

Improving Spectral Resolution Using Basis Selection

B. D. Rao, K. Kreutz-Delgado and S. Dharanipragada*
Electrical and Computer Engineering Dept.
University of California, San Diego
La Jolla, CA 92093-0407
e-mail: {brao, kkreutzd}@ucsd.edu

ABSTRACT

In this paper, we develop high resolution nonparametric spectrum estimation methods using basis selection methodology. As opposed to standard minimization of the ℓ_2 norm of the solution, it is shown that by minimizing suitable diversity measures associated with the linear representation problem one can obtain high resolution spectrum estimates. Algorithms for this purpose are discussed with attention being paid to the robustness issue. In particular, methods are developed to accommodate noise in measurements using a Bayesian framework, and to incorporate statistical averaging using a novel multiple measurement vector framework.

1. INTRODUCTION

The goal of this paper is to develop robust, high resolution nonparametric spectrum estimation methods. This has been a subject of much recent research and interest, and the methods most relevant to this work are those developed in the context of bandlimited extrapolation or reconstruction from partial information [1, 2, 3, 4]. Our goal in this paper is to examine this problem as a basis selection problem [5, 6, 7], and to exploit the developments in this field to provide new insights, and to explore potentially new solutions to this problem. We first motivate the need for basis selection (Section 2), and then show the ability of basis selection methods to provide high resolution spectrum estimates (Section 3). Robust basis selection algorithms are then developed for this purpose in Sections 4 and 5.

2. FFT BASED SPECTRUM COMPUTATION

The DFT is a widely used transform for computing the spectrum of a signal. Given a sequence $y[l]$ of duration m , n samples of the Fourier transform are computed by

*With IBM T. J. Watson Research Center, Yorktown Heights, NY.

zero padding the sequence and computing the DFT via the FFT algorithm. Usually $n \geq m$, and the relationship between the samples $y[l]$ and its DFT coefficients, $Y[k]$, is given by

$$y[l] = \frac{1}{n} \sum_{k=0}^{n-1} Y[k] e^{j \frac{2\pi}{n} kl}, \quad 0 \leq l \leq (m-1). \quad (1)$$

The transform relation, (1), can be written in matrix form, and consists of a *underdetermined* linear system of equations of the form,

$$Ax = b. \quad (2)$$

A is an $m \times n$ matrix with $m \leq n$, and $\text{rank}(A) = m$. The columns of A are formed from the Fourier vectors $\frac{1}{n} [1, e^{j\omega_k}, e^{j2\omega_k}, \dots, e^{j(m-1)\omega_k}]^T$, with the frequencies ω_k uniformly spaced on the unit circle, i.e. $\omega_k = \frac{2\pi}{n}k$, $k = 0, 1, 2, \dots, (n-1)$. b is the $m \times 1$ data vector formed from the data $y[l]$, and x is the $n \times 1$ DFT vector containing the spectrum. The DFT coefficients $Y[k]$, $0 \leq k \leq (n-1)$, are computed as

$$Y[k] = \sum_{l=0}^{n-1} y[l] e^{-j \frac{2\pi}{n} kl} = \sum_{l=0}^{m-1} y[l] e^{-j \frac{2\pi}{n} kl}. \quad (3)$$

The last simplification in (3) is because of the zero padding. Viewed as a solution to the linear inverse problem (2), the DFT solution of (3) can be written as $x_{DFT} = n A^H b$. Because of the orthogonality of the rows of A , the Moore-Penrose pseudo-inverse of A is given by $A^+ = A^H (A A^H)^{-1} = n A^H$. Hence the DFT coefficients can be expressed as the particular solution

$$x_{DFT} = A^+ b. \quad (4)$$

Note that, in fact, there are *many* solutions to the system of equations (2). We see from (4) that zero padding and computing the FFT corresponds to choosing a representation with the *smallest 2-norm* solution from among many possible choices. This choice may be adequate for

most purposes but *not* necessarily for obtaining a high resolution spectrum estimate. In particular for high resolution estimates, imposition of *sparsity* on the solution x is more appropriate than minimizing the 2-norm.

We now demonstrate this fact with the help of a simple example. A sequence $y[l]$ of 32 samples is generated which is a sum of six exponentials and has the form $y[l] = \sum_{r=1}^6 c_r e^{j\frac{2\pi}{32} k_r l}$. The true spectrum of the signal is shown in Fig. 1(a). The corresponding 64 point FFT of the sequence or the minimum 2-norm solution is shown in Fig. 1(b). As is clearly evident, the FFT solution does not have the representational simplicity required of a high resolution estimate. This does not change with the use of a larger length FFT [8]. In terms of the under-determined system of equations, A is a 32×64 matrix and the frequencies that make up the data vector b are contained in the columns of A . Consequently, b can be represented exactly by six columns of A , and a solution vector with only six non-zero entries is a feasible solution to (2). The minimum 2-norm criteria is not appropriate for identifying this solution; minimizing the 2-norm favors a solution with many small non-zero entries. In order to obtain a high resolution estimate, a sparse solution is desired. Hence, basis selection methods offer a potential solution [6, 7, 8].

3. HIGH RESOLUTION SPECTRUM VIA BEST BASIS SELECTION

Best basis selection corresponds to identifying a few columns of the matrix A that best represent the data vector b [5]. This corresponds to finding a solution to (2) with few nonzero entries. In the context of spectrum estimation, this corresponds to identifying a small number of exponentials that best describe the data sequence thereby leading to a high resolution spectrum estimate. The motivation for basis selection, in general, is that a minimal spanning set of basis vectors is usually only adequate to efficiently represent a small class of signals, while forming an overcomplete dictionary using a carefully chosen set of redundant basis vectors can represent a larger class of signals compactly.

Finding an optimal solution to the basis selection problem generally requires a combinatorial search which is computational unattractive. Therefore suboptimal techniques are usually employed and we discuss one such method called FOCUSS, for FOCal Underdetermined System Solver [7, 3, 2], and explore its use in spectrum estimation.

The FOCUSS method was motivated by the observation that if a sparse solution is desired then choosing

a solution based on the smallest 2-norm is not appropriate. The minimum 2-norm criteria favors solutions with many small nonzero entries, a property that is contrary to the goal of sparsity and high resolution [6, 7]. Consequently there is a need to consider the minimization of alternative measures that promote sparsity. In this context, of particular interest are diversity measures, functionals which measure the lack of concentration/sparsity, and algorithms for minimizing them to obtain high resolution solutions. A popular diversity measure is the $\ell_{(p \leq 1)}$ diversity measure given by [9, 10, 11, 12],

$$E^{(p)}(x) = \text{sgn}(p) \sum_{i=1}^n |x[i]|^p, \quad p \leq 1. \quad (5)$$

Minimizing these measures naturally leads to the algorithm FOCUSS, whose iterations are as follows [7, 13, 12]:

$$x_{k+1} = W_{k+1} (AW_{k+1})^+ b, \quad (6)$$

where $W_{k+1} = \text{diag}(|x_k[i]|^{1-\frac{p}{2}})$. Intuitively, the algorithm can be explained by noting that there is competition between the columns of A to represent b . In each iteration, certain columns get emphasized while others are deemphasized. In the end a few columns survive to represent b providing a high resolution estimate.

If the FOCUSS algorithm, with $p = 0$, is applied to the data samples containing the six exponentials, described in the previous section, it converges in four iterations to the true sparse solution. This indicates the potential of this method for high resolution spectrum estimation. More simulation results documenting the properties of FOCUSS can be found in [7, 13]. Results related to minimizing the ℓ_1 norm can be found in [6].

4. BASIS SELECTION IN THE PRESENCE OF NOISE

The previous discussion did not account for noise in the data, i.e. it is assumed that there is a perfect match between the data b and a linear combination of a few columns of A . We extend the FOCUSS method to deal with noise in the measurements using a Bayesian framework. As we will see, the stochastic framework provides theoretical insights and assists in developing robust methods. For this discussion, we assume that

$$b = Ax + v,$$

where v is an additive noise vector. Furthermore, in this formulation x is also assumed to be a random vector independent of v . Under these assumptions, a Maximum

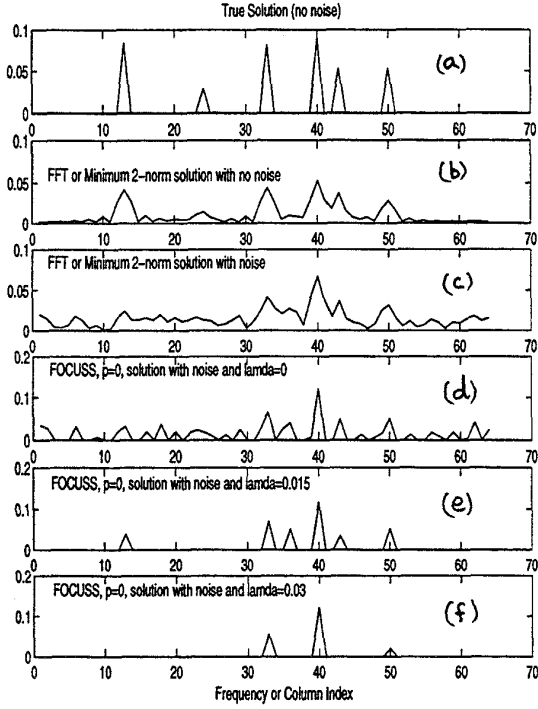


Figure 1: (a) The true spectrum consisting of six exponentials. (b)-(c) The FFT spectrum without and with noise. (d)-(f) The FOCUSS solution for $p = 0$ with no regularization ($\lambda = 0$) and with different choices of λ .

A Posteriori (MAP) estimate of x can be obtained,

$$\begin{aligned} x_{map} &= \arg \max_x \ln p(x|b) \\ &= \arg \max_x [\ln p(b|x) + \ln p(x)] \\ &= \arg \max_x [\ln p_v(b - Ax) + \ln p(x)]. \end{aligned}$$

This formulation is general with considerable flexibility. In order to proceed further, some assumptions about the noise v and the solution vector x have to be made. The distribution of v is not very critical to the approach except for analytical and computational tractability. We assume that v is a Gaussian random vector with i.i.d elements¹, i.e. $p_v(v) = c_1 e^{-\frac{\|v\|^2}{2\sigma^2}}$. The distribution of x is quite important for the generation of sparse solutions. For this purpose, the elements $x[i]$ are assumed to be i.i.d. random variables with density function concentrated near the origin to promote sparsity [14]. A distribution consistent with the $\ell_{(p \leq 1)}$ diversity measures is $p_x(x) = c_2 e^{-\frac{\beta}{2} \text{sgn}(p) \sum_{i=1}^n |x[i]|^p}$, where $\beta > 0$

¹More general Gaussian distributions can be also easily dealt with.

and c_2 is the normalizing constant. Substituting these densities in the expression for the MAP estimate results in

$$x_{map} = \arg \min_x J(x),$$

$$\text{where } J(x) = \left[\|Ax - b\|^2 + \text{sgn}(p)\sigma^2\beta \sum_{i=1}^n |x[i]|^p \right].$$

Note that $p = 2$ gives rise to the standard regularized least squares problem. For $p \leq 1$ it can be shown that the minima of $J(x)$ are sparse. Following the factored-gradient approach in [13, 12], an iterative algorithm can be derived to minimize $J(x)$. A necessary condition for the optimum solution x_* is that it satisfies

$$\nabla_x J(x_*) = 2A^T A x_* - 2A^T b + 2\lambda \Pi(x_*) x_* = 0, \quad (7)$$

where $\lambda = \frac{|p|}{2} \sigma^2 \beta$ and $\Pi(x) = \text{diag}(|x[i]|^{p-2})$. For ease of manipulation, we define a scaling matrix $W(x) = \text{diag}(|x[i]|^{1-\frac{p}{2}})$. Substituting $\Pi(x) = W^{-2}(x)$ in (7) and after some simple manipulation we have

$$((AW(x_*))^T (AW(x_*) + \lambda I) W^{-1}(x_*) x_* = (AW(x_*))^T b.$$

Hence the optimum solution satisfies

$$x_* = W(x_*) ((AW(x_*))^T (AW(x_*) + \lambda I)^{-1} (AW(x_*))^T b.$$

This suggests the following iterative algorithm²:

$$x_{k+1} = W_{k+1} (A_{k+1}^T A_{k+1} + \lambda I)^{-1} A_{k+1}^T b, \quad (8)$$

where $A_{k+1} = AW_{k+1}$ with $W_{k+1} = \text{diag}(|x_k[i]|^{1-\frac{p}{2}})$. The algorithm (8) is the same as that in [7], wherein the procedure is presented as Tikhonov regularization applied to (6). Here, the derivation is based on a Bayesian framework followed by an algorithm derivation employing the factored gradient approach. When the noise level is reduced, $\sigma \rightarrow 0$, then $\lambda \rightarrow 0$ and the algorithm reduces to the original FOCUSS algorithm (6).

The results obtained by applying the regularized algorithm to the data consisting of six exponentials corrupted by white noise are shown in Fig. 1. The minimum norm solution corresponding to the noisy data is shown in Figure 1(c). The results of FOCUSS for $\lambda = 0$ (the original unregularized FOCUSS) and regularized FOCUSS with $\lambda = 0.015$ and 0.03 are shown in Figs. 1(d)-1(f). The solutions obtained are sparse, with the solution becoming more sparse as λ is increased. This is consistent with our expectations and with the proper choice of λ , one can deal effectively with white noise perturbations. Choice of λ is an important issue for obtaining a good solution. The L-curve criteria developed in [15], and used in [7], provides a potential solution.

²When the elements of A and b are complex, the transpose operation has to be replaced by the Hermitian transpose

5. BASIS SELECTION EMPLOYING DATA SEGMENTATION

Another approach for obtaining statistical reliability is to segment the data and to compute a meaningful average of the solutions found for the individual segments, e.g. the modified periodogram. In the basis selection framework, this requires solving a sparse solution to the following system of equations [16]:

$$Ax^{(l)} = b^{(l)}, l = 1, \dots, L, \text{ or } AX = B, \quad (9)$$

where $X = [x^{(1)}, \dots, x^{(L)}]$, and $B = [b^{(1)}, \dots, b^{(L)}]$. This problem is referred to as the Multiple Measurement Vector (MMV) problem. L is the number of segments into which the data is segmented, and is assumed to be much less than m . For a fixed data record, increasing the number of segments reduces the segment data length. Overlapping the segments is a useful way to address this difficulty.

For high resolution, the desired solution requires not only that the individual columns of X , $x^{(l)}$, have a sparse structure but that they share the structure and have a common sparsity profile, i.e. the indices of the nonzero entries are independent of l . In terms of the matrix solution X , this implies having a few nonzero rows as opposed to a few non zero entries. To deal with this problem there is a need to develop new diversity measures that reflect the requirement of few nonzero rows as opposed to simply few nonzero entries. We discuss an extension that meets the sparsity requirements and also results in a simple algorithm [16],

$$J^{(p)}(X) = \sum_{i=1}^n \left(\sum_{l=1}^L |x^{(l)}[i]|^2 \right)^{\frac{p}{2}}.$$

This is an extension of the " $\ell_{(p \leq 1)}$ diversity measures" and as p approaches zero, it provides a count of the number of nonzero rows in X .

Starting from this measure, as in Section 4, the factored-gradient approach is used to develop an algorithm to minimize it. Due to space limitations, we omit the details and present only the algorithm:

$$\begin{aligned} W_{k+1} &= \text{diag}(c_k[i]^{1-\frac{p}{2}}), \text{ where } c_k[i] = \left(\sum_{l=1}^L |x_k^{(l)}[i]|^2 \right)^{\frac{1}{2}} \\ Q_{k+1} &= A_{k+1}^+ B, \text{ where } A_{k+1} = AW_{k+1} \\ X_{k+1} &= W_{k+1} Q_{k+1}. \end{aligned} \quad (10)$$

The algorithm can be initialized by using the minimum Frobenius norm solution or any other suitable solution. The range of p is $p \leq 2$, with p close to zero corresponding to the numerosity measure. The algorithm is

a generalization of the FOCUSS class of algorithms with $L = 1$ corresponding to the original FOCUSS algorithm.

Some simulations supporting the usefulness of the algorithm (10) can be found in [16]. To deal with noise, one can use Tikhonov regularization or Truncated SVD based regularization. The Tikhonov regularization procedure can be derived formally using the Bayesian formulation as in Section 4.

6. REFERENCES

- [1] A. Papoulis and C. Chamzas. "Detection of Hidden Periodicities by Adaptive Extrapolation". *IEEE Trans. on ASSP*, Oct. 1979.
- [2] S. D. Cabrera and T. W. Parks. "Extrapolation and Spectral Estimation with Iterative weighted norm modification". *IEEE Trans. on ASSP*, April 1991.
- [3] H. Lee, et al. "Improvement of discrete band-limited signal extrapolation by iterative subspace modification". In *Proc. ICASSP*, April 1987.
- [4] A. K. Jain and S. Ranganath. "Extrapolation Algorithms for Discrete Signals with Application in Spectral Estimation". *IEEE Trans. on ASSP*, 1981.
- [5] S. G. Mallat and Z. Zhang. "Matching Pursuits with Time-Frequency Dictionaries". *IEEE Trans. ASSP*, 41(12):3397-3415, Dec. 1993.
- [6] S. Chen and D. Donoho. "Application of Basis Pursuit in Spectrum Estimation". In *Proc. ICASSP*, May 1998.
- [7] I.F. Gorodnitsky and B.D. Rao. "Sparse Signal Reconstructions from Limited Data using FOCUSS: A Reweighted Minimum Norm Algorithm". *IEEE Trans. on Signal Processing*, 45:600-616, March 1997.
- [8] B.D. Rao. "Signal Processing with the Sparseness Constraint". In *Proc. ICASSP*, Seattle, WA, May 1998.
- [9] M. V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*. A. K. Peters, Wellesley, MA, 1994.
- [10] B. D. Jeffs. "Sparse Inverse Solution Methods for Signal and Image Processing Applications". *ICASSP 98*.
- [11] G. Harikumar and Y. Bresler. "A New Algorithm for Computing Sparse Solutions to Linear Inverse Problems". In *Proc. ICASSP 96*.
- [12] K. Kreutz-Delgado and B. D. Rao. "Measures and Algorithms for Best Basis Selection". *ICASSP 98*.
- [13] B. D. Rao and K. Kreutz-Delgado. "An Affine Scaling Methodology for Best Basis Selection". *to appear in the IEEE Trans. on Signal Processing*, 1997.
- [14] B. A. Olshausen and D. J. Field. "Sparse Coding with an Overcomplete Basis Set: A strategy employed in V1". *In Press*, 1997.
- [15] P. C. Hansen. "Analysis of the Discrete Ill-Posed Problems by Means of the L-Curve". *SIAM Review*, 1992.
- [16] B.D. Rao and K. Kreutz-Delgado. "Sparse Solutions to Linear Inverse Problems with Multiple Measurement vectors". In *IEEE DSP Workshop*, Aug. 1998.