DIVERSITY MEASURE MINIMIZATION BASED METHOD FOR COMPUTING SPARSE SOLUTIONS TO LINEAR INVERSE PROBLEMS WITH MULTIPLE MEASUREMENT VECTORS

B. D. Rao, K.Engan† and S. Cotter

Electrical and Computer Engineering Department
University of California, San Diego
La Jolla, California 92093-0407
brao@ece.ucsd.edu
Tel: (858)534-6186 Fax: (858)534-1004

ABSTRACT

We address the problem of finding sparse solutions to linear inverse problems when there are Multiple Measurement Vectors (MMV) and the solutions are assumed to have a common, but unknown, sparsity profile. This is an important extension to the single measurement sparse solution problem that has been extensively studied in the past. Of particular interest are methods based on minimizing diversity measures. A measure appropriate for the multiple measurement problem is developed, and an algorithm is derived based on its minimization. The algorithm developed, M-FOCUSS, generalizes the FOcal Underdetermined System Solver (FOCUSS) algorithm developed for the single measurement case. The convergence of the algorithm is established and a simulation study is conducted to evaluate its effectiveness. The results clearly show the ability of M-FOCUSS to utilize multiple measurement vectors to accurately identify the sparsity structure and compute sparse solutions.

1. INTRODUCTION

The problem of computing sparse solutions (i.e., solutions where only a very small number of entries are nonzero) to linear inverse problems arises in a large number of application areas [1]. The problem can be stated as: Represent a signal of interest using the minimum number of vectors from an overcomplete dictionary (set of vectors), and has been shown to be NP-hard. Consequently a number of sub-optimal low complexity algorithms have been developed. Popular approaches include greedy sequential search techniques such as Matching Pursuit [2] and methods based on minimizing diversity measures such as Basis Pursuit [3] and FOCUSS (FOCal Underdetermined System Solver) [4, 5, 6]. The later class of diversity minimization methods are of interest in this work. A discussion on extension of sequential search methods can be found in [7, 8]

In this paper, we consider an important variation of the sparse linear inverse problem: the computation of sparse solutions when there are *Multiple Measurement Vectors* (MMV) and the solutions are assumed to have a *common*, but unknown, sparsity profile. Our initial interest in solving the MMV problem was motivated by the need to solve the neuromagnetic inverse problem that arises in Magnetoencephalography (MEG), a modality for imaging the brain [4, 9]. It is assumed that the MEG signal is the result of activity at a small number of possible activation regions in the brain. When several snapshots (measurement vectors) are obtained over a small time period, the assumption is made that the variation in brain activity is such that, while the activation magnitudes change, the activation sites themselves do not. This naturally leads to the formulation of the MMV problem studied in this paper. The formulation is also useful in array processing, non-parametric spectrum analysis, and equalization of sparse channels. Despite the considerable application potential, the MMV problem has not received proper attention. As our results show, one can greatly improve upon the ability to provide sparse signal representations by utilizing MMV.

1.1. Problem Formulation

The MMV problem can be stated as solving the following L underdetermined systems of equations:

$$A\mathbf{x}^{(l)} = \mathbf{b}^{(l)} + \mathbf{n}^{(l)}, \quad l = 1, \cdots, L\,, \tag{1}$$

where $A \in \mathcal{C}^{m\times n}$, $m < n$, and often $m \ll n$. It is assumed that A has full row rank (rank$(A) = m$). L is the number of measurement vectors and it is usually assumed that $L \ll m$. The quantities $\mathbf{b}^{(l)} \in \mathcal{C}^m, l = 1, \cdots, L$ are the measurement vectors and $\mathbf{x}^{(l)} \in \mathcal{C}^n, l = 1, \cdots, L$ are the corresponding source vectors. $\mathbf{n}^{(l)} \in \mathcal{C}^m$ represent the additive noise. Since the matrix A is common to each of the L representation problems, we can succinctly rewrite (1) as

$$AX = B + N, \tag{2}$$

where $X = [\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(L)}]$, $B = [\mathbf{b}^{(1)}, \cdots, \mathbf{b}^{(L)}]$, and $N = [\mathbf{n}^{(1)}, \cdots, \mathbf{n}^{(L)}]$. In the MMV problem, we make the following distinct and important assumptions about the desired solution: 1) The solution vectors $\mathbf{x}^{(l)}$, $l = 1, \cdots, L$ are sparse, i.e., most of the entries are zero. 2) The solution vectors $\mathbf{x}^{(l)}$, $l = 1, \cdots, L$ are assumed to have the *same*, but unknown, sparsity profile, so that the indices of the nonzero entries are independent of l. However, the values in each nonzero position can be very different. The number of non-zero rows is referred to as the sparsity of the solution.

*THIS RESEARCH WAS PARTIALLY SUPPORTED BY THE NSF GRANT NO. CCR-9902961.

†Stavanger University College, Norway, Kjersti.Engan@tn.his.no

Even in the noisefree case ($N = 0$), the problem is challenging. Because of the assumption that A has full row rank, the system (2) is consistent and always has a solution. The issue is how to find a *sparse* solution from among the infinity of solutions which exist because $m < n$ (and usually $m \ll n$). Unfortunately, it has been shown for $L = 1$ that finding the solution which has the minimum number of nonzero entries is NP-hard. The common sparsity profile requirement of the MMV problem further complicates the problem. The suboptimal algorithms we develop seek a good compromise between complexity and optimality of solution. In the presence of noise, an additional complicating factor one has to consider is the tradeoff between quality of fit, i.e. $\|AX - B\|$, and the sparsity of the solution.

2. DIVERSITY MINIMIZATION METHODS

Our emphasis in this work is on methods based on minimizing diversity measures, measures whose minimization lead to sparse solution. Such approaches for $L = 1$ have been found to be promising [10, 3, 5, 6]. For simplicity, we start with the noise free case ($N = 0$) and then address noise in Section 2.3.

2.1. Diversity Measures for the MMV Problem

The underdetermined system of equations 2 has many solutions and a popular choice is the minimum 2-norm solution. The minimum 2-norm solutions of Equation 2 are non-sparse. Thus we need to consider alternate functionals referred to as diversity measures which lead to sparse solutions when minimized. A popular diversity measure for the single measurement vector ($L = 1$) is $E^{(p)}(\mathbf{x})$ [3, 5, 10] where

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

Due to the close connection to ℓ_p norms, these measures are referred to as "$\ell_{(p\le 1)}$ diversity measures" or "p-norm-like diversity measures." The diversity measure for $p = 0$ is of special interest because it is a *direct* measure of sparsity. It provides a count of the number of nonzero components in $\mathbf{x}$: $E^{(0)}(\mathbf{x}) = \#\{i : x[i] \neq 0\}$. Finding a global minimum to this measure requires an enumerative search that is NP-hard. Consequently, alternate diversity measures that are more amenable to optimization techniques are of interest. The $E^{(p)}(\mathbf{x})$ measures for $0 < p \le 1$ are useful candidate measures in this context.

All the above measures are relevant to the single measurement case ($L = 1$) and not much work is available for the MMV problem. There is a need to extend the measures to the MMV scenario, and a good starting point is to consider suitably extending diversity measures developed for $L = 1$ such as the $\ell_{(p\le 1)}$ diversity measure, Gaussian or Shannon entropy measures among others [10, 5]. A general and comprehensive study of diversity measures for the MMV problem is outside the scope of this work. Instead, we present one measure which our study has shown to hold much promise. It is an extension of the $\ell_{(p\le 1)}$ diversity measure which has often been found to produces better results than other diversity measures for $L = 1$ [5]. The modified measure is given by

$$J^{(p,q)}(X) = \sum_{i=1}^{n} (\|\mathbf{x}[i]\|_q)^p, \ 0 \le p \le 1, q \ge 1 \qquad (3)$$

where[1] $\mathbf{x}[i] = [x^{(1)}[i], x^{(2)}[i], ..., x^{(L)}[i]]$ is the ith row of X and the row norm is given by $\|\mathbf{x}[i]\|_q = \left(\sum_{l=1}^{L} |x^{(l)}[i]|^q\right)^{\frac{1}{q}}$. For simplicity we consider the case $q = 2$ in the rest of this paper and denote $J^{(p,2)}(X)$ by $J^{(p)}(X)$, i.e.,

$$J^{(p)}(X) = \sum_{i=1}^{n} (\|\mathbf{x}[i]\|_2)^p = \sum_{i=1}^{n} \left(\sum_{l=1}^{L} |x^{(l)}[i]|^2\right)^{\frac{p}{2}}. \qquad (4)$$

This choice of cost function may be motivated in two ways. Firstly, it may be seen that as p approaches zero, it provides a count of the number of nonzero *rows* in X. A nonzero row gets penalized as p is reduced which promotes a common sparsity profile across the columns of X. Secondly, pragmatic considerations such as computational complexity also favor its utilization. The minimization of the $J^{(p)}(X)$ measure (4) will be found to lead to a reasonably low complexity computational algorithm.

2.2. The M-FOCUSS Algorithm

To obtain sparse solutions with common sparsity profile, in the noise free case, we minimize $J^{(p)}(X)$ subject to the equality constraint $AX = B$. For this purpose, we employ the factored-gradient approach [5]. The details are omitted due to space limitations. The algorithm is summarized as follows:

$$\begin{aligned} W_{k+1} &= \text{diag}(c_k[i]^{1-\frac{p}{2}}), \ \ p \in [0, 2] \\ \text{where } \ c_k[i] &= \|\mathbf{x}_k[i]\| = \left(\sum_{l=1}^{L} (x_k^{(l)}[i])^2\right)^{\frac{1}{2}}, \\ Q_{k+1} &= A_{k+1}^{\dagger} B, \ \text{ where } \ A_{k+1} = AW_{k+1} \\ X_{k+1} &= W_{k+1} Q_{k+1}. \end{aligned} \qquad (5)$$

This algorithm represents an extension of the FOCUSS class of algorithms developed for $L = 1$ to the MMV case [5]. Therefore, it is referred to as M-FOCUSS. The algorithm is terminated once a convergence criterion has been satisfied, e.g.,

$$\frac{\|X_{k+1} - X_k\|_F}{\|X_k\|_F} < \delta,$$

where δ is a user-selected parameter[2]. This algorithm can be proven to reduce $J^{(p)}(X)$ in each iteration (Section 2.4). Insight into the algorithm can be obtained by viewing each iteration as computation of a weighted minimum norm solution wherein the column weighting matrix W_{k+1} is computed from the row norms of the solution obtained in the previous iteration. Intuitively, columns corresponding to rows with smaller norm are likely to be deemphasized if they are not relevant in fitting the data and vice versa.

2.3. The Regularized M-FOCUSS Algorithm

Now we generalize the algorithm to deal with additive noise by developing the Regularized M-FOCUSS algorithm, a generalization

[1] The range of p can be extended to include negative values [5], but is not deemed useful enough to be pursued here

[2] In our experiments δ was chosen as 0.01.

of Regularized FOCUSS [6]. The algorithm is as follows:

$$\begin{aligned} W_{k+1} &= \text{diag}(c_k[i]^{1-\frac{p}{2}}), \quad p \in [0,2] \\ \text{where } c_k[i] &= \left(\sum_{l=1}^{L}(x_k^{(l)}[i])^2\right)^{\frac{1}{2}}, \\ Q_{k+1} &= A_{k+1}^T(A_{k+1}A_{k+1}^T + \lambda I)^{-1}B \qquad (6) \\ \text{where } A_{k+1} &= AW_{k+1} \text{ with } \lambda \geq 0 \\ X_{k+1} &= W_{k+1}Q_{k+1}. \end{aligned}$$

Note that M-FOCUSS algorithm corresponds to setting λ to zero in Regularized M-FOCUSS. There are two useful ways to view the Regularized M-FOCUSS algorithm. One is by viewing the algorithms, Regularized M-FOCUSS and M-FOCUSS, as solving at each iteration a weighted least squares (WLS) problem with the Regularized M-FOCUSS algorithm providing a more robust solution. This can be seen by examining the difference between the two underlying WLS problems by comparing equation 6 with equation 5. The problem of equation 6 can be regarded as a Tikhonov regularization problem:

$$Q_{k+1} = \arg\min_{Q} \left(\|AW_{k+1}Q - B\|_F^2 + \lambda \|Q\|_F^2 \right),$$

where $\|.\|_F$ is the Frobenius norm. Alternately,

$$\begin{aligned} X_{k+1} &= \arg\min_{X} G_{k+1}(X), \qquad (7) \\ \text{where } G_{k+1}(X) &= \|AX - B\|_F^2 + \lambda \|W_{k+1}^{-1}X\|_F^2. \end{aligned}$$

$$\text{If } X_{k+1} \neq X_k, \text{ then } G_{k+1}(X_{k+1}) < G_{k+1}(X_k). \qquad (8)$$

A second interpretation of Regularized M-FOCUSS is as an iterative algorithm designed to minimize the regularized cost function

$$\begin{aligned} C(X) &= \|AX - B\|_F^2 + \gamma J^{(p)}(X), \qquad (9) \\ \text{with } \gamma &= \lambda \frac{2}{|p|} \geq 0. \end{aligned}$$

This can be shown by adapting a factored gradient approach to minimize this regularized cost function. We omit the derivation, and refer the reader to the derivation for $L = 1$ in [6] of which this is an extension. An interesting consequence of this interpretation is that the tradeoff between quality of fit and sparsity made by the algorithm becomes readily evident. A larger γ emphasizes sparsity over quality of fit and vice versa.

The challenge in the Regularized M-FOCUSS algorithm is finding the regularization parameter λ. This parameter has to be found for every iteration of the algorithm to ensure that the algorithm does a reasonable trade-off between finding a solution as sparse as possible and with as small error as possible. Fortunately, the *modified l-curve method* described in [6] as a method of choosing the regularization parameter also performs well in this context. The modified l-curve method is based on the l-curve method introduced in [11] as a method for finding the parameter λ, and more details can be found in [6].

2.4. Convergence of Regularized M-FOCUSS Algorithm

Fortunately, the Regularized M-FOCUSS algorithm given by (6) can be shown to reduce the regularized cost function given by (9) indicating that the algorithm will likely converge to a local minimum. This result is summarized in the following theorem.

Theorem 1. *For the Regularized M-Focuss algorithm given by (6), if $X_{k+1} \neq X_k$, then the regularized cost function $C(X)$ given by (9) decreases, i.e. $C(X_{k+1}) < C(X_k)$.*

Proof. The result is shown using the concavity of the $\ell_{(p\leq 1)}$ diversity measure. Details can be found in [8].

Experimentally, the algorithm has always converged to a sparse solution. However, unlike the $L = 1$ case, no rigorous proof of such a property appears feasible. Additional parameters in the M-FOCUSS algorithm performance are the parameter p and the initial condition. These parameters play the same role as in the $L = 1$ case. The choice of p is dictated by the speed of convergence and the sparsity of the solution generated. In practice, values of p between 0.8 and 1.1 have been found to represent a good compromise between speed of convergence and the sparsity of the generated solution. Another parameter to be chosen is the initial condition. Often in engineering applications, good initial solutions can be postulated using domain specific knowledge and should be used for initialization. If no good starting points are available, then the minimum Frobenius norm solution is a good initializer [8].

3. EXPERIMENTS

In order to evaluate the methods, the true sparse solution has to be known and this is often hard to know with real data. To circumvent this problem, we generate synthetic data wherein the true sparse solution is known and stored as a reference for comparison with the solutions obtained by the algorithms. An advantage of this approach is simplicity as well as it can facilitate exhaustive testing enabling one to draw reliable conclusions. We now describe the data generation process.

A random $m \times n$ matrix A is created whose entries are each Gaussian random variables with mean 0 and variance 1. A known sparse matrix X_0 with L columns and only r rows with nonzero entries is created. The indices of the r nonzero rows are chosen randomly from a discrete uniform distribution, and the amplitudes of the row entries are chosen randomly from a standard Gaussian distribution. The MMV matrix B is computed by first forming the product $\hat{B} = AX_0$ and adding noise where the components of the noise sequence, $\mathbf{n}^{(\ell)}$, $\ell = 1, \cdots, L$ are i.i.d and Gaussian with variance σ^2 determined from a specified SNR level as

$$\sigma^2 = \frac{1}{m}\|\hat{\mathbf{b}}\|^2 10^{-SNR/10}. \qquad (10)$$

3.1. Experimental Details

Two quantities are varied in this experiment: SNR and the number of measurement vectors L. In a Monte Carlo simulation, 500 trials are run with the dimensions set to $m = 20, n = 30$ and the sparsity to $r = 7$. In each trial, a different realization of the generating dictionary A, the solution matrix X_0, and the noise vectors are used. The performance of the M-FOCUSS, and Regularized M-FOCUSS algorithms are evaluated based on these trials. Even though M-FOCUSS was derived assuming no noise, we test it on noisy data to get an indication of its robustness and to better understand the improvements afforded by Regularized M-FOCUSS.

After terminating the algorithms three situations may occur; exactly, less than or more than r vectors (columns) from the dictionary are selected. This corresponds to the selected non-zero rows of X. In the first situation we use the r selected vectors. If

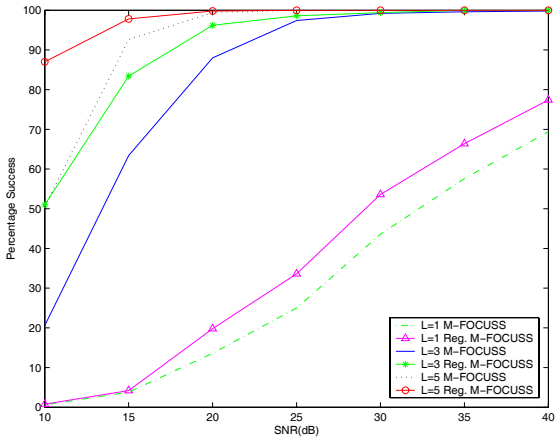

Fig. 1. Plot of Percentage of trials where *all* r=7 of the generating vectors were successfully obtained using different number of measurement vectors (L) and varying SNR for M-FOCUSS and Regularized M-FOCUSS (p=0.8, m=20, n=30, r=7)

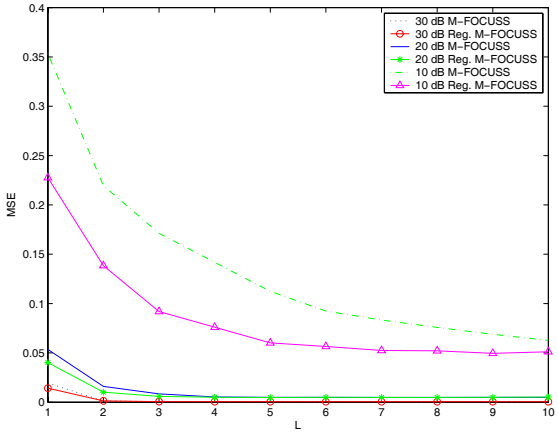

Fig. 2. Plot of MSE obtain with different SNR (10, 20, and 30 dB) as L is varied for M-FOCUSS and Regularized M-FOCUSS (p=0.8, m=20, n=30, r=7)

less than r vectors are selected we continue choosing vectors using an extended version of ORMP (M-ORMP) [8] on the residual until we have exactly r selected vectors. This is however a situation that is not likely to occur using the M-FOCUSS algorithm, but more likely to happen for the Regularized M-FOCUSS algorithm. A more frequent situation for both the M-FOCUSS and the Regularized M-FOCUSS is that more than r vectors are selected. In this case, the r rows in X which yielded the largest magnitudes row norms are chosen. Having selected the columns, the solution X is computed using a least squares approach.

3.2. Measurement of Algorithm Performance

The algorithms are run over a large number of trials and their performance is measured in the following two ways: Percentage success and relative mean squared error. Percentage success refers to the percentage of trials in which *all* of the r columns used to generate $\hat{B}$, and only these r columns, were correctly identified by the MMV algorithm. These results are plotted for different values of SNR and L in Figure 1. This gives us a performance comparison among the algorithms when used in a component detection problem, i.e., here we are trying to identify the sparsity pattern. The relative mean squared error (MSE) between the true and the estimated solution is calculated as

$$MSE = E\left(\frac{\|X - X_0\|_F^2}{\|X_0\|_F^2}\right), \tag{11}$$

where X is the solution matrix found using the FOCUSS algorithms, and X_0 is the true sparse matrix used to generate the vectors of observations. The expectation is replaced by an average over the number of trials run, and results are plotted in Figure 2.

The simulation results clearly indicate that multiple measurement vectors greatly improve the ability to identify the sparse structure and also to accurately compute the sparse solution. The regularized M-FOCUSS is shown to be quite effective in combatting the effects of noise.

4. SUMMARY

We have addressed the important problem of finding sparse solutions to linear inverse problems when there are Multiple Measurement Vectors (MMV) and the solutions are assumed to have a common, but unknown, sparsity profile. A diversity measure appropriate for the multiple measurement problem is developed, an algorithm (M-FOCUSS) is derived based on its minimization, and its convergence is established. The simulation results clearly demonstrate the usefulness of the algorithms developed.

5. REFERENCES

[1] B.D. Rao, ""Signal processing with the sparseness constraint"," in *Proc. ICASSP*, Seattle, WA, May 1998, vol. III, pp. 1861–4.

[2] S. G. Mallat and Z. Zhang, ""Matching pursuits with time-frequency dictionaries"," *IEEE Trans. on Signal Processing*, Dec. 1993.

[3] S.S. Chen, D.L. Donoho, and M.A. Saunders, ""Atomic decomposition by basis pursuit"," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[4] I.F. Gorodnitsky and B.D. Rao, ""Sparse signal reconstructions from limited data using FOCUSS: a re-weighted minimum norm algorithm"," *IEEE Trans. on Signal Processing*, March 1997.

[5] B. D. Rao and K. Kreutz-Delgado, ""An affine scaling methodology for best basis selection"," *IEEE Trans. on Sig. Proc.*, Jan. 1999.

[6] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Trans. Signal Processing*, pp. 760–770, Mar. 2003.

[7] B.D. Rao and K. Kreutz-Delgado, ""Sparse solutions to linear inverse problems with multiple measurement vectors"," in *IEEE Digital Signal Processing Workshop*, Bryce Canyon, Utah, Aug. 1998.

[8] S. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *Submitted to IEEE Trans. Signal Processing*, 2003.

[9] J. W. Phillips, R. M.Leahy, and J. C. Mosher, "Meg-based imaging of focal neuronal current sources," *IEEE Trans. Medical Imaging*, vol. 16, no. 3, pp. 338–348, 1997.

[10] M. V. Wickerhauser, *Adapted wavelet analysis from theory to software*, A. K. Peters, Wellesley, MA, 1994.

[11] P. C. Hansen, "Analysis of Discrete Ill-posed Problems by Means of the L-curve," *SIAM Review*, vol. 34, pp. 561–580, Dec. 1992.