

SPARSE SIGNAL RECOVERY IN THE PRESENCE OF CORRELATED MULTIPLE MEASUREMENT VECTORS

Zhilin Zhang and Bhaskar D. Rao

Department of Electrical and Computer Engineering,
University of California at San Diego, La Jolla, CA 92093-0407, USA
{z4zhang, brao}@ucsd.edu

ABSTRACT

Sparse signal recovery algorithms utilizing multiple measurement vectors (MMVs) are known to have better performance compared to the single measurement vector case. However, current work rarely consider the case when sources have temporal correlation, a likely situation in practice. In this work we examine methods to account for temporal correlation and its impact on performance. We model sources as AR processes, and then incorporate such information into the framework of sparse Bayesian learning for sparse signal recovery. Experiments demonstrate the superiority of the proposed algorithms. They also show that the performance of existing algorithms are limited by temporal correlation, and that if such correlation can be fully exploited, as in our proposed algorithms, the limitation can be overcome.

Index Terms— Sparse Bayesian Learning, Sparse Signal Recovery, Compressive Sensing, Multiple Measurement Vectors

1. INTRODUCTION

Recovering sparse signals from multiple measurement vectors (MMVs) is an important problem in sparse signal recovery and compressive sensing. The model can be expressed as

$$\mathbf{T} = \Phi \mathbf{W} + \mathbf{E}, \quad (1)$$

where $\Phi \in \mathcal{R}^{N \times M}$ ($N \leq M$) is the dictionary matrix, $\mathbf{T} \in \mathcal{R}^{N \times L}$ is the measurement matrix consisting of L measurement vectors, $\mathbf{W} \in \mathcal{R}^{M \times L}$ is the source matrix with each row representing a possible source, and $\mathbf{E}$ is the noise matrix with white Gaussian noise entries with zero mean and variance σ^2 . The MMV problem is often encountered in practical applications. For example, in neuroelectromagnetic source localization the event-related potentials to identify exist at least through 5 measurement vectors if the sampling frequency is 500 Hz.

It has been shown [1, 2, 3, 4, 5] that compared to the case of single measurement vector (SMV), recovery rate can be greatly improved using MMV. Early work was done by Cotter et al.[1] who showed that given the problem dimensions, i.e. M and N , one can significantly improve the chances of recovering the sparse solution utilizing multiple measurement vectors. The work was extended by Chen et al.[2], giving the sparse signal recovery conditions for the MMV problem when using ℓ_1 versus ℓ_0 minimization. Block-sparsity model [4, 6] was first proposed for the SMV problem, and

was shown by Eldar et al.[4] that the MMV model can be transformed into the block-sparsity model with relaxed recovery conditions compared to the SMV case.

In addition to the theoretical work, various algorithms were proposed by extending the SMV based algorithms using row norms and row sparsity, such as M-FOCUSS [1], M-OMP [1, 2] and Greedy Pursuit [7]. However, these algorithms ignored the dynamics of sources, which widely exists in practical signals, such as biomedical signals and video signals. Only recently this issue has received attention. In [8, 9] spatial correlation among sources was analyzed and exploited in their algorithms for better performance. Temporal correlation was also *implicitly* considered in [10, 11]. In [10] the temporal correlation was formed as a smoothness constraint and incorporated into the M-FOCUSS algorithm. In [11] the Kalman prediction is combined with a traditional sparse signal recovery algorithm, but the hybrid algorithm still uses a SMV algorithm to find the nonzero rows of $\mathbf{W}$.

Sparse Bayesian learning (SBL) based methods [3, 12], as a branch of sparse signal recovery methods, have been shown to compare favorably with traditional ℓ_1 and ℓ_p based methods [3, 9, 13]. However, existing work ignored the case where sources have temporal correlation. In this work we *explicitly* model sources as AR processes, transform the MMV model into the block-sparsity model, and then derive algorithms (called AR-SBL) utilizing the SBL framework. Experiments show that when sources have temporal correlation, AR-SBL has superior performance and performs remarkably well even in the case of extremely high correlation.

2. MODEL DESCRIPTION

As in the past work [1, 3], we make the assumption that the multiple measurement vectors have the same, but unknown, sparsity structure, i.e. common sparsity [1]. This results in several rows of $\mathbf{W}$ being zero, i.e. row sparsity. An additional consideration in this work is the correlation among the entries in a non-zero row, referred to here as a source. The sources (i.e. rows of $\mathbf{W}$) are mutually independent, but each source satisfies an AR(1) model¹ given by

$$\mathbf{W}_{i,k+1} = \beta \mathbf{W}_{i,k} + \sqrt{1 - \beta^2} \mathbf{n}_{i,k}, \quad i = 1, \dots, M; k = 1, \dots, L \quad (2)$$

where $\beta \in (-1, 1)$ is the AR coefficient², and we assume $\mathbf{n}_{i,k} \sim \mathcal{N}(0, \alpha_i)$ and $\mathbf{W}_{i,k} \sim \mathcal{N}(0, \alpha_i)$. The assumption of Gaussianity

¹Higher order models can be readily included in the framework but they increase the complexity of the estimation problem. AR(1) represents a good compromise between complexity and performance.

²Due to space limit we only present results for the case where all the sources have the same temporal correlation, but this can be generalized to the case where each source has a different temporal correlation.

and parameterized variance α_i , the hyperparameters, is motivated by and consistent with the SBL framework [12, 3]. A value of α_i of zero results in a row with zero entries promoting sparsity. Obviously, if $\beta = 0$ the MMV model becomes the one with i.i.d. sources, which is widely considered in literature [1, 2, 3, 4, 6, 7]. If $\beta = \pm 1$, the MMV model is equivalent to the SMV model in terms of recovery performance, and the benefit from multiple measurement vectors is only signal to noise ratio (SNR) enhancement.

With the AR(1) modeling assumption, the joint distribution of $\mathbf{W}_i = [\mathbf{W}_{i1}, \dots, \mathbf{W}_{iL}]$ is given by

$$p(\mathbf{W}_i; \alpha_i, \beta) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i) \quad (3)$$

where $\mathbf{R}_i = \alpha_i \mathbf{B}^{-1}$ and $\mathbf{B}$ is defined as

$$\mathbf{B} \equiv \begin{bmatrix} 1 & \beta & \dots & \beta^{L-1} \\ \beta & 1 & \dots & \beta^{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta^{L-1} & \beta^{L-2} & \dots & 1 \end{bmatrix}^{-1}. \quad (4)$$

Under these assumptions, we now develop algorithms to solve for a row sparse $\mathbf{W}$.

3. AR-SBL ALGORITHMS

By letting $\mathbf{t} = \text{vec}(\mathbf{T}^T) \in \mathcal{R}^{NL \times 1}$, $\mathbf{D} = \mathbf{\Phi} \otimes \mathbf{I}_L$, $\mathbf{w} = \text{vec}(\mathbf{W}^T) \in \mathcal{R}^{ML \times 1}$, and $\epsilon = \text{vec}(\mathbf{E}^T)$, we can transform the MMV model (1) to the block-sparsity model:

$$\mathbf{t} = \mathbf{D}\mathbf{w} + \epsilon. \quad (5)$$

We use Bayesian methods to solve the inverse problem.

Note that the likelihood of the model (5) is

$$p(\mathbf{t}|\mathbf{w}; \sigma^2) \sim \mathcal{N}(\mathbf{D}\mathbf{w}, \sigma^2 \mathbf{I}). \quad (6)$$

On the other hand, based on our assumptions in the previous section, the prior for $\mathbf{w}$ is given by

$$p(\mathbf{w}; \alpha_1, \dots, \alpha_M, \beta) = \prod_{i=1}^M p(\mathbf{W}_i; \alpha_i, \beta) = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_0^{-1}), \quad (7)$$

where $\mathbf{\Sigma}_0 \equiv \mathbf{\Gamma} \otimes \mathbf{B}$ and $\mathbf{\Gamma} \equiv \text{diag}\{\frac{1}{\alpha_1}, \dots, \frac{1}{\alpha_M}\}$. Thus the posterior density of $\mathbf{w}$ is also Gaussian, and given by

$$p(\mathbf{w}|\mathbf{t}; \sigma^2, \alpha_1, \dots, \alpha_M, \beta) \sim \mathcal{N}(\mu_w, \mathbf{\Sigma}_w), \quad (8)$$

where

$$\begin{cases} \mu_w = \frac{1}{\sigma^2} \mathbf{\Sigma}_w \mathbf{D}^T \mathbf{t} \\ \mathbf{\Sigma}_w = (\mathbf{\Sigma}_0 + \frac{1}{\sigma^2} \mathbf{D}^T \mathbf{D})^{-1}. \end{cases} \quad (9)$$

From the posterior, we directly have the MAP estimation of $\mathbf{w}$:

$$\mathbf{w}_{\text{MAP}} = \frac{1}{\sigma^2} \mathbf{\Sigma}_w \mathbf{D}^T \mathbf{t} \quad (10)$$

where σ^2 , $\alpha_1, \dots, \alpha_M$, and β are estimated using evidence maximization or type-II maximum likelihood. This involves marginalizing over the weights and then performing ML optimization. The likelihood function can be shown to be given by:

$$p(\mathbf{t}; \sigma^2, \alpha_1, \dots, \alpha_M, \beta) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_t) \quad (11)$$

with $\mathbf{\Sigma}_t \equiv \sigma^2 \mathbf{I} + \mathbf{D} \mathbf{\Sigma}_0^{-1} \mathbf{D}^T$. We now develop two methods to maximize it and obtain the parameters; one is the EM method and the other is the fixed-point method.

3.1. EM based AR-SBL

Treating $\mathbf{w}$ as hidden variables, the Q function in the EM framework is given by

$$Q = E_{\mathbf{w}|\mathbf{t}; \{\alpha, \beta, (\sigma^2)\}^{(\text{old})}} [\log p(\mathbf{t}, \mathbf{w}; \alpha, \beta, \sigma^2)]. \quad (12)$$

To estimate α , the Q function reduces to

$$Q(\alpha) = L \log(|\mathbf{\Gamma}|) - \text{Tr}[(\mathbf{\Gamma} \otimes \mathbf{B}) \hat{\mathbf{C}}_w], \quad (13)$$

where $\hat{\mathbf{C}}_w = \hat{\mathbf{\Sigma}}_w + \hat{\mu}_w \hat{\mu}_w^T$, and $\hat{\mathbf{\Sigma}}_w$ and $\hat{\mu}_w$ are evaluated using $\alpha^{(\text{old})}$, $\beta^{(\text{old})}$ and $(\sigma^2)^{(\text{old})}$. Taking the derivative of (13) w.r.t. α_i ($i = 1, \dots, M$), we obtain

$$\alpha_i^{(\text{new})} = \frac{1}{L} \text{Tr}[\mathbf{B} \hat{\mathbf{C}}_w^i], \quad (14)$$

where $\hat{\mathbf{C}}_w^i = (\hat{\mathbf{C}}_w)_{[(i-1)L+1:iL, (i-1)L+1:iL]}$.

To estimate β , the Q function (12) reduces to

$$Q(\beta) = M \log(|\mathbf{B}|) - \text{Tr}[(\mathbf{\Gamma} \otimes \mathbf{B}) \hat{\mathbf{C}}_w]. \quad (15)$$

Taking the derivative w.r.t. β , the resulting gradient can be used to develop a gradient-descent based estimation procedure.

$$\beta^{(\text{new})} = \beta + \eta \text{Tr}[(\mathbf{\Gamma} \otimes (\mathbf{B} \mathbf{F} \mathbf{B})) \hat{\mathbf{C}}_w - M \mathbf{B} \mathbf{F}], \quad (16)$$

where $\mathbf{F} = \partial(\mathbf{B}^{-1})/\partial\beta$, and η is a pre-fixed small step size, or can be determined by line search methods. In our experiments we use the classic backtracking method to determine η .

For estimating σ^2 the Q function (12) reduces to

$$\begin{aligned} Q(\sigma^2) &= -NL \log \sigma^2 - \frac{1}{\sigma^2} [\|\mathbf{t} - \mathbf{D} \hat{\mu}_w\|^2 \\ &\quad + (\sigma^2)^{(\text{old})} [ML - \text{Tr}(\hat{\mathbf{\Sigma}}_w \mathbf{\Sigma}_0)]]. \end{aligned} \quad (17)$$

After have taken its derivative, we obtain

$$(\sigma^2)^{(\text{new})} = \frac{\|\mathbf{t} - \mathbf{D} \hat{\mu}_w\|^2 + (\sigma^2)^{(\text{old})} [ML - \text{Tr}(\hat{\mathbf{\Sigma}}_w \mathbf{\Sigma}_0)]}{NL}. \quad (18)$$

However, as pointed out in [3, 13], σ^2 actually is a tradeoff parameter between signal fit and sparsity. It is not necessary to estimate it using (18), but instead one can choose a value (may be different to the true σ^2) using simple heuristic methods.

3.2. Fixed-Point Based AR-SBL

The EM algorithm is computationally complex and has slow convergence. Motivated by the work in the SMV case [12], at the expense of proven convergence, we derive a fixed-point learning rule for α_i , which typically results in faster convergence. Following the method in [12], from the log of (11) we take the derivative w.r.t. α_i and form a fixed-point equation, obtaining the following α learning rule:

$$\alpha_i^{(\text{new})} = \frac{(\mu_w^i)^T \mathbf{B}^T \mu_w^i}{L - \frac{1}{\alpha_i} \text{Tr}(\mathbf{B} \mathbf{\Sigma}_w^i)}, \quad (19)$$

where we define $\mathbf{\Sigma}_w^i = (\mathbf{\Sigma}_w)_{[(i-1)L+1:iL, (i-1)L+1:iL]}$ and $\mu_w^i = (\mu_w)_{[(i-1)L+1:iL]}$. In addition, from the log of (11) we can derive the β learning rule by the gradient method:

$$\begin{aligned} \beta^{(\text{new})} &= \beta - \eta_1 [\text{Tr}(\mathbf{D}(\mathbf{\Gamma}^{-1} \otimes \mathbf{F}) \mathbf{D}^T \mathbf{\Sigma}_t^{-1}) \\ &\quad - \mathbf{t}^T \mathbf{\Sigma}_t^{-1} \mathbf{D}(\mathbf{\Gamma}^{-1} \otimes \mathbf{F}) \mathbf{D}^T \mathbf{\Sigma}_t^{-1} \mathbf{t}]. \end{aligned} \quad (20)$$

Similarly, the learning rule for σ^2 is given by

$$(\sigma^2)^{(new)} = \sigma^2 - \eta_2 [\text{Tr}(\Sigma_t^{-1}) - \mathbf{t}^T (\Sigma_t^{-1})^2 \mathbf{t}]. \quad (21)$$

Based on our experiments (not shown in this paper), the algorithm exhibits faster convergence but has slightly lower performance than the EM based one.

4. EXPERIMENTS

Similar to [3], four experiments were carried out. First, a dictionary $\Phi \in \mathcal{R}^{N \times M}$ was created with columns uniformly drawing from the surface of a unit hypersphere. The source matrix $\mathbf{W}_{\text{gen}} \in \mathcal{R}^{M \times L}$ were randomly generated with D nonzero rows. The nonzero amplitudes of each row were generated according to (2). Then the measurement matrix was constructed by $\mathbf{T} = \Phi \mathbf{W}_{\text{gen}} + \mathbf{E}$, where noise were spatially and temporally uncorrelated Gaussian noise. The goal of each algorithm was to recover the source matrix from $\mathbf{T}$ and Φ .

For the high SNR case (SNR = 100dB), the performance measurement is $\text{MSE} = E(\|\hat{\mathbf{W}} - \mathbf{W}_{\text{gen}}\|_{\mathcal{F}}^2 / \|\mathbf{W}_{\text{gen}}\|_{\mathcal{F}}^2)$, since in this case we are interested in the performance both in finding the locations of sources and in recovering their magnitudes. Here $\hat{\mathbf{W}}$ is the estimated source matrix. For the moderate SNR case (SNR = 20dB), the metric is the ability to identify the sparse rows measurement. This is measured in terms of the failure rate obtained by aligning D largest estimated row-norms with the sparsity profile of $\mathbf{W}_{\text{gen}}$.

Algorithms compared were M-SBL algorithm [3], SOB-M-FOCUSS [10], and the proposed EM based AR-SBL algorithm.

Experiment 1 studied the case when L varied from 1 to 5. Here $N = 25$, $M = 50$, $D = 16$, and SNR = 100dB³. For AR-SBL, two cases were considered, i.e. β was given or unknown. In the latter case, we used the initial value $\beta_0 = 0.5$ to learn β . For SOB-M-FOCUSS, we used the parameter $p = 1$ (p-norm) and the k -th order smoothness approach [10] for $L = k + 1$ ($k = 1, \dots, 4$) to achieve the best results. The experiment was repeated 5000 times for $L = 1, 2, 3$ and 40000 times for $L = 4, 5$. Fig.1 present the results when the true β was 0, 0.5, 0.9, and 0.99, respectively.

Experiment 2 studied the similar case under SNR = 20dB. All the experiment settings and algorithm initialization were the same to the previous one, except to $D = 12$. Results are shown in Fig.2.

Experiment 3 studied the case when D varied from 10 to 20. Here N , M , SNR and the initial parameters of algorithms were the same as Experiment 1. $L = 3$. The experiment was repeated 100000 times for $D = 10, \dots, 12$, 50000 times for $D = 13, \dots, 15$, and 5000 times for $D = 16, \dots, 20$. Results are shown in Fig.3. In Fig.3 (d) the performance of M-SBL with $L = 1$ is also shown for comparison purposes.

Experiment 4 studied the case when the redundancy M/N varied from 1 to 4, where $N = 25$, $L = 3$, $D = 16$ and SNR = 100dB. The experiment was repeated 5000 times. Results for $\beta = 0.7$ and 0.9 are shown in Fig.4.

Based on the experiments, the following observations can be made. (1) AR-SBL has superior performance in all cases, and its performance in the case where β unknown is close to the case where β given. (2) When $\beta = 0.99$ the performance of M-SBL and SOB-M-FOCUSS is approaching to the one in the case of $L = 1$ (see Fig.1(d) and Fig.3(d)), and performance improvement is very limited with L increasing. In contrast, AR-SBL roughly remains the same performance with β increasing. With increasing L , the performance

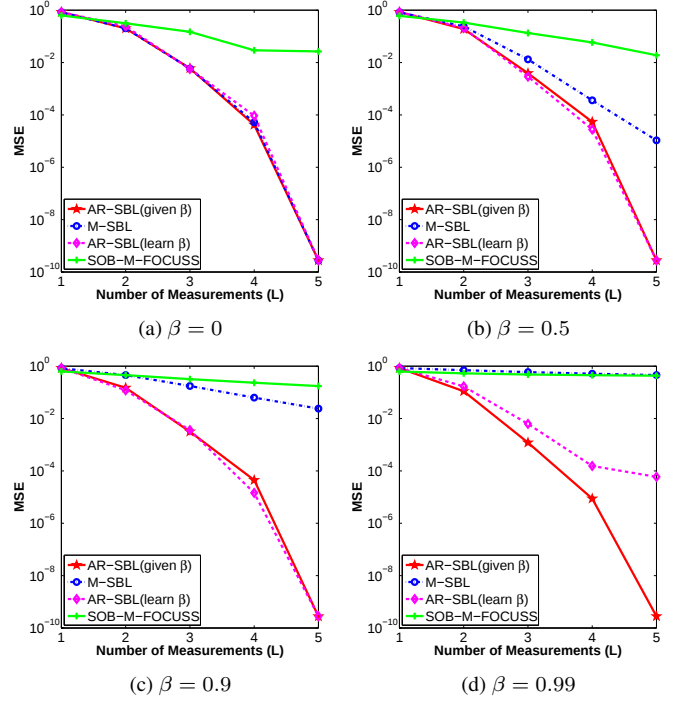


Fig. 1. Results when L varied from 1 to 5 and SNR = 100dB.

improvement is significant. (3) Exploiting temporal correlation not only can help recover signal magnitudes, but also can help correctly find the locations of active sources (see Fig.2 where the performance measurement is the failure rate to correctly find the locations, ignoring the magnitudes of sources).

The noisy counterparts of Experiment 3 and Experiment 4 show similar trends and are not shown here due to the space limit. Also, we don't present the comparison results with other popular algorithms, such as the Basis Pursuit and the Orthogonal Match Pursuit for the MMV case. However, we note that in [3] the two algorithms have been compared with M-SBL and show inferior performance to M-SBL.

To explain the success in the presence of temporal correlation, we draw upon some of the results in [5]. In [5], a parallel was drawn between the MMV sparse recovery problem and the MIMO communication problem. Using this connection the performance improvements were connected to the capacity improvements in MIMO channels which grows as $\min(D, L)$. The MIMO channel matrix was obtained from the non-zero rows of $\mathbf{W}$. Now from MIMO capacity results, it is also known that the MIMO capacity still grows as $\min(D, L)$ at high SNR in correlated channel environments. This provides some insight into why the algorithms developed perform well even in correlated environment.

5. CONCLUSION

Temporal correlation widely exists in natural signals, such as EEG/MEG signals and video signals. Motivated by the fact, we propose two AR-SBL algorithms for the MMV case incorporating temporal correlation. The superiority of the algorithms have been verified by extensive experiments.

Note that under the assumption of i.i.d. sources, current theories

³So this experiment can be viewed as an approximation of noise-free case.

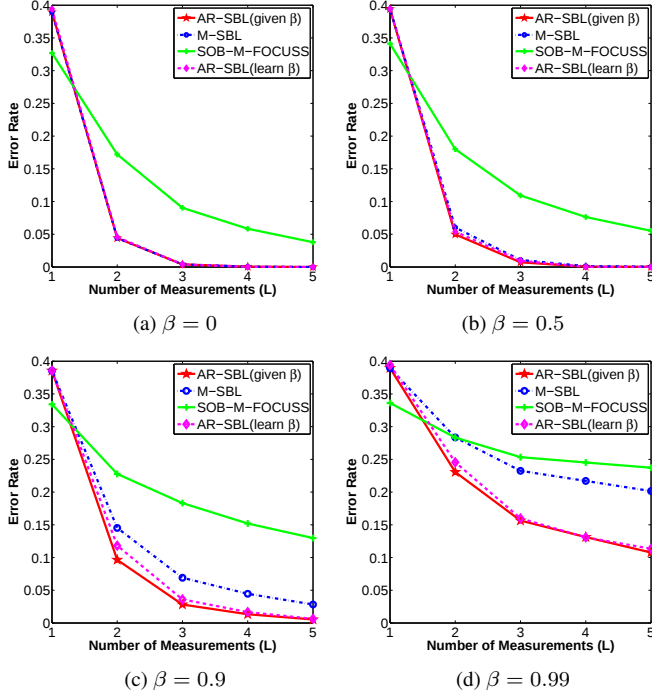


Fig. 2. Results when L varied from 1 to 5 and SNR = 20dB.

and algorithms have shown [1, 2, 4] that recovery performance improves as the number of measurement vectors increases. However, in the presence of temporal correlation, the existing MMV algorithms show less benefits, especially when the temporal correlation is high. Our work shows that if temporal correlation can be properly modeled and incorporated in algorithms, performance improvement is also significant even in highly correlated environments.

6. REFERENCES

- [1] S. F. Cotter and B.D. Rao et al., "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, 2005.
- [2] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, 2006.
- [3] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, 2007.
- [4] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.
- [5] Y. Jin and B.D. Rao, "Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory," in *ICASSP*, 2008.
- [6] M. Stojnic et al., "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, 2009.
- [7] Joel A Tropp et al., "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, pp. 572–588, 2006.
- [8] D. Baron et al, "Distributed compressive sensing," *submitted, available at <http://www.dsp.ece.rice.edu/cs>*, 2009.
- [9] D. Wipf et al., "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG," *NeuroImage*, vol. 49, 2010.
- [10] R. Zdunek and A. Cichocki, "Improved M-FOCUSS algorithm with overlapping blocks for locally smooth sparse signals," *IEEE Trans. Signal Process.*, vol. 56, no. 10, 2008.
- [11] N. Vaswani, "Kalman filtered compressed sensing," in *ICIP*, 2008.
- [12] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. of Mach. Learn. Res.*, vol. 1, 2001.
- [13] D. Wipf and S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," *NeuroImage*, vol. 44, 2009.

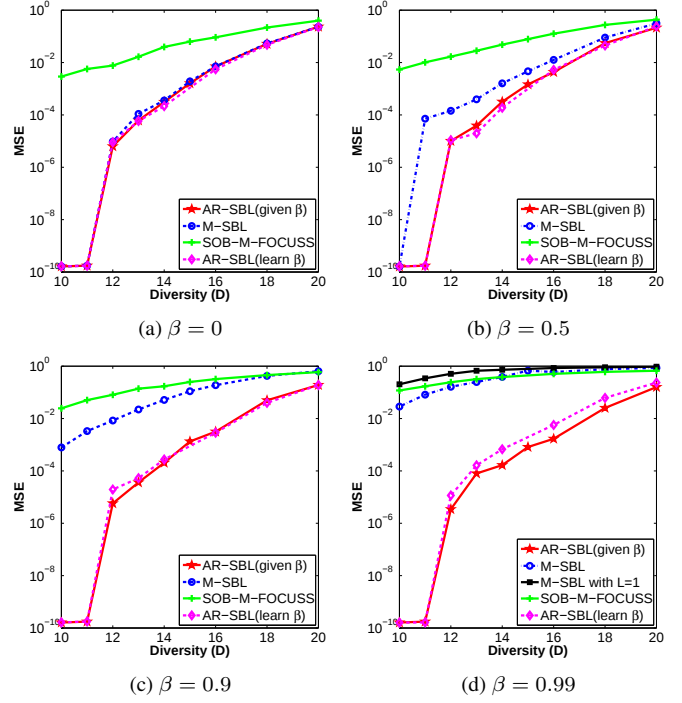


Fig. 3. Results when D varied from 10 to 20 and SNR = 100dB.

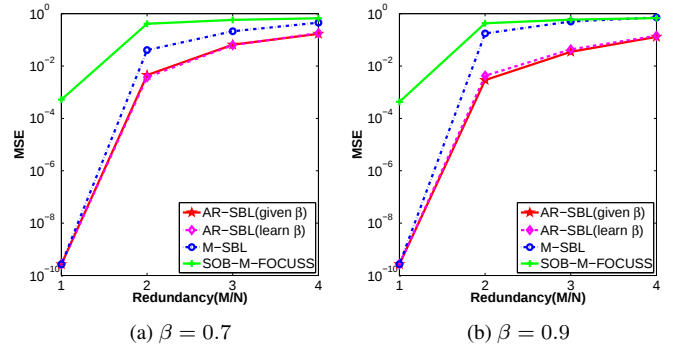


Fig. 4. Results when M/N varied from 1 to 4 and SNR = 100dB.