Maximum Entropy Image Reconstruction

STEPHEN J. WERNECKE, STUDENT MEMBER, IEEE, AND LARRY R. D'ADDARIO, MEMBER, IEEE

Abstract—**Two-dimensional digital image reconstruction is an important imaging process in many of the physical sciences. If the data are insufficient to specify a unique reconstruction, an additional criterion must be introduced, either implicitly or explicitly, before the best estimate can be computed. Here we use a principle of maximum entropy, which has proven useful in other contexts, to design a procedure for reconstruction from noisy measurements. Implementation is described in detail for the Fourier synthesis problem of radio astronomy. The method is iterative and hence more costly than direct techniques; however, a number of comparative examples indicate that a significant improvement in image quality and resolution is possible with only a few iterations. A major component of the computational burden of the maximum entropy procedure is shown to be a two-dimensional convolution sum, which can be efficiently calculated by fast Fourier transform techniques.**

Index Terms—**Digital image processing, Fourier synthesis, image processing, image reconstruction, maximum entropy, radio telescopes, statistical estimation theory.**

INTRODUCTION

THE reconstruction problem considered here is the estimation of a two-dimensional function $f(x,y)$ from a finite number of noise-corrupted, linear measurements $\{m_i, i = 1,2,\cdots,M\}$ of the form

$$m_i = \int\int_D h_i(x,y)f(x,y)dxdy + e_i, \tag{1}$$

where e_i is an error term and f is zero outside the known region D. We borrow optics terminology by calling the unknown function f an object and its reconstruction $\hat{f}$ an image. Specification of the measurement kernel $h_i(x,y)$ depends on the application. Problems of this type have arisen in a number of diverse fields—radiography, radio astronomy, and optics, to mention a few—and it is only within the last few years that the various disciplines have appreciated their common interest in reconstruction. This discovery has revealed a large number of reconstruction algorithms and, inevitably, a considerable duplication of research effort.

We restrict ourselves to two-dimensional reconstruction; however, this is not a serious limitation. All of the results in this paper can be straightforwardly generalized to higher dimensions, but to avoid unnecessary mathematical abstraction we have not done so. The bulk of current research is directed toward digital reconstruction, at least as a simulation tool for the development of special-purpose hardware, and the most noticeable effect of reconstruction in higher dimensions is the rapid increase in the processing burden. Even in applications where the structures of interest are inherently three-dimensional, e.g., radiography, two-dimensional reconstruction is an important activity because it is possible to design measurement apparatus so parallel two-dimensional slices can be reconstructed separately and "stacked" to yield a three-dimensional reconstruction.

The assumption that the measurement process is linear, apart from errors, is not uncommon, and the extent to which it reflects physical reality depends on the application. Now that the theory and practice of reconstruction from linear measurements is maturing, attempts to incorporate nonlinearities into the measurement model have begun [1], [2]. To some extent, the linearity assumption is responsible for the success of interdisciplinary cooperation in reconstruction research. Nonlinearities tend to be particular to each application, and efforts to deal with significant departures from linearity become quite specialized.

The measurement model (1) is general enough to show that a variety of seemingly different applications involve similar mathematics. If, for example, we choose

$$h_i(x,y) = \text{rect}\left[\frac{x\cos\theta_i + y\sin\theta_i - R_i}{w_i}\right] \tag{2a}$$

we describe the radiographic problem of reconstruction from projection measurements of the linear attenuation coefficient integrated along the path of a collimated X-ray beam. The geometry for this activity is shown in Fig. 1.

Use of a complex exponential kernel

$$h_i(x,y) = \exp\left[-j2\pi(u_ix + v_iy)\right] \tag{2b}$$

with $j = \sqrt{-1}$ models the measurement m_i as a sample of the two-dimensional Fourier transform of the object at spatial frequency (u_i,v_i). This situation is important in radio astronomy where reconstruction is sometimes called Fourier synthesis.

If we let

$$h_i(x,y) = h(x_i - x, y_i - y) \tag{2c}$$

measurements become samples of the convolution of the image and a space-invariant point spread function $h(\cdot,\cdot)$.

Manuscript received December 1, 1975; revised April 10, 1976. This work was supported in part by the National Science Foundation under Grant DCR75-15140 and in part by the National Radio Astronomy Observatory. Portions of this work have been presented at the Image Processing for 2-D and 3-D Reconstruction from Projections: Theory and Practice in Medicine and the Physical Sciences Meeting, Stanford University, Stanford, CA, August 1975.

S. J. Wernecke is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

L. R. D'Addario is with the National Radio Astronomy Observatory, Socorro, NM 87801.

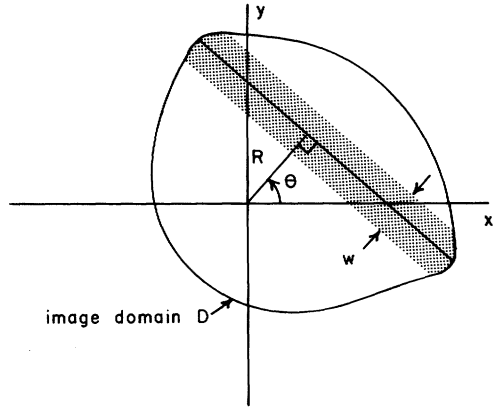


Fig. 1. A projection measurement with radius R, angle θ, and width w determines the mean image brightness in the shaded strip.

Here the reconstruction problem becomes one of deconvolution or restoration of a blurred object.

In pointing out the mathematical similarity among these three examples, we do not imply that the problems are identical nor do we claim that an algorithm developed for one application is guaranteed suitable for use in another. For example, the restoration problem (2c) involves space-invariant linear filtering, and this permits processing techniques that would not be appropriate in the other cases. We can, however, say that techniques to invert the general measurement equation (1) will be applicable to all of the specialities of (2). In some cases, these general methods will enjoy simplification in the context of a particular reconstruction application.

Statistical Reconstruction Techniques

The error term is deliberately included in (1) because we intend to discuss certain statistical reconstruction techniques that are designed to acknowledge and, to some extent, compensate for measurement errors. To explicitly consider measurement errors in the formulation of new reconstruction schemes is, we feel, important. Many of the currently used reconstruction algorithms have been derived without this consideration. The effect of noise on linear reconstruction procedures can be explored by superposition; however, superposition does not hold for the nonlinear algorithms we discuss here, and this makes conventional error analysis more difficult. We note, also, that reconstructions are often categorized according to some resolution figure-of-merit, a number that is frequently determined by measurement apparatus geometry alone despite the fact that the useful resolution of a reconstruction is strongly influenced by noise. It would be helpful if properly designed, error-acknowledging algorithms could "sense" the resolution inherent in a set of noisy measurements and yield reconstructions lacking spurious detail.

Also important is the incorporation of *a priori* knowledge into the reconstruction process. One constraint operative in most reconstruction applications is that the objects—representing absorptivities, densities, emissivities, and the like—cannot, for physical reasons, assume negative values. It is difficult to introduce this property into the simpler reconstruction procedures; however, constraints of this sort can be naturally included in statistically motivated techniques.

One aspect common to all reconstruction problems is the insufficiency of a finite set of measurements to specify a unique reconstruction, even permitting the luxury of error-free measurements for the moment. This is due to the fact that a continuous function possesses, in general, an infinite number of degrees of freedom. Of course, in practice f can often be adequately described by a finite number of parameters, and such an approximation is necessary if we are to consider digital reconstruction. Choosing the dimension of a suitable finite representation is a nontrivial problem whose solution depends on the structures of interest, the final purpose for which reconstruction is needed, the speed and memory offered by available computing facilities, etc. We will not consider this problem here, but we will assume that an intelligent choice has been made so we can concentrate on other aspects of reconstruction. Thus, the object is represented by N parameters $(f_1,f_2,\cdots,f_N)$; frequently, these parameters are just samples of the object $f_i = f(x_i,y_i)$.

It often happens that the number of parameters needed for an adequate description is considerably greater than the number of available measurements. In this event, even the finite representation of the object is underdetermined, and there is an infinity of possible solutions. A fundamental decision must be made: should we redefine the reconstruction task so the data are sufficient to uniquely specify the solution to the new problem, or should we stick with the original formulation and decide which of the solutions is "best" in some sense? In this paper, we deal with the latter option.

One approach is to use Bayes' theorem

$$P(\hat{f}|\{m_i\}) = \frac{P(\{m_i\}|\hat{f})P(\hat{f})}{P(\{m_i\})} \tag{3}$$

which calculates the *a posteriori* probability that any reconstruction $\hat{f}$ is correct given the available measurements. This conditional probability can be maximized once the terms in the numerator of the right-hand side of (3) are specified. The denominator is a constant and does not affect the optimization. Assuming the measurement equation (1) and the noise statistics are known, determination of $P(\{m_i\}|\hat{f})$ is usually straightforward. The difficulty with Bayes' theorem is that it depends on the *a priori* probability density $P(\hat{f})$. To find this, there must be a statistical model that gives, for every possible object, an *a priori* probability of occurrence.

Several workers, studying restoration of blurred photographs [3] and reconstruction from projections [4], have used Bayesian methods with the *a priori* image probability density taken to be multivariate Gaussian. While this assignment is made primarily for analytic convenience, Hunt and Cannon [5] have shown the Gaussian model to be a good one provided the mean object is not taken to be uniformly gray. They adopt a context-dependent mean object, obtained by blurring one member of the image ensemble. Although the use of a Gaussian prior does not explicitly constrain images to be nonnegative, negative values occur with low probability if fluctuations about a positive mean image are not too large.

Maximum Entropy Reconstruction

In this paper, we will consider an approach motivated by non-Bayesian, but still probabilistic, arguments. In abstract terms, the procedure begins with the assignment of an *entropy measure* to indicate the randomness or uncertainty in a statistical environment. Each measurement or piece of *a priori* information, assuming these data to be free of contradiction, reduces the randomness and restricts statistical freedom. If the data are insufficient to eliminate all uncertainty, a principle of maximum entropy can be used to assure that the admittedly incomplete description of the situation reflects only the statistical structure imposed by the available information. This maximum entropy description retains all of the uncertainty not removed by the data, and it has been interpreted as the description that is most objective [6] or maximally noncommittal with respect to missing information [7], [8].

The concept of entropy is a familiar one in information theory, probability theory, and thermodynamics [9], although its place in the present problem of image reconstruction is not yet clear. To define an entropy measure that is useful in the context of reconstruction we must formulate a statistical model for the imaging process. Two such models, leading to different entropy measures, have been suggested in the literature, and each has attracted followers. Once a given model is adopted, maximum entropy (ME) reconstruction proceeds by finding the reconstruction that maximizes the entropy measure and is compatible with all the available data, both measurements and *a priori* information.

To develop the first model, we require a discrete representation of the object in terms of picture elements (pixels). Let the object be partitioned into N pixels, each of area ΔA, and let f_i be the average brightness in the ith pixel. We suppose that the brightness arises from the random emission of discrete particles (call them photons), each of energy e, and we define the brightness as

$$f_i = \frac{e}{\Delta A} r_i \tag{4}$$

where r_i is the average rate of emission of photons from the ith pixel. Under this model, the probability that a photon was emitted from the ith pixel, given that it was emitted from the object, is

$$P_i = \frac{r_i}{\Sigma_i r_i} = \frac{f_i}{\Sigma_i f_i} = \frac{f_i}{F} \tag{5}$$

where $F = \Sigma_i f_i$ is the total intensity.

The entropy of the discrete probability distribution (5) is

$$H_1 = -\sum_{i=1}^{N} P_i \log P_i = -\sum_{i=1}^{N} \frac{f_i}{F} \log \frac{f_i}{F}. \tag{6}$$

Formula (6) is solidly grounded in information theory [6], [10] and measures the uncertainty as to which pixel emitted a given photon. We emphasize that it is the random emission model which enables us to make the connection between discrete pictures and the entropy measure H_1; this model may not be appropriate for all reconstruction problems.

Several authors [11], [12] have used an "entropy measure" similar to H_1, namely,

$$-\sum_{i=1}^{N} f_i \log f_i. \tag{7}$$

Note that maximization of H_1 is equivalent to maximization of (7) if the latter is done under the constraint F = constant, i.e., if we regard the total intensity as exactly known. We prefer to avoid this restriction.

The second statistical model is based on an extension of Burg's work on spectral analysis of stationary random processes [13], [14]. There the problem is to estimate the power spectrum $S(\nu)$, which is the Fourier transform of the autocorrelation function $R(\tau)$ of a stationary random process $x(t)$. In many applications it happens that observations include only a finite number of samples of $R(\tau)$ or of a particular realization of $x(t)$. The samples are not sufficient to specify $S(\nu)$ uniquely, just as in the reconstruction problem where $\{m_i\}$ cannot specify $f(x,y)$ uniquely. Also, $S(\nu)$ must be nonnegative, as must $f(x,y)$ in all applications considered here. The maximum entropy

method of power spectrum estimation is based on the fact that the entropy rate E of a stationary, band-limited random process is related in a simple way to its power spectrum, namely,

$$E = \int_{-\nu_c}^{\nu_c} \log S(\nu) d\nu + C \tag{8}$$

where ν_c is the cutoff frequency, and C depends on higher order statistics of the process and is independent of $S(\nu)$ [8], [15].

The ME philosophy is to assume that the process under investigation is as random, or statistically structureless, as possible. Applying this principle to the spectral analysis problem, we select as our spectral estimate the function $S(\nu)$ that maximizes the entropy rate (8) and is consistent with the available observations. The basic mathematics and an important algorithm for ME spectral analysis were developed by Levinson [16]; however, the usefulness of this technique was not widely appreciated until it was interpreted and extended by Burg [13], [14], and others [17], [18]. For a tutorial discussion of ME theory, the reader should consult [8].

Generalizing (8) to two dimensions and letting $f(x,y)$ correspond to $S(\nu)$, we obtain a second entropy measure

$$H_2 = \int \int_D \log f(x,y) dx dy \tag{9}$$

which has been adopted by Ables [7] and Ponsonby [19]. If the imaging process can be regarded as two-dimensional spectral estimation, the appropriateness of H_2 as an entropy measure follows immediately. This is the case in at least one important reconstruction application in radio astronomy. Here, an incoherent radio source gives rise to a random electric field whose spatial autocorrelation function is sampled by radio interferometers. The relationship between the unknown radio brightness distribution, which is to be reconstructed, and the available interferometer measurements is embodied in the van Cittert–Zernike theorem [20] and is a Fourier transformation. Hence, reconstruction in radio astronomy can be described as power spectral analysis of the electric field, and use of H_2 as an entropy measure is justified.

The existence of two different entropy measures for image formation is interesting. Which entropy measure, if either, is correct for a given application depends on the physics of the measurement process. There has been speculation (Ponsonby, private communication) about the relationship between the two measures introduced here; however, no clear statement of this relationship has yet emerged. Frieden [11] has pointed out that reconstructions found using H_1 tend to be smooth in a certain sense. Wernecke [21] has interpreted both H_1 and H_2 as simple smoothness indicators and has suggested other criteria that could also serve this purpose. Thus, different formulations of ME reconstruction can be regarded as special cases in the larger framework of reconstruction with maximum smoothness. It is felt, however, that entropy criteria have a stronger theoretical foundation than *ad hoc* smoothness criteria since the former are grounded in specific statistical models.

Application to Fourier Synthesis

For concreteness, we will restrict the remainder of the discussion to ME reconstruction using H_2 in the context of Fourier synthesis as it arises in radio astronomy (see, e.g., [22]). We are interested in reconstructing a radio brightness distribution from noisy samples of its two-dimensional Fourier transform,

$$m_i = \int \int_D f(x,y) \exp\left[-j2\pi(u_i x + v_i y)\right] dx dy + e_i. \tag{10}$$

Since measurements are made in the transform domain, it is natural to consider reconstruction by Fourier inversion. This would be a simple matter if the image transform were known everywhere in the (u,v)-plane; however, this is usually not the case in practice. For economic and technical reasons, measurement coverage of the (u,v)-plane is seldom as complete as would be desired in the absence of these constraints.

The dilemma is how to form an image from a finite number of Fourier transform samples, too few to specify a unique reconstruction. It is instructive to consider what a proposed reconstruction technique implies about the transform at points where measurements are not available. It is not uncommon in this field to see reconstructions defined by equations that assume, at least implicitly, that unmeasured values are zero. An example of this treatment is afforded by what has been called the direct transform reconstruction

$$\hat{f}_D(x,y) = \sum_{i=1}^{M} \alpha_i \{m_i^{(R)} \cos\left[2\pi(u_i x + v_i y)\right] - m_i^{(I)} \sin\left[2\pi(u_i x + v_i y)\right]\} \tag{11}$$

where $m_i = m_i^{(R)} + j m_i^{(I)}$. The real apodizing constants α_i can be chosen to improve the appearance of the point source response in a fashion analogous to use of "lag windows" in classical power spectrum estimation [23].

Methods of this sort work well if measurement coverage of the (u,v)-plane is nearly complete or if the missing data would have had values close to zero had these extra numbers been available. If coverage is restricted or irregular, however, images reconstructed by such algorithms can suffer from serious sidelobe artifacts. There will also be areas in which the reconstructed brightness is negative and hence physically inadmissible.

The existence of regions of negative brightness on a radio map is not, in itself, disastrous; radio astronomers have been successfully analyzing such maps for years. The defect does indicate, however, that the reconstruction algorithm has not used all the available information. Inclusion

of the *a priori* knowledge that a physically admissible reconstruction cannot go negative is not sufficient, in general, to permit a unique reconstruction, but this constraint does represent additional information no less valid than the actual measurements made. The difficulty, of course, is including this information in the reconstruction process.

The ME technique ensures a nonnegative reconstruction since the entropy measure H_2 diverges to $-\infty$ if more than a countable number of points in $\hat{f}$ approach zero. If the measurements were error-free, we would formulate ME reconstruction as the constrained optimization

$$\text{maximize} \int\int_D \log \hat{f}(x,y)dxdy \tag{12}$$

subject to

$$\int\int_D \hat{f}(x,y) \exp\left[-j2\pi(u_i x + v_i y)\right]dxdy = m_i, \quad i = 1,2,\cdots,M \tag{13}$$

$$\text{and } \hat{f}(x,y) \geq 0 \qquad \text{for } (x,y) \in D. \tag{14}$$

The constraints depend only on the available data so no unwarranted assumptions about unavailable measurements are made.

Physical measurements are never free from errors, and modification of (12)–(14) to reflect that uncertainty is desirable. If the data are sufficiently noisy, (12)–(14) may have no solution, i.e., there may be no nonnegative image that, when Fourier transformed, agrees exactly with all the measurements. Ables [7] has noted that information about the error statistics constitutes important *a priori* knowledge which should be included in the problem formulation.

If we suppose the errors are independent, zero mean random variables with variances σ_i^2, we could replace (13) with the single constraint

$$\sum_{i=1}^{M} \frac{1}{\sigma_i^2} |m_i - \hat{m}_i|^2 = M \tag{15}$$

where

$$\hat{m}_i = \int\int_D \hat{f}(x,y) \exp\left[-j2\pi(u_i x + v_i y)\right]dxdy. \tag{16}$$

By the Central Limit Theorem, (15) will be very nearly satisfied when $\hat{f}(x,y)$ represents the true object, provided that M is large.

No explicit solution is known for either of the constrained optimization problems just posed, and numerical solution of such problems is very difficult. Furthermore, there is still a nonzero probability that the constraints (14) and (15) are inconsistent for a particular set of measurements. Therefore, in seeking a practical algorithm, we replace the constrained maximization with the unconstrained (except for nonnegativity) maximization of

$$\int\int_D \log \hat{f}(x,y)dxdy - \lambda \sum_{i=1}^{M} \frac{1}{\sigma_i^2} |m_i - \hat{m}_i|^2 \tag{17}$$

where λ is chosen so that (15) is satisfied to sufficient accuracy. This problem has a nonnegative solution for any $\lambda > 0$. λ plays the role of a Lagrange multiplier in the maximization of (12) subject to (15). From another point of view, the maximization of (17) can be regarded as an attempt to simultaneously maximize the entropy measure and minimize the total squared residual, with the relative importance of the two being specified by the parameter λ.

Even the unconstrained maximization of (17) has no known explicit solution. We therefore seek a numerical solution, and for that purpose write a discrete version of (17) by partitioning the image $\hat{f}(x,y)$ into N pixels, each of area ΔA, and approximating the integrals by sums, giving

$$J(\hat{f}_1,\hat{f}_2,\cdots,\hat{f}_N) = \Delta A \sum_{k=1}^{N} \log \hat{f}_k - \lambda \sum_{i=1}^{M} \frac{1}{\sigma_i^2} \left| m_i - \Delta A \sum_{k=1}^{N} \hat{f}_k \exp\left[-j2\pi(u_i x_k + v_i y_k)\right] \right|^2 \tag{18}$$

where (x_k,y_k) is the location of the kth pixel. The objective function $J(\hat{f}_1,\cdots\hat{f}_N)$ is a concave function (one always underestimated by linear interpolation), and it therefore has at the most one local maximum.

Numerical Methods

Differentiating (18) with respect to the pixel values, we obtain

$$\frac{\partial J}{\partial \hat{f}_l} = \frac{\Delta A}{\hat{f}_l} + 2\Delta A\lambda \sum_{i=1}^{M} \frac{1}{\sigma_i^2} m_i^{(R)} \cos\left[2\pi(u_i x_l + v_i y_l)\right] - 2\Delta A\lambda \sum_{i=1}^{M} \frac{1}{\sigma_i^2} m_i^{(I)} \sin\left[2\pi(u_i x_l + v_i y_l)\right] - 2\Delta A^2\lambda \sum_{i=1}^{M} \frac{1}{\sigma_i^2} \sum_{k=1}^{N} \hat{f}_k \cdot \cos\left\{2\pi\left[u_i(x_l - x_k) + v_i(y_l - y_k)\right]\right\}. \tag{19}$$

By changing the order of the double summation, we arrive at the alternate form

$$\frac{\partial J}{\partial \hat{f}_l} = \frac{\Delta A}{\hat{f}_l} + \Delta A d_l - \Delta A^2 \sum_{k=1}^{N} p_{l,k} \hat{f}_k \tag{20}$$

where

$$d_l = 2\lambda \sum_{i=1}^{M} \frac{1}{\sigma_i^2} \left\{m_i^{(R)} \cos\left[2\pi(u_i x_l + v_i y_l)\right] - m_i^{(I)} \sin\left[2\pi(u_i x_l + v_i y_l)\right]\right\} \tag{21}$$

and

$$p_{l,k} = 2\lambda \sum_{i=1}^{M} \frac{1}{\sigma_i^2} \cos\left\{2\pi\left[u_i(x_l - x_k) + v_i(y_l - y_k)\right]\right\}. \tag{22}$$

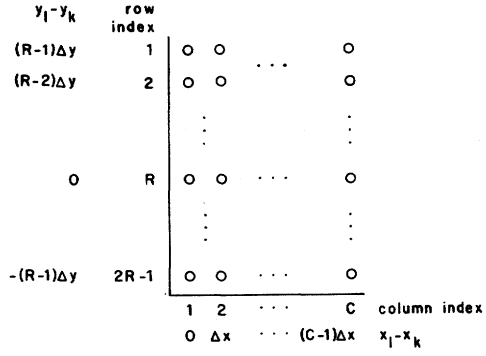


Fig. 2. The distinct values of $p_{l,k}$ are stored in an array with $2R - 1$ rows and C columns. The correspondence between array elements and pixel location differences is shown.

We can express the objective function (18) in terms of $\{d_l\}$ and $\{p_{l,k}\}$ as

$$J = \Delta A \sum_{k=1}^{N} \log \hat{f}_k + \Delta A \sum_{k=1}^{N} d_k \hat{f}_k - \frac{\Delta A^2}{2} \sum_{l=1}^{N} \sum_{k=1}^{N} p_{l,k} \hat{f}_k \hat{f}_l - \lambda \sum_{i=1}^{M} \frac{1}{\sigma_i^2} |m_i|^2. \quad (23)$$

There are a number of advantages, notation notwithstanding, to the definition of the constants in (21) and (22). Since they are independent of the pixel values, they can be computed once at the beginning of the reconstruction procedure and stored with modest memory requirements.[1] There are N numbers defined by (21). Although the double subscripted $p_{l,k}$ in (22) implies N^2 elements, there is a considerable redundancy among the values of these constants because they depend only on differences in pixel locations $(x_l - x_k)$ and $(y_l - y_k)$. This reduces the number of distinct values calculable by (22) to at most $(2R - 1)(2C - 1)$ for rectangularly arranged pixels with R rows and C columns. The actual storage required can be reduced by another factor of 2 by observing that $p_{l,k} = p_{k,l}$.

In practice, it is convenient to store the distinct values of $p_{l,k}$ in an array dimensioned as $P(2R - 1, C)$ as shown in Fig. 2. With this scheme, there are some duplicated elements in the first column of P; however, the amount of redundancy is small ($R - 1$ elements), and the total storage required for different values of $p_{l,k}$ is slightly under $2N$ elements. A particular value of $p_{l,k}$ is extracted from P by

$$p_{l,k} = P[R - \text{sgn}\,(x_l - x_k)(y_l - y_k)/\Delta y, 1 + |x_l - x_k|/\Delta x] \quad (24)$$

where row and column spacings are Δy and Δx, respectively.

[1] What constitutes modest memory requirements depends, of course, on available computing facilities. The minimum configuration anticipated includes a main memory at least several times larger than that required to store the reconstruction and auxiliary random-access storage such as a disk. With these assets, array storage proportional to the number of pixels (N) can be considered reasonable. It is usually possible to structure program flow and memory allocation so not all arrays are needed in main memory simultaneously. Algorithms demanding storage proportional to N^2 or MN are, in our view, unreasonable for most image processing problems.

Both (21) and (22) have strong physical interpretations. As shown by comparison of (11) and (21), $\{d_l\}$ is a direct transform reconstruction from the measurements with the ith scaling constant equal to $2\lambda/\sigma_i^2$. A direct transform reconstruction is equivalent to the output of a two-dimensional, linear, space-invariant filter whose input is the original object [24]. If we evaluate (21) with $m_i^{(R)} = 1$ and $m_i^{(I)} = 0$, we find the point spread function of the filter. This permits us to identify $p_{l,k}$ as the filter response, in the lth pixel, to a point source in the kth pixel.

From (20) we see that the gradient of the ME reconstruction objective function (18) is completely characterized by these direct transform reconstruction parameters. That the nonlinear ME reconstruction procedure should have such ties with the linear direct transform techniques is both interesting and fortunate. By exploiting our understanding of linear image processing operations, we might hope to gain insight on the nonlinearities of ME reconstruction.

By recognizing the summations

$$\sum_{k=1}^{N} p_{l,k} \hat{f}_k \qquad l = 1,2,\cdots,N \quad (25)$$

as *two-dimensional* convolution of the reconstruction and the point spread function, we can use fast Fourier transform (FFT) techniques for efficient calculation of the objective function in the form (23) and its gradient, whose elements are given by (20). To identify the convolution form of (25), we temporarily drop the one-dimensional ordering of pixels.

Let the center position of the kth pixel be

$$(x_k, y_k) = (c_k \Delta x, r_k \Delta y) \quad (26)$$

where $1 \le c_k \le C$ and $1 \le r_k \le R$. The parameters r_k and c_k indicate row and column locations, respectively, within the pixel matrix. We avoid the imposition of a particular data structure (e.g., row-sequential or column-sequential) for generality and notational brevity. Since $p_{l,k}$ is a function of $(x_l - x_k)$ and $(y_l - y_k)$, we can write

$$p_{l,k} = p(x_l - x_k, y_l - y_k). \quad (27)$$

From (25)–(27) we obtain

$$\sum_{k=1}^{N} p_{l,k}\hat{f}_k = \sum_{k=1}^{N} \cdot p[(c_l - c_k)\Delta x,(r_l - r_k)\Delta y]\hat{f}(c_k\Delta x,r_k\Delta y). \quad (28)$$

Since the sum ranges over all pixels, we have

$$\sum_{k=1}^{N} p_{l,k}\hat{f}_k = \sum_{c_k=1}^{C}\sum_{r_k=1}^{R} \cdot p[(c_l - c_k)\Delta x,(r_l - r_k)\Delta y]\hat{f}(c_k\Delta x,r_k\Delta y) \quad (29)$$

which is the form of two-dimensional convolution.

We will not belabor the details of high-speed convolution via FFT calculations; the interested reader should consult [25], [26]. Of interest, though, are the storage requirements and the speed advantage of FFT use. We can only give rough guidelines here since both aspects depend on the computer installation and the FFT realization.

The minimum transform size is $(2C - 1)$ by $(2R - 1)$, and these figures will usually be adjusted upward to the nearest power of 2. The burden of such a transform increases roughly as $4RC \log_2 (4RC) \simeq 4N \log_2 N$. We can precompute and store the FFT of the point spread function array; this gives a discrete approximation to the direct reconstruction filter transfer function. The number of distinct values in this discrete transfer function and in the point spread function array can be shown to be equal so no extra storage is involved.

Each evaluation of (25) consequently requires only two FFT's, one of the reconstruction, properly augmented with zeros, and the other an inverse FFT which yields the desired convolution values. Hence, the burden of FFT evaluation of (25) is about $8N \log_2 N$ as compared with N^2 for routine programming of the convolution sum. For a 128 by 128 picture ($N = 16\,384$), the speedup factor is on the order of 100.

The reduction of the problem to the unconstrained optimization of (18) required the introduction of the Lagrange multiplier λ, whose value must be determined separately. In general it may be necessary to compute a solution for several trial values of λ in order to find the one which allows constraint (15) to be satisfied with sufficient accuracy. Larger values of λ cause the total squared residual [left-hand side of (15)] of the solution to be smaller.

We have found the following argument useful for determining an initial estimate for λ. If there were only one measurement m_1 (with variance σ_1^2) at $(u_1,v_1) = (0,0)$, then m_1 is an estimate of the total object intensity. In this case, the ME reconstruction is uniformly gray with brightness $\hat{f}$, where, from (17), $\hat{f}$ is the number that maximizes

$$J = A \log \hat{f} - \lambda(m_1 - A\hat{f})^2/\sigma_1^2 \quad (30)$$

where $A = N\Delta A$ is the total area of the object. The necessary condition

$$\frac{\partial J}{\partial \hat{f}} = \frac{A}{\hat{f}} + 2\lambda A(m_1 - A\hat{f})/\sigma_1^2 = 0 \quad (31)$$

is satisfied when

$$\lambda = \frac{-\sigma_1^2}{2\hat{f}(m_1 - A\hat{f})} = \frac{A\sigma_1^2}{2(m_1 + \delta)\delta} \quad (32)$$

where $\delta = A\hat{f} - m_1$ is the discrepancy between the reconstruction and the measurement. Since the variance of the intensity measurement m_1 is known to be σ_1^2, setting $\delta = \sigma_1$ seems reasonable; this gives

$$\lambda = \frac{A\sigma_1^2}{2(m_1 + \sigma_1)\sigma_1} = \frac{A}{2(1 + m_1/\sigma_1)}. \quad (33)$$

Using λ from (33) in the optimization of (18) does not guarantee that the solution will satisfy constraint (15) when there is more than one measurement, but the approximation has been adequate for our work thus far. If necessary, λ can be increased to obtain closer agreement with the measurements. It should be realized, however, that there may be no λ for which (15) is satisfied; e.g., if one or more measurements are unusually noisy, or if the variances $\{\sigma_i^2\}$ assumed by the (optimistic) experimenter are unrealistically small.

Although ME reconstruction is a nonlinear procedure, it does maintain linearity with respect to measurement scale if λ is chosen according to (33). The loss of this valuable property would be disturbing: it would imply a nontrivial sensitivity of the reconstruction to the physical units of the measurements. To demonstrate scaling linearity, we consider a set of measurements $\{m_i\}$ and variances $\{\sigma_i^2\}$ leading to the ME reconstruction $\{\hat{f}_i\}$. Of necessity, all partial derivatives (20) of the objective function must vanish. We now scale the measurements by $m_i' = tm_i$; where $t > 0$. This scale change induces, from (33), (21) and (22)

$$(\sigma_i^2)' = t^2\sigma_i^2, \quad (34)$$

$$\lambda' = \lambda, \quad (35)$$

$$d_l' = d_l/t, \quad (36)$$

$$p_{l,k}' = p_{l,k}/t^2, \quad (37)$$

and it is seen by substitution of (36) and (37) into (20) that $\{t\hat{f}_i\}$ is the ME reconstruction in the new measurement system.

Optimization Algorithms

Lacking an explicit solution to the nonlinear maximization of (18), we carry out an iterative search for the maximum. The general problem of such nonlinear optimization has been well studied (see [27]–[29] and others), and many algorithms are available. As mentioned earlier, the objective function has a unique local maximum; as a result, convergence to that point can be guaranteed for most reasonable search algorithms. The process is hampered, however, by the large dimensionality of typical imaging problems: there are as many independent variables as there are pixels. This causes practical difficulties

due to slow convergence, large storage requirements, and limited numerical precision.

We consider algorithms in which a sequence of one-dimensional searches is executed in the N-dimensional solution space. These can be classified according to the amount of local information about the objective function that is used in determining the search direction. Zero-, first-, and second-order methods are those which use evaluations of, respectively, the function itself; the function and its gradient; and the function, gradient, and Hessian matrix of second partial derivatives. Second-order methods require storage proportional to N^2, which is so large in our application that we will not consider them further. Quasi second-order methods, e.g., [30], which build up an estimate of the Hessian from function and gradient evaluations, have the same storage problem.

On the other hand, the storage required for both zero- and first-order methods is proportional to N, so the choice between them depends mainly on computational considerations. Equations (20) and (23) show that the major burden of computing the objective function and its derivatives is due to the convolution sum (25). Once this has been computed, either by routine programming or FFT techniques, the number of extra operations needed for evaluation of the objective function and the complete gradient is proportional to N. Consequently, exact calculation of the gradient each time the objective function is evaluated is almost free, and this discourages use of search algorithms that do not employ derivatives or rely on finite difference approximations. We are thus led to first-order procedures such as steepest ascent or the conjugate gradient method [31], [27].

We have used both of these methods with some success. In steepest ascent, the search is in the direction of the objective function gradient, ∇J; in the conjugate gradient method, the search direction is

$$S = \nabla J - S' \, |\nabla J|/|\nabla J'| \tag{38}$$

where primes refer to the values in the previous iteration. The conjugate gradient method requires more storage since it retains the previous search direction as well as the current gradient. Asymptotically, steepest ascent converges linearly and conjugate gradient quadratically, making the latter apparently more powerful; however, with the large dimensionality of the present objective function, most of the computation takes place before the asymptotic rate is achieved, and the superiority of either method has yet to be demonstrated.

In any such methods, the one-dimensional search procedure is quite important. A very accurate search requires many function and gradient evaluations, which is expensive, but the use of approximate searches may require more iterations for convergence. We have used the following compromise: take a step of predetermined length in the search direction, evaluate the function and gradient there, and estimate the location of the one-dimensional search maximum by quadratic or cubic interpolation.[2]

A practical problem becomes apparent when some pixel values are close to zero. The objective function is highly nonlinear in this region, and its value is undefined for negative pixel values. Convergence is slowed since steps that would further reduce these already small values must be short to maintain nonnegativity. This problem can be alleviated by the exponential transformation $f_i = \exp(g_i)$, $i = 1,2,\cdots,N$. The objective function is now defined for both positive and negative values of the new independent variables g_i.

Another modification, which has proven more useful than the exponential transformation, is to deflect the search direction so pixel values below a certain cutoff, say 1 percent of the maximum pixel value, are not changed unless they would be increased by a move in the original search direction. The justification for this is that relatively small pixel values have little effect on either reconstruction-measurement consistency or the appearance of the image. An algorithm that uses deflected gradients for most of the iterations and occasionally perturbs the reconstruction in the true gradient direction has been found to converge faster than exact steepest ascent.

Another procedure showing potential for ME reconstruction in some applications involves univariate moves. If only one pixel brightness, $\hat{f}_l$, is changed at a time, the optimal change can be found analytically by setting the corresponding partial derivative to zero. From (20) we find

$$\hat{f}_l = \frac{d_l - \Delta A \sum_{k \neq l} p_{l,k}\hat{f}_k + \sqrt{\left(d_l - \Delta A \sum_{k \neq l} p_{l,k}\hat{f}_k\right)^2 + 4\Delta A p_{l,l}}}{2\Delta A p_{l,l}} \tag{39}$$

to be the value that maximizes the objective function while holding all other variables constant. By adjusting each pixel separately, we can thus avoid the need for a search to find the optimal step size. Another advantage is that (39) automatically satisfies the nonnegativity constraint since λ is positive.

There are two disadvantages to the univariate approach. First, the sequential handling of pixel values induces patterns, characteristic of the order in which variables are adjusted, in the reconstruction during the early iterations. This is particularly apparent when working with images

[2] One can also consider moves that are a constant fraction of the gradient vector, as in [3]. This simplifies programming and reduces the number of function evaluations in each iteration; however, the number of iterations will generally be increased and convergence cannot be guaranteed unless the fraction is chosen carefully. Also, it is necessary to consider the action to be taken should such a fixed-fraction move drive a pixel value negative.

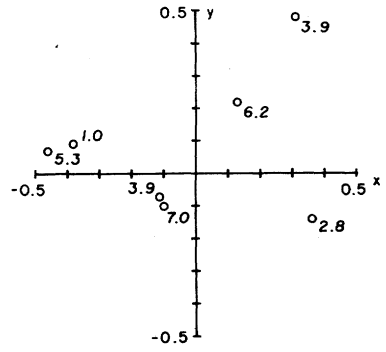


Fig. 3. A phantom consisting of seven point sources. Source locations are shown by circles, and relative source strengths appear in italics.

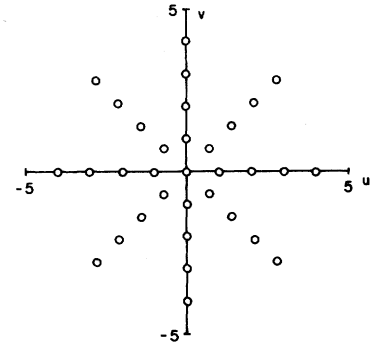


Fig. 4. Measurement coverage in the spatial frequency domain for the seven point source phantom.

possessing a high degree of symmetry. Gradient methods, moving all variables simultaneously, tend to maintain any symmetry present and yield better reconstructions for the first few iterations. Randomization of the order in which univariate methods update the variables is helpful, although true randomization increases the program complexity. A reasonable compromise is to define several deterministic, but different, adjustment orders, e.g., first across rows and then down columns, so patterns break up sooner.

The second disadvantage is more serious. Each time a variable is changed, the convolution sum (25) is altered. With gradient methods, one use of the FFT for high-speed convolution permits adjustment of all pixel values. It is inefficient to compute the complete convolution for each single variable step; consequently, an exact univariate search requires an effort proportional to N^2 to update all pixel values. While this makes univariate optimization unattractive in large problems, it may not be a handicap in all reconstruction applications. Since the gradient is not stored, less memory is needed, and the difficulty of reconstruction on moderate-sized computers is eased. Also, if the number of pixels is relatively small, FFT techniques may offer no computational advantage. Even in cases where FFT use could be profitable, the increased program complexity and additional storage requirements may restrict its application.

EXAMPLES

The ME procedure has been used with encouraging results for reconstruction from both phantom (artificial) data and actual interferometer measurements. Programming was done in Fortran on a Hewlett–Packard 2116B computer with 16K 16 bit words of main memory and 1.2M words of auxiliary disk storage. It is convenient to divide the reconstruction task into three independent programs. The first program inputs the Fourier transform samples, calculates the direct transform parameters defined in (21) and (22), and stores the parameters on disk. The second program uses these parameters for iterative maximization of the ME objective function. This program can be terminated after any number of iterations; the last reconstruction estimate is saved in a disk file. A third program is responsible for display and for checking the reconstruction consistency with the individual measurements. If the agreement is unsatisfactory, the image is reinput by the second program for continuing iteration. Adjustment of the parameter λ can easily be handled in this program if necessary.

Reconstructions have been computed on a 21 by 21 image grid, resulting in 441 independent variables in the optimization. The examples shown here have been computed with the univariate search algorithm discussed above. The exact optimizing step for each variable is calculated from (39). The values λ/σ_i^2 needed in (21) and (22) were calculated according to (33) with $\sigma_i^2 \equiv \sigma_1^2$ and $m_1/\sigma_1 = 10$.

A useful quantitative measure of the discrepancy between a reconstruction and the measurements is the factor

$$\Delta = \sqrt{\frac{\sum_{i=1}^{M} \left| m_i - \Delta A \sum_{k=1}^{N} \hat{f}_k \exp\left[-j2\pi(u_i x_k + v_i y_k)\right] \right|^2}{\sum_{i=1}^{M} |m_i|^2}}, \tag{40}$$

that is, the ratio of the rms error to the rms amplitude of the data. This figure is quoted for the reconstructions shown here. The discrepancy of direct transform reconstructions is not zero because the image brightness is assumed to be zero outside the image domain. In these examples, we have multiplicatively scaled each direct transform reconstruction so the reconstruction intensity $\Delta A \sum_{k=1}^{N} \hat{f}_k$ agrees exactly with the measured value.

The first example uses a phantom composed of seven point sources whose strengths and positions, shown in Fig. 3, were chosen randomly. The Fourier transform of this phantom was evaluated at the locations shown in Fig. 4, and these data was used for reconstruction. Contour plots of the direct transform reconstruction, with areas of negative brightness shaded, and of the ME reconstruction after 10 iterations—that is, 10 adjustments to each pixel brightness—are compared in Fig. 5. The initial image estimate for the ME algorithm is taken to be the direct transform reconstruction with negative pixel values set to zero. We do not claim that the ME procedure has con-

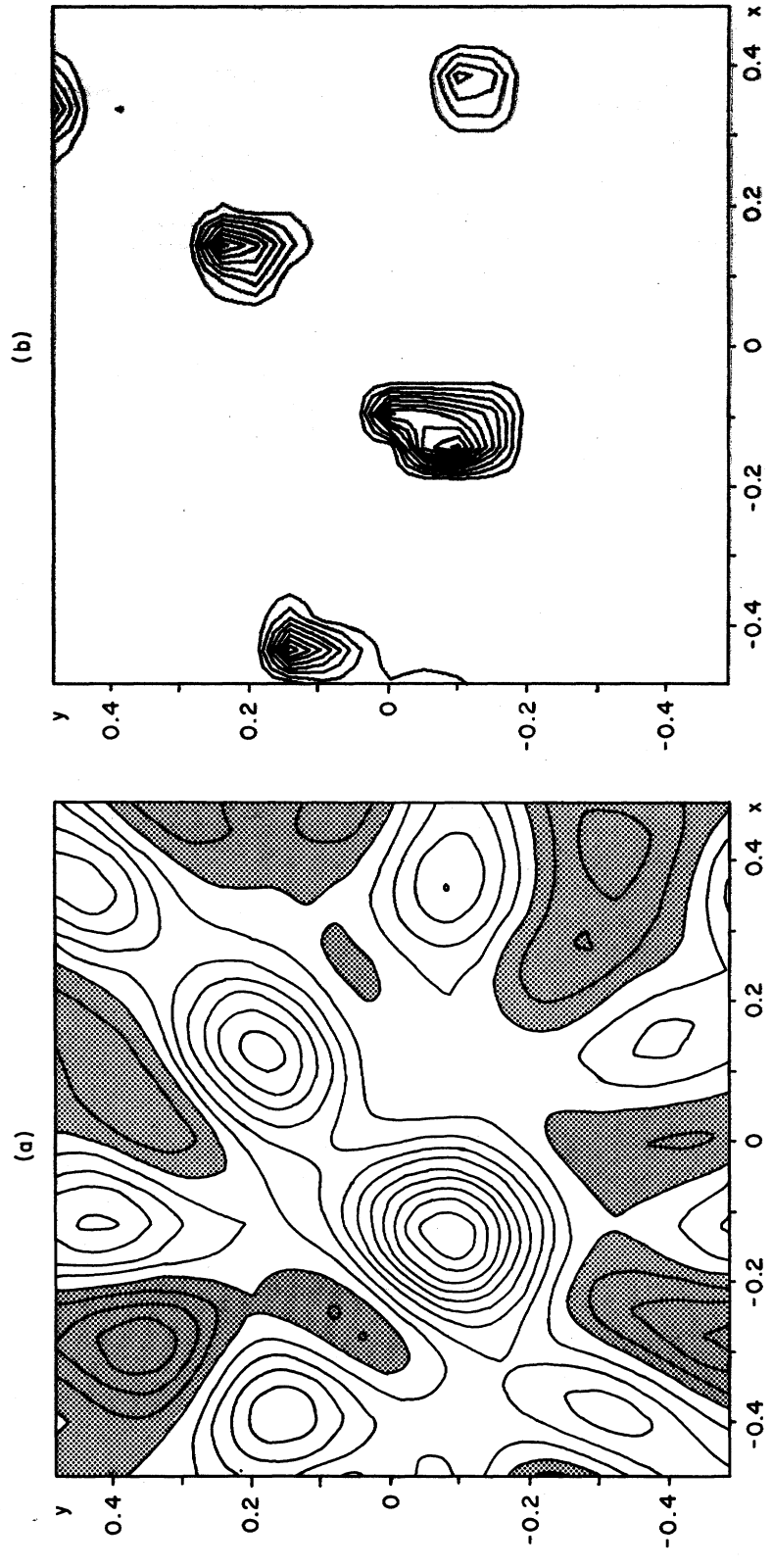


Fig. 5. Reconstructions of the seven point source phantom (a) by direct Fourier transformation (Δ = 19.0 percent), and (b) after 10 iterations of the maximum entropy technique (Δ = 10.8 percent). In these and all other contour plots, the contour interval is 10 percent of the maximum brightness.

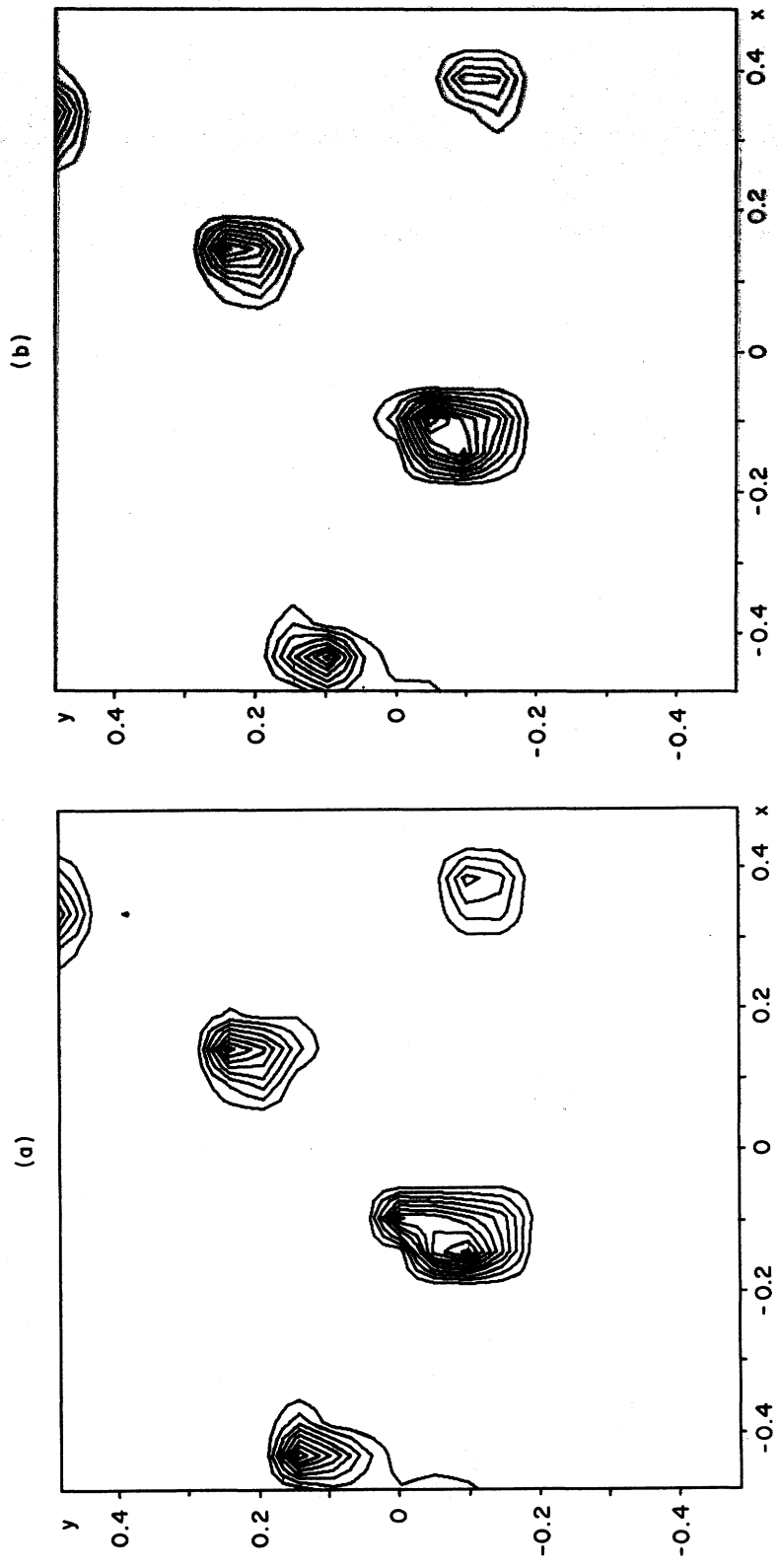


Fig. 6. Maximum entropy reconstructions of the seven point source phantom after (a) 10 (Δ = 10.8 percent) and (b) 20 (Δ = 10.0 percent) iterations.

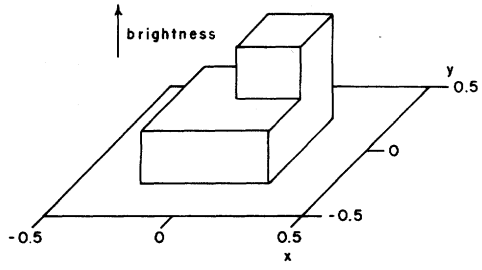


Fig. 7. A piecewise-constant "stacked blocks" phantom. The image brightness is zero for $|x| > 0.25$, $|y| > 0.25$.

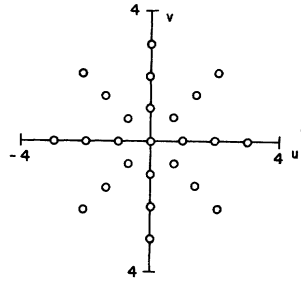


Fig. 8. Spatial frequency domain coverage for reconstruction of the stacked blocks phantom.

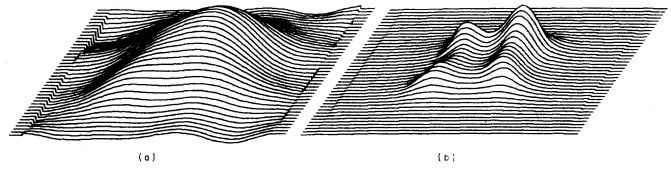


Fig. 9. Reconstruction of the "stacked blocks" phantom (a) by direct Fourier transformation (Δ = 14.1 percent) and (b) after 16 iterations of the maximum entropy procedure (Δ = 10.1 percent). A cubic spline was used for interpolation along each row before plotting.

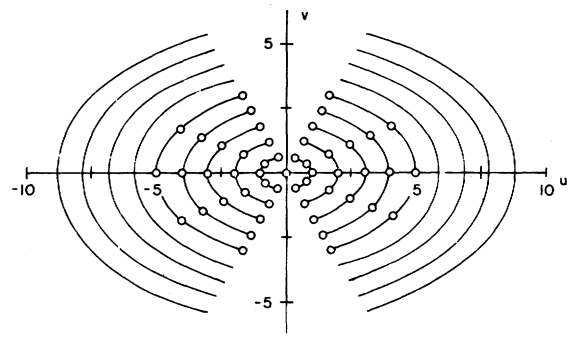


Fig. 10. Interferometer measurements for an observation of the radio source Cygnus A (3C405) lie on nine concentric ellipses. Measurements used by the maximum entropy procedure are circled.

verged in these few iterations; however, it has settled down enough that iteration-to-iteration changes in the reconstruction are slight. In this example, the rms pixel value change in going from iteration 9 to iteration 10 was only 4 percent of the rms value of the reconstruction. The effect of increasing the number of iterations is demonstrated in Fig. 6 where the ME reconstruction from the same data after 20 iterations is shown.

It is seen from Fig. 5 that the ME reconstruction is of higher resolution than the direct transform image although there are two pairs of sources—at approximately (−0.4, 0.1) and (−0.1, −0.1)—that are not resolved by either technique. The ME procedure eliminates much of the background fluctuation that characterizes direct transform reconstructions. It is interesting to note that in the direct transform reconstruction, the peak at (−0.1, 0.4) appears with a higher amplitude than the peak at (0.4, −0.1) even though there is no source at the former location but there is one at the latter. The fictitious peak disappears (at the 10 percent level) from the maximum entropy reconstruction.

Point source images are ideal candidates for ME reconstruction because they provide the most opportunity for resolution gain over direct procedures; however, improved reconstruction quality is also obtained for the high-contrast, piecewise-constant phantom shown in Fig. 7. Its Fourier transform is calculated at points indicated in Fig. 8, and a comparison of reconstructions is given in Fig. 9. The improved resolution of the ME reconstruction is manifest in steeper edges and a stronger indication of the extra brightness in the back right quadrant.

The final example uses real data taken with the Stanford University five-element interferometer during an observation of the well-mapped radio double Cygnus A (3C405). The instrument, described in detail in [32], operates at a wavelength of 2.8 cm and provides measurements of the two-dimensional Fourier transform of the source under investigation. These measurements lie on nine concentric ellipses as shown in Fig. 10. Because of the way in which data are accumulated, measurement coverage is essentially continuous along each arc.

Only a fraction of the available data has been input to the ME algorithm, but the direct transform program has been allowed to use all the measurements. This is indicated in Fig. 10, where it is seen that the data given to the ME procedure have been sampled at approximately 30° intervals in angle and, more importantly, include none of the data from the outer four ellipses. As a consequence of the latter action, the natural resolution of the data provided to the ME algorithm is poorer, by almost a factor of 2, than the resolution of the data given to the direct transform procedure.

Reconstructions using the two methods are displayed in Fig. 11. The east–west peak widths in the ME reconstruction are slightly greater than in the direct transform reconstruction, although the increase is not as great as might be expected considering the differences in the data given to the two routines. The north–south widths are smaller in the ME reconstruction, resulting in brightness contours that are more nearly circular. Once again, the ME technique has suppressed much of the background activity apparent in the direct transform reconstruction.

Conclusion

Although we have concentrated most of our attention on the imaging problem of radio astronomy, maximum entropy reconstruction has potential application in other fields concerned with reconstruction from incomplete data,

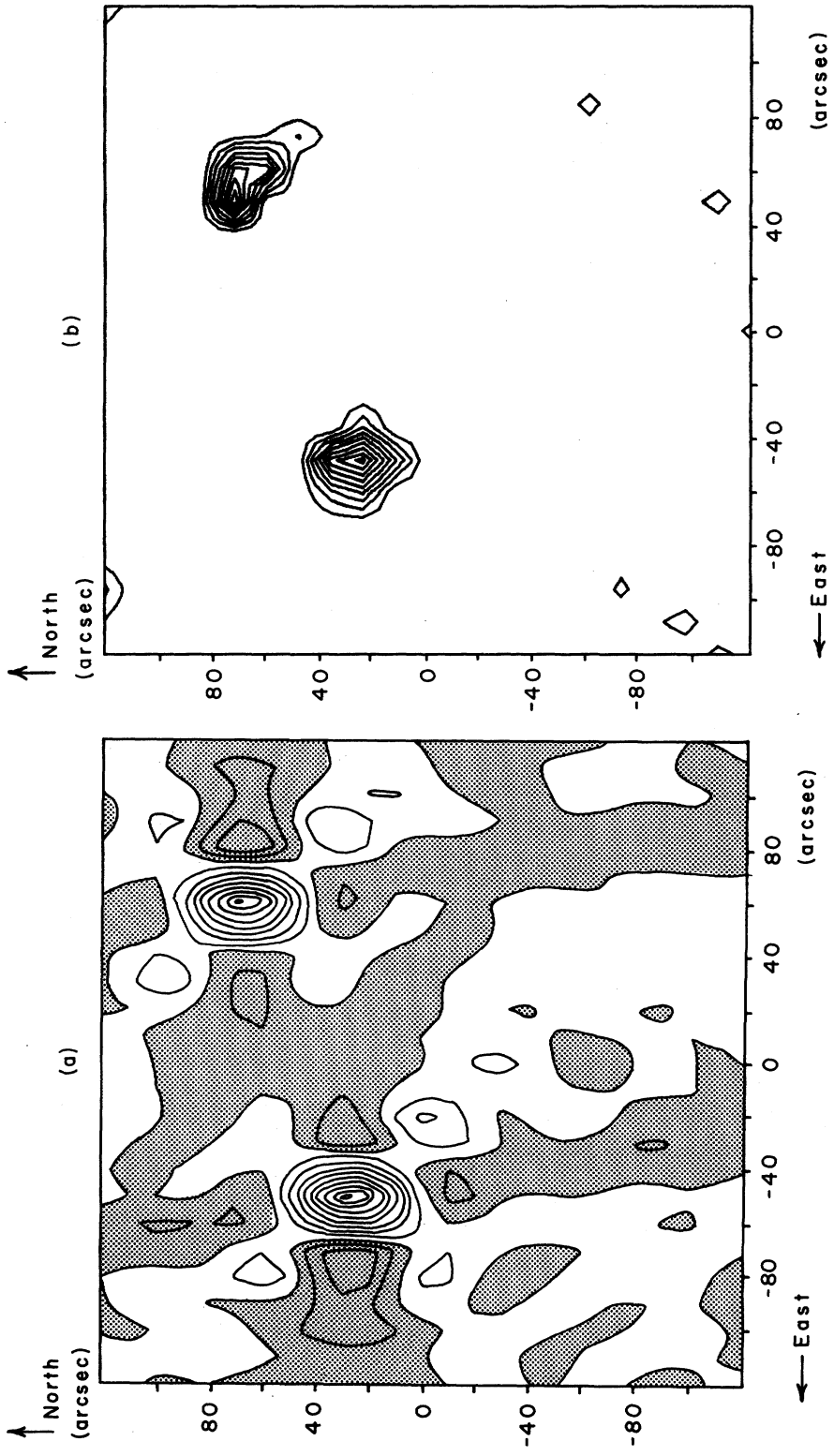


Fig. 11. Reconstruction of Cygnus A by (a) direct Fourier transformation of all measurements (Δ unknown), and (b) after 23 iterations of the maximum entropy technique using the limited measurements indicated in Fig. 10 (Δ = 15.7 percent).

e.g., electron microscopy. Even in radiography, where measurements are usually plentiful, there are situations in which data are missing; for example, when the number of projections is intentionally reduced to limit X-ray exposure to the patient or when imaging moving objects. As discussed in the Introduction, the essential difference between reconstruction in these varied disciplines is in the form of the kernel in the measurement equation (1). Incorporation of the appropriate kernel into the maximum entropy objective function (18) adapts the method to new specialties.

References

[1] B. R. Hunt, "Digital image processing," *Proc. IEEE,* vol. 63, pp. 693–708, 1975.

[2] A. Macovski, R. E. Alvarez, J. L.-H. Chan, and J. P. Stonestrom, "Correction for spectral shift-artifacts in X-ray computerized tomography," presented at Image Processing for 2-D and 3-D Reconstruction from Projections: Theory and Practice in Medicine and the Physical Sciences Meeting, Stanford, CA, Aug. 4–7, 1975.

[3] B. R. Hunt, "Bayesian methods of nonlinear digital image restoration," *IEEE Trans. Comput.,* to be published.

[4] G. T. Herman and A. Lent, "A computer implementation of a Bayesian analysis of image reconstruction," State Univ. of New York, Buffalo, NY, Tech. Rep. 91, 1974.

[5] B. R. Hunt and T. M. Cannon, "Nonstationary assumptions for Gaussian models of images," *IEEE Trans. Syst., Man, and Cybern.,* to be published.

[6] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Syst. Sci. Cybern.,* vol. SSC-4, pp. 227–241, 1968.

[7] J. G. Ables, "Notes on maximum entropy spectral analysis," *Astron. Astrophy. Suppl.,* vol. 15, 1974.

[8] D. E. Smylie, G. K. C. Clarke, and T. J. Ulrych, "Analysis of the irregularities in the earth's rotation," in *Methods in Computational Physics,* vol. 13, B. Alder, S. Feinbach, and B. A. Bolt, Eds. New York: Academic Press, 1973.

[9] L. Brillouin, *Science and Information Theory.* New York: Academic Press, 1956.

[10] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication.* Urbana, IL: University of Illinois Press, 1949.

[11] B. R. Frieden, "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Amer.,* vol. 62, no. 4, pp. 511–518, 1972.

[12] R. Gordon and G. T. Herman, "Reconstruction of pictures from their projections," *Quarterly Bull. Center for Theor. Biol.,* vol. 4, pp. 71–151, 1971.

[13] J. P. Burg, "Maximum entropy spectral analysis," presented at the 37th Meeting Soc. of Exploration Geophysicists, Oklahoma City, OK, 1967.

[14] J. P. Burg, "A new analysis technique for time series data," presented at NATO Adv. Study Inst. on Signal Processing with Emphasis on Underwater Acoustics, 1968.

[15] M. S. Bartlett, *An Introduction to Stochastic Processes.* London, England: Cambridge University Press, 1966.

[16] N. Levinson, "The Wiener rms (root mean square) error criterion in filter design and prediction," *J. Math. and Phys.,* vol. 25, pp. 261–278, 1947.

[17] T. J. Ulrych and O. G. Jensen, "Cross-spectral analysis using maximum entropy," *Geophysics,* vol. 39, pp. 353–354, 1974.

[18] A. Van den Bos, "Alternative interpretation of maximum entropy spectral analysis," *IEEE Trans. Inform. Theory,* vol. 17, pp. 493–494, 1971.

[19] J. E. B. Ponsonby, "An entropy measure for partially polarized radiation and its application to estimating radio sky polarization distributions from incomplete 'aperture synthesis' data by the maximum entropy method," *Mon. Not. R. Astr. Soc.,* vol. 163, pp. 369–380, 1973.

[20] M. Born and E. Wolf, *Principles of Optics.* Oxford: Pergamon Press, 1970.

[21] S. J. Wernecke, "Maximum entropy image reconstruction," presented at Image Processing for 2-D and 3-D Reconstruction from Projections: Theory and Practice in Medicine and the Physical Sciences Meeting, Stanford, CA, Aug. 1975.

[22] G. W. Swenson and N. C. Mathur, "The interferometer in radio astronomy," *Proc. IEEE,* vol. 56, p. 2114, 1968.

[23] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra.* New York: Dover, 1958.

[24] R. N. Bracewell and A. R. Thompson, "The main beam and ringlobes of an east-west rotation-synthesis array," *Astrophys. J.,* vol. 182, pp. 77–94, 1973.

[25] R. M. Mersereau and D. E. Dudgeon, "Two-dimensional digital filtering," *Proc. IEEE,* vol. 63, pp. 610–623, 1975.

[26] B. R. Hunt, "Minimizing the computation time for using the technique of sectioning for digital filtering of pictures," *IEEE Trans. Comput.,* vol. C-21, pp. 1219–1222, 1972.

[27] R. L. Fox, *Optimization Methods for Engineering Design.* Reading, MA: Addison-Wesley, 1971.

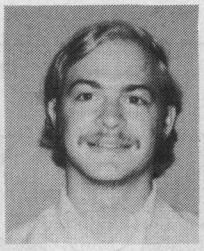
[28] W. Murray, Ed., *Numerical Methods for Unconstrained Optimization.* New York: Academic Press, 1972.

[29] E. J. Beltrami, *An Algorithmic Approach to Nonlinear Analysis and Optimization.* New York: Academic Press, 1970.

[30] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *Comput. J.,* vol. 6, pp. 163–168, 1963.

[31] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *Comput. J.,* vol. 7, pp. 149–154, 1964.

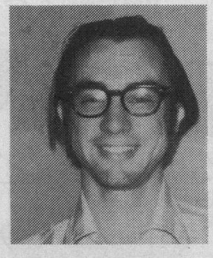
[32] R. N. Bracewell, R. S. Colvin, L. R. D'Addario, C. J. Grebenkemper, K. M. Price, and A. R. Thompson, "The Stanford five-element radio telescope," *Proc. IEEE,* vol. 61, pp. 1249–1257, Sept. 1973.



Stephen J. Wernecke (S'72) was born in Evanston, IL on October 20, 1950. He received the B.S. (with distinction), M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1972, 1974, and 1976, respectively.

From 1972 to 1976, he was a Research Assistant with the Stanford Electronic Laboratories. During that time he also held appointments as Teaching Fellow and Acting Instructor of Electrical Engineering at Stanford. He is currently a Research Associate at Stanford where his interests include image reconstruction, digital signal processing and statistical estimation theory. In 1972 he received the F. E. Terman Award for Scholastic Achievement.

Dr. Wernecke is a member of Phi Beta Kappa, Tau Beta Pi, and the Optical Society of America.



Larry R. D'Addario (S'70–M'74) was born in East Orange, NJ, on December 26, 1946. He received the S.B. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1968, and the M.S. and Ph.D. degrees, both in electrical engineering, from Stanford University, Stanford, CA, in 1969 and 1974, respectively.

From 1971 through August 1974, he was with the Radio Astronomy Institute, Stanford University, where he developed the calibration methods and major portions of the software for the five-element synthesis telescope. Since September 1974, he has been with the National Radio Astronomy Observatory, Charlottesville, VA, engaged in basic research in data processing for radio telescopes. In January 1976, he transferred to the Observatory's Very Large Array Project, Socorro, NM, where he is responsible for evaluating electronic systems tests.

Dr. D'Addario is a member of Tau Beta Pi, Sigma Xi, and Eta Kappa Nu, and of Commission V of URSI.