Image reconstruction: a unifying model for resolution enhancement and data extrapolation. Tutorial

Hsin M. Shieh

Department of Electrical Engineering, Feng Chia University, 100 Wenhwa Rd., Seatwen, Taichung, Taiwan 40724

Charles L. Byrne

Department of Mathematical Sciences, University of Massachusetts Lowell, One University Avenue, Lowell, Massachusetts 01854

Michael A. Fiddy

Center for Optoelectronics and Optical Communications, The University of North Carolina at Charlotte, 9201 University City Blvd., Charlotte, North Carolina 28223

Received April 20, 2005; accepted May 31, 2005

In reconstructing an object function $F(r)$ from finitely many noisy linear-functional values $\int F(r)G_n(r)\mathrm{d}r$ we face the problem that finite data, noisy or not, are insufficient to specify $F(r)$ uniquely. Estimates based on the finite data may succeed in recovering broad features of $F(r)$, but may fail to resolve important detail. Linear and nonlinear, model-based data extrapolation procedures can be used to improve resolution, but at the cost of sensitivity to noise. To estimate linear-functional values of $F(r)$ that have not been measured from those that have been, we need to employ prior information about the object $F(r)$, such as support information or, more generally, estimates of the overall profile of $F(r)$. One way to do this is through minimum-weighted-norm (MWN) estimation, with the prior information used to determine the weights. The MWN approach extends the Gerchberg–Papoulis band-limited extrapolation method and is closely related to matched-filter linear detection, the approximation of the Wiener filter, and to iterative Shannon-entropy-maximization algorithms. Nonlinear versions of the MWN method extend the noniterative, Burg, maximum-entropy spectral-estimation procedure. © 2006 Optical Society of America

OCIS codes: 100.2980, 100.3020.

1. INTRODUCTION

The problem of object-function reconstruction from limited data is the following. We have noisy measurements of the linear-functional values

$$d_n = \int F(r)\overline{G_n(r)}\mathrm{d}r, \qquad (1)$$

for $n=1,2,\ldots,N$. $G_n(r)$ are known functions and the overbar denotes complex conjugate. On the basis of these data values, we wish to estimate the function $F(r)$. The variable r is allowed to be multivariate. The region over which the integration is performed is, as yet, unspecified, but will be made explicit in special cases to be discussed later.

The data we have are generally insufficient to specify $F(r)$ uniquely. Among all functions $H(r)$ consistent with the data, the minimum-norm (MN) solution, that is, the one having the smallest energy $\int|H(r)|^2\mathrm{d}r$, has the form

$$H(r) = F_{\mathrm{MN}}(r) = \sum_{n=1}^{N} a_n G_n(r), \qquad (2)$$

where the coefficients a_n are determined by the data-consistency equations

$$d_m = \sum_{n=1}^{N} a_n \int G_n(r)\overline{G_m(r)}\mathrm{d}r, \qquad (3)$$

for $m=1,2,\ldots,N$.

As an example, consider the problem of reconstructing $F(r)$ from finitely many values of its Fourier transform. In this example $G_n(r)=(1/2\pi)\exp(ix_n r)$ for some x_n and $n=1,2,\ldots,N$. Then the data values are

$$d_m = \frac{1}{2\pi}\int F(r)\exp(-ix_m r)\mathrm{d}r = f(x_m), \qquad (4)$$

where $f(x)$ denotes the Fourier transform of $F(r)$,

$$f_{(x)} = \frac{1}{2\pi}\int F(r)\exp(-ixr)\mathrm{d}r, \qquad (5)$$

Then the MN solution is

$$F_{\mathrm{MN}}(r) = \sum_{n=1}^{N} a_n \exp(ix_n r), \qquad (6)$$

with

$$f(x_m) = \sum_{n=1}^{N} a_n \int \exp[i(x_n - x_m)r]\frac{\mathrm{d}r}{2\pi}, \qquad (7)$$

for $m=1,2,\ldots,N$. In the particular one-dimensional case in which $F(r)$ is supported on the interval $[-\pi,\pi]$ and $x_m=m$ we find that $a_n=f(x_n)=f(n)$ and the MN estimate becomes

$$F_{\mathrm{DFT}}(r) = \sum_{n=1}^{N} f(n)\exp(inr), \qquad (8)$$

which we shall refer to as the discrete Fourier transform (DFT) of the finite data. Note that the term DFT is commonly used to denote the finite-length vector obtained by evaluating the function in Eq. (8) at N equispaced points within $[-\pi,\pi]$. Since it is also common practice first to zero-pad, that is, to append zeros to the data vector, before calculating the vector DFT, there is ambiguity in the use of DFT to describe finite vectors associated with the data. For that reason, we believe that the DFT defined to be the function in Eq. (8) is the more fundamental notion.

As is well known, the ability of the DFT to resolve closely spaced peaks in the function $F(r)$ is limited by the value of N. This does not necessarily mean that information about these peaks is unavailable in the data, just that the DFT has failed to make the best use of this information. Indeed, high-resolution methods, such as Gerchberg–Papoulis band-limited extrapolation[1,2] and Burg's maximum-entropy spectral-estimation procedure,[3–5] illustrate this point by resolving peaks left unresolved by the DFT. Resolution limits, properly understood, must include degrees of freedom and signal-to-noise ratio in the data.[6] Such resolution enhancement is often achieved through the explicit or implicit use of models for $F(r)$. These models incorporate prior information about the function $F(r)$ to be reconstructed. Of particular interest here are those models derived through minimum-weighted-norm (MWN) estimation; band-limited extrapolation is one special case.

2. BAND-LIMITED EXTRAPOLATION

We assume throughout this section that $F(r)$ is defined for real r and is nonzero only for $|r|\leqslant\Omega$, where $0<\Omega<\pi$. In addition, we assume that our data are $f(m), m=-M,\ldots,M$, that is, we have (possibly noisy) measurements of its Fourier transform $f(x)$. Because $\Omega<\pi$, the data are over-sampled; the Nyquist rate is $\Delta=\pi/\Omega>1$. The minimum-norm estimate F_{DFT} in Eq. (8) does not involve the value Ω and estimates $F(r)$ over the interval $[-\pi,\pi]$ associated with the actual sampling rate of unity. As simulations readily illustrated in Fig. 1, this DFT estimate can do a poor job of recovering the finer detail of $F(r)$ within its true support $[-\Omega,\Omega]$ because it wastes its limited degrees of freedom describing the values $F(r)=0$ that occur for r outside $[-\Omega,\Omega]$. In addition, if we simply restrict the DFT to variables r within $[-\Omega,\Omega]$ and set our estimate of $F(r)$ to zero outside, we have an estimate that is no longer consistent with the data. The goal of band-limited extrapolation is to find a function that is both consistent with the measured Fourier-transform data and supported on the interval $[-\Omega,\Omega]$.

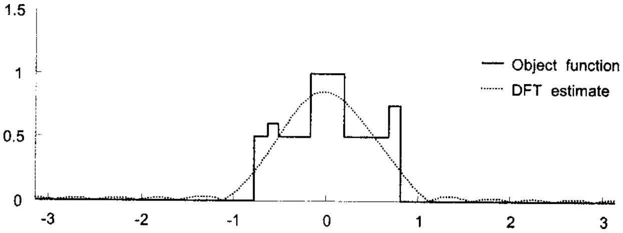

Fig. 1. DFT estimation from noiseless data.

Because $F(r)$ is zero outside the interval $[-\pi,\pi]$ it has a Fourier-series representation within $[-\pi,\pi]$:

$$F(r) = \sum_{m=-\infty}^{\infty} f(m)\exp(imr). \qquad (9)$$

The Gerchberg–Papoulis (GP) band-limited extrapolation method[1,2] is an iterative procedure for estimating those $F(m)$ for $|m|>M$. The procedure begins with zeros in place of all the $f(m)$ for $|m|>M$ and the data for the others. By use of those Fourier coefficients the resulting Fourier series is the DFT. The DFT is then truncated outside $[-\Omega,\Omega]$ and the Fourier coefficients of this new function are computed. These new coefficients no longer match the measured data for $|m|\leqslant M$ and so are replaced by the measured data; the other coefficients are left unchanged and the new Fourier series is formed. The resulting function of r is no longer zero outside $[-\Omega,\Omega]$ and so is truncated, as before. Repeating this, we obtain an iterative algorithm that converges to a function $H(r)$ that is both consistent with the measured data and supported on the interval $[-\Omega,\Omega]$. Its Fourier series has coefficients that extrapolate the measured data, hence the name of the procedure.

The GP algorithm as just described is not a practical method; it requires that we calculate an infinite set of Fourier coefficients at each step. One way around this is to discretize the problem and represent $F(r)$ as a finite vector. The iteration can then be performed using the fast Fourier transform; this is the actual GP algorithm as used in practice. There is a second way around the practical problem, however. As pointed out in Ref. 7 (see also Ref. 8, p. 209), the function $H(r)$ to which the theoretical GP method converges has the form

$$H(r) = \chi_{\Omega}(r)\sum_{n=-M}^{M} a_n \exp(inr), \qquad (10)$$

where $\chi_{\Omega}(r)=1$ for $|r|\leqslant\Omega$ and is zero otherwise, and the a_m are such as to make $H(r)$ consistent with the measured data. This means that the a_n satisfy the equations

$$f(m) = \sum_{n=-M}^{M} a_n \frac{\sin\left[\Omega(m-n)\right]}{\pi(m-n)}, \qquad (11)$$

for $n=-M,\ldots,M$; note that the value of the function $\sin(x)/x$ is defined by continuity to be one at $x=0$.

The Fourier transform of $F_{\mathrm{MDFT}}(r)$ is

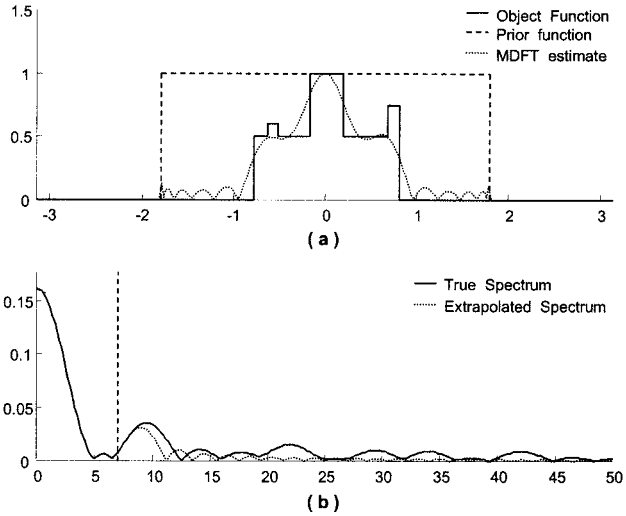

Fig. 2. MDFT estimate with $\Omega=1.8$ from noiseless data (a) in spatial domain, (b) in spectrum domain.

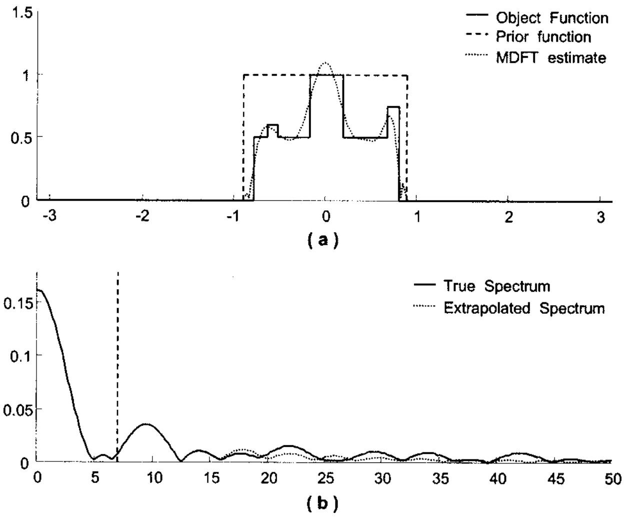

Fig. 3. MDFT estimate with $\Omega=0.9$ from noiseless data (a) in spatial domain, (b) in spectrum domain.

$$f\mathrm{MDFT}(x)=\sum_{n=-M}^{M}a_n\frac{\sin\left[\Omega(x-n)\right]}{\pi(x-n)}, \tag{12}$$

We can use the Fourier transform of the MDFT estimate as a data-extrapolation procedure, whereby $f(x)$ is estimated by $f_{\mathrm{MDFT}}(x)$. The behavior of this reconstruction method depends heavily on properties of the matrix B with entries $B_{mn}=\sin[\Omega(m-n)]/[\pi(m-n)]$, as we shall see.

This noniterative implementation of the theoretical GP algorithm, called the modified DFT (MDFT) in Ref. 7, can be viewed as a MWN solution to the reconstruction problem. The $H(r)=F_{\mathrm{MDFT}}(r)$ described by Eqs. (10) and (11) is the function consistent with the data that minimizes the energy over the interval $[-\Omega,\Omega]$, given by $\int^{\Omega}{}_{-\Omega}|H(r)|^2\mathrm{d}r$. It is also the optimal approximation of $F(r)$ of its form, in the sense that the coefficients a_n obtained using Eq. (11) minimize

$$\int^{\Omega}{}_{-\Omega}|F(r)-\sum_{n=-M}^{M}a_n\exp(inr)|^2\mathrm{d}r.$$

These iterative and noniterative methods are usually called superresolution techniques in the signal-processing literature. Similar methods applied in sonar and radar array processing are called superdirective methods.[9]

In Figs. 1–3 we see the improvement, both in resolution achieved by the MDFT, compared with the DFT, and in the accuracy of the extrapolated Fourier-transform values. The vertical dashed line in each spectrum figure indicates the boundary of the data support.

In particular, the form of the estimator in Eq. (10) is suggestive and leads to a more general MWN estimation procedure, called the prior DFT (PDFT) method (see below).

3. PRIOR DISCRETE FOURIER TRANSFORM

Prior information about the support of the function $F(r)$ is incorporated in the $F_{\mathrm{MDFT}}(r)$ estimator in Eq. (10) through the multiplicative factor $\chi_{\Omega}(r)$. If we have prior information about the shape of the function $|F(r)|$ we can incorporate this information in a function $P(r)\geqslant 0$ and replace $\chi_{\Omega}(r)$ with $P(r)$ as the first factor in the estimator. Because the second factor has the algebraic form of the DFT we call this new estimator the PDFT.[10,11] The PDFT estimate of $F(r)$ is then

$$F_{\mathrm{PDFT}}(r)=P(r)\sum_{n=-M}^{M}b_n\exp(inr), \tag{13}$$

with the b_n satisfying the equations

$$f(m)=\sum_{n=-M}^{M}b_np(m-n), \tag{14}$$

for $n=-M,\ldots,M$, and

$$p(x)=\frac{1}{2\pi}\int_{-\pi}^{\pi}P(r)\exp(-ixr)\mathrm{d}r. \tag{15}$$

Note that, if $P(r)=\chi_{\Omega}(r)$, then $p(x)=\sin(\Omega x)/(\pi x)$.

As in the case of the MDFT, the PDFT can be viewed as a data-extrapolation procedure. With the coefficients b_n determined by Eq. (14), the Fourier transform of $F_{\mathrm{PDFT}}(r)$ is

$$f_{\mathrm{PDFT}}(x)=\sum_{n=-M}^{M}b_np(x-n). \tag{16}$$

The data extrapolation is achieved here through the use of a model of $f(x)$ as a sum of translations of a positive-definite kernel function. In a recent paper[12] Poggio and Smale discuss the use of positive-definite kernels for interpolation, in the context of artificial intelligence and supervised learning.

The weighted energy of $F_{\mathrm{PDFT}}(r)$, given by $\int|F_{\mathrm{PDFT}}(r)|^2P(r)^{-1}\mathrm{d}r$, is the smallest among all functions defined on the support of $P(r)$ and consistent with the data; therefore, the PDFT estimate is a MWN estimate. In addition, the coefficients b_n found using Eq. (14) minimize

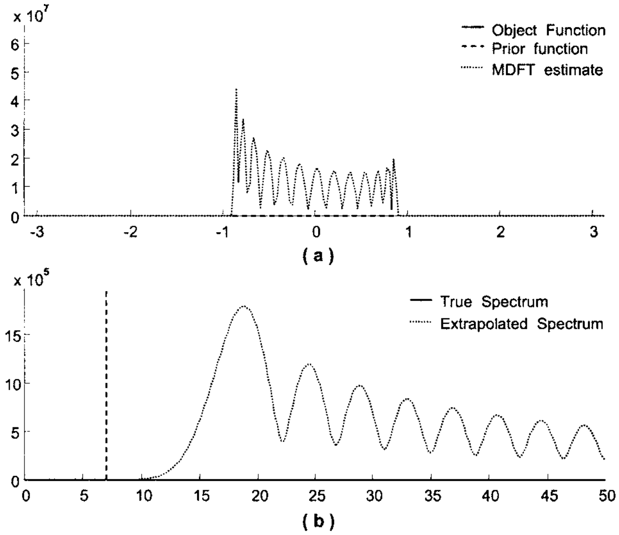

Fig. 4. MDFT estimate with $\Omega=0.9$ from noiseless data (the true support is between $-7\pi/8$ and $-3\pi/8$ (a) in spatial domain, (b) in spectrum domain.

$$\int_{-\pi}^{\pi} |F(r)-P(r)\sum_{n=-M}^{M} b_n \exp(inr)|^2 P(r)^{-1}\mathrm{d}r, \qquad (17)$$

so the PDFT estimate is the function of its form closest to $F(r)$, in the weighted-distance sense. Once again, the behavior of the estimation procedure will depend heavily on the properties of the matrix P whose entries are $P_{mn} = p(m-n)$.

Although we have discussed the MDFT and PDFT within the context of reconstructing $F(r)$ from Fourier-transform values, both procedures extend in an obvious manner to any linear-functional data and to multivariate r.

Of particular interest is the reconstruction of nonnegative functions $F(r)$ from finitely many Fourier-transform values. The PDFT cannot include a nonnegativity constraint. However, the PDFT can be viewed as a linearized approximation of the minimum cross-entropy solution, which does impose nonnegativity. If we minimize the cross-entropy

$$\int H(r)\log\frac{H(r)}{P(r)} + P(r) - H(r)\mathrm{d}r$$

over all nonnegative $H(r)$ consistent with the Fourier-transform data, the solution is

$$F_{\mathrm{MCE}}(r) = P(r)\exp\left[\sum_{n=-M}^{M} c_n \exp(inr)\right],$$

with the c_n chosen for consistency with the data.[13–15] When we replace the exponential factor with a first-order linear approximation we get the PDFT.

The goal of estimating $F(r)$ from finite linear-functional data is somewhat paradoxical. The finite data we have tell us nothing, by themselves, about the values $f(n)$ we have not measured. Using the MDFT, we can define $f(M+1)$ any way we wish and still construct an $F_{\mathrm{MDFT}}(r)$ supported on the interval $[-\Omega,\Omega]$, and consistent with the original data and with this chosen value of $f(M+1)$. In a similar sense our finite data also tell us nothing about the value of Ω; we can select any interval $[a,b]$ and find a function $H(r)$ supported on $[a,b]$ whose $h(x)$ is consistent with the data.

But this is not quite the whole story; finite data cannot rule out anything, but they can suggest strongly that certain things are false. For example, let us select an interval $[a,b]$ disjoint from $[-\Omega,\Omega]$, and find the function $H(r)$ consistent with the data, that is, with

$$f(m) = \int_a^b H(r)\exp(-imr)\frac{\mathrm{d}r}{2\pi},$$

for $m=-M,\dots,M$, for which the energy over $[a,b], \int_a^b |H(r)|^2\mathrm{d}r$, is minimum. Then this function $H(r)$ will probably have large energy compared with that of the MDFT; that is, the integral $\int_a^b |H(r)|^2\mathrm{d}r$ will be much larger than $\int_{-\Omega}^{\Omega} |F_{\mathrm{MDFT}}(r)|^2\mathrm{d}r$, as clearly shown in Fig. 4. We can use this fact to help us decide if we have chosen a good value for Ω. In Ref. 16 this same idea was used to obtain an iterative algorithm for solving the phase-retrieval problem.

4. SENSITIVITY TO NOISE AND MODEL ERROR

To use the MDFT we need the data to be oversampled and we need a decent estimate of the true support of the function $F(r)$. The more oversampled the data and the more accurately we know the true support, the greater the improvement in resolution, but also the greater the sensitivity to noise and model error. Our goal in this section is to see why this is the case.

The matrix B used in the MDFT has the entries $B_{mn} = \sin[\Omega(m-n)]/[\pi(m-n)]$, with $B_{mm}=\Omega/\pi$. Loosely speaking, B has $(\Omega/\pi)(2M+1)$ eigenvalues near one and the remaining eigenvalues near zero, as shown in Fig. 5. Solving Eqs. (11) is therefore an ill-conditioned problem, as Ω grows smaller. For the remainder of this section we denote by $\lambda_1 > \lambda_2 > \cdots > \lambda_{2M+1} > 0$ the eigenvalues of B, with associated orthonormal eigenvectors $u^n = (u^n_{-M},\dots,u^n_M)^T$, for $n=1,\dots,2M+1$. The matrix B then has the form

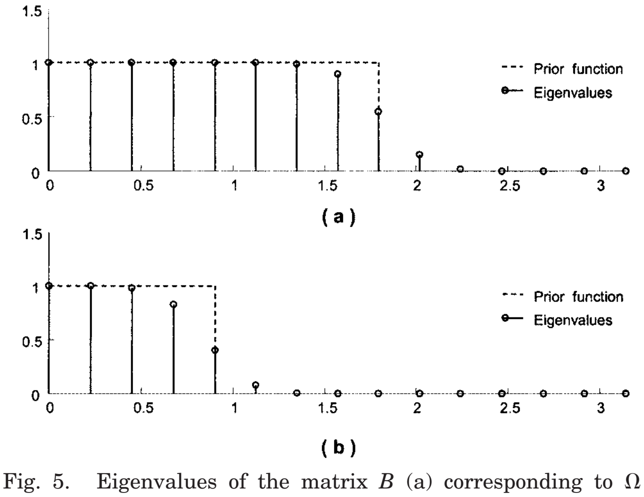

Fig. 5. Eigenvalues of the matrix B (a) corresponding to $\Omega = 1.8$, (b) corresponding to $\Omega=0.9$.

$$B = \sum_{n=1}^{2M+1} \lambda_n u^n (u^n)^\dagger, \tag{18}$$

so that

$$B^{-1} = \sum_{n=1}^{2M+1} \lambda_n^{-1} u^n (u^n)^\dagger. \tag{19}$$

We also denote by $U_n(r)$ the functions

$$U_n(r) = \sum_{m=-M}^{M} u_m^n \exp(imr).$$

Since the eigenvectors are orthonormal we have

$$\int_{-\pi}^{\pi} U_k(r)\overline{U_n(r)}\mathrm{d}r = 0,$$

for $k \neq n$ and

$$\int_{-\pi}^{\pi} U_n(r)\overline{U_n(r)}\mathrm{d}r = \int_{-\pi}^{\pi} |U_n(r)|^2 \mathrm{d}r = 1.$$

Also

$$\int_{-\Omega}^{\Omega} |U_n(r)|^2 \mathrm{d}r = (u^n)^\dagger Q u^n = \lambda_n.$$

Therefore, λ_n is the proportion of the energy of $U_n(r)$ for r in the interval $[-\Omega,\Omega]$. The function $U_1(r)$ is the most concentrated in that interval, while $U_{2M+1}(r)$ is the least. In Fig. 6 the typical behaviors of $U_n(r)$ for an example with 15 data are demonstrated.

The function $U_1(r)$ has a single large main lobe and no zeros within $[-\Omega,\Omega]$. The function $U_2(r)$, being orthogonal to $U_1(r)$, but still largely concentrated within $[-\Omega,\Omega]$, has a single zero in that interval. Each succeeding function has one more zero within $[-\Omega,\Omega]$ and is somewhat less concentrated there than its predecessors. At the other extreme, the function $U_{2M+1}(r)$ has $2M$ zeros in $[-\Omega,\Omega]$, is near zero throughout that interval, and is concentrated mainly outside the interval.

Because the eigenvectors are orthonormal the DFT estimate in Eq. (8) can be written in terms of these $U_n(r)$:

$$F_{\mathrm{DFT}}(r) = \sum_{n=1}^{2M+1} \left[(u^n)^\dagger d\right] U_n(r). \tag{20}$$

Similarly, using Eq. (19), the MDFT estimate in Eq. (10) can be written as

$$F_{\mathrm{MDFT}}(r) = \sum_{n=1}^{2M+1} \lambda_n^{-1} \left[(u^n)^\dagger d\right] U_n(r). \tag{21}$$

Comparing Eqs. (20) and (21) we see that the MDFT places greater emphasis on those $U_n(r)$ corresponding to larger values of n. These are the functions least concentrated within the interval $[-\Omega,\Omega]$, but they are also those with the greatest number of zeros within that interval. That means that these functions are much more oscillatory within $[-\Omega,\Omega]$ and better suited to resolve closely spaced peaks in $F(r)$. Because the inner product $(u^n)^\dagger d$ can be written as

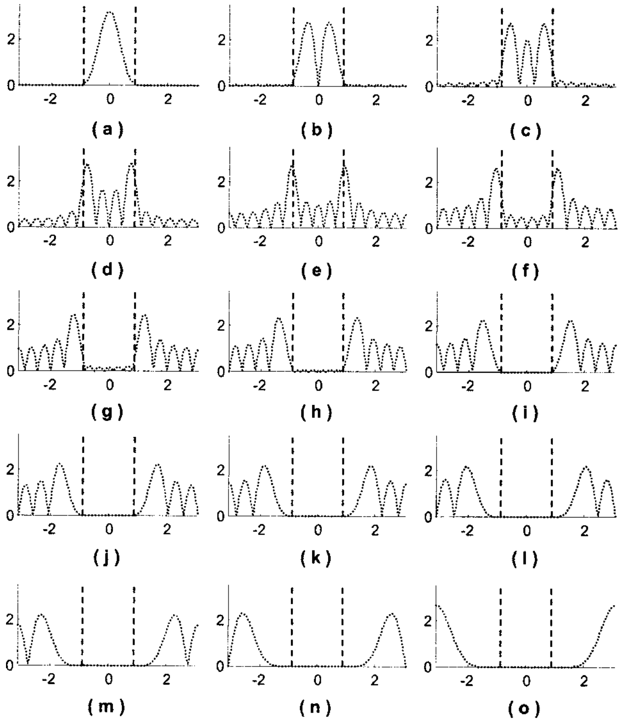

Fig. 6. Example showing the absolute values of $U_n(r)$ for $n=1, 2,\ldots,15$, where their corresponding eigenvalues are (a) 1.00, (b) 9.99×10^{-1}, (c) 9.82×10^{-1}, (d) 8.29×10^{-1}, (e) 4.02×10^{-1}, (f) 7.82×10^{-2}, (g) 6.83×10^{-3}, (h) 3.52×10^{-4}, (i) 1.21×10^{-5}, (j) 2.91×10^{-7}, (k) 4.85×10^{-9}, (l) 5.53×10^{-11}, (m) 4.12×10^{-13}, (n) 1.72×10^{-15}, (o) 4.28×10^{-18}. Two vertical dashed lines indicate the boundaries of the prior support.

$$(u^n)^\dagger d = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(r)\overline{U_n(r)}\mathrm{d}r,$$

this term will be relatively small for the larger values of n and so the product $\lambda_n^{-1}[(u^n)^\dagger d]$ will not be excessively large, provided that we have selected Ω properly. If Ω is too small and the support of $F(r)$ extends beyond the interval $[\Omega,\Omega]$, then the term $(u^n)^\dagger d$ will not be as small and the product $\lambda_n^{-1}[(u^n)^\dagger d]$ will be too large. This is what happens when there is noise in the data; the object function corresponding to the noisy data is not simply $F(r)$ but contains a component that can be viewed as extending throughout the interval $[-\pi,\pi]$, as shown in Fig. 7.

To reduce the sensitivity to noise while not sacrificing resolution, we regularize. The simplest way to do this is to add a small positive quantity, $\epsilon > 0$, to each of the diagonal elements of the matrix B. This is equivalent to modifying the MDFT to a PDFT in which the prior $P(r)$ consists of two components, one the original $\chi_\Omega(r)$, the second a small positive multiple of $\chi_\pi(r)$. The effect of this regularization is to increase each of the λ_n by ϵ without altering the eigenvectors or the $U_n(r)$. Since we now have $1/(\lambda_n+\epsilon)$ instead of the (potentially) much larger $1/\lambda_n$ in Eq. (21), the sensitivity to noise and to poor selection of the Ω is reduced, as shown in Fig. 8. At the same time, however, we have reduced the importance for the MDFT of the $U_n(r)$ for larger values of n; this will lead to a loss of resolution and an MDFT that behaves like the DFT if ϵ is too large, as shown in Fig. 9. Selecting the proper ϵ is a bit of an art;

it will certainly depend on what the eigenvalues are and on the signal-to-noise ratio. The eigenvalues, in turn, will depend on the ratio Ω/π.

We have focused here on a particular case in which the variable r is one-dimensional and the prior $P(r)=\chi_{\Omega}(r)$ describes only the support of the object function $F(r)$. For other choices of $P(r)$, the eigenvalues of the corresponding matrix P are similarly behaved and ill conditioning is still an important issue, although the distribution of the eigenvalues may be somewhat different. For two-dimensional r the support of $F(r)$ can be described using a prior $P(r)$ that is nonzero on a rectangle, on a circle, on an ellipse, or on a more general region. For each of those choices, the corresponding matrix will have eigenvalues that decay toward zero, but perhaps at different rates.

5. PHASE PROBLEM

In optical image processing and elsewhere we find that we are unable to measure the complex values of the Fourier transform $f(x_m)$, only the magnitudes $|f(x_m)|$. Estimating $F(r)$ from these magnitude-only values is called the phase problem.[17–21] Such problems can arise in optical imaging through turbulent atmosphere.[22] One solution to the phase problem in crystallography led to a Nobel Prize in 1985 for Jerome Karle.

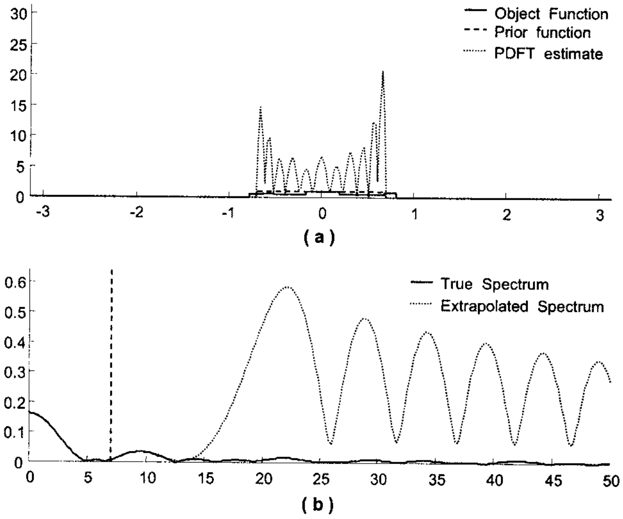

Fig. 7. PDFT estimate with $\Omega=0.7$ (smaller than the true support) from noiseless data (a) in spatial domain, (b) in spectrum domain.

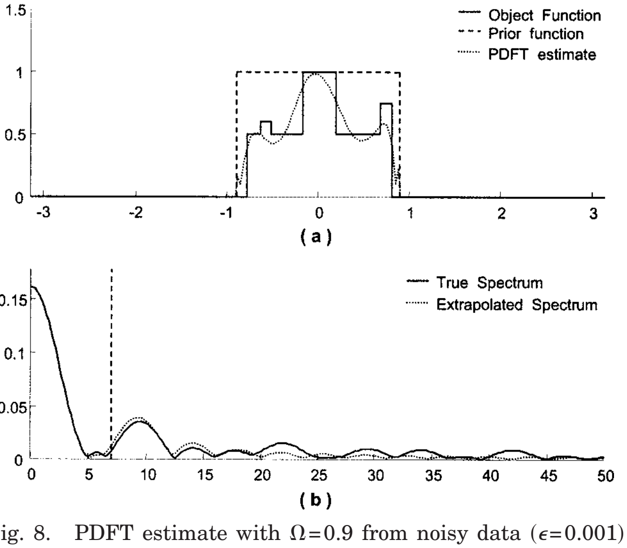

Fig. 8. PDFT estimate with $\Omega=0.9$ from noisy data ($\epsilon=0.001$) (a) in spatial domain, (b) in spectrum domain.

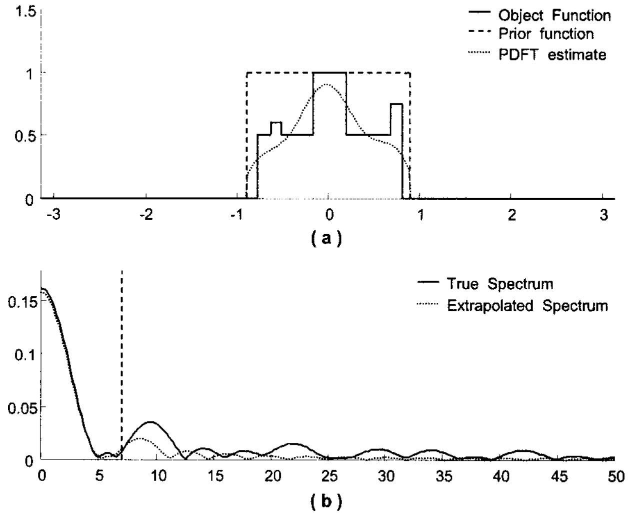

Fig. 9. PDFT estimate with $\Omega=0.9$ from noisy data ($\epsilon=0.1$) (a) in spatial domain, (b) in spectrum domain.

Assume now that $F(r)=0$ for $|r|>\Omega$. We can select an arbitrary collection of phases θ_m to combine with the magnitudes, to form the complex pseudodata $|f(x_m)|e^{i\theta_m}$. If we have some idea of the proper choice of Ω, we calculate the estimate $F_{\mathrm{MDFT}}(r)$ corresponding to the pseudodata and again monitor the energy integral. For good choices of the phases, the energy should not be too large, while, for inappropriate choices, the energy should be much larger, particularly if the data are oversampled. The reconstruction process can be implemented as an iterative optimization procedure, in which we select a new collection of phases at each step in such a way as to reduce the energy in the band-limited extrapolation that results. In Ref. 16 we show how to do this in an efficient manner. When the extrapolation energy is sufficiently small, the resulting estimate is typically acceptable, particularly when the data are oversampled.

When we have only magnitude measurements, we can at least be sure that if $|f(x_m)|=0$ then $f(x_m)=0$. This suggests that we might try to estimate the function $F(r)$ from the zeros of its Fourier transform. In Ref. 23 we showed that this approach has some promise for solving the phase problem.

6. CALCULATING THE PRIOR DISCRETE FOURIER TRANSFORM

When the data set is large, as usually happens in multidimensional problems such as image reconstruction, solving Eqs. (11) and (14) is sometimes done iteratively. Nevertheless, these algorithms still differ from the GP method based on a discretized model in that we are still extrapolating infinitely many values of $f(n)$ and obtaining a continuous-function estimate; we are just doing it using a finite parameter model.

Constructing the matrix P used in Eq. (14) can be difficult when the data sets are large. In such cases we can employ an iterative discrete implementation of the PDFT, the DPDFT, which allows us to avoid having to form this large matrix.[24] The DPDFT reconstructs a finite-vector approximation of the function $F(r)$. The linear-functional data are represented as inner products of this vector with finitely many known vectors. The number of data points we have is smaller than the dimension of the discretized object, so the reconstruction problem, which is now a system of linear equations to be solved, is still underdetermined. We use a discretized version of the reciprocal of the prior function $P(r)$ to determine the weights and then calculate a MWN solution to the underdetermined system of equations. This is done using the algebraic reconstruction technique.[13]

7. PRIOR DISCRETE FOURIER TRANSFORM AND OPTIMAL LINEAR DETECTION

The problem of detecting a signal in additive noise uses the following model. The data vector $z=(z_1,\dots,z_N)^T$ is assumed to be the sum

$$z = \gamma s + q$$

of a signal component γs, where $s=(s_1,\dots,s_N)^T$ and $\gamma>0$, and a noise component $q=(q_1,\dots,q_N)^T$ with mean $E(q)=0$ and covariance matrix $E(qq^\dagger)=Q$.[25] The estimation problem is to estimate γ, given s and Q. The detection problem is to decide if γ is zero or not.

In certain applications s is not known exactly, but is assumed to be a member of a parametrized family. For example, in the problem of detecting a sinusoidal component at an unknown frequency ω we take $e(\omega)$ to be the column vector with entries $e(\omega)_m=\exp(imw)$. For fixed ω the optimal linear filter for estimating γ is

$$b = \frac{1}{e(\omega)^\dagger Q^{-1} e(\omega)} Q^{-1} e(\omega),$$

and the optimal estimate of γ is

$$\hat{\gamma} = b^\dagger z = \frac{1}{e(\omega)^\dagger Q^{-1} e(\omega)} e(\omega)^\dagger Q^{-1} z.$$

The factor $1/e(\omega)^\dagger Q^{-1} e(\omega)$ can be viewed as an estimate of the power spectrum associated with the covariances in Q,[25] while the factor $e(\omega)^\dagger Q^{-1} z$ has the form of a DFT. Comparing this estimate of γ with the PDFT estimate of $F(r)$ we see that the first factor, $1/e(\omega)^\dagger Q^{-1} e(\omega)$, is playing a role analogous to that played by $P(r)$, the matrix Q is analogous to the matrix P, and the factor $e(\omega)^\dagger Q^{-1} z$ corresponds to the sum that appears as the second factor in the PDFT. Although the matrix Q need not be Toeplitz, as P always is, the correspondence is interesting.

8. PRIOR DISCRETE FOURIER TRANSFORM AND WIENER FILTER APPROXIMATION

Suppose now that the discrete stationary random process to be filtered is the doubly infinite sequence $\{z_n=s_n+q_n\}_{n=-\infty}^{\infty}$, where $\{s_n\}$ is the signal component with autocorrelation function $r_s(k)=E(s_{n+k}\overline{s_n})$ and power spectrum $R_s(\omega)$ defined for ω in the interval $[-\pi,\pi]$, and $\{q_n\}$ is the noise component with autocorrelation function $r_q(k)$ and power spectrum $R_q(\omega)$ defined for ω in $[-\pi,\pi]$. We assume that for each n the random variables s_n and q_n have mean zero and that the signal and noise are independent of one another. Then the autocorrelation function for the signal-plus-noise sequence $\{z_n\}$ is

$$r_z(n) = r_s(n) + r_q(n)$$

for all n, and

$$R_z(\omega) = R_s(\omega) + R_q(\omega)$$

is the signal-plus-noise power spectrum.

Let $h=\{h_k\}_{k=-\infty}^{\infty}$ be a linear filter with transfer function

$$H(\omega) = \sum_{k=-\infty}^{\infty} h_k e^{ik\omega},$$

for ω in $[-\pi,\pi]$. Given the sequence $\{z_n\}$ as input to this filter, the output is the sequence

$$y_n = \sum_{k=-\infty}^{\infty} h_k z_{n-k}. \tag{22}$$

The goal of Wiener filtering is to select the filter h so that the output sequence $\{y_n\}$ approximates the signal sequence $\{s_n\}$ as well as possible. Specifically, we seek h so as to minimize the expected squared error, $E(|y_n-s_n|^2)$, which, because of stationarity, is independent of n. Minimizing $E(|y_n-s_n|^2)$ with respect to the function $H(\omega)$ leads to the equation

$$R_z(\omega) H(\omega) = R_s(\omega),$$

so that the transfer function of the optimal filter is

$$H(\omega) = R_s(\omega)/R_z(\omega).$$

The Wiener filter is then the sequence $\{h_k\}$ of the Fourier coefficients of this function $H(\omega)$.

Since $H(\omega)$ is a nonnegative function of ω, therefore real-valued, its Fourier coefficients h_k will be conjugate symmetric; that is, $h_{-k}=\overline{h_k}$. Therefore, the Wiener filter is not causal; this poses a problem when the random process z_n is a discrete time series, with z_n denoting the measurement recorded at time n. To remedy this we can obtain the best causal approximation of the Wiener filter h.

Even having a causal filter does not completely solve the problem, since we would have to record and store the infinite past. Instead, we can decide to use a filter $f=\{f_k\}_{k=-\infty}^{\infty}$ for which $f_k=0$ unless $-K\leqslant k\leqslant L$ for some positive integers K and L. This means we must store L values and wait until time $n+K$ to obtain the output for time n. Such a linear filter is a finite-memory–finite-delay filter, also called a finite-impulse-response (FIR) filter. Given the input sequence $\{z_n\}$ the output of the FIR filter is

$$v_n = \sum_{k=-K}^{L} f_k z_{n-k}.$$

To obtain such an FIR filter f that best approximates the Wiener filter, we find the coefficients f_k that minimize the quantity $E(|y_n-v_n|^2)$, or, equivalently,

$$\int_{-\pi}^{\pi} |H(\omega) - \sum_{k=-K}^{L} f_k e^{ik\omega}|^2 R_z(\omega)\mathrm{d}\omega. \tag{23}$$

The orthogonality principle tells us that the optimal coefficients must satisfy the equations

$$r_s(m) = \sum_{k=-K}^{L} f_k r_z(m-k), \quad \text{for } -K \leqslant m \leqslant L. \tag{24}$$

In Ref. 26 it was pointed out that the minimization in relation (23) is analogous to that in relation (17), with $R_z(\omega)$ and $R_s(\omega)$ playing the roles of $P(r)$ and $F(r)$, respectively. In the PDFT we do not require that $F(r)$ be nonnegative-valued, however. If we switch the roles, viewing $R_z(\omega)$ and $R_s(\omega)$ as $F(r)$ and $P(r)$, respectively, we obtain nonlinear reconstruction methods containing, as a particular case, the Burg entropy-maximization estimator for the case of nonnegative $F(r)$.[27]

9. CONCLUSIONS

In summary, an important concern in image reconstruction is the validity and reliability of the resulting image. When one has only limited, sampled noisy data, it makes no sense to consider image restoration or superresolution in the absence of a model that incorporates prior knowledge of the scene or target being imaged. The incorporation of prior knowledge has deep implications because it impacts how one might interpret the information capacity of an optical imaging system while fixing the numbers of degrees of freedom that the system possesses. Image reconstruction techniques also demand measures of image quality; these are very difficult to agree upon and yet are needed and important. By offering a more complete view of the methods for image restoration and data extrapolation here, and in showing the relationships that exist between such methods, we are laying a foundation to address the fundamental problem of assessing the information content of data and of the resulting processed image.

We have shown that improved resolution in reconstructing an image from limited, noisy data can be achieved through the use of minimum-weighted-norm (MWN) data extrapolation. The weighted norm is chosen to incorporate prior knowledge of object support or overall shape. The resulting object-function estimate is a data-consistent model that can be viewed as extrapolating the data, thereby improving resolution. To understand precisely how improved resolution is achieved, why the reconstruction can be sensitive to noise in the data, and how this sensitivity is reduced, we have examined the eigenvectors and eigenvalues of the Gramian matrix describing the data-collection process. We found that the eigenvectors associated with the smallest eigenvalues are responsible both for the improved resolution and the greater sensitivity. Regularization enables us to maintain higher resolution while reducing sensitivity. The improvement in reconstruction is observed in object space, where the reconstruction is compared to the original object, as well as in data space, where we can judge success in extrapolating beyond the data window. The MWN paradigm for reconstruction includes band-limited extrapolation of Fourier-transform data; is closely related to Wiener-filter approximation, which suggests a nonlinear reconstruction procedure leading to entropy-maximization techniques; and provides a unifying model for resolution enhancement and data extrapolation.

Hsin M. Shieh's e-mail address is hmshieh@fcu.edu.tw.

REFERENCES

1. R. W. Gerchberg, "Super-restoration through error energy reduction," Opt. Acta **21**, 709–720 (1974).
2. A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," IEEE Trans. Circuits Syst. **22**, 735–742 (1975).
3. J. Burg, "Maximum entropy spectral analysis," presented at the 37th Annual Society of Exploration Geophysicists Meeting, Oklahoma City, Oklahoma, July 1967.
4. J. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra," Geophysics **37**, 375–376 (1972).
5. J. Burg, "Maximum Entropy Spectral Analysis," Ph.D. thesis (Stanford University, Stanford, California, 1975).
6. M. Bertero, "Sampling theory, resolution limits and inversion methods," in *Inverse Problems in Scattering and Imaging*, M. Bertero and E. R. Pike, eds. (Malvern Physics Series, Adam Hilger, IOP Publishing, 1992), pp. 71–94.
7. C. L. Byrne and R. M. Fitzgerald, "A unifying model for spectrum estimation," presented at the Rome Air Development Center Workshop on Spectrum Estimation, Griffiss Air Force Base, Rome, New York, October 3–5, 1979.
8. H. Stark and Y. Yang, *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets and Optics* (Wiley, 1998).
9. H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," J. Acoust. Soc. Am. **54**, 771–785 (1973).
10. C. L. Byrne and R. M. Fitzgerald, "Reconstruction from partial information, with applications to tomography," SIAM J. Appl. Math. **42**, 933–940 (1982).
11. 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).
12. T. Poggio and S. Smale, "The mathematics of learning: dealing with data," Not. Am. Math. Soc. **50**, 537–544 (2003).
13. R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," J. Theor. Biol. **29**, 471–481 (1970).
14. C. L. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," IEEE Trans. Image Process. **IP-2**, 96–103 (1993).
15. C. L. Byrne, "Erratum and addendum to 'Iterative image reconstruction algorithms based on cross-entropy minimization'," IEEE Trans. Image Process. **IP-4**, 225–226 (1995).
16. C. L. Byrne and M. A. Fiddy, "Estimation of continuous object distributions from Fourier magnitude measurements," J. Opt. Soc. Am. A **4**, 412–417 (1987).
17. M. A. Fiddy, "The phase retrieval problem," in *Inverse Optics*, A. Devaney, ed., Proc. SPIE **413**, 176–181 (1983).
18. J. C. Dainty and M. A. Fiddy, "The essential role of prior knowledge in phase retrieval," Opt. Acta **31**, 325–330 (1984).
19. J. Fienup, "Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint," J. Opt. Soc. Am. A **4**, 118–123 (1987).
20. R. Lane, "Recovery of complex images from Fourier magnitude," Opt. Commun. **63**, 6–10 (1987).

21. J. Cederquist, J. Fienup, C. Wackerman, S. Robinson, and D. Kryskowski, "Wave-front phase estimation from Fourier intensity measurements," J. Opt. Soc. Am. A **6**, 1020–1026 (1989).
22. J. Fienup, "Space object imaging through the turbulent atmosphere," Opt. Eng. (Bellingham) **18**, 529–534 (1979).
23. C.-W. Liao, M. A. Fiddy, and C. L. Byrne, "Imaging from the zero locations of far-field intensity data," J. Opt. Soc. Am. A **14**, 3155–3161 (1997).
24. H. M. Shieh, M. E. Testorf, C. L. Byrne, and M. A. Fiddy, "Iterative image reconstruction using prior knowledge," submitted to J. Opt. Soc. Am. A.
25. C. L. Byrne, *Signal Processing: A Mathematical Approach* (AK Peters, 2005).
26. C. L. Byrne and M. A. Fiddy, "Images as power spectra; reconstruction as Wiener filter approximation," Inverse Probl. **4**, 399–409 (1988).
27. C. L. Byrne and R. M. Fitzgerald, "Spectral estimators that extend the maximum entropy and maximum likelihood methods," SIAM J. Appl. Math. **44**, 425–442 (1984).