JOURNAL OF COMPUTATIONAL PHYSICS 3, 167–175 (1968)

Error Estimates for Iterative Unfolding[1]

RAYMOND GOLD AND EDGAR F. BENNETT

Argonne National Laboratory, Argonne, Illinois 60439

Received April 8, 1968

ABSTRACT

Some approximate formulas are developed for errors arising from the application of iterative unfolding to experimental measurements. For the practical case of a dominantly diagonal response matrix of large order, it is shown that the relative error of the unfolded solution is approximately that of the original measurement. To demonstrate this behavior, actual experimental results from a proton-recoil proportional counter detection system have been included.

I. INTRODUCTION

The matrix formulation of the unfolding problem for detection systems is well known [1] and within this framework iterative unfolding has been considered in detail [2]. Applications of iterative unfolding to detection system problems in the nuclear sciences have been numerous [3]–[10]. The development of some simple error estimates that accrue from the iterative unfolding process are therefore of interest.

Elements of the output vector of a given detection system (e.g. the measured spectrum) will possess random statistical error. This fact implies the existence of random error in the iterative solution (i.e. the input vector). Expressions and estimates for this error are desirable. It must be emphasized at the outset that the present treatment will be applicable only within the context of two important assumptions. To begin with, one must assume that the iterative unfolding method furnishes *appropriate* solutions for the given detection system of interest [2]. In the event *appropriate* solutions cannot be found, the method of iterative unfolding fails to determine the desired physical solutions and for such cases error estimates cannot be regarded as meaningful. It will also be assumed that the response matrix is an exact representation of the systematic behavior of the

[1] Work performed under the auspices of the U.S. Atomic Energy Commission.

detection system and therefore does not possess random statistical error. In many instances, this latter assumption is justified since the actual random error in the response matrix is negligible compared with that of the output vector.

II. Analysis

Let the matrix representation of the detection system take the form [1]

$$\mathbf{Y} = A\mathbf{X}, \tag{1}$$

where A is the response matrix (of order n) and $\mathbf{Y}$ is the output vector. In addition, let $\mathbf{E}$ represent the random error of the output vector $\mathbf{Y}$. Hence, one need generally consider systems of the form

$$\mathbf{Y} + \mathbf{E} = \mathbf{Y}' = A\mathbf{X}. \tag{2}$$

Application of iterative unfolding for Eqs. (1) and (2) will lead to, under rather broad physical conditions [2], the convergent sequences $\{\mathbf{Y}^{(m)}\}$ and $\{\mathbf{Y}^{(m)'}\}$, respectively. Namely,

$$\lim_{m\to\infty} Y^{(m)} = \mathbf{Y}, \tag{3a}$$

$$\lim_{m\to\infty} \mathbf{Y}^{(m)'} = \mathbf{Y}' = \mathbf{Y} + \mathbf{E}. \tag{3b}$$

It follows from Eqs. (3a) and (3b) that $\mathbf{E}$ also represents the random error in $\mathbf{Y}^{(m)}$, for sufficiently large m. In other words, any given iterate $\mathbf{Y}^{(m)}$, for sufficiently large m, possesses the same random statistical behavior attributable to the output vector $\mathbf{Y}$. As an example, consider the case where the output vector $\mathbf{Y}$ is governed Poisson statistics. This assumption is commonly employed for many-particle detection systems and counting experiments that arise in nuclear measurements. For this case, the variance $S^2_{y_i} = y_i$, $i = 1, 2 \ldots n$. In this event, the variance $S^2_{y_i}(m)$ of $y_i^{(m)}$ is given by $S^2_{y_i}(m) = y_i^{(m)} = S^2_{y_i} = y_i$, $i = 1, 2 \ldots n$, for sufficiently large m.

The variance $S^2_{x_i}(m)$ of any given element $x_i^{(m)}$, in the iterative solution $\mathbf{X}^{(m)}$, can be calculated directly from the matrix relation

$$\mathbf{Y}^{(m)} = A\mathbf{X}^{(m)}. \tag{4}$$

One finds

$$S^2_{x_i}(m) = \sum_{j=1}^{n} [(a^{-1})_{ij}]^2 S^2_{y_j}(m), \qquad i = 1, 2 \ldots n, \tag{5a}$$

and therefore, for sufficiently large m,

$$S^2_{x_i}(m) = \sum_{j=1}^{n} [(a^{-1})_{ij}]^2 S^2_{y_i}, \qquad i = 1, 2 \ldots n, \tag{5b}$$

where $(a^{-1})_{ij}$ are the elements of the inverse matrix, A^{-1}. For the case of Poisson statistics, Eqs. (5a) and (5b) become

$$S^2_{x_i}(m) = \sum_{j=1}^{n} [(a^{-1})_{ij}]^2 y_j^{(m)}, \qquad i = 1, 2 \ldots n, \tag{6a}$$

and

$$S^2_{x_i}(m) = \sum_{j=1}^{n} [(a^{-1})_{ij}]^2 y_j, \qquad i = 1, 2 \ldots n. \tag{6b}$$

These expressions are cumbersone and difficult to employ since they involve elements of the inverse matrix, A^{-1}. Consequently, a less precise but more useful result, not involving A^{-1}, will be developed below. This analysis is based on the observation that most response matrices possess dominant diagonal elements and relatively small off-diagonal elements. Moreover, for diagonal response matrices, it follows from Eqs. (4) and (5a) that the relative error in the elements of $\mathbf{X}^{(m)}$ is identical with that of the corresponding elements of $\mathbf{Y}^{(m)}$, hence $\mathbf{Y}$. Therefore, one may qualitatively infer for dominantly diagonal response matrices that the relative error in the elements of $\mathbf{X}^{(m)}$ is approximately that of the corresponding elements of $\mathbf{Y}$.

To establish this conjecture rigorously, let the matrix A be partitioned in the form

$$A = A_d + A_o, \tag{7}$$

where the matrices A_d and A_o are defined as

$$\left.\begin{aligned} (A_d)_{ii} &= a_{ii}, & i &= 1, 2 \ldots n, \\ (A_d)_{ij} &= 0, & i &\neq j, \quad i, j = 1, 2 \ldots n \end{aligned}\right\} \tag{8a}$$

$$\left.\begin{aligned} (A_o)_{ii} &= (a_o)_{ii} = 0, & i &= 1, 2 \ldots n, \\ (A_o)_{ij} &= (a_o)_{ij} = a_{ij}, & i &\neq j, \quad i, j = 1, 2 \ldots n. \end{aligned}\right\} \tag{8b}$$

Consequently, A_d is simply a diagonal matrix whose elements coincide with the diagonal elements of A and the matrix A_o contains the remaining off-diagonal elements of A.

Taking differentials in Eq. (4) and using Eq. (7), one finds

$$\delta\mathbf{Y} = (A + A_o)\, \delta\mathbf{X}. \tag{9}$$

Since it has been assumed that m is sufficiently large to have established convergent results, the superscript m of the vectors $\mathbf{X}^{(m)}$ and $\mathbf{Y}^{(m)}$ has been omitted. In Eq. (9), the differential vector $\delta\mathbf{Y}$ can be identified with the random error vector $\mathbf{E}$, i.e., $\delta y_i = e_i$, $i = 1, 2 \ldots n$. Let us form the ensemble average $\langle \delta y_i \cdot \delta y_{i+1} \rangle$, using the representation given in Eq. (9). The quantities δy_i and δy_{i+1} are independent random variables, one therefore has

$$\langle \delta y_i \cdot \delta y_{i+1} \rangle = \left\langle \left[a_{ii}\, \delta x_i + \sum_{j=1}^{n} (a_0)_{ij}\, \delta x_j \right] \left[a_{i+1,i+1}\, \delta x_{i+1} + \sum_{k=1}^{n} (a_0)_{i+1,k}\, \delta x_k \right] \right\rangle = 0, \tag{10a}$$

which leads to

$$\langle \delta x_i \cdot \delta x_{i+1} \rangle = \frac{-1}{a_{ii} a_{i+1,i+1}} \sum_{j,k=1}^{n} (a_0)_{ij}\, (a_0)_{i+1,k} \langle \delta x_j\, \delta x_k \rangle. \tag{10b}$$

Since the off-diagonal elements of response matrices are no larger than n^{-1}, it is clear that the right-hand side of Eq. (10b) will become negligible for sufficiently large n. This conclusion is based on physical conditions implied by the matrix representation of the detection system. Namely, the physical description implies that the correlation between any two elements x_i and x_ℓ decreases as $| i - \ell |$ increases. It follows that the correlation between δx_i and δx_ℓ is even a more rapidly decreasing function of $| i - \ell |$. One may therefore assume that the quantities δx_i are independent random variables for sufficiently large n.

The validity of this assumption can also be demonstrated by using the relation

$$\delta x_i = \sum_{i=j}^{n} (a^{-1})_{ij}\, \delta y_j\,, \tag{11}$$

to obtain a direct comparison of $\langle (\delta x_i)^2 \rangle$ and $\langle \delta x_i \cdot \delta x_\ell \rangle$. One finds

$$\langle (\delta x_i)^2 \rangle = \sum_{i=1}^{n} [(a^{-1})_{ij}]^2 \langle (\delta y_j)^2 \rangle, \tag{12a}$$

and

$$\langle \delta x_i \cdot \delta x_\ell \rangle = \sum_{j=1}^{n} (a^{-1})_{ij}\, (a^{-1})_{\ell j} \langle (\delta y_j)^2 \rangle. \tag{12b}$$

Introducing appropriate mean values, m_d^2 and m_0^2, for the set of variances $\langle (\delta y_i)^2 \rangle$ which appear in these expressions, one has

$$\langle (\delta x_i)^2 \rangle = m_d^2 \sum_{j=1}^{n} [(a^{-1})_{ij}]^2, \tag{13a}$$

and

$$\langle \delta x_i \cdot \delta x_\ell \rangle = m_0^2 \sum_{j=1}^{n} (a^{-1})_{ij} (a^{-1})_{\ell j} . \tag{13b}$$

The resulting summations which appear in Eqs. (13a) and (13b) now have a simple interpretation. In particular, these sums represent elements of the matrix $A^{-1} \cdot \tilde{A}^{-1} = (\tilde{A}A)^{-1}$. Consequently one can write

$$\langle (\delta x_i)^2 \rangle = m_d^2 [(\tilde{A}A)^{-1}]_{ii} , \tag{14a}$$

and

$$\langle \delta x_i \cdot \delta x_\ell \rangle = m_0^2 [(\tilde{A}A)^{-1}]_{i\ell} . \tag{14b}$$

It follows from Eqs. (14a) and (14b) that $\langle (\delta x_i)^2 \rangle \gg |\langle \delta x_i \cdot x_\ell \rangle|$ for dominantly diagonal response matrices. Moreover, Eq. (14b) establishes the nature of the rapid decrease in $\langle \delta x_i \cdot \delta x_\ell \rangle$ with increasing $|i - \ell|$. Thus, when A is dominantly diagonal so is A^{-1}. Furthermore for A dominantly diagonal $\tilde{A}A$ will be even more dominantly diagonal. Consequently, as one moves away from the dominant diagonal, the off-diagonal elements of $(\tilde{A}A)^{-1}$ can be expected to decrease quite rapidly.

Calculation of the variance of y_i can be significantly simplified by employing this result. Equation (9) yields

$$\langle (\delta y_i)^2 \rangle = (a_{ii})^2 \langle (\delta x_i)^2 \rangle + \sum_{j=1}^{n} [(a_0)_{ij}]^2 \langle (\delta x_j)^2 \rangle . \tag{15}$$

Using once more the fact that off-diagonal matrix elements, $(a_0)_{ij}$, vanish at least as rapidly as n^{-1}, Eq. (15) reduces to the approximate condition

$$\langle (\delta y_i)^2 \rangle \cong a_{ii}^2 \langle (\delta x_i)^2 \rangle \tag{16}$$

for sufficiently large n. One can therefore write

$$\frac{\langle (\delta x_i)^2 \rangle^{1/2}}{x_i} \cong \frac{y_i}{a_{ii} x_i} \frac{\langle (\delta y_i)^2 \rangle^{1/2}}{y_i} . \tag{17}$$

Hence, for dominantly diagonal response matrices of sufficiently large order, the relative error in an element $x_i \in \mathbf{X}$ is approximately equal to the relative error in the corresponding element $y_i \in \mathbf{Y}$. For the case of Poisson statistics, Eq. (17) reduces to

$$\langle (\delta x_i)^2 \rangle^{1/2} \cong (y_i)^{1/2} / a_{ii} . \tag{18}$$

Some mention can be made of relaxing the condition that the response matrix be dominantly diagonal. In order to establish the (approximate) independence of the random variables $\{\delta x_i\}$, this condition may be too conservative for many applications. This observation is based on the behavior of the exact solution, $\mathbf{X} = A^{-1}\mathbf{Y}$, which is invariably beset with violent and unphysical oscillations. It follows that the elements of the inverse of a response matrix will generally be both positive and negative. One can therefore anticipate a great deal of cancellation in the sum of Eq. (13b) for sufficiently large n. Consequently, the condition, $\langle(\delta x_i)^2\rangle \gg |\langle \delta x_i \, \delta x_\ell \rangle|$, may actually hold in many applications under less restrictive assumptions. In this event, Eq. (15) provides the basis for a matrix representation of an unfolding problem for the error estimates, namely

$$\mathbf{S}_y = B\mathbf{S}_x \,, \tag{19}$$

where the elements of the vectors $\mathbf{S}_y$ and $\mathbf{S}_x$ are the variances $\{\langle(\delta y_i)^2\rangle\}$ and $\{\langle(\delta x_i)^2\rangle\}$, respectively. The matrix elements b_{ij} of B are defined by

$$b_{ij} = (a_{ij})^2 \qquad i, j = 1, 2 \ldots n. \tag{20}$$

This new unfolding problem is an analog of the original unfolding problem. Moreover, iterative unfolding is directly applicable since all pertinent physical conditions are obviously met [2]. The essential feature of this error estimate unfolding problem is that the new response matrix elements are merely the square of the elements in the original response matrix [cf. Eq. (20)].[2] For such cases, iterative unfolding can be employed, in principle, to determine error estimates and thereby one may again avoid the problems associated with both calculating and employing inverse matrix elements.

III. Application

The simple conclusion concerning the approximate preservation of the relative error can be readily demonstrated. The data displayed in Fig. 1 represent iterative unfolding results for an actual experimental case. For this system, the **Y** vector is the pulse-height distribution arising from the ionization of fast neutron-induced recoil protons in a hydrogen-filled proportional counter. The response matrix in this application is an approximate representation of so-called "wall-and-end-effect" distortion. Thus, the measured spectrum (or **Y** vector) contains recoil-proton events which leave the finite sensitive volume of the detector and thereby do not

[2] Here, the applicability of Poisson statistics would imply an even closer analogy, since it follows that $\mathbf{S}_y = \mathbf{Y}$ for Poisson statistics.

register the full ionization (or equivalent pulse-height) that normally corresponds to the event. Iterative unfolding yields the distortion-free solution **X** depicted in Fig. 1. This solution represents the infinite-medium proton-recoil spectrum that is desired, i.e. that spectrum attained in the limit of infinitely large sensitive volume. A more detailed account of this system can be found in Ref. [10].

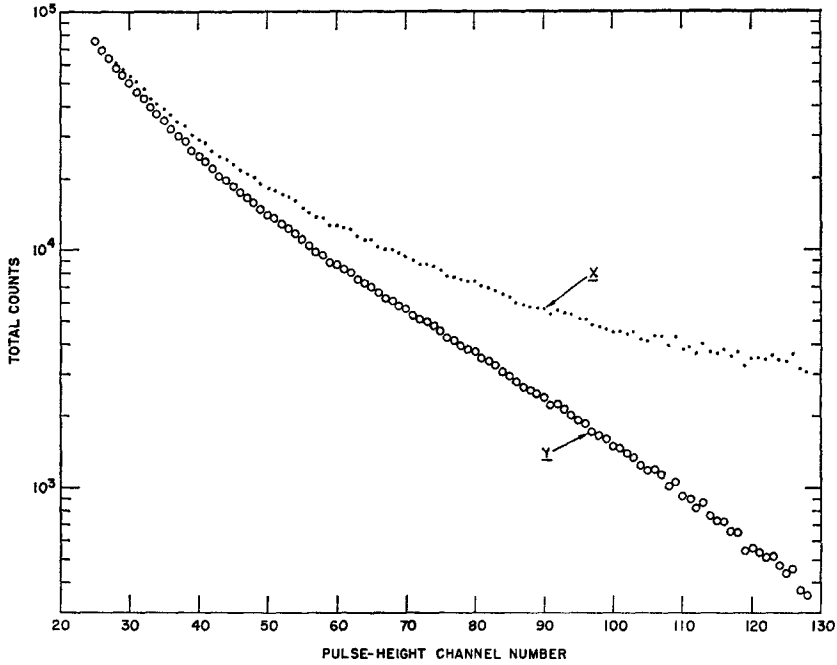

FIG. 1. Comparison of a proportional counter proton-recoil ionization spectrum **Y** and the distortion-free solution **X** obtained by iterative unfolding.

Table I presents the lower corner of the response matrix. It can be seen that this response matrix is dominantly diagonal and of large order. Hence, the approximations introduced above are fulfilled.

For this detection system, the spectrum **Y** arises from a nuclear counting experiment; consequently the elements of **Y** are governed by Poisson statistics. This behavior can be seen in Fig. 1. Here the monotonically decreasing nature of the measured **Y** spectrum with increasing pulse-height channel number gives rise to a random (Poisson) relative error which grows with increasing channel number. Inspection of Fig. 1 reveals that the statistical behavior of **X** mirrors almost precisely the random fluctuations in **Y** and thereby conclusively demonstrates that the relative error has been little affected by the iterative unfolding process.

TABLE I

RESPONSE MATRIX (LOWER CORNER)

Row No.	Column No. 94	95	96	97	98
78	2.4215 — 003	2.6248 — 003	2.8200 — 003	3.0039 — 003	3.1732 — 003
79	2.2881 — 003	2.4994 — 003	2.7047 — 003	2.9005 — 003	3.0835 — 003
80	2.1514 — 003	2.3694 — 003	2.5835 — 003	2.7903 — 003	2.9862 — 003
81	2.0116 — 003	2.2348 — 003	2.4566 — 003	2.6732 — 003	2.8810 — 003
82	1.8687 — 003	2.0959 — 003	2.3239 — 003	2.5491 — 003	2.7676 — 003
83	1.7231 — 003	1.9527 — 003	2.1856 — 003	2.4180 — 003	2.6460 — 003
84	1.5749 — 003	1.8055 — 003	2.0417 — 003	2.2799 — 003	2.5162 — 003
85	1.4244 — 003	1.6544 — 003	1.8925 — 003	2.1350 — 003	2.3780 — 003
86	1.2716 — 003	1.4998 — 003	1.7383 — 003	1.9835 — 003	2.2318 — 003
87	1.1170 — 003	1.3419 — 003	1.5792 — 003	1.8257 — 003	2.0777 — 003
88	9.6060 — 004	1.1810 — 003	1.4157 — 003	1.6618 — 003	1.9159 — 003
89	8.0276 — 004	1.0174 — 003	1.2481 — 003	1.4924 — 003	1.7470 — 003
90	6.4369 — 004	8.5151 — 004	1.0769 — 003	1.3178 — 003	1.5712 — 003
91	4.8364 — 004	6.8362 — 004	9.0247 — 004	1.1385 — 003	1.3893 — 003
92	3.2284 — 004	5.1413 — 004	7.2532 — 004	9.5520 — 004	1.2018 — 003
93	1.6155 — 004	3.4343 — 004	5.4595 — 004	7.6843 — 004	1.0093 — 003
94	1.9563 — 001	1.7192 — 004	3.6491 — 004	5.7883 — 004	8.1260 — 004
95	0.0000 + 000	1.9021 — 001	1.8274 — 004	3.8709 — 004	6.1250 — 004
96	0.0000 + 000	0.0000 + 000	1.8480 — 001	1.9391 — 004	4.0979 — 004
97	0.0000 + 000	0.0000 + 000	0.0000 + 000	1.7939 — 001	2.0534 — 004
98	0.0000 + 000	0.0000 + 000	0.0000 + 000	0.0000 + 000	1.7399 — 001

ACKNOWLEDGMENT

The authors are indebted to Ingeborg Olson for assistance with computer programs and numerical computations.

REFERENCES

1. R. GOLD and N. E. SCOFIELD, *Bull. Am. Phys. Soc.* **2**, 276 (1960).
2. R. GOLD, Argonne National Laboratory Report No. ANL-6984 (1964).
3. J. F. MOLLENAUER, University of California Radiation Laboratory Report No. UCRL-9748 (1961).
4. N. E. SCOFIELD, Paper (3-2), *in* "Applications of Computers to Nuclear and Radiochemistry," NAS-NS 3107, OTS. Department of Commerce, Washington, D.C. (1963).
5. R. SANNA, K. O'BRIEN, M. ALBERG, S. ROTHENBERG, and J. MC LAUGHLIN, U.S.A.E.C. Health and Safety Laboratory Report No. HASL-162 (1964).

6. B. S. J. DAVIES, Paper (4-3), *in* "Radiation Measurements in Nuclear Power." Berkeley Nuclear Laboratories; The Institute of Physics and the Physical Society, London (1966).
7. M. ALBERG, K. O'BRIEN, and J. MC LAUGHLIN, *Nucl. Sci. Eng.* **25**, 303 (1966).
8. D. MILLER, P. SCHLOSSER, J. BURT, D. D. GLOWER, and J. M. MC NEILLY, *IEEE Trans. Nucl. Sci.* **NS-14**, 245 (1967).
9. W. N. MCELROY, S. BERG, and G. GIGAS, *Nucl. Sci. Eng.* **27**, 533 (1967).
10. R. GOLD and E. F. BENNETT, *Nucl. Instr. Methods* **63**, 285 (1968).