SPARSE SOLUTIONS TO LINEAR INVERSE PROBLEMS WITH MULTIPLE MEASUREMENT VECTORS

B.D. Rao and K. Kreutz-Delgado

Electrical and Computer Engineering Department
University of California, San Diego
La Jolla, California 92093-0407
E-Mail: {brao, kreutz}@ece.ucsd.edu

ABSTRACT

This paper describes algorithms for computing sparse solutions to linear inverse problems when there are multiple measurement vectors. Extensions to the forward sequential basis selection methods (such as the matching pursuit and order recursive matching methods) are developed for this purpose. In addition, new diversity measures are defined, and algorithms are developed to minimize them and obtain sparse solutions. The resulting algorithms are an extension of the FOCUSS class of algorithms. Computer simulations are provided to support the methods.

1. INTRODUCTION

The problem of computing sparse solutions to linear inverse problems has received attention because of its application to signal representation, signal reconstruction, signal coding etc. Both sequential and parallel methods have been developed to deal with this problem [1, 2, 3, 4, 5, 6, 7]. We consider an important variation of this problem, namely that of computing sparse solutions to linear inverse problems when there are **m**ultiple **m**easurement **v**ectors (MMV). This formulation was initially motivated by the need to solve the neuromagnetic inverse problems that arises in Magnetoencephalography (MEG), a modality for imaging the brain [8]. However, the framework is quite general and the algorithms developed have applications in signal representation, signal reconstruction, and array processing, among others. Because of the intimate connection between the problem of computing sparse solutions to linear inverse problems and the best basis selection problem, the methods developed also extend the best basis selection framework.

2. PROBLEM FORMULATION

Mathematically, the MMV problem can be stated as solving the following system of equations [9].

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

where $X = [x^{(1)}, .., x^{(L)}]$, and $B = [b^{(1)}, .., b^{(L)}]$. L is the number of measurement vectors and is usually assumed to be much less than m. Past work has mainly dealt with the problem of one measurement vector, i.e. $L = 1$. Here we concentrate on the case when $L \geq 1$ with $b^{(l)}$ denoting the lth measurement vector, and $x^{(l)}$ the corresponding solution. A is a $m \times n$ matrix which is known and obtained using the physics of the problem. It is also assumed that $m << n$ and rank$(A) = m$.

There are many solutions to (1), and the desired solution is characterized by two distinct and important features.

1. The solution vectors $x^{(l)}$ are sparse, i.e. most of the entries are zero. This requirement is the same as that imposed in the single measurement vector case. Often, the sparsest solution will be of interest.
2. The solution vectors $x^{(l)}$ are assumed to have a common sparsity profile, i.e. the indices of the nonzero entries are independent of l. This requirement provides the coupling between the vectors and is the source of the additional complexity.

Some results concerning the theoretical uniqueness of such sparse solutions can be found in [10, 9]. Based on the results related to the problem of using a single measurement vector it is clear that an approach based on a exhaustive search strategy is computationally infeasible calling for simpler and yet reliable techniques. The advantage of having more measurement vectors is the enhanced ability to find sparse solutions having the correct sparsity profile through suboptimal techniques and potential robustness in the presence of noise. To address this problem, we extend both the basis pursuit methodology and methods based on minimizing diversity measures [11, 12].

3. FORWARD SEQUENTIAL SELECTION METHODS

The methods described in this section find a sparse solution by selecting a small subset of columns vectors from A to best represent B. Selection of a column corresponds to selecting a nonzero row of X. The vectors are selected sequentially, i.e. the basis set is built up one vector at a time. The variants differ mainly in the criteria used to select a column vector, and result in methods of differing complexity and ability to select a compact basis set.

To facilitate the presentation, we first develop some notation. The residual vector after the pth iteration is denoted by B_p, with $B_0 = B$. The lth column of B_p is denoted by $b_p^{(l)}$. The indices of the p vectors selected are stored in the index set denoted by I_p, where $I_p = \{k_1, k_2, .., k_p\}$, $I_0 = \emptyset$, and the vectors themselves are stored as columns in the matrix S_p, i.e. $S_p = \left[a_{k_1}, a_{k_2}, ..., a_{k_p}\right]$, $S_0 = \emptyset$. The orthogonal projection matrix onto the range space of S_p is denoted by P_{S_p} and its orthogonal complement $P_{S_p}^{\perp} = \left(I - P_{S_p}\right)$, $P_{S_0} = 0$, $P_{S_0}^{\perp} = I$. Without loss of generality, it is assumed that the columns of the matrix A are of unit norm.

3.1. The MMV Basic Matching Pursuit (M-BMP)

This method is an extension of the matching pursuit approach suggested in [1, 13]. In this selection method, in the pth iteration the vector most closely aligned with the residual B_{p-1} is chosen. This is accomplished by examining the residual $E_{p,k} = P^{\perp}_{a_k} B_{p-1}$, where $P_{a_k} = a_k a_k^H$ is the projection matrix onto the space spanned by the vector a_k, and selecting a column vector from A that minimizes the Frobenius norm of the error the most.

$$\begin{aligned} \|E_{p,k}\|_F^2 &= \text{tr}(E_{p,k}^H E_{p,k}) = \text{tr}(B_{p-1}^H P^{\perp}_{a_k} B_{p-1}) \\ &= \|B_{p-1}\|_F^2 - \text{tr}(B_{p-1}^H P_{a_k} B_{p-1}). \end{aligned}$$

The minimization is achieved by maximizing the second term, $\text{tr}(B_{p-1}^H P_{a_k} B_{p-1})$, in the above expression. Using the fact that $P_{a_k} = a_k a_k^H$, the computation involved for the selection is

$$k_p = \arg\max_r \sum_{l=1}^{L} |a_r^H b_{p-1}^{(l)}|^2. \tag{2}$$

If $k_p \notin I_{p-1}$, then the index and basis sets are updated, i.e. $I_p = I_{p-1} \cup k_p$, and $S_p = [S_{p-1}, a_{k_p}]$. Otherwise $I_p = I_{p-1}$ and $S_p = S_{p-1}$. The new residual vector is then computed as $B_p = P^{\perp}_{a_k} B_{p-1}$ or more explicitly

$$b_p^{(l)} = P^{\perp}_{a_{k_p}} b_{p-1}^{(l)} = b_{p-1}^{(l)} - (a_{k_p}^H b_{p-1}^{(l)}) a_{k_p}, \;\; l = 1, \ldots, L. \tag{3}$$

Equations (2) and (3) give the M-BMP algorithm (with $B_0 = B$). The procedure terminates when either $p = r$ (for specified sparsity index r) or $\|B_p\|_F \leq \epsilon$ (for specified ϵ).

3.2. The MMV Order Recursive Matching Pursuit (M-ORMP)

This method employs a more comprehensive strategy for selecting a new column, and is an extension of the methodology developed in [3, 4, 13]. In this method, the pursuit of the matching pth basis vector conceptually involves solving $(n - p + 1)$ order recursive least squares problems of the type $\min_Y \|[S_{p-1}, a_r]Y - B\|_F^2$, and selecting the vector $a_r \notin S_{p-1}$ that reduces the residual the most. Because of the similarity with the derivation with the single measurement vector case [5], and due to space considerations, we omit the details and only summarize the main steps in the algorithm.

With the initialization $a_k^{(0)} = a_k, k = 1, \ldots, n$ and $B_0 = B$, the index selection criteria in the pth iteration is given by

$$k_p = \arg\max_r \frac{\sum_{l=1}^{L} |(a_r^{(p-1)})^H b_{p-1}^{(l)}|^2}{\|a_r^{(p-1)}\|^2}, \; r \notin I_{p-1}, \tag{4}$$

resulting in $I_p = I_{p-1} \cup k_p$, $S_p = [S_{p-1}, a_{k_p}]$, and $P_{S_p} = P_{S_p,k_p} = P_{S_{p-1}} + q_p q_p^H$ where

$$q_p \equiv \frac{a_{k_p}^{(p-1)}}{\|a_{k_p}^{(p-1)}\|}. \tag{5}$$

The pth iteration is completed by projecting the column vectors of A and B.

$$a_l^{(p)} = P^{\perp}_{S_p} a_l^{(p-1)} = a_l^{(p-1)} - (q_p^H a_l^{(p-1)}) q_p. \tag{6}$$

Similarly, the residual vector $b_p^{(l)}$ are recursively computed as

$$b_p^{(l)} = P^{\perp}_{S_p} b_{p-1}^{(l)} = b_{p-1}^{(l)} - (q_p^H b_{p-1}^{(l)}) q_p, \; l = 1, \ldots, L. \tag{7}$$

Equations (4)–(7) constitute the M-ORMP algorithm. The termination procedure is the same as that for the M-BMP algorithm.

3.3. The MMV Modified Matching Pursuit (M-MMP)

This procedure, also referred to as the orthogonalized matching pursuit [14], is a minor modification of the BMP method and seeks to improve the computation of the residue B_{p-1} [5]. The index selection procedure involves computation as in (2) of M-BMP, but the residual matrix B_p is computed as $P^{\perp}_{S_p} B_{p-1}$ as opposed to $P^{\perp}_{a_{k_p}} B_{p-1}$. This residual computation is carried out by first carrying out a Modified Gram-Schmidt type of procedure on the vector a_{k_p} selected, More precisely, with the initialization $\hat{a}_{k_p}^{(0)} = a_{k_p}, q_0 = 0$, we have $P_{S_p} = P_{S_p,k_p} = P_{S_{p-1}} + q_p q_p^H$ where

$$\begin{aligned} \hat{a}_{k_p}^{(\ell)} &= \hat{a}_{k_p}^{(\ell-1)} - (q_{\ell-1}^H \hat{a}_{k_p}^{(\ell-1)}) q_{\ell-1}, \; \ell = 1, .., p \qquad (8) \\ q_p &= \frac{\hat{a}_{k_p}^{(p)}}{\|\hat{a}_{k_p}^{(p)}\|} \end{aligned}$$

The residual B_p is updated via

$$b_p^{(l)} = P^{\perp}_{S_p} b_{p-1}^{(l)} = b_{p-1}^{(l)} - (q_p^H b_{p-1}^{(l)}) q_p. \tag{9}$$

Equations (2), (8) and (9) define the M-MMP algorithm. The stopping rules are the same as for M-BMP.

4. DIVERSITY MINIMIZATION METHODS

4.1. Background

In this approach, all the vectors are initially selected and are eliminated (asymptotically) till only a few select columns remain [6]. To understand this methodology, it is useful to examine the solution set to (1). Any solution can be expressed as

$$X = X_{mn} + V,$$

where X_{mn} is the minimum Frobenius norm solution and is given by $x_{mn} = A^+ b$, where A^+ denotes the Moore-Penrose pseudoinverse. The lth column of X_{mn} is the minimum 2-norm solution to the system of equations $Ax^{(l)} = b^{(l)}$. The matrix V is a specially constrained matrix. Its column vectors $v^{(l)}$ are arbitrary as long as they lie in $\mathcal{N}(A)$, the null space of A. In this case A has a nontrivial nullspace of dimension $(n - m)$.

In many situations, a popular approach has been to set $V = 0$ and to select X_{mn} as the desired solution. This has two main drawbacks. The first is that the minimum 2-norm solutions that make up this solution is based on a criteria that favors solutions with many small nonzero entries, a property that is contrary to the goal of sparsity/concentration [2, 6]. The second drawback is that the solutions for each of the measurement vectors are computed independently thereby being unable to enforce any common sparsity structure across the measurement vectors as required by the problem. The first problem, 2-norm criteria, has been addressed in the recent past, but the the second problem, common sparsity enforcement has not been addressed and is dealt with in this section.

4.2. Diversity Measures for the MMV Problem

To alleviate the problems encountered by minimum 2-norm solutions, i.e. that of many non-zero entries, alternate functionals referred to here as diversity measures have been defined which when

optimized lead to sparse solutions. A popular diversity measure is $E^{(p)}(x)$ [15, 16, 17, 18, 11, 12], where

$$E^{(p)}(x) = \sum_{i=1}^{n} |x[i]|^p, \quad 0 \leq p \leq 1.$$

Due to the close connection to ℓ_p norms, these measures are referred to as "$\ell_{(p\leq 1)}$ diversity measures" and often, more simply, as the "p-norm-like diversity measures." The diversity measure for $p = 0$, the *numerosity* discussed in [16], is of special interest because it is a *direct* measure of sparsity, providing a count of the number of nonzero elements of a vector x:

$$E^{(0)}(x) = \#\{i : x[i] \neq 0\}.$$

Finding a global minimum to the numerosity measures requires an enumerative search and is NP hard [4]. Consequently, alternate diversity measures that are more amenable to optimization techniques are of interest and the $E^{(p)}(x)$ measures for $p \leq 1, p \neq 0$ are useful candidate measures in this context. For discussion of these diversity measures for $0 \leq p \leq 1$ the reader is referred to [15, 16, 17, 12].

A new diversity measures is now defined for the MMV problem which when minimized leads to a sparse X, with the solution vectors (the columns) sharing a common sparsity profile. There are many ways in which these diversity measures can be defined [12]. Among the several choices, we choose an extension that appears to meet the sparsity requirements and also results in a simple computational algorithm,

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

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. A nonzero row gets penalized as p gets small thereby promoting a common sparsity profile across the columns of X.

4.3. The M-FOCUSS Algorithm

Starting from this measure, the factored-gradient approach is used to develop a algorithm to minimize it [11, 12]. This leads to an interesting and useful class of iterative algorithms which reduces to the the class of algorithms called FOCUSS when $L = 1$ [6]. Since this algorithm represents an extension of the FOCUSS class of algorithms to the MMV case, it is referred to as M-FOCUSS. Due to space limitations, we omit the details and present only the algorithm which is as follows:

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

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.

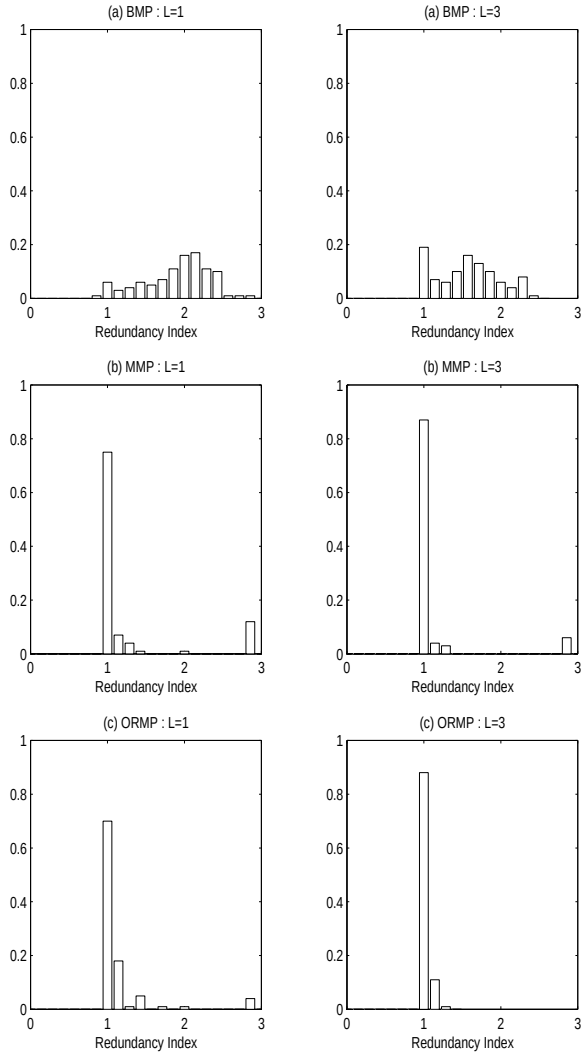

Figure 1: Comparison of M-BMP, M-MMP, and M-ORMP test data with sparsity $r = 7$ and no noise. $m = 20, n = 30$. The number of measurement vectors are $L = 1$ and $L = 3$.

5. SIMULATIONS AND COMPARISON

Computer simulations are conducted to provide a comparative evaluations of the methods. In this test case, a random $m \times n$ matrix A is created whose entries are each Gaussian random variables with mean zero and variance 1. A known sparse matrix, X_s, with L columns and a specified number of nonzero entries rows r is created; the indices of the r nonzero rows are chosen randomly and the amplitudes of the row entries are also chosen randomly. The MMV vector B is then computed as $B = AX_s$. Note that choosing a nonzero row of X is equivalent to selecting a column of A. With a known sparse solution, X_s, now at hand to provide a benchmark, the various best basis selection methods are evaluated. The number of vectors chosen by the methods are compared with the actual number, r, used to generate the data. The experiment is repeated 100 times, and a histogram of a *redundancy index*, defined as the ratio of the number of distinct columns chosen by the

method to the number of columns actually used to generate the data, is plotted. Algorithms with a redundancy index histogram concentrated around 1 indicate a good procedure.

The results of these experiments give insight into the capabilities of the various methods. The results obtained using the sequential methods are shown in Figure 1, and the results obtained using M-FOCUSS are shown in Figure 2. Comparing M-BMP and M-MMP, the M-MMP appears to do a better job of subset selection with negligible additional computational complexity. It avoids the problem of vectors being reselected that can occur with M-BMP which unnecessarily increases the number of iterations. Secondly, the stopping criteria based on the norm of the residual vector is more meaningful with the M-BMP as it does represents the true residual matrix. Of all the three forward selection methods, the M-ORMP is the most capable in terms of identifying a sparse solution. The computational complexity though higher can be greatly reduced by conducting a more careful scrutiny of the steps [19]. The M-FOCUSS algorithm with the proper choice of p in the diversity measure yields the best result on the test data. For instance. when comparing the M-FOCUSS for $p = 0$ with M-ORMP, though M-ORMP appears to have a slight edge with $L = 1$, the M-FOCUSS does better with $L = 3$. The M-FOCUSS algorithm is, however, more complex than the sequential methods. Some procedures for reducing the complexity is suggested in [9]. In addition to these results, the M-FOCUSS approach appears to provide better sparse solutions when the A matrix has structured columns [5].

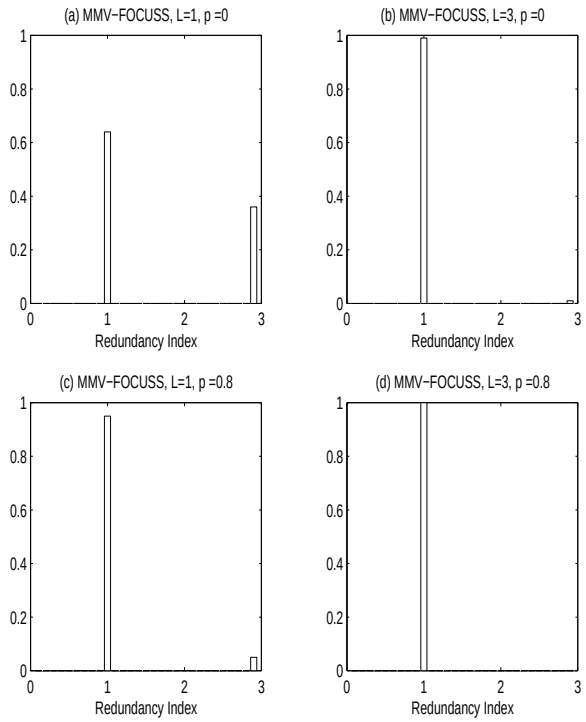

Figure 2: Comparison of M-FOCUSS for choices of $p = 0$ and $p = 0.8$ on test data with sparsity $r = 7$ and no noise. $m = 20, n = 30$. The number of measurement vectors are $L = 1$ and $L = 3$.

6. ACKNOWLEDGMENT

The authors wish to thank Mr. Shane Cotter for his assistance with the simulations.

7. REFERENCES

[1] S. G. Mallat and Z. Zhang. "Matching Pursuits with Time-Frequency Dictionaries". *IEEE Trans. ASSP*, Dec. 1993.

[2] S. Chen and D. Donoho. "Basis Pursuit". In *Twenty-Eighth Asilomar Conference, Vol. I*, pages 41–44, Nov. 1994.

[3] S. Chen and J. Wigger. *IEEE Trans. on Signal Processing*, 43(7):1713–1715, July 1995.

[4] B. K. Natarajan. *SIAM Journal on Computing*, April 1995.

[5] J. M. Adler, B. D. Rao, and K. Kreutz-Delgado. In *Proc. of the 30th Asilomar Conference*, Nov. 1996.

[6] I.F. Gorodnitsky and B.D. Rao. *IEEE Trans. on Signal Processing*, 45:600–616, March 1997.

[7] B.D. Rao. "Signal Processing with the Sparseness Constraint". In *Proc. ICASSP*, Seattle, WA, May 1998.

[8] I.F. Gorodnitsky, J.S. George, and B.D. Rao. *Jour. of Electroenceph. and Clinical Neurophysiology*, Oct. 1995.

[9] B. D. Rao. "Analysis and Extensions of the FOCUSS Algorithm". In *Proc. of the 30th Asilomar Conf.*, Nov. 1996.

[10] M. Wax and I. Ziskind. *IEEE Trans. ASSP*, ASSP-37(7):996–1000, July 1989.

[11] B. D. Rao and K. Kreutz-Delgado. "An Affine Scaling Methodology for Best Basis Selection". *In press.*

[12] K. Kreutz-Delgado and B. D. Rao. "Measures and Algorithms for Best Basis Selection". In *Proc. ICASSP 98.*

[13] S. Singhal and B. S. Atal. "Amplitude Optimization and pitch prediction in multipulse coders". *IEEE Trans. ASSP*, 1989.

[14] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. In *27th Asilomar Conference on Signal, Syst., and Comp.*, Nov. 1993.

[15] M. V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software.* A. K. Peters, Wellesley, MA, 1994.

[16] D. Donoho. In *Wavelets: Theory, Algorithms, and Applications, edited by C. K. Chui et al*, Academic Press.

[17] B. Jeffs and M. Gunsay. *IEEE Trans. on Image Processing*, 1993.

[18] G. Harikumar and Y. Bresler. In *Proc. ICASSP 96*, vol. III, pages 1331–1334, May 1996.

[19] S. F. Cotter, M. N. Murthi, and B. D. Rao. "Fast Basis Selection Methods". In *31st Asilomar Conference*, Nov. 1997.