Prior Knowledge and Resolution Enhancement

Charles Byrne
(Charles_Byrne@uml.edu)
http://faculty.uml.edu/cbyrne/cbyrne.html
Department of Mathematical Sciences
University of Massachusetts Lowell
Lowell, MA 01854, USA

June 3, 2008

1 Overview

A common problem in signal and image processing is the reconstruction of a function $f(x)$ of one or several variables from finitely many integrals

$$d_n = \int_a^b f(x)\overline{g_n(x)}dx,$$

for $n = 1, ..., N$. If we take as our inner product of functions $u(x)$ and $v(x)$

$$\langle u, v\rangle = \int u(x)\overline{v(x)}dx,$$

then the minimum-norm solution of this under-determined problem is

$$\hat{f}(x) = c_1 g_1(x) + ... + c_N g_N(x),$$

where the coefficients are chosen so as to make the reconstructed function consistent with the data.

For any positive function $p(x)$ we can write

$$d_n = \int_a^b f(x)\overline{g_n(x)}p(x)p(x)^{-1}dx,$$

so that, using the inner product

$$\langle u, v\rangle = \int_a^b u(x)\overline{v(x)}p(x)^{-1}dx,$$

the minimum-norm solution becomes

$$\hat{f}(x) = p(x)\Big(c_1 g_1(x) + ... + c_N g_N(x)\Big),$$

with the c_n again chosen so as to make $\hat{f}(x)$ consistent with the data. We can incorporate our prior knowledge of prominent features of the true $f(x)$, such as its support, through the appropriate selection of the $p(x)$.

2 Introduction

An important point to keep in mind when doing signal and image processing is that, while the data is usually limited, the information we seek may not be lost. Although processing the data in a reasonable way may suggest otherwise, other processing methods may reveal that the desired information is still available in the data. Figure 1 illustrates this point. The top right is the original discrete simulated head slice. The data are some of the FFT values nearest the origin. The bottom right is the DFT, or minimum-norm solution, obtained by replacing with zeros the FFT values not being used, and then inverse Fourier transforming. The top left is the prior weighting function, and the bottom left is the minimum weighted norm reconstruction, from the same data.

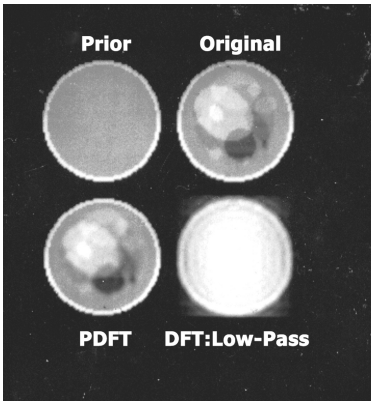

Figure 1: Minimum Norm and Minimum Weighted Norm Reconstruction.

3 The Basic problem

The problem is to reconstruct a (possibly complex-valued) function $f : R^D \rightarrow C$ from finitely many measurements d_n, $n = 1, ..., N$, pertaining to f. The function $f(r)$ represents the physical object of interest, such as the spatial distribution of acoustic energy in sonar, the distribution of x-ray-attenuating material in transmission tomography, the distribution of radionuclide in emission tomography, the sources of reflected radio waves in radar, and so on. Often the reconstruction, or estimate, of the function f takes the form of an image in two or three dimensions; for that reason, we also speak

of the problem as one of *image reconstruction*. The data are obtained through measurements. Because there are only finitely many measurements, the problem is highly under-determined and even noise-free data are insufficient to specify a unique solution.

4 The Optimization Approach

One way to solve such under-determined problems is to replace $f(r)$ with a vector in C^N and to use the data to determine the N entries of this vector. An alternative method is to model $f(r)$ as a member of a family of linear combinations of N preselected basis functions of the multi-variable r. Then the data is used to determine the coefficients. This approach offers the user the opportunity to incorporate prior information about $f(r)$ in the choice of the basis functions. Such finite-parameter models for $f(r)$ can be obtained through the use of the minimum-norm estimation procedure, as we shall see. More generally, we can associate a *cost* with each data-consistent function of r, and then minimize the cost over all the potential solutions to the problem. Using a norm as a cost function is one way to proceed, but there are others. These optimization problems can often be solved only through the use of discretization and iterative algorithms.

5 A Hilbert Space Formulation

In many applications the data are related linearly to f. To model the operator that transforms f into the data vector, we need to select an ambient space containing f. Typically, we choose a Hilbert space. The selection of the inner product provides an opportunity to incorporate prior knowledge about f into the reconstruction. The inner product induces a norm and our reconstruction is that function, consistent with the data, for which this norm is minimized. We shall illustrate the method using Fourier-transform data and prior knowledge about the support of f and about its overall shape.

Our problem, then, is to estimate a (possibly complex-valued) function $f(r)$ of D real variables $r = (r_1, ..., r_D)$ from finitely many measurements, d_n, $n = 1, ..., N$. We shall assume, in this article, that these measurements take the form

$$d_n = \int_S f(r)\overline{g_n(r)}dr, \tag{1}$$

where S denotes the support of the function $f(r)$, which, in most cases, is a bounded set. For the purpose of estimating, or reconstructing, $f(r)$, it is convenient to view Equation (1) in the context of a Hilbert space, and to write

$$d_n = \langle f, g_n \rangle, \tag{2}$$

where the usual Hilbert space inner product is defined by

$$\langle u, v \rangle_2 = \int_S u(r)\overline{v(r)}dr, \tag{3}$$

for functions $u(r)$ and $v(r)$ supported on the set S. Of course, for these integrals to be defined, the functions must satisfy certain additional properties, but a more complete

discussion of these issues is outside the scope of this chapter. The Hilbert space so defined, denoted $L^2(S)$, consists (essentially) of all functions $f(r)$ for which the norm

$$||f||_2 = \sqrt{\int_S |f(r)|^2 dr} \tag{4}$$

is finite.

5.1 Minimum-Norm Solutions

Our estimation problem is highly under-determined; there are infinitely many functions in $L^2(S)$ that are consistent with the data and might be the right answer. Such under-determined problems are often solved by acting conservatively, and selecting as the estimate that function consistent with the data that has the smallest norm. At the same time, however, we often have some prior information about f that we would like to incorporate in the estimate. One way to achieve both of these goals is to select the norm to incorporate prior information about f, and then to take as the estimate of f the function consistent with the data, for which the chosen norm is minimized.

The data vector $d = (d_1, ..., d_N)^T$ is in C^N and the linear operator $\mathcal{H}$ from $L^2(S)$ to C^N takes f to d; so we write $d = \mathcal{H}f$. Associated with the mapping $\mathcal{H}$ is its adjoint operator, $\mathcal{H}^\dagger$, going from C^N to $L^2(S)$ and given, for each vector $a = (a_1, ..., a_N)^T$, by

$$\mathcal{H}^\dagger a(r) = a_1 g_1(r) + ... + a_N g_N(r). \tag{5}$$

The operator from C^N to C^N defined by $\mathcal{H}\mathcal{H}^\dagger$ corresponds to an N by N matrix, which we shall also denote by $\mathcal{H}\mathcal{H}^\dagger$. If the functions $g_n(r)$ are linearly independent, then this matrix is positive-definite, therefore invertible.

Given the data vector d, we can solve the system of linear equations

$$d = \mathcal{H}\mathcal{H}^\dagger a \tag{6}$$

for the vector a. Then the function

$$\hat{f}(r) = \mathcal{H}^\dagger a(r) \tag{7}$$

is consistent with the measured data and is the function in $L^2(S)$ of least norm for which this is true. The function $w(r) = f(r) - \hat{f}(r)$ has the property $\mathcal{H}w = 0$. It is easy to see that

$$||f||_2^2 = ||\hat{f}||_2^2 + ||w||_2^2 \tag{8}$$

The estimate $\hat{f}(r)$ is the *minimum-norm solution*, with respect to the norm defined in Equation (4). If we change the norm on $L^2(S)$, or, equivalently, the inner product, then the minimum-norm solution will change.

For any continuous linear operator $\mathcal{T}$ on $L^2(S)$, the adjoint operator, denoted $\mathcal{T}^\dagger$, is defined by

$$\langle \mathcal{T}u, v\rangle_2 = \langle u, \mathcal{T}^\dagger v\rangle_2. \tag{9}$$

The adjoint operator will change when we change the inner product.

6 A Class of Inner Products

Let $\mathcal{T}$ be a continuous, linear and invertible operator on $L^2(S)$. Define the $\mathcal{T}$ inner product to be

$$\langle u, v\rangle_{\mathcal{T}} = \langle \mathcal{T}^{-1}u, \mathcal{T}^{-1}v\rangle_2. \tag{1}$$

We can then use this inner product to define the problem to be solved. We now say that

$$d_n = \langle f, t^n\rangle_{\mathcal{T}}, \tag{2}$$

for known functions $t^n(r)$. Using the definition of the $\mathcal{T}$ inner product, we find that

$$d_n = \langle f, g^n\rangle_2 = \langle \mathcal{T}f, \mathcal{T}g^n\rangle_{\mathcal{T}}. \tag{3}$$

The adjoint operator for $\mathcal{T}$, with respect to the $\mathcal{T}$-norm, is denoted $\mathcal{T}^*$, and is defined by

$$\langle \mathcal{T}u, v\rangle_{\mathcal{T}} = \langle u, \mathcal{T}^*v\rangle_{\mathcal{T}}. \tag{4}$$

Therefore,

$$d_n = \langle f, \mathcal{T}^*\mathcal{T}g^n\rangle_{\mathcal{T}}. \tag{5}$$

Lemma 1. *We have* $\mathcal{T}^*\mathcal{T} = \mathcal{T}\mathcal{T}^\dagger$.

Consequently, we have

$$d_n = \langle f, \mathcal{T}\mathcal{T}^\dagger g^n\rangle_{\mathcal{T}}. \tag{6}$$

7 Minimum-$\mathcal{T}$-Norm Solutions

The function $\tilde{f}$ consistent with the data and having the smallest $\mathcal{T}$-norm has the algebraic form

$$\hat{f} = \sum_{m=1}^{N} a_m \mathcal{T}\mathcal{T}^\dagger g^m. \tag{1}$$

Applying the $\mathcal{T}$-inner product to both sides of Equation (1), we get

$$d_n = \langle \hat{f}, \mathcal{T}\mathcal{T}^\dagger g^n\rangle_{\mathcal{T}} \tag{2}$$

$$= \sum_{m=1}^{N} a_m \langle \mathcal{T}\mathcal{T}^\dagger g^m, \mathcal{T}\mathcal{T}^\dagger g^n\rangle_{\mathcal{T}}. \tag{3}$$

Therefore,

$$d_n = \sum_{m=1}^{N} a_m \langle \mathcal{T}^\dagger g^m, \mathcal{T}^\dagger g^n\rangle_2. \tag{4}$$

We solve this system for the a_m and insert them into Equation (1) to get our reconstruction. The Gram matrix that appears in Equation (4) is positive-definite, but is often ill-conditioned; increasing the main diagonal by a percent or so usually is sufficient regularization.

8 The Case of Fourier-Transform Data

To illustrate these minimum-$\mathcal{T}$-norm solutions, we consider the case in which the data are values of the Fourier transform of f. Specifically, suppose that

$$d_n = \int_S f(x) e^{-i\omega_n x} dx, \tag{1}$$

for arbitrary values ω_n.

8.1 The $L^2(-\pi, \pi)$ Case

Assume that $f(x) = 0$, for $|x| > \pi$. The minimum-2-norm solution has the form

$$\hat{f}(x) = \sum_{m=1}^{N} a_m e^{i\omega_m x}, \tag{2}$$

with

$$d_n = \sum_{m=1}^{N} a_m \int_{-\pi}^{\pi} e^{i(\omega_m - \omega_n)x} dx. \tag{3}$$

For the equispaced values $\omega_n = n$ we find that $a_m = d_m$ and the minimum-norm solution is

$$\hat{f}(x) = \sum_{n=1}^{N} g_n e^{inx}. \tag{4}$$

8.2 The Over-Sampled Case

Suppose that $f(x) = 0$ for $|x| > A$, where $0 < A < \pi$. The Nyquist sample spacing is then $\Delta = \pi/A$. In many applications we can take as many samples as we wish, but must take them within some fixed interval. If we take samples at the rate of $\Delta = \pi/A$, we may not get very many samples to work with. Instead, we may sample at a faster rate, say $\Delta = 1$, to get more data points. How we process this over-sampled data is important.

If we use as our ambient Hilbert space $L^2(-\pi, \pi)$, the minimum-norm reconstruction wastes a lot of effort reconstructing $f(x)$ outside $[-A, A]$, where we already know it to be zero. Instead, we use $L^2(-A, A)$ as the ambient Hilbert space.

For the simulation in Figure 2, $f(x) = 0$ for $|x| > A = \pi/30$. The top graph is the minimum-norm estimator, with respect to the Hilbert space $L^2(-A, A)$, called the *modified* DFT (MDFT); the bottom graph is the DFT, the minimum-norm estimator with respect to the Hilbert space $L^2(-\pi, \pi)$. The MDFT is a non-iterative variant of Gerchberg-Papoulis [18, 19] band-limited extrapolation.

8.3 Using a Prior Estimate of f

Suppose that $f(x) = 0$ for $|x| > \pi$ again, and that $p(x)$ satisfies

$$0 < \epsilon \leq p(x) \leq E < +\infty, \tag{5}$$

for all x in $[-\pi, \pi]$. Define the operator $\mathcal{T}$ by $(\mathcal{T}f)(x) = \sqrt{p(x)}f(x)$. The $\mathcal{T}$-norm is then

$$\langle u, v \rangle_{\mathcal{T}} = \int_{-\pi}^{\pi} u(x)\overline{v(x)}p(x)^{-1}dx. \tag{6}$$

It follows that

$$d_n = \int_{-\pi}^{\pi} f(x)p(x)e^{-inx}p(x)^{-1}dx, \tag{7}$$

so that the minimum $\mathcal{T}$-norm solution is

$$\hat{f}(x) = \sum_{m=1}^{N} a_m p(x)e^{imx} = p(x)\sum_{m=1}^{N} a_m e^{imx}, \tag{8}$$

where

$$d_n = \sum_{m=1}^{N} a_m \int_{-\pi}^{\pi} p(x)e^{i(m-n)x}dx. \tag{9}$$

If we have prior knowledge about the support of f, or some idea of its shape, we can incorporate that prior knowledge into the reconstruction through the choice of $p(x)$.

The reconstruction in Equation (8) was presented in [5], where it was called the PDFT method. The PDFT was based on an earlier non-iterative version of the Gerchberg-Papoulis bandlimited extrapolation procedure [4]. The PDFT was then applied to image reconstruction problems in [6]. An application of the PDFT was presented in [8].

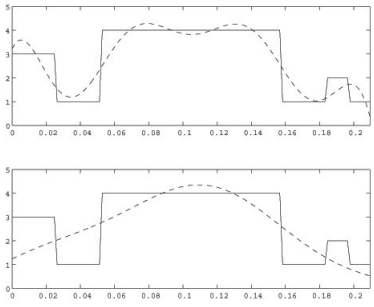

Figure 2: The non-iterative band-limited extrapolation method (MDFT) (top) and the DFT (bottom); 30 times over-sampled.

In [7] we extended the PDFT to a nonlinear version, the indirect PDFT (IPDFT), that generalizes Burg's maximum entropy spectrum estimation method [1, 2, 3]. The PDFT was applied to the phase problem in [10] and in [11] both the PDFT and IPDFT were examined in the context of Wiener filter approximation. More recent work on these topics is discussed in the books [12, 13].

In the next section we sketch the non-linear indirect PDFT estimator, the IPDFT.

9 The Non-Linear Indirect PDFT (IPDFT)

Suppose that $r(x) \geq 0$, for $|x| \leq \pi$, with

$$r(x) = \sum_{n=-\infty}^{\infty} R(n) e^{inx},$$

and we want to reconstruct this function from the data $R(n)$, $|n| \leq N$. The goal here is to obtain a non-linear estimator along the lines of Burg's maximum entropy method [1, 2, 3].

9.1 Reconstructing the Additive Causal Part

We begin by considering the problem of reconstructing its *additive causal part*,

$$r(x)_+ = \sum_{n=0}^{\infty} R(n) e^{inx},$$

from data $R(n)$, for $n = 0, 1, ..., N$. We use the prior $p(x)$ and the PDFT, obtaining the estimate

$$\hat{r}(x) = p(x) \sum_{n=0}^{N} c_n e^{inx}.$$

To obtain the c_n we need to solve the system

$$\begin{bmatrix} P(0) & P(-1) & ... & P(-N) \\ P(1) & P(0) & ... & P(-N+1) \\ \cdot & \cdot & ... & \cdot \\ \cdot & \cdot & ... & \cdot \\ P(N) & P(N-1) & ... & P(0) \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \cdot \\ \cdot \\ c_N \end{bmatrix} = \begin{bmatrix} R(0) \\ R(1) \\ \cdot \\ \cdot \\ R(N) \end{bmatrix}.$$

Suppose now that we switch the roles of $r(x)$ and $p(x)$, "estimating" $p(x)_+$ using $r(x) \geq 0$ as the prior.

9.2 Switching Roles

Now we need to solve the system

$$\begin{bmatrix} R(0) & R(-1) & ... & R(-N) \\ R(1) & R(0) & ... & R(-N+1) \\ \cdot & \cdot & ... & \cdot \\ \cdot & \cdot & ... & \cdot \\ R(N) & R(N-1) & ... & R(0) \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \cdot \\ \cdot \\ c_N \end{bmatrix} = \begin{bmatrix} P(0) \\ P(1) \\ \cdot \\ \cdot \\ P(N) \end{bmatrix}.$$

Since $R(-n) = \overline{R(n)}$, we know all the entries of the matrix.

Our "estimate"of $p(x)_+$ is then

$$\hat{p}(x)_+ = r(x) \sum_{n=0}^{N} c_n e^{inx} = r(x)c(x).$$

The additive causal part of the right side is

$$\Big(r(x)c(x)\Big)_+ = r(x)_+ c(x) + \sum_{m=0}^{N-1} \Big(\sum_{k=1}^{N-m} R(-k)c_{m+k} \Big) e^{imx}$$

$$= r(x)_+ c(x) + j(x).$$

Therefore, we can say that

$$p(x)_+ \approx r(x)_+ c(x) + j(x),$$

so that

$$r(x)_+ \approx q(x) = \frac{p(x)_+ - j(x)}{c(x)}.$$

Our IPDFT estimate of $r(x)$ is then

$$\hat{r}(x) = 2\text{Real}(q(x)) - R(0).$$

When $p(x) = 1$ for all x we get Burg's maximum entropy estimator, which is always non-negative and consistent with the original data. The IPDFT is always real-valued. It is not guaranteed to be non-negative, but seems to be, most of the time. The big question is "Is the IPDFT estimate consistent with the original data?"

If $c(x)^{-1}$ is causal, that is,

$$c(x)^{-1} = d_0 + d_1 e^{ix} + d_2 e^{2ix} + ...,$$

then our estimate $q(x)$ of $r(x)_+$ is causal and the IPDFT is consistent with the data. It was difficult, but not impossible, for me to find an example in which the function $c(x)^{-1}$ is not causal. This leads to

Open Problem: When is $c(x)^{-1}$ causal?

10 Poisson Mixture Problems

A *compound Poisson* probability function on the non-negative integers has

$$p(n) = \frac{1}{n!} \int_0^\infty c(\lambda) e^{-\lambda} \lambda^n d\lambda,$$

as the probability that the non-negative integer n will occur; here the non-negative function $c(\lambda)$ is the *compounding probability density function*. Measured counts provide estimates of $p(n)$, for $n = 0, 1, ..., N$. On the basis of this data we want to estimate the function $c(\lambda)$. Both the PDFT and IPDFT approaches can be used for this purpose [8, 14].

11 Discretizing the Problem

Suppose we select $J > N$ and replace the functions $f(x)$ and $g_n(x)$ with finite (column) vectors,

$$\mathbf{f} = (f_1, ..., f_J)^T,$$

and

$$\mathbf{g}^n = (g_1^n, ..., g_N^n)^T,$$

and model the data as

$$d_n = f_1 g_1^n + ... + f_N g_N^n.$$

Then a vector $\mathbf{f}$ is data consistent if it solves the under-determined system

$$A\mathbf{f} = \mathbf{d},$$

where the entries of the matrix A are

$$A_{n,j} = g_j^n.$$

11.1 Minimum-Weighted-Two-Norm Solutions

The PDFT estimator minimizes the weighted two-norm

$$\int |f(x)|^2 p(x)^{-1} dx,$$

subject to data consistency. In the discrete formulation of the reconstruction problem, we seek a solution of a system of equations $A\mathbf{f} = \mathbf{d}$ for which the weighted two-norm

$$\sum_{j=1}^J |f_j|^2 w_j^{-1}$$

is minimized, where the weight vector $\mathbf{w}$ is a discretization of the function $p(x)$. This solution can be found without forming the matrix H, using, say, the iterative algebraic reconstruction technique (ART).

12 Minimum One-Norm Solutions

Finding sparse solutions to under-determined systems of linear equations is an increasingly important problem in a variety of applications, such as *compressed sampling* or *compressed sensing* (CS) [16].

12.1 Sparse Solutions

Suppose that A is a real M by N matrix, with $M < N$, and that the linear system $Ax = b$ has infinitely many solutions. For any vector x, we define the *support* of x to be the subset S of $\{1, 2, ..., N\}$ consisting of those n for which the entries $x_n \neq 0$.

Consider the problem P_0: among all solutions x of the consistent system $y = Ax$, find one, call it $\hat{x}$, having the minimum number of non-zero entries. Obviously, there will be at least one such solution having minimal support, but finding one, however, is a combinatorial optimization problem and is generally NP-hard. It is important, therefore, to have a computationally tractable method for finding maximally sparse solutions.

12.2 Minimum One-Norm Solutions

Instead, we can seek a *minimum one-norm* solution, that is, solve the problem P_1: minimize

$$||x||_1 = \sum_{n=1}^{N} |x_n|,$$

subject to $Ax = b$. Problem P_1 can be formulated as a linear programming problem, so is more easily solved. The big questions are: when does P_1 have a unique solution, and when does that solution solve P_0? The problem P_1 will have a unique solution if and only if A is such that the one-norm satisfies

$$||\hat{x}||_1 < ||\hat{x} + v||_1,$$

for all non-zero v in the null space of A.

12.3 Why the One-Norm?

When a system of linear equations $Ax = b$ is under-determined, we can find the *minimum-two-norm solution* that minimizes the square of the two-norm,

$$||x||_2^2 = \sum_{n=1}^{N} x_n^2,$$

subject to $Ax = b$. One drawback to this approach is that the two-norm penalizes relatively large values of x_n much more than the smaller ones, so tends to provide non-sparse solutions. Alternatively, we may seek the solution for which the one-norm,

$$||x||_1 = \sum_{n=1}^{N} |x_n|,$$

is minimized. The one-norm still penalizes relatively large entries x_n more than the smaller ones, but much less than the two-norm does. As a result, it often happens that the minimum one-norm solution actually solves P_0 as well.

12.4 Comparison with the PDFT

If our weights w_n are reasonably close to $|\hat{x}_n|$, then

$$\sum_{n=1}^{N} |\hat{x}_n| = \sum_{n=1}^{N} |\hat{x}_n|^2 |\hat{x}_n|^{-1} \approx \sum_{n=1}^{N} |\hat{x}_n|^2 w_n^{-1},$$

and the minimum weighted-two-norm solution of $Ax = b$ should be reasonably close to the minimum one-norm solution.

12.5 Iterative Reweighting

We want each weight w_n to be a good prior estimate of the reciprocal of $|\hat{x}_n|$. Because we do not yet know $\hat{x}$, we may take a sequential-optimization approach, beginning with weights $w_n^0 > 0$, finding the PDFT solution using these weights, then using this PDFT solution to get a (we hope!) a better choice for the weights, and so on. This sequential approach was successfully implemented in the early 1980's by Michael Fiddy and his students [17].

In [15], the same approach is taken, but with respect to the one-norm. Since the one-norm still penalizes larger values disproportionately, balance can be achieved by minimizing a weighted-one-norm, with weights close to the reciprocals of the $|\hat{x}_n|$. Again, not yet knowing $\hat{x}$, they employ a sequential approach, using the previous minimum-weighted-one-norm solution to obtain the new set of weights for the next minimization.

It is interesting to note that an on-going debate among users of the PDFT has been the nature of the prior weighting. Does w_n approximate $|x_n|$ or $|x_n|^2$? This is close to the issue treated in [15], the use of a weight in the minimum-one-norm approach.

It should be noted again that finding a sparse solution is not usually the goal in the use of the PDFT, but the use of the weights has much the same effect as using the one-norm to find sparse solutions: to the extent that the weights approximate the entries of $\hat{x}$, their use reduces the penalty associated with the larger entries of an estimated solution.

13 Acknowledgments

This work was supported, in part, by the National Institutes of Health (NIH), under grant CA23452. The contents of this presentation are solely the responsibility of the author and do not necessarily reflect the official views of NIH.

References

[1] Burg, J. (1967) "Maximum entropy spectral analysis."*paper presented at the 37th Annual SEG meeting, Oklahoma City, OK.*

[2] Burg, J. (1972) "The relationship between maximum entropy spectra and maximum likelihood spectra."*Geophysics* **37**, pp. 375–376.

[3] Burg, J. (1975) *Maximum Entropy Spectral Analysis*, Ph.D. dissertation, Stanford University.

[4] Byrne, C. and Fitzgerald, R. (1979) "A unifying model for spectrum estimation."in *Proceedings of the RADC Workshop on Spectrum Estimation- October 1979*, Griffiss AFB, Rome, NY.

[5] Byrne, C. and Fitzgerald, R. (1982) "Reconstruction from partial information, with applications to tomography."*SIAM J. Applied Math.* **42(4)**, pp. 933–940.

[6] Byrne, C., Fitzgerald, R., Fiddy, M., Hall, T. and Darling, A. (1983) "Image restoration and resolution enhancement."*J. Opt. Soc. Amer.* **73**, pp. 1481–1487.

[7] Byrne, C. and Fitzgerald, R. (1984) "Spectral estimators that extend the maximum entropy and maximum likelihood methods."*SIAM J. Applied Math.* **44(2)**, pp. 425–442.

[8] Byrne, C., Levine, B.M., and Dainty, J.C. (1984) "Stable estimation of the probability density function of intensity from photon frequency counts."*JOSA Communications* **1(11)**, pp. 1132–1135.

[9] Byrne, C., and Wells, D. (1985) "Optimality of certain iterative and non-iterative data extrapolation procedures."*Journal of Mathematical Analysis and Applications* **111 (1)**, pp. 26–34.

[10] Byrne, C. and Fiddy, M. (1987) "Estimation of continuous object distributions from Fourier magnitude measurements."*JOSA A* **4**, pp. 412–417.

[11] Byrne, C. and Fiddy, M. (1988) "Images as power spectra; reconstruction as Wiener filter approximation."*Inverse Problems* **4**, pp. 399–409.

[12] Byrne, C. (2005) *Signal Processing: A Mathematical Approach*, AK Peters, Publ., Wellesley, MA.

[13] Byrne, C. (2007) *Applied Iterative Methods*, AK Peters, Publ., Wellesley, MA.

[14] Byrne, C., Haughton, D., and Jiang, T. (1993) "High-resolution inversion of the discrete Poisson and binomial transformations."*Inverse Problems* **9**, pp. 39–56.

[15] Candès, E., Wakin, M., and Boyd, S. (2007) "Enhancing sparsity by reweighted l_1 minimization" preprint available at http://www.acm.caltech.edu/ emmanuel/publications.html .

[16] Donoho, D. (2006) "Compressed sampling" *IEEE Transactions on Information Theory*, **52 (4)**. (download preprints at http://www.stat. stanford.edu/ donoho/Reports).

[17] Fiddy, M. (1983) private communication.

[18] Gerchberg, R. W. (1974) "Super-restoration through error energy reduction."*Optica Acta* **21**, pp. 709–720.

[19] Papoulis, A. (1975) "A new algorithm in spectral analysis and band-limited extrapolation."*IEEE Transactions on Circuits and Systems* **22**, pp. 735–742.

[20] Shieh, M., Byrne, C., Testorf, M., and Fiddy, M. (2006) "Iterative image reconstruction using prior knowledge." *Journal of the Optical Society of America, A*, **23(6)**, pp. 1292–1300.

[21] Shieh, M., Byrne, C., and Fiddy, M. (2006) "Image reconstruction: a unifying model for resolution enhancement and data extrapolation: Tutorial." *Journal of the Optical Society of America, A*, **23(2)**, pp. 258–266.

[22] Shieh, M., and Byrne, C. (2006) "Image reconstruction from limited Fourier data." *Journal of the Optical Society of America, A*, **23(11)**.