FERD AND FERDOR TYPE UNFOLDING CODES*

Walter R. Burrus**
Neutron Physics Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37830

May 1976

*Research performed at Oak Ridge National Laboratory operated by Union carbide Corporation for the Energy Research and Development Administration

**Consultant, Mini-Max Services, Inc., 312 W. Broadway Lenoir City, Tennessee 37771

BLANK PAGE

FERD AND FERDOR TYPE UNFOLDING CODES

Walter R. Burrus*
Neutron Physics Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37830

ABSTRACT

FERD and FERDO are unfolding codes which were developed at the Neutron Physics Division of Oak Ridge National Laboratory in 1965 and 1966. FERDO variants such as FERDOR and FORIST have been widely used, and many useful supplementary procedures have been developed for neutron and gamma-ray spectroscopy and other diverse applications. FERD has not been well documented but has been used for various ORNL problems.

The pulse-height distribution b_i (i=1,2,...m) is related to an unknown spectrum by the equation

$$b_i = \int A_i(E)\, x(E)dE \qquad i=1,2,\ldots,m \qquad (1)$$

where the $A_i(E)$ functions are the instrument response functions. Both types of codes seek confidence intervals (p_k^{lo}, p_k^{up}) for parameters p_k which are related to the unknown spectrum by the equation

$$p_k = \int W_k(E)\, x(E)dE \qquad k=1,2,\ldots,t \quad . \qquad (2)$$

where the $W_k(E)$ functions are chosen by the experimenter and are referred to as "window functions". In classical linear regression, confidence intervals may only be found for "estimable" functions which must be exact linear combinations of the instrument response functions of the form

$$L_k(E) = \sum_{i=1}^{m} U_{ki} A_i(E) \quad . \qquad (3)$$

*Consultant, Mini-Max Services, Inc., 312 W. Broadway, Lenoir City, Tennessee 37771

Generally, it is not possible to find "estimable" $L_k(E)$ functions which agree with the desired window functions. Both FERD and FERDO codes get around this problem by bracketing the desired window function between two "estimable" functions such that

$$L_k^{lo}(E) \leq W_k(E) \leq L_k^{up}(E) \quad . \tag{4}$$

The fact that $x(E)>0$ (on physical *a priori* grounds) allows valid confidence intergals to be found for the values of p_k, even though they are not formally "estimable".

Generally, if a tight match is sought by minimizing only the difference between $L_k(E)$ and $W_k(E)$, large coefficients U_{ki} result such that some terms in Eq. (3) have large positive values and others have large negative values. Then the statistical uncertainty due to the experimental estimates $\hat{b}_i$ of b_i dominates the confidence interval widths. If a loose match (with appreciable "slop") between the $L_k(E)$ and $W_k(E)$ functions is sought by modifying the minimization criterion, then smaller coefficients U_{ki} can be found which reduce the statistical contribution to the confidence interval width at the expense of a greater width due to the "slop" uncertainty between $L_k(E)$ and $W_k(E)$.

The key to the FERD and FERDO techniques is to approximately minimize the total width of the confidence interval due to both sources of uncertainty, balancing one type of uncertainty against the other. FERDOR type codes use a least squares minimization technique with a set of weight factors (which in mathematical statistics terminology define a "metric") which approximately balances the two types of errors. Part of the metric is determined by observational errors in the $\hat{b}_i$, and the other part of the metric is determined by means of some indirect inequality techniques which depend upon $x(E)\geq 0$. Irrespective of the approximations or assumptions made in deriving the minimization criterion, it is still possible to determine conservative estimates

of the two uncertainty contributions once the U_{ki} coefficients have been determined numerically. FERDO type codes usually do not explicitly obtain the $L_k(E)$ functions (although the U_{ki} coefficients are available), and usually very conservatively overestimate the "slop" type errors to save computation.

FERDO type codes do calculate $L_k^{lo}(E)$ and $L_k^{up}(E)$ explicitly, and use a more complex inequality technique (motivated by some methods based upon Linear Programming) to determine the "slop" uncertainty.

In the application of any unfolding code, many problems must be worked through by the experimenters starting with the design of the instrument, the philosophy of "what is a solution", and going through the communication of the "interpreted" results. A wide variety of divergent viewpoints hampers understanding of common problems. Some "Axioms of Data Analysis" are discussed which guided the philosophy behind the development of the ORNL codes.

INTRODUCTION

The origins of radiation spectrum unfolding go back to the 1950's when radiation spectra began to be measured. One set of origins traces back to the use of fission and threshold foils for reactor spectral measurements. Some of the early work was done in about 1952, and since then many papers have appeared.[1-3]

The other set of origins stems back to the first practicable many channel spectrometer -- the NaI detectors developed in the early 1950's. The work done by Russell Heath[4] in radioisotope assay is somewhat out of the scope of of this conference, although the techniques are certainly pertinent. Corrections for non-perfect line shapes in various kinds of spectrometers has been studied for years, with an early reference being to Carl Eckert in 1937.[5] Resolution improvement was treated by Sir Arthur

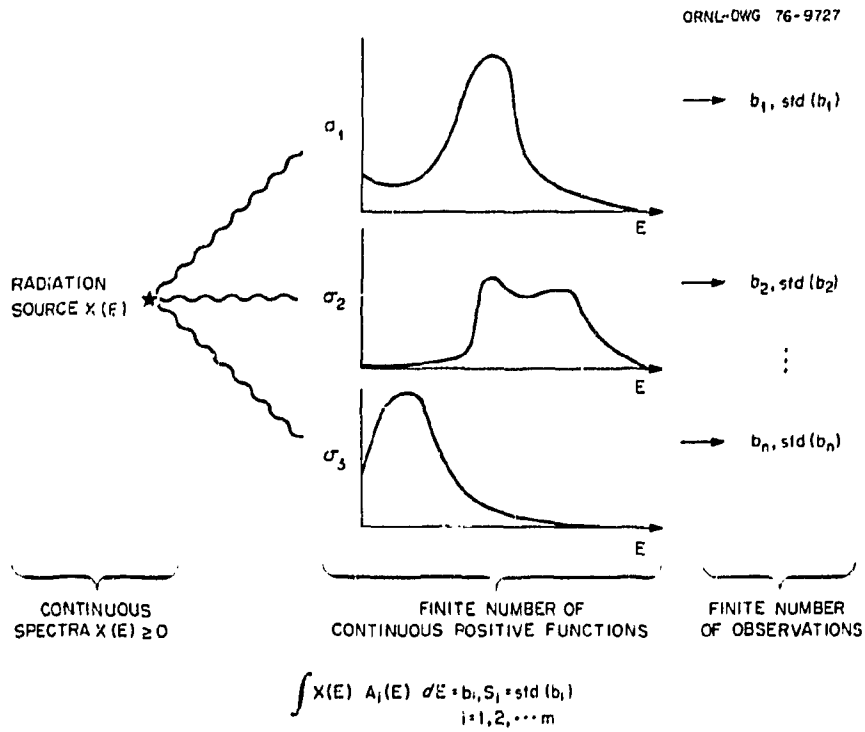
Eddington in connection with an astronomical problem in 1913.[6] The concept of window function was used by A. Sommerfeld in 1891.[7] Linden and Starfelt had developed a technique for unfolding continuous gamma-ray and X-ray spectra in 1954.[8] Much pertinent work has been done in the astronomy and geophysics areas, as well as the communications field, where the key work was done by Norbert Wiener in the early 1940's.

My own work in this area was during the time span 1960 to 1968. The ORNL Neutron Physics Division recognized that adequate methods for unfolding did not exist. E. P. Blizard in particular and later Fred Maienschein strongly encouraged the attempts at developing a more powerful approach at unfolding methodology. I am especially indebted to the experimental physicists who collaborated closely with me in risking their data to the new methods,[9-11] and to my numerical colleagues.[12-13]

THE UNFOLDING PROBLEM

Figure 1 illustrates the forward or *faltung* problem in the case of a few-channel spectrometer, such as a set of threshold foils. The radiation source is denoted by x(E) where x(E) is the number of photons (per unit area or angle) interacting with a finite number of detectors which have response functions σ_1, σ_2, ... σ_m. In the case of threshold foils, the detector response functions are just the reaction cross sections of the foils. In the case of the scintillation spectrometer, the response functions are the probability functions that an incident photon of energy E will produce a reaction (or count) in channel i (where i goes from 1 to m). The experimental results are denoted by b_i (i=1, ..., m), and the error in the result by S_i=std(b_i). Note that x(E) is generally a continuous function, and b_i is measured for a finite number of channels.

Figure 2 illustrates the information flow in folding and unfolding in a more schematic fashion. A continuous function (in Hilbert space) is difficult to visualize; however, we depict two directions in Hilbert space as the first two axes, and represent all the other directions by the axis labeled by an ellipsis (3 dots). The point in X-space denotes a "true" radiation spectrum x(E). The results of the measurement is a vector of observations in B-space, also illustrated as a 3-axis schematic representation.



$$\int X(E)\ A_i(E)\ dE = b_i, S_i = \text{std}\,(b_i) \qquad i = 1, 2, \cdots m$$

Fig. 1. Instrument Response

ORNL-DWG 76-9721R

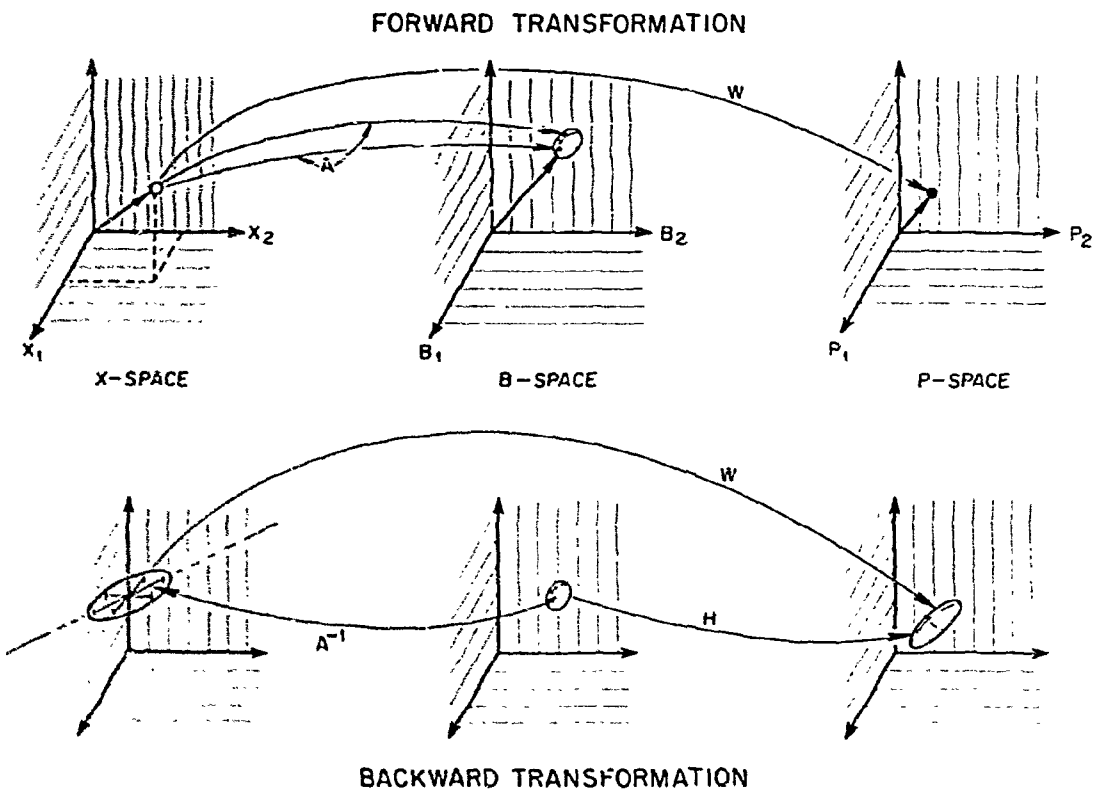


Fig. 2. Information Flow in Hilbert Space

The transformation shown is generally multiple valued because of the counting and systematic errors in making the observation, so that a true spectrum $x(E)$ would give different results for $\hat{b}$ on each measurement, and the family of b's could be described by a region which I will call the "B-ball".

The reverse transformation is shown on the bottom of Figure 2. The relation between b_i (i=1, ... m) and $x(E)$ is generally undetermined since one cannot solve a finite system of equations (one for each b_i) for an effectively infinite number of points of the continuous function $x(E)$. A single point within the B-ball is consistent with points within some subspace of X-space, illustrated by a line. Similarly, each point within the B-ball is consistent with another subspace. The family of all such subspaces in X-space contains all the $x(E)$ functions which are consistent with the observed b_i (i=1,...,m).

Since the reverse transformation from b_i (i=1, ...,m) back to $x(E)$ is multivalued (even if $\hat{b}_i$ could be measured without any error), some techniques must be utilized to obtain an acceptable representation of consistent solutions.

One often used technique is to solve not for $x(E)$, but for some related function (denoted schematically by a p vector in P-space (or solution space). The p_k's may represent a smoothed version of $x(E)$, for example, or they may represent the integral spectrum (integrated from E to ∞). The unfolding procedure, denoted by H, may be considered to operate upon b_i (i=1, ...,m) to yield p_k (k=1, ...,t). It is convenient to think of solving (or unfolding) for values p_k which are functions of the unknown spectrum $x(E)$ related by

$$p_k = \int W_k(E)\, dE$$

where the $W_k(E)$ functions are referred to as window functions. The relation between X-Space and P-Space is denoted as W.

Figure 3 illustrates the "windows" from a different perspective. The response functions $A_i(E)$ connect the unknown spectrum $x(E)$ to the observations

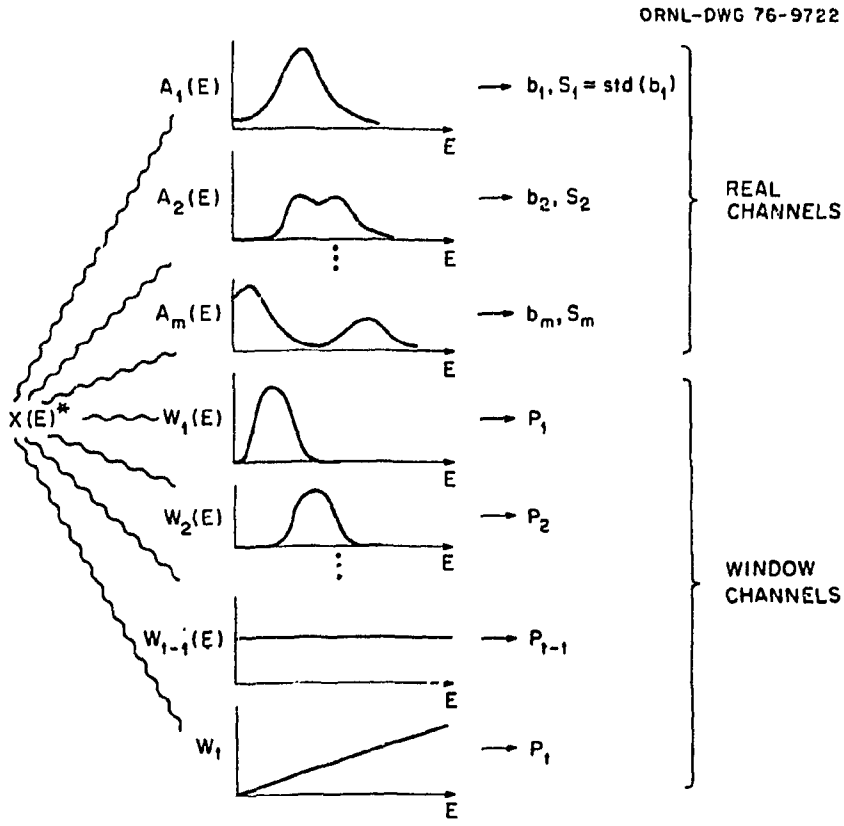


Fig. 3. Window Functions

$\hat{b}_i$, while the window functions $W_k(E)$ connect the unknown spectrum to the "solution" p_k. Since the window functions are fictional, we are not restricted to think of one window function per real response function.

More generally, one may think of a window function as any linear function of the spectrum x(E) that one wishes to solve for. Simple functions are the flux function, the "energy" function, and a damage function as illustrated.

UNFOLDING

The simplest unfolding methods are the "triangular" methods in which peaks in the response functions are replaced by discrete elements, and smooth edges are replaced by sharp edges, as in Fig. 4. This is equivalent to replacing the Response Kernel (A in matrix notation) by TW where T is an upper triangular matrix and W is a window matrix. The solution of the triangular system of equations T x = b may be carried out without difficulty; however, Wx is obtained for a solution and not x itself. This is the basis for the SEGO and High Speed Sego methods developed at ORNL and for the "differentiation" type methods for neutron spectra.

The triangular methods operate only on the observed data b_i (i=1, ...m). This works out OK in many-channel problems where something like the instrumental FWHM is acceptable in the solutions; however, they ignore certain other classes of information which are available.

Information

Information about the spectrum can be categorized into three classes: *A posteriori* information, *a priori* information, and intuitive information. *A posteriori* means "after the fact" and denotes the information which is measured--i.e., the observation vector $\hat{b}_i$ plus any additional measurements which might have been made at the time (such as a monitor chamber or foil). *A priori* information denotes information which is known about x(E) before the measurement was made. What can we really know about the spectrum before it is measured? If we know the source of the spectrum, we may have a good bit of theoretical knowledge of the spectrum. For example, we know that a neutron capture spectrum is a combination of continuous background

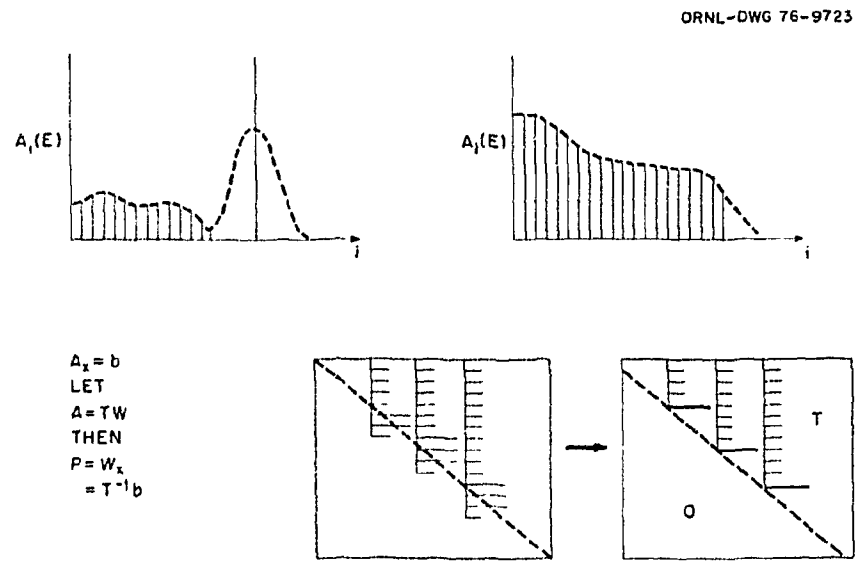


Fig. 4. Triangular Methods

and discrete components. Further we may know that there are a finite number of discrete components at energies permitted by the relevant nuclear energy levels. As another illustration of *a priori* knowledge, we may know that the prompt gamma rays from fission are from moving fission fragments and are therefore Doppler broadened. In any event, it is generally known *a priori* that $x(E) \geq 0$, on physical grounds.

There is another class of information which does not fall into the *a posteriori* or *a priori* class. This kind of information can be called "intuitive". Very likely the spectrum which is to be unfolded is not the first spectrum which has ever been considered by the user. Frequently the same types of spectrum have been measured by other experimenters, some using quite different types of apparatus. Thus, the experimenter approaches the problem with preconceived notions based on previous work and physical insights into what the admissible class of solutions should be.

The distinction between *a priori* information and intuitive information is this: *A priori* information can be described in mathematical form such as:

> Number of peaks is less than N ,
> $x(E) > 0$, or
> continuum limited to energy E max.

Intuitive information is much more subtle than this. Intuitive knowledge cannot be mathematically quantified by the experimenter. That is not to say that the intuitive knowledge might be quantifiable, but only that the experimenter does not know how to do it. For example, before the development of linear programming ideas, it was not clear how the *a priori* constraint of non-negativity might be incorporated as a constraint on admissible solutions. However, experimental physicists devised many *ad hoc* procedures for forcing this into the analysis of their data. Some forms of information are on the borderline between intuitive and *a priori*. But it is clear that there are many feelings that the experimenter has which are more properly intuitive than *a priori*.

An example of intuitive knowledge is the experimenter's preconceived idea about the complexity of the admissible spectra. Many spectra are possible which are consistent with the *a posteriori* and *a priori* information but the experimenter may reject some of these as pathological (i.e., not apt to occur). What is needed is a way to mathematically quantify this feeling. The Maximum Entropy methods (such as MAZE) incorporate this information by including *ad hoc* terms in the optimization procedure which suppress large order derivatives.[14]

An interesting way to incorporate this intuitive information is to limit the "variation" of the spectrum x(E). The variation is the total "up and down" variation of the spectrum as shown in Figure 5. Linear programming statements of unfolding permit the experimenter to set a maximum limit to the variation as a constraint on the admissible spectra in X-space. Even though the experimenter does not know *a priori* what this limit is, he may try several unfoldings with different values for the variation limit and then choose subjectively between them. If he chooses a limit which is equal or greater than the true value, his results are valid. If he chooses a limit which is smaller than the true value, then his results are invalid (to the extent that they depend upon this restriction).

This is obviously "risky business", but in few-channel problems such risk must be taken to get useful results. The real test of an unfolding procedure is not that it has no risk, but that it produces useful results in the hands of an expert practitioner. In fact, it has been recognized for years that published errors in measurements made by independent teams tend to be inconsistent by several times the stated amount. In my view, this is a reflection of the experimenters' bias (or optimism) in treating an overall analysis of systematic errors, or stated another way, it is a reflection of the fact that experimenters use physical judgement (i.e., intuition) in publishing results which cannot be strictly justified.

Figure 6 summarizes the situation with respect to the three kinds of information. The Venn diagram shows the set of all possible x(E) and

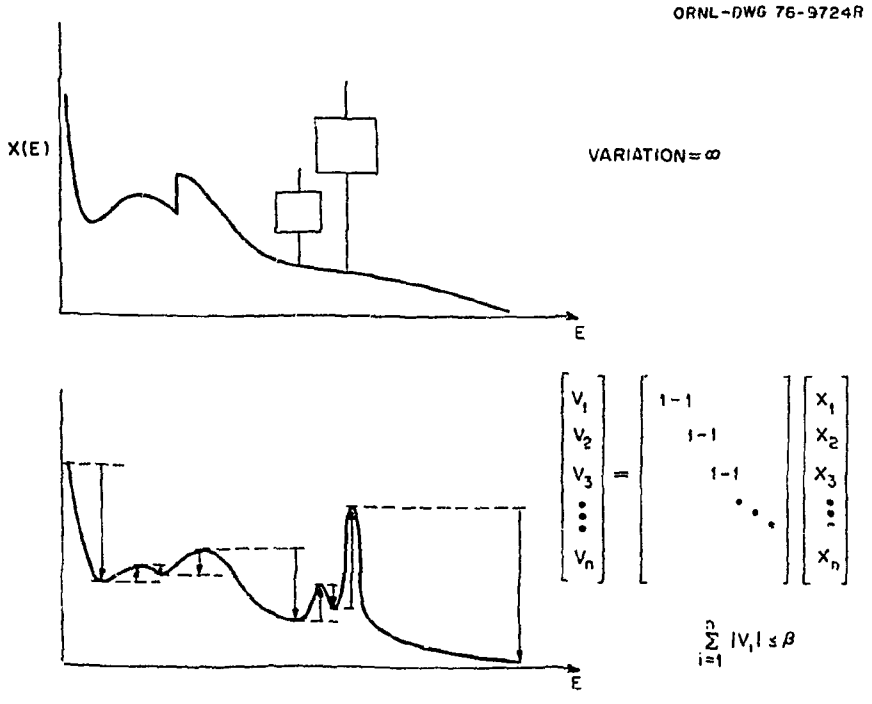


Fig. 5. Variation of the Spectrum

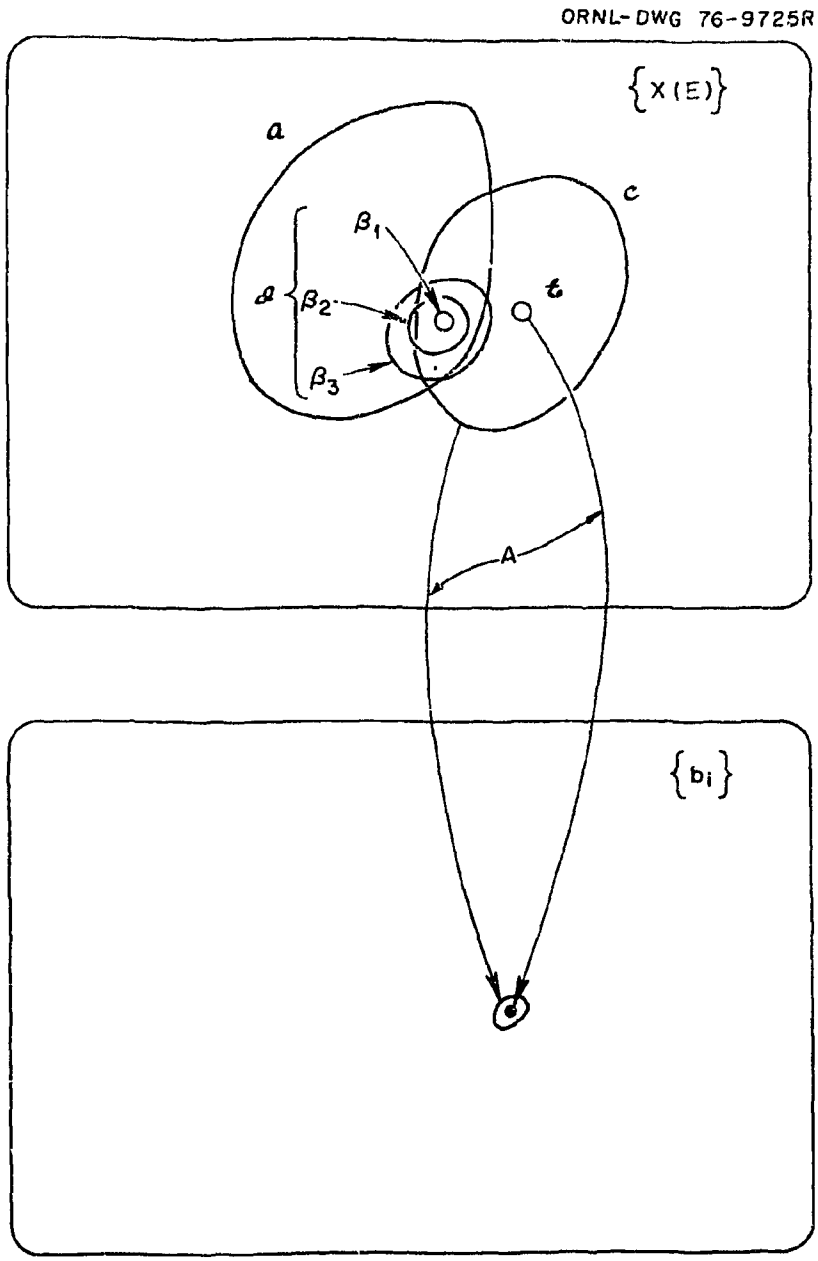


Fig. 6. Venn Diagram of All Sets

the set of all possible b_i. The E subset is the set of spectra which produce exactly the observation vector $\hat{b}_i$. The C subset is the set of spectra which are statistically consistent (i.e., close) to the observation vector. The A set is the *a priori* set (say non-negativity). Note that in this illustration the *a priori* A set has no point in common with the C set. Thus, the spectrum which produces the observations exactly is inadmissible, while one which is only close to the observation is admissible. Finally, the family of subsets denoted by I illustrate the effect of an intuitive constraint as the value of a parameter β is varied. The validity of the subsets A and C are firm, while the validity of the subset I is subjective (reflected in one or more subjective parameters).

RULES OF THE UNFOLDING GAME

In game theory, the first step in examining a specific game is to codify the rules. In fact, the rule constitutes the game. Once the game has been defined, then the second step is to consider methods for solution or playing of the game. In other words, it is difficult to play the game without knowing the rules. The big problem in unfolding which has caused much controversy over the years is that while solutions are many, rules are few. Incidentally, the unfolding problem can be cast in several different game theoretical frameworks, as a one person (solitaire type) game or as a two person game (between "mother nature" and the "experimenter".) The intuitive aspects of unfolding make it particularly difficult to obtain agreement on the rules. According to the established wisdom, within the "scientific method", intuitive interpretations of data are bad and objective results are good which can be verified by objective computation techniques.

In any event, we can categorize unfolding procedures into the three classes previously discussed. The first class utilizes only *a posteriori* data. Some of these methods (i.e., the Triangular methods) when properly implemented should be non-controversial. Proper implementation would consist in having an adequate error analysis and a proper recognition of the "resolution" of the solution.

The second class utilizes *a posteriori* and also *a priori* data. The use of *a priori* data such as non-negativity should be non-controversial; however, the extensions to provide non-equality constraints to the classical regression analysis have been slow to be accepted. Just recently has the numerical mathematics community begun to seriously consider ill-conditioned systems. Bert Rust in the next paper will comment about some of the recent contemporary mathematical interest evidenced in some special meetings on this topic.[15]

The third class utilizes *a posteriori*, *a priori*, and intuitive information and will probably always be on the borderline of respectability. Nevertheless, one can make a good argument that many respectable processes such as curve fitting are really intuitive since the analyst may choose subjectively between many different fitting techniques (such as least square, maximum deviation, etc.).

THE FERD AND FERDOR METHODS

The FERD method has been described in several places. An RSIC "package" is under preparation which compares the two methods and gives the key numerical algorithms. Both of these methods have essentially the same input and the same output. In fact, another ORNL method, OPTIMO, based on linear programming, also has essentially the same input and output. However, the OPTIMO[16] code was never developed to the point where it could be successfully transferred to other groups. Perhaps one day OPTIMO will be rediscovered, or an improved version of FERD will be developed which will obsolete it.

Figure 1 illustrates the input and output of the algorithm. The response data and the window data is prepared as matrices A and W. The observation vector B and errors S are entered as vectors. A and W are thought of as being finite in the pulse height direction and continuous in the x(E) direction. The output consist of bounds (p_k^{lo}, p_k^{up}) for the response of the window functions. No assumptions are made about the structure of the x(E) function. The only assumptions are that the $A_i(E)$ and $W_k(E)$ are adequately represented at a sufficient number NC of "comparison points".

Figure 7 also illustrates a typical $A_i(E)$ function and a typical $W_k(E)$ function. These are tabulated at the comparison points shown. The codes form upper and lower bounds to the $W_k(E)$ function by linear combinations of the $A_i(E)$ functions. If the comparison points are too far apart, then inequality at these points may not be sufficient to insure inequality at all intermediate points. The exact spacing of the points is arbitrary, and it is not necessary or desirable to interpret the matrix values of averages over an interval dE (although some practitioners insist on such an interpretation).

Intuitive Information

The FERD, FERDO, and OPTIMO methods are of the second type in that they utilize only *a posteriori* and firm *a priori* information (e.g., $x(E) \geq 0$). However, they may be generalized easily by a pretreatment of the input data to impose additional constraints upon the class of admissible solutions. Instead of requiring only that $x(E) \geq 0$, it is possible to require that $x = Rq$, where $q \geq 0$ instead. References 13 and 17 outline many different constraints which can be devised by suitable choice of the regularizing function R. In particular, if R represents a Gaussian smoothing function, the modified constraint condition requires that the admissible spectra be at least as smooth as the Gaussian functions. Figure 8 illustrates the relation between q, x, b, and p.

UNFOLDING LAWS

The following set of LAWS is suggested as a set of desirable properties for an unfolding method. There are probably no current methods which live up to these LAWS. While there is a much better understanding today of the problems of unfolding (as typified by the papers at this meeting), there is still much that remains to be done. In particular, the attempt to define some meaningful benchmarks for the many-channel and the few-channel problems will force some close scrutiny of the desirable properties. The tentative set of Unfolding Laws is:

1st LAW The "solution" should be representative of the entire set (or family) of admissible solutions consistent with the constraints incorporated (*a posteriori*, *a priori*, and/or intuitive). A single member of the consistent solutions is unacceptable.

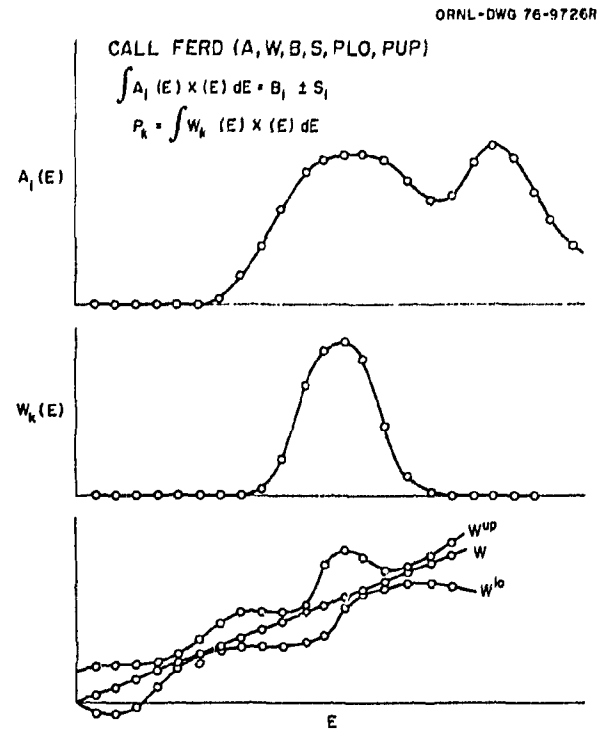


Fig. 7. Input-Output of FERD Algorithm

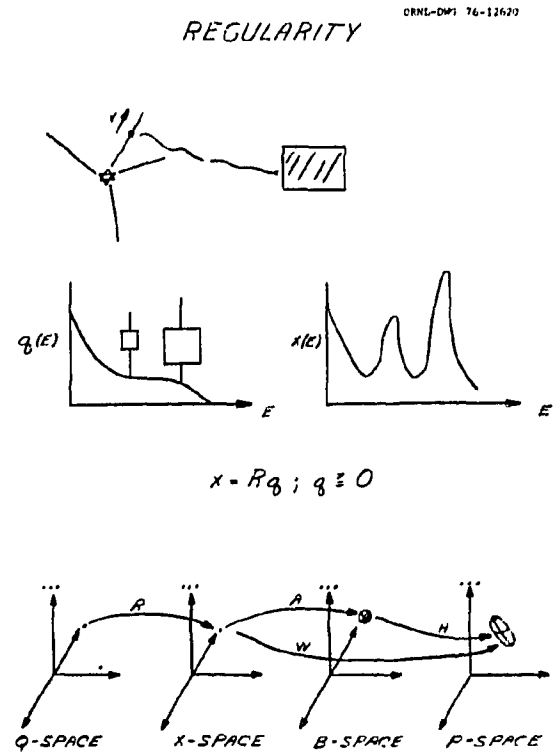


Fig. 8. Regularity Conditions

2nd LAW The relationship between the characterization of the "solution" and the spectrum x(E) should be available in comprehensible mathematical form. As a minimum, the "resolution" of the solution characterization should be known as a function of energy.

3rd LAW The characterization of the solution should have a confidence level interpretation with a specified percentage confidence level, such that a specified percentage of independent experiments would agree at least with the stated fraction of trials. In other words, the confidence interpretation should be conservative and not too optimistic.

Two additional "minor" laws are:

4th LAW In the unfolding procedure for spectral representations, it should be possible to control the inherent trade-off between "resolution" and "errors".

5th LAW If intuitive information is included in the constraints on the solution (which by definition cannot be verified firmly), then it should be possible to relax the intuitive constraints in degrees till the solution is based only upon *a posteriori* and *a priori* constraints. (This at least allows an informed subjective judgement to be made on the strength of one or more subjective constants).

CONCLUSION

The toughest problems are probably the few-channel problem and the super-resolution problem. Intuitive constraints are probably necessary in both cases. However, much work needs to be done to make these constraints as plausible as possible and to simplify the exercise of the required subjective judgement as much as possible. Alternately, improved formulations of *a priori* information can be sought which accomplish the same objectives as intuitive information but in a strictly defensible manner.

REFERENCES

1. J. B. Trice, "Measurements of Neutron Energy Spectra with Threshold Detectors - Part II. Results of Reaction Neutron Energy Measurements with Fission Probes and Activation Detectors,"

2. W. N. McElroy, S. Bert, T. Crockett, and R. G. Hawkins, "SAND II - A Computer-Automated Iterative Method for Neutron Flux Determination by Foil Activation," AFWL-TR-67-41 (1967).

3. F. B. K. Kam and F. W. Stallman, "CRYSTAL BALL - A Computer Program for Determining Neutron Spectra from Activation Measurements," ORNL-TM-4601 (1974).

4. R. L. Heath, "Data Processing Techniques for Routine Application of Gamma-Ray Scintillation Spectrometry," USAEC Total Absorption Gamma-Ray Spectrometry Symposium Proceedings, pp. 147-158, 1960, Gatlinburg, Tennessee.

5. C. Eckart, "The Correction of Continuous Spectra for the Finite Resolution of the Spectrometer," Phys. Rev., 51, 735-738, May 1937.

6. A. S. Eddington, "On the Formula for Correcting Statistics for the Effects of a Known Probable Error of Observation," Monthly Notices Roy. Astron. Soc., 73, 359-360, March 1913.

7. A. Sommerfeld, Ph.D. Dissertation, 1891.

8. K. Linden and N. Starfelt, "Scintillation Spectroscopy of Continuous Gamma and X-Ray Spectra Below 1 MeV," Arkiv Fysik 2, 427-457, 1954.

9. R. W. Peelle, "Techniques Used at Oak Ridge National Laboratory for Unfolding Neutron and Gamma-Ray Pulse-Height Spectra," ORNL Report TM-3463 (August, 1971).

10. V.V. Verbinski, W. R. Burrus, T. A. Love, W. Zobel, N.W. Hill, and R. Textor, "Calibration of an Organic Scintillator for Neutron Spectrometry," Nucl. Instr. & Meth, 65, 8-25, 1968.

11. C. E. Clifford, *et al*, "Measurement of the Spectra of Uncollided Fission Neutrons Transmitted Through Thick Sampler of Nitrogen, Oxygen, Carbon, and Lead: Investigation of the Minima in Total Cross Sections," Nucl. Sci. Eng. 27, 299-307, 1967.

12. B. W. Rust and W. R. Burrus, "Suboptimal Methods for Solving Constrained Estimation Problems," DASA-2604 (1971).

13. J. Replogle, B. D. Holcomb, and W. R. Burrus, "The Use of Mathematical Programming for Solving Singular and Poorly Conditioned Systems of Equations," Jour. Math. Anal. and Appl., 20(2), Nov. 1967.

14. Louis Huszar, John H. Reed, and Martin Sperling, "MAZE3 - A Sequential Computer Spectral Enhancement Code," Defense Nuclear Agency Report DNA 3274F (December 1973).

15. B. W. Rust, "Constrained Internal Estimation for Ill Posed Problems," Invited Paper presented at SIAM Fall Meeting, Dec. 1975.

16. W. R. Burrus, ORNL Neutron Physics Division Annual Progress Report, ORNL-3193, pp. 44-52 (September 1, 1961).

17. Bert W. Rust and Walter R. Burrus, "Mathematical Programming and the Numerical Solution of Linear Equations," American Elsevier, New York, 1972.