Image restoration and resolution enhancement

Charles L. Byrne and Raymond M. Fitzgerald
Code 5120, U.S. Naval Research Laboratory, Washington, D.C. 20375

Michael A. Fiddy, Trevor J. Hall, and Angela M. Darling
Department of Physics, Queen Elizabeth College, London W8 7AH, U.K.

Received February 11, 1983

The ill-posed problem of restoring object information from finitely many measurements of its spectrum can be solved by using the best approximation in Hilbert spaces appropriately designed to include a priori information about object extent and shape and noise statistics. The procedures that are derived are noniterative, the linear ones extending the minimum-energy band-limited extrapolation methods (and thus related to Gerchberg–Papoulis iteration) and the nonlinear ones generalizing Burg's maximum-entropy reconstruction of nonnegative objects.

1. INTRODUCTION

We consider the restoration of object information from finitely many measurements of the object's spectrum, with particular emphasis on the design of new linear and nonlinear methods that incorporate a priori information. The linear methods extend the minimum-energy reconstruction of band-limited functions, which is the limit of Gerchberg–Papoulis iterative extrapolation in the discrete-data case. The nonlinear methods apply to the reconstruction of nonegative objects and include Burg's maximum-entropy method (MEM) as a particular case.

Even in the absence of noise the finiteness of the data implies that there are infinitely many mathematically permissible solutions to the reconstruction problem. This inherent ambiguity is overcome by using a priori knowledge to reduce the class of allowed solutions and to design appropriate optimality criteria that lead to useful reconstruction models.

Because inverse problems of the sort that we consider are ill posed, the discrete formulations typically involve ill-conditioned systems of linear equations. Restoration methods based on solutions of these systems are often sensitive to noise. Some form of regularization, involving a compromise between fidelity to the data and smoothness in the restored object, is usually needed. Although our primary concern in this paper is the ambiguity that is due to finite data, stability in the reconstruction is essential and, as we shall show, regularization is included in the methods that we offer.

The purpose of this paper is to demonstrate the feasibility of incorporating a priori information in reconstruction through the use of suitably designed Hilbert spaces. By using optimal approximation in these spaces we are able to derive linear and nonlinear adaptive reconstruction procedures that include, and considerably extend, several procedures already in the literature. We assume throughout that the user has available a priori information about the object. Our goal is to derive flexible models that can incorporate such prior information. The origin of the information or its correctness is not discussed. It is important to note that these are not simple procedures to clean up noise or to undo systematic distortion; the a priori information need not be derived from noise statistics or analysis of the imaging system.

2. LINEAR RESTORATION MODELS

The (one-dimensional) object function $o(x)$ is related to the image $i(y)$ by the imaging equation

$$i(y) = \int_{-\infty}^{\infty} o(x)s(y-x)dx,$$  \hspace{1cm} (2.1)

where $s(y-x)$ is the system point-spread function. We assume from now on that $o(x) = 0$ for $|x| > X$ (finite object support) so that the infinite integral in Eq. (2.1) can be replaced by a finite one. If the imaging system is spatially invariant and ideal diffraction limited we have

$$i(y) = \frac{1}{\pi} \int_{-\infty}^{\infty} o(x) \sin(\Omega(y-x)) dy,$$  \hspace{1cm} (2.2)

where $[-\Omega, \Omega]$ is the passband of the system. The properties of the eigenvalues and eigenfunctions of Eq. (2.2) have been studied thoroughly. The eigenfunctions $\phi_n(x)$ are the prolate spheroidal wave functions for $\Omega$ and $X$, and the associated eigenvalues $\lambda_0 > \lambda_1 > \ldots > 0$ tend to zero rapidly for $n > 2\Omega X/\pi$. The object function can be expanded as

$$o(x) = \sum_{n=0}^{\infty} \lambda_n^{-1} \left[ \int_{-X}^{X} i(y) \phi_n(y)dy \right] \phi_n(x),$$  \hspace{1cm} (2.3)

from which it is clear that slight errors in measuring $i(y)$ can lead to large changes in the restored object function.

In principle, because $o(x)$ has finite support, its spectrum or Fourier transform,

$$O(\omega) = \int_{-\infty}^{\infty} o(x) \exp(-ix\omega)dx/2\pi,$$  \hspace{1cm} (2.4)

is analytic and can be analytically continued beyond $[-\Omega, \Omega]$ to provide higher resolution. The behavior of the eigenvalues

0030-3941/83/11481-07801.00 © 1983 Optical Society of America
is a statement of the difficulty of achieving stable continuation with continuous data. For finite data no such unique continuation exists, and one must employ models. Let us assume, for the remainder of Section 2, that the data are \( O(m, \Delta), m = 1, \ldots, M \), for some fixed \( \Delta, 0 < \Delta \leq \pi = \pi / \Delta \). If \( \Delta < d \), then the sampling rate is above the Nyquist rate.

From the Shannon sampling theorem we know that, for \( |x| \leq X \),

\[
o(x) = d \sum_{n=-\infty}^{\infty} O(n, d) \exp(in \Delta x),
\]

so that

\[
o(x) = d \sum_{n=-\infty}^{\infty} O(n, d) \frac{\sin[X(\omega - nd)]}{\pi(\omega - nd)},
\]

To recover \( o(x) \) in Eq. (2.5) we need to solve Eq. (2.6) for the \( O(n, d) \). Use of \( \omega = m \Delta \) in Eq. (2.6), for each \( m = 1, \ldots, M \), leads to an undetermined linear system of equations in the unknowns \( O(n, d) \):

\[
o(m, \Delta) = d \sum_{n=-\infty}^{\infty} O(n, d) \frac{\sin[X(m \Delta - nd)]}{\pi(m \Delta - nd)}.
\]

If we arbitrarily decide to set \( O(n, d) = 0 \) for all \( n \) except \( n = 1, \ldots, M \), we have the method of Harris. If, instead, we seek the solution having the least energy \( \text{minimum} L^2(-X, X) \| |o(x)| \| \), we find that we must solve the system of equations

\[
o(m, \Delta) = \sum_{n=1}^{M} a_n \frac{\sin[X(m \Delta - nd)]}{\pi(m \Delta - nd)},
\]

Once the \( a_n \) are determined, the estimator of \( o(x) \) can be written in closed form as

\[
o(x) = p(x) \sum_{n=1}^{M} a_n \exp(i \Delta x),
\]

where

\[
p(x) = \begin{cases} 1, & |x| \leq X \\ 0, & |x| > X \end{cases}
\]

We find it convenient to define \( p(a, b, x) = 1 \) for \( a \leq |x| \leq b \) and 0 otherwise, so that Eq. (2.10) becomes \( p(x) = r(-X, X, x) \). If, instead of solving Eq. (2.8), we merely use \( a_n = O(n, \Delta) \), we have the discrete Fourier transform (DFT) estimator of \( o(x) \). The estimator [Eqs. (2.8)–(2.10)] is the closed-form limit of Gerchberg's and Papoulis's iterative extrapolation. As is well known, this estimator is unstable; because the model requires the restored object to vanish off \( [-X, X] \), any broadband noise in the data can spuriously cause spurious energy. Because Harris's model has even more energy, it too is unstable. Some examples will help to illustrate this point.

In Fig. 1 (and subsequent figures) the object to be reconstructed is the nonnegative function shown in the background. The object contains a small flat component (amplitude is 0.0001) extending from \(-\pi\) to \( \pi\) that is not visible, so \( X = \pi \). Our \( \Delta \) is 1 and the complex data are \( O(m), |m| \leq 6 \). Figure 1 shows the DFT estimate of \( o(x) \). In Fig. 2 we see the instability caused by the small [\(-\pi, \pi\)] component when the support of \( o(x) \) is mistakenly chosen to be \([-1.8, 1.8]\). To convince ourselves that it is not the matrix inversion that is at fault we modify the data, removing the \([-\pi, \pi]\) component. This amounts to changing \( O(0) \) from 0.345933 to 0.345933. Figure 3 is the restored object from the modified data, using the same support as in Fig. 2. The problem is to achieve the improvement of Fig. 3 (over Fig. 1) without modifying the data. We can regularize the method to avoid the spurious features of Fig. 2.

The algebraic form of estimator (2.9) suggests that the \( p(x) \) be used to incorporate more general \( a \) priori information than simply a guess of the support of the object. Indeed, we must use more than that if we are to improve on the DFT while avoiding the oscillations that are due to error in estimating the actual object support. Let us modify the \( p(x) \) in Eq. (2.10) as follows: For some positive \( \varepsilon \) take

\[
p(x) = \begin{cases} 1, & |x| \leq X \\ \varepsilon, & X < |x| < \pi / \Delta \end{cases}
\]

or \( p(x) = r(-X, X, x) + \varepsilon(-\pi / \Delta, \pi / \Delta, x) \). This is a more-realistic \( a \) priori estimate of the object that produced the data we have measured. Our estimate of \( o(x) \) is now Eq. (2.9), with \( p(x) \) as in Eq. (2.11) and the \( a_n \) chosen so as to make Eq. (2.9) data consistent.

In Fig. 4 we see the resulting estimate of \( o(x) \), using Eq. (2.11) with \( X = 1.8 \) and \( \varepsilon = 0.0001 \). The reconstruction within \([-1.8, 1.8]\) is as good as in Fig. 3; the procedure has been regularized to ensure stability against noise. This method has been shown to be equivalent to a form of Miller regularization (see Ref. 12). Note that the estimator is invariant to a change of scale; multiplying \( p(x) \) by a positive constant has no effect. Consequently only relative amplitudes are expressed in \( p(x) \). It is not the case, therefore, that \( \varepsilon = 0.0001 \) is the correct noise level or that near-perfect estimation of this noise is required. This \( \varepsilon \) is not correct with regard to the ratio of maximum amplitudes and outside \([-1.8, 1.8]\), and it is not correct with regard to the jumps in the object that occur at \( \pm \pi / 2 \). In fact, the behavior of this estimator is roughly unchanged over a wide range of \( \varepsilon \). When \( \varepsilon \) is very big (so that
1 + ε and ε are about equal), the estimator looks like the DFT. If ε is nearly equal to machine zero, the behavior of Fig. 2 reappears. The behavior shown in Fig. 4 is typical whenever these two extremes are avoided.

We can do more with the \( p(x) \) in Eq. (2.9) besides including a regularizing component. Any other information about the general shape of \( o(x) \) can be included to free degrees of freedom for the polynomial factor of Eq. (2.9) to describe the less prominent features. Let us then permit \( p(x) \) in Eq. (2.9) to be any positive function and find the \( a_n \) for which Eq. (2.9) is data consistent. The equations that we must solve are

\[
O(m\Delta) = \sum_{n=1}^{M} a_n \int p(x) \exp[i(m - n)\Delta x] dx. \tag{2.12}
\]

The resulting estimator from Eq. (2.9) is optimal in minimizing the weighted error

\[
\text{error} = \int |o(x) - \delta(x)|^2 p^{-1}(x) dx, \tag{2.13}
\]

where the integral is taken over the support of \( p(x) \), which we always take to contain the true support of \( o(x) \). The appearance in Eq. (2.13) of the reciprocal weighting suggests an increase in sensitivity to less prominent features. This shows up in Figs. 5 and 6.

In Fig. 5 we use the same estimator as in Fig. 4, except that we take \( X = \pi/2 \). This additional \( a \text{ priori} \) information about the location of the abrupt jumps leads to a somewhat better reconstruction of the low-contrast detail. If we then include in \( p(x) \) the large box in the middle, we see improvement in the positioning of the smaller box on the left-hand side and a better estimate of the 0.5 amplitude level at the base on each side of the large box.
The linear procedures just described are for the reconstruction of \( o(x) \) from spectral values. But the basic idea of using \textit{a priori} information to select appropriate Hilbert spaces within which to reconstruct from linear functional data has a variety of applications.\textsuperscript{15}

The matrix inversion required by Eq. (2.12) would seem to prohibit the use of this method for multidimensional reconstruction or for any problem involving large data sets. However, recent work\textsuperscript{14} suggests that fast Fourier transform and convolution techniques can be applied to provide approximate solutions to Eq. (2.12) in such cases and that the resulting estimates are reasonable.

When the object \( o(x) \) is nonnegative there are available additional, nonlinear techniques for reconstruction. Many of these methods come from power-spectrum analysis of time series or from information theoretic analysis of probability distributions and stochastic inference. In Section 3 we present an adaptive nonlinear reconstruction procedure that includes Burg’s MEM as a particular case.

3. NONLINEAR RESTORATION MODELS

We make the assumption from now on that \( o(x) \) is nonnegative for all \( x \) and that the data values are \( O(n,\Delta), |n| \leq M \). This is a physically reasonable constraint in many cases, and the value of a positivity constraint on the resulting reconstruction has been discussed by numerous authors (e.g., Frieden\textsuperscript{19}). The importance of the Burg MEM\textsuperscript{15–18} is that it is the only known noniterative method that provides a necessarily positive reconstruction of \( o(x) \) consistent with the data above.

The key word here is necessarily. As we demonstrate shortly, MEM is but one member of a large class of adaptive nonlinear reconstruction methods that frequently—but perhaps not always—provide positive, data-consistent reconstructions without iteration. In addition, these methods permit inclusion of \textit{a priori} information through a weighting function, in much the same way as in Section 2. These methods are derived without using stochastic assumptions, through Hilbert space approximation, and, consequently, a nonstochastic justification for MEM estimation is obtained in the process.

Burg’s MEM takes, as the estimator of \( o(x) \), the function \( \hat{o}(x) \) for which the entropy integral

\[
H = \int_{-\pi/\Delta}^{\pi/\Delta} \log \hat{o}(x) \, dx
\]

is maximized, subject to the data constraints

\[
O(m,\Delta) = \int_{-\pi/\Delta}^{\pi/\Delta} \hat{o}(x) \exp(-\imath m \Delta x) \, dx / 2\pi,
\]

where the coefficients satisfy

\[
\sum_{m=0}^{M} c_m O(n,\Delta - m,\Delta) = \begin{cases} 1, & n = 0 \\ 0, & n = 1, \ldots, M \end{cases}
\]

The MEM estimator is necessarily positive throughout \([-\pi/\Delta,\pi/\Delta]\) and is data consistent.

When the object \( o(x) \) is (exactly or nearly) supported on a smaller portion of \([-\pi/\Delta,\pi/\Delta]\), that is, when the data are over-sampled, MEM has a tendency to produce spurious features. In Fig. 7 we have the MEM reconstruction of the same object used earlier. The object is nearly supported on \([-\pi/2,\pi/2]\), so that the oversampling is only by a factor of 2. This tendency of MEM to produce spurious features has convinced a number of users that \( H \) is not the entropy to be used and that the Shannon entropy is to be preferred (see Ref. 15). As we show below, however, it is possible to generalize MEM in such a way that \textit{a priori} information about object concentration, for example, can be included. The resulting estimators do not exhibit the spurious oscillation, and yet, because they are nonlinear, they are capable of providing a high degree of resolution of sharply peaked objects. We begin...
our discussion of these nonlinear models with an approximation theoretic rederivation of MEM.

Let us assume that \( o(x) \) is sufficiently well behaved\(^9\) that its reciprocal admits a factorization, for \( |x| \leq \pi/\Delta \), of

\[
o^{-1}(x) = |h(x)|^2, \tag{3.5}\]

where

\[
h(x) = \sum_{n=0}^{\infty} H_n \exp(in \Delta x) \tag{3.6}\]

is the restriction to \( z = \exp(i \Delta x) \) of a function that is analytic and without zeros inside and on the unit circle \( |z| = 1 \). It can then be shown that a constant \( c \) exists so that, for \( n = 0, 1, 2, \ldots \),

\[
0 = \int_{-\pi/\Delta}^{\pi/\Delta} [o^{-1}(x) - ch(x)]o(x)\exp(-in \Delta x) \, dx. \tag{3.7}\]

Therefore \( ch(x) \) is the orthogonal projection of \( o^{-1}(x) \) onto the subspace of \( L \) spanned by \( \{\exp(in \Delta x), n = 0, 1, 2, \ldots\} \), where by \( L \) we mean the Hilbert space whose inner product is given by

\[
(f, g) = \int_{-\pi/\Delta}^{\pi/\Delta} f(x)g(x)*o(x) \, dx, \tag{3.8}\]

with * denoting a complex conjugate. Let us then compute the polynomial

\[
q_M(x) = \sum_{m=0}^{M} c_m \exp(im \Delta x) \tag{3.9}\]

for which the error

\[
\text{error} = \int_{-\pi/\Delta}^{\pi/\Delta} \left[ p(x) o^{-1}(x) - q_M(x)^2 o(x) \right] \, dx. \tag{3.10}\]

is minimum. This polynomial \( q_M(x) \) can then be viewed as an estimate of \( o(x) \). The equations to be solved are easily seen to be Eq. (3.4). The estimate of \( o(x) \), which is obtained by using Eq. (3.5), is Burg's MEM estimate [Eq. (3.3)].

If we have a priori information about \( o(x) \) in the form of a weighting function \( p(x) \), we can incorporate it in our estimator by modifying the error in Eq. (3.10) to read

\[
\text{error} = \int_{-\pi/\Delta}^{\pi/\Delta} \left[ p(x) o^{-1}(x) - q_M(x)^2 o(x) \right] \, dx. \tag{3.11}\]

The polynomial \( q_M(x) \) is the orthogonal projection of \( p(x) o^{-1}(x) \) onto the subspace of \( L \) spanned by \( \{\exp(in \Delta x), n = 0, 1, \ldots, M\} \). The coefficients of \( q_M(x) \) must then satisfy

\[
\sum_{m=0}^{M} c_m O(n \Delta - m \Delta) = \int_{-\pi/\Delta}^{\pi/\Delta} p(x) \exp(-in \Delta x) \, dx = P(n \Delta), \tag{3.12}\]

The task now is to relate \( q_M(x) \) to \( o(x) \) to get our estimator of the true object.

As \( M \) grows, \( q_M(x) \) converges to \( q(x) \), the orthogonal projection of \( p(x) o^{-1}(x) \) onto the subspace of \( L \) spanned by \( \{\exp(in \Delta x), n = 0, 1, \ldots\} \). Because

\[
0 = \int_{-\pi/\Delta}^{\pi/\Delta} \left[ p(x) o^{-1}(x) - q(x) \exp(-in \Delta x) o(x) \right] \, dx
= \int_{-\pi/\Delta}^{\pi/\Delta} \left[ p(x) - q(x) o(x) \right] \exp(-in \Delta x) \, dx \tag{3.13}\]

for \( n = 0, 1, 2, \ldots \), it follows that \( p(x) \) and \( q(x) o(x) \) have the same causal parts, where the causal part of \( p(x) \) is

\[
p(x)^+ = \Delta \sum_{n=0}^{\infty} P(n \Delta) \exp(in \Delta x). \tag{3.14}\]

By viewing \( q_M(x) \) as an estimate of \( q(x) \), we can treat the causal part of \( q_M(x) o(x) \) as an estimate of the known function \( p(x)^+ \).

From a simple calculation involving the Fourier series we obtain the causal part of \( q_M(x) o(x) \):

\[
[q_M(x) o(x)]^+ = q_M(x) o(x)^+ + j(x), \tag{3.15}\]

with

\[
j(x) = c_1 O(-\Delta) + c_2 [O(-2 \Delta) + O(-\Delta) \exp(i \Delta x)] + \ldots + c_M [O(-M \Delta) + \ldots + O(-\Delta) \exp(M \Delta - i \Delta x)]. \tag{3.16}\]

By solving for \( o(x)^+ \) in Eq. (3.15) and by replacing \( q_M(x) o(x)^+ \) with \( p(x)^+ \) we obtain our estimate of \( o(x)^+ \):

\[
\hat{o}(x) = 2 \text{Re}[\hat{o}(x)^+] - O(0). \tag{3.17}\]

Our estimate of \( o(x) \) is then

\[
\hat{o}(x) = 2 \text{Re}[\hat{o}(x)^+] - O(0). \tag{3.18}\]

If the analytic function

\[
f(z) = \sum_{m=0}^{M} c_m z^m, \tag{3.19}\]

obtained by extending \( q_M(x) \) to the complex plane, has all its zeros outside the unit circle, then the reconstruction in Eq. (3.18) is data consistent. The MEM estimator is a particular case of Eq. (3.18) obtained by choosing \( p(x) = 1, |x| \leq \pi/\Delta, p(x) = 0, |x| > \pi/\Delta \). As is well known, the \( f(z) \) in the MEM case always has its zeros outside the unit circle; hence it is always data consistent.

Returning to our examples, we have in Fig. 8 the reconstruction given by Eq. (3.18) using an a priori weighting function that is concentrated within \([-\pi/2, \pi/2]\). The spurious
Fig. 9. Nonlinear restoration from \( O(n), |n| \leq 6 \) using \( P(x) = 0.5 \)
\( r(-\pi/6, \pi/6, x) + 0.5 r(-\pi/2, \pi/2, x) + 0.0001 r(-\pi, \pi, x) \) in Eq.
(3.18).

Other features of MEM (Fig. 7) have been removed. We see from Fig. 7 that the MEM does a good job of indicating the object support. By incorporating support (or better, concentration) information in the \( p(x) \), we improve the reconstruction of the other features. Figure 9 is based on a \( p(x) \) that includes a further concentration within \([\pi/6, \pi/6]\).

4. COMMENTS AND CONCLUSIONS

We have described two general restoration schemes that incorporate a priori knowledge of object shape and extent. Both procedures are noniterative and are based on optimal approximation in suitably designed Hilbert spaces. The adaptive nature of these methods suggests their use in a succession of reconstructions in which information gained at a previous step is incorporated in the next reconstruction. The linear procedure presented extends the minimum-norm band-limited extrapolation technique and so is related to Geregberg-Papoulis iteration, whereas our nonlinear methods contain and considerably extend Burg's MEM.

The linear procedures offered here are easily extended to higher dimensions and to other sorts of linear functional data. Recent work on simplifying the computation involved suggests that this extension can be made practical. The nonlinear method does not extend so easily. A nonlinear procedure for multidimensional reconstruction is being developed by the authors.

In a number of cases of practical importance the data available do not include phase information (e.g., structure determination from x-ray, optical, and electron scattering). This additional ambiguity is related to the location of complex zeros of the spectrum and is an active area of research.20,21 The use of a priori information, such as known object features, to recover the phase appears crucial. The methods presented here offer a possible solution to this phase problem, since the phase of the spectrum of an a priori estimate can be combined with the known intensity data to provide a good initial guess for some iterative procedure, such as that of Fienup.22

The methods presented here have been applied to real and simulated data, such as Fourier transforms of projection data taken over limited angles. The application of these techniques to data compression also appears promising.23

ACKNOWLEDGMENTS

A. M. Darling gratefully acknowledges the financial support of the Procurement Executive, Ministry of Defense, U.K. Some of this work was carried out while M. A. Fiddy was supported by the U.S. Air Force Office of Scientific Research under grant AFOSR 81 0003. Those authors at Queen Elizabeth College are indebted to the U.K. Science and Engineering Research Council for support.

C. L. Byrne is also with the Department of Mathematics, The Catholic University of America, Washington, D.C. 20064.

REFERENCES