Inverse Problems 4 (1988) 399–409. Printed in the UK

Images as power spectra; reconstruction as a Wiener filter approximation

Charles L Byrne† and Michael A Fiddy‡

† Department of Mathematics, University of Lowell, Lowell, MA 01854, USA
‡ Department of Electrical Engineering, University of Lowell, Lowell, MA 01854, USA

Received 10 August 1987, in final form 17 September 1987

Abstract. The problem of reconstructing a non-negative function from finitely many values of its Fourier transform is a problem of approximating one function by another and, as such, is analogous to the design of finite-impulse-response approximations to the Wiener filter. Using this analogy we obtain reconstruction methods that are computationally simpler approximations of entropy-based procedures. Our linear estimators allow for the inclusion of prior information about oversampling rate, i.e. support information, as well as other prior knowledge of the general shape of the object. Our nonlinear methods, designed to recover spiky objects, make use of prior information about non-uniformity in the background to avoid bias in the estimation of peak locations.

1. Introduction

The problem of reconstructing a non-negative function $f(a, b)$ of two real variables from finitely many values of its Fourier transform (FT) arises in a number of applications. These include recovering an image or object distribution from its spectrum, a power spectrum from its autocorrelation function, a distribution of energy in bearing from cross-sensor correlations or a bivariate probability density from its characteristic function. In many cases of interest the function $f(a, b)$ is non-negative and we shall make that assumption here. The problem of limited data can arise for a variety of reasons: to remove the effects of a known convolution-filter degradation one can divide by the filter transfer function in the spectral domain, but must avoid dividing by small quantities; in the case of sensor array processing one is limited to spatial separations provided by the array geometry.

Because the data are finite there will always be infinitely many reconstructions consistent with the data values. Some of these reconstructions will be reasonably good, while others will not; the data constraints, by themselves, will not automatically lead to a good reconstruction unless the number of data values is large. There are several methods based on minimising some cost function, such as entropy; one problem with such approaches is that it is not always clear just how the resulting reconstruction is related to the correct answer. The methods we present here are based on the theory of best approximation in Hilbert space and make clear how the reconstruction is related to the original, unknown, correct object function.

0266-5611/88/020399+11 $02.50

In order to obtain a good reconstruction it is necessary to incorporate additional information about the function being reconstructed; in some cases support information is used or positivity is enforced; in others upper and lower bounds are employed. In a number of methods one uses a prior estimate of $f(a, b)$; this is done in cross-entropy minimisation and it has been shown that the Burg maximum entropy method employs (tacitly) a uniform prior estimate [1]. In earlier work [1] we extended the Burg method for the one-dimensional case to incorporate other prior information and considered numerical examples; our purpose here is to provide a theoretical justification for that procedure, based on analogy with the design of approximate Wiener filters. This approach allows for generalisation to higher dimensions, which we also consider. We present both linear and nonlinear methods.

One of the difficulties with methods that incorporate prior knowledge is that it is not always clear what the prior estimate is estimating. As our development here reveals, the role of the prior estimate is different in the linear and the nonlinear methods. The chief virtue of the Wiener filter design approach is that it gives us a clear picture of the role being played by the prior estimate. Loosely speaking, in the case of linear methods the prior estimate is an estimate of the whole function associated with the data, including any noise background component, while for nonlinear methods (to be used mainly for high-resolution reconstruction of spiky objects) the prior should estimate the smooth component only; linear methods such as superresolution become unstable when the prior estimates only the support-limited object and ignores any noise background, while nonlinear methods that employ a uniform prior estimate, such as Burg's maximum entropy [1], become unstable when the background is non-uniform.

In reference [1] we presented methods for the reconstruction of 1D objects from limited FT data. Here we extend these methods to 2D objects and present a unified interpretation of both cases in terms of the finite-impulse-response approximation to a Wiener filter; in this way we are led naturally to the particular Hilbert spaces used earlier [1], where they may have seemed somewhat *ad hoc*.

It is important to note that, while the Wiener filter and its finite-impulse-response approximations are used to motivate the reconstruction methods presented here, we do not employ a statistical model for the functions being reconstructed.

Throughout the paper we denote by $f(a, b)$ a non-negative function supported on the square $|a|\leqslant\pi$, $|b|\leqslant\pi$. The Fourier series representation for the function f on $|a|$, $|b|\leqslant\pi$ is

$$f(a, b)=\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} F(m, n)\exp(\mathrm{i}ma+\mathrm{i}nb). \tag{1.1}$$

We assume that we have the data $F(m, n)$ for $|m|\leqslant M$, $|n|\leqslant N$, from which we are to reconstruct (estimate) $f(a, b)$. A commonly used estimate is the truncated Fourier series (also sometimes referred to as the 'discrete Fourier transform' because the summation replaces the integration); for $|a|$, $|b|\leqslant\pi$ define the DFT(a, b) to be

$$\mathrm{DFT}(a, b)=\sum_{-M}^{M}\sum_{-N}^{N} F(m, n)\exp(\mathrm{i}ma+\mathrm{i}nb). \tag{1.2}$$

Note that the DFT is defined here as a function of two continuous variables; one sometimes sees 'DFT' used to denote a sampled version of (1.2).

In many applications the DFT will be unsatisfactory, particularly if the function f is supported on a smaller interval within $[-\pi, \pi]\times[-\pi, \pi]$, or if f is a spiky function and the number of data values is not large. The DFT is consistent with the original data, in the sense that the Fourier series of DFT (a, b) has the data values in the proper positions, but may fail to be non-negative or to resolve closely spaced peaks. The objective of high-resolution processing is to employ prior information to obtain better reconstructions than the DFT.

For completeness we discuss the Wiener filter and its approximations, for the 1D case (for notational simplicity), and then discuss the use of Wiener filter approximation for the reconstruction of 1D functions. We then turn to the 2D case, the main differences stemming from difficulties in extending the concept of 'causal filter'. Finally, we discuss briefly the connections between these methods and those based on the minimisation of cross-entropy.

2. Wiener filtering: the one-dimensional case

The Wiener filter [2] is a procedure designed to produce as output an estimate of 'signal' when presented with input 'signal plus noise'. Assume that $\{s(n)\}$, $\{u(n)\}$ are independent, mean-zero stationary random sequences with autocorrelation functions $r_{ss}(m)$, $r_{uu}(m)$ and power spectra $R_{ss}(\alpha)$, $R_{uu}(\alpha)$, respectively, with $|\alpha|\leqslant\pi$; the sequence $\{r_{ss}(m)\}$ are the Fourier coefficients of $R_{ss}(\alpha)$, and similarly for $R_{uu}(\alpha)$. The Wiener filter is a doubly infinite sequence $\{h(k)\}$ designed as follows: given the random sequence $x(n)=s(n)+u(n)$ as input and $y(n)$ as output, where

$$y(n)=\sum_{-\infty}^{\infty} h(k)\, x(n-k) \qquad -\infty<n<\infty \tag{2.1}$$

select $\{h(k)\}$ so as to minimise the expected mean square error, $E|s(n)-y(n)|^2$. The well known result is that the optimal choice of sequence $\{h(k)\}$ is the sequence of Fourier coefficients of the function $H(\alpha)=R_{ss}(\alpha)/R_{xx}(\alpha)$, where $R_{xx}(\alpha)=R_{ss}(\alpha)+R_{uu}(\alpha)$, and $H(\alpha)$ is defined to be zero if $R_{xx}(\alpha)=0$.

The Wiener filter is not a causal filter, since we do not have $h(k)=0$ for $k<0$. We can ask for the causal filter $\{g(k)\}$ ($g(k)=0$, $k<0$) that best approximates the Wiener filter, or, going further, the finite-impulse-response filter $\{d(k)\}$ ($d(k)=0$ unless $K\leqslant k\leqslant L$) that best approximates the Wiener filter. To obtain these optimal approximations we minimise the expected mean square difference between the outputs of the Wiener filter and the approximation. These optimisation problems are equivalent to best approximations in a Hilbert space with weighted inner product.

To obtain the best causal approximation to the Wiener filter we minimise the distance

$$\int_{-\pi}^{\pi} \left| H(\alpha)-\sum_{k=0}^{\infty} g(k)\exp(ik\alpha)\right|^2 R_{xx}(\alpha)\,\mathrm{d}\alpha \tag{2.2}$$

over all causal sequences $\{g(k)\}$. Similarly, to obtain the optimal finite-impulse-response filter with support $K\leqslant k\leqslant L$ we minimise the error

$$\int_{-\pi}^{\pi} \left| H(\alpha)-\sum_{k=K}^{L} d(k)\exp(ik\alpha)\right|^2 R_{xx}(\alpha)\,\mathrm{d}\alpha \tag{2.3}$$

over all finite sequences $\{d(k)\}$. From the orthogonality principle in Hilbert space [3] it follows that the optimal $\{g(k)\}$ and $\{d(k)\}$ must satisfy the following systems of linear equations:

$$r_{ss}(m) = \sum_{k=0}^{\infty} g(k)\, r_{xx}(m-k) \qquad m \geqslant 0 \tag{2.4}$$

$$r_{ss}(m) = \sum_{k=K}^{L} d(k)\, r_{xx}(m-k) \qquad K \leqslant m \leqslant L. \tag{2.5}$$

Equations (2.4) are the discrete Wiener–Hopf equations. Having solved these equations we write, for $|\alpha| \leqslant \pi$,

$$G(\alpha) = \sum_{k=0}^{\infty} g(k) \exp(ik\alpha) \qquad D_K^L(\alpha) = \sum_{k=K}^{L} d(k) \exp(ik\alpha). \tag{2.6}$$

It is worth noting that the best finite impulse-response filter approximating $\{g(k)\}$, for $0 \leqslant K \leqslant k \leqslant L$, is the $\{d(k)\}$ above; that is, this choice of $d(k)$ minimises the approximation error

$$\int_{-\pi}^{\pi} \left| G(\alpha) - \sum_{k=K}^{L} d(k) \exp(ik\alpha) \right|^2 d\alpha \tag{2.7}$$

viewed as a function of the $d(k)$. Therefore, the function $D_K^L(\alpha)$ is simultaneously the best approximation, of its form, of $H(\alpha)$ and of $G(\alpha)$, in the Hilbert space with inner product weighted by $R_{xx}(\alpha)$.

The error (2.3) is of interest for the reconstruction problem because, for the case $0 \leqslant K$, a non-negative function ($H(\alpha)$) is being approximated by a necessarily non-real trigonometric polynomial, in a Hilbert space with weighted inner product. As shown in reference [1], this is precisely what happens in the maximum entropy method (MEM) of Burg [4], that the finite polynomial also approximates $G(\alpha)$ is implicit in the MEM in the spectral factorisation [5].

In the next section we employ these approximation theoretic aspects of Wiener filter design to obtain reconstruction methods.

3. Wiener filter approximation and reconstruction: 1D case

We consider the problem of reconstructing the non-negative function $f(a)$, $|a| \leqslant \pi$, from finitely many values of its Fourier coefficients, $F(m)$, $|m| \leqslant M$. We present first linear methods and then nonlinear ones.

3.1. Linear methods

Assume that we have a prior estimate of the broad features of $f(a)$, in the form of a non-negative function $p(a)$, such that $p(a) = 0$ only if $f(a) = 0$; of course, in practice we will not know where the support of f is, exactly, so $p(a)$ should be positive everywhere. A rough idea of the support of f can be indicated by concentrating p in that region. Let p play the role of R_{xx}, f the role of R_{ss}; we are effectively assuming that, for some $\varepsilon > 0$, we have $p(a) \geqslant \varepsilon f(a)$ for all a, and that (apart from the scaling) $p(a)$

overestimates $f(a)$ everywhere; the scale factor cancels in the end, so is not needed. Then $H=f/p$ is approximated by the polynomial D^M_{-M} and the equations that must be solved are

$$F(m)=\sum_{k=-M}^{M} d(k)\,P(m-k) \qquad -M\leqslant m\leqslant M \tag{3.1}$$

where $P(m)$ are the Fourier coefficients of $p(a)$. These equations arise when we minimise the approximation error

$$\int_{-\pi}^{\pi}\left| f(a)-p(a)\sum_{k=-M}^{M} d(k)\exp(ika)\right|^2 p(a)^{-1}\,da \tag{3.2}$$

as a function of the $d(k)$. The resulting estimator of $f(a)$ is the PDFT [6], so called because of its form:

$$\text{PDFT}(a)=p(a)D^M_{-M}(a). \tag{3.3}$$

If the prior estimate $p(a)=\text{constant}$, $|a|\leqslant\pi$, then $d(k)=F(k)$ and the PDFT reduces to the DFT.

If the data are oversampled relative to the actual support of $f(a)$ then including information about this support in the $p(a)$ can result in significant improvement, so long as regularisation to avoid sensitivity to noise is used [1]. Note that, although the PDFT is not necessarily non-negative, it is data consistent; it extrapolates values of $F(m)$ beyond the data window.

3.2. Nonlinear methods

We assume now that $f(a)$ consists of two components, a discrete (delta functions) component, which is the object of interest, and a background (continuous) component, about which we have some prior information; let $p(a)$ be our non-negative prior estimate of the background component. Letting $p(a)$ play the role of R_{ss} and $f(a)$ the role of R_{xx}, we see that $H=p/f$; for the filter function D^L_K to remove from $R_{xx}=f$ the component associated with $R_{uu}=f-p$ it must place nulls near the values of the support of the discrete component. We perform the calculations to obtain the D^L_K and then examine the nulls. We have some freedom in the choice of the K and L; two choices are of particular importance: (i) $M/2\geqslant L=-K$; (ii) $L=M$, $K=0$. The first has as a special case the symmetric linear predictor (SLP) [7, 8], while the second includes Burg's MEM.

In case (i) we solve the equations

$$P(m)=\sum_{k=-L}^{L} d(k)\,F(m-k) \qquad |m|\leqslant L=M/2 \quad (M\text{ even}) \tag{3.4}$$

and use the fact that D^L_{-L} approximates $H=p/f$ to obtain, as the estimate of f, the centred inverted PDFT (CIPDFT):

$$\text{CIPDFT}(a)=p(a)/D^L_{-L}(a). \tag{3.5}$$

If the prior $p(a)=\text{constant}$, then (3.5) becomes the symmetric linear predictive method of Johnson [7]. Note that if the support of f is properly contained within the support of p then, in order to obtain the $d(k)$ that are optimal for that f and p, it is necessary to replace $P(m)$ in (3.4) by the corresponding Fourier coefficient of the function, that is $p(a)$ on the support of f and zero otherwise. In practice one does not know the support of f; our point is rather that (3.4) does not provide the optimal $d(k)$ for such pairs p and f.

The role of the prior $p(a)$, in the nonlinear methods, is to reduce bias in the estimate of peak locations; if we estimate the background component badly then the filter, as it tries to eliminate the $f(a)-p(a)$ features, must null out {true background $-\,p$} as well as the discrete component. With limited freedom to place nulls, bias is unavoidable. This has been shown to be a problem with MEM, when used on oversampled data [1], and is due to the assumption, implicit in MEM, that the background is constant over $[-\pi, \pi]$. We consider MEM next, as a special case of (ii).

For case (ii) we have $K=0$, $L=M$ and we solve equations

$$P(m)=\sum_{k=0}^{M} d(k)\,F(m-k) \qquad 0\leqslant m\leqslant M \tag{3.6}$$

to obtain the filter function D_0^M; we view this function now as an approximation of G, not of $H=p/f$. The discrete Wiener–Hopf equations (2.4) are equivalent to the statement $(R_{ss})_+=(R_{xx}G)_+$, where by $(R_{ss})_+$ we mean the causal part of the Fourier series

$$(R_{ss})_+(a)=\sum_{m=0}^{\infty} r_{ss}(m)\exp(ima) \qquad |a|\leqslant\pi \tag{3.7}$$

and similarly for other functions. Equations (3.6) tell us that the two causal functions p_+ and $(f\,D_0^M)_+$ have identical Fourier coefficients, out to index $m=M$. Because D_0^M is a finite polynomial we can rewrite $(fD_0^M)_+$ as $(fD_0^M)_+=f_+D_0^M+j_+$, where j_+ is a finite causal polynomial involving only known values:

$$j_+(a)=\sum_{m=1}^{M-1}\left(\sum_{k=1}^{M-m} F(-k)\,d(m-k)\right)\exp(ima). \tag{3.8}$$

From $p_+\simeq(fD_0^M)_+=f_+D_0^M+j_+$ we obtain an estimate q of f_+:

$$q(a)=(p_+(a)-j_+(a))/D_0^M(a); \tag{3.9}$$

from $f=2\mathrm{Re}(f_+)-F(0)$ we obtain the inverse PDFT (IPDFT) estimate of f itself:

$$\mathrm{IPDFT}(a)=2\mathrm{Re}(q(a))-F(0). \tag{3.10}$$

Consider the complex polynomial $D(z)=d(0)+d(1)z+\ldots+d(M)z^M$. If the roots of $D(z)$ are outside the unit circle (the minimum phase, or MP, property) then $1/D_0^M(a)$ is also causal and so is $q(a)$. It can be shown easily that, if $1/D_0^M(a)$ is causal, then IPDFT(a) is data consistent, although it may not be non-negative. Although it is not always the case that $D(z)$ has the MP property, it is frequently the case in practice, and the IPDFT is usually data consistent. If the prior $p(a)=$ constant, then the IPDFT reduces to Burg's MEM, the $D(z)$ has the MP and the MEM is data consistent (as well as non-negative).

As remarked earlier, the MEM has been observed to perform poorly when the function f is concentrated in a smaller region of $[-\pi, \pi]$; this is because the $p(a)$ is a constant, while the background is not evenly distributed over all of $[-\pi, \pi]$. Because the IPDFT is free to take on negative values it could be used to gauge the accuracy of the prior being used; significant negative values should indicate that our $p(a)$ is not accurate. We have not obtained a quantitative measure of the significance of negative values, however.

We consider now the extensions of these methods to the two-dimensional case. Although much remains essentially the same the absence of an obvious generalisation of the notion of causality affects the extension of the IPDFT.

4. The two-dimensional case

As in the one-dimensional case, the power spectrum of the input, $R_{xx}(\alpha, \beta)$, is the sum of two components, $R_{xx}(\alpha, \beta) = R_{ss}(\alpha, \beta) + R_{uu}(\alpha, \beta)$, and the Wiener filter is the doubly indexed sequence $\{h(j, k)\}$ of Fourier coefficients of the function $H(\alpha, \beta) = R_{ss}(\alpha, \beta)/R_{xx}(\alpha, \beta)$. To obtain a finite-impulse-response approximation to the Wiener filter we minimise the following error of approximation, as a function of the $d(j, k)$:

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |H(\alpha,\beta) - D_{I,K}^{J,L}(\alpha,\beta)|^2 R_{xx}(\alpha,\beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta \qquad (4.1)$$

where

$$D_{I,K}^{J,L}(\alpha,\beta) = \sum_{j=I}^{J}\sum_{k=K}^{L} d(j,k)\exp(\mathrm{i}j\alpha + \mathrm{i}k\beta). \qquad (4.2)$$

In the two-dimensional case the problem is to reconstruct the non-negative function $f(a, b)$, $|a|$, $|b| \leq \pi$, from finitely many values of its Fourier coefficients, $F(m, n)$, $|m| \leq M$, $|n| \leq N$. As in the one-dimensional case we consider estimates of $f(a, b)$ obtained by analogy with the problem of approximating the Wiener filter.

4.1. *Linear methods*

Assume that a prior estimate of the general shape of $f(a, b)$ is given by the positive function $p(a, b)$, and that $\{P(m, n)\}$ are its Fourier coefficients. As before, we let p play the role of R_{xx}, f the role of R_{ss}, so that the Wiener filter function is $H = f/p$.

For fixed I, J, K, L the optimal finite-impulse-response filter function is $D(\alpha, \beta) = D_{I,K}^{J,L}(\alpha, \beta)$ where the coefficients of D satisfy the equations

$$F(m,n) = \sum_{j=I}^{J}\sum_{k=K}^{L} d(j,k)\, P(m-j, n-k) \qquad |m| \leq M, \quad |n| \leq N. \qquad (4.3)$$

Having found the $d(j, k)$ we use the fact that D approximates $H = f/p$ to obtain our estimate of f:

$$\mathrm{PDFT}(a,b) = p(a,b)\, D(a,b). \qquad (4.4)$$

The obvious choices for I, J, K, L are $-I = J = M$, $-K = L = N$.

4.2. *Nonlinear methods*

We assume now that $f(a, b)$ has a discrete component of interest, as well as a background component estimated by the non-negative function $p(a, b)$. As in the 1D case we let p play the role of R_{ss}, f the role of R_{xx}, so that the approximate Wiener filter attempts to null out the discrete component. If the support of f contains the support of p then the equations to be solved for the optimal finite-impulse-response filter $\{d(j, k)\}$ are

$$P(m, n) = \sum_{j=I}^{J} \sum_{k=K}^{L} d(j, k) F(m-j, n-k) \qquad I \leqslant m \leqslant J, \quad K \leqslant n \leqslant L. \tag{4.5}$$

The choices for I, J, K, L will be restricted by the available data, since the data make up the entries of the matrix that appears in the system of equations to be solved. We consider here two possibilities.

(i) Let $J = -I = M/2$, $L = -K = N/2$ (M, N even). Solving (4.4) for the $d(j, k)$ we view $D = D_{I,K}^{J,L}$ as an approximation of $H = p/f$, so that our estimate of f is the two-dimensional version of (3.5):

$$\text{CIPDFT}\,(a, b) = p(a, b)/D(a, b). \tag{4.6}$$

(ii) Let $I = K = 0$, $J = M$, $L = N$. Then D can be viewed as an estimate of the first-quadrant-indexed component of H, which we denote by H_{++}. With the first-quadrant-indexed component of $p(a, b)$ given by

$$p(a, b)_{++} = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} P(m, n) \exp(ima + inb) \tag{4.7}$$

equations (4.4) state that $p(a, b)_{++}$ and $[f(a, b)D(a, b)]_{++}$ have the same Fourier coefficients, for indices $0 \leqslant m \leqslant M$, $0 \leqslant n \leqslant N$. As in the 1D case, we can write $[f(a, b)D(a, b)]_{++} = f(a, b)_{++}D(a, b) + j(a, b)_{++}$, where $j(a, b)_{++}$ is a first-quadrant-indexed function that involves only known values. Our estimate of $f(a, b)_{++}$ is then $q(a, b) = [p(a, b)_{++} - j(a, b)_{++}]/D(a, b)$, which may not itself be first-quadrant-indexed. Repeating this procedure three more times, for each quadrant, we estimate $f(a, b)$ by summing the four estimates so obtained, taking care to subtract components included in more than one estimator. The resulting estimator we call the IPDFT.

5. Relation to other methods

The reconstruction problem considered here is to obtain the function $f(a, b)$ from the values $F(m, n)$, $|m| \leqslant M$, $|n| \leqslant N$, where

$$F(m, n) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(a, b) \exp[-(ima + inb)] \, da \, db/4\pi^2; \tag{5.1}$$

that is, we are attempting to solve the integral equation. The survey paper by Frieden [9] describes a number of approaches to this problem.

When the $p(a, b)$ is chosen to incorporate support information, so that $p(a, b) = 1$, $|a| < A < \pi$, $|b| < B < \pi$, and $p(a, b) = 0$ elsewhere, a small amount of noise in the data can cause degradation of the PDFT estimator. It is safer to make $p(a, b) = \varepsilon > 0$, instead of $p(a, b) = 0$. This is a form of regularisation and is in keeping with the requirement that the support of p be no smaller than that of f. Since the PDFT performs an approximation of the function f in the (a, b) domain and smooths the effects of noise (if regularised), it resembles the methods of Phillips [10] and Twomey [11]. The main differences are that the PDFT retains the continuous formulation, rather than discretising $f(a, b)$, and employs a prior estimate of the function f.

It might appear that the PDFT is related to the Helstrom–Wiener 'sharpness-constrained' method [12]. The latter is based, however, on a statistical, or ensemble, model for the restoration problem and employs power spectra of f and p; the mean squared error is calculated in the usual L^2 norm, rather than with a weighted norm in (a, b) space. In the approach presented here we do not postulate the existence of an ensemble of object functions f to be restored and the idea of Wiener filtering is introduced only to borrow the weighted error criterion used for approximating non-negative functions by polynomials. The Backus–Gilbert [13] method is similar in philosophy to the Helstrom–Wiener approach but there is only a superficial connection to the estimators presented here.

Because the finite data are typically insufficient to determine a single, unique solution to the reconstruction problem, one is faced with the task of selecting, from among the many possibilities, one particular answer. The general feeling, which we share, is that the selection should not be arbitrary but should be guided by some reasonable principles of inference. At this point there is some disagreement concerning which principles of inference are to be called reasonable. In an attempt to resolve the situation Shore and Johnson [14] developed an axiomatic basis for the principle of cross-entropy minimisation and Jones [15] has recently provided an approximation-theoretic argument for the same method.

Among all functions $g(a, b)>0$ consistent with our data we could select the one for which the Shannon entropy

$$\text{entropy}\ (g) = -\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(a, b)\log g(a, b)\, \mathrm{d}a\, \mathrm{d}b \tag{5.2}$$

is maximised. Generally, there is no closed-form solution and iterative procedures are employed.

If there is available a prior estimate, $p(a, b)$, of $f(a, b)$ then (5.2) is replaced by the cross-entropy of g, given p:

$$\text{cross-entropy}\ (g|p) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(a, b)\log[g(a, b)/p(a, b)]\, \mathrm{d}a\, \mathrm{d}b. \tag{5.3}$$

The method of 'minimisation of cross-entropy' (MCE) has us select, as the estimate of $f(a, b)$, that data-consistent $g(a, b)>0$ for which the integral in (5.3) is minimised. The optimal solution then has the form

$$\text{MCE}\ (a, b) = p(a, b)\exp\left(\sum_{j=-M}^{M}\sum_{k=-N}^{N} c(j, k)\exp(\mathrm{i}ja + \mathrm{i}bk)\right). \tag{5.4}$$

If $p(a, b)$ is a good prior estimate then the sum in the exponential term will be near zero; approximating $\exp(x)$ by $1+x$ leads to an estimator of the PDFT form. If it is known that the function $f(a, b)$ is spiky, then the sum in the exponential term will have significant negative values. If we estimate $\exp(x)$ by $1/(1-x)$, then this is better than $1+x$ for negative x; making this approximation in (5.4) leads to the CIPDFT form.

When the $p(a, b)$ is constant over the support of the object function $f(a, b)$ the PDFT (4.4) provides a minimum energy extrapolation of the data, consistent with the support constraint. In reference [16] we considered the problem of reconstructing

$f(a, b)$ from only the magnitude data, $|F(m, n)|$, $|m| \leq M$, $|n| \leq N$, that is, the phase retrieval problem. When arbitrary phases are assigned to the magnitude data and the PDFT energy calculated one finds the energy to be dependent on that choice of phases and therefore to provide a useful cost function to direct the search for the correct phases.

6. Conclusions

In this paper we have considered the reconstruction of a non-negative function from finitely many values of its Fourier transform. We have extended to the 2D case methods previously presented for 1D reconstruction [1] and obtained a new derivation of these estimators based on analogy with the design of approximate Wiener filters, in which the object function to be reconstructed and our prior estimate play the roles of input and output power spectra. To obtain linear estimators we let our prior estimate play the role of the input power spectrum, allowing the filter to extract those features not found in the true object. To obtain nonlinear estimators for spiky objects with continuous backgrounds we estimate the background function, and then let it play the role of the output power spectrum; the true object function then plays the role of the input power spectrum, so that the filter attempts to null out the discrete component.

The linear methods are extrapolation procedures that are particularly useful when the data are oversampled. The nonlinear methods generalise the Burg maximum entropy method, for the 1D case, and provide computationally inexpensive 2D approximations to other entropy-based methods.

The methods presented here are derived using the best approximation in weighted Hilbert spaces; the linear equations to be solved in each case are the normal equations that arise from such a best approximation and we know what is being approximated by what in each case.

Acknowledgments

The authors acknowledge the financial support of the Office of Naval Research, through Contract N00014-87-K-0394 and NATO grant 725/84. They also wish to thank Professor James Howard and the staff of the Human Performance Laboratory, The Catholic University of America, for their assistance.

References

[1] Byrne C L, Fitzgerald R M, Fiddy M A, Hall T J and Darling A M 1983 Image restoration and resolution enhancement *J. Opt. Soc. Am.* **73** 1481–7

[2] Ekstrom M 1982 Realisable Wiener filtering in two dimensions *IEEE Trans. Acoust., Speech Signal Proc.* **ASSP-30** 31

[3] Papoulis A 1977 *Signal Analysis* (New York: McGraw-Hill)

[4] Burg J P 1975 Maximum entropy spectral analysis *PhD Dissertation* Stanford University

[5] Byrne C L and Fitzgerald R M 1983 An approximation theoretic approach to maximum entropy spectral analysis *IEEE Trans. Acoust., Speech Signal Proc.* **ASSP-31** 734–6

[6] Byrne C L and Fitzgerald R M 1984 Spectral estimators that extend the maximum entropy and maximum likelihood methods *SIAM J. Appl. Math.* **44** 425–42

[7] Johnson D 1982 The application of spectral estimation methods to bearing estimation problems *Proc. IEEE* **70** 1018–28

[8] Cox H, Zeskind R and Owen M 1985 The effect of random errors on linear predictive array processors *Bolt Beranek and Newman, Inc. Technical Report No* 5595

[9] Frieden B R 1978 Image enhancement and restoration *Picture Processing and Digital Filtering* (*Topics in Applied Physics* vol. 6) ed. T S Huang (New York: Springer)

[10] Phillips D L 1962 A technique for the numerical solution of certain integral equations of the first kind *J. Assoc. Comput. Mach.* **9** 84–97

[11] Twomey S 1963 On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature *J. Assoc. Comput. Mach.* **10** 97–101

[12] Helstrom C W 1967 Image restoration by the method of least squares *J. Opt. Soc. Am.* **57** 297–303

[13] Backus G and Gilbert F 1970 Uniqueness in the inversion of gross earth data *Phil. Trans. R. Soc.* A **266** 123–92

[14] Shore J E and Johnson R W 1980 Axiomatic derivation of the principle of maximum entropy and minimum cross entropy *IEEE Trans. Inform. Theory* **IT-26** 926–43

[15] Jones L K 1987 Approximation theoretic uniqueness for the principle of minimum cross entropy *IEEE Trans. Inform. Theory* to be published

[16] Byrne C L and Fiddy M A 1987 Estimation of continuous object distributions from limited Fourier magnitude measurements *J. Opt. Soc. Am.* A **4** 112–7