SPARSE INVERSE SOLUTION METHODS FOR SIGNAL AND IMAGE PROCESSING APPLICATIONS

Brian D. Jeffs *

Department of Electrical and Computer Engineering
459 CB, Brigham Young University
Provo, UT 84602, USA. E-mail: bjeffs@ee.byu.edu.

ABSTRACT

This paper addresses image and signal processing problems where the result most consistent with prior knowledge is the minimum order, or "maximally sparse" solution. These problems arise in such diverse areas as astronomical star image deblurring, neuromagnetic image reconstruction, seismic deconvolution, and thinned array beamformer design. An optimization theoretic formulation for sparse solutions is presented, and its relationship to the MUSIC algorithm is discussed. Two algorithms for sparse inverse problems are introduced, and examples of their application to beamforming array design and star image deblurring are presented.

1. INTRODUCTION

Many important signal processing applications are formulated as ill-posed inverse problems which have no unique solution unless some regularizing prior constraint is imposed. In a surprising number of these cases, the appropriate constraint which agrees with our prior knowledge is that the solution must be "sparse." By sparse we mean that the solution is minimum order, or has the maximum number of zero valued elements consistent with observed data and the system model.

Examples of applications best solved using a sparseness requirement include blurred point source image restoration (e.g. star field images), neuromagnetic image reconstruction to map brain electrical activity, seismic deconvolution, overlapping specular echo resolution in SONAR or RADAR, and thinned array beamformer design.

Unfortunately, sparse solutions are usually both analytically and computationally more challenging than typical optimization problems. In this paper we review some of our work in the area, and demonstrate by example how widely applicable sparse optimization methods are in realistic DSP applications. Two very different algorithms for finding sparse solutions to ill-posed inverse problems will be presented.

In each of the applications discussed below, we adopt the discrete linear observation model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \eta \tag{1}$$

*The author would like to acknowledge the contributions to this work by Thomas Hebert (who suggested looking at linear programming to solve the maximally sparse problem), Metin Gunsay, and Richard Leahy.

where $\mathbf{y}$ is the $M \times 1$ observed data vector, $\mathbf{x}$ is the $N \times 1$ desired, unobserved source vector, η is additive noise or measurement error, and $\mathbf{H}$ is the system response matrix. Typically $\mathbf{H}$ is singular, and the problem of finding $\mathbf{x}$ given $\mathbf{y}$ is highly ill posed.

The minimum order solution consistent with the observed data and signal model (1) is given by the following simple and heuristically satisfying formulation

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} I\{x_n\} \text{ such that } \psi\{\mathbf{y} - \mathbf{H}\mathbf{x}\} \leq \epsilon \tag{2}$$

where $I\{\cdot\}$ is the indicator function, which under the summation counts the number of non-zero elements in $\mathbf{x}$, and $\psi\{\cdot\}$ is some distance metric, chosen to be appropriate for the application of interest. Enforcing $\psi\{\mathbf{y} - \mathbf{H}\mathbf{x}\} \leq \epsilon$ provides a fidelity constraint on the solution. We shall refer to $\hat{\mathbf{x}}$ obtained from equation (2) as the *maximally sparse solution.* Using $I\{\cdot\}$ as the optimization objective forces $\hat{\mathbf{x}}$ to have the fewest non-zero terms without violating the fidelity constraint.

2. ALGORITHMS FOR SPARSE SOLUTIONS

2.1. Maximally Sparse Simplex Search

If we define the fidelity distance constraint based on element-wise absolute values, then $\psi\{\cdot\} = |\cdot|$, and a computationally efficient simplex search algorithm related to linear programming is possible. Equation (2) becomes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} |x_n|^p \text{ s.t. } |\mathbf{y} - \mathbf{H}\mathbf{x}| \leq \epsilon,\ 0 < p < 1 \tag{3}$$

where ϵ is a non-negative vector. Note that the indicator function has been replaced with the approximation $|x_n|^p$, which is differentiable (almost everywhere) and simplifies algorithm development and analysis. We have shown that for $0 < p < 1$, $\arg\min\{\sum_{n=1}^{N} |x_n|^p\} \approx \arg\min\{\sum_{n=1}^{N} I\{x_n\}\}$, with equality holding for a sufficiently small value of p [3].

To solve equation (3) we express it in the equivalent equality constraint form [3] [2]

$$\hat{\mathbf{u}} \quad = \quad \arg\min_{\mathbf{u}} \sum_{i=1}^{2N} (u_i)^p \text{ s.t. } \mathcal{H}\mathbf{u} = \mathbf{b},\ \mathbf{u} \geq \mathbf{0} \tag{4}$$

0-7803-4428-6/98 $10.00 © 1998 IEEE

$$\mathcal{H} = \begin{bmatrix} \mathbf{H} & -\mathbf{H} & \mathbf{I} & \mathbf{0} \\ \mathbf{H} & -\mathbf{H} & \mathbf{0} & -\mathbf{I} \end{bmatrix}, \mathbf{u} = \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \\ \mathbf{s}^+ \\ \mathbf{s}^- \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{y} + \epsilon \\ \mathbf{y} - \epsilon \end{bmatrix}$$

where $\mathbf{s}^+$ and $\mathbf{s}^-$ are arbitrary, non-negative "slack variables" which are non-zero only when the constraint of equation (3) is satisfied in the inequality rather than the equality sense. The estimate of $\mathbf{x}$ which solves (3) is given by $\hat{\mathbf{x}} = \hat{\mathbf{x}}^+ - \hat{\mathbf{x}}^-$.

Since $\sum_{i=1}^{2N}(u_i)^p$ is concave over $\mathbf{u} \geq \mathbf{0}$, the solution to (4) must lie at one of the finite number of extreme points of the constraint set, i.e. at a vertex of the convex polytope defined by $\mathcal{H}\mathbf{u} = \mathbf{b}$ [4] [3]. These extreme points are called *basic solutions*, $\mathbf{u}_B$, and have at most $2M$ non-zero elements. A basic solution may be computed by partitioning $\mathcal{H}$ as $\mathcal{H} = [\mathbf{C}|\mathbf{D}]$, where $\mathbf{C}$ has dimensions $2M \times 2M$ and contains the basis vectors for the solution. Multiplying the constraint of (4) by $\mathbf{C}^{-1}$ leads directly to the obvious basic solution,

$$\begin{aligned} [\mathbf{I}|\mathbf{C}^{-1}\mathbf{D}]\mathbf{u} &= \mathbf{C}^{-1}\mathbf{b}, \\ \mathbf{u}_B &= [(\mathbf{C}^{-1}\mathbf{b})^T, 0, \cdots, 0]^T \end{aligned} \quad (5)$$

Any choice of $2M$ columns from $\mathcal{H}$ for which $\mathbf{C}$ is non-singular is acceptable for a solution $\mathbf{u}_B$, and elements of $\mathbf{u}$ are reordered to correspond to the ordering of columns in $\mathbf{C}$. The simplex search algorithm finds a globally optimal solution by repeatedly moving from one basic solution vertex to some neighboring vertex which reduces the objective function $\sum_{i=1}^{2N}(u_i)^p$. This is accomplished using computationally efficient pivoting operations as in standard linear programming [4]. In each iteration of the simplex search, one column from the basis, $\mathbf{C}$, is swapped (pivoted) with a column from $\mathbf{D}$. Columns are chosen to reduce $\sum_{i=1}^{2N}(u_i)^p$ with each such pivot. Pivoting continues until no neighboring vertices reduce to objective.

2.2. Eigen-space Parametric Methods

In this section we will show an equivalence relationship between maximally sparse optimization and parametric subspace decomposition algorithms (such as MUSIC [5]) commonly used in direction of arrival (DOA) estimation. The benefit of this relationship is that MUSIC and other related algorithms are more computationally efficient than the simplex search of Section 2.1..

The usual signal model for parametric methods is

$$\begin{aligned} \mathbf{y} &= \mathbf{A}(\Theta)\,\mathbf{s} + \eta, \text{ where} \\ \mathbf{A}(\Theta) &= [\mathbf{a}(\theta_1)|\cdots|\mathbf{a}(\theta_P)],\ \Theta = [\theta_1, \ldots \theta_P]^T. \end{aligned} \quad (6)$$

$\mathbf{A}(\Theta)$ is the parametric system matrix, Θ is the parameter vector to be estimated, and $\mathbf{s}$ is the $P \times 1$ source amplitude vector. For DOA estimation, θ_p is the direction of arrival for a plane wave signal from the p^{th} source with corresponding amplitude s_p. $\mathbf{a}(\theta_p)$ is the response across a fixed array of sensors to a unit amplitude source in direction θ_p. P, Θ, and $\mathbf{s}$ are unknown. Implicit in this model is that all s_p are non-zero, $P < M$, and that the $\mathbf{a}(\theta_p)$ are mutually linearly independent. Though seldom recognized, this model has all the earmarks of a sparse system.

Wax proposed an optimal ML joint detection and estimation method based on the Akaike Information Criterion [7]. By detection we mean determining the number of sources present (P), and by estimation we mean finding the corresponding θ_p's. Assuming η is i.i.d. Gaussian noise, Wax's joint estimator becomes

$$[\hat{P}, \hat{\Theta}, \hat{\mathbf{s}}]_{ML} = \arg\min_{P,\Theta,\mathbf{s}} d(P) + \lambda||\mathbf{y} - \mathbf{A}(\Theta)\,\mathbf{s}||^2. \quad (7)$$

where $d(P)$ is the number of degrees of freedom in Θ as a function of P, and λ is a constant that depends on noise variance.

In practical implementations of MUSIC, DOA estimates are formed by "scanning" θ on a discrete sample grid, ϕ_n, $1 \leq n \leq N$. In other words, in evaluating the parametric model, only sources located in a finite number of directions are considered or permitted. With θ thus quantized, (6) can be represented in the form of (1), i.e. $\mathbf{y} = \mathbf{H}\mathbf{x} + \eta$, where

$$\begin{aligned} \mathbf{H} &= [\mathbf{a}(\phi_1)|\cdots|\mathbf{a}(\phi_N)], \text{ and} \\ \mathbf{x} &= [0\cdots 0, s_1, 0\cdots 0, s_2, \cdots 0, s_P, 0\cdots 0]^T. \end{aligned} \quad (8)$$

$\mathbf{x}$ has exactly P non-zero terms, but the positions and amplitudes of these terms are unknown. $\theta_p = \phi_{n_p}\ \forall p,\ 1 \leq p \leq P$, though the indices n_p, $1 \leq n_p \leq N$, are unknown and must be estimated.

Consider substituting (1) into (7). Each column of $\mathbf{H}$ corresponds to a fixed parameter value, $\theta = \phi_n$. Assuming that $\mathbf{A}(\Theta)$ is rank P, and columns of $\mathbf{H}$ are linearly independent, then $d(P)$ is equal to the number of columns of $\mathbf{H}$ that are not multiplied by zero in the product $\mathbf{H}\mathbf{x}$. It follows then that $d(P) = \sum_{n=1}^{N} I\{x_n\}$. Also, minimizing with respect to Θ and $\mathbf{s}$ in (7) is equivalent to minimizing with respect to $\mathbf{x}$ alone. This is true since by choosing any x_{n_p} to be non-zero, a corresponding $\theta_p = \phi_{n_p}$ is implicitly selected as the parameter value associated with s_p and column n_p of $\mathbf{H}$. Likewise, by solving for amplitudes, x_{n_p}, optimizing over the corresponding s_p is accomplished. Thus, an equivalent representation for equation (7) is

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \sum_{n=1}^{N} I\{x_n\} + \lambda||\mathbf{y} - \mathbf{H}\mathbf{x}||^2, \quad (9) \\ [\hat{\Theta}, \hat{\mathbf{s}}]_{ML} &= \{\phi_{n_p}, \hat{x}_{n_p} \,|\, \forall\, n_p \text{ s.t. } \hat{x}_{n_p} \neq 0\}, \\ \hat{P}_{ML} &= \sum_{n=1}^{N} I\{x_n\} \end{aligned}$$

The constrained minimization of equation (2) may be solved as an unconstrained minimization using the penalty method [4]. Defining $\psi\{\mathbf{y} - \mathbf{H}\mathbf{x}\} = ||\mathbf{y} - \mathbf{H}\mathbf{x}||^2$, and using the penalty method, equation (2) yields equation (9) exactly, with λ as the penalty weight. Thus we see that for a discrete parameter space, and Gaussian noise, optimal joint detection and estimation using the parametric model of (6) is equivalent to maximally sparse optimization with a squared error constraint.

A MUSIC based solution to (7) is

$$\hat{\Theta}_{MU} = [\arg \text{ local} \min_{\phi_n} \mathbf{a}(\phi_n)^H \mathbf{E}_\eta \mathbf{E}_\eta^H \mathbf{a}(\phi_n)]^T \quad (10)$$
$$\hat{\mathbf{s}}_{MU} = \mathbf{A}^\dagger(\hat{\Theta})\mathbf{y},\ \hat{P}_{MU} = \arg\min_P AIC\{P\}$$

where $\mathbf{R}_y[\mathbf{E}_P|\mathbf{E}_\eta] = [\mathbf{E}_P|\mathbf{E}_\eta]\Lambda$, $\mathbf{E}_P$ is the matrix of eigenvectors corresponding to the P largest eigenvalues of covariance matrix $\mathbf{R}_y$, Λ is the diagonal matrix of ordered eigenvalues [5]. $AIC\{\cdot\}$ indicates the Akaike Information Criterion (or other order estimation metric).

MUSIC achieves efficiency by decoupling the detection and estimation parts of the problem, thus avoiding the exhaustive search implied by (9). Stoica and Nehorai showed that if the s_p are mutually uncorrelated, $\hat{\Theta}_{MU}$ is a large sample realization of $\hat{\Theta}_{ML}$ [6]. Therefor, MUSIC is a practical approximate solution to the maximally sparse problem of (2), when the covariance of $\mathbf{y}$ can be estimated.

3. APPLICATIONS

3.1. Sparse Beamforming Array Design

The problem of designing antenna arrays in which the elements are not equally spaced has been studied widely over the last 30+ years. A prime motivation for this work is the possibility of improved resolution or reduced sidelobe levels as compared to an equispaced array with the same number of elements. In this section we present a method for designing symmetric narrowband beamforming arrays using the minimum number of elements required to satisfy directional response specifications [3].

Let x_n be the real beamformer weight (shade) for a sensor element at 3-D position $\mathbf{r}_n$ in an arbitrary N element symmetric array. Let $\mathbf{s}_{\theta,\varphi}$ be a unit direction vector for azimuth θ and elevation φ. Assume the array is pre-steered to a desired mainlobe direction $\mathbf{s}_d = \mathbf{s}_{\theta_d,\varphi_d}$ by complex phasing. The beamformer response to a far field unit amplitude source in direction (θ_m, φ_m) is then

$$y(\theta_m, \varphi_m) = \sum_{n=1}^{N} 2x_n \cos k\mathbf{r}_n \cdot (\mathbf{s}_{\theta_m,\varphi_m} - \mathbf{s}_d) \qquad (11)$$

where k is the propagation wavenumber, and '·' indicates vector dot product.

For sparse design, the desired response is specified on a sufficiently dense lattice of constraint directions, $\mathbf{y}_d = [y(\theta_1, \varphi_1), \cdots y(\theta_M, \varphi_M)]^T$, and a dense grid of potential element positions, $\mathbf{r}_n$ is specified. Equation (11) becomes, $\mathbf{y} = \mathbf{Hx}$, where the m^{th} row of $\mathbf{H}$ is given by $\mathbf{h}_m^T = 2[\cos k\mathbf{r}_1 \cdot (\mathbf{s}_{\theta_m,\varphi_m} - \mathbf{s}_d), \cdots \cos k\mathbf{r}_N \cdot (\mathbf{s}_{\theta_m,\varphi_m} - \mathbf{s}_d)]^T$. The sparse array design can be expressed as in equation (3) by simply substituting $\mathbf{y}_d$ for $\mathbf{y}$. The simplex search algorithm is then used to solve (3). The algorithm sets as many element weights as possible to zero on the dense initial array grid. Zero weighted elements can then be removed from the array. The result is a minimum order, non-uniform array design.

Figure 1 shows the final positions for the 16 element solution of an example of sparse array design. The initial dense element grid was a filled disc, including 60 co-planar sensors as shown. The beam is steered in the plane of the sensors, along the positive y axis. The desired response specification was that the mainlobe width and maximum sidelobe level were not to be degraded as compared to that obtained from the dense initial array using unity shading.

Figure 2 shows the beam response pattern for the thinned array of Figure 1. Both element shading and position were obtained using the simplex search algorithm. The desired mainlobe and sidelobe constraints were achieved with a savings in array complexity of nearly 4 to 1 as compared with the initial array.

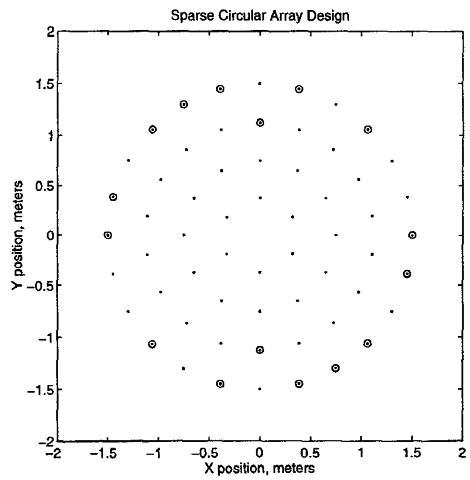

Figure 1. Thinned array design obtained using the simplex search algorithm. ' · ' indicates the original element grid. 'o' indicates an element in the sparse design.

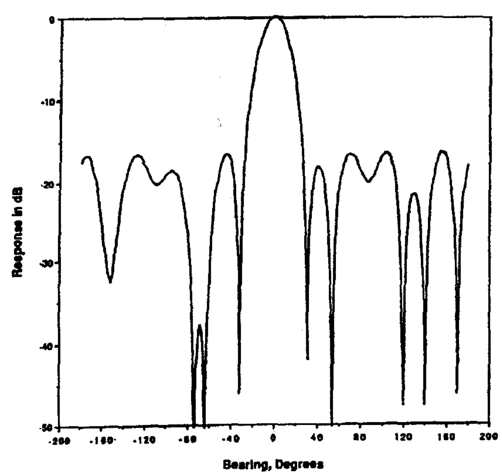

Figure 2. Beam response pattern for the sparse array of Figure 1.

3.2. Point Source Image Restoration

Most well know restoration algorithms are ill suited for blurred point source images, such as star fields or IR target tracking images. Methods which incorporate prior knowledge that the image is sparse, or point-like, are much better able to resolve closely spaced objects like binary stars blurred by atmospheric turbulence. This section presents two approaches to point-source image deblurring; one based on maximally sparse simplex search, and one using eigenspace parametric methods.

Application of the simplex search to image restoration is straightforward, with minor adaptations of the existing formulation [2]. Assuming the blur can be modeled as a shift invariant linear operation, then $\mathbf{H}$ of equation (1) is simply the doubly block Toeplitz convolution matrix constructed from the blur discrete point spread function (psf). For intensity images, we must constrain $\mathbf{x}$ to be non-negative, so we drop $\mathbf{x}^-$ from $\mathbf{u}$, and the corresponding block column from $\mathcal{H}$ in (4).

Figures 3 and 4 present an example of star field deblurring using the simplex search algorithm on telescopic data.

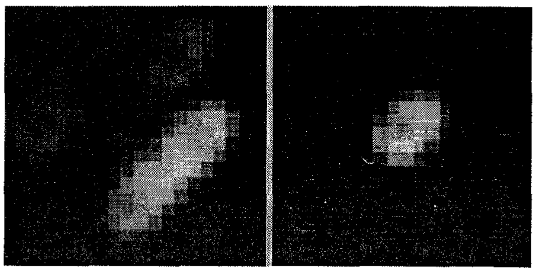

Figure 3. Star cluster image from the CWRU/NOAO observatory, with 2.1 arcseconds resolution per pixel. a) [left] blurred image. b) [right] psf estimate from an isolated star in the same frame.

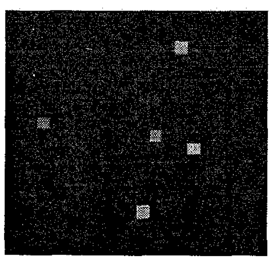

Figure 4. Maximally sparse simplex search restoration of Figure 3a. Note that five distinct stars were identified.

The MUSIC based eigenspace approach can also be used to restore blurred point source images [1]. However, often only a single image frame is available, and the underlying point sources are constant valued and deterministic (rather than random processes as assumed in the parametric model). This implies that direct computation of $\mathbf{R}_y$ would yield a degenerate rank-one signal subspace, $\mathbf{E}_P$, and MUSIC would fail.

In [1] we presented a method for transforming the image into the frequency domain and performing generalized 2-D smoothing to form a full rank covariance matrix. Using this smoothed covariance, the MUSIC approach described above may be applied. One advantage of this approach is that the point source positions are not limited to the original pixel sample grid, and may be scanned at any desired resolution. We have shown this permits sub-pixel super resolution in determining source positions [1].

Figure 5 shows a 2-D generalized frequency domain MUSIC reconstruction of the image data in Figure 3. Note that stars are located on a grid with four times the resolution of Figure 4 in each dimension. One of the stars identified with the simplex search is not included in the MUSIC result. This implies that P may have been under-estimated with the AIC criterion. Also, a higher setting for ϵ in (3) causes the simplex search to drop this middle star. Experiments with synthetic data where the "ground truth" is known show that both algorithms do an excellent job of recovering the true image.

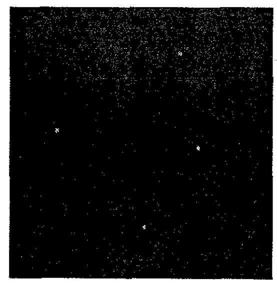

Figure 5. MUSIC based eigenspace sparse restoration of the data in Figure 4.

REFERENCES

[1] M. Gunsay and B.D. Jeffs, "Point Source Localization in Blurred Images by a Frequency Domain Eigenvector Based Method," *IEEE Trans. on Image Proc.*, scheduled for publication, December, 1995.

[2] B.D. Jeffs and M. Gunsay, "Restoration of Blurred Star Field Images by Maximally Sparse Optimization," *IEEE Trans. on Image Proc.*, Vol. 2, No. 2, Apr. 1993, pp. 202- 211.

[3] R.M. Leahy and B.D. Jeffs, "On the Design of Maximally Sparse Beamforming Arrays," *IEEE Trans. on Antennas and Prop.*, Vol. 39, No. 8, Aug. 1991, pp. 1178-1187.

[4] D.G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, 1984, Reading Mass.

[5] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," *Proc. RADC Spectral Est. Workshop*, Oct. 1979, pp. 243-258.

[6] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood, and Cramer-Rao Bound," *IEEE Trans. on Acous., Speech, and Signal Proc.*, Vol. 37, No. 5, May 1989, pp. 720-741.

[7] M. Wax, *Detection and Estimation of Superimposed Signals*, Ph.D. Dissertation, Stanford University, March, 1985.