Comment

GABOR T. HERMAN, YAIR CENSOR, DAN GORDON, and ROBERT M. LEWITT*

We wish to place the development presented by Vardi, Shepp, and Kaufman into the framework of the research area that has become known as "image reconstruction from projections." In the terminology of that field, the technique presented by Vardi et al. is a finite series-expansion method. A recent tutorial on such methods was given by Censor (1983). Figure 1 gives an overview of the methodology of the finite series-expansion methods.

The first significant step is the discretization of the problem. Vardi et al. chose to do this based on the traditional approach, in which the function of two variables, which is to be reconstructed, is assumed to be constant in uniform square-shaped regions of space, called pixels (short for picture elements). Such a discretization for image reconstruction has been used since the beginnings of computerized tomography (e.g., see Gordon et al. 1970), and it is certainly the most popular approach to discretization. However, there are some very different alternatives (e.g., see Lewitt 1983 and Eggermont 1983). Although we have no basic objection to pixel-based discretization, we cannot agree with Vardi et al. that "for the maximum likelihood method of Section 2.1, there seems to be no essential difference in the final result if we discretize either at the outset or after deriving a functional equation (using a limiting argument) that defines the estimate in the continuous case" (p. 10). In our experience (Gray et al. 1982) the error introduced in the model by discretization at the outset can far outweigh that due to the limited number of photons. The maximum-likelihood approach takes proper care of the error due to photon statistics, but it ignores the discretization error.

The next step is to set up a system of algebraic equations. The attitude in finite series-expansion image reconstruction (Censor 1983) is usually that these equations are only *approximate* due to both discretization and statistical errors. Thus we

* Gabor T. Herman is Professor and Chief of Medical Imaging, Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104; Yair Censor is Senior Lecturer, Department of Mathematics, University of Haifa, Mt. Carmel, Haifa 31999, Israel; Dan Gordon is Visiting Associate Professor, Department of Computer Science, University of Cincinnati, OH 45221; and Robert M. Lewitt is Research Associate Professor, Medical Image Processing Group, Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104.
The work of the authors in this area was supported by National Institutes of Health Grant HL-28438 and National Science Foundation Grant ECS-8117908.

Reprinted from: © 1985 American Statistical Association
Journal of the American Statistical Association
March 1985, Vol. 80, No. 389, Applications

would write, instead of (2.18) of Vardi et al., the approximate system

$$\mathbf{n}^* \simeq \mathbf{P}'\boldsymbol{\lambda}, \tag{1}$$

where $\mathbf{P}'$ is a $D \times B$ matrix and $\mathbf{n}^*$ is the observed measurement vector.

The next step is to choose a criterion according to which we can decide whether or not a particular $\boldsymbol{\lambda}$ is a "solution" of (1). This is typically an optimization criterion of the form

$$\text{optimize } l(\boldsymbol{\lambda}), \quad \text{subject to } \boldsymbol{\lambda} \in Q. \tag{2}$$

The log-likelihood function optimized in the article by Vardi et al. (ignoring a constant term) is

$$l(\boldsymbol{\lambda}) = \sum_{d=1}^{D} [-\lambda^*(d) + n^*(d)\log \lambda^*(d)], \tag{3}$$

where

$$\lambda^*(d) = \sum_{b=1}^{B} \lambda(b)p(b, d). \tag{4}$$

The constraint set Q in Equation (2) is the non-negative orthant of the B-dimensional Euclidean space. As far as we know, the first proposal to use the maximum-likelihood optimization criterion in emission image reconstruction from projections was made by Rockmore and Macovski (1976).

There have been many alternative optimization criteria proposed in the literature. One that we have found potentially useful in transmission tomography with low-photon count (Gray et al. 1982) is

$$l(\boldsymbol{\lambda}) = r^2\|\mathbf{n}^* - \mathbf{P}'\boldsymbol{\lambda}\|^2 + \|\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}\|^2, \tag{5}$$

where $\hat{\boldsymbol{\lambda}}$ is some predetermined picture (it could, e.g., be the expected prior value of $\boldsymbol{\lambda}$ in a Bayesian approach) and r is a predetermined constant (reflecting our belief in the accuracy of the measurements relative to the importance of the prior distribution). We call this the regularized least squares (RLS) criterion. This and other criteria were discussed in Censor (1983).

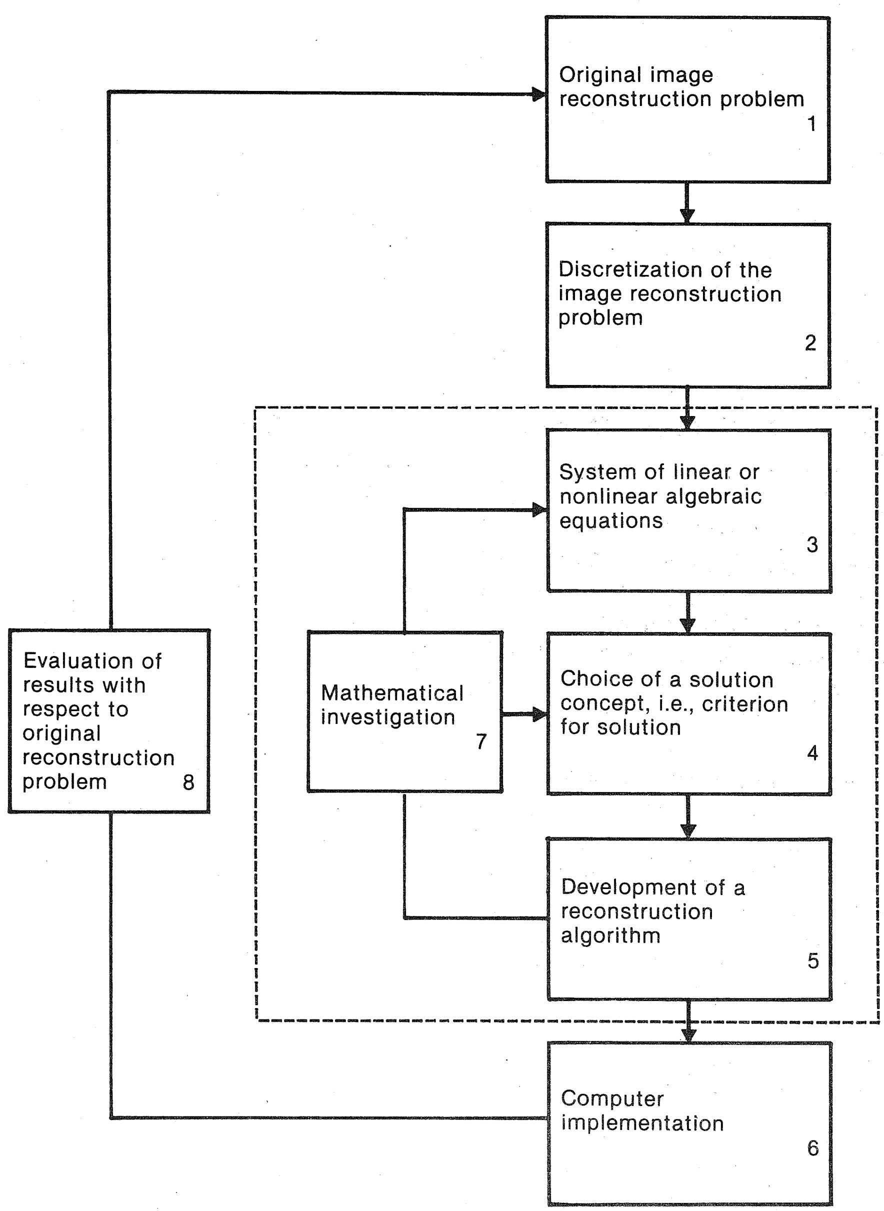

Figure 1. Methodology of the Series-Expansion Approach. Reprinted from Censor (1983) (© 1983 IEEE).

The next step is the development of a reconstruction algorithm. Vardi et al. propose the EM algorithm for maximizing (3); the iterative step of this algorithm is equation (2.13) of their article. In the same notation, the algorithm proposed for minimizing (5) by Herman et al. (1980) is the following (it makes use of a D-dimensional dual variable $\boldsymbol{\mu}$, whose dth component is denoted by $\mu(d)$, as well as the B-dimensional primal variable $\boldsymbol{\lambda}$).

1. Start with an initial estimate of $\boldsymbol{\lambda}^{\text{old}}$, which is $\hat{\boldsymbol{\lambda}}$, and an initial estimate of $\boldsymbol{\mu}^{\text{old}}$ with all components equal to zero.

2. If $\boldsymbol{\lambda}^{\text{old}}$ and $\boldsymbol{\mu}^{\text{old}}$ denote the current estimates of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, define the new estimates, $\boldsymbol{\lambda}^{\text{new}}$ and $\boldsymbol{\mu}^{\text{new}}$, by the following process: Let $\boldsymbol{\lambda}^{(0)} = \boldsymbol{\lambda}^{\text{old}}$. For each $d = 0, \ldots, D-1$, do the following three steps:

$$c = \rho \frac{r\left[n^*(d) - \sum_{b=1}^{B} \lambda^{(d)}(b)p(b, d)\right] - \mu^{\text{old}}(d)}{1 + r^2 \sum_{b=1}^{B} [p(b, d)]^2}, \quad (6)$$

$$\lambda^{(d+1)}(b) = \lambda^{(d)}(\mathrm{b}) + \mathrm{rcp}(\mathrm{b}, d), \qquad b = 1, \ldots, B, \quad (7)$$

and

$$\mu^{\text{new}}(d) = \mu^{\text{old}}(d) + c, \quad (8)$$

where ρ is a constant called the *relaxation parameter*. Finally, let $\boldsymbol{\lambda}^{\text{new}} = \boldsymbol{\lambda}^{(D)}$.

This algorithm looks somewhat more complicated than it really is due to the complexity of notation (compare Censor 1983 and Herman et al. 1980). We have adopted this notation, since it allows us to point out the similarities and differences between it and the EM algorithm of Vardi et al. The similarities are that in each algorithm (a) getting from $\boldsymbol{\lambda}^{\text{old}}$ to $\boldsymbol{\lambda}^{\text{new}}$ involves the calculation of an inner product between an estimate of $\boldsymbol{\lambda}$ with a $p(\cdot, d)$ vector D times [see (2.13) of Vardi et al. and our Equation (6)] and (b) the calculation of $\lambda^{\text{new}}(b)$ involves (for $d = 1, \ldots, D$) the multiplication of $p(b, d)$ by a previously calculated value and then adding the product to an accumulated partial sum [see (2.13) of Vardi et al. and our Equation (7)]. Thus the computer cost of one iterative step from $\boldsymbol{\lambda}^{\text{old}}$ to $\boldsymbol{\lambda}^{\text{new}}$ is likely to be similar for the two algorithms. The differences are three-fold:

1. Our algorithm makes use of both a dual-vector and a relaxation parameter, neither of which appears in the EM algorithm of Vardi et al.

2. The EM update is "simultaneous," whereas ours is "row-by-row." In the simultaneous technique an update is done based on all of the measurements, whereas the estimate is updated separately for each measurement before going to the next one in the row-by-row technique [see (7)]. That simultaneous and row-by-row techniques can sometimes be considered extreme cases of more general block-iterative techniques has been demonstrated in Eggermont et al. (1981).

3. The updating in the EM algorithm is multiplicative, whereas the updating in Equation (7) is additive. Thus the EM algorithm is more similar to the multiplicative algebraic reconstruction technique (MART) than to the additive one (ART), both of which were introduced in Gordon et al. (1970) and were discussed in Censor (1983).

The next stage depicted in Figure 1 is the mathematical investigation. A major part of this is proving that the solution specified by the solution concept exists and is unique, and that the proposed algorithm converges in the limit to the desired solution as specified by the optimization criterion. This is the content of the theorem in Section 2.1, which in particular gives a condition for the uniqueness of the maximum likelihood solution. We point out that the RLS solution of Equation (5) always exists and is always unique, and that the algorithm described using Equations (6), (7), and (8) converges to this unique solution, provided $0 < \rho < 2$ (Herman et al. 1980).

The next stage in the process is computer implementation. In a previous paper, Shepp and Vardi (1982) published a computer code for their EM algorithm. A modification of this code was implemented by G. Muehllehner in our department on a VAX computer (with some kind advice from Vardi). Muehllehner has given us access to this computer code, and due to the similarity of the two algorithms we easily made a version of it that is an implementation of the RLS algorithm as described before. As expected, an iterative step of either algorithm takes approximately the same time.

The next stage of investigation is the evaluation of the results with respect to the original problem. One aspect of such work is the experimental evaluation of the method for the type of application for which it was intended, as was done by Shepp et al. (1984). An alternative interesting problem is that of the relative performance of an algorithm in terms of its own criterion. For example, one may ask whether alternative iterative algorithms [which may not maximize the log-likelihood function (3) in the limit] may in their early stages achieve a faster increase of the log-likelihood function than the EM algorithm of Vardi et al. Since these iterative image-reconstruction techniques are expensive compared with some alternative methods (one iteration typically costs as much as a whole reconstruction by what Vardi et al. call the convolution back-projection technique; see Herman 1980) and since reconstructions have to be performed for a number of cross sections for each individual patient, these considerations of cost are relevant to the original medical problem. In an earlier paper (Herman 1982) we have found that an iterative algorithm (ART) originally designed for minimizing the norm (subject to equality and non-negativity constraints) does as well in its early iterates in maximizing the entropy functional defined by

$$I(\boldsymbol{\lambda}) = -\sum_{b=1}^{B} \lambda(b)\log \lambda(b) \quad (9)$$

as the already mentioned MART, which is known to converge in the limit to the maximum entropy solution.

Based on these considerations we decided to compare the early behavior of the EM algorithm of Vardi et al. with that of the RLS minimizing algorithm of Equations (6), (7), and (8). For this purpose, we have taken a particular data set and applied the two algorithms for 10 iterations each, starting with the same $\hat{\boldsymbol{\lambda}}$. The data set is based on the so-called Derenzo resolution phantom, which we have used in previous studies (Muehllehner et al. 1983), and the geometry of data collection

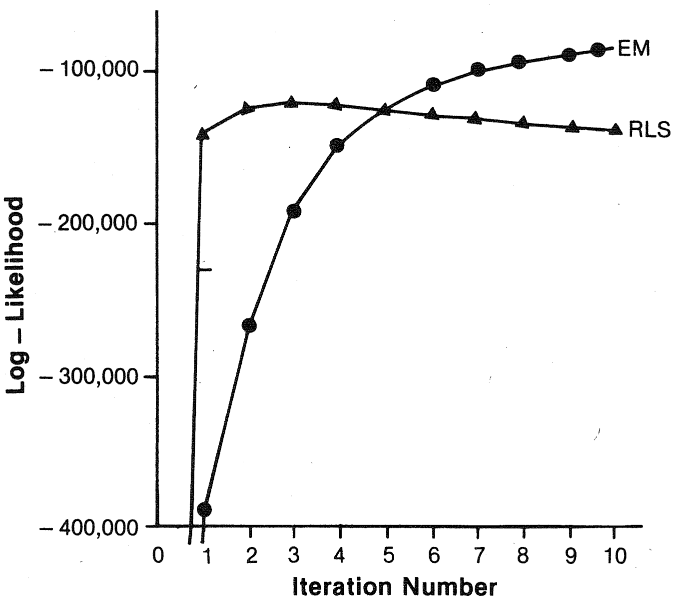

Figure 2. Plots of the Log-Likelihood Function. The dots indicate the values of the log-likelihood function using the EM algorithm of Vardi et al. The triangles indicate the values of the same function using the RLS algorithm of Equations (6), (7), and (8) on the same data, starting with the same initial estimate (0th iterate), for which the value of the log-likelihood is less than $-2{,}000{,}000$.

simulates that of the hexagonal Anger-camera-based PET device built by Muehllehner and his associates (Muehllehner et al. 1983). In the RLS algorithm we set $r = 10$ and $\rho = .05$. Figure 2 shows plots of the log-likelihood function for the first 10 iterates (not counting the 0th iterate $\hat{\boldsymbol{\lambda}}$). As can be seen, the EM algorithm behaves as expected based on the discussion of Vardi et al. The RLS algorithm has an interesting behavior from the point of view of log-likelihood: Initially it far outperforms the EM algorithm, it peaks at the third iterate, after which it slowly declines, and it is eventually overtaken by the EM algorithm at the sixth iterate. (We note in passing that the RLS functional is steadily reduced by the RLS algorithm.)

In this Comment we have attempted to put the very interesting work of Vardi et al. into the general framework of finite series expansion methods for image reconstruction from projections. We hope that this might lead to deeper understanding and medically useful generalizations of the method.

ADDITIONAL REFERENCES

Censor, Y. (1983), "Finite Series-Expansion Reconstruction Methods," *Proceedings of the IEEE,* 71, 409–419.

Eggermont, P. P. B. (1983), "Tomographic Reconstruction on a Logarithmic Polar Grid," *IEEE Transactions on Medical Imaging,* MI-2, 40–48.

Eggermont, P. P. B., Herman, G. T., and Lent, A. (1981), "Iterative Algorithms for Large Partitioned Linear Systems, With Applications to Image Reconstruction," *Linear Algebra and Its Applications,* 40, 37–67.

Gordon, R., Bender, R., and Herman, G. T. (1970), "Algebraic Reconstruction Techniques (ART) for Three-Dimensional Electron Microscopy and X-ray Photography," *Journal of Theoretical Biology,* 29, 471–481.

Gray, J. E., Hanson, D. P., Herman, G. T., Lewitt, R. M., Reynolds, R. A., Robb, R. A., Smith, B. and Tuy, H. (1982), "Reconstruction Algorithms for Dose Reduction in X-ray Computed Tomography," Medical Image Processing Group Technical Report MIPG63, University of Pennsylvania, Dept. of Radiology.

Herman, G. T. (1980), *Image Reconstruction From Projections: The Fundamentals of Computerized Tomography,* New York: Academic Press.

——— (1982), "Mathematical Optimization Versus Practical Performance: A Case Study Based on the Maximum Entropy Criterion in Image Reconstruction," *Mathematical Programming Studies,* 20, 96–112.

Herman, G. T., Lent, A., and Hurwitz, H. (1980), "A Storage-Efficient Algorithm for Finding the Regularized Solution of a Large, Inconsistent System of Equations," *Journal of the Institute of Mathematics and Its Applications,* 25, 361–366.

Lewitt, R. M. (1983), "Reconstruction Algorithms: Transform Methods," *Proceedings of the IEEE,* 71, 390–408.

Muehllehner, G., Colsher, J. G., and Lewitt, R. M. (1983), "A Hexagonal Bar Positron Camera: Problems and Solutions," *IEEE Transactions on Nuclear Science,* NS-30, 652–660.

Rockmore, A. J., and Macovski, A. (1976), "A Maximum Likelihood Approach to Emission Image Reconstruction From Projections," *IEEE Transactions on Nuclear Science,* NS-23, 1428–1432.