25

Reprinted from Journal of the Optical Society of America A, Vol. 4, page 112, January 1987

Estimation of continuous object distributions from limited Fourier magnitude measurements

Charles L. Byrne* and Michael A. Fiddy†

The Catholic University of America, Washington, D.C. 20064

Received May 21, 1986; accepted July 29, 1986

From finite complex spectral data one can construct a continuous object with a given support that is consistent with the data. Given Fourier magnitude data only, one can choose the phases arbitrarily in the above construction. The energy in the extrapolated spectrum is phase dependent and provides a cost function to be used in phase retrieval. The minimization process is performed iteratively, using an algorithm that can be viewed as a combination of Gerchberg–Papoulis and Fienup error reduction.

INTRODUCTION

The problem of recovering a signal or images from the Fourier-transform magnitude (power spectrum) has an important role in many applications and has been the subject of many reviews (e.g., Refs. 1–7). The focus of the discussion has been largely on questions of uniqueness when noise-free continuous data are available from either continuous or discrete objects of compact support. The analytic properties of the Fourier transform of such an object provide a convenient model for determining the existence of Fourier phase ambiguities by means of non-self-conjugate factor flipping. For one-dimensional objects, many complex factors and thus phase ambiguities can be expected to be present. For two- or higher-dimensional discrete objects, it is highly likely that no factors occur that generate nontrivial phase ambiguities.[8] For continuous objects in more than one dimension, there exists a much richer variety of prime factors than in one dimension,[9] but their probability of occurring is small, unless the object has some symmetry properties.[10] If the Fourier transform of the object has only one factor, i.e., is irreducible, then there is a unique phase to be associated with the Fourier magnitude,[5,11] and thus the object can be recovered. It follows from the work of Hayes and McClellan[8] that reducibility of the Fourier function is an unstable property. For discrete objects the set of reducible Fourier transforms is of measure zero, but one could imagine a sparse set, in the sense of a collection of isolated points, a curve, or a set that is dense in the manner of the rationals in the reals. Sanz and Huang[11] showed that the set, in fact, resembles a curve (a surface in four-space, for example) and a small perturbation is likely to result in an irreducible neighbor's being found. A similar but weaker result exists for continuous objects. One consequence of this, however, is that with noise on the Fourier magnitude data (or, more exactly, the Fourier intensity data) this function is highly likely to be irreducible, thus bringing into question the very existence of a solution at all, if the intensity does not factor into the Fourier transform of the object function F multiplied by its complex conjugate F^*.

In practice, of course, not only will there be noise on the Fourier magnitude data but only a finite sequence of samples of this function will be measured. An immediate consequence of having finite data, even complex, noise-free Fourier values, is that an infinite ambiguity arises for the object function. Iterative and noniterative extrapolation methods exist that provide estimates of the object function that are both consistent with the data and limited to the object support[12,13]; random phase values could be assigned to the Fourier magnitudes and such an estimator computed.[14] It is therefore necessary to adopt an estimation approach to the practical phase-retrieval problem and to seek a method that is in some sense optimal. The approach that we adopt is based on minimizing the energy required to obtain a data- and a support-consistent estimation of the object, using this energy as a cost function to guide us to an appropriate choice of phases.

MINIMUM-ENERGY EXTRAPOLATION AND PHASE RETRIEVAL

Let $f(t)$, $|t| \leq \pi$ be a compactly supported object function with spectrum $F(x)$ given by

$$F(x) = \int_{-\pi}^{\pi} f(t)\exp(ixt)\mathrm{d}t/2\pi. \tag{1}$$

We limit ourselves to one-dimensional objects only to keep the notation simple; all of what we say applies equally to higher-dimensional cases, and we shall discuss particularly the two-dimensional case when we consider algorithm implementation. Because the support is $[-\pi, \pi]$ the unit sampled values of F suffice to reconstruct $f(t)$:

$$f(t) = \sum_{n=-\infty}^{\infty} F(n)\exp(-int), \qquad |t| \leq \pi. \tag{2}$$

Given the limited complex data $F(n)$, $|n| \leq N$, the minimum energy estimate of the object $f(t)$ is the discrete Fourier transform (DFT) of the data,

$$\mathrm{DFT}(t) = \sum_{n=-N}^{N} F(n)\exp(-int), \qquad |t| \leq \pi. \tag{3}$$

This estimate may or may not be satisfactory, depending on the amount of detail in $f(t)$ to be recovered and the size of N.

0740-3232/87/010112-06$02.00

If the actual support of $f(t)$ is $V = [-v, v]$ for some $v < \pi$, then the DFT may fail to recover f over V adequately. This follows because N is finite, and the DFT is a minimum-energy estimate for f over $|t| \leq \pi$; hence it wastes algebraic freedom by describing the zero region outside V.

For each fixed N there is always a positive lower bound to the amount of energy that the DFT must have outside V. With $\mathbf{S}$ the $2N + 1$ by $2N + 1$ matrix with values $S_{m,n} = \sin[(m - n)v]/\pi(m - n)$, it is easily seen that, with $\mathbf{F} = [F(-N), \ldots, F(N)]^T$, and † denoting conjugate transpose,

$$\int_{-v}^{v} |\mathrm{DFT}(t)|^2 \mathrm{d}t = \mathbf{F}^\dagger \mathbf{S} \mathbf{F}. \tag{4}$$

Therefore the proportion of energy outside V must be

$$1 - \mathbf{F}^\dagger \mathbf{S} \mathbf{F} / \mathbf{F}^\dagger \mathbf{F} \geq 1 - \lambda_{\max} > 0, \tag{5}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{S}$. As N increases, this value approaches 1, and so the lower bound goes to zero. For moderate N, however, the DFT can be unsatisfactory. It is interesting to note that the lower bound is attained only when $\mathbf{F}$ is an eigenvector of $\mathbf{S}$ associated with $\lambda_{\max}$, which means that the DFT is maximally concentrated on V but with very little structure within V. If the object $f(t)$ is quite structured within V, we would not expect the lower bound to be attained, so structure within V is linked to an increase in energy outside V for the DFT.

An alternative estimation procedure that provides a data-consistent function that can be made to vanish off V is the the following estimator, referred to as the PDFT estimate[15,16] since it has a somewhat similar form to the DFT estimate of Eq. (3), apart from the weighting factor or previous estimate $p(t)$. When $p(t)$ tends to unity over $|t| \leq \pi$, then PDFT(t) tends to DFT(t):

$$\mathrm{PDFT}(t) = p(t) \sum_{n=-N}^{N} a_n \exp(-int), \qquad |t| \leq \pi, \tag{6}$$

where $p(t) \geq 0$ is a previous profile function and the a_n are chosen to make PDFT(t) data consistent. Denoting by $P(n)$ the values of the spectrum of $p(t)$, letting $\mathbf{P}$ be the $(2N + 1)$ by $(2N + 1)$ matrix with entries $\mathbf{P}_{m,n} = P(m - n)$, and letting $\mathbf{a} = (a_{-N}, \ldots, a_N)^T$, the data-consistency condition becomes $\mathbf{F} = \mathbf{Pa}$. If $p(t) = 1$, $|t| \leq v$, 0 elsewhere, then $\mathbf{P} = \mathbf{S}$; this choice can be ill conditioned if v is significantly smaller than π, and it is always advisable to use $\mathbf{S} + \epsilon\mathbf{I}$ instead, for small $\epsilon > 0$, where $\mathbf{I}$ is the identity matrix.[17]

For the remainder of the paper we focus on the choice $\mathbf{P} = \mathbf{S}$, so the $p(t)$ incorporates just the information about the support of $f(t)$. The resulting PDFT(t) has extrapolated infinitely many nonzero spectrum values, given by

$$\hat{F}(m) = \sum_{n=-N}^{N} a_n P(m - n), \qquad |m| > N, \tag{7}$$

and is a closed-form, noniterative implementation of the familiar band-limited extrapolation commonly achieved by using Gerchberg–Papoulis iteration.[18,19]

For a given data vector $\mathbf{F}$ the PDFT(t) has more energy than the DFT but has minimum energy subject to the data and support constraints. The energy of the PDFT(t) is given by

$$\int_{-v}^{v} |\mathrm{PDFT}(t)|^2 \mathrm{d}t = \mathbf{F}^\dagger \mathbf{S}^{-1} \mathbf{F} = \mathbf{a}^\dagger \mathbf{S} \mathbf{a}. \tag{8}$$

If the original object $f(t)$ is well modeled by Eq. (6) and the coefficients a_n are not abnormally large, then we would expect Eq. (8) to be reasonably small. If, on the other hand, $\mathbf{F}$ happens to be a vector chosen randomly, then $\mathbf{a} = \mathbf{S}^{-1}\mathbf{F}$ could have sizable entries, because some of the eigenvalues of $\mathbf{S}$ are very close to zero. This is a significant point when we consider using Eq. (8) as a way to retrieve lost phase values of the $F(n)$.

The phase problem (in its one-dimensional formulation) is to recover $f(t)$ from the magnitude data $|F(n)|$; that is, the phases ϕ_n in $F(n) = |F(n)| \exp(i\phi_n)$ have been lost. For any choice of phases θ_n in $[-\pi, \pi]$, and $\boldsymbol{\theta} = (\theta_{-N}, \ldots, \theta_N)$, the complex vector $\mathbf{G}(\boldsymbol{\theta})$ given by $\mathbf{G}(\boldsymbol{\theta}) = [|F(-N)| \exp(i\theta_{-N}), \ldots, |F(N)| \exp(i\theta_N)]^T$ can replace $\mathbf{F}$ in the PDFT (with $\mathbf{P} = \mathbf{S}$) to obtain an extrapolation of the spectrum beyond the data window and thus a data-consistent estimate of $f(t)$, supported on V. The energy in this estimate is, from Eq. (8),

$$E(\boldsymbol{\theta}) = \mathbf{G}(\boldsymbol{\theta})^\dagger \mathbf{S}^{-1} \mathbf{G}(\boldsymbol{\theta}). \tag{9}$$

As we remarked earlier, for the correct choice of phases we get $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{G}(\boldsymbol{\phi}) = \mathbf{F}$, and we expect $E(\boldsymbol{\phi})$ to be moderate sized. But, for other choices of $\boldsymbol{\theta}$, the vector $\mathbf{G}(\boldsymbol{\theta})$ may not resemble a vector of spectrum values from an object well modeled by Eq. (6), and the $E(\boldsymbol{\theta})$ should reflect this by growing larger. Because $\mathbf{S}$ is nondiagonal, $E(\boldsymbol{\theta})$ depends on the phases chosen and provides a reasonable cost function for phase retrieval.

Based on Eq. (9) the energy of the PDFT(t) was calculated as a function of increasing phase noise added to the true Fourier phases. The norm of the estimate (the numbers are for illustrative purposes only) increases rapidly as the percentage phase noise also increases. This can be seen in Fig. 1a, where the noise range increases from 0 to 50%. Figure 1b shows, expanded, the range from 0 to 0.01%. These curves are typical of those obtained for a variety of one- and two-dimensional objects. The percentage error is defined to be

$$\{\text{average(true phase-noisy phase)/average(true phase)}\} \cdot (100)\%, \tag{10}$$

where the phase errors were additive and were mean-zero independent with Gaussian random variables. As can be seen, the norm of the estimate increases dramatically with the phase error but not quite monotonically for small percentage values. For the very small phase errors, the norm merely fluctuates about a value equal to that for perfect data; to the eye, none of these estimates can be distinguished from the original object.

One way to implement the minimization of $E(\boldsymbol{\theta})$ as a method of phase retrieval would be to undertake a search in $\boldsymbol{\theta}$ space for a minimum. We shall report on such methods in a later paper. Our focus in the remainder of this paper is on iterative procedures that resemble those of Gerchberg–Papoulis and Fienup and offer some hope of at least partial phase retrieval with relatively less computation. Before going on to consider algorithms to implement this procedure, it is worth noting the relationship between the PDFT and the minimum-cross-entropy (or maximum-Shannon-entropy) approach to phase retrieval. If $g(t) \geq 0$ is a nonnegative object and $p(t) \geq 0$ the prior profile function, then the cross entropy from p to g is $H(g, p) = \int g \log(g/p)$. The minimum-cross-entropy (MCE) approach selects as the op-

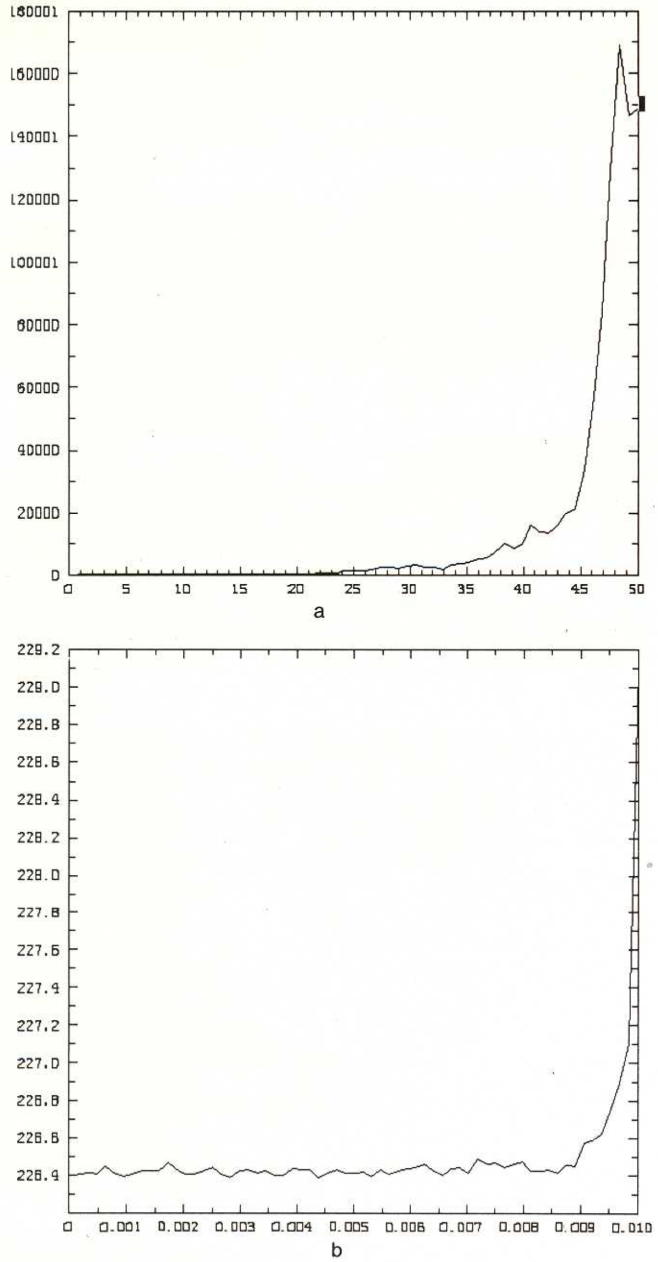


Fig. 1. The dependence of the norm [Eq. (8)] on phase errors ranging from a, 0 to 50% and b, 0 to 0.01%, for a two-dimensional object function.

timal estimate of (nonnegative) $f(t)$ that $g(t)$ for which $H(g, p)$ is minimized, among all data-consistent g. The MCE solution has the algebraic form

$$\mathrm{MCE}(t) = p(t)\exp\left[\sum_{n=-N}^{N} b_n \exp(-int)\right], \qquad (11)$$

with the b_n chosen (iteratively) to make MCE(t) data consistent.

If $p(t)$ is a reasonable approximation to $f(t)$, then the sum in the exponent of Eq. (11) will be around zero, and the estimate $\exp(x) \simeq 1 + x$ for x near zero shows that MCE(t) is nearly of the form described in Eq. (6). Writing out the expression for $H = H[\mathrm{MCE}(t), p(t)]$, we have (ignoring constants)

$$H = \int_{-v}^{v} \mathrm{MCE}(t)\left[\sum_{n=-N}^{N} b_n \exp(-int)\right]$$

$$\approx \int_{-v}^{v} \left\|\left[\sum_{n=-N}^{N} b_n \exp(-int)\right]\right\|^2, \qquad (12)$$

since $p(t)$ is 1 on V and 0 off. Thus the MCE solution for a fixed vector $\mathbf{G}(\theta)$ has cross entropy nearly equal to $E(\theta)$, at least for θ near the true phases. This suggests that our approach can be viewed as a less computationally expensive way to implement the MCE approach to phase retrieval.[20]

ITERATIVE ALGORITHMS FOR MINIMUM-ENERGY PHASE RETRIEVAL

For the purpose of presenting our iterative algorithm to minimize the extrapolation energy $E(\theta)$ for phase retrieval, we introduce the following spaces and operators: $L^2(-\pi, \pi)$ is the usual Hilbert space of all finite energy objects g supported on $[-\pi, \pi]$. All such objects have Fourier series representations, as in Eq. (2), and the sequence of Fourier coefficients will be denoted $\mathcal{F}g$. Given a sequence of Fourier coefficients $G = \{G(n)\}$, the function g is $\mathcal{F}^{-1}\mathbf{G}$. The operator $\mathbf{V}$ is the orthogonal linear projection of $L^2(-\pi, \pi)$ on $L^2(-\mathbf{v}, \mathbf{v})$, which associates with each g in $L^2(-\pi, \pi)$ the function $\mathbf{V}g$, equal to g on V and zero outside $[-v, v]$. The set of indices $A = \{-N, \ldots, N\}$ is our data window, and $\mathbf{A}$ is the operator that sets to zero the sequence values $G(n)$ for n not in A. The operator $\mathbf{D}$ replaces $|G(n)|$ with $|F(n)|$, our data values, for n in A; note that $\mathbf{DA} = \mathbf{AD}$.

The iterative algorithm starts with $g_0 = \mathbf{V}\mathcal{F}^{-1}\mathbf{G}(\theta)$, θ arbitrary; we identify $\mathbf{G}(\theta)$ and the infinite sequence having only the entries of $\mathbf{G}(\theta)$ as nonzero values. Then, for $n \geq 1$,

$$g_{n+1} = \mathbf{V}\mathcal{F}^{-1}(\mathbf{I} - \mathbf{A})\mathcal{F}g_n + \mathbf{V}\mathcal{F}^{-1}\mathbf{DA}\mathcal{F}g_n. \qquad (13)$$

Writing

$$\mathbf{P}_1 g = \mathbf{V}g, \qquad \mathbf{P}_2 g = \mathcal{F}^{-1}(\mathbf{I} - \mathbf{A})\mathcal{F}g + \mathcal{F}^{-1}\mathbf{DA}\mathcal{F}g,$$

we have

$$g_{n+1} = \mathbf{P}_1\mathbf{P}_2 g_n. \qquad (14)$$

Here $\mathbf{P}_1$ is the orthogonal linear projection onto a linear subspace. The $\mathbf{P}_2$ is also a projection ($\mathbf{P}_2{}^2 = \mathbf{P}_2$, since $\mathbf{DA} = \mathbf{AD}$), and $\mathbf{P}_2$ is the closest data-consistent function to g, for each g, but the range of $\mathbf{P}_2$ is not a linear variety (is not even convex), and so we cannot rely on the standard results on iterated projections[21,22] to give us convergence. However, we can show that the limit, if it exists, is in the range of $\mathbf{P}_2$ but not that it is the closest function to f that is both data consistent and supported on $\mathbf{V}$; that is, we cannot be sure that the limit has the minimum energy subject to the constraints or that it minimizes $E(\theta)$.

As Levi and Stark[23] have shown, however, all is not lost when one of the projections has nonconvex range. They define the set-distance error

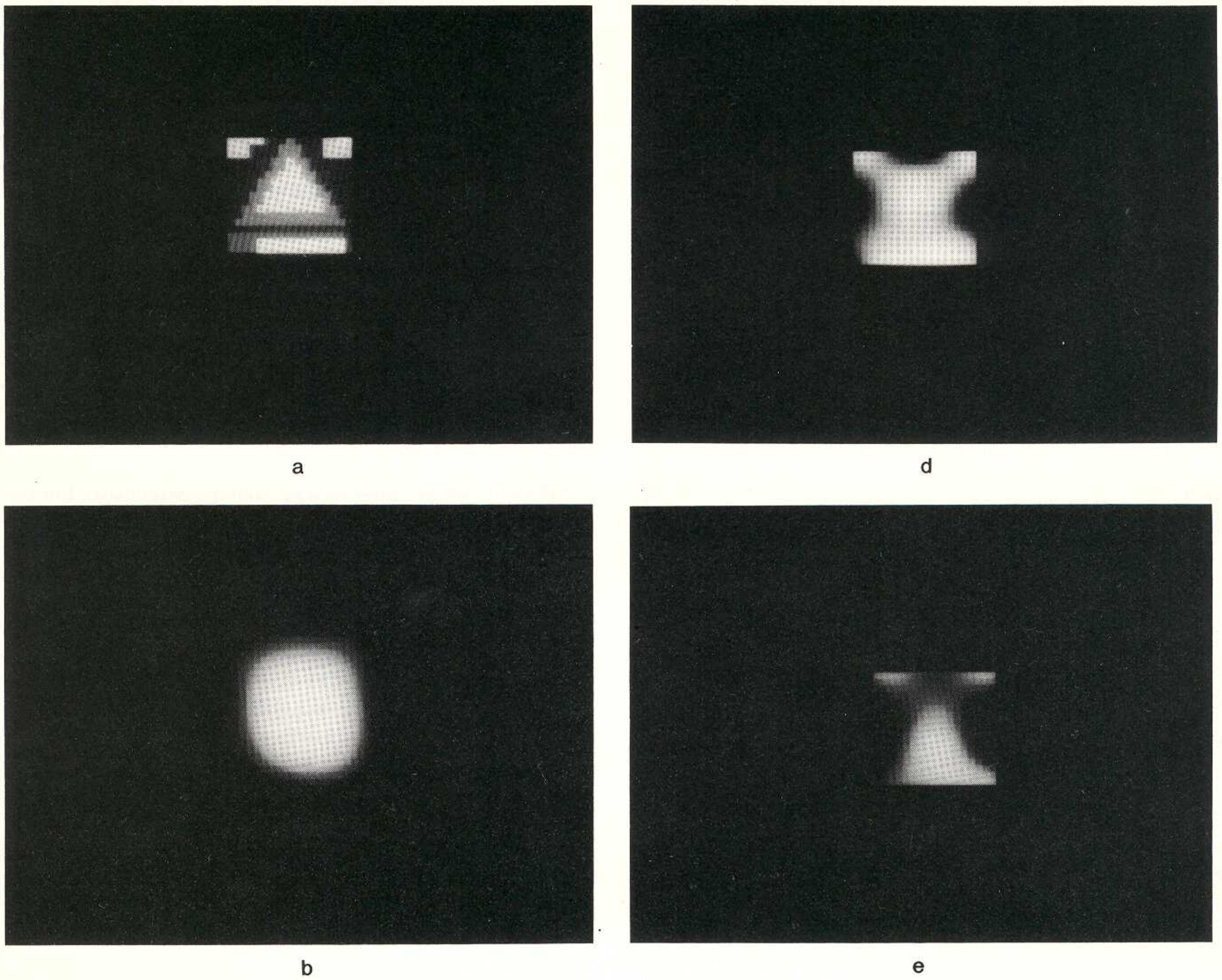
$$J(g) = \|\mathbf{P}_1 g - g\| + \|\mathbf{P}_2 g - g\|, \qquad \|\ \| \text{ the } L^2(-\pi, \pi)\text{norm} \qquad (15)$$

and prove that $J(g_{n+1}) \leq J(\mathbf{P}_2 g_n) \leq J(g_n)$. So, in our situation, g_{n+1} is at least as close to being data consistent as g_n is, and $\mathbf{P}_2 g_{n+1}$ is at least as close to being supported on V as $\mathbf{P}_2 g_n$ is.

The iterative step [Eq. (13)] seems to require the use of the projection $\mathbf{I} - \mathbf{A}$, involving infinitely many Fourier coefficients; however, writing Eq. (11) as

$$g_{n+1} = g_n - \mathbf{V}\mathcal{F}^{-1}\mathbf{A}\mathcal{F}g_n + \mathbf{V}\mathcal{F}^{-1}\mathbf{D}\mathbf{A}\mathcal{F}g_n \tag{16}$$

shows that only finite Fourier-transform values are needed.



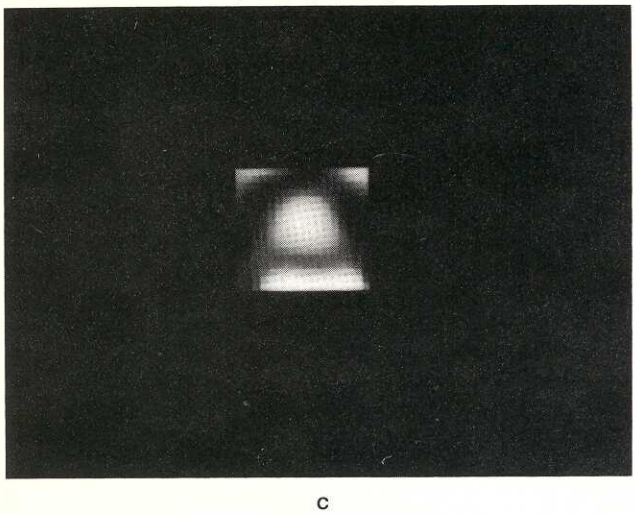


Fig. 2. a, Original 17 × 17 (nonzero) object embedded in 256 × 256 zero array; b, DFT of 32 × 32 complex Fourier samples; c, PDFT calculated from Eqs. (6) and (7); d, an approximation to c after only 100 iterations and that converges to the latter; e, the reconstruction from Fourier magnitude alone after 100 iterations.

In practice, the implementation of the operation $\mathcal{F}^{-1}$ would be done discretely, and we would have to decide how many sampled values to use. If we choose only $2N + 1$, so that $\mathcal{F}$ and $\mathcal{F}^{-1}$ are simply viewed as $2N + 1$ point fast-Fourier-transform (FFT) operations, then Eq. (16) becomes

$$g_{n+1} = \mathbf{V}\mathcal{F}^{-1}\mathbf{D}\mathcal{F}g_n, \tag{17}$$

which is Fienup's error-reduction iterative method.[24]

We can avoid the use of a discrete implementation of $\mathcal{F}^{-1}$, however, by noticing that each g_n has the form $g_n = \mathbf{V}\mathcal{F}^{-1}\mathbf{a}_n$, where $\mathbf{a}_n$ is a sequence that is nonzero only within $\mathbf{A}$. Identifying the infinite sequence a_n and the $2N + 1$ by 1 vector of its nonzero entries, we can write an iterative procedure for the a_n:

$$\mathbf{a}_{n+1} = \mathbf{a}_n - \mathbf{S}\mathbf{a}_n + \mathbf{D}\mathbf{S}\mathbf{a}_n, \tag{18}$$

and $\mathbf{a}_0 = \mathbf{G}(\boldsymbol{\theta})$. The reconstruction procedure is now to iterate Eq. (18) until convergence to $\mathbf{a}$ and then to use $g = \mathbf{V}\mathcal{F}^{-1}\mathbf{a}$ as the answer; here the implementation of $\mathcal{F}^{-1}$ is purely a graphing decision and does not affect the computed solution.

In the two-dimensional case the data are taken over a grid of points and then vectorized. If the grid is $(2N + 1)$ by $(2N + 1)$, then the vector of data is $(2N + 1)^2$ by 1. The $\mathbf{P}$ matrices involved are then $(2N + 1)^2$ by $(2N + 1)^2$. In the case of separable prior profiles, however, a considerable reduction in computation can be achieved. In particular, with $\mathbf{P}$ representing a square support region in the plane, we can bypass the vectorizing step entirely and keep the data $\mathbf{F}$ and the coefficients in Eq. (6) in the form of $(2N + 1)$ by $(2N + 1)$ matrices, which we denote by $\mathbf{F}$ and $\mathbf{C}$, respectively. The operation denoted earlier by $\mathbf{Pa}$ reduces to $\mathbf{SCS}$, where $\mathbf{S}$ is the same sinc matrix used in the one-dimensional problem. The inversion required to get the PDFT becomes $\mathbf{C} = \mathbf{S}^{-1}\mathbf{F}\mathbf{S}^{-1}$.

When the iterative method [Eq. (13)] is applied to the two-dimensional problem of phase retrieval in the form given by Eq. (18), the matrix multiplications required can be performed by using the same $\mathbf{S}$ matrix as in the one-dimensional case; the iteration becomes

$$\mathbf{C}_{n+1} = \mathbf{C}_n - \mathbf{S}\mathbf{C}_n\mathbf{S} + \mathbf{D}\mathbf{S}\mathbf{C}_n\mathbf{S}, \tag{19}$$

$\mathbf{C}_0 = \mathbf{G}(\boldsymbol{\theta})$, as a matrix. This permits a reduction in computation from $(2N + 1)^4$ to $(2N + 1)^3$ for any separable prior profile function $p(t)$.

Returning to the Fienup error-reduction method briefly, the iterative procedure [Eq. (17)] is guaranteed a nondecreasing set-distance error [Eq. (15)] but may stagnate. The limit of the g_n, when it exists, cannot be data consistent, since it is the restriction to V of a polynomial that is itself data consistent. Consequently, if the iterative procedure is carried out by using a FFT algorithm, convergence is sought on the basis of consistency with Fourier magnitude data and the inverse transform of $\mathbf{D}\mathcal{F}g_n$ being zero at the discrete points outside V. It is our conjecture that the DFT of $\mathbf{D}\mathcal{F}g_n$ may attain these goals even when g_n represents f poorly and there is considerable energy in the DFT outside V. This could provide a mechanism for stagnation that could be avoided through the use of the iterative scheme proposed here; our estimate is identically zero off V.

NUMERICAL EXPERIMENTS

Figure 2a shows a two-dimensional object having dimensions 17×17 and embedded in a 256×256 array of zeros, corresponding to a Fourier oversampling factor of 15. Figure 2b shows the DFT of 32×32 complex Fourier samples and is clearly a low-pass-filtered version of the original object, as expected. In the DFT estimate, considerable energy lies outside the true object support, but this is not visible in the figure. Thus, even with correct Fourier phase values, this is the best reconstruction that one could hope for by taking the Fourier transform of such a limited data set. Using the Fourier data and a previous profile function describing the known object support, the resulting PDFT was calculated from Eqs. (6) and (7) and is shown in Fig. 2c. A small degree of regularization was introduced by multiplying the diagonal elements of $\mathbf{P}$ in Eq. (7) by 1.00005. Figure 2d, to be compared with Fig. 2c, shows the reconstructed object after 100 iterations of the procedure given by Eq. 18 but with the $\mathbf{D}$ operator absent and the original complex data reintroduced. This estimator will converge to that shown in Fig. 2c as the number of iterations increases. After the same number of iterations as used for Fig. 2d, but including the $\mathbf{D}$ operator, Fig. 2e was obtained. For the examples studied, it was found that stagnation could sometimes occur when a random or symmetric initial phase guess was used. The stagnation seemed to be toward a twin-image configuration. However, it appears that this can be avoided through the use of an initial phase guess corresponding to an asymmetric object with the same support; in the example shown, the phase function was obtained from an object having the same support but containing a linear ramp.

CONCLUSIONS

The problem of reconstructing a function from its Fourier magnitude was considered. The particular problem of recovering a continuous object estimate from limited Fourier magnitude samples was addressed from the point of view of determining an optimal solution consistent both with the data and with the support constraint. Because the data are finite, there are ambiguities in continuous solutions satisfying these constraints, and we believe that these outweigh any ambiguities that may exist owing to zero-factor flipping when continuous data are available. We note that Frost *et al.*[25] also studied an iterative algorithm that combined phase recovery with spectral extrapolation.

A method was presented for a minimum energy object estimate to be determined, and it was demonstrated that the norm of this estimate is sensitive to the choice of the Fourier phase. A search in phase space to minimize this norm is ideally the procedure to be adopted, and future work will concentrate on this. In the special case of a prior weighting function defining only support information, an iterative procedure can be used to minimize the norm of the estimate. If a separable prior function is used, then a simple and fast algorithm can be implemented, and the results presented here were based on this.

The relationship of this approach to Fienup's error-reduction algorithm was demonstrated. An important application of the new method is to limited Fourier magnitude data sets, for which the DFT, even with the correct Fourier phases, provides a poor estimate.

In particular, when such an estimate has considerable energy outside the anticipated object support region, a possible cause of stagnation of an iterative algorithm is that the estimate be forced to zero on a grid of points outside the object support. This could be achieved at the price of energy still residing between the grid points; hence the rationale behind the method presented here, which ensures an estimate that is identically zero beyond the object support.

In the examples processed by this algorithm, the main cause of stagnation seems to be convergence toward a twin image. The phase chosen to start the iterative procedure is clearly an important factor in causing this mode of stagnation, and rather than a random or a symmetric initial phase guess a phase obtained from a simple ramp object proved appropriate. For the examples presented here it follows that the relative importance of the support constraint increases with the oversampling factor.

ACKNOWLEDGMENTS

The authors are grateful to U.S. Office of Naval Research for its financial support through contract N00014-85-K-0402 and NATO grant 725/84.

* Permanent address, Department of Mathematics, University of Lowell, Lowell, Massachusetts 01854.

† Permanent address, Department of Physics, Kings College, London WC2R 2LS, United Kingdom.

REFERENCES AND NOTES

1. H. A. Ferwerda, "The phase reconstruction problem for wave amplitudes and coherence functions," in *Inverse Source Problems in Optics*, H. P. Baltes, ed. (Springer-Verlag, Berlin, 1978), Vol. 9, p. 13.
2. L. S. Taylor, "The phase retrieval problem," IEEE Trans Antennas Propag. **AP-29,** 317–343 (1981).
3. M. A. Fiddy, "The phase retrieval problem," in *Inverse Optics*, A. J. Devaney, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **413,** 176–181 (1983).
4. J. Rosenblatt, "Phase retrieval," Commun. Math. Phys. **95,** 317–343 (1984).
5. J. L. C. Sanz, "Mathematical considerations for the problem of Fourier transform phase retrieval from magnitude," SIAM J. Appl. Math. **45,** 651–664 (1985).
6. I. S. Stefanescu, "On the phase retrieval problem in two dimensions," J. Math. Phys. **26,** 2141–2160 (1985).
7. R. Barakat and G. Newsam, "Algorithms for reconstruction of partially known, band-limited Fourier-transform pairs from noisy data," J. Opt. Soc. Am. A **2,** 2027–2039 (1985).
8. M. H. Hayes and J. H. McClellan, "Reducible polynomials in more than one variable," Proc. IEEE **70,** 197–198 (1982).
9. R. Barakat and G. Newsam, "Necessary conditions for a unique solution to two dimensional phase recovery," J. Math. Phys. **25,** 3190–3193 (1984).
10. W. Lawton "Uniqueness results for the phase retrieval problem for radial functions," J. Opt. Soc. Am. **71,** 1519–1522 (1981).
11. J. L. C. Sanz and T. S. Huang, "Unique reconstruction of a band-limited multidimensional signal from its phase or magnitude," J. Opt. Soc. Am. **73,** 1446–1450 (1983).
12. T. S. Huang, ed., *Advances in Computer Vision and Image Processing* (JAI, New York, 1984), Vol. 1.
13. C. R. Smith and W. T. Grandy, *Maximum Entropy and Bayesian Methods in Inverse Problems* (Reidel, Dordrecht, The Netherlands, 1985).
14. A. M. Darling, H. V. Deighton, and M. A. Fiddy, "Phase ambiguities in more than one dimension," in *Inverse Optics*, A. J. Devaney, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **413,** 197–201 (1983).
15. C. L. Byrne and R. M. Fitzgerald, "Spectral estimates that extend the maximum entropy and maximum likelihood methods," SIAM J. Appl. Math **44,** 425–442 (1984).
16. C. L. Byrne, R. M. Fitzgerald, M. A. Fiddy, T. J. Hall, and A. M. Darling, "Image restoration and resolution enhancement," J. Opt. Soc. Am. **73,** 1481–1487 (1983).
17. A. M. Darling, T. J. Hall, and M. A. Fiddy, "Stable, noniterative object reconstruction from incomplete data using prior knowledge," J. Opt. Soc. Am. **73,** 1466–1469 (1983).
18. R. W. Gerchberg, "Super-resolution through error energy reduction," Opt. Acta **21,** 709–720 (1974).
19. A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," IEEE Trans. Circuits Syst. **CAS-22,** 733–742 (1975).
20. C. L. Byrne and L. K. Jones, SIAM J. Appl. Math. (to be published).
21. D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," IEEE Trans. Circuits Syst. **CAS-25,** 694–702 (1978).
22. D. C. Youla and H. Webb, "Image restoration by the method of projections onto convex sets," IEEE Trans. Med. Imaging **TMI-1,** 81–94 (1982).
23. A. Levi and H. Stark, "Image restoration by the method of generalized projections with application to restoration from magnitude," J. Opt. Soc. Am. A **1,** 932–943 (1984).
24. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21,** 2758–2769 (1982).
25. R. L. Frost, C. K. Rushforth, and B. S. Baxter, "High resolution astronomical imaging through the turbulent atmosphere," Tech. Rep. UTEC 79-081 (University of Utah, Salt Lake City, 1979). Frost *et al.* developed an algorithm with some similarities to that presented here. Their procedure did not correct the magnitude spectrum to the measured magnitude each iteration, as in Fienup's error-reduction algorithm. Instead, the magnitude and phase spectra were alternately extended through the frequency space. The phase-spectrum extension was allowed to converge, and then the new magnitude spectrum was scaled to agree with the measured magnitude below the old cutoff frequency. The reason for doing this was to limit instabilities in the extrapolation due to reintroducing the noise present in the known portion of the spectrum. They studied the consequences of varying the rate of extrapolation on the relative contribution of the noise on the final extrapolated estimate. Results presented by them looked promising, but further development and optimization of the procedure were not pursued.