_{i1} \\ r_{i1}^{-1}\, r_{12} + r_{i2}^{-1}\, r_{22} + \ldots + r_{iN}^{-1}\, r_{N2} &= \delta_{i2} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots& \\ r_{i1}^{-1}\, r_{1N} + r_{i2}^{-1}\, r_{2N} + \ldots + r_{iN}^{-1}\, r_{NN} &= \delta_{iN}\,. \end{aligned} \qquad 2.3.3$$

The symbol δ_{ij} is a Kroenecker delta having the property

$$\delta_{ij} = 0 \text{ for } i \neq j \,,$$
$$= 1 \text{ for } i = j \,. \qquad 2.3.4$$

As an initial step the first equation of group (2.3.3) is divided by r_{11}. Next it is multiplied by r_{12} and subtracted from the second equation, which gives us a new second equation with the unknown variable r_{i1}^{-1} eliminated. In a similar way we can eliminate r_{i1}^{-1} from all succeeding equations, which concludes the first stage. In the next stage we divide the new second equation by the coefficient of r_{i2}^{-1} and then proceed to eliminate this variable from all succeeding equations using the same process as described in stage one . After N stages of such calculations we end up with a triangular set of equations.

$$\begin{aligned} r_{i1}^{-1} + r_{i2}^{-1} a_{21} + r_{i3}^{-1} a_{31} + \cdots + r_{iN}^{-1} a_{N1} &= b_{i1} \\ r_{i2}^{-1} + r_{i3}^{-1} a_{32} + \cdots + r_{iN}^{-1} a_{N2} &= b_{i2} \qquad 2.3.5 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ r_{iN}^{-1} &= b_{iN} \,. \end{aligned}$$

The coefficients b_{ik} are zero for $k < i$ since up to the i^{th} stage there is no contribution from the Kroenecker deltas.

It is interesting to note the number of multiplications and divisions performed. These operations require a greater effort than subtractions and hence provide an indication of the over-all difficulty. For the moment we shall consider only the operations performed on the left side of the equality sign. In the k^{th} stage we are operating on (N-k+1) equations. There are (N-k) numerical divisions by the coefficient of r_{ik}^{-1} . The subsequent eliminations will require (N-k) multiplications performed for each of the (N-k) equations. Thus the total

number of operations performed in the k^{th} stage is

$$M_k = (N - k) + (N - k)^2.$$

Summing the operations of all stages we obtain the result

$$M_T = \sum_{k=1}^{N} M_k$$

$$= \frac{N}{3}(N^2 - 1). \qquad 2.3.6$$

These M_T operations define all coefficients a_{mn} which need not be recalculated.

Calculation of coefficients b_{ij} on the right side of equality signs will require additional operations the number of which will depend on index i. Consider the k^{th} stage of elimination, where $k \geqslant i$. There will be one division and (N-k) multiplications. Summing over all stages with $k \geqslant i$ we get the number of operations

$$N_i = \sum_{k=i}^{N} (N - k + 1)$$

$$= \frac{i^2}{2} - (N + \frac{3}{2})i + (N + 1)(\frac{N}{2} + 1).$$

To obtain the full inverse of matrix $\underline{R}$ we shall need to obtain a triangular form akin to (2.3.5) for each value of index i. Therefore we sum the N_i's with the result

$$N_T = \sum_{i=1}^{N} N_i$$

$$= N(\frac{N^2}{6} + \frac{N}{2} + \frac{1}{3}) \qquad 2.3.7$$

Hence the total number of multiplications and divisions required to obtain the N triangular sets (2.3.5) is given by

$$T = M_T + N_T$$

$$= \frac{N^2}{2}(N + 1) . \qquad 2.3.8$$

The final calculation of inverse matrix elements is performed by a process of back-substitution. In equations (2.3.5) the element r_{ik}^{-1} is given by

$$r_{ik}^{-1} = b_{ik} - \sum_{\ell=k+1}^{N} r_{i\ell}^{-1} a_{\ell k} ,$$

which requires (N - k) multiplications. For all values of k the number of multiplications is

$$B_i = \sum_{k=1}^{N} (N-k)$$

$$= \frac{N}{2}(N-1) .$$

This number is the same for all N sets of equations (2.3.5), so that the total back-substitution effort can be represented by

$$B = \frac{N^2}{2}(N - 1) . \qquad 2.3.9$$

We are now in a position to write down the overall effort factor for complete matrix inversion

$$E = T + B$$

$$= N^3 . \qquad 2.3.10$$

This cubic relationship presents a serious practical limitation on the size of matrices which can be inverted by the direct methods.

In the foregoing derivation it was assumed that the divisor used

in the k^{th} stage of elimination was non-zero. When this condition does not hold it is necessary to use a modified method. Even when the divisor is finite but small the results may be unsatisfactory due to accumulation of large rounding-off errors. A way out of these difficulties is to search each equation for the largest coefficient and then use it as the divisor for the purpose of elimination. This method is sometimes referred to as the "Gauss' Method with Selection of the Pivotal Element".

Another method popular in automatic computer calculations is the Gauss-Jordan method of elimination. During the k^{th} stage of elimination the unknown r_{ik}^{-1} is eliminated from all the preceding as well as the succeeding equations. Thus the equations (2.3.3) are reduced to diagonal form and no back-substitution is necessary. However a larger number of operations is required during the process of elimination with the result that the total number of multiplications and divisions remains N^3.

There are various other direct methods of inversion which shall not be discussed here. In particular, when the matrix is symmetrical there are special techniques available which provide a greater economy in effort. These techniques are amply discussed in the literature and the interested reader may find them in the references quoted above.

The special case of a triangular matrix shall be considered since some of the response matrices in this thesis take that form. Consider the upper-triangular matrix $\underline{R}$ with elements $r_{ij} = 0$ for $i > j$. In this case the matrix equation (2.3.1) can be represented by the set of algebraic equations

$$\sum_{i=1}^{j} r_{ki}^{-1} \; r_{ij} = \delta_{kj} \; . \qquad 2.3.11$$

Note that in equation (2.3.11) when $k > j$ we also have $k > i$. Therefore the right hand side is zero and the equation is satisfied by setting

$$r_{ki}^{-1} = 0 \qquad \text{for } k > i.$$

In other words, the inverse matrix is also upper-triangular and equation (2.3.11) can be replaced by

$$\sum_{i=k}^{j} r_{ki}^{-1} \; r_{ij} = \delta_{kj} \; . \qquad 2.3.12$$

Triangular matrices are particularly easy to invert since the process can be carried out row by row independently of other rows. Consider a general k^{th} row of the inverse matrix. Using expression (2.3.12) we can write a set of $(N - k + 1)$ equations which already are in triangular form

$$\begin{aligned}
(j=N) \quad & r_{kk}^{-1} \; r_{kN} + r_{kk+1}^{-1} \; r_{k+1\,N} + \cdots + r_{kN}^{-1} \; r_{NN} = 0 \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \qquad 2.3.13 \\
(j=k+1) \quad & r_{kk}^{-1} \; r_{kk+1} + r_{kk+1}^{-1} \; r_{k+1\,k+1} = 0 \\
(j=k) \quad & r_{kk}^{-1} \; r_{kk} = 1 \; .
\end{aligned}$$

Thus the inverse elements of k^{th} row can be calculated by a simple process of back-substitution. Evaluation of r_{kk}^{-1} will require one division. The next term $r_{k\,k+1}^{-1}$ will require one multiplication and one division. It is easily seen that the number of operations increases by

one in each step. Hence the total number of operations required is given by

$$B_k = \sum_{\ell=1}^{(N-k+1)} \ell$$

$$= \frac{k^2}{2} - (N + \frac{3}{2})k + \frac{N}{2}(N + 3) + 1.$$

The number of operations for all rows of the inverse matrix is obtained by summing over k, i.e.

$$B = \sum_{k=1}^{N} B_k$$

$$= N(\frac{N^2}{6} + \frac{N}{2} + \frac{1}{3}) . \qquad 2.3.14$$

For large values of N this effort factor is about one sixth of the factor for general matrix inversion.

In many response matrices the diagonal elements are large in magnitude compared to other elements. Under these conditions the iterative methods may have good convergence properties[15] and lead to fairly rapid calculation of the inverse matrix. An example of iterative techniques is provided by the Gauss-Seidel method discussed below.

Equations (2.3.3) can be rewritten in the following form:

$$r_{i1}^{-1} = \frac{1}{r_{11}} (\delta_{i1} - r_{i2}^{-1} r_{21} - \cdots - r_{iN}^{-1} r_{N1})$$

$$r_{i2}^{-1} = \frac{1}{r_{22}} (\delta_{i2} - r_{i1}^{-1} r_{12} - \cdots - r_{iN}^{-1} r_{N2}) \qquad 2.3.15$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$r_{iN}^{-1} = \frac{1}{r_{NN}} (\delta_{iN} - r_{i1}^{-1} r_{12} - \cdots - r_{iN-1}^{-1} r_{N-1\,N}) .$$

One starts with initial guesses for the unknowns r_{ij}^{-1} and substitutes them at the right hand sides of equations (2.3.15). The resultant values at the left are the first approximations $(r_{ij}^{-1})_1$. In the second step these new values can be substituted at the right sides, yielding second approximations $(r_{ij}^{-1})_2$. This process is repeated until the changes become insignificant. The last estimates $(r_{ij}^{-1})_n$ give the i^{th} row elements of the inverse matrix.

2.4 Application to β-Ray Spectra Obtained with Organic Scintillators

One of the first attempts to correct β-ray spectra for resolution effects was made by Freedman et al[6]. Their studies revealed that the response to monoenergetic electrons was of the form shown in Figure 3 where the peak is associated with the totally absorbed β-rays. In addition to the peak there is a constant-height tail which arises from partial energy losses due to scattering from the crystal[20].

The method proposed by Freedman et al[6] was an iterative procedure. One makes a first estimate $\underline{T}_1$ for the true spectrum and then obtains a second estimate $\underline{T}_2$, from the relation

$$\underline{T}_2 = \underline{T}_1 + \underline{M}' - \underline{R}\,\underline{T}_1 , \qquad 2.4.1$$

where $\underline{M}'$ is the observed spectrum and $\underline{R}$ is the response matrix. This process is continued using the algorithm

$$\underline{T}_{m+1} = \underline{T}_m + \underline{M}' - \underline{R}\,\underline{T}_m \ . \qquad 2.4.2$$

By using progressive back-substitution we can express this last equation in terms of the first estimate $\underline{T}_1$. The result is

$$\underline{T}_{m+1} = (\underline{I} - \underline{R})^m \underline{T}_1 + \underline{R}^{-1}\left[\underline{I} - (\underline{I}-\underline{R})^m\right]\underline{M}' \ . \qquad 2.4.3$$

Here $\underline{I}$ is the identity matrix and the inverse $\underline{R}^{-1}$ is assumed to exist.

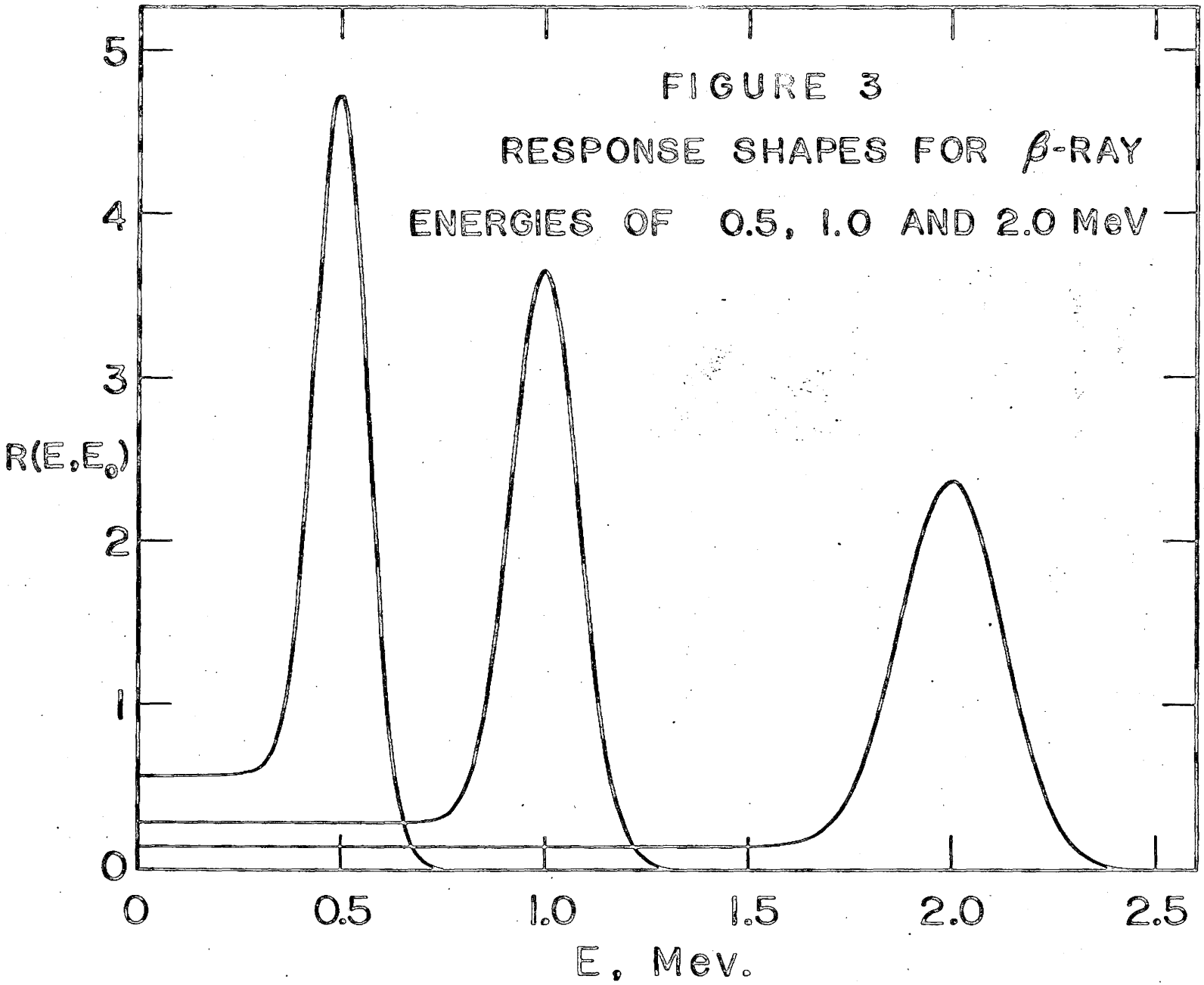


FIGURE 3

RESPONSE SHAPES FOR β-RAY ENERGIES OF 0.5, 1.0 AND 2.0 MeV

If $\lim_{m \to \infty} (\underline{I} - \underline{R})^m = 0$,

then $\lim_{m \to \infty} \underline{T}_{m+1} = \underline{R}^{-1} \underline{M}'$

and successive estimates approach the unfolded spectrum. Freedman et al[6] found that for a good initial estimate only two to four iterations were required to obtain a spectrum $\underline{T}_{m+1}$ such that $\underline{R}\,\underline{T}_{m+1}$ reproduced the observed spectrum $\underline{M}'$ within about 2%.

This iterative procedure has the advantage that the inverse matrix $\underline{R}^{-1}$ does not have to be evaluated directly. As indicated in the last section, the inversion of matrices having sizes corresponding to data fields of 200 or more channels can be quite difficult. Even if the full response matrix could be inverted there are reasons why its use might be undesirable. The response shapes shown in Figure 3 contain the Gaussian resolution function. As will be shown in section 2.5, correction for this effect involves an inverse matrix with very large positive and negative elements. Consequently the statistical uncertainties originally present in $\underline{M}'$ can be grossly magnified.

When the iterative formula (2.4.2) is used to correct for the Gaussian response alone, the statistical deviations show a roughly linear increase with each iteration. If the results finally converge to the required unfolded spectrum then this expansion of statistical deviations can be expected to continue until a levelling-off is reached near values corresponding to multiplication by the inverse matrix of Gaussian response. It is therefore important to obtain a very good first estimate which would allow the iterations to be kept down to a minimum.

The method of response correction presented here was first described in a paper by Slavinskas, Kennett and Prestwich[21]. It is

based on writing the full response matrix as a product of an upper-triangular matrix and a Gaussian matrix. The triangular matrix represents the tail part of the response and can be inverted in closed form. After correction for this response effect is made, the Fermi-plot of the β-ray spectrum is found to be very nearly linear. This allows a very good first guess T_1 of the true spectrum in formula (2.4.1). Thus only one iteration is required to correct for the remaining Gaussian resolution. It will be shown later in this section that multiplication by the inverse of the triangular matrix produces only a small effect on the statistical deviations.

The response functions shown in Figure 3 can be represented by the convolution of a Gaussian $G(E,E_o)$ with a second function $L(E,E_o)$, namely

$$R(E,E_o) = \int L(E,E_1)G(E_1,E_o)dE_1 \,, \qquad 2.4.4$$

where

$$G(E,E_o) = \frac{1}{\sqrt{2\pi}\,\sigma_o} \exp\left[\frac{-(E-E_o)^2}{2\sigma_o^2}\right] \qquad 2.4.5$$

and

$$L(E,E_o) = \frac{k_o + \delta(E-E_o)}{k_o E_o + 1} \qquad \text{for } E \leq E_o \qquad 2.4.6$$

$$= 0 \qquad \text{for } E > E_o \,.$$

Parameter k_o is related to the tail height of response to β rays with energy E_o and σ_o is the standard deviation of the Gaussian resolution at the same energy (full width at half-maximum $\approx 2.35\,\sigma_o$). All three functions are normalized to have unit area.

Using the digitizing approximations discussed in section 2.1 we can replace the response functions by matrices of rank N. $L(E,E_o)$ is replaced by an upper-triangular matrix $\underline{L}$ in which the rows are indexed by values of E and the columns indexed by E_o. Similarly we can replace $G(E,E_o)$ by a matrix $\underline{G}$. The overall response matrix $\underline{R}$ can then be written in the form

$$\underline{R} = \underline{L}\,\underline{G} \,. \qquad 2.4.7$$

In accordance with equation (2.1.10) we can express the observed spectrum as

$$\underline{M}' = \underline{L}\,\underline{G}\,\underline{T} + \underline{S}' \,, \qquad 2.4.8$$

where $\underline{T}$ is the true spectrum without response effects and $\underline{S}'$ is a random vector of statistical deviations. Operation on $\underline{M}'$ by the inverse matrix $\underline{L}^{-1}$ will result in

$$\underline{G}\,\underline{T} + \underline{L}^{-1}\,\underline{S}' \,.$$

Apart from statistical deviations this is just the true spectrum $\underline{T}$ as "seen" through a Gaussian resolution. Since β-ray spectra are continuous and usually have widths large compared to that of Gaussian resolution, the spectrum $\underline{G}\,\underline{T}$ is very similar to the true spectrum $\underline{T}$.

Evaluation of inverse matrix $\underline{L}^{-1}$ is straight-forward and can be written in closed form. This is a great advantage since we are no longer limited to small arrays and one can obtain the inverse for almost any size data field. The delta function in equation (2.4.6) represents β-rays which deposit their full energy within the crystal. If we define the area under this full-energy peak by α_j we can write

$$\alpha_j = \frac{1}{k_j E_j + 1} \,. \qquad 2.4.9$$

The area under the tail is given by the complement

$$\beta_j = 1 - \alpha_j \,. \qquad 2.4.10$$

With these definitions matrix $\underline{L}$ has the form

$$\underline{L} = \begin{Vmatrix} \alpha_1 & \beta_2 & \frac{\beta_3}{2} & \cdots & \frac{\beta_N}{N-1} \\ 0 & \alpha_2 & \frac{\beta_3}{2} & \cdots & \frac{\beta_N}{N-1} \\ 0 & 0 & \alpha_3 & \cdots & \frac{\beta_N}{N-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \alpha_N \end{Vmatrix} \,. \qquad 2.4.11$$

Derivation of the inverse matrix is given in Appendix III. If we define the quantity

$$\gamma_j = 1 - \frac{\beta_j}{(j-1)\alpha_j} \,, \qquad 2.4.12$$

where $j > 1$, then the inverse matrix elements are given by

$$\begin{aligned} \ell_{ii}^{-1} &= \frac{1}{\alpha_i} \,, \\ \ell_{ij}^{-1} &= 0 \qquad \text{for} \quad i > j \,, \\ \ell_{ij}^{-1} &= \frac{1}{\alpha_i} (\gamma_j - 1) \prod_{k=i+1}^{j-1} \gamma_k \qquad \text{for} \quad i < j. \end{aligned} \qquad 2.4.13$$

Trial and error calculations with experimental spectra having end-point energies up to 2.27 MeV indicated that linear Fermi-plots can be obtained by using constant values for parameters α_j and β_j. This fact implies that the fraction of electrons scattered outside

the crystal is independent of incident energy. Experimental support for this conclusion is provided by the work of Bothe[22] who found that for a given scattering material the back-scattering coefficient of low-energy electrons is constant. The values used in the succeeding calculations are $\alpha = 0.72$ and $\beta = 0.28$.

From equation (2.4.12) it is apparent that all values of γ_j are less than unity. This means that in any given row of $\underline{L}^{-1}$ the diagonal element $1/\alpha = 1.39$ has the largest numerical value and all elements to the right progressively become smaller. Therefore the magnification of statistical uncertainties produced by the operator $\underline{L}^{-1}$ is small enough as to be of no practical significance.

Sample calculations of resolution correction were performed on three essentially single group spectra of ^{90}Y, ^{91}Y, ^{143}Pr and a mixture of ^{90}Sr with ^{90}Y in secular equilibrium. All experimental spectra were obtained under identical conditions with the sources mounted 1 cm from a 5 x 5 cm type NE-102 organic scintillator coupled to an EMI 9536 photo-multiplier tube. The pulses were fed to a DD2 amplifier and recorded in a multi-channel analyzer. Sources were impregnated into filter paper and had a thickness corresponding to approximately 3 mg/cm^2.

Before applying resolution corrections the spectra were corrected for analyzer nonlinearity by the use of the technique described in section 2.2. The parameters A and B in formula (2.2.1) were determined from internal conversion lines of ^{137}Cs, ^{207}Bi and initial estimates of the β-ray spectra end-points. All β-ray spectra were then converted from the original channel base to a linear energy scale with bins

ΔE = 10 keV wide. Values used for the two parameters were

$$A = 9.444 \text{ keV/channel}$$

and $$B = 12.0 \text{ channels} .$$

^{143}Pr has a $\frac{5}{2}+$ ground state which decays by β^- to the $\frac{7}{2}-$ ground state in ^{143}Nd[23]. Thus we have a change in spin $\Delta J = 1$ and a parity change, the combination of which corresponds to a first forbidden transition[24]. Many first forbidden transitions have spectra corresponding to those of allowed transitions[25] and the case of ^{143}Pr appears to fall in this class. The energy distribution of allowed transition β-rays is given by

$$T(E) = D\ PW\ F(Z,E)(E_m - E)^2 , \qquad 2.4.14$$

where D is a constant and F(Z,E) is a Coulomb effect correction factor dependent on the atomic number Z and electron kinetic energy E. Numerical values of F(Z,E) which included electron screening effects were obtained from reference[26]. The factor F(Z,E) is essentially constant at high β-ray energies and has little effect on the shape of the spectrum. Other factors in formula (2.4.14) are the electron momentum

$$P = \frac{1}{c}\sqrt{E^2 + 2Em_oc^2}$$

and the total electron energy

$$W = E + m_oc^2 .$$

Parameter E_m is the total transition energy and thus represents the upper energy limit of the β-ray. When $E < E_m$ the remaining energy is carried away by the neutrino (nuclear recoil energy can usually be neglected).

It is customary to represent the data in the form of so-called Fermi (or Fermi-Kurie) plots. These can be obtained by plotting $\sqrt{T/(PWF)}$ as a function of E. The result should be a straight line with an intercept $E = E_m$ on the abscissa. Thus one obtains a convenient indication of transition energy E_m.

The β^- transitions of ^{90}Sr, ^{90}Y and ^{91}Y are characterized by $\Delta J = 2$ and a change in parity. Thus they fall in the class of first forbidden transitions. On account of the spin change $\Delta J = 2$ the spectra are expected to have unique shapes[25] which are distinguished from the allowed shapes (2.4.14) by having an extra factor S. The shape factor was calculated from the relation

$$S = Q^2 + \lambda_1 P^2 , \qquad 2.4.15$$

where Q is the neutrino momentum and λ_1 is an energy-dependent parameter (close to unity) tabulated by Kotani and Ross[27].

Langer et al[28] have reported the necessity of an additional shape correction for ^{90}Y and ^{91}Y. Based on their precise magnetic spectrometer measurements, this additional factor takes the form

$$C = 1 + \frac{b}{w} , \qquad 2.4.16$$

where w is the total electron energy in $m_o c^2$ units and b is an empirical parameter having a numerical value between 0.2 and 0.4. In the calculations presented here the intermediate value b = 0.3 was used. Due to these additional factors the expression $\sqrt{T/(PWFSC)}$ was used to calculate Fermi plots for ^{90}Sr, ^{90}Y and ^{91}Y.

Fermi plots of ^{143}Pr, ^{91}Y and ^{90}Y are shown in Figures 4, 5 and 6 respectively. Each Figure contains three plots which illustrate the various stages of resolution correction. Plots labelled (a)

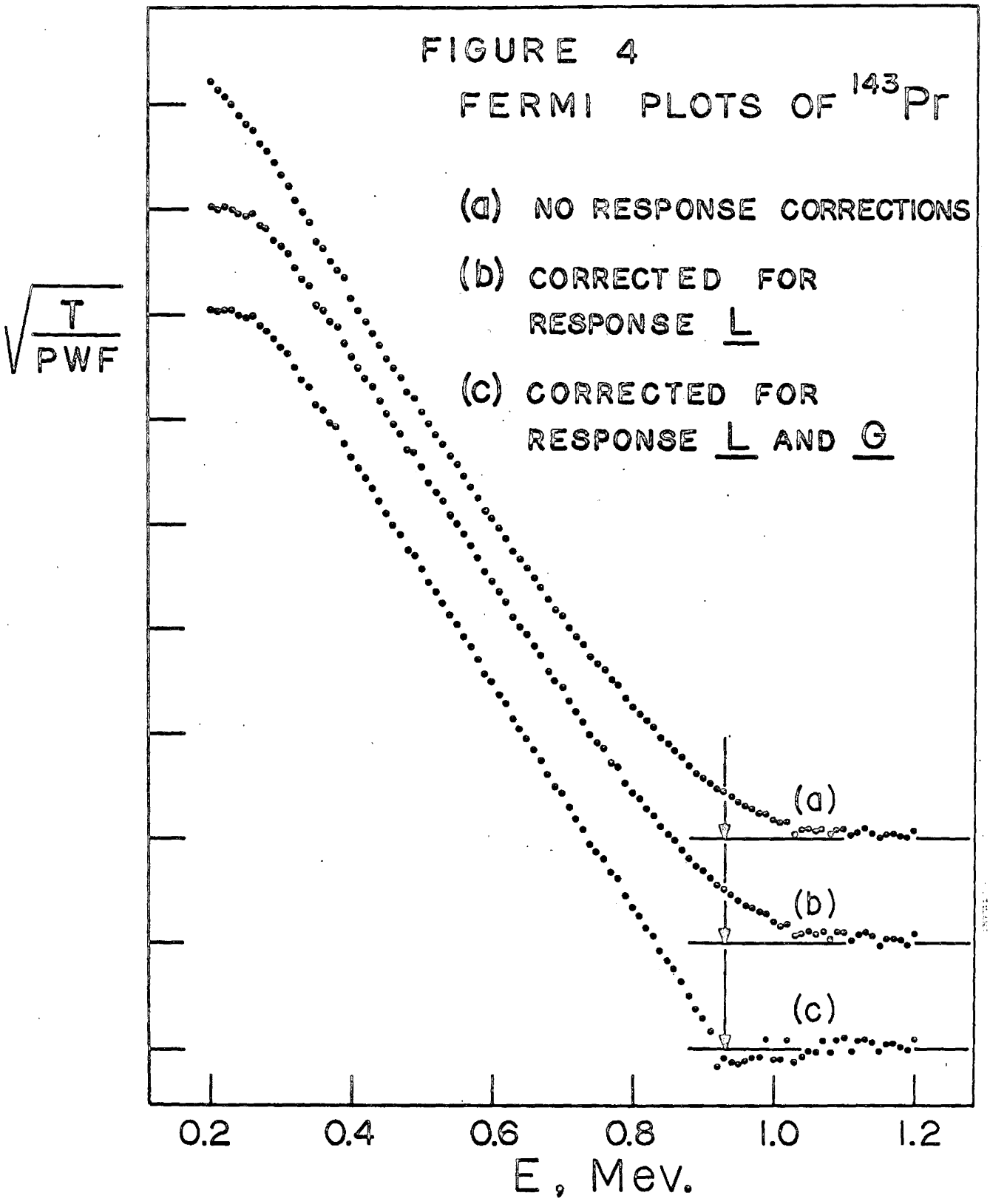


FIGURE 4
FERMI PLOTS OF ^{143}Pr
(a) NO RESPONSE CORRECTIONS
(b) CORRECTED FOR RESPONSE L
(c) CORRECTED FOR RESPONSE L AND G
$\sqrt{\frac{T}{PWF}}$
(a)
(b)
(c)
0.2
0.4
0.6
0.8
1.0
1.2
E, Mev.

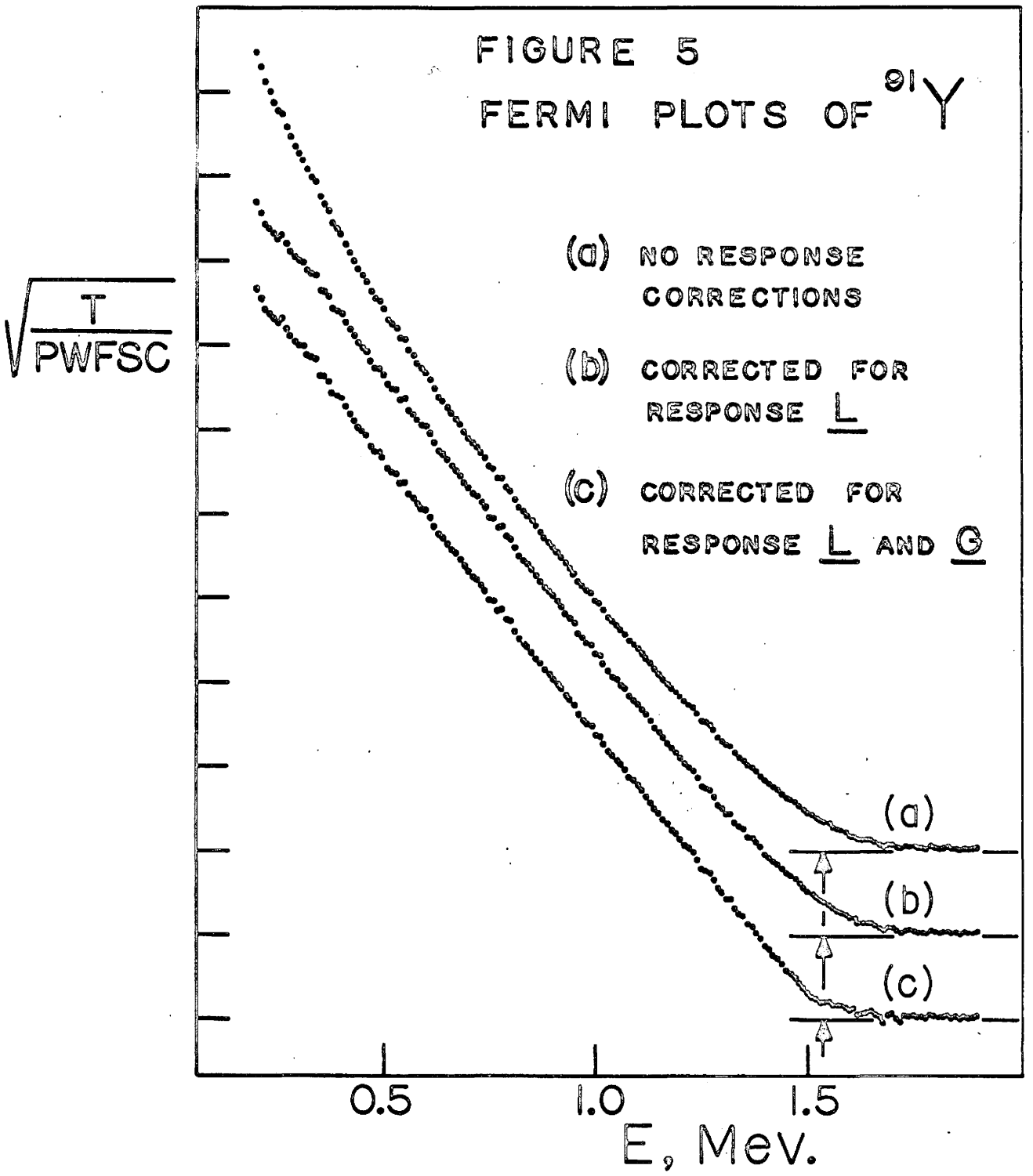


FIGURE 5
FERMI PLOTS OF ^{91}Y
(a) NO RESPONSE CORRECTIONS
(b) CORRECTED FOR RESPONSE L
(c) CORRECTED FOR RESPONSE L AND G
$\sqrt{\frac{T}{PWFSC}}$
(a)
(b)
(c)
0.5
1.0
1.5
E, Mev.

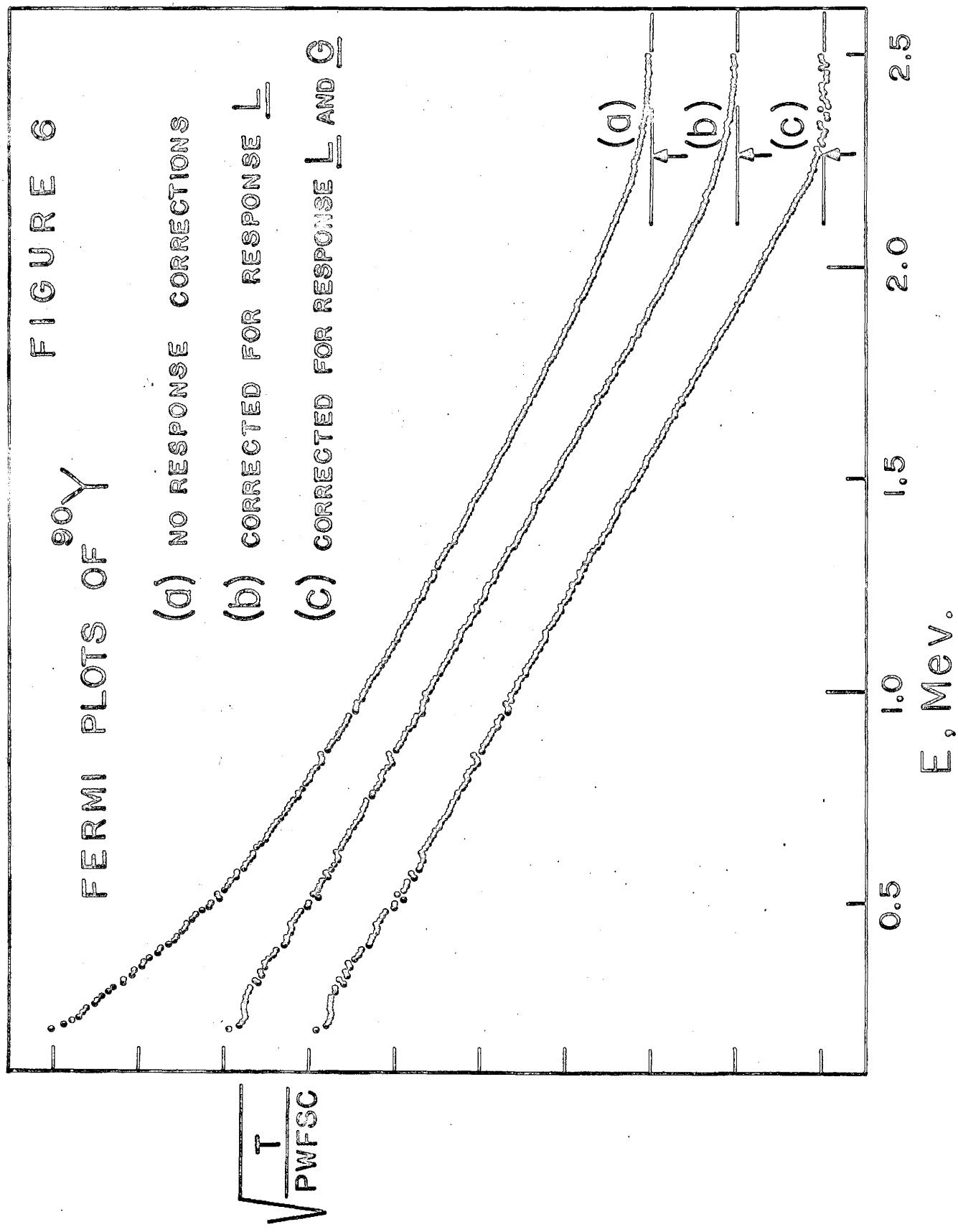


FIGURE 6
FERMI PLOTS OF ^{90}Y
(a) NO RESPONSE CORRECTIONS
(b) CORRECTED FOR RESPONSE L
(c) CORRECTED FOR RESPONSE L AND G
(a)
(b)
(c)
0.5
1.0
1.5
2.0
2.5
E, Mev.
$\sqrt{\frac{T}{PWFSC}}$

were calculated before application of any resolution corrections. They tend to show an upward curvature for decreasing energies. This effect is caused mainly by the low-energy tail in the scintillator response and makes the determination of end-point energy uncertain. Fermi plots (b) were obtained after multiplying the spectra by inverse matrix $\underline{L}^{-1}$. A considerable improvement in linearity is evident, so that accurate end-point determinations can now be made. The remaining slight curvature near maximum energies is presumably caused by Gaussian resolution effects. This residual has the effect of slightly modifying the apparent end-point energy. Calculations with theoretical model spectra gave an estimate of the error magnitude. A number of theoretical spectra with different maximum energies were multiplied by the Gaussian resolution matrix. Fermi plots were then calculated and "best" straight lines were drawn through the resulting points. It was found that these lines gave correct intercepts with the ordinate axis but underestimated the end-point energy by about 8 keV.

Although the two corrected intercepts thus obtained are sufficient to determine the required experimental spectrum parameters, further calculations were applied in an attempt to correct for the Gaussian resolution. A quadratic variation of σ_o^2 with energy was assumed. After a few trial calculations the particular form adopted was

$$\sigma_o^2 = 0.0027E^2 + E + 3000 , \qquad 2.4.17$$

where both σ_o and E are expressed in keV. One iteration

as defined by equation (2.4.1) was applied to each of the three spectra. The first estimates $\underline{T}_1$ were calculated by the use of corrected intercepts of Fermi plots (b). Results are shown in part (c) of Figures 4, 5 and 6. Linearity near the end-points is seen to be considerably improved. Some points lie below the E-axis because imaginary values (obtained where the unfolded spectrum was negative) are plotted as negative numbers.

A summary of end-point energy results is given in Table I. The largest deviation from Nuclear Data Sheet values[29] (column 2) occurs for the end-point of ^{90}Sr, the difference being 20 keV. Errors in other isotopes are considerably smaller. This is partly caused by the fact that the experimental end-points were included in the determination of the energy scale. Since all spectra were obtained under identical experimental conditions, the energies should be internally consistent. This requirement is satisfied within the estimated uncertainty of $\pm$ 20 keV.

The two-component spectrum of ^{90}Sr and ^{90}Y was analyzed for relative intensities as well as for end-point energies. Intensities are expected to be equal since the system was in secular equilibrium with the 28 year ^{90}Sr decaying to the 64 hour ^{90}Y. Fermi plots of the results are given in Figure 7. Part (a) shows the total spectrum after multiplication by $\underline{L}^{-1}$. The long linear part of the ^{90}Y component provided good estimates of the intercepts. These were used to calculate a theoretical ^{90}Y spectrum which was multiplied by the Gaussian response matrix and then subtracted from the total. A Fermi plot of the remaining ^{90}Sr spectrum is shown in part (b). Finally, part (c)

TABLE I

End-Point Energies

Isotope	E_m (MeV)	
	Nuclear Data Sheets (29)	Present Work
^{90}Sr	0.54	0.56 ± 0.02
^{143}Pr	0.93	0.93 ± 0.02
^{91}Y	1.54	1.54 ± 0.02
^{90}Y	2.27	2.27 ± 0.02

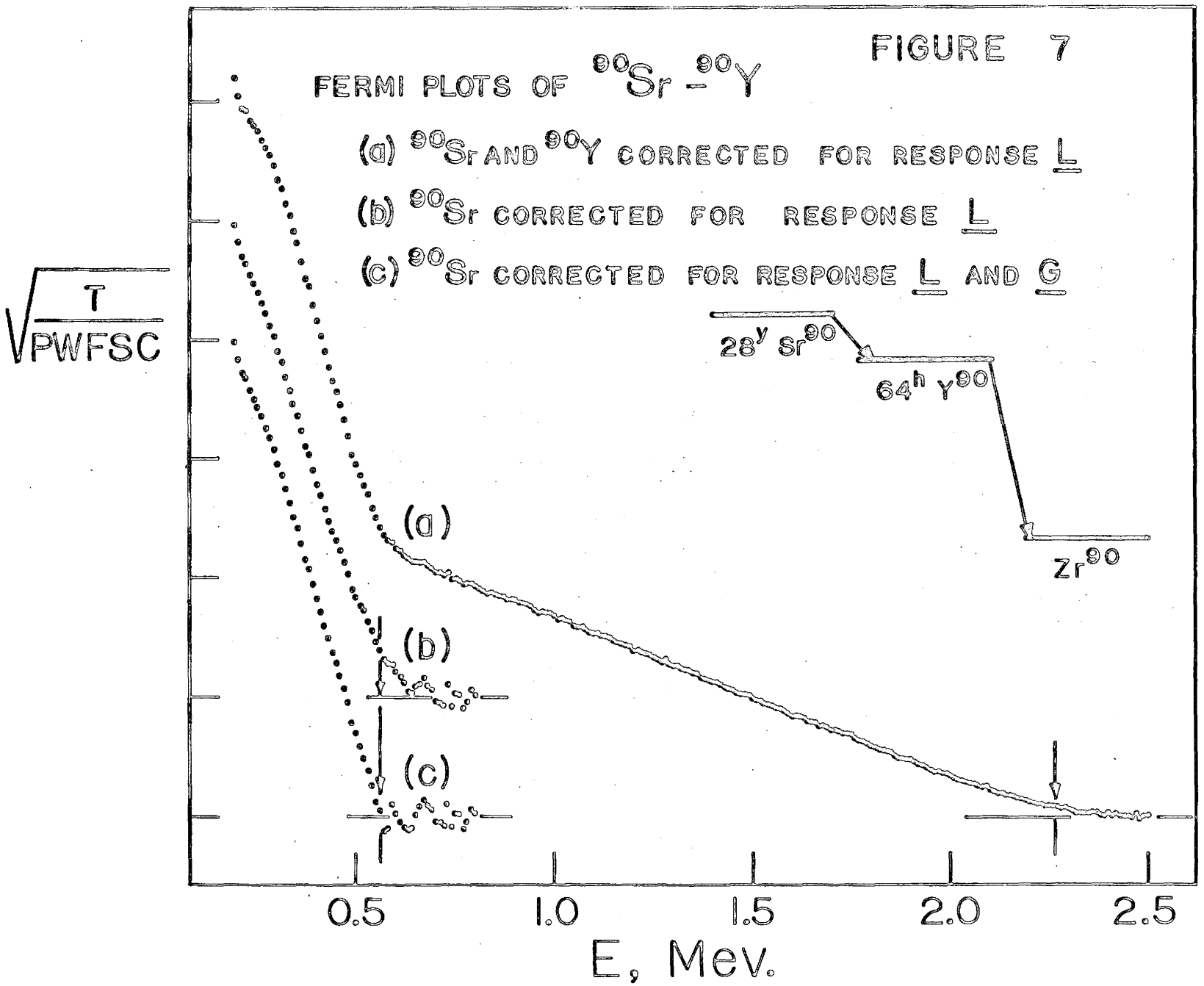


FIGURE 7
FERMI PLOTS OF $^{90}Sr - ^{90}Y$
(a) ^{90}Sr AND ^{90}Y CORRECTED FOR RESPONSE L
(b) ^{90}Sr CORRECTED FOR RESPONSE L
(c) ^{90}Sr CORRECTED FOR RESPONSE L AND G
$\sqrt{\frac{T}{PWFSC}}$
$28^{y}\ Sr^{90}$
$64^{h}\ Y^{90}$
Zr^{90}
(a)
(b)
(c)
0.5
1.0
1.5
2.0
2.5
E, Mev.

shows the Fermi plot of ^{90}Sr after correction for Gaussian resolution. Experimental intercepts obtained for ^{90}Sr and ^{90}Y were used to calculate the corresponding theoretical spectra. These were then summed over all channels to obtain a measure of intensities. The results gave a total of $(1.02 \pm .10) \times 10^6$ counts for ^{90}Sr and $(1.04 \pm 0.05) \times 10^6$ counts for ^{90}Y, which indicates good agreement with the expected equal intensity condition. In principle, this procedure of "stripping" could be applied to a spectrum with any number of components. The limitation, of course, is that the end-points have separations sufficiently large to leave recognizable straight line segments in the Fermi plot after the operation $\underline{L}^{-1}$.

It is shown in Appendix III that the number of multiplicative and divisional operations required to invert $\underline{L}^{-1}$ has a quadratic dependence on matrix size N. This is a considerable improvement over the cubic dependence (2.3.14) which obtains for general triangular matrices. Thus one can perform rapid analysis even when the data fields are quite large. For example, using the IBM 7040 electronic computer, corrected Fermi plots were calculated in 2 min. and 16 sec. for a 64 by 64 β-γ coincidence spectrum.

2.5 Application to Energy-Independent Response Shapes

In this section we shall consider response functions of the form $R(x-y)$, which implies a fixed response shape for all incident energies y. Nuclear radiation detectors usually do not satisfy this requirement; for example, the Gaussian resolution width in spectra obtained with NaI(Tℓ) detectors varies roughly as the square root of incident γ-ray energy. However, it is possible that future develop-

ments of analyzers[30] will provide the option of storing the spectra on a channel base proportional to $\sqrt{E}$. Obtaining thus an approximately constant resolution width would result in most efficient use of the available memory storage capacity (or number of channels). In some instances (e.g. lithium-drifted Ge detectors) the variation in resolution width is very small over the energy range covering one peak. With some modifications the method of this section might still be used to apply resolution corrections for such spectra.

The technique described below need not be limited to nuclear spectra alone. There is a wide class of other applications where some signal may have suffered distortion due to undesirable frequency component filtering - or equivalently convolution - effects. Such may be the case when data are transmitted, for example, over long telephone lines. In some instances a short pulse may be purposefully distorted into a long wave train, which technique permits transmission of increased energy when the peak power of a transmitting device is limited. Examples of this technique are found in some radar systems[31] and in seismic exploration methods using surface sources[32]. When this extended signal is received, some method of de-convolution is usually required in order to obtain the original short pulse. The present section investigates a method of de-convolution which is applicable to data having digitized form.

In the present section we shall derive a general expression for the inverse of an infinite rank response matrix $\underline{R}$ having elements r_{ij} which are dependent on (i-j) only. This matrix consists of identical columns which are shifted by one row with respect to their neigh-

bours. The derivation of inverse matrix elements will be accomplished by the use of Fourier transforms and the convolution theorem[33].

We start by considering the convolution integral

$$M(x) = \int_{-\infty}^{\infty} R(x-y)\ T(y)dy\ , \qquad 2.5.1$$

and then try to find an inverse function $R_{in}(x)$ such that

$$T(x) = \int_{-\infty}^{\infty} R_{in}(x-y)\ M(y)dy. \qquad 2.5.2$$

For this purpose we shall need the Fourier transform pair

$$M(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} m(\omega)\ e^{i\omega x}\ d\omega\ , \qquad 2.5.3\ a$$

$$m(\omega) = \int_{-\infty}^{\infty} M(x)\ e^{-i\omega x}\ dx. \qquad 2.5.3\ b$$

Capital and lower case letters shall be employed to represent the functions in the x-domain and the ω-domain respectively. A short derivation of the convolution theorem is given in Appendix IV. Given the convolution relationship in the x-domain

$$F_3(x) = \int_{-\infty}^{\infty} F_1(x-y)\ F_2(y)dy, \qquad 2.5.4\ a$$

the theorem states that in the ω-domain we have the product

$$f_3(\omega) = f_1(\omega)\ f_2(\omega)\ . \qquad 2.5.4\ b$$

Similarly, for convolution in the ω-domain

$$g_3(\omega) = \int_{-\infty}^{\infty} g_1(\omega-\lambda)\ g_2(\lambda)d\lambda \qquad 2.5.5\ a$$

we have

$$G_3(x) = 2\pi\, G_1(x)\, G_2(x) \,. \qquad 2.5.5\ b$$

The method of finding $R_{in}(x)$ can now be outlined as follows. Applying the convolution theorem to equations (2.5.1) and (2.5.2) we can write the products

$$m(\omega) = r(\omega)\, t(\omega) \qquad 2.5.6$$

and

$$t(\omega) = r_{in}(\omega)\, m(\omega) \,. \qquad 2.5.7$$

To satisfy these last two relations we must have

$$r_{in}(\omega) = \frac{1}{r(\omega)} \,. \qquad 2.5.8\ a$$

Thus the relation (2.5.8 a) gives us the Fourier transform of the required inverse function $R_{in}(x)$. Accordingly we can transform back to the x-domain which yields the result

$$R_{in}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\omega x}}{r(\omega)}\, d\omega \,. \qquad 2.5.8\ b$$

Often $r_{in}(\omega)$ is unbounded and the inverse function $R_{in}(x)$ does not exist. This is usually the case with continuous functions $R(x)$. As an example we can consider the Gaussian response function

$$R(x) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left[-\frac{x^2}{2\sigma^2}\right], \qquad 2.5.9\ a$$

which has the Fourier transform (see Appendix IV)

$$r(\omega) = \exp\left[-\frac{\sigma^2\omega^2}{2}\right] . \qquad 2.5.9\ b$$

It is clear that the reciprocal $1/r(\omega)$ approaches infinity together with ω; hence the value $R_{in}(0)$ as defined by (2.5.8 b) is infinite

and the desired inverse function does not exist. However, the inverse can frequently be found in matrix form when the response functions are digitized into discrete channel bins. The Gaussian response falls into this class of invertable digitized functions.

A form of digitization can be accomplished by making use of the following set of equally spaced delta functions

$$D(x) = \sum_{n=-\infty}^{\infty} \delta(x-n) \ , \qquad 2.5.10\ a$$

which, according to Appendix IV, has the Fourier transform

$$d(\omega) = 2\pi \sum_{n=-\infty}^{\infty} \delta(\omega-2\pi n). \qquad 2.5.10\ b$$

Thus the digitized response function can be written as the product

$$R^d(x) = R(x)D(x) \ . \qquad 2.5.11$$

According to the scheme outlined above we will try to find the inverse function $R^d_{in}(x)$ by calculating the Fourier transform of $R^d(x)$ and then transforming the reciprocal $1/r^d(\omega)$ back to the x-domain. Proceeding with the execution of the first step we write the Fourier transform

$$r^d(\omega) = \int_{-\infty}^{\infty} R(x)\, e^{-i\omega x} \sum_{n=-\infty}^{\infty} \delta(x-n)dx$$

$$= \sum_{n=-\infty}^{\infty} R(n)e^{-i\omega n} \ . \qquad 2.5.12$$

This expression contains a summation which may consist of many terms when R(x) covers a wide range of integral values of x. For such cases it might be more convenient to express $r^d(\omega)$ in another form which can be particularly useful when the function $r(\omega)$ is easily obtained.

Applying the convolution theorem to expression (2.5.11) we can write

$$r^d(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} r(\omega-\lambda)\ 2\pi \sum_{n=-\infty}^{\infty} \delta(\lambda-2\pi n)\ d\lambda$$

$$= \sum_{n=-\infty}^{\infty} r(\omega-2\pi n) \qquad 2.5.13$$

From expression (2.5.13) it is clear that, although $r(\omega)$ may approach zero for large values of ω, $r^d(\omega)$ will not do so on account of the periodicity inherent in the summation. Except for special cases in which $r(2\pi n) = 0$ for all values of n, the reciprocal $1/r^d(\omega)$ will remain finite everywhere. Figure 8 illustrates the effect of digitization by showing the reciprocal functions $1/r(\omega)$ and $1/r^d(\omega)$ obtained for a Gaussian response with $\sigma = 1$.

Transformation to the x-domain is accomplished by noting that $1/r^d(\omega)$ is periodic in ω with period 2π. Thus after defining the truncated function

$$f(\omega) = \frac{1}{r^d(\omega)} \quad \text{for} \quad |\omega| < \pi$$

$$= \frac{1}{2}\frac{1}{r^d(\omega)} \quad \text{for} \quad |\omega| = \pi \qquad 2.5.14$$

$$= 0 \quad \text{for} \quad |\omega| > \pi$$

we can express

$$\frac{1}{r^d(\omega)} = \sum_{n=-\infty}^{\infty} f(\omega-2\pi n)$$

$$= \int_{-\infty}^{\infty} f(\omega-\lambda) \sum_{n=-\infty}^{\infty} \delta(\lambda-2\pi n)\ d\lambda$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\omega-\lambda)\ d(\lambda)\ d\lambda\ . \qquad 2.5.15$$

The required inverse function $R_{in}(x)$ is given by the Fourier transform of $1/r^d(\omega)$. Using (2.5.15) and the convolution theorem we can

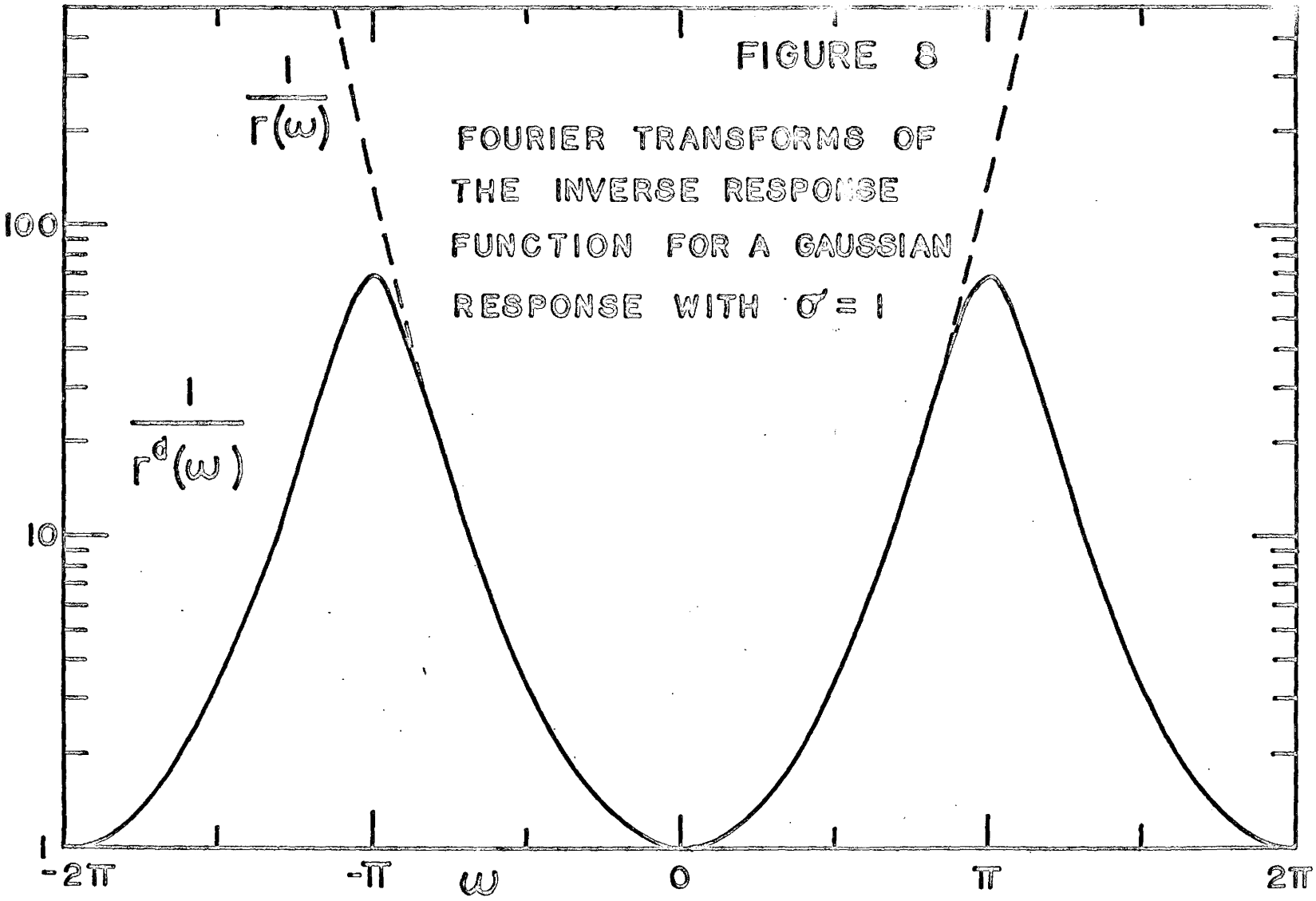


FIGURE 8
FOURIER TRANSFORMS OF THE INVERSE RESPONSE FUNCTION FOR A GAUSSIAN RESPONSE WITH σ = 1
1/Γ(ω)
1/Γ^d(ω)
100
10
1
-2π
-π
ω
0
π
2π

express the inverse function in the form of the product

$$R_{in}(x) = F(x)\ D(x)\ , \qquad 2.5.16$$

where

$$F(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\omega x}}{r^{d}(\omega)}\ d\omega\ . \qquad 2.5.17$$

The inverse function (2.5.16) is seen to contain the original digitizing function D(x). This factor is instrumental in replacing the de-convolution integral (2.5.2) by a summation so that the process becomes analogous to multiplication by an inverse matrix of infinite dimension. Carrying out the substitution in (2.5.2) we obtain

$$T(x) = \int_{-\infty}^{\infty} F(x-y) \sum_{n=-\infty}^{\infty} \delta(x-y-n)\ M(y)dy$$

$$= \sum_{n=-\infty}^{\infty} F(n)\ M(x-n)\ . \qquad 2.5.18$$

Since the response function was digitized, the spectrum M(x-n) will have non-zero values only for integral values of the argument (x-n). Thus we need consider only integral values of x, and upon making the transformation m = x-n in equation (2.5.18) we get the result

$$T(x) = \sum_{m=-\infty}^{\infty} F(x-m)\ M(m)\ . \qquad 2.5.19$$

If the digitized response function (2.5.11) is substituted in the original convolution integral (2.5.1), we obtain

$$M(x) = \sum_{n=-\infty}^{\infty} R(x-n)\ T(n)\ . \qquad 2.5.20$$

Expressions (2.5.19) and (2.5.20) can be replaced by the matrix equations

$$\underline{T} = \underline{F}\ \underline{M} \qquad 2.5.21$$

and

$$\underline{M} = \underline{R}\,\underline{T} \ , \qquad 2.5.22$$

where $\underline{F}$ is an infinite square matrix having elements

$$f_{ij} = F(i-j) \qquad 2.5.23$$

and $\underline{R}$ is likewise an infinite square matrix with elements

$$r_{k\ell} = R(k-\ell) \ . \qquad 2.5.24$$

$\underline{T}$ and $\underline{M}$ are infinite one-column vectors. It is clear that

$$\underline{F} = \underline{R}^{-1} \ , \qquad 2.5.25$$

and after collecting the results of equations (2.5.23), (2.5.17) and (2.5.13) we can write down the inverse matrix elements

$$r^{-1}_{k\ell} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\omega(k-\ell)}}{\sum_{n=-\infty}^{\infty} r(\omega-2\pi n)}\, d\omega \ , \qquad 2.5.26$$

where $r(\omega)$ is the Fourier transform of the response function $R(x)$. The denominator of the integrand can be replaced by the alternative expression (2.5.12) if the latter proves to be more convenient.

The foregoing derivation used response functions which were digitized by sampling $R(x)$ at integral values of x corresponding to channel positions. However, if the data are of the histogram type, it is necessary to include the effects of integration over channel width, as is done in the model (2.1.8). Using unit channel width we can write the new digitized response function

$$R^{d}(x) = D(x)\int_{x-\frac{1}{2}}^{x+\frac{1}{2}} R(y)dy. \qquad 2.5.27$$

The channel profile is assumed to be flat-topped and is given by the

function

$$P(x) = 1 \quad \text{for} \quad |x| < \frac{1}{2} \quad ,$$
$$= \frac{1}{2} \quad \text{for} \quad |x| = \frac{1}{2} \quad , \qquad 2.5.28\ a$$
$$= 0 \quad \text{for} \quad |x| > \frac{1}{2} \quad .$$

As shown in Appendix IV, the Fourier transform of P(x) is

$$p(\omega) = \frac{2}{\omega} \sin\left(\frac{\omega}{2}\right) \qquad 2.5.28\ b$$

Expression (2.5.27) can be put in the form of a convolution integral

$$R^d(x) = D(x) \int_{-\infty}^{\infty} P(x-y)\, R(y)\, dy, \qquad 2.5.29$$

which enables us to use the convolution theorem and obtain the Fourier transform

$$r^d(\omega) = \sum_{n=-\infty}^{\infty} r(\omega-2\pi n)\, \frac{\sin \frac{1}{2}(\omega-2\pi n)}{\frac{1}{2}(\omega-2\pi n)} \quad . \qquad 2.5.30$$

Therefore the inverse matrix elements for channel-integrated response are given by

$$r^{-1}_{k\ell} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\omega(k-\ell)}}{\sum_{n=-\infty}^{\infty} \left[r(\omega-2\pi n)\, \frac{\sin \frac{1}{2}(\omega-2\pi n)}{\frac{1}{2}(\omega-2\pi n)} \right]} \, d\omega \quad . \qquad 2.5.31$$

Some examples of matrices and their inverses are provided in Table II. An integrated expression could be obtained only for the first entry (damped sine-wave response). The integrals in other entries must be evaluated by numerical methods. In some cases the inverse matrix does not exist for certain values of the response parameters. For example, in the matrix for damped sinusoid response we must satisfy $a \geq \pi$ and in the matrix corresponding to a triangular response funct-

TABLE II

Examples of Inverse Matrices

r_{ij}	r^{-1}_{ij}
Damped Sine Wave $r_{ij} = A \dfrac{\sin a(i-j)}{a(i-j)}$; $a = n\pi + k$, $n = 1, 2, 3, \cdots$ $0 \leq k < \pi$	$r^{-1}_{ij} = \dfrac{C_{ij}\, a}{A\pi^2\, n(n+1)} \dfrac{\sin[k(i-j)]}{(i-j)}$ for $i \neq j$, $= \dfrac{a[(n+1)\pi - k]}{A\pi^2\, n(n+1)}$ for $i = j$. $C_{ij} = (-1)^{i-j+1}$ for n odd, $= 1$ for n even.
Rectangle $r_{ij} = \dfrac{1}{2N+1}$ for $\lvert i-j\rvert \leq N$ $= 0$ for $\lvert i-j\rvert > N$	$r^{-1}_{ij} = \dfrac{2N+1}{\pi} \displaystyle\int_0^{\pi} \dfrac{\sin\left(\frac{\omega}{2}\right) \cos\left[w(i-j)\right]}{\sin\left(\frac{\omega}{2}\right) + 2\cos\left[\frac{\omega}{2}(N+1)\right] \sin\left(\frac{N\omega}{2}\right)} d\omega$
Triangle $R_{ij} = A(q - \lvert i-j\rvert)$ for $\lvert i-j\rvert \leq q$, $= 0$ for $\lvert i-j\rvert > q$. $q = n + k$, $n = 1, 2, 3, \cdots$ $0 < k < 1$	$r^{-1}_{ij} = \dfrac{1}{A\pi} \displaystyle\int_0^{\pi} \dfrac{\cos\left[\omega(i-j)\right]}{q + 2\sum_{\ell=1}^{n}\left[q-\ell\right]\cos(\omega\ell)} d\omega$

-continued-

TABLE II(cont'd)

Examples of Inverse Matrices

r_{ij}	r_{ij}^{-1}
Gaussian $r_{ij} = \frac{A}{\sigma\sqrt{2\pi}} \exp\left[-(i-j)^2/(2\sigma^2)\right]$	$r_{ij}^{-1} = \frac{1}{A\pi}\int_0^{\pi} \frac{\exp(\sigma^2\omega^2/2)\cos\left[\omega(i-j)\right]}{\sum_{\ell=-\infty}^{\infty}\exp\left[2\ell\pi\sigma^2(\omega-\ell\pi)\right]}\, d\omega$
Channel-Integrated Gaussian $r_{ij} = \frac{1}{\sigma\sqrt{2\pi}}\int_{i-j-\frac{1}{2}}^{i-j+\frac{1}{2}} \exp\left[-x^2/(2\sigma^2)\right] dx$	$r_{ij}^{-1} = \frac{1}{\pi}\int_0^{\pi} \frac{\exp(\sigma^2\omega^2/2)\cos\left[\omega(i-j)\right]}{\sum_{\ell=-\infty}^{\infty}\exp\left[2\ell\pi\sigma^2(\omega-\ell\pi)\right]\frac{\sin\left[\frac{1}{2}(\omega-2\ell\pi)\right]}{\frac{1}{2}(\omega-2\ell\pi)}}\, d\omega$

tion the parameter q cannot have integral values other than unity. The last two entries contain the factor exp $(\sigma^2\omega^2/2)$ in the inverse matrix integrands. This may lead to very large inverse matrix elelements for high values of σ and thus render application of the inverse matrix unpractical when the spectrum contains statistical deviations.

The matrices considered in this section have the restriction that their elements depend only on their distance $(k-\ell)$ from the principal diagonal. This restriction reflects the assumption that the shape of response functions is independent of the incident energy (e.g. constant value of σ for Gaussian response functions). In general a response matrix has elements

$$ {}^{(j)}r_{ij} = {}^{(j)}R(i-j) \, , \qquad 2.5.32 $$

where the superscript j signifies the dependence of response function on column number. Under certain conditions it is possible to obtain a good approximation to the inverse of this generalized matrix. Suppose we fix the superscript j at a certain value ℓ and then obtain the inverse matrix ${}^{(\ell)}\underline{R}^{-1}$ by the method outlined above. This inversion can be carried out in turn for various other fixed values of ℓ. Then we construct a matrix $\underline{R}_o^{-1}$ in which the ℓ^{th} column is given by the corresponding column of ${}^{(\ell)}\underline{R}^{-1}$. If the elements of ${}^{(\ell)}\underline{R}^{-1}$ are given by

$$ {}^{(\ell)}r_{ij}^{-1} = {}^{(\ell)}F(i-j) \, , \qquad 2.5.33 $$

then the condition that

$$ \underline{R}\,\underline{R}_o^{-1} \longrightarrow \underline{I} $$

can be stated in the detailed form

$$\sum_j {}^{(j)}R(i-j)\ {}^{(k)}F(j-k) \longrightarrow \delta_{ik} \,. \qquad 2.5.34$$

Equality would obtain in this last expression if we could replace ${}^{(j)}R(i-j)$ by ${}^{(k)}R(i-j)$. If the response shape varies slowly with energy, then close to the value of k there is a range of j-values for which ${}^{(j)}R(i-j) \approx {}^{(k)}R(i-j)$. Suppose that outside this range of j-values the contribution to the sum in (2.5.34) is negligible; then $\underline{R}_o^{-1}$ can be a good approximation to the true inverse matrix $\underline{R}^{-1}$. In the case of Gaussian response with the standard deviation $\sigma(E)$ these necessary conditions are satisfied when $d\sigma/dE \ll 1$.

2.6 Reduction of Statistical Deviations

As was remarked in Chapter I, the spectra of nuclear counting experiments contain statistical deviations which have very nearly Gaussian frequency functions. Each channel of the observed spectrum therefore has an associated variance $v_j^2 = m_j$, where m_j is the expectation of the contents of the j^{th} channel. Since the contents of the channels are mutually statistically independent, all covariances v_{ij}^2 are zero. When the observed spectrum is multiplied by an inverse matrix $\underline{R}^{-1}$, the variances in the unfolded spectrum will naturally have magnitudes different from the original variances. According to the error propagation formula given in section 3.2, the new variance in channel i will be

$$u_i^2 = \sum_j \left[r_{ij}^{-1}\right]^2 v_j^2 \,. \qquad 2.6.1$$

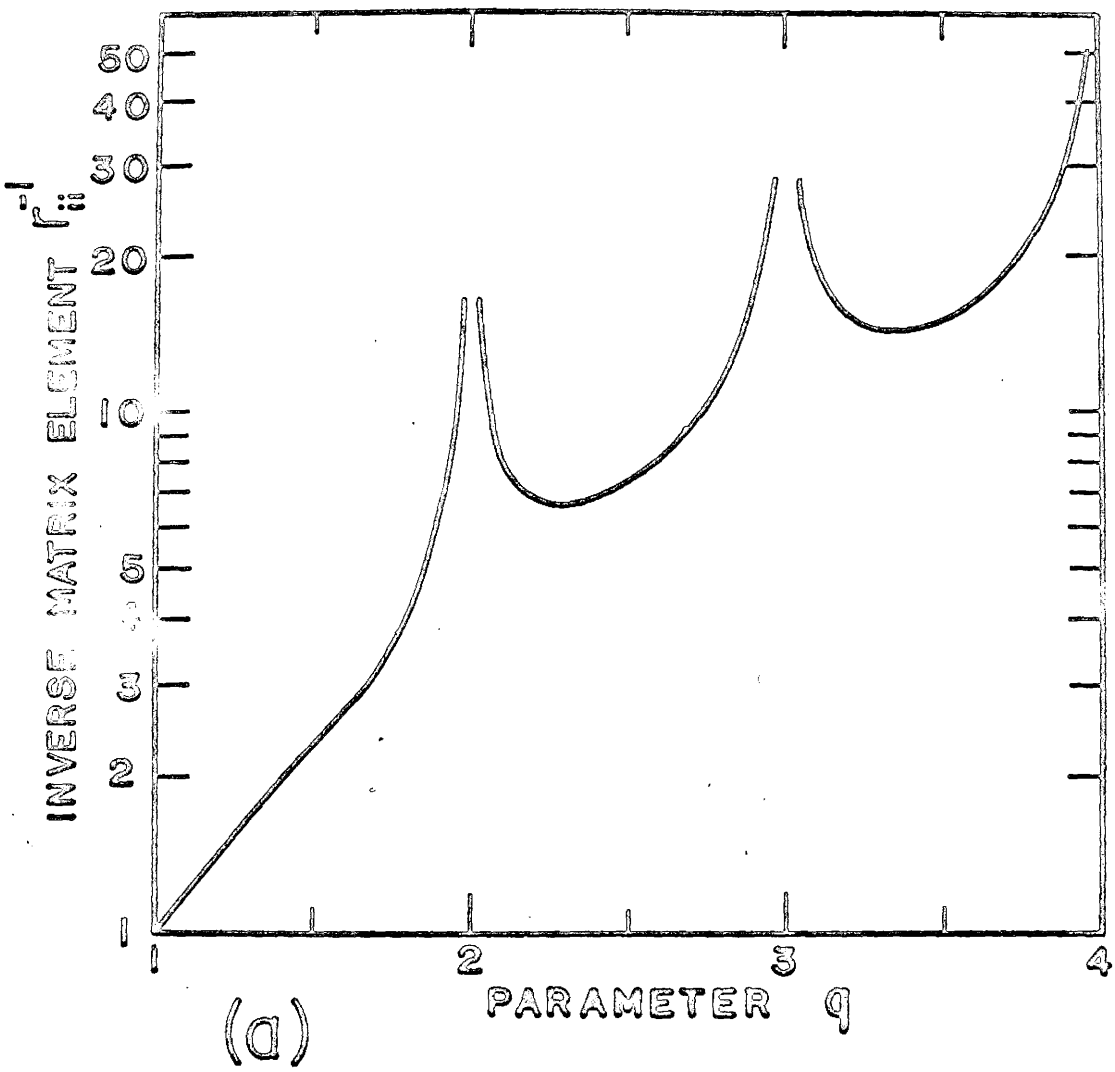
From this expression it is clear that when the inverse matrix $\underline{R}^{-1}$ contains elements of large magnitude, the unfolded spectrum will have greatly magnified statistical uncertainties. A lower limit on the

ratio of standard deviations in channel i is given by

$$\left(\frac{u_i}{v_i}\right)_{min} = \left|r_{ii}^{-1}\right| . \qquad 2.6.2$$

Expressions (2.5.26) and (2.5.31) defining the inverse matrix elements, show that the diagonal elements always have the largest numerical value. It is therefore of some interest to consider their magnitudes for various types of response functions and various values of response parameters. Figure 9a shows the diagonal inverse matrix element as a function of parameter q for a triangular response function. The response matrix and inverse matrix are defined in the third entry of Table II, where an amplitude factor A is included so as to conserve the normalization condition $\sum_j r_{ij} = 1$. It is seen that, as parameter q takes on integral values greater than unity, the element r_{ii}^{-1} becomes infinite. This has the significance that for these values of q there is no unique unfolded solution. Figure 9b exemplifies this fact by showing two initially different spectra which become identical after a triangular response with q=2 is folded in.

The inverse matrix element r_{ii}^{-1} for a channel-integrated Gaussian response (see entry 5 of Table II) is shown as a function of σ in Figure 10. The extremely rapid rise with increasing σ is dictated by the factor $\exp(\sigma^2\omega^2/2)$ in the integrand of the expression for inverse matrix element. Since, according to equation (2.6.2), the magnification in statistical uncertainties becomes rapidly untenable, the practical application of this inverse matrix is limited to values of σ not much greater than unity. However, in practice it is desirable to have a high channel density for the purpose of



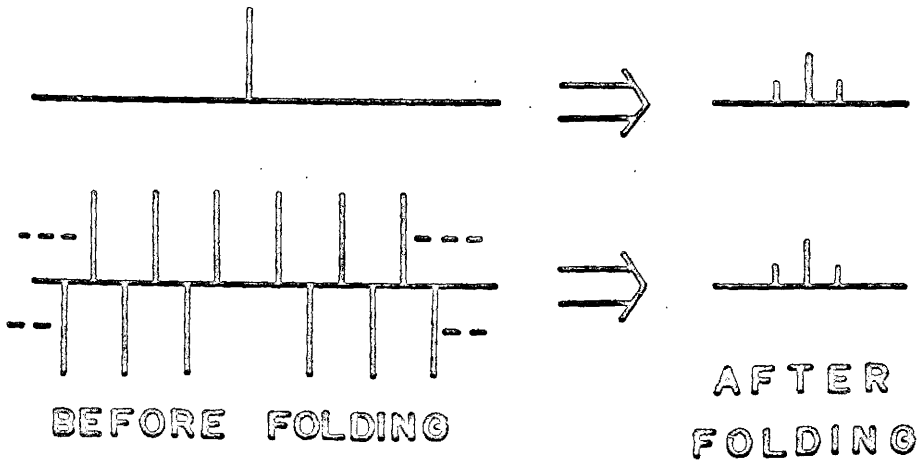


FIGURE 9

TRIANGULAR RESPONSE FUNCTION

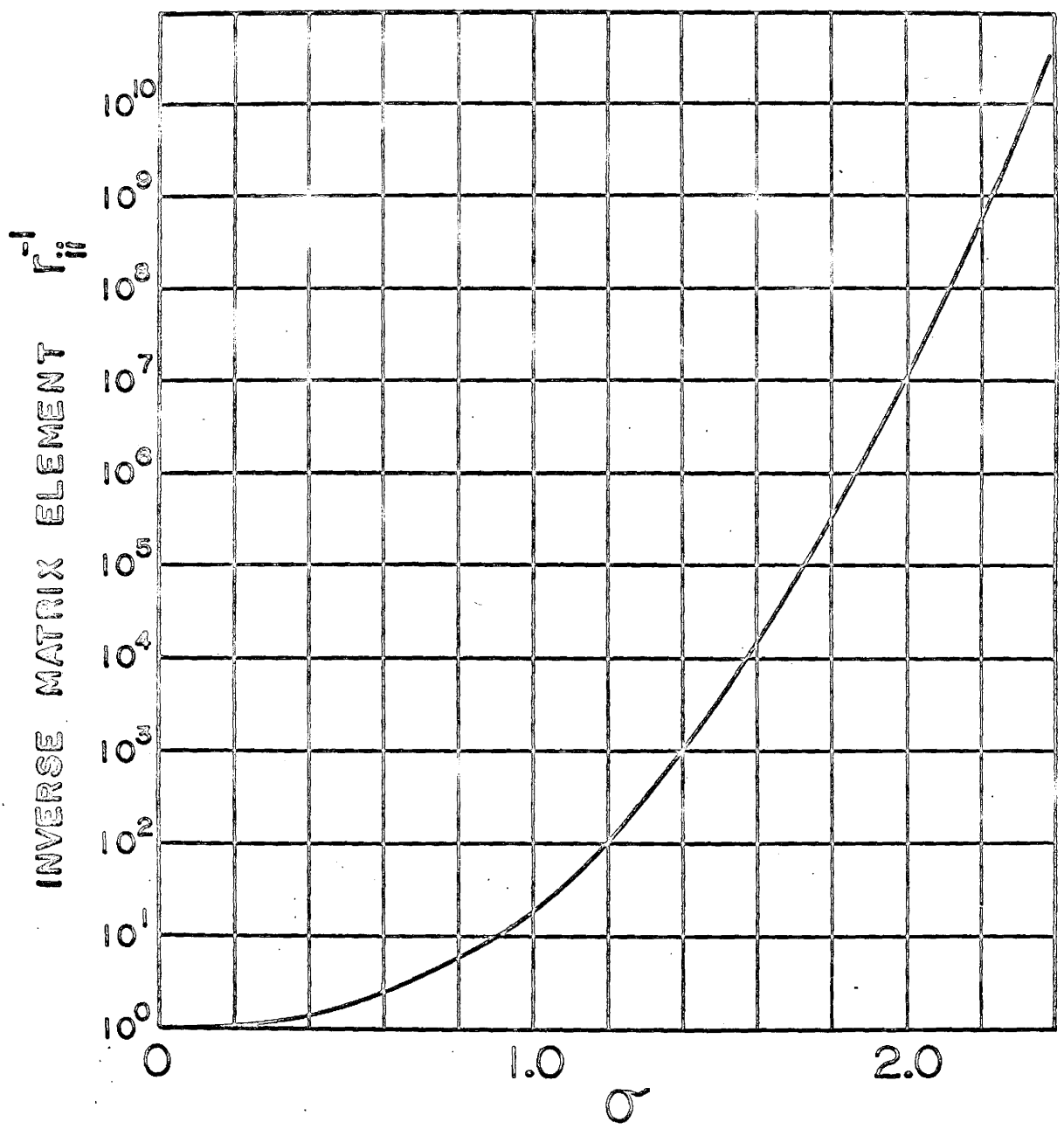


FIGURE 10

CHANNEL-INTEGRATED GAUSSIAN RESPONSE

preserving as much of the fine structure as possible. This leads to high values of σ (which is measured in channel units) and we are thus faced with conflicting requirements.

Since the merits of fine channel mesh cannot be denied, the difficulty might be resolved by modifying the inverse matrix. Instead of attempting to "squeeze" the Gaussian into one channel, one may have to settle for partial unfolding which would leave some residual but narrower response function $R_n(x)$. The elements of a "partial inversion" matrix (see Appendix IV) are given by

$$r^{-1}_{k\ell} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\omega(k-\ell)}}{r^d(\omega)} \sum_{n=-\infty}^{\infty} r_n(\omega-2\pi n)d\omega \ , \qquad 2.6.3$$

where $r^d(\omega)$ is defined by (2.5.12) and (2.5.13). For a Gaussian response the value of $1/r^d(\omega)$ is especially large near the integration limits $\omega = \pm \pi$. Therefore, if we are to keep the magnitude of $r^{-1}_{k\ell}$ down to relatively low values, we must choose the residual $R_n(x)$ in such a way that its Fourier transform $r_n(\omega)$ is very small or zero near $\omega = \pm \pi$. It may also be desirable that $R_n(x)$ have no negative values. This can be ensured by choosing a $r_n(\omega)$ such that it is the convolution of some other function with itself. According to the convolution theorem $R_n(x)$ is then some function squared and hence cannot have negative values. Within these requirements there is considerable room for variation in the particular form of $R_n(x)$ chosen. Some future effort may be warranted in search of an optimum shape such that for a given reduction in resolution width the magnification of statistical uncertainties is minimized.

A powerful method of reducing statistical deviations can be

derived by making use of the a priori knowledge that intensity spectra cannot have negative values. If the observed spectrum vector is $\underline{M}'$ (as defined in (2.1.10)), then, due to the random nature of $\underline{S}'$, the unfolded spectrum $\underline{R}^{-1}\ \underline{M}'$ will generally contain some negative as well as positive components. The condition of non-negativity can be imposed by adding a vector of corrections $\underline{C}$ to $\underline{M}'$ such that the result gives

$$\underline{R}^{-1}(\underline{M}' + \underline{C}) \geq 0. \qquad 2.6.4$$

Equation (2.6.4) defines an infinite set of allowed vectors $\underline{C}$. Out of this set we select a particular vector so as to maximize the likelihood function

$$L = \left[\prod_i \frac{1}{\sqrt{2\pi}\, v_i}\right] \exp\left[-\frac{1}{2}\sum_i \frac{c_i^2}{v_i^2}\right], \qquad 2.6.5$$

where c_i are the components of vector $\underline{C}$ and v_i^2 are the variances in spectrum $\underline{M}'$. Maximum L is obtained by minimizing the exponent, i.e. we have the condition

$$\sum_i \frac{c_i^2}{v_i^2} \longrightarrow \min. \qquad 2.6.6$$

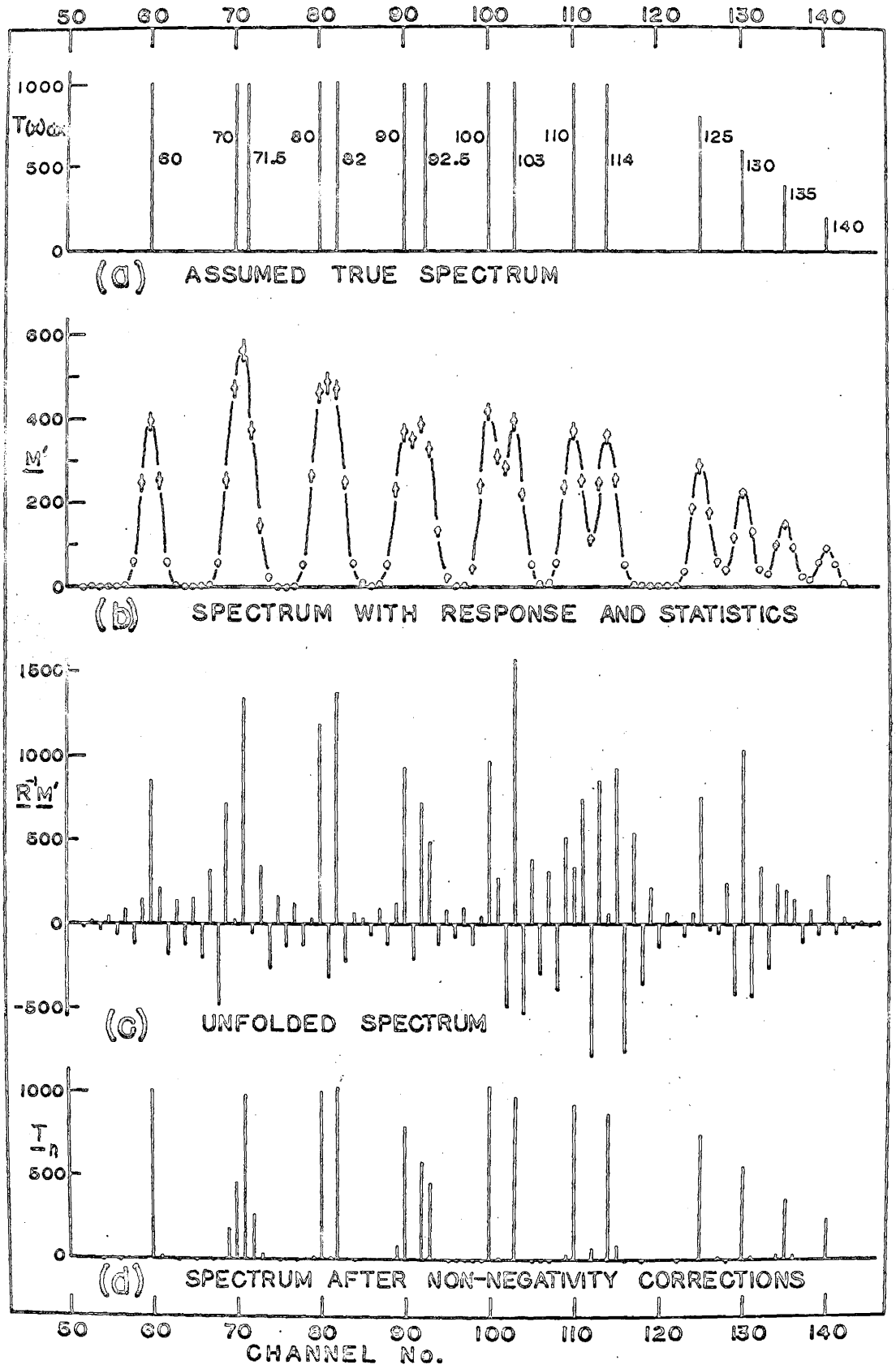
It will be shown in section 3.4 that good estimates of the weights $1/v_i^2$ are provided by $1/(m_i' + 1)$, where m_i' are the components of the observed spectrum $\underline{M}'$.

The problem as stated above can be given a geometrical interpretation. If we consider spectra limited to N channels, then expression (2.6.4) represents N linear inequalities with N independent variables c_i. Each inequality defines a region to one side of a hyper-plane in N-dimensional space. All inequalities combined define a feasible region for the vector $\underline{C}$ in this space. The solution

is given by the particular point in this feasible region which will minimize the objective function $\Sigma_i(c_i^2/v_i^2)$. This is a problem in mathematical programming for which there are several methods of solution available.[34] However, the difficulty of solution is comparable to numerical matrix inversion and in the following sample calculation we shall use another approach which is simple-minded but perhaps, for this application, more practical.

An example of unfolding by matrix inversion is given in Figure 11. Part (a) shows the assumed true spectrum T(x) in the form of a number of delta functions. Note that T(x) includes two delta functions at the non-integral positions 71.5 and 92.5 for which there is no representation in the matrix model, as explained in section 2.1. Part (b) shows the spectrum $M'(x)$ obtained by folding in a channel-integrated Gaussian response with $\sigma = 1$ and adding appropriate statistical deviations. The deviations were obtained by sampling from a set of normally distributed random numbers. After multiplication by the inverse matrix $\underline{R}^{-1}$ we obtain the spectrum shown in part (c). Large positive and negative oscillations caused by magnification of statistical uncertainties are seen to obscure the true spectrum almost completely. Part (d) shows the result after application of non-negativity calculations, which were based on a very rapid iterative procedure. In each iteration a search was made for the largest negative value in the spectrum, which was found, say, at channel y. Then the following combination of inverse functions F(x) was added to the spectrum:

$$A\,F(x-y) - \frac{A}{2}\left[F(x-y-1) + F(x-y+1)\right] . \qquad 2.6.7$$



EXAMPLE OF UNFOLDING

FIGURE 11

The amplitude A was chosen so as to make the spectrum component at channel y equal to zero. Inclusion of the last two terms avoids the introduction of a positive bias. The iterations were continued by searching again for the largest negative component and repeating the same process as above. If after n iterations the spectrum is $T_n(x)$, then the addition of expression (2.6.7) to $T_n(x)$ affects only three channels in the spectrum $\underline{R}\ \underline{T}_n$. The value A is added to channel y and $\frac{A}{2}$ is subtracted from the channels (y-1) and (y+1).

The spectrum in Figure 11(d) was obtained after 1200 such iterations performed on a base of 200 channels. The result is a considerable improvement over the spectrum in part (c). All doublets, except the one near channel 70 are now clearly resolved. It is interesting to note that the delta function originally placed at position 92.5 ended up distributed between channels 92 and 93 in about equal proportions. In principle, the results should be better if the constraint of non-negativity were applied properly by minimizing the objective function (2.6.6).

CHAPTER III

LEAST SQUARES APPROACH

3.1 Model Equation and Solution for Parameters

In Chapter II we dealt with the matrix model in which it was assumed that the true spectrum can be represented by the components of a vector $\underline{T}$. Various methods of response correction were suggested for the purpose of obtaining a spectrum which would approximate $\underline{T}$ as closely as possible. An advantage associated with these methods is that no previous knowledge of the mathematical form of $\underline{T}$ is required. If the form of response matrix $\underline{R}$ is known, then the same set of inversion calculations should be applicable to all spectra, independently of the number or intensity of components present. However, as was shown in section 2.1, the matrix model is limited in that some types of spectra cannot be exactly represented by the vector $\underline{T}$. For instance, $\underline{T}$ cannot include transitions having energies which do not correspond to channel mid-points. This limitation is particularly important when it is desired to make accurate transition energy determinations.

The least squares approach does not have this limitation, since it permits consideration of any energy positions. In fact, the energies can be used as continuously variable least squares solution parameters. However, the least squares approach requires the use of a model function which involves the response functions and a parametrized form of an assumed true spectrum. Construction

of this model function usually requires a detailed inspection of the observed spectrum followed by an educated guess as to the number and positions of the energy transitions present.

Application of least squares methods to the analysis of one-dimensional nuclear spectra has received considerable attention in previous literature [36,37, ..., 48]. In the case of NaI(Tℓ) γ-ray spectra one great practical difficulty is encountered in the generation of a library of accurate response functions. The usual procedure [47,48] is to observe the response shapes of a number of standard isotopes having the simplest possible spectra, preferably consisting of a single energy transition. These response shapes are divided into segments which can be fitted by some convenient parametrized functions. The parameters thus obtained are in turn fitted to some appropriate functions of incident γ-ray energy. By methods such as these one obtains a set of parameters which permit calculation of response functions $R(E,E_o)$ for a continuous range of incident energies E_o. Since particular details of these methods are given elsewhere [47,48], they will not be discussed in this thesis and the subsequent development will proceed with the assumption that the response surface $R(E,E_o)$ is known.

The surface $R(x,y)$ is assumed here to be a continuous function of both its arguments. In a nuclear γ-ray spectrum we need consider only a set of n discrete γ-rays having energies y_i, where $i = 1, 2, \ldots, n$. Each γ-ray will have an associated response function which is digitized by integration into channel bins. If each channel is assigned unit width, then these response functions

have the form

$$X_i(x) = \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} R(x_1, y_i)\,dx_1 \,, \qquad 3.1.1$$

where x is allowed to have only integral values. Summing over the contributions of all γ-rays we can write down the observed spectrum model function

$$m_x = \sum_{i=1}^{n} a_i\, X_i(x) \,. \qquad 3.1.2$$

The parameters a_i are indicative of the γ-ray intensities and can be expressed by

$$a_i = \varepsilon_i\, \mathcal{N}_i \,, \qquad 3.1.3$$

where $\mathcal{N}_i$ is the number of transitions γ_i occurring in observation time T, and ε_i is the detection efficiency of transition γ_i.

According to the statistical model derived in Chapter I, the observed spectrum m'_x contains statistical deviations s'_x and can be written as the sum

$$m'_x = m_x + s'_x \,. \qquad 3.1.4$$

For large numerical values of m_x the frequency function of s'_x can be assumed to be normal with zero mean and variance m_x. This frequency function is given by $f(s'_x, m_x)$ in the expression (1.1.3). Since the N channels under consideration are statistically independent, one can obtain an N-dimensional frequency function for all the deviations s'_x by simply writing the product

$$L = \prod_{x=1}^{N} f(s'_x, m_x)$$

$$= (\text{const}). \exp \left\{ -\frac{1}{2} \sum_{x=1}^{N} \frac{1}{m_x} \left[m_x^{\circ} - \sum_{i=1}^{n} a_i X_i(x) \right]^2 \right\}. \qquad 3.1.5$$

L is known as the likelihood function of parameters a_i and finds extensive application to statistical estimation problems.

We can estimate the coefficients a_i by the method of "maximum likelihood". It will be shown in the following section that this method, based on maximizing L, is particularly advantageous in this application since it provides unbiased estimates with minimum possible variances. Obviously, L is maximized by minimizing the exponent in expression (3.1.5). This requirement leads directly to the well known condition of weighted least squares, namely

$$\sum_{x=1}^{N} \frac{1}{m_x} \left[m_x^{\circ} - \sum_{i=1}^{n} a_i X_i(x) \right]^2 \longrightarrow \min. \qquad 3.1.6$$

Differentiating (3.1.6) with respect to a_k and equating the result to zero yields n normal equations

$$\sum_{x=1}^{N} \frac{m_x^{\circ} X_k(x)}{m_x} = \sum_{i=1}^{n} a_i^{\circ} \sum_{x=1}^{N} \frac{X_i(x) X_k(x)}{m_x}, \qquad 3.1.7$$

where $k = 1, 2, \ldots, n$. If the weights $1/m_x$ were known, then we would be in position to use equations (3.1.7) for solution of the n unknown quantities a_i°. In section 3.4 it will be shown that the expression $1/(m_x^{\circ} + 1)$ provides good estimates of the weighting factors; thus for all practical purposes we may assume that the weights are known and proceed with the solution for parameters a_i.

Since equations (3.1.7) are linear, they can be conveniently represented in matrix notation. We start by defining the N-row and n-column matrix

$$\underline{X} = \begin{Vmatrix} X_1(1) & X_2(1) & \cdots & X_n(1) \\ X_1(2) & X_2(2) & \cdots & X_n(2) \\ \cdots & \cdots & \cdots & \cdots \\ X_1(N) & X_2(N) & \cdots & X_n(N) \end{Vmatrix} . \qquad 3.1.8$$

After also defining the vectors

$$\underline{M} = \begin{Vmatrix} m_1 \\ m_2 \\ \cdot \\ \cdot \\ \cdot \\ m_N \end{Vmatrix} \qquad 3.1.9$$

and

$$\underline{A} = \begin{Vmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{Vmatrix} \qquad 3.1.10$$

we can replace model equations (3.1.2) by

$$\underline{M} = \underline{X}\,\underline{A} . \qquad 3.1.11$$

The statistical observed vector $\underline{M}'$ was already defined in (2.1.1). Since the least squares solution for $\underline{A}$ will have statistical properties (indicated by the use of a prime), we define the solution vector

$$\underline{A}' = \begin{Vmatrix} a_1' \\ a_2' \\ \cdot \\ \cdot \\ \cdot \\ a_n' \end{Vmatrix} \quad . \qquad 3.1.12$$

We now can write the vector of residuals

$$\underline{D}' = \underline{M}' - \underline{M}'' \quad , \qquad 3.1.13$$

where

$$M'' = X\,A' \qquad 3.1.14$$

is the least squares estimate of vector $\underline{M}$. Upon introduction of the diagonal matrix of weights

$$\underline{W} = \begin{Vmatrix} \frac{1}{m_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{m_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{m_3} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \frac{1}{m_N} \end{Vmatrix} \quad , \qquad 3.1.15$$

the condition of least squares (3.1.6) can be rewritten

$$\underline{D}'^{T}\ \underline{W}\ \underline{D}' \longrightarrow \min. \quad , \qquad 3.1.16$$

where the superscript T indicates the transposition of vector $\underline{D}'$. Finally, we are in position to rewrite the normal equations (3.1.7) in the new form

$$\underline{X}^T\ \underline{W}\ \underline{D}' = 0 \quad . \qquad 3.1.17$$

If we define the design matrix

$$\underline{B} = \underline{X}^T\ \underline{W}\ \underline{X} \quad , \qquad 3.1.18$$

then equation (3.1.17) becomes

$$\underline{B}\,\underline{A}' = \underline{X}^T\,\underline{W}\,\underline{M}' . \qquad 3.1.19$$

To obtain a solution for $\underline{A}'$ we can multiply this last equation by $\underline{B}^{-1}$, provided the inverse matrix exists. The result is

$$\underline{A}' = \underline{B}^{-1}\,\underline{X}^T\,\underline{W}\,\underline{M}' . \qquad 3.1.20$$

Matrix $\underline{B}$ has some important properties which we shall now investigate. First of all, it is easily seen that $\underline{B}$ is symmetric about the principal diagonal. In fact, from the definition (3.1.18) we have

$$\underline{B}^T = (\underline{X}^T\,\underline{W}\,\underline{X})^T = \underline{X}^T\,\underline{W}\,\underline{X} = \underline{B} .$$

It is shown in textbooks on statistics (e.g. Cramér[49]) that, provided the columns of matrix $\underline{X}$ form a set of n linearly independent vectors (i.e. $\underline{X}$ has rank n), then $\underline{B}$ is a positive definite matrix. This means that $\det(B) > 0$ and that the inverse matrix $\underline{B}^{-1}$ exists. The condition that response functions $X_i(x)$ be linearly independent should always be satisfied in a properly set up model. Linear dependence indicates that one response function can be written as a linear combination of others and is therefore redundant. We can therefore assume that the inverse $\underline{B}^{-1}$ exists and the solution (3.1.20) is obtainable. Since the inverse of a symmetric matrix is also symmetric, we have

$$(\underline{B}^{-1})^T = \underline{B}^{-1} .$$

3.2 Statistical Properties of the Least Squares Solution

The last section dealt with a method of obtaining estimates for the coefficients a_i and showed that the required solution exists.

In matrix notation this solution was expressed by equation (3.1.20). Now we turn to the investigation of statistical properties of the solution vector $\underline{A}'$.

First we shall show that the estimate $\underline{A}'$ is unbiased, i.e. the expectation in $\underline{A}'$ or, in other words, the average value that would be obtained after a large number (approaching infinity) of identical experiments and solutions, is given by $\underline{A}$. Using the symbol E for expectation we can write

$$E(\underline{A}') = E(\underline{B}^{-1}\,\underline{X}^{T}\,\underline{W}\,\underline{M}')$$

$$= \underline{B}^{-1}\,\underline{X}^{T}\,\underline{W}\,E(\underline{M}') \,. \qquad 3.2.1$$

Matrices $\underline{B}^{-1}$, $\underline{X}^{T}$ and $\underline{W}$ were moved to the left of the expectation sign, since they are definite quantities. Since $\underline{M}'$ is assumed to be normally distributed with mean value $\underline{M}$, we have

$$E(\underline{M}') = \underline{M} = \underline{X}\,\underline{A} \,. \qquad 3.2.2$$

Thus equation (3.2.1) can be rewritten to yield the result

$$E(\underline{A}') = \underline{B}^{-1}\,\underline{B}\,\underline{A} = \underline{A} \,, \qquad 3.2.3$$

which proves the assertion that estimates $\underline{A}'$ are unbiased.

The statistical properties of a normally distributed vector, such as $\underline{M}'$ can be conveniently summarized by the use of a covariance matrix $\underline{C}_{M'}$. Individual elements of this matrix are defined by the expression

$$\left\{ C_{M'} \right\}_{ij} = E\left[(m_i' - m_i)\,(m_j' - m_j) \right] \,. \qquad 3.2.4$$

According to this definition the diagonal elements represent variances of individual components m_i', whereas the off-diagonal elements are covariances between different components m_i' and m_j'. In the

particular case of observed vector $\underline{M}^{\circ}$, the components are statistically independent and hence all covariances are zero. Since the weight matrix $\underline{W}$ is diagonal and has diagonal elements equal to reciprocals of variances m_i, we immediately have the result

$$\underline{C}_{M^{\circ}} = \underline{W}^{-1} \quad . \qquad 3.2.5$$

The covariances in vector $\underline{A}^{\circ}$ are generally not zero and we shall now derive the full covariance matrix $\underline{C}_{A^{\circ}}$. If we make the substitution

$$\underline{P} = \underline{B}^{-1} \underline{X}^{T} \underline{W} \quad ,$$

then we can rewrite equation (3.1.20) in the simple form

$$\underline{A}^{\circ} = \underline{P} \underline{M}^{\circ} \quad . \qquad 3.2.6$$

There is a theorem[50] which states that, for a normally distributed vector $\underline{M}^{\circ}$, the vector $\underline{A}^{\circ}$ is also normally distributed and has the covariance matrix

$$\underline{C}_{A^{\circ}} = \underline{P} \underline{C}_{M^{\circ}} \underline{P}^{T} \quad . \qquad 3.2.7$$

Accordingly we can write

$$\underline{C}_{A^{\circ}} = \underline{B}^{-1} \underline{X}^{T} \underline{W} \underline{W}^{-1} \underline{W}^{T} \underline{X} \underline{B}^{-1}$$

$$= \underline{B}^{-1} \quad . \qquad 3.2.8$$

Equation (3.2.8) and the expectation statement (3.2.3) completely define the statistical properties of solution vector $\underline{A}^{\circ}$.

It is possible to show that the least squares solution $\underline{A}^{\circ}$ provides a best unbiased estimate of the true vector $\underline{A}$ in the sense that components a_j° have minimum possible variances. For this purpose we construct Fisher's information matrix $\underline{H}$ which is defined[51] to have

elements

$$h_{ij} = -E\left(\frac{\partial^2 \log L}{\partial a_i \, \partial a_j}\right) \quad . \qquad 3.2.9$$

Carrying out this double differentiation on the exponent of likelihood function L we obtain the elements

$$h_{ij} = \sum_{x=1}^{N} \frac{X_i(x)\, X_j(x)}{m_x} \quad ,$$

which are identical to the elements of design matrix $\underline{B}$. We now let $\underline{A}$ have some other unbiased estimator $\underline{A}''$ with a corresponding covariance matrix $\underline{C}_{A''}$. Then according to an important theorem[52] we have

$$\underline{V}^T \underline{C}_{A''} \underline{V} \geq \underline{V}^T \underline{H}^{-1} \underline{V} \quad , \qquad 3.2.10$$

where $\underline{V}$ is any n-dimensional vertical vector. Since $\underline{H} = \underline{B}$ and, consequently, $\underline{H}^{-1} = \underline{C}_{A'}$, we can write the inequality

$$\underline{V}^T \underline{C}_{A''} \underline{V} \geq \underline{V}^T \underline{C}_{A'} \underline{V} \quad . \qquad 3.2.11$$

If we choose a particular vector $\underline{V}$ in which all but the j^{th} component are zero, the above inequality reduces to

$$\left\{ C_{A''} \right\}_{jj} \geq \left\{ C_{A'} \right\}_{jj} \quad . \qquad 3.2.12$$

This last expression proves that the estimate $\underline{A}'$ has minimum possible variances.

In some applications it is necessary to deal with a nonlinear function of the components a_i'. When the variances of a_i' can be assumed to be small, it is possible to derive approximate expressions for the statistical properties of the function $f'(a_1', a_2', \ldots, a_n')$. Using a Taylor series expansion, in which quadratic and higher order

terms are neglected, we can write

$$f' - f = \sum_{i=1}^{n} \frac{\partial f}{\partial a_i} (a_i' - a_i) \, . \qquad 3.2.13$$

Squaring both sides and taking their expectation values we obtain

$$E(f'-f)^2 = \sigma_{f'}^2$$

$$= \sum_{i,j=1}^{n} \frac{\partial f}{\partial a_i} \frac{\partial f}{\partial a_j} \left[E(a_i' - a_i) (a_j' - a_j) \right] \, . \qquad 3.2.14$$

The quantity within square brackets is, by definition, the covariance between components a_i' and a_j' (or variance if i = j) and is equal to an element of the covariance matrix $\underline{C}_{A'}$. Hence we can write down the variance of function f',

$$\sigma_{f'}^2 = \sum_{i,j=1}^{n} \frac{\partial f}{\partial a_i} \frac{\partial f}{\partial a_j} \left\{ C_{A'} \right\}_{ij} \, . \qquad 3.2.15$$

The expectation in f' is given by $f = f(a_1, a_2, \ldots a_n)$ and within the range of values where the linear assumption (3.2.13) is good, the probability density function of f' is

$$P(f') = \frac{1}{\sqrt{2\pi}\ \sigma_{f'}} \exp \left[-\frac{(f' - f)^2}{2\ \sigma_{f'}^2} \right] \, . \qquad 3.2.16$$

A commonly used measure of "goodness of fit" is the function

$$\chi'^2 = \underline{D}'^T \underline{W} \underline{D}'$$

$$= \sum_{x=1}^{N} \frac{1}{m_x} \left[m_x' - \sum_{i=1}^{n} a_i' X_i(x) \right]^2 \, . \qquad 3.2.17$$

The expectation in χ'^2 is equal to the number of degrees of freedom, (N-n), and the degree of departure from this value provides a measure of confidence in the correctness of the least squares model

employed. Excessively low values would indicate that the estimates of variances in the original data were too high, whereas too large values of χ'^2 might be due to the model (3.1.2) being incorrect. The tolerable limits for values of χ'^2 can be found in tables of most textbooks on statistics.

3.3 Application to Two-Dimensional Coincidence Spectra

In this section we shall consider the least-squares analysis of two-dimensional time correlation experiments performed with detectors x and y. Two methods of solution for the coincidence coefficients will be discussed. The first method is based on a full least squares solution which in some instances may require the inversion of an impractically large design matrix. The second method avoids this difficulty by breaking up the model equation into parts and then performing a least squares calculation for each part in turn. In general, the two solutions lead to estimates having different statistical properties. Full covariance matrices will be derived for both methods of solution and comparisons of relative variances will be made.

Some practical uses of the coincidence coefficients a_{ij} will be also discussed together with some sample calculations. It will be shown how they can be used to determine γ-ray cascades, branching ratios and even absolute decay rates without requiring knowledge of detector efficiencies.

Suppose that detector x produces response functions $X_i(x)$, where $i = 1, 2, \ldots, n_x$. The subscript i indexes the n_x types of nuclear transitions to which detector x is sensitive. Similarly we

let detector y have n_y response functions $Y_j(y)$. In general the numbers n_x and n_y may be different. We could have, say, detector x sensitive to β-rays and detector y to γ-rays. The function $X_i(x)$ would in this case represent the experimental β-ray spectrum with end-point E_i, whereas $Y_j(y)$ would be the response to a γ-ray of energy E_j. In the particular case of a γ-γ coincidence spectrum we expect to have $n_x = n_y$. We assume that the normalization condition $\Sigma_x X_i(x) = \Sigma_y Y_j(y) = 1$ applies. With this stipulation the response functions can be thought of as being frequency functions over their respective channel bases.

If counters x and y operate freely without coincidence gating we have the one-dimensional model spectra

$$M_x(x) = \sum_{i=1}^{n_x} a_{xi} X_i(x) \qquad 3.3.1$$

and

$$M_y(y) = \sum_{j=1}^{n_y} a_{yj} Y_j(y) , \qquad 3.3.2$$

where the coefficients can be expressed in the form

$$a_{xi} = \varepsilon_{xi} \mathcal{N}_i , \qquad 3.3.3$$

$$a_{yj} = \varepsilon_{yj} \mathcal{N}_j . \qquad 3.3.4$$

The number of nuclear transitions γ_i in time T is given by $\mathcal{N}_i$ and the total detection efficiencies are represented by ε_{xi} and ε_{yj}.

When coincidences between detectors x and y are demanded the resultant spectrum fills a two-dimensional array of size $N_x N_y$. Since detector x is independent of detector y, the two-dimensional frequency function for a coincidence (i,j) is simply the product $X_i(x)\, Y_j(y)$.

Upon summing the coincidences between all possible pairs (i,j) we obtain the spectrum model

$$M_2(x,y) = \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} a_{ij} X_i(x) Y_j(y) \, . \qquad 3.3.5$$

The coincidence coefficients are given by

$$a_{ij} = \varepsilon_{xi} \varepsilon_{yj} \left[\bar{W}_{ij}(\Theta) \mathcal{N}_{ij} + k \mathcal{N}_i \mathcal{N}_j \right] , \qquad 3.3.6$$

where $\bar{W}_{ij}(\Theta)$ is the angular correlation function averaged over all relative angles subtended by the two detectors, $\mathcal{N}_{ij}$ is the number of coincident transitions (i,j) and k is related to the coincidence resolving time τ through the expression

$$2\tau = k T . \qquad 3.3.7$$

The second term in equation (3.3.6) represents the contribution due to chance coincidences.

The function $\bar{W}_{ij}(\Theta)$ is usually unknown during the initial stages of a decay-scheme study. However, its variations are relatively small and for most practical purposes we can assume[53]

$$\bar{W}_{ij}(\Theta) \approx 1 \, . \qquad 3.3.8$$

Solution for coefficients a_{ij} can be effected by the application of least squares techniques to model equation (3.3.5). There are $n_x n_y$ coefficients; hence the full least squares solution requires the inversion of an $n_x n_y$ by $n_x n_y$ size design matrix. Clearly, this matrix can become prohibitively large when n_x and n_y reach the neighbourhood of 10 or more.

For large scale problems we can use a solution by parts in which the dimensions of x and y are treated separately. Model equat-

ion (3.3.5) can be rewritten in the form of two coupled equations[1]

$$M_2(x,y) = \sum_{i=1}^{n_x} q_i(y)\, X_i(x) , \qquad 3.3.9$$

$$q_i(y) = \sum_{j=1}^{n_y} a_{ij}\, Y_j(y) . \qquad 3.3.10$$

If the value of y is fixed, equation (3.3.9) can be used as a model for a least squares fit in the x-dimension. This calculation, yielding n_x coefficients $q_i(y)$, can be carried out for all the N_y values of y in the data field. As a result, least squares estimates are obtained for all $n_x N_y$ intermediate coefficients $q_i(y)$. Next we can fix the value of i and, using model equation (3.3.10), perform a least squares fit in the y-dimension. After n_x such calculations one obtains estimates of all $n_x n_y$ values of a_{ij}.

The solution by parts requires a total of $N_y + n_x$ least squares calculations, each involving the inversion of a size n_x by n_x or n_y by n_y design matrix which is considerably smaller than the $n_x n_y$ by $n_x n_y$ design matrix in the full solution. This reduction in matrix size is a decided advantage in that a larger number of nuclear transitions can be considered in the analysis.

There is another practical advantage to the solution by parts. The simultaneous solution does not provide estimates of intermediate coefficients $q_i(y)$ which represent a one-dimensional spectrum in coincidence with the i^{th} nuclear transition. These spectra are important because they provide additional insight into the model. For instance, they may provide a basis for improving the accuracy of employed response functions or they may even reveal the existence of

new transitions which were not included in the original model.

When we write down definite (non-statistical) quantities, the coupled equations (3.3.9) and (3.3.10) are mathematically equivalent to the model equation (3.3.5). In the least squares calculations we are dealing with statistical intermediate coefficients $q_i^0(y)$ which in general do not satisfy an equation like (3.3.10) but are distributed around its values. Therefore the solution by parts can be expected to yield estimates of a_{ij} somewhat different from the true least squares estimates of the full solution. The character of this difference shall be investigated below. It shall be shown that both types of estimates are unbiased and that the solution by parts gives estimate variances greater than or equal to the full solution variances. This reduction in accuracy is the price exacted for the practical advantages offered in the solution by parts. Fortunately it turns out that the increase in variances is usually small - only a few percent in the examples considered below. Full covariance matrices of the coefficients a_{ij} will be derived for both methods of solution.

As was done in the foregoing sections, we assume that the observed spectrum $M_2^0(x,y)$ is normally distributed with the variance in each channel being $M_2(x,y)$. The contents of different channels are assumed to be statistically independent. We could again write down a likelihood function and maximize it with respect to the parameters. However we go directly to the ensuing least squares condition

$$\sum_{x=1}^{N_x}\sum_{y=1}^{N_y}\left[M_2'(x,y) - \sum_{i=1}^{n_x}\sum_{j=1}^{n_y} a_{ij}\, X_i(x)\, Y_j(y)\right]^2 \Big/ M_2(x,y) \longrightarrow \min. \qquad 3.3.11$$

Differentiating (3.3.11) with respect to $a_{k\ell}$ and equating the result to zero we obtain $n_x n_y$ normal equations

$$\sum_{x=1}^{N_x}\sum_{y=1}^{N_y} X_k(x)\, Y_\ell(y)\, M_2'(x,y) / M_2(x,y)$$

$$= \sum_{i=1}^{n_x}\sum_{j=1}^{n_y} a_{ij}' \sum_{x=1}^{N_x}\sum_{y=1}^{N_y} X_k(x)\, Y_\ell(y)\, X_i(x)\, Y_j(y) / M_2(x,y) \quad , \qquad 3.3.12$$

where $k = 1, 2, \ldots, n_x$ and $\ell = 1, 2, \ldots, n_y$.

Expression (3.3.12) represents the normal equations of the full least squares solution. In order to discuss the solution and its properties in a systematic manner, it is convenient to use methods of matrix algebra. Some matrices will consist of smaller submatrices and their detailed expression would normally require cumbersome spatial arrangements of elements. For this reason a special type of notation is adopted. Matrices and vectors will be represented by enclosing a general element between double vertical bars. Two superscripts outside the bars will indicate the number of rows and columns. Curly braces will be used to denote a general matrix element, whether it be just a number or a submatrix. For example, we define the matrix of line shapes

$$\underline{Y} = \left\| \begin{matrix} Y_1(1) & Y_2(1) \ldots Y_{n_y}(1) \\ Y_1(2) & Y_2(2) \ldots Y_{n_y}(2) \\ \ldots\ldots\ldots\ldots\ldots \\ Y_1(N_y) & Y_2(N_y) \ldots Y_{n_y}(N_y) \end{matrix} \right\|$$

$$= \left\| \left\{ Y_j(y) \right\}_{y,j} \right\|^{N_y, n_y} . \qquad 3.3.13$$

The subscripts y and j indicate the row and column of the general element. Carrying this notation one step further we define a matrix of submatrices

$$\underline{Z} = \left\| \left\{ \left\| \left\{ X_i(x)\, Y_j(y) \right\}_{y,j} \right\|^{N_y, n_y} \right\}_{x,i} \right\|^{N_x, n_x} . \qquad 3.3.14$$

The outer superscripts N_x and n_x refer to the number of rows and columns of submatrices. For elements the number of rows and columns is $N_x N_y$ and $n_x n_y$ respectively.

The spectrum $M_2(x,y)$ can be represented by the vertically arranged vector (one column matrix)

$$\underline{M}_2 = \left\| \left\{ \left\| \left\{ M_2(x,y) \right\}_{y,1} \right\|^{N_y, 1} \right\}_{x,1} \right\|^{N_x, 1} . \qquad 3.3.15$$

Likewise, for the coefficients a_{ij} we use the vector

$$\underline{A}_2 = \left\| \left\{ \left\| \left\{ a_{ij} \right\}_{j,1} \right\|^{n_y, 1} \right\}_{i,1} \right\|^{n_x, 1} . \qquad 3.3.16$$

With these definitions we can write model equation (3.3.5) in the simplified matrix notation

$$\underline{M}_2 = \underline{Z}\, \underline{A}_2 . \qquad 3.3.17$$

To rewrite the normal equations (3.3.12) we make use of two further definitions. Let $\underline{D}_2'$ be the vector of residuals

$$\underline{D}_2' = \underline{M}_2' - \underline{Z}\,\underline{A}_2' \,, \qquad 3.3.18$$

where the primes on $\underline{M}_2$ and $\underline{A}_2$ indicate the coresponding statistical vectors. We also introduce the diagonal matrix of weights

$$\underline{W}_2 = \left\| \left\{ \left\| \left\{ \frac{\delta_{yj}}{M_2(x,y)} \right\}_{y,j} \right\|^{N_y N_y} \delta_{xi} \right\}_{x,i} \right\|^{N_x, N_x} . \qquad 3.3.19$$

Now the normal equations (3.3.12) can be replaced by the equivalent expression

$$\underline{B}_2\,\underline{A}_2' = \underline{Z}^T\,\underline{W}_2\,\underline{M}_2' \,, \qquad 3.3.20$$

where

$$\underline{B}_2 = \underline{Z}^T\,\underline{W}_2\,\underline{Z} \qquad 3.3.21$$

is the symmetric design matrix. If we obtain the inverse of this matrix, then we can write down the full least squares solution

$$\underline{A}_2' = \underline{B}_2^{-1}\;\underline{Z}^T\,\underline{W}_2\,\underline{M}_2' \,. \qquad 3.3.22$$

A sufficient condition for the existence of inverse matrix $\underline{B}_2^{-1}$ is that the response functions $X_i(x)$ and $Y_j(y)$ form two sets of linearly independent vectors (see Appendix V).

The same arguments as were used in section 3.2 can be applied here to derive the statistical properties of solution vector $\underline{A}_2'$. We have the obvious result that the normally distributed estimate $\underline{A}_2'$ is unbiased and that its covariance matrix is given by

$$\underline{C}_{A_2'} = \underline{B}_2^{-1} \quad . \qquad 3.3.23$$

Since $\underline{A}_2'$ is the result of a full least squares solution, its individual components have minimum possible variances.

We now turn to the problem of the solution by parts and consideration of its statistical properties. As stated before, this solution is expected to yield estimates $\underline{A}_2''$ different from the full solution estimates $\underline{A}_2'$. It will be necessary to introduce the following quantities: the matrix of response functions in the x-dimension

$$\underline{X} = \left\| \left\{ X_i(x) \right\}_{x,i} \right\|^{N_x, n_x} , \qquad 3.3.24$$

the data vectors

$$\underline{M}_y' = \left\| \left\{ M_2'(x,y) \right\}_{x,1} \right\|^{N_x, 1} , \qquad 3.3.25$$

vectors of intermediate coefficients

$$\underline{Q}_y' = \left\| \left\{ q_i'(y) \right\}_{i,1} \right\|^{n_x, 1} , \qquad 3.3.26$$

the diagonal matrix of weights

$$\underline{W}_y = \left\| \left\{ \frac{\delta_{xj}}{M_2(x,y)} \right\}_{x,j} \right\|^{N_x, N_x} , \qquad 3.3.27$$

and the positive definite design matrix

$$\underline{F}_y = \underline{X}^T \underline{W}_y \underline{X} \, . \qquad 3.3.28$$

Using model equation (3.3.9) we can obtain least squares estimates of the intermediate coefficients

$$\underline{Q}_y' = \underline{F}_y^{-1} \underline{X}^T \underline{W}_y \underline{M}_y' \, . \qquad 3.3.29$$

It can be easily shown that the estimates $\underline{Q}_y^{\prime}$ are unbiased and that they have the covariance matrices

$$\underline{C}_{Q_y^{\prime}} = \underline{F}_y^{-1} \quad . \qquad 3.3.30$$

Next we turn to model equation (3.3.10) which we use for fitting the data in the y-dimension. In the set of $n_x N_y$ coefficients $q_i^{\prime}(y)$ we will now have to fix the value of i and let y vary. Therefore the coefficients must be rearranged into a new set of vectors

$$\underline{Q}_i^{\prime} = \left\| \left\{ q_i^{\prime}(y) \right\}_{y,1} \right\|^{N_y,1} \quad . \qquad 3.3.31$$

Since all components of $\underline{Q}_i^{\prime}$ are calculated from statistically independent sets of data, the vector has a diagonal covariance matrix. In fact we have

$$\underline{C}_{Q_i^{\prime}} = \left\| \left\{ \left\{ F_y^{-1} \right\}_{i,i} \delta_{yj} \right\}_{y,j} \right\|^{N_y,N_y} , \qquad 3.3.32$$

where $\left\{ F_y^{-1} \right\}_{i,i}$ is the i^{th} diagonal element of the inverse matrix $\underline{F}_y^{-1}$.

We define a matrix of weights

$$\underline{W}_i = \underline{C}_{Q_i^{\prime}}^{-1} \qquad 3.3.33$$

and a vector of estimates for coincidence coefficients,

$$\underline{A}_i^{\prime\prime} = \left\| \left\{ a_{ij}^{\prime\prime} \right\}_{j,1} \right\|^{n_y,1} \quad . \qquad 3.3.34$$

Then by the use of model equation (3.3.10) we can obtain the solutions

$$\underline{A}_i^{\prime\prime} = \underline{K}_i^{-1} \underline{Y}^T \underline{W}_i \underline{Q}_i^{\prime} \quad , \qquad 3.3.35$$

where the positive definite matrix $\underline{K}_i$ is defined by

$$\underline{K}_i = \underline{Y}^T \underline{W}_i \underline{Y} \, . \qquad 3.3.36$$

As before, we have the result that the solutions are unbiased estimates of the true coefficients a_{ij} and that the covariance matrices are given by

$$\underline{C}_{A''_i} = \underline{K}_i^{-1} \, . \qquad 3.3.37$$

From this last expression one can obtain all variances of the estimates a''_{ij} and also the covariances between terms with equal i. The other covariances are generally not zero and are as yet undetermined.

In Appendix V it is shown that the full covariance matrix of the vector

$$\underline{A}''_2 = \left\| \left\{ \left\| \left\{ a''_{ij} \right\}_{j,1} \right\|^{n_y,1} \right\}_{i,1} \right\|^{n_x,1} \qquad 3.3.38$$

is given by

$$\underline{C}_{A''_2} = \left\| \left\{ \underline{K}_i^{-1} \underline{Y}^T \underline{U}_{ij} \underline{Y} \underline{K}_j^{-1} \right\}_{i,j} \right\| \, . \qquad 3.3.39$$

The diagonal matrix $\underline{U}_{ij}$ is defined by

$$\underline{U}_{ij} = \left\| \left\{ \rho_{ij}(y) \, \delta_{ky} \right\}_{k,y} \right\|^{N_y,N_y} \, , \qquad 3.3.40$$

where

$$\rho_{ij}(y) = \frac{\left\{ F_y^{-1} \right\}_{i,j}}{\left\{ F_y^{-1} \right\}_{i,i} \left\{ F_y^{-1} \right\}_{j,j}} \, . \qquad 3.3.41$$

Consider now the diagonal submatrices of $\underline{C}_{A''_2}$. Putting j=i we obtain

$$\left\{ \underline{K}_i^{-1} \underline{Y}^T \underline{U}_{ii} \underline{Y} \underline{K}_i^{-1} \right\}_{i,i} = \underline{K}_i^{-1} \underline{Y}^T \underline{W}_i \underline{Y} \underline{K}_i^{-1}$$

$$= \underline{K}_i^{-1} \, . \qquad 3.3.42$$

This result agrees with the covariance matrices given by equation (3.3.37). The off-diagonal submatrices of $\underline{C}_{A''_2}$ give covariances which were not obtainable before, namely covariances between coefficients a''_{ij} having different values of i.

According to the inequality (3.2.12), the variances in $\underline{A}''_2$ are greater than or equal to the corresponding variances in the full solution vector $\underline{A}'_2$. An exact comparison of variances can be made only after inverting the full solution design matrix $\underline{B}_2$. This, however, may be impractical in a routine calculation since one of the reasons for going to the solution by parts is precisely to avoid such a matrix inversion. It is therefore desirable to have some simple method of determining lower limits to the variances in $\underline{A}'_2$. The design matrix $\underline{B}_2$ can be written in terms of submatrices $\underline{B}_{ij}$,

$$\underline{B}_2 = \left\| \left\{ \underline{B}_{ij} \right\}_{ij} \right\|^{n_x, n_x}, \qquad 3.3.43$$

where

$$\underline{B}_{ij} = \left\| \left\{ \sum_{x,y} X_i(x)\, X_j(x)\, Y_p(y)\, Y_q(y) \,/\, M_2(x,y) \right\}_{p,q} \right\|^{n_y, n_y}. \qquad 3.3.44$$

It can be shown that for variances $D(a'_{k\ell})$ in the vector $\underline{A}'_2$ the following inequalities hold (see Appendix V)

$$D(a'_{k\ell}) \geq \left\{ \underline{B}^{-1}_{kk} \right\}_{\ell,\ell} \geq \frac{1}{\left\{ \underline{B}_2 \right\}_{i,i}}, \qquad 3.3.45$$

where the index i is given by

$$i = (k-1)\, n_y + \ell .$$

In expression (3.3.45) the second term represents the ℓ^{th} diagonal element in the inverse of k^{th} diagonal submatrix $\underline{B}_{kk}$. It is the

stronger of the two given lower limits. The weaker limit (last term in (3.3.45))is just the reciprocal of i^{th} diagonal element in matrix $\underline{B}_2$.

These limits can be used to establish upper bounds on the magnification of variances introduced through the solution by parts. Some examples of actual variance magnifications and calculated upper limits are given below.

The models chosen for the purpose of comparing variances are based on the simple scheme of two γ-rays in cascade. A computer program for generating response functions corresponding to a 3" x 3" NaI(Tℓ) detector at 10 cm. from a γ-ray source was used (Archer[47]). Response functions were assumed to be identical in both dimensions and were generated for an assumed analyzer gain of 40 keV/channel. Assuming that $a_{11} = a_{22} = 0$ and that $a_{12} = a_{21} = 100$, the model coincidence spectrum is

$$M_2(x,y) = 100 \left[X_1(x)\, X_2(y) + X_2(x)\, X_1(y) \right] . \qquad 3.3.46$$

A set of 13 models characterized by different response functions $X_2(x)$ was used. Function $X_1(x)$ was kept fixed with its photo-peak centeredin channel 20. The photo-peak position of $X_2(x)$ was varied from channel 30 to 20.1 in steps as shown in column 1 of Table III.

The spectrum field size used was 33 x 33 channels. With the given response functions (having shapes similar to the ones in Figure 13) the spectrum has some regions of zero content. Consequently the weighting function $1/M_2(x,y)$ contains singularities. This difficulty was eliminated by slightly modifying the response functions with the addition of a constant equal to 10^{-3}.

TABLE III

Comparison of Variances

Peak Positions, Channel Nos.	Singularity Parameter s	Average Variance $D(a'_{ij})$	Average Variance Ratio r	Stronger Upper Limit ℓ_s	Weaker Upper Limit ℓ_w
20,30	.465	130	1.030	1.227	1.502
20,29	.445	136	1.033	1.240	1.538
20,28	.424	143	1.036	1.255	1.582
20,27	.407	148	1.040	1.271	1.620
20,26	.421	144	1.041	1.262	1.601
20,25	.495	123	1.042	1.218	1.466
20,24	.590	103	1.040	1.167	1.339
20,23	.592	103	1.042	1.163	1.336
20,22	.379	154	1.054	1.295	1.661
20,21	3.71×10^{-2}	702	1.036	2.19	5.13
20, 20.5	4.41×10^{-4}	8.05×10^{3}	1.004	6.68	47.2
20,20.2	2.25×10^{-6}	1.23×10^{5}	1.0004	26.2	667
20,20.1	2.2×10^{-7}	4.16×10^{5}	1.0001	46.9	2240

Variances of the full solution estimates a'_{ij} and the solution by parts estimates a''_{ij} were calculated for each model. Column 3 of Table III gives the average variance

$$\overline{D(a'_{ij})} = \frac{1}{4} \sum_{i,j=1}^{2} D(a'_{ij}) \qquad 3.3.47$$

which is seen to increase quite rapidly as the two photo-peaks reach close proximity. The average ratio of variances

$$r = \frac{1}{4} \sum_{i,j=1}^{2} \left[D(a''_{ij}) \,/\, D(a'_{ij}) \right] \qquad 3.3.48$$

is given in column 4. It is seen that this ratio is very close to unity, the greatest departure from that value being only 5.4%.

The second column of Table III provides a measure of the approach to singularity of design matrix $\underline{B}_2$. Since $\underline{B}_2$ is positive definite and symmetric, the ratio of its determinant to the product of diagonal elements cannot be greater than unity. In equation form we have the singularity parameter

$$s = \frac{\det (\underline{B}_2)}{\prod_i \left\{ \underline{B}_2 \right\}_{i,i}} \leq 1 . \qquad 3.3.49$$

If $\underline{B}_2$ were diagonal this ratio would equal unity. However, diagonality cannot be had in $\underline{B}_2$ since the response functions used do not possess any orthogonality properties. For photo-peak separations greater than one channel the value of s is near 0.5. As the separation is reduced, matrix $\underline{B}_2$ approaches singularity and the ratio s tends to zero. Uncertainties in least squares solution become very large but the ratio of variances, r appears to approach unity.

Upper limits for the ratio r were calculated by the use of inequalities (3.3.45). Column 5 of Table III gives values of the stronger upper limit

$$\ell_s = \frac{1}{4} \sum_{i,j=1}^{2} \left[D(a''_{ij}) \Big/ \left\{ B_{ii}^{-1} \right\}_{j,j} \right] . \qquad 3.3.50$$

The weaker limit, given in column 6, is

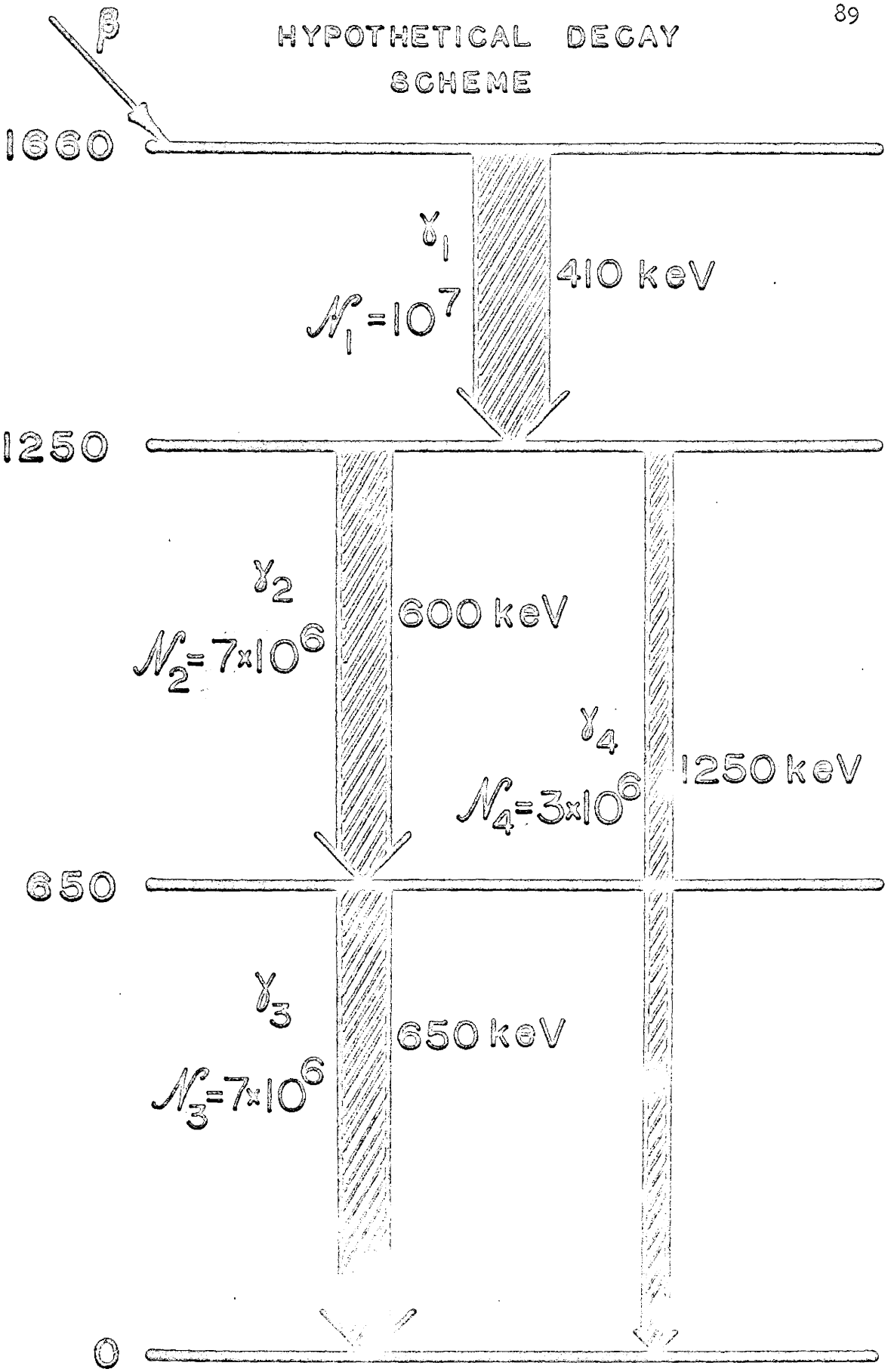
$$\ell_w = \frac{1}{4} \sum_{i,j=1}^{2} D(a''_{ij}) \sum_{x,y=1}^{33} \left[X_i^2(x)\, X_j^2(y) \Big/ M_2(x,y) \right] . \qquad 3.3.51$$

For most of the models tested both limits remain close to unity and therefore are useful. However, these limits provide little information when s approaches zero. This limitation is not serious, since, in the cases where s is very small, the uncertainties are so great that the practical application of a least squares solution may become unfeasible.

Variances resulting from the two methods of solution were calculated also for the example given below, which has four γ-ray transitions and consequently 16 coincidence coefficients a_{ij}. The average variance ratio r was 1.01. It would thus appear that in most practical cases the magnification of uncertainties introduced by the solution by parts is insignificant.

In order to illustrate the potential power of coincidence analysis an example of calculations was carried to conclusion. The hypothetical decay scheme used is shown in Figure 12. There are four γ-ray transitions indexed from 1 to 4 in order of their energies 410, 600, 650 and 1250 keV. Intensities are indicated by the total numbers of transitions N_i which occurred during the time of experiment. The first transition γ_1 is shown to have occurr-

FIGURE 12


HYPOTHETICAL DECAY SCHEME
β
1660
γ1
N1=10^7
410 keV
1250
γ2
N2=7x10^6
600 keV
γ4
N4=3x10^6
1250 keV
650
γ3
N3=7x10^6
650 keV
0

ence of 10^7 counts. Following this there are 2 branches. Transitions γ_2 and γ_3 in cascade have an occurrence of 70%, whereas the single γ_4 takes up the remaining 30%.

In order of increasing energy the total efficiencies used were 0.294, 0.293, 0.293, 0.283 for detector x and 0.265, 0.265, 0.265, 0.254 for detector y. The parameter k in equation (3.3.6) was fixed for 10% chance coincidences by setting $k\mathcal{N}_1 = 0.1$. Response functions were generated using the different gains of 20 keV/ch. in the x-dimension and 25 keV/ch. in the y-dimension. The set of functions for the y-dimension, $Y_j(y)$ is shown in Figure 13. A data field of 3100 channels was used with $N_x = 62$ and $N_y = 50$.

Model one-dimensional and two-dimensional (coincidence) spectra were calculated by the use of equations (3.3.1) through (3.3.6). Angular correlation effects were neglected by setting $\bar{W}_{ij}(\theta) = 1$. To simulate experimental spectra statistical deviations sampled from appropriate normal distributions were added to the models. Resultant values were rounded off to the nearest integer to give the final channel content M'. If M' came out negative, it was arbitrarily set to zero.

The one-dimensional spectrum in the y-dimension, $M'_y(y)$, is shown in Figure 14. All four response functions of Figure 13 show their presence in various intensities. A least squares calculation produced estimates a'_{yj} of 2.6528×10^6, 1.8537×10^6, 1.8556×10^6 and 7.6189×10^5 in order of increasing j. The largest fractional deviation from model value $\varepsilon_{yj}\mathcal{N}_j$ occurred in the first coefficient a'_{y1} and was 0.11%. The content of each channel in spectrum $M'_y(y)$ is quite large ($M' > 4000$). Consequently the statistical

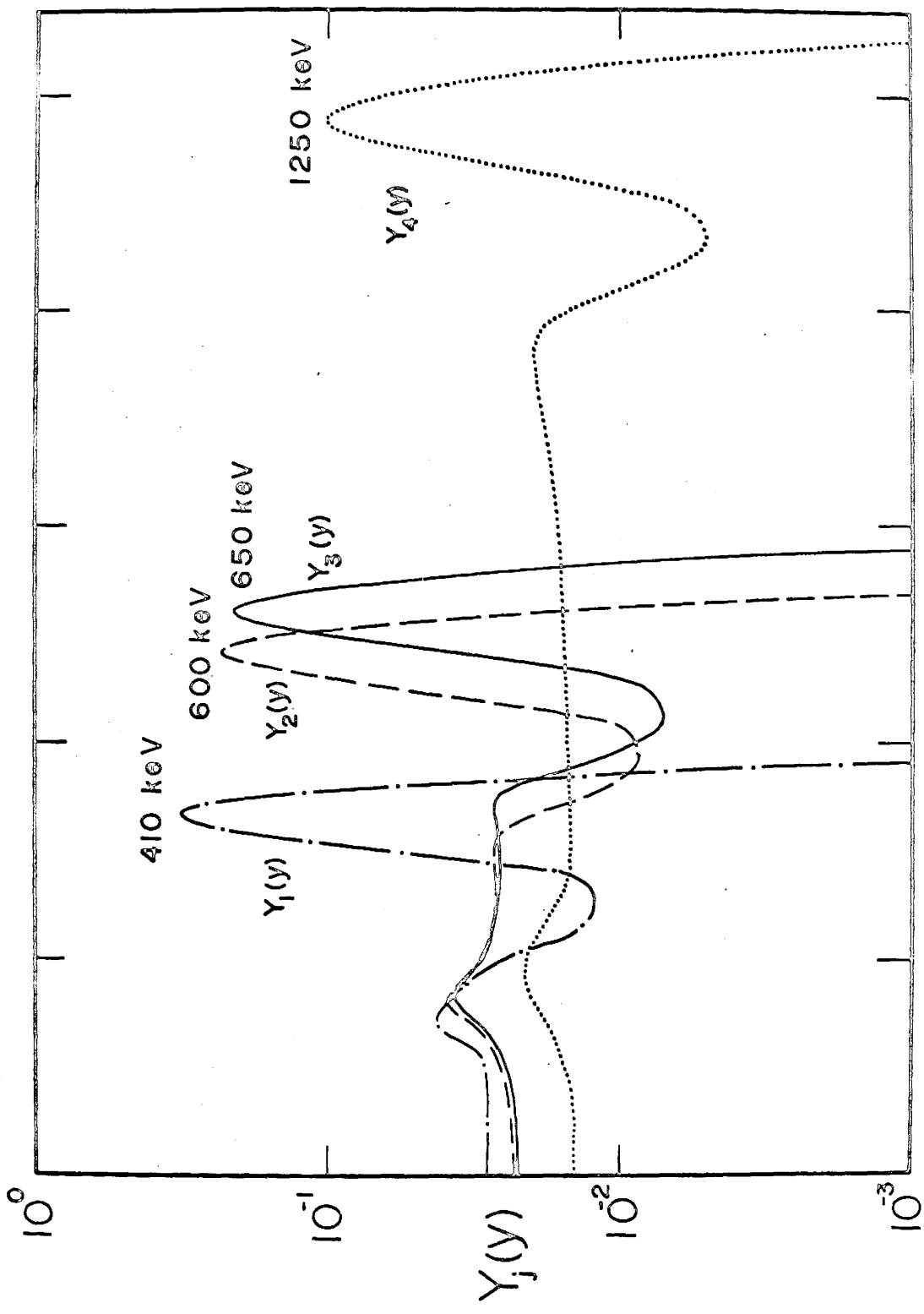


FIGURE 13

RESPONSE FUNCTIONS

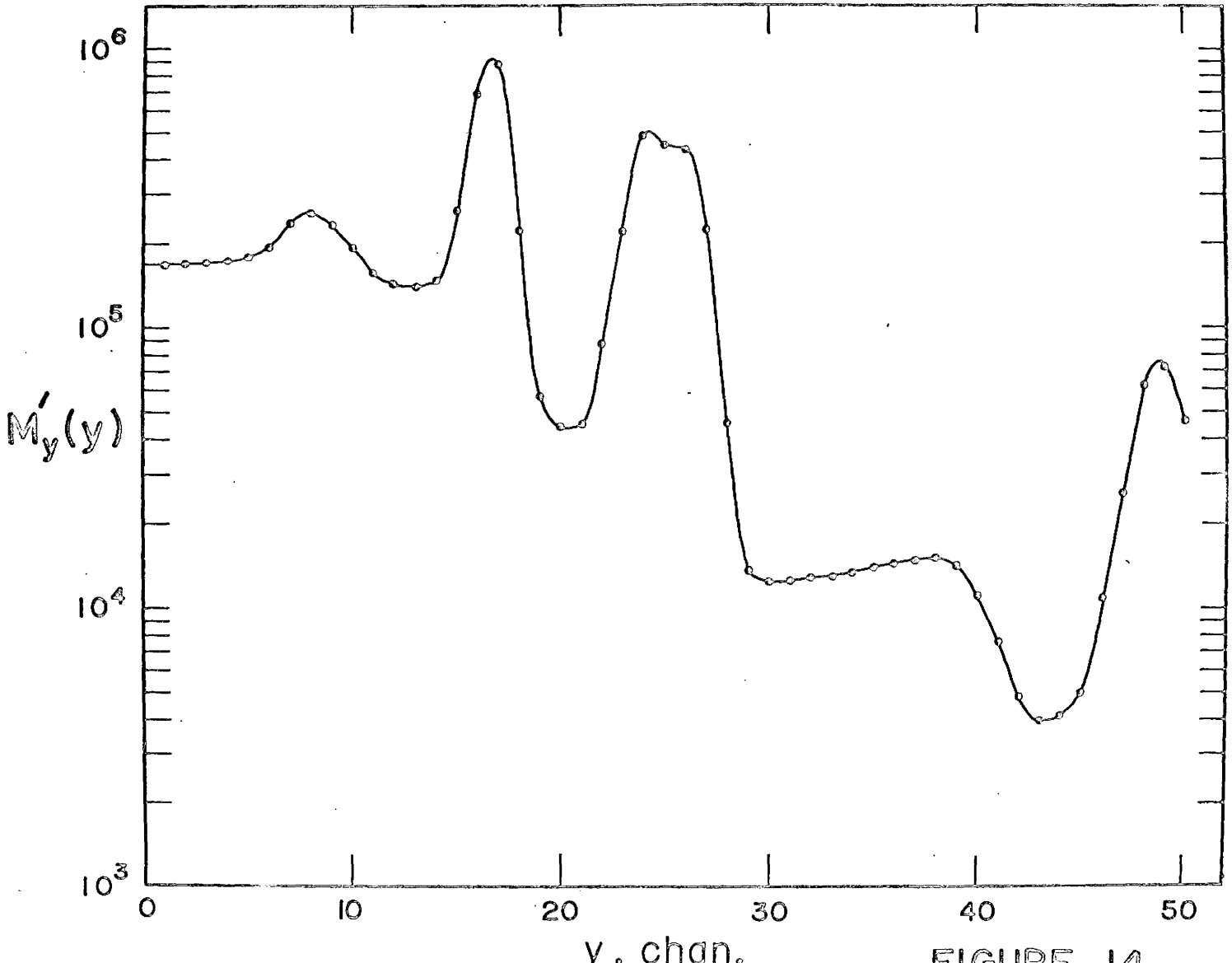


FIGURE 14

ONE-DIMENSIONAL SPECTRUM
IN DETECTOR y

deviations which are of the order $\sqrt{M'}$ represent a small percentage - too small to be visible in Figure 14. With these "good statistics" the use of weighting factor $1/M'$ in the least squares calculation provided an adequate approximation to the true weights $1/M$. A similar analysis of the spectrum $M_x'(x)$ in the x-dimension yielded coefficients a'_{xi} of 2.9417×10^6, 2.0473×10^6, 2.0564×10^6 and 8.4843×10^5. The largest fractional deviation from the model value was 0.26% in the third coefficient. These coefficients will be of use in determining the absolute intensities $\mathcal{N}_i$ later on.

The generated coincidence spectrum $M_2'(x,y)$ contains areas of very low channel content. For these channels $1/M'$ can be a very poor approximation of the true weight. An iterative process was used to obtain better weight estimates. Weights $1/M'$ were used in a preliminary least squares solution by parts. Channels with $M' < 5$ were ignored in the calculation. The ensuing coincidence coefficients were then substituted in model equation (3.3.5) to obtain the improved weight estimates $1/M''$. A second least squares solution by parts (rejecting channels with $M'' < 5$) yielded the final estimates a''_{ij}. Comparison of results obtained with various weight estimates is made in the following section.

Intermediate coefficients $q_i'(y)$ are plotted in Figure 15. Each shown function corresponds to a spectrum in the y-dimension which is in coincidence with the i^{th} transition recorded in detector x. Except for small contribution due to chance coincidences, we see only

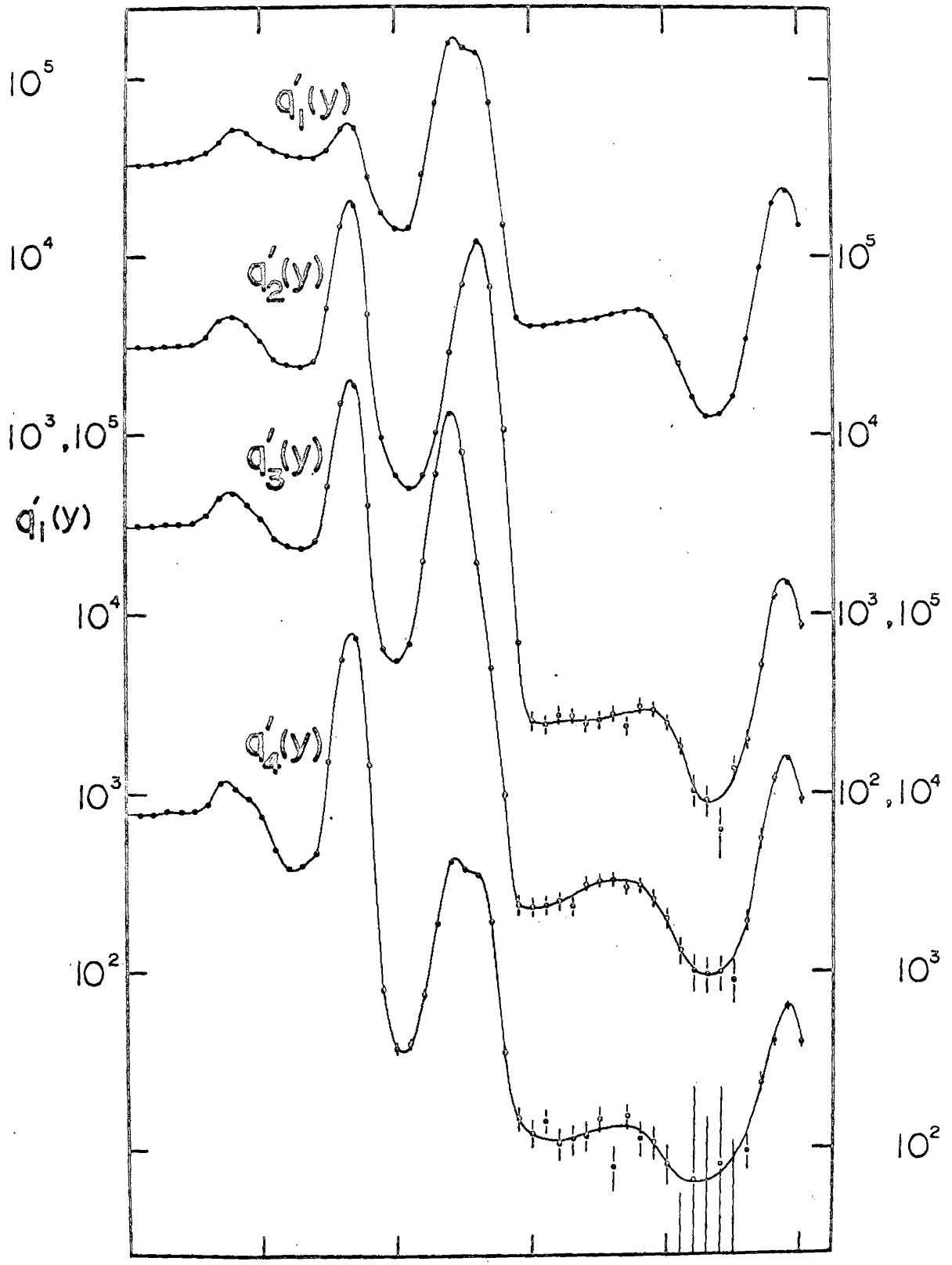


FIGURE 15

INTERMEDIATE COEFFICIENTS $q'_i(y)$

transitions γ_2, γ_3 and γ_4 in $q_1'(y)$; γ_1 and γ_3 in $q_2'(y)$; γ_1 and γ_2 in $q_3'(y)$ and finally - only γ_1 in $q_4'(y)$. These spectra are simpler than the one-dimensional spectrum of Figure 14 and therefore can be advantageous in studying response function shapes or in searching for new transitions not included in the old model.

Using equations (3.3.3), (3.3.4) and (3.3.6) we can obtain

$$\frac{a_{ij}}{a_{xi}\, a_{yj}} = \frac{\mathcal{N}_{ij}}{\mathcal{N}_i\, \mathcal{N}_j} + k . \qquad 3.3.52$$

Since $\mathcal{N}_{ii} = 0$, the diagonal elements in the above array determine the parameter k. This can be subtracted from the whole array leaving the simplified elements $\mathcal{N}_{ij} / (\mathcal{N}_i\ \mathcal{N}_j)$. The 4 by 4 matrix of these simplified elements (calculated from the results of least squares solutions) is given in Table IV. The value of the chance parameter k was calculated to be $k' = 1.017 \times 10^{-8}$, which is 1.7% above the model value.

Diagonal elements of Table IV are expected to be zero. The degree of their departure from this value is a rough indication of the overall accuracy of the entries in the table. The large values of second, third and fourth elements in row 1 and column 1 indicate that transition γ_1 has cascade coincidences with all other transitions. Thus γ_1 cannot have any competing branches. Further examination of Table IV reveals that transitions γ_2 and γ_3 are in cascade, whereas γ_4 forms a separate branch . These conclusions are in agreement with the originally assumed decay scheme of Figure 12.

For $\mathcal{N}_i \geq \mathcal{N}_j$ it is possible to write

TABLE IV

Matrix of Values $\left[\mathcal{N}'_{ij} / (\mathcal{N}'_i \ \mathcal{N}'_j)\right] \times 10^8$

i \ j	1	2	3	4
1	-0.018	10.023	9.932	9.975
2	9.982	0.018	14.276	-0.039
3	9.923	14.223	-0.011	-0.033
4	10.032	-0.003	-0.041	0.010

$$\mathcal{N}_{ij} / (\mathcal{N}_i \mathcal{N}_j) = \mathcal{N}_i \alpha_{ij} / (\mathcal{N}_i \alpha_{ij} \mathcal{N}_i)$$

$$= 1/\mathcal{N}_i , \qquad 3.3.53$$

where α_{ij} is the fraction of transitions γ_i leading to transitions γ_j. Averaging elements with (i,j) equal to (1,2), (1,3), (1,4), (2,1), (3,1) and (4,1) we thus obtain the value 9.978×10^{-8}, the reciprocal of which yields

$$\mathcal{N}_1^{\circ} = 1.0022 \times 10^7 .$$

This result is only 0.22% above the model value $\mathcal{N}_1$. Similarly, from elements (2,3) and (3,2) we obtain the estimate

$$\mathcal{N}_2^{\circ} = \mathcal{N}_3^{\circ} = 7.018 \times 10^6 ,$$

which exceeds the model value 7×10^6 by only 0.26%. An estimate of the intensity in the competing branch γ_4 is obtained by simply taking the difference $\mathcal{N}_1^{\circ} - \mathcal{N}_2^{\circ}$.

These conclusions, derived from Table IV, required no knowledge of the detector efficiencies. Since the latter quantities are either unknown or difficult to obtain, their elimination can be a great advantage in studies of nuclear decay schemes.

3.4 Choice of Weighting Function

The general least squares solution discussed in the first two sections of the present chapter requires advance knowledge of the weighting function. This information is often not available and one is forced to resort to the use of some estimate of the proper weights. When the data vector is normally distributed, the required weights are equal to the reciprocals of variances. Use of these weights accounts for the fact that various data points have un-

equal uncertainties and hence must be assigned different degrees of importance in the least squares analysis.

In nuclear counting experiments the content M' in any given channel has a Poisson frequency function (see Chapter I). If the mean, or expectation, value is denoted by M, then the variance is given by $E(M' - M)^2 = M$. For large values of M (say $M > 5$) the Poisson frequency function assumes a nearly Gaussian shape; consequently the maximum likelihood criterion leads to the method of least squares.

Since the true weights $1/M$ are unknown, a common practice is to use the estimate $1/M'$. This can lead to some difficulties, especially when the spectrum contains channels with low counts. The value of M' may be zero, leading to a meaningless infinite weight even when the true weight $1/M$ is quite small. This necessitates special treatment (or complete exclusion) of channels with zero content. Moreover, the weight expectation $E(1/M')$ is generally different from the more realistic weight $1/E(M')$. Using the Poisson frequency function

$$p(M',M) = (M^{M'} / M'!\ )\, e^{-M}, \qquad 3.4.1$$

we have

$$E(1/M') = \sum_{M'=M_\ell}^{\infty} p(M',M)\ (1/M')\ , \qquad 3.4.2$$

where M_ℓ is the lowest channel content not rejected from the least squares fit. Ratios of $E(1/M')$ to the true weight $1/M$ are shown by the solid curves in Figure 16. The six curves were calculated for $M_\ell = 1,2,3,4,5$ and 10. It is seen that for most values of M the estimates $1/M'$ tend to be too high. For $M_\ell = 1$ the excess is almost 30% at $M = 4$ and gradually diminishes with increasing M. At the high value $M = 100$ the excess is about 1%.

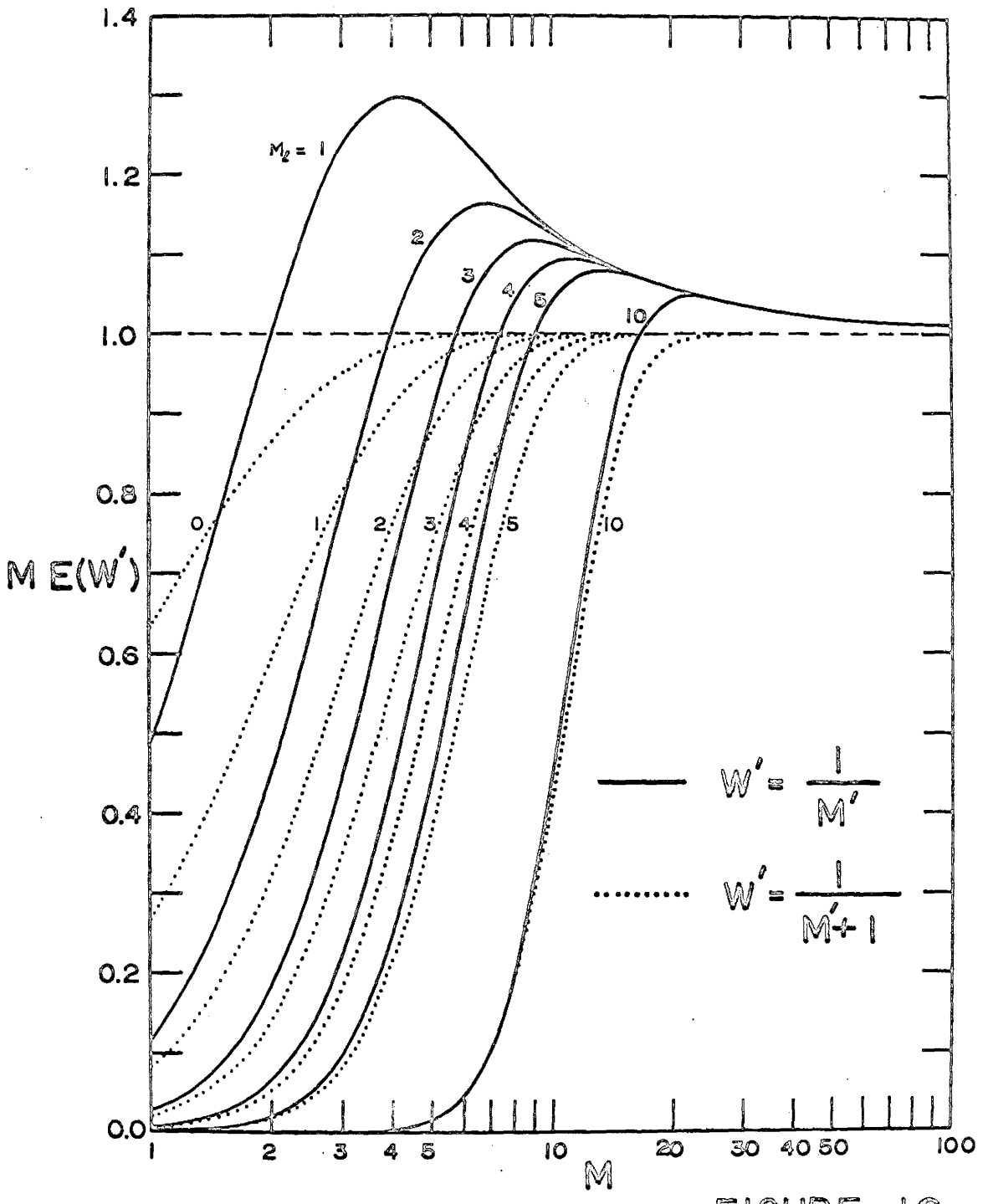


FIGURE 16

EXPECTATIONS IN WEIGHT ESTIMATES

It appears that a better estimate of weights can be obtained by using $1/(M'+1)$. With this estimate channels with zero content require no special consideration. For a given value of M we can express the weight expectation

$$E\left[1/(M'+1)\right] = \sum_{M'=M_\ell}^{\infty} p(M',M)\left[1/(M'+1)\right]$$

$$= \frac{1}{M}\left[1 - e^{-M}\sum_{n=0}^{M_\ell}\frac{M^n}{n!}\right]. \qquad 3.4.3$$

When no channels are rejected (i.e. $M_\ell = 0$) the ratio of $E\left[1/(M'+1)\right]$ to $1/M$ is simply $1 - e^{-M}$, which rapidly approaches unity with increasing M. These ratios are shown in Figure 16 by the dotted curves calculated for $M_\ell = 0, 1, 2, 3, 4, 5$ and 10. It is apparent that the averages $E\left[1/(M'+1)\right]$ approach true weights $1/M$ much faster than is done by values $E(1/M')$ of the solid curves.

Better weights than either of the two above estimates can be obtained by a process of iteration. The above weights can be used in an initial solution for the least squares parameters. The latter can then be substituted in the model equation to get model spectrum estimates M'' for each point in the data field. A second least squares calculation can then be performed using the more accurate weight estimates $1/M''$.

Various weighting schemes were tested by using the previously discussed (section 3.3) two-dimensional spectrum with 16 coefficients a_{ij}. Least squares solutions by parts were performed with three varieties of weights and a number of acceptance limits M_ℓ. A basis of comparison was obtained by defining the dispersion parameter

$$d = \sum_{i,j=1}^{4}\frac{(a''_{ij} - a_{ij})^2}{\sigma^2_{ij}}. \qquad 3.4.4$$

The same set of variances σ^2_{ij} was used in all calculations of d. This set was obtained from the covariance matrices $\underline{K}_i^{-1}$ in the solution with weights $W' = 1/M''$ and acceptance limit $M_\ell = 5$. In a good fit d is expected to have a value near 16. Larger values will result when the estimates a''_{ij} have excessive deviations from the model values a_{ij}.

Results are presented in Table V. Column 1 shows the acceptance limit M_ℓ ranging from 0 to 25. Columns 2 and 3 give the number of channels which fell below the limit M_ℓ and were excluded from the least squares calculation. Results in column 2 obtain in calculations with weights $W' = 1/M'$ and $W' = 1/(M'+1)$, whereas column 3 corresponds to the weights $W' = 1/M''$. The last three columns give the parameter d obtained with various weights. The iterative weights $1/M''$ were calculated from the results of an initial solution with $W' = 1/M'$ and $M_\ell = 5$.

In all cases the weights $1/(M'+1)$ appear superior to the weights $1/M'$ in that their associated dispersion parameter d has smaller values. As expected, the difference diminishes with increasing M_ℓ. Also expected is the observed general tendency for d to decrease at higher values of M_ℓ. On the other hand, the iterative weights $1/M''$ yield values of d relatively independent of M_ℓ within the range of values used. It thus appears that in cases where channels with very low content must be used, the iterative weights $1/M''$ are preferable. When M_ℓ can be set at 15 or 20, the weights $1/(M'+1)$ seem to be adequate. The use of weights $1/M'$ is not recommended for any condition.

TABLE V

Test of Various Weighting Functions

Acceptance Limit M_ℓ	No. of Channels Rejected out of 3100		Dispersion Parameter d		
	$M' < M_\ell$	$M'' < M_\ell$	$W' = 1/M'$	$W' = 1/(M'+1)$	$W' = 1/M''$
0	0	-	-	55.0	-
1	167	193	38.5	26.8	12.5
5	588	570	25.3	24.9	13.6
10	750	765	33.4	32.3	14.7
15	836	835	14.9	14.7	14.9
20	890	897	16.0	15.9	14.7
25	939	947	13.8	13.6	14.7

3.5 Application to Nonlinear Models

The models considered thus far had the property of being linear functions of the least squares parameters. This led to linear normal equations which could be solved by relatively simple matrix methods. However, the use of more complicated model functions usually results in a set of normal equations which are very difficult to solve. Some modification in approach is required to facilitate the practical solution for the pertinent parameters.

Nonlinear models are essential for accurate determination of energies in nuclear spectra. Many detector response functions involve a peak which has a shape well approximated by a Gaussian function (e.g. the photo-peak in Figure 1). In order to determine the central position-and hence the energy associated with the Gaussian peak - it is necessary to perform a least squares calculation based on this nonlinear model.

One practical method of solution[40,44,47] linearizes the model function by expanding it in a Taylor series and neglecting terms which are higher than the first order. With this method it is necessary to provide initial estimates of the least squares parameters. If the estimates are good, then the required corrections turn out to be small and the linear approximation is fairly accurate. The calculations can be repeated with the new corrected values being used as estimates. After several such interations the calculated correctionsusually become insignificant and the process can be stopped.

We define a general model function

$$M(x) = M(a_1, a_2, \ldots, a_n ; x) \qquad 3.5.1$$

and denote by $M^o(x)$ the function obtained when initial estimates a_i^o are substituted in (3.5.1). The function M(x) can be approximated by the truncated Taylor series

$$M^1(x) = M^o(x) + \sum_{i=1}^{n} \frac{\partial M^o(x)}{\partial a_i^o} \Delta^1 a_i \,, \qquad 3.5.2$$

where

$$\Delta^1 a_i = a_i - a_i^o \,. \qquad 3.5.3$$

We can now state the least squares condition

$$\sum_{x=1}^{N} \frac{1}{M^1(x)} \left[M'(x) - M^o(x) - \sum_{i=1}^{n} \frac{\partial M^o(x)}{\partial a_i^o} \Delta^1 a_i \right]^2 \longrightarrow \min, \qquad 3.5.4$$

where, as before, the data vector is represented by $M'(x)$. Treating the corrections $\Delta^1 a_i$ as our least squares parameters, we obtain the set of n normal equations

$$\sum_{x=1}^{N} \frac{1}{M^1(x)} \left[M'(x) - M^o(x) \right] \frac{\partial M^o(x)}{\partial a_k^o} = \sum_{x=1}^{N} \frac{1}{M^1(x)} \frac{\partial M^o(x)}{\partial a_k^o} \sum_{i=1}^{n} \frac{\partial M^o(x)}{\partial a_i^o} \Delta^{1'} a_i \,, \qquad 3.5.5$$

where the index k can have values 1, 2, ..., n.

These results can be gathered into a neater matrix representation. Using the symbolism of section 3.4 we define the matrices and vectors

$$\underline{M}^1 = \left\| \left\{ M^1(x) - M^o(x) \right\}_{x,1} \right\|^{N,1} \,, \qquad 3.5.6$$

$$\underline{D}^o = \left\| \left\{ \frac{\partial M^o(x)}{\partial a_i^o} \right\}_{x,i} \right\|^{N,n} \,, \qquad 3.5.7$$

$$\underline{\Delta}^1 = \left\| \left\{ \Delta^1 a_i \right\}_{i,1} \right\|^{n,1} , \qquad 3.5.8$$

$$\underline{W}^1 = \left\| \left\{ \frac{\delta_{ix}}{M^1(x)} \right\}_{i,x} \right\|^{N,N} , \qquad 3.5.9$$

$$\underline{B}^o = \underline{D}^{oT} \underline{W}^1 \underline{D}^o$$

$$= \left\| \left\{ \sum_{x=1}^{N} \frac{1}{M^1(x)} \frac{\partial M^o(x)}{\partial a^o_i} \frac{\partial M^o(x)}{\partial a^o_j} \right\}_{i,j} \right\|^{n.n} . \qquad 3.5.10$$

The model equation (3.5.2) becomes now

$$\underline{M}^1 = \underline{D}^o \underline{\Delta}^1 \qquad 3.5.11$$

and the normal equations (3.5.5) can be replaced by

$$\underline{B}^o \underline{\Delta}^{1'} = \underline{D}^{oT} \underline{W}^1 \underline{M}^{1'} , \qquad 3.5.12$$

where

$$\underline{M}^{1'} = \left\| \left\{ M'(x) - M^o(x) \right\}_{x,1} \right\|^{N,1} . \qquad 3.5.13$$

Solution for the vector of corrections $\underline{\Delta}^{1'}$ can be effected by the inversion of design matrix $\underline{B}^o$. The components of this vector are added to the initial estimates a^o_i to obtain a new set of estimates

$$a^{1'}_i = a^o_i + \Delta^1 a_i . \qquad 3.5.14$$

If all superscripts are increased by unity, then the same set of equations (3.5.2) to (3.5.14) will describe the second iteration. The iterations are continued until the added corrections in equation

(3.5.14) have negligible effect.

In general there is no assurance that the process converges to give estimates a_i' which will minimize the function

$$\chi'^2 = \sum_x \frac{1}{\sigma^2} \left[x'(x) - M(a_1', a_2', \ldots, a_n'; x) \right]^2 . \quad 3.5.15$$

There may exist local minima which provide non-optimum estimates a_i', or the process may not converge at all. Much depends on the model function used and the choice of initial estimates. Fortunately a Gaussian model has good convergence properties and its parameters are usually easy to estimate by a visual inspection of the spectrum. Thus the method is well suited for determination of transition energies. However, when the model includes more than one peak located within the width σ, the results are not always predictable. In such cases use might be made of the "parabolic method" discussed in detail by Archer[47]. Near its minimum value, χ'^2 is nearly a quadratic function of the parameters a_i'. By calculating χ'^2 at various values of a_i' and fitting the results to quadratic functions it is often possible to find a set of parameters a_i' which will minimize χ'^2.

In general the estimates a_i' are not normally distributed when the model function is nonlinear. Their exact distributions are not easily determined and may require numerical Monte Carlo calculations. However, if certain conditions are satisfied, the actual distributions can be nearly normal. Consider a Taylor series expansion of M(x) about the unknown true values of parameters

a_i. We have

$$M(a_1', a_2', \ldots a_n' ; x) = M(x) + \sum_{i=1}^{n} \frac{\partial M(x)}{\partial a_i} (a_i' - a_i)$$

$$+ \text{(higher order terms)} . \qquad 3.5.16$$

If we assume now that the variances in estimates a_i' are small, then over the range of most probable values of a_i' the differences $(a_i' - a_i)$ will also be small. Consequently we can neglect the higher order terms and expression (3.5.16) becomes linear in the parameter estimates a_i'. When the initial estimates a_i^o are not too different from the true values, the iterative solution is independent of the values a_i^o and converges to the same set a_i'. Therefore we can assume that our initial estimates were the true values a_i without altering the distribution in a_i'. With the linear approximation in (3.5.16) only one iteration will be required to obtain the least squares estimates a_i' and this solution will have the same statistical properties as the solutions of the linear model in sections 3.1 and 3.2. Consequently, we can say that the estimates a_i' are normally distributed, that they have expectation values given by

$$E(a_i') = a_i \ , \qquad 3.5.17$$

and that their covariance matrix is

$$\underline{C}_{A'} = \underline{B}^{-1} \ , \qquad 3.5.16$$

where

$$\underline{B} = \left\| \left\{ \sum_{x=1}^{N} \frac{1}{M(x)} \frac{\partial M(x)}{\partial a_i} \frac{\partial M(x)}{\partial a_j} \right\}_{i,j} \right\|^{n,n} . \qquad 3.5.19$$

Since Gaussian functions find frequent application in non-

linear least squares calculations, it is perhaps worthwhile to derive an explicit expression for the covariance matrix of this particular model. As our model function we take the expression

$$M(x) = a_1 \exp\left[-\frac{(x - a_2)^2}{2a_3^2}\right] \qquad 3.5.20$$

and assume that the width a_3 is sufficiently large to permit replacement of the summation in (3.5.19) by an integral sign.

We shall consider two particular cases. In the first instance it will be assumed that the available data extends over the full range of the Gaussian peak and the limits of integration used will be $-\infty$ to $+\infty$. Another case of interest occurs when only the right half of the peak approximates a Gaussian shape. For example, the response shapes of NaI(Tℓ) detectors may contain small angle scattering events and contributions from the Compton scattering distribution which tend to distort the left half of the Gaussian photo-peak. For this case only the right half of the peak will be considered by using the limits of integration a_2 to ∞.

We start by considering the full Gaussian peak. Substitution of M(x) and its partial derivatives in expression (3.5.19) yields the design matrix

$$\underline{B} = \begin{Vmatrix} \frac{N_o}{a_1^2} & 0 & \frac{N_o}{a_1 a_3} \\ 0 & \frac{N_o}{a_3^2} & 0 \\ \frac{N_o}{a_1 a_3} & 0 & \frac{3N_o}{a_3^2} \end{Vmatrix}, \qquad 3.5.21$$

where N_o is the total number of counts under the peak, given by

$$N_o = \sqrt{2\pi}\ a_1\ a_3. \qquad 3.5.22$$

The matrix is easily inverted to obtain the covariance matrix

$$\underline{C}_{A'} = \left\| \begin{array}{ccc} \frac{3\,a_1^2}{2N_o} & 0 & -\frac{a_1 a_3}{2N_o} \\ 0 & \frac{a_3^2}{N_o} & 0 \\ \frac{-a_1 a_3}{2N_o} & 0 & \frac{a_3^2}{2N_o} \end{array} \right\| . \qquad 3.5.23$$

Of some special interest is the central covariance matrix element a_3^2/N_o which provides the variance in estimate of the peak position a_2, $\sigma_{a'_2}^2$. To take a practical example, consider a Gaussian peak consisting of $N_o = 10^6$ counts. Its full width at half maximum is given by $2.35\ a_3$ and its central position is determined within an accuracy of $\pm\ a_3/\sqrt{N_o}$. The ratio of position uncertainty to the peak width is 1/2350. This high precision demonstrates the power in the method of least squares analysis.

Next we turn to the case where usable data covers only half of the peak. Using the integration limits a_2 to ∞ we obtain the design matrix

$$\underline{B} = \left\| \begin{array}{ccc} \frac{N_o}{2a_1^2} & 1 & \frac{N_o}{2a_1 a_3} \\ 1 & \frac{N_o}{2a_3^2} & \frac{2\,a_1}{a_3} \\ \frac{N_o}{2a_1 a_3} & \frac{2a_1}{a_3} & \frac{3\,N_o}{2a_3^2} \end{array} \right\| . \qquad 3.5.24$$

After a considerable amount of algebraic manipulation it is possible to obtain the expression for the inverse matrix.

$$\underline{C}_{A'} = \left\| \begin{array}{c|c|c} \left(\frac{3\pi-8}{\pi-3}\right)\frac{a_1^2}{N_o} & -\frac{1}{(\pi-3)} & \frac{(4-\pi)}{(\pi-3)}\frac{a_1 a_3}{N_o} \\ \hline -\frac{1}{(\pi-3)} & \frac{2\pi}{(\pi-3)}\frac{a_3^2}{N_o} & \frac{-\sqrt{2\pi}}{(\pi-3)}\frac{a_3^2}{N_o} \\ \hline \frac{(4-\pi)}{(\pi-3)}\frac{a_1 a_3}{N_o} & \frac{-\sqrt{2\pi}}{(\pi-3)}\frac{a_3^2}{N_o} & \frac{(\pi-2)}{(\pi-3)}\frac{a_3^2}{N_o} \end{array} \right\| \quad . \qquad 3.5.25$$

From the above result we see that the use of only half the peak has increased the variance $\sigma^2_{a'_2}$ by a factor $2\pi/(\pi-3) = 44.3$. The uncertainty, or standard deviation, is increased by the factor $\sqrt{44.3} = 6.66$. Uncertainties in the estimates of a_1 and a_3 have increased by the factors 2.59 and 4.02 respectively.

In the foregoing discussion it was assumed that the Gaussian width a_3 is sufficiently large to permit neglection of channel integration effects. When a_3 is small it becomes necessary to replace equation (3.5.20) by the more accurate expression

$$M(x) = a_1 \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} \exp\left[-\frac{(y-a_2)^2}{2a_3^2}\right] dy$$

$$= \frac{N_o}{2}\left\{ \operatorname{erf}\left[\frac{x+\frac{1}{2}-a_2}{\sqrt{2}\,a_3}\right] - \operatorname{erf}\left[\frac{x-\frac{1}{2}-a_2}{\sqrt{2}\,a_3}\right]\right\} , \qquad 3.5.26$$

where the error function is defined by

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt \, . \qquad 3.5.27$$

It also becomes necessary to retain the summation sign in design matrix definition (3.5.19), since the use of integrals would result

in a too crude approximation. Consequently simple expressions for the matrix elements are not available and the calculations were performed by numerical methods.

Elements of design matrix $\underline{B}$ were calculated on the assumption that data are available over the full range of the peak. Model equation (3.5.26) contains the three constants N_o , a_2 and a_3 which were used as the least squares parameters. Upon inversion of the design matrix $\underline{B}$ it was possible to obtain numerical values of $N_o \sigma^2_{a_2'}$ independently of the value N_o. Therefore it was necessary to vary only the values of a_2 and a_3 in order to study the behaviour of $\sigma^2_{a_2'}$ at various parameter values.

The quantity $N_o \sigma^2_{a_2'} / a_3^2$ is plotted as a function of width a_3 in Figure 17. Two curves are shown: one for the peak placed in the middle of a channel ($a_2 = 0$) and one for a peak position halfway between channel centers ($a_2 = \frac{1}{2}$). Below the value $a_3 = 0.5$ the curves show considerable divergence. The curve with $a_2 = \frac{1}{2}$ remains close to unity, conforming with the previous approximate result in covariance matrix (3.5.23), whereas the curve for $a_2 = 0$ displays a sharp rise with decreasing values of the width a_3.

The significance of the variance $\sigma^2_{a_2'}$ at very low values of a_3 is questionable since the probability density function of estimate a_2' loses its Gaussian shape. In fact, as the width a_3 vanishes, the peak position becomes indeterminate within the limits of channel width and the probability density function of a_2' assumes the shape of channel profile P(x), as defined by (2.5.28a). In this case the actual variance becomes

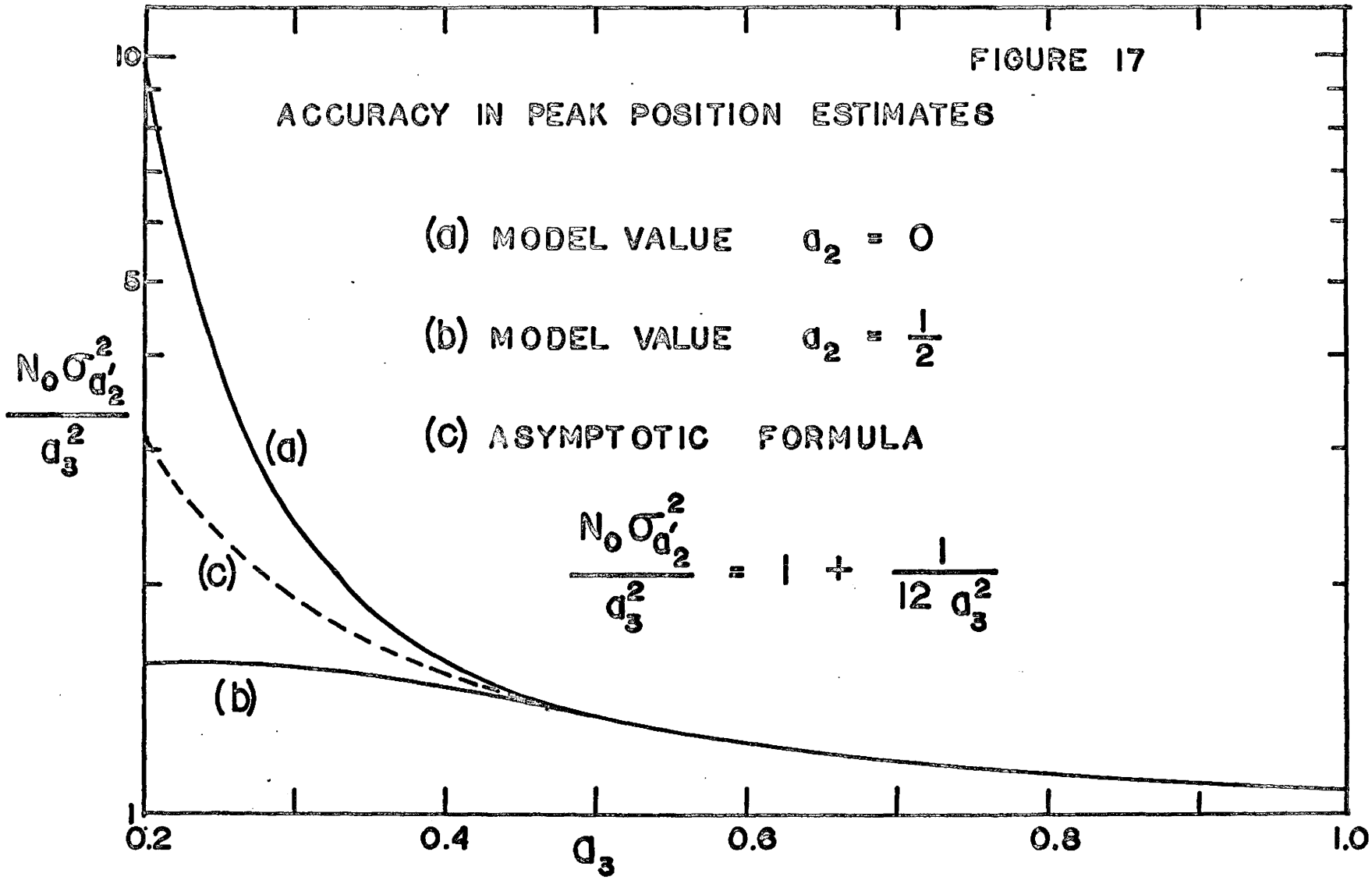


FIGURE 17
ACCURACY IN PEAK POSITION ESTIMATES
(a) MODEL VALUE $a_2 = 0$
(b) MODEL VALUE $a_2 = \frac{1}{2}$
(C) ASYMPTOTIC FORMULA
$\frac{N_0 \sigma^2_{a'_2}}{a_3^2} = 1 + \frac{1}{12\, a_3^2}$
$\frac{N_0 \sigma^2_{a'_2}}{a_3^2}$
10
5
1
(a)
(c)
(b)
0.2
0.4
0.6
0.8
1.0
a_3

$$\sigma^2_{a_2'} = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^2 \, dx = \frac{1}{12} \qquad 3.5.28$$

and is independent of N_o. This channel variance manifests itself at all values of a_3 and the asymptotic expression for variance in estimate a_2' is

$$\sigma^2_{a_2'} \sim \frac{1}{N_o} \left[a_3^2 + \frac{1}{12} \right]. \qquad 3.5.29$$

A curve corresponding to this asymptotic expression is included in Figure 17. It is seen that all three curves converge very rapidly at $a_3 = 0.5$.

In order to test the actual distributions of estimates a_2', a number of Monte Carlo calculations were performed with various values of width parameter a_3. Spectra were generated by using values obtained from the model (3.5.26) and adding normally distributed random numbers having variances equal to M(x). Three cases were studied with a_3 fixed at 4.0, 0.5 and 0.2. In each case the model values of N_o and a_2 remained constant at 100 counts and 0 channels respectively. Figure 18 shows histograms of the least squares solutions a_2'. Gaussian probability density functions with variances $\sigma^2_{a_2'}$ taken from Figure 17 (curve with $a_2 = 0$) are also plotted for the sake of obtaining a comparison. Except for the case with the low value $a_3 = 0.2$, the estimates a_2' appear to be normally distributed.

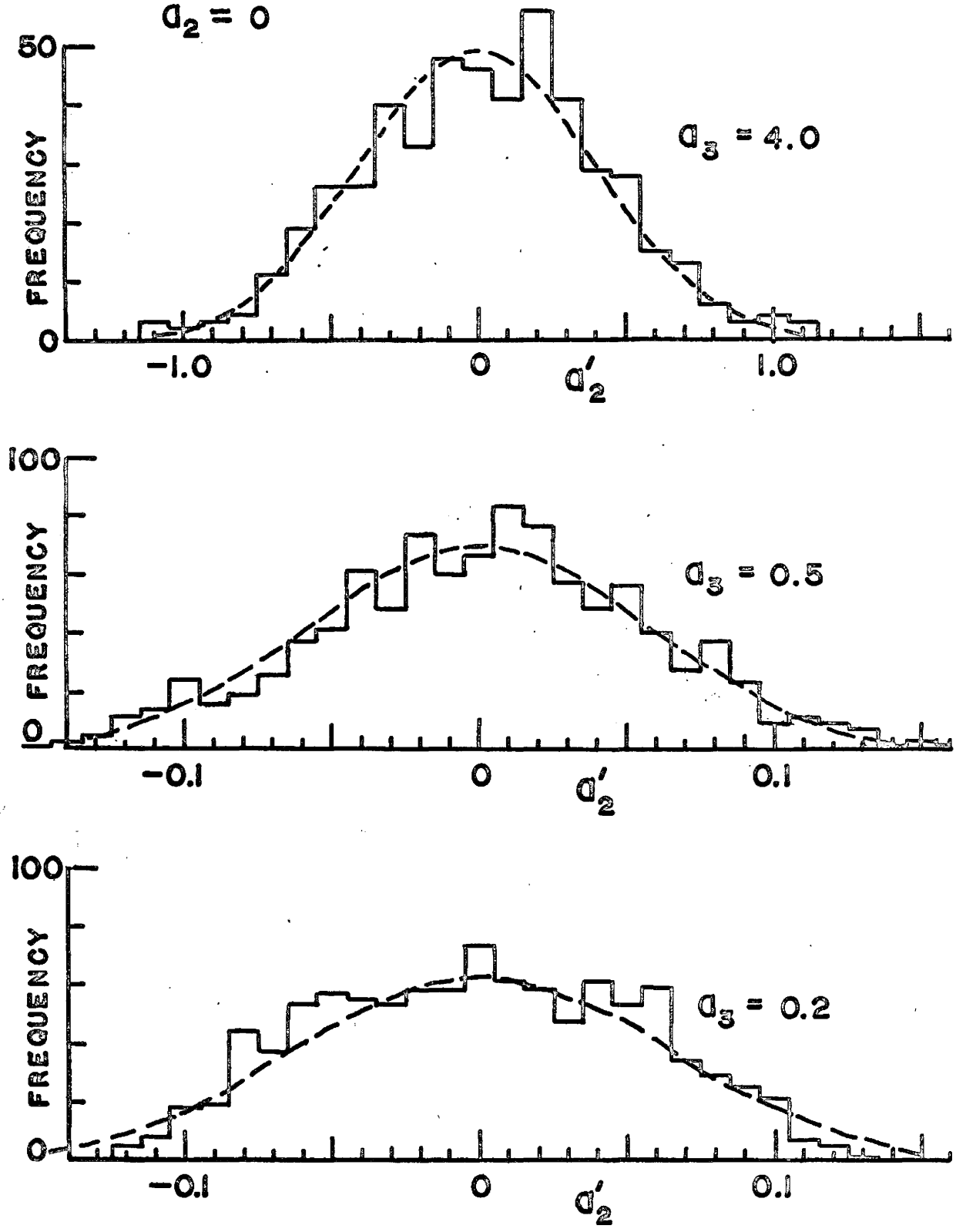
Results of the above calculations can be summed up as follows. The nonlinear least squares solutions provide normally distributed estimates of Gaussian peak positions for values of N_o as low as 100

FIGURE 18

HISTOGRAMS OF PEAK POSITION ESTIMATES

$N_0 = 100$

$a_2 = 0$



counts (or lower) and values of width a_3 as low as 0.5 channels. When $a_3 > 0.5$, the variances in position estimates a_2' are very closely approximated by the asymptotic expression (3.5.29). Values of a_3 smaller than 0.5 channels should be avoided due to the possibility of a large loss in accuracy. Since a_3 is measured in channel widths, its value can be increased by redesigning the experiment to yield a finer channel mesh.

CHAPTER IV

CONCLUSIONS

The methods of analysis presented in this thesis were discussed with special emphasis to application in nuclear spectroscopy. However, use of this emphasis was not intended to be indicative of a limitation in possible applications. The techniques presented can be generally applied to digitized statistical (or nonstatistical) data which contain the distorting effects of apparatus response functions. Statistical considerations were based on Poisson and Gaussian frequency functions which are found to occur in a wide class of physical measurements.

Two essentially different methods of approach were discussed. Both are concerned with the same problem of obtaining the physically meaningful parameters with least possible error. The matrix inversion approach requires no initial knowledge of the form the unfolded spectrum might take. It is thus convenient in cases where this information is unavailable or difficult to obtain. On the other hand, the least squares approach generally requires a model of the unfolded spectrum to the extent that the number of components and their approximate positions must be specified in advance. This often requires a detailed inspection of the original data followed by an educated guess of the pertinent unfolded spectrum parameters. However, the least squares approach has also some important advantages. Unlike the matrix model approach, it is not limited to integral values of the spectrum components. The removal of this restriction is important when accurate position determinations of narrow peaks

are required. In some applications the two approaches might be combined to take full advantage of their relative merits. A preliminary application of response matrix inversion may provide a useful visual representation of the unfolded spectrum and thus delineate the appropriate least squares model with a smaller element of uncertainty.

It was shown in Chapter II that the application of some inverse response matrices may lead to a gross magnification in statistical uncertainties. Two methods of combating this undesirable effect were suggested. It is possible to obtain a "partial inverse" matrix which will in effect replace the originally broad response functions in the observed spectrum by a set of narrower ones of some specified shape. Since this "partial inverse" matrix has elements smaller in magnitude than the full inverse matrix, the statistical uncertainties in the unfolded spectrum can be greatly reduced. Another method of dealing with statistical deviations is based on the a priori knowledge of non-negativity in spectral intensities. Corrections are added to the observed spectrum so as to render the unfolded spectrum positive (or zero) in every channel. The particular set of corrections to be added is determined by the condition that the sum of the weighted squares of their individual values should be minimized.

When the data vector is multiplied by the inverse matrix it is a simple matter to calculate the statistical variances in the resultant unfolded spectrum (see formula (2.6.1)). However, after application of non-negativity the determination of statistical uncertainties is somewhat more difficult. After non-negativity corrections are added to the observed spectrum, the result is a linear combination of response

functions with non-negative amplitude coefficients. These coefficients - one for each channel - represent the unfolded non-negative spectrum. The same unfolded spectrum would be obtained in a least squares calculation if one used a model equation which contains one response function with non-negative amplitude for each channel. The model could avoid negative amplitudes by the use of quadratic coefficients, in which case the non-linear least squares techniques discussed in section 3.5 must be applied. Therefore, in principle, one could obtain the statistical properties of the completely unfolded non-negative spectrum from the inverse of the corresponding least squares design matrix.

Similar arguments can be applied to cases where a "partial inverse" matrix is used in conjunction with the application of non-negativity. In order to obtain a least squares analogy it is first necessary to construct a set of complementary response functions. These can be obtained by unfolding the residual response functions from the full width original response shapes. The corresponding least squares model is then constructed by placing one complementary response function with non-negative amplitude in each data channel. Since the least squares solution gives a spectrum from which, in effect, the complementary functions have been unfolded, the remaining response shapes will be the same as in the application of partial matrix inversion. The covariance matrix (inverse of the design matrix) of this least squares model determines the statistical properties of the partially unfolded spectrum.

It is thus seen how the two basically different methods of approach can become identical under certain conditions. However, when the data fields are large, it is not practical to apply ordinary least squares

techniques to the solution of intensities in each channel. The required inversion of the design matrix becomes unfeasible due to its large size. In these cases the response matrix approach can be of decisive advantage when the inverse matrix elements are obtainable in closed form. The proper application of non-negativity conditions may require extensive numerical calculations based on methods of mathematical programming. However, it was demonstrated in section 2.6 that at least in some cases a very rapid iterative method can be used.

APPENDIX I

STATISTICAL PROPERTIES OF CHANNEL CONTENT

Assume that for a large number (approaching infinity) of identical experiments, the average count in channel i is given by m_i. It is desired to find the relative probabilities of obtaining various numbers of counts m_i'. To do this we postulate a large number N ($N >> m_i$) of "potential" counts. Each potential count has a probability $\frac{m_i}{N}$ of becoming a "real" count and is statistically independent of all others. The probability for a particular set of m_i' potential counts materializing and the remainder being not counted is given by

$$\left(\frac{m_i}{N}\right)^{m_i'} \left(1 - \frac{m_i}{N}\right)^{N-m_i'} .$$

Multiplying the last expression by the total number of such sets we obtain the probability of getting m_i' real counts in channel i, namely

$$\frac{N!}{(N-m_i')!\ (m_i')!} \left(\frac{m_i}{N}\right)^{m_i'} \left(1 - \frac{m_i}{N}\right)^{N-m_i'} . \qquad \text{Al.1}$$

We now let N approach infinity, under which condition

$$\frac{N!}{(N-m_i')!\ N^{m_i'}} \longrightarrow 1 ,$$

and

$$\left(1 - \frac{m_i}{N}\right)^{N-m_i'} \longrightarrow e^{-m_i} .$$

Therefore, in the limit, we obtain the required frequency function for m_i',

$$p(m_i', m_i) = \frac{m_i'^{\,m_i'}}{(m_i')!}\, e^{-m_i} \quad . \qquad \text{Al.2}$$

It should be noted that this frequency function also holds true for the contents of histogram bins, provided the histogram is made up of a large number of statistically independent experiments and is spread out over many bins of non-zero content. If the histogram is concentrated around one bin, then it is necessary to use expression (Al.1) with N set equal to the total number of experiments.

When the average value m_i is large, expression (Al.2) approaches a Gaussian frequency function. To show this we use Stirling's formula[54]

$$n! \sim e^{-n}\, n^n \sqrt{2\pi n} \, . \qquad \text{Al.3}$$

We can thus write

$$\log p \sim (m_i' + \tfrac{1}{2}) \log\left(\frac{m_i}{m_i'}\right) - (m_i - m_i') - \tfrac{1}{2} \log (2\pi\, m_i).$$

Next we use the series expansion[55]

$$\log\left(\frac{m_i}{m_i'}\right) = 2\left\{ \frac{m_i - m_i'}{m_i + m_i'} + \frac{1}{3}\left(\frac{m_i - m_i'}{m_i + m_i'}\right)^3 + \cdots \right\} ,$$

in which the cubic and higher order terms can be neglected since, for values of p significantly greater than zero, we have

$$\frac{m_i - m_i'}{m_i + m_i'} \ll 1 \quad . \qquad \text{Al.4}$$

We now can write the result

$$\log p \sim (2m_i' + 1)\frac{(m_i - m_i')}{(m_i + m_i')} - (m_i - m_i') - \frac{1}{2}\log(2\pi m_i)$$

$$= \frac{-(m_i' - m_i)^2}{2m_i}\,\frac{2m_i}{(m_i + m_i')} - \frac{1}{2}\log(2\pi m_i) \quad .$$

When the strong inequality (A1.4) holds we have also the approximate relation

$$\frac{2m_i}{m_i + m_i'} \approx 1 \quad ,$$

in which case we can write the final asymptotic frequency function

$$p(m_i', m_i) \sim \frac{1}{\sqrt{2\pi m_i}}\, e^{\frac{-(m_i' - m_i)^2}{2m_i}} \quad . \qquad \text{A1.5}$$

A comparison between the Gaussian and Poisson frequency functions is given in Figure 19 for various values of m_i. It is seen that the similarity is fairly good at $m_i = 5$ and becomes increasingly better at higher values of m_i.

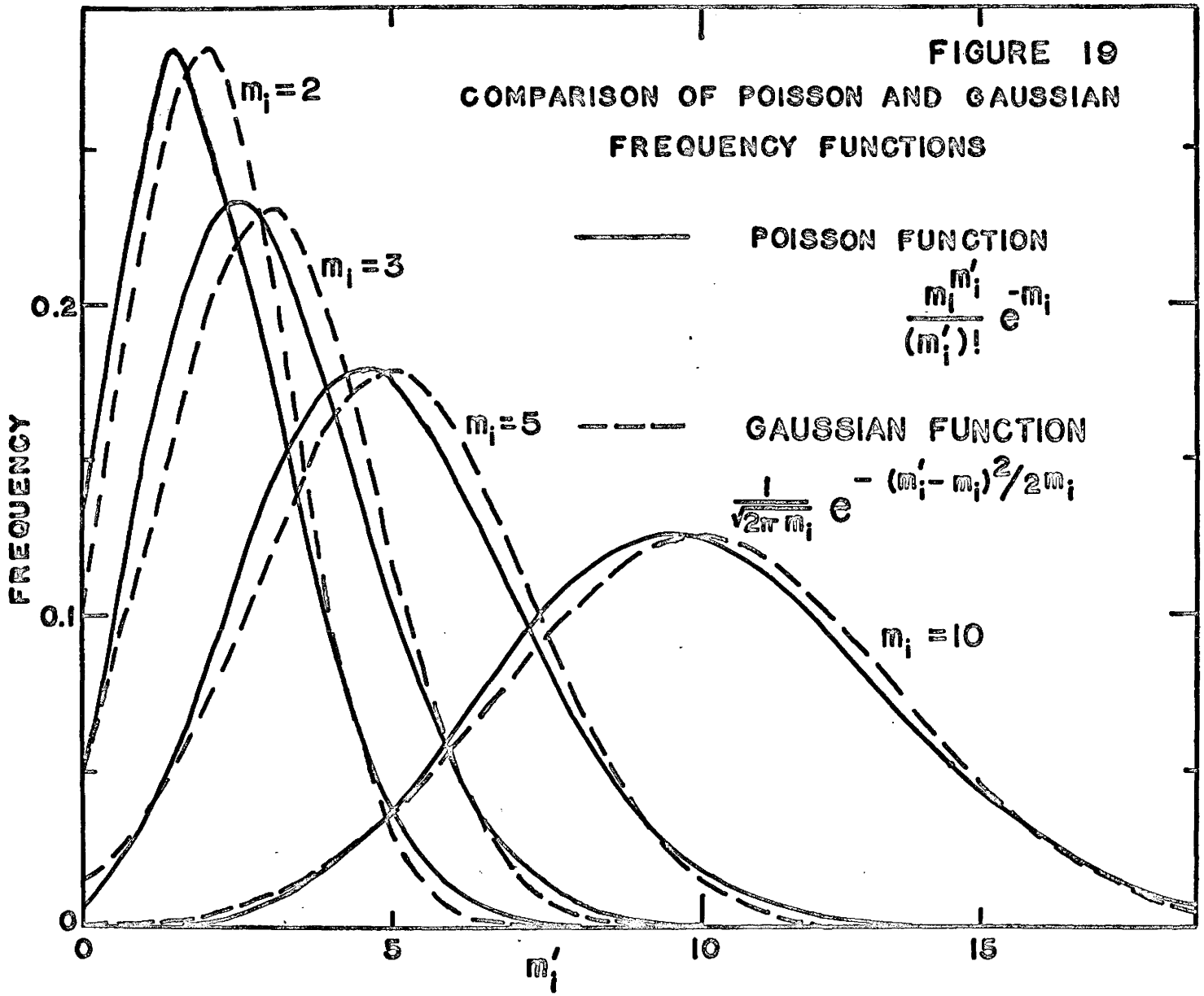


FIGURE 19
COMPARISON OF POISSON AND GAUSSIAN
FREQUENCY FUNCTIONS
POISSON FUNCTION
$\frac{m_i^{m_i'}}{(m_i')!} e^{-m_i}$
GAUSSIAN FUNCTION
$\frac{1}{\sqrt{2\pi m_i}} e^{-(m_i'-m_i)^2/2m_i}$
$m_i = 2$
$m_i = 3$
$m_i = 5$
$m_i = 10$
FREQUENCY
0.2
0.1
0
0
5
10
15
m_i'

APPENDIX II

CORRECTIONS FOR COINCIDENCE AND CHANCE SUMMING

The number of prompt coincidence summing events is given by

$$N_{cs} = \sum_{i,j} \varepsilon_i \; \varepsilon_j \; \mathcal{N}_{ij} \; W_{ij} \; , \qquad \text{A2.1}$$

where $\mathcal{N}_{ij}$ is the number of transitions γ_i followed by γ_j, W_{ij} is the angular correlation function integrated over the solid angle Ω subtended by the detector and ε_i is the overall detection efficiency for γ_i. Since efficiencies are proportional to Ω , we can write

$$\varepsilon_i = k_i \Omega \quad , \qquad \text{A2.2}$$

which results in

$$N_{cs} = \Omega^2 \sum_{i,j} k_i \, k_j \; \mathcal{N}_{ij} \, W_{ij} \; . \qquad \text{A2.3}$$

The number of single events is simply

$$N_s = \Omega \sum_i k_i \; \mathcal{N}_i \; . \qquad \text{A2.4}$$

Consequently the relative contribution from coincident summing, N_{cs}/N_s is proportional to the solid angle Ω . Therefore the effect can be reduced by placing the source at a greater distance from the detector.

Contribution due to random summing of uncorrelated events can be calculated from the following considerations. We have N_s single events occurring during observation time T. Each of these events has the probability $2 \tau N_s/T$ of being followed or preceded by another event within the electronic resolution time τ . Therefore the expectation of the total random summing events is

$$N_{rs} = \frac{2\tau}{T} N_s^2 . \qquad A2.5$$

The relative contribution N_{rs}/N_s is thus proportional to the counting rate N_s/T and the effect can be reduced by using weaker sources or smaller solid angle.

Practical considerations often require relatively high counting rates which could necessitate correction for summing effects. It is possible to make these corrections by using special experimental techniques or by applying appropriate post-experimental calculations.

Random sum pulses can be rejected to some extent by electronic circuitry which senses the pulse broadening[4]. However this method is not 100% efficient and has no effect on coincident sum events.

A powerful experimental technique can be applied by the use of two identical detectors[3]. The mixed output from both detectors is stored in one subgroup of the analyzer memory. Another memory subgroup receives the gated sum of pulses whenever both detectors respond in coincidence. The contents of the gated subgroup can then be subtracted from the mixed spectrum. A great advantage of this technique is that correction can be made for both random and coincident sum events[5].

An expression for the spectrum of coincident sum events has been given by Heath[2]. Let the response functions to transitions γ_i and γ_j be given by $R(x,y_i)$ and $R(x,y_j)$ respectively. The two-dimensional probability density function for γ_i to produce a pulse height x and for γ_j to produce a pulse-height z-x is given by the product

$$R(x,y_i)\, R(z-x,y_j).$$

The one-dimensional probability density function of obtaining two pulses

which add up to the value z is given by the integral

$$\int_0^z R(x,y_i)\, R(z - x, y_j)dx \,.$$

To obtain the total coincidence sum spectrum we must multiply this last expression by the number of coincidences $\mathcal{N}_{ij}$, the efficiencies ε_i, ε_j and then sum over all values of i and j. Thus we get the result

$$M_{cs}(z) = \sum_{i,j} \varepsilon_i \varepsilon_j \mathcal{N}_{ij} W_{ij} \int_0^z R(x,y_i)\, R(z-x,y_j)dx \,. \qquad A2.6$$

One limitation of this formula is that advance knowledge of the decay properties (in particular the values of $\mathcal{N}_{ij}$) is required.

The spectrum due to random summing events is more difficult to calculate since, due to possible time-separation, the resultant pulse height may be not equal to the sum of individual pulse heights. A method taking this effect into account has been described in detail by Kennett et al[5] and shall not be discussed in this thesis. This random summing spectrum is independent of cascades and can be calculated without knowledge of coincidences $\mathcal{N}_{ij}$.

APPENDIX III

DERIVATION OF INVERSE MATRIX $\underline{L}^{-1}$

We wish to find the elements of matrix $\underline{L}^{-1}$, such that

$$\underline{L}^{-1}\,\underline{L} = \underline{I}\,. \tag{A3.1}$$

Since matrix $\underline{L}$ is upper triangular, the inverse $\underline{L}^{-1}$ also has the same property and we can immediately write

$$\ell^{-1}_{ij} = 0 \quad \text{for} \quad i > j\,. \tag{A3.2}$$

To find the diagonal elements of $\underline{L}^{-1}$ we multiply the k^{th} row of $\underline{L}^{-1}$ by the k^{th} column of $\underline{L}$. The result is

$$\ell^{-1}_{kk}\ \ell_{kk} = 1\,. \tag{A3.3}$$

By definition (2.4.11) for matrix $\underline{L}$ we can express the elements

$$\begin{aligned} \ell_{kk} &= \alpha_k, \\ \ell_{mn} &= \frac{\beta_n}{n-1} \quad \text{for } n > m\,, \\ &= 0 \quad \text{for } n < m\,. \end{aligned} \tag{A3.4}$$

Consequently, by virtue of (A3.3) we have the diagonal elements

$$\ell^{-1}_{kk} = \frac{1}{\alpha_k}\,. \tag{A3.5}$$

It remains to find the upper off-diagonal elements of $\underline{L}^{-1}$. To accomplish this, consider the product of the k^{th} row of $\underline{L}^{-1}$ by the $(k+j)^{th}$ column of $\underline{L}$, where $0 < j \leq N - k$. Writing out this product in detail we have

$$\sum_{i=0}^{j} \ell^{-1}_{kk+i} \, \ell_{k+ik+j} = 0. \qquad \text{A3.6}$$

Substitution of elements (A3.4) in (A3.6) yields the result

$$\ell^{-1}_{kk+j} = \frac{-\beta_{k+j}}{\alpha_{k+j}\,(k+j-1)} \sum_{i=1}^{j-1} \ell^{-1}_{kk+i} \, . \qquad \text{A3.7}$$

By using similar arguments we can arrive at the expression

$$\ell^{-1}_{kk+j+1} = -\frac{\beta_{k+j+1}}{\alpha_{k+j+1}\,(k+j)} \sum_{i=1}^{j} \ell^{-1}_{kk+i} \, . \qquad \text{A3.8}$$

Combining (A3.7) with (A3.8) we can obtain the recursion relation

$$R_{kj} \equiv \frac{\ell^{-1}_{kk+j+1}}{\ell^{-1}_{kk+j}} = \frac{\beta_{k+j+1}}{\alpha_{k+j+1}\,(k+j)} \left[\frac{\alpha_{k+j}\,(k+j-1)}{\beta_{k+j}} - 1 \right] . \qquad \text{A3.9}$$

Using the quantity γ_j defined by (2.4.12) the above ratio can be written in the form

$$R_{kj} = \frac{(\gamma_{k+j+1} - 1)}{(\gamma_{k+j} - 1)} \gamma_{k+j} \, . \qquad \text{A3.10}$$

We already have determined the diagonal inverse element in the k^{th} row (see expression (A3.5)). To obtain the next element in this row we multiply the k^{th} row of $\underline{L}^{-1}$ by the $(k+1)^{th}$ column of $\underline{L}$. This results in the expression

$$\ell^{-1}_{kk+1} = \frac{1}{\alpha_k} (\gamma_{k+1} - 1) \, . \qquad \text{A3.11}$$

By the recursion relation (A3.10) the next inverse element in the k^{th} row is given by

$$\ell^{-1}_{kk+2} = \frac{1}{\alpha_k} (\gamma_{k+2} - 1) \gamma_{k+1} \, . \qquad \text{A3.12}$$

It is easily seen that by repeated application of the recursion formula we obtain a general inverse matrix element

$$\ell_{kj}^{-1} = \frac{1}{\alpha_k} (\gamma_j - 1) \prod_{i=k+1}^{j-1} \gamma_i \quad \text{for } k < j . \qquad \text{A3.13}$$

This completes the derivation of expression (2.4.13).

It is of interest to derive an effort factor (total number of multiplications and divisions) required for inversion of matrix $\underline{L}$. There are (N-1) values of γ_i. Each requires one multiplication and one division to evaluate. This results in a total number of operations

$$M_\gamma = 2(N-1) . \qquad \text{A3.14}$$

The recursion relation (A3.10) involves a total of (N-2) factors. Since each factor is evaluated by two operations we have a total of operations

$$M_R = 2(N-2) . \qquad \text{A3.15}$$

Now we consider the evaluation of inverse elements. Each of the $\frac{N}{2}(N+1)$ non-zero elements requires either one division or one multiplication. Consequently the total number of operations is given by

$$M = \frac{N^2}{2} + \frac{N}{2} + M_\gamma + M_R$$

$$= \frac{N^2}{2} + \frac{9N}{2} - 6 . \qquad \text{A3.16}$$

This is a considerable improvement over the cubic relation (2.3.14) for the inversion of a general triangular matrix.

When parameters α and β are assumed constant the required effort is somewhat less than the effort given by (A3.16). After determination of β/α each γ_i requires only one division. Therefore in this case

$$M_\gamma^c = (N-1) + 1 = N . \qquad \text{A3.17}$$

Since the γ_i's are different the number M_R remains the same as in (A3.15).

The diagonal elements are all equal to $1/\alpha$ and thus require only one division. All off-diagonal elements are obtained by $N(N-1)/2$ multiplications. Thus the total number of operations is

$$M^c = N + 1 + \frac{N^2}{2} - \frac{N}{2}$$

$$= \frac{N^2}{2} + \frac{N}{2} + 1 \,. \qquad \text{A3.18}$$

APPENDIX IV

FOURIER TRANSFORMS AND CONVOLUTION

First we prove the convolution theorem, which is used in section 2.5. We have the convolution expression (2.5.4a) in the x-domain and wish to derive the product (2.5.4b) in the ω-domain. By substituting the Fourier transforms under the convolution integral we can write

$$F_3(x) = \frac{1}{(2\pi)^2} \iiint_{-\infty}^{\infty} f_1(\omega_1) e^{i\omega_1(x-y)} f_2(\omega_2) e^{i\omega_2 y}\, dy\, d\omega_1\, d\omega_2$$

$$= \frac{1}{2\pi} \iint_{-\infty}^{\infty} f_1(\omega_1)\, f_2(\omega_2)\, e^{i\omega_1 x} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(\omega_2-\omega_1)y}\, dy\right] d\omega_1\, d\omega_2 .$$

The quantity inside the brackets is just one form of the definition[35] for the delta function $\delta(\omega_2-\omega_1)$. Thus we can make the substitution $\omega_2=\omega_1$ and drop the integration over ω_2. This leads to the expression

$$F_3(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_1(\omega_1)\, f_2(\omega_1) e^{i\omega_1 x}\, d\omega_1 . \qquad \text{A4.1}$$

Therefore $F_3(x)$ is the Fourier transform of $f_1(\omega)\, f_2(\omega)$; but by definition it is also the same Fourier transform of $f_3(\omega)$. Hence we can write the equality

$$f_3(\omega) = f_1(\omega)\, f_2(\omega) . \qquad \text{A4.2}$$

When we have convolution in the ω-domain (as in (2.5.5a)), it is possible to arrive at the product expression (2.5.5b) by using arguments identical to the ones presented above. This would complete the proof of both aspects of the convolution theorem.

Next we derive the Fourier transforms of some expressions used in section 2.5. Consider the Gaussian function of (2.5.9a). Its Fourier transform is given by

$$r(\omega) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{2\sigma^2} - i\omega x\right] dx \quad . \qquad A4.3$$

We can complete the square in the exponent and write it in the form

$$-\left(\frac{x}{\sqrt{2}\,\sigma} + \frac{i\omega\sigma}{\sqrt{2}}\right)^2 - \frac{\omega^2\sigma^2}{2} \quad .$$

Then, upon introducing the transformation

$$y = \frac{x}{\sqrt{2}\,\sigma} + \frac{i\omega\sigma}{\sqrt{2}} \quad ,$$

we can rewrite expression (A4.3) ,

$$r(\omega) = \frac{e^{-\frac{\sigma^2\omega^2}{2}}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2} dy \quad . \qquad A4.4$$

The definite integral has the numerical value $\sqrt{\pi}$. Hence the Fourier transform of the Gaussian response function (2.5.9a) is given by

$$r(\omega) = e^{-\frac{\sigma^2\omega^2}{2}} \quad . \qquad A4.5$$

Consider now the set of delta functions

$$D(x) = \sum_{n=-\infty}^{\infty} \delta(x-n) \quad . \qquad A4.6$$

Since this set forms a periodic structure symmetrical about the point $x = 0$, we can expand it in a Fourier cosine series,

$$D(x) = a_o + \sum_{n=1}^{\infty} a_n \cos 2\pi n x \quad . \qquad A4.7$$

We have

$$a_o = \int_{-\frac{1}{2}}^{\frac{1}{2}} D(x)dx = 1 ,$$

$$a_n = 2 \int_{-\frac{1}{2}}^{\frac{1}{2}} D(x) \cos 2\pi nx \, dx$$

$$= 2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \delta(x) \cos 2\pi nx$$

$$= 2 . \qquad \text{A4.8}$$

Therefore we can write

$$D(x) = 1 + 2 \sum_{n=1}^{\infty} \cos 2\pi nx. \qquad \text{A4.9}$$

Taking the Fourier transform of D(x) we obtain

$$d(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} dx + \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} (e^{i2\pi nx} + e^{-i2\pi nx}) e^{-i\omega x} dx$$

$$= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{ix(2\pi n-\omega)} dx$$

$$= 2\pi \sum_{n=-\infty}^{\infty} \delta(\omega - 2\pi n) , \qquad \text{A4.10}$$

which establishes the relation (2.5.10b).

Now consider the channel profile P(x) as defined in expression (2.5.28a). The Fourier transform is

$$p(\omega) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i\omega x} \, dx \, . \qquad \text{A4.11}$$

According to de Moivre's theorem, we have

$$e^{-i\omega x} = \cos \omega x - i \sin \omega x \, .$$

Being assymetrical, the second term has no contribution to the integral in (A4.11); hence we can write

$$p(\omega) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos \omega x \, dx$$

$$= \frac{2}{\omega} \sin \left(\frac{\omega}{2} \right) , \qquad \text{A4.12}$$

which proves the relation (2.5.28b).

It remains now to derive the result (2.6.3) for "partial inversion" matrix elements. We start by considering again the convolution integral (2.5.1),

$$M(x) = \int_{-\infty}^{\infty} R(x-y) \, T(y) \, dy \, . \qquad \text{A4.13}$$

However, this time we try to find a "partial inverse" function $R_{pin}(x)$ such that the unfolded spectrum retains a resolution function $R_n(x)$, which is narrower than the original response $R(x)$. Therefore, instead of (2.5.2), we must write

$$\int_{-\infty}^{\infty} R_n(x-y) \, T(y) dy = \int_{-\infty}^{\infty} R_{pin}(x-y) \, M(y) dy \, . \qquad \text{A4.14}$$

Using the convolution theorem we can express these two equations in the ω-domain,

$$m(\omega) = r(\omega) \, t(\omega) \qquad \text{A4.15}$$

and

$$r_n(\omega)\, t(\omega) = r_{pin}(\omega)\, m(\omega) . \qquad \text{A4.16}$$

These two relations are satisfied when

$$r_{pin}(\omega) = \frac{r_n(\omega)}{r(\omega)} . \qquad \text{A4.17}$$

When R(x) was digitized by the factor D(x) we had the result

$$r^d(\omega) = \sum_{n=-\infty}^{\infty} r(\omega - 2\pi n) , \qquad \text{A4.18}$$

as given by (2.5.13). In a similar way we digitize now the function $R_n(x)$, and obtain

$$r_n^d(\omega) = \sum_{n=-\infty}^{\infty} r_n(\omega - 2\pi n) . \qquad \text{A4.19}$$

After these digitizations we replace expression (A4.17) by

$$r_{pin}(\omega) = \frac{r_n^d(\omega)}{r^d(\omega)} . \qquad \text{A4.20}$$

Following the same procedure as in section 2.5, we define a truncated function

$$\begin{aligned} f_n(\omega) &= r_{pin}(\omega) && \text{for } |\omega| < \pi \\ &= \tfrac{1}{2}\, r_{pin}(\omega) && \text{for } |\omega| = \pi \\ &= 0 , && \text{for } |\omega| > \pi \end{aligned} \qquad \text{A4.21}$$

which can be used to obtain the result

$$R_{pin}(x) = F_n(x)\ D(x) , \qquad \text{A4.22}$$

where

$$F_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\omega x}}{r^d(\omega)}\, r_n^d(\omega)\, d\omega . \qquad \text{A4.23}$$

If we now substitute (A4.22) in expression (A4.14) and use the digitized expression for $R_n(x)$, we have

$$\sum_{m=-\infty}^{\infty} R_n(m)\, T(x-m) = \sum_{m=-\infty}^{\infty} F_n(m)\, M(x-m) . \qquad \text{A4.24}$$

By using the transformation $x-m = k$ this last expression can be rewritten as

$$\sum_{k=-\infty}^{\infty} R_n(x-k)\, T(k) = \sum_{k=-\infty}^{\infty} F_n(x-k)\, M(k) . \qquad \text{A4.25}$$

This is analogous to a matrix equation

$$\underline{R}_n \; \underline{T} = \underline{F}_n \; \underline{M} , \qquad \text{A4.26}$$

where $\underline{R}_n$ is a matrix containing the reduced width response functions and $\underline{F}_n$ is a "partial inverse" matrix with elements

$$(f_n)_{k\ell} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\omega(k-\ell)}}{r^d(\omega)}\, r_n^d(\omega)\, d\omega . \qquad \text{A4.27}$$

The relation (2.6.3) is thus established.

APPENDIX V

SUPPLEMENTARY PROOFS FOR SECTION 3.3

It is required to prove that the design matrix $\underline{B}_2$, as defined by (3.3.21), has an inverse $\underline{B}_2^{-1}$, provided the response functions $X_i(x)$ and $Y_j(y)$ form two sets of linearly independent vectors. The proof will be complete if we can show that the vectors formed by the columns of matrix $\underline{Z}$ are linearly independent. We start by assuming that the columns are linearly dependent and then proceed to show that this leads to an impossible condition. The linear dependence in $\underline{Z}$ implies that we can find a non-zero vector

$$\underline{V} = \left\| \left\{ \left\| \left\{ v_{ij} \right\}_{j,1} \right\|^{n_y,1} \right\}_{i,1} \right\|^{n_x,1}, \qquad \text{A5.1}$$

such that

$$\underline{Z}\,\underline{V} = \left\| \left\{ 0 \right\}_{i,1} \right\|^{N_y N_x,1} . \qquad \text{A5.2}$$

It is possible to write the null-vector $\underline{Z}\,\underline{V}$ in the more detailed form

$$\left\| \left\{ \sum_{i=1}^{n_x} \left\| \left\{ X_i(x)\, Y_j(y) \right\}_{y,j} \right\|^{N_y,n_y} \left\| \left\{ v_{ij} \right\}_{j,1} \right\|^{n_y,1} \right\}_{x,1} \right\|^{N_x,1}$$

$$= \underline{0}^{N_x N_y,1} \qquad \text{A5.3}$$

If we look only at the subvectors indexed by x we can say that

$$\left\| \left\{ Y_j(y) \right\}_{y,j} \right\|^{N_y,n_y} \left\| \left\{ \sum_{i=1}^{n_x} X_i(x)\, v_{ij} \right\}_{j,1} \right\|^{n_y,1} = \underline{0}^{N_y,1} \qquad \text{A5.4}$$

for every $x = 1, 2, \ldots, N_x$. The second factor in equation (A5.4) cannot be a null vector for every x since $\underline{V}$ is not a null vector and the $X_i(x)$ are linearly independent. But then equation (A5.4) implies that the $Y_j(y)$ are linearly dependent, which is not true. Hence we must reject the assumption that the columns of $\underline{Z}$ are linearly dependent. According to Cramér (49) the linear independence in $\underline{Z}$ is sufficient to insure that the inverse matrix $\underline{B}_2^{-1}$ exists.

Next we turn to the derivation of expression (3.3.39) which gives the full covariance matrix of estimate vector $\underline{A}_2''$. We join all data vectors (3.3.25) to form one long vertical vector defined by

$$\underline{M}_o' = \left\| \left\{ \left\| \left\{ M_2'(x,y) \right\}_{x,1} \right\|^{N_x,1} \right\}_{y,1} \right\|^{N_y,1} . \qquad \text{A5.5}$$

Likewise, we string out the vectors of intermediate coefficients by defining

$$\underline{Q}' = \left\| \left\{ \left\| \left\{ q_i'(y) \right\}_{i,1} \right\|^{n_x,1} \right\}_{y,1} \right\|^{N_y,1} . \qquad \text{A5.6}$$

Making use of these definitions we propose to find an $n_x n_y$ by $N_x N_y$ size matrix $\underline{P}$ such that

$$\underline{A}_2'' = \underline{P}\ \underline{M}_o' . \qquad \text{A5.7}$$

According to a theorem[50] mentioned in the text the required covariance matrix is then expressible in the form

$$\underline{C}_{A_2''} = \underline{P}\ \underline{C}_{M_o'}\ \underline{P}^T . \qquad \text{A5.8}$$

Equation (3.3.29) gives N_y least squares solutions in the x-dimension. These solutions can be grouped together and expressed in one matrix equation

$$\underline{Q}' = \underline{H}_o \, \underline{M}_o' , \qquad \text{A5.9}$$

where

$$\underline{H}_o = \left\| \left\{ \left\| \underline{F}_y^{-1} \, \underline{X}^T \, \underline{W}_y \right\|^{n_x, N_x} \delta_{yi} \right\}_{y,i} \right\|^{N_y, N_y} . \qquad \text{A5.10}$$

Before we can group solutions (3.3.35) for the y-dimension, we must rearrange the intermediate vector $\underline{Q}'$ so that it consists of a string of vectors $\underline{Q}_i'$. This is accomplished by a rearranging matrix $\underline{R}_o$ such that

$$\underline{Q}_o' = \left\| \left\{ \left\| \left\{ q_i'(y) \right\}_{y,1} \right\|^{N_y, 1} \right\}_{i,1} \right\|^{n_x, 1}$$

$$= \underline{R}_o \, \underline{Q}' . \qquad \text{A5.11}$$

The desired rearranging matrix can be defined by

$$\underline{R}_o = \left\| \left\{ \left\| \left\{ \delta_{ij} \, \delta_{ky} \right\}_{k,i} \right\|^{N_y, n_x} \right\}_{j,y} \right\|^{n_x, N_y} , \qquad \text{A5.12}$$

as substitution in (A5.11) will readily show.

After introducing one more definition,

$$\underline{R}_1 = \left\| \left\{ \left\| \underline{K}_i^{-1} \, \underline{Y}^T \, \underline{W}_i \right\|^{n_y, N_y} \delta_{ij} \right\}_{i,j} \right\|^{n_x, n_x} , \qquad \text{A5.13}$$

we can write down the final solution vector

$$\underline{A}_2'' = \underline{R}_1 \, \underline{R}_o \, \underline{H}_o \, \underline{M}_o' . \qquad \text{A5.14}$$

Thus the covariance matrix of estimate $\underline{A}_2''$ is given by

$$\underline{C}_{A_2''} = \underline{R}_1 \underline{R}_o \underline{H}_o \underline{C}_{M_o'} \underline{H}_o^T \underline{R}_o^T \underline{R}_1^T , \qquad \text{A5.15}$$

which reduces to expression (3.3.39) after a certain amount of matrix algebra.

It remains now to prove the inequalities in expression (3.3.45). It is shown by Cramér[49] that the reciprocal of a diagonal element in the positive definite design matrix is not greater than the value of corresponding diagonal element in the inverse matrix. This fact immediately establishes the inequality

$$D(a'_{k\ell}) \geq \frac{1}{\left\{ \underline{B}_2 \right\}_{i,i}} , \qquad \text{A5.16}$$

where $i = (k-1)n_y + \ell$.

Since the diagonal submatrix $\underline{B}_{kk}$ is also positive definite, we have the second inequality

$$\left\{ \underline{B}_{kk}^{-1} \right\}_{\ell,\ell} \geq \frac{1}{\left\{ \underline{B}_2 \right\}_{i,i}} . \qquad \text{A5.17}$$

Next we consider the inequality between expressions $D(a'_{k\ell})$ and $\left\{ \underline{B}_{kk}^{-1} \right\}_{\ell\ell}$. Matrix $\underline{B}_2$ can be partitioned according to the following scheme:

$$\underline{B}_2 = \left\| \begin{array}{c|c} \underline{A}_{11} & \underline{A}_{12} \\ (n_y \text{ by } n_y) & \left[n_y \text{ by } (n_x-1)n_y \right] \\ \hline \underline{A}_{21} & \underline{A}_{22} \\ \left[(n_x-1)n_y \text{ by } n_y \right] & \left[(n_x-1)n_y \text{ by } (n_x-1)n_y \right] \end{array} \right\| . \qquad \text{A5.18}$$

In a similar way we partition the inverse matrix $\underline{B}_2^{-1}$ and label the submatrices by $\underline{C}_{ij}$. Upon multiplication of these partitioned matrices we can write

$$\underline{A}_{11}\ \underline{C}_{11} + \underline{A}_{12}\ \underline{C}_{21} = \underline{I}^{n_y}\ , \qquad \text{A5.19}$$

$$\underline{C}_{21}\ \underline{A}_{11} + \underline{C}_{22}\ \underline{A}_{21} = \underline{O}\ , \qquad \text{A5.20}$$

$$\underline{C}_{21}\ \underline{A}_{12} + \underline{C}_{22}\ \underline{A}_{22} = \underline{I}^{(n_x-1)n_y}\ , \qquad \text{A5.21}$$

where $\underline{I}$ is a diagonal identity matrix (with rank indicated by the superscript). From (A5.19) we have

$$\underline{C}_{11} = \underline{A}_{11}^{-1} - \underline{A}_{11}^{-1}\ \underline{A}_{12}\ \underline{C}_{21}\ , \qquad \text{A5.22}$$

and from (A5.20) we obtain

$$\underline{C}_{21} = -\ \underline{C}_{22}\ \underline{A}_{21}\ \underline{A}_{11}^{-1}\ . \qquad \text{A5.23}$$

These last two expressions can be combined to yield the result

$$\underline{C}_{11} = \underline{A}_{11}^{-1} + \underline{A}_{11}^{-1}\ \underline{A}_{12}\ \underline{C}_{22}\ \underline{A}_{21}\ \underline{A}_{11}^{-1}\ . \qquad \text{A5.24}$$

We now let the positive definite matrix $\underline{C}_{22}$ have elements c_{ij} and define another set of elements t_{ij} such that

$$\underline{A}_{11}^{-1}\ \underline{A}_{12} = \left\| \left\{ t_{ij} \right\}_{i,j} \right\|^{n_y,(n_x-1)n_y}\ . \qquad \text{A5.25}$$

The second term in (A5.24) can now be written in the form

$$\left\| \left\{ t_{ij} \right\}_{i,j} \right\|^{n_y,(n_x-1)n_y} \left\| \left\{ c_{ij} \right\}_{i,j} \right\|^{(n_x-1)n_y,(n_x-1)n_y} \left\| \left\{ t_{ij} \right\}_{j,l} \right\|^{(n_x-1)n_y,n_y}$$

$$= \left\| \left\{ \sum_{k,j=1}^{(n_x-1)n_y} t_{\ell k}\ c_{kj}\ t_{ij} \right\}_{\ell,i} \right\|^{n_y,n_y}\ . \qquad \text{A5.26}$$

The ℓ^{th} diagonal element of this matrix is given by

$$\sum_{k,j=1}^{(n_x-1)n_y} t_{\ell k}\ c_{kj}\ t_{\ell j}\ ,$$

which is a positive definite quadratic form by virtue of matrix $\underline{C}_{22}$ being

positive definite. Therefore, going back to equation (A5.24), we see that the diagonal elements of $\underline{C}_{11}$ are greater than or equal to the corresponding diagonal elements in $\underline{A}_{11}^{-1}$. This establishes the inequality

$$D(a'_{k\ell}) \geq \left\{ \underline{B}_{kk}^{-1} \right\}_{\ell,\ell}$$

for the index value k = 1. Similar arguments can be used to prove the inequality for other values of k.

REFERENCES

(1) W. V. Prestwich, Thesis, McMaster University, Hamilton, Ont. (1963) p. 31a.

(2) R. L. Heath, ID016784/TID 450D.

(3) N. R. Johnson, E.Eichler, G.D. O'Kelley, J. W. Chase and J. T. Wasson, Phys. Rev. 122 (1961) p.1546.

(4) R. E.Segel, Bull. Am. Phys. Soc., 7 (1962) p.542.

(5) T. J. Kennett, W. V. Prestwich and G. L. Keech, Nucl. Instr. and Meth., 29 (1964) p.325.

(6) M. S. Freedman, T.B. Novey, F. T. Porter and F. Wagner Jr., Rev. Sci. Instr., 27 (1956) p.716.

(7) N. E.Scofield, U.S. At. En. Comm. NAS-NS 3107, p.108.

(8) W. R. Burrus and D. Bogert, Ibid, p 127.

(9) V. D. Bogert and W. R. Burrus, Neutron Phys. Div. Ann. Pr. Rep. Sept. 1, 1962, ORNL-3360, p. 22

(10) D. L. Summers and D. D. Babb, U.S. At. En. Comm. NAS-NS 3107, p. 143.

(11) W. R. Dixon and J. H. Aitken, Can. J. Phys. 36 (1958) p.1624.

(12) J. G. Herriot, Methods of Mathematical Analysis and Computation, John Wiley & Sons (1963).

(13) J. S. Berezin and N. P. Zhidkov, Computing Methods, vol. 2 Pergamon Press (1965).

(14) E. Bodewig, Matrix Calculus, North-Holland Publ. Co. (1959).

(15) R. S. Varga, Matrix Iterative Analysis, Prentice-Hall (1962).

(16) V. N. Faddeeva, Computational Methods of Linear Algebra, Dover Publications (1959).

(17) A. M. Turing, Quart. J. Mech. Appl. Math., 1 (1948) p.287.

(18) J. von Neumann and H. H. Goldstine, Bull. Amer. Math. Soc., 53 (1947) p. 1021.

(19) Ibid, Proc. Amer. Math. Soc., 2 (1951) p. 188.

(20) T. J. Kennett and G. L. Keech, Nucl. Instr. and Meth., 24 (1963) p.142.

(21) D. D. Slavinskas, T. J. Kennett and W. V. Prestwich, Nucl. Instr. and Meth. 37 (1965) p.36.

(22) W. Bothe, Z. Naturforsch., 4a (1949) p.542.

(23) Nuclear Data Sheets (Natl. Acad. Sci. Natl. Res. Counc., Publ. Washington, D.C.), Mass Number 143.

(24) M. A. Preston, Physics of the Nucleus, Addison-Wesley Publ. Co. (1963) p.421.

(25) Ibid, p. 431.

(26) Tables for the Analysis of Beta Spectra, (U.S. Dept. of Commerce, Natl. Bureau of Standards).

(27) T. Kotani and M. Ross, Phys. Rev., 113 (1959) p.622.

(28) L. M. Langer and M. Ross, Phys. Rev., 113 (1959) p.622.

(29) Nuclear Data Sheets (Natl. Acad. Sci. Natl. Res. Counc., Publ. Washington, D.C.) Mass Numbers 90, 91 and 143.

(30) M. G.Strauss and R. Brenner, Rev.Sc. Instr. 36 (1965) p.1857 .

(31) J. R. Klauder, A.C. Price, S. Darlington and W. J. Albersheim, Bell Syst. Tech. J., 39 (1960) p.745.

(32) F. A. Van Melle, D. L. Faass, S. Kaufman, G. W. Postma and A. J. Seriff, Geophysics, 28 (1963) p.466.

(33) I. N. Sneddon, Fourier Transforms, McGraw-Hill (1951).

(34) R. L. Graves and P. Wolfe, Recent Advances in Mathematical Programming, McGraw-Hill (1963).

(35) Albert Messiah, Quantum Mechanics, vol. 1, p.470, John Wiley and Sons (1961).

(36) M. H. Lietzke, U.S. At. En. Comm., NAS-NS 3107, p.1.

(37) Walter E. Nervik, Ibid, p.9.

(38) J. B. Cumming, Ibid, p.25.

(39) S. B. Burson, R.G. Helmer and T. Gedayloo, Ibid, p. 51.

(40) Berol L. Robinson, Ibid, p. 67.

(41) A. J. Ferguson, Ibid, p. 157.

(42) L. Salmon, Ibid, p.165.

(43) J. I. Trombka, Ibid, p. 183.

(44) R. O. Chester, R. W. Peelle and F. C. Maienschein, Ibid, p.201.

(45) W. L. Nicholson, J. E. Schlosser and F. P. Brauer, Ibid, p.254.

(46) W. B.Seefeldt, Ibid, p. 287.

(47) N. P. Archer, Thesis, McMaster University, Hamilton, Ont. (1965).

(48) R. L. Heath, R. G. Helmer, L. A. Schmittroth and G. A. Cazier, IDO-17017 (1965).

(49) H. Cramér, Mathematical Methods of Statistics, Princeton University Press (1946).

(50) C. R. Rao, Advanced Statistical Methods in Biometric Research, Wiley (1952) p. 53.

(51) Y. V. Linnik, Method of Least Squares and Principles of the Theory of Observations, Pergamon Press (1961) p.84.

(52) Ibid, p.85.

(53) N. R. Johnson and G. D. O'Kelley, Phys. Rev. 108 (1957) p.85.

(54) M. Abramovitz and I. A. Stegun, Handbook of Mathematical Functions, Natl. Bur. of Stds. (1965) p.257.

(55) Ibid, p.68