ITERATIVE REWEIGHTED ALGORITHMS FOR SPARSE SIGNAL RECOVERY WITH TEMPORALLY CORRELATED SOURCE 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

Iterative reweighted algorithms, as a class of algorithms for sparse signal recovery, have been found to have better performance than their non-reweighted counterparts. However, for solving the problem of multiple measurement vectors (MMVs), all the existing reweighted algorithms do not account for temporal correlations among source vectors and thus their performance degrades significantly in the presence of the correlations. In this work we propose an iterative reweighted sparse Bayesian learning (SBL) algorithm exploiting the temporal correlations, and motivated by it, we propose a strategy to improve existing reweighted ℓ_2 algorithms for the MMV problem, i.e. replacing their row norms with Mahalanobis distance measure. Simulations show that the proposed reweighted SBL algorithm has superior performance, and the proposed improvement strategy is effective for existing reweighted ℓ_2 algorithms.

Index Terms— Sparse Signal Recovery, Compressed Sensing, Iterative Reweighted ℓ_2 Algorithms, Sparse Bayesian Learning, Temporal Correlation

1. INTRODUCTION

The multiple measurement vector (MMV) model for sparse signal recovery is given by [1]

$$\mathbf{Y} = \mathbf{\Phi X} + \mathbf{V}, \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{N \times M}(N \ll M)$ is the dictionary matrix whose any N columns are linearly independent, $\mathbf{Y} \in \mathbb{R}^{N \times L}$ is the measurement matrix consisting of L measurement vectors, $\mathbf{X} \in \mathbb{R}^{M \times L}$ is the source matrix with each row representing a possible source, and $\mathbf{V}$ is the white Gaussian noise matrix with each entry satisfying $\mathbf{V}_{ij} \sim \mathcal{N}(0, \lambda)$. The key assumption under the MMV model is that the support (i.e. locations of nonzero entries) of every column vector $\mathbf{X}_{\cdot i}$ $(\forall i)$ [1] is identical (referred as *the common sparsity assumption* in the literature [1]). The MMV problem is often encountered in practical applications, such as neuroelectromagnetic source localization and direction-of-arrival estimation.

Among many algorithms for the MMV problem, iterative reweighted methods have received attention because of their improved performance compared to their non-reweighted counterparts [2, 3]. In [3], an iterative reweighted ℓ_1 minimization framework is employed. The framework can be directly used for the MMV problem and many MMV algorithms based on mixed norm optimization

can be improved via the framework. On the other hand, iterative reweighted ℓ_2 algorithms were also proposed [2, 4]. The reweighted ℓ_2 minimization framework for the MMV problem (in noisy case) computes the solution at the $(k + 1)$-th iteration as follows [2]:

$$
\begin{aligned}
\mathbf{X}^{(k+1)} &= \arg\min_{\mathbf{x}} \|\mathbf{Y} - \mathbf{\Phi X}\|_{\mathcal{F}}^2 + \lambda \sum_i w_i^{(k)} (\|\mathbf{X}_{i \cdot}\|_q)^2 \quad (2) \\
&= \mathbf{W}^{(k)} \mathbf{\Phi}^T (\lambda\mathbf{I} + \mathbf{\Phi W}^{(k)} \mathbf{\Phi}^T)^{-1} \mathbf{Y} \tag{3}
\end{aligned}
$$

where typically $q = 2$, $\mathbf{W}^{(k)}$ is a diagonal weighting matrix at the k-th iteration with i-th diagonal element being $1/w_i^{(k)}$, and $w_i^{(k)}$ depends on the previous estimate of $\mathbf{X}$. Recently, Wipf et al [2] unified most existing iterative reweighted algorithms as belonging to the family of *separable* reweighted algorithms, whose weighting w_i of a given row $\mathbf{X}_{i \cdot}$ at each iteration is only a function of that individual row from the previous iteration. Further, they proposed *non-separable* reweighted algorithms via variational approaches, which outperform many existing separable reweighted algorithms.

In our previous work [5, 6] we showed that temporal correlations [3] in sources $\mathbf{X}_{i \cdot}$ seriously deteriorates recovery performance of existing algorithms and proposed a block sparse Bayesian learning (bSBL) framework, in which we incorporated temporal correlations and derived effective algorithms. These algorithms operate in the hyperparameter space, not in the source space as most sparse signal recovery algorithms do. Therefore, it is not clear what the connection of the bSBL framework is to other sparse signal recovery frameworks, such as the reweighted ℓ_2 in (2). In this work, based on the cost function in the bSBL framework, we derive an iterative reweighted ℓ_2 SBL algorithm with superior performance, which directly operates in the source space. Furthermore, motivated by the intuition gained from the algorithm and analytical insights, we propose a strategy to modify existing reweighted ℓ_2 algorithms to incorporate the temporal correlations, and use two typical algorithms as illustrations. The strategy is shown to be effective.

2. THE BSBL FRAMEWORK

The block sparse Bayesian learning (bSBL) framework [5, 6] transforms the MMV problem to a single measurement vector problem. This makes the modeling of temporal correlations much easier. First, we assume the rows of $\mathbf{X}$ are mutually independent, and the density of each row $\mathbf{X}_{i \cdot}$ is multivariate Gaussian, given by

$$p(\mathbf{X}_{i \cdot}; \gamma_i, \mathbf{B}_i) \sim \mathcal{N}(\mathbf{0}, \gamma_i \mathbf{B}_i), \quad i = 1, \cdots, M$$

---

[1] The i-th column of $\mathbf{X}$ is denoted by $\mathbf{X}_{\cdot i}$. The i-th row of $\mathbf{X}$ is denoted by $\mathbf{X}_{i \cdot}$ (also called the i-th source).

[2] For convenience, we omit the superscript, k, on the right hand side of learning rules in the following.

[3] By temporal correlations, we mean the correlations of elements in each nonzero row in $\mathbf{X}$.

where γ_i is a nonnegative hyperparameter controlling the row sparsity of $\mathbf{X}$ as in the basic sparse Bayesian learning [7, 8]. When $\gamma_i = 0$, the associated i-th row of $\mathbf{X}$ becomes zero. $\mathbf{B}_i$ is an unknown positive definite matrix.

By letting $\mathbf{y} = \mathrm{vec}(\mathbf{Y}^T) \in \mathbb{R}^{NL \times 1}$, $\mathbf{D} = \mathbf{\Phi} \otimes \mathbf{I}_L$ [4], $\mathbf{x} = \mathrm{vec}(\mathbf{X}^T) \in \mathbb{R}^{ML \times 1}$ and $\mathbf{v} = \mathrm{vec}(\mathbf{V}^T)$, where $\mathrm{vec}(\mathbf{A})$ denotes the vectorization of the matrix $\mathbf{A}$ formed by stacking its columns into a single column vector, we can transform the MMV model (1) to the block single vector model as follows

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{v}. \tag{4}$$

To elaborate on the block sparsity model (4), we rewrite it as $\mathbf{y} = [\mathbf{\Phi}_1 \otimes \mathbf{I}_L, \cdots, \mathbf{\Phi}_M \otimes \mathbf{I}_L][\mathbf{x}_1^T, \cdots, \mathbf{x}_M^T]^T + \mathbf{v} = \sum_{i=1}^{M}(\mathbf{\Phi}_i \otimes \mathbf{I}_L)\mathbf{x}_i + \mathbf{v}$, where $\mathbf{\Phi}_i$ is the i-th column of $\mathbf{\Phi}$, $\mathbf{x}_i \in \mathbb{R}^{L \times 1}$ is the i-th block in $\mathbf{x}$ and it is the transposed i-th row of $\mathbf{X}$ in the original MMV model (1), i.e. $\mathbf{x}_i = \mathbf{X}_i^T$. K nonzero rows in $\mathbf{X}$ means K nonzero blocks in $\mathbf{x}$. Thus we refer to $\mathbf{x}$ as block-sparse.

For the block model (4), the Gaussian likelihood is $p(\mathbf{y}|\mathbf{x}; \lambda) \sim \mathcal{N}_{y|x}(\mathbf{D}\mathbf{x}, \lambda\mathbf{I})$. The prior for $\mathbf{x}$ is given by $p(\mathbf{x}; \gamma_i, \mathbf{B}_i, \forall i) \sim \mathcal{N}_x(\mathbf{0}, \mathbf{\Sigma}_0)$, where $\mathbf{\Sigma}_0$ is a block diagonal matrix with the i-th diagonal block $\gamma_i \mathbf{B}_i$ ($\forall i$). Given the hyperparameters $\Theta \triangleq \{\lambda, \gamma_i, \mathbf{B}_i, \forall i\}$, the Maximum-A-Posterior (MAP) estimate of $\mathbf{x}$ can be directly obtained from the posterior of the model. To estimate these hyperparameters, we can use the Type-II maximum likelihood method [8], which marginalizes over $\mathbf{x}$ and then performs maximum likelihood estimation, leading to the cost function:

$$
\begin{aligned}
\mathcal{L}(\Theta) &\triangleq -2\log \int p(\mathbf{y}|\mathbf{x}; \lambda)p(\mathbf{x}; \gamma_i, \mathbf{B}_i, \forall i)d\mathbf{x} \\
&= \log|\lambda\mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T| + \mathbf{y}^T(\lambda\mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T)^{-1}\mathbf{y}, (5)
\end{aligned}
$$

where $\boldsymbol{\gamma} \triangleq [\gamma_1, \cdots, \gamma_M]^T$. We refer to the whole framework including the solution estimation of $\mathbf{x}$ and the hyperparameter estimation as the bSBL framework. Note that in contrast to the original SBL framework, the bSBL framework models the temporal correlation structures of sources in the prior density via the matrix $\mathbf{B}_i$ ($\forall i$).

3. ITERATIVE REWEIGHTED SPARSE BAYESIAN LEARNING ALGORITHM

Based on the cost function (5), we can derive efficient algorithms that exploit temporal correlations of sources [5, 6]. But these algorithms directly operate in the hyperparameter space (i.e. the $\boldsymbol{\gamma}$-space). So, it is not clear what their connection is to other sparse signal recovery algorithms which directly operate in the source space (i.e. the $\mathbf{X}$-space) by minimizing penalties on the sparsity of $\mathbf{X}$. Particularly, it is interesting to see if we can transplant the benefits gained from the bSBL framework to other sparse signal recovery frameworks such as the iterative reweighted ℓ_2 minimization framework (2), and improve algorithms belonging to those frameworks. Following the approach developed by Wipf et al [2] for the single measurement vector problem, in the following we use the duality theory [9] to obtain a penalty in the source space, based on which we derive an iterative reweighted algorithm for the MMV problem.

3.1. Algorithm Development

First, we find that assigning a different covariance matrix $\mathbf{B}_i$ to each source $\mathbf{X}_i$. will result in overfitting in the learning of the hyperpa-

---

rameters. To overcome the overfitting, we simplify and consider using one matrix $\mathbf{B}$ to model all the source covariance matrixes. Thus $\mathbf{\Sigma}_0 = \mathbf{\Gamma} \otimes \mathbf{B}$ with $\mathbf{\Gamma} \triangleq \mathrm{diag}(\boldsymbol{\gamma})$. Simulations will show that this simplification leads to good results even if different sources have different temporal correlations (see Section 5).

In order to transform the cost function (5) to the source space, we use the identity: $\mathbf{y}^T(\lambda\mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T)^{-1}\mathbf{y} \equiv \min_{\mathbf{x}}\left[\frac{1}{\lambda}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \mathbf{x}^T\mathbf{\Sigma}_0^{-1}\mathbf{x}\right]$, by which we can upper-bound the cost function (5) and obtain the bound

$$\mathfrak{L}(\mathbf{x}, \boldsymbol{\gamma}, \mathbf{B}) = \log|\lambda\mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T| + \frac{1}{\lambda}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \mathbf{x}^T\mathbf{\Sigma}_0^{-1}\mathbf{x}.$$

By first minimizing over $\boldsymbol{\gamma}$ and $\mathbf{B}$ and then minimizing over $\mathbf{x}$, we can get the cost function in the source space:

$$\mathbf{x} = \arg\min_{\mathbf{x}}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda g_{\mathrm{TC}}(\mathbf{x}), \tag{6}$$

where the penalty $g_{\mathrm{TC}}(\mathbf{x})$ is defined by

$$g_{\mathrm{TC}}(\mathbf{x}) \triangleq \min_{\boldsymbol{\gamma} \succeq \mathbf{0}, \mathbf{B} \succ \mathbf{0}} \mathbf{x}^T\mathbf{\Sigma}_0^{-1}\mathbf{x} + \log|\lambda\mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T|. \tag{7}$$

From the definition (7) we have

$$
\begin{aligned}
g_{\mathrm{TC}}(\mathbf{x}) &\leq \mathbf{x}^T\mathbf{\Sigma}_0^{-1}\mathbf{x} + \log|\lambda\mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T| \\
&= \mathbf{x}^T\mathbf{\Sigma}_0^{-1}\mathbf{x} + \log|\mathbf{\Sigma}_0| + \log\left|\frac{1}{\lambda}\mathbf{D}^T\mathbf{D} + \mathbf{\Sigma}_0^{-1}\right| + NL\log\lambda \\
&\leq \mathbf{x}^T\mathbf{\Sigma}_0^{-1}\mathbf{x} + \log|\mathbf{\Sigma}_0| + \mathbf{z}^T\boldsymbol{\gamma}^{-1} - f^*(\mathbf{z}) + NL\log\lambda
\end{aligned}
$$

where in the last inequality we have used the conjugate relation

$$\log\left|\frac{1}{\lambda}\mathbf{D}^T\mathbf{D} + \mathbf{\Sigma}_0^{-1}\right| = \min_{\mathbf{z} \succeq 0} \mathbf{z}^T\boldsymbol{\gamma}^{-1} - f^*(\mathbf{z}). \tag{8}$$

Here we denote $\boldsymbol{\gamma}^{-1} \triangleq [\gamma_1^{-1}, \cdots, \gamma_M^{-1}]^T$, $\mathbf{z} \triangleq [z_1, \cdots, z_M]^T$, and $f^*(\mathbf{z})$ is concave conjugate of $f(\boldsymbol{\gamma}^{-1}) \triangleq \log|\frac{1}{\lambda}\mathbf{D}^T\mathbf{D} + \mathbf{\Sigma}_0^{-1}|$. Finally, reminding of $\mathbf{\Sigma}_0 = \mathbf{\Gamma} \otimes \mathbf{B}$, we have

$$
\begin{aligned}
g_{\mathrm{TC}}(\mathbf{x}) \leq \quad & NL\log\lambda - f^*(\mathbf{z}) + M\log|\mathbf{B}| + \\
& \sum_{i=1}^{M}\left[\frac{\mathbf{x}_i^T\mathbf{B}^{-1}\mathbf{x}_i + z_i}{\gamma_i} + L\log\gamma_i\right]. \tag{9}
\end{aligned}
$$

Therefore, to solve the problem (6) with (9), we can perform the coordinate descent method over $\mathbf{x}, \mathbf{B}, \mathbf{z}$ and $\boldsymbol{\gamma}$, i.e,

$$
\begin{aligned}
\min_{\mathbf{x}, \mathbf{B}, \mathbf{z} \succeq 0, \boldsymbol{\gamma} \succeq 0} \quad & \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda\Big[\sum_{i=1}^{M}\Big(\frac{\mathbf{x}_i^T\mathbf{B}^{-1}\mathbf{x}_i + z_i}{\gamma_i} \\
& + L\log\gamma_i\Big) + M\log|\mathbf{B}| - f^*(\mathbf{z})\Big]. \tag{10}
\end{aligned}
$$

Compared to the framework (2), $1/\gamma_i$ can be seen as the weighting for the corresponding $\mathbf{x}_i^T\mathbf{B}^{-1}\mathbf{x}_i$. But instead of applying ℓ_q norm on $\mathbf{x}_i$ (i.e. the i-th row of $\mathbf{X}$) as done in existing iterative reweighted ℓ_2 algorithms, our algorithm computes $\mathbf{x}_i^T\mathbf{B}^{-1}\mathbf{x}_i$, i.e. the quadratic *Mahalanobis distance* of $\mathbf{x}_i$ and its mean.

By minimizing (10) over $\mathbf{x}$, the updating rule for $\mathbf{x}$ is given by

$$\mathbf{x}^{(k+1)} = \mathbf{\Sigma}_0\mathbf{D}^T(\lambda\mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T)^{-1}\mathbf{y}. \tag{11}$$

According to the dual property [9], from the relation (8), the optimal $\mathbf{z}$ is directly given by

$$
\begin{aligned}
z_i &= \frac{\partial\log|\frac{1}{\lambda}\mathbf{D}^T\mathbf{D} + \mathbf{\Sigma}_0^{-1}|}{\partial(\gamma_i^{-1})} \\
&= L\gamma_i - \gamma_i^2\mathrm{Tr}\Big[\mathbf{B}\mathbf{D}_i^T\big(\lambda\mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T\big)^{-1}\mathbf{D}_i\Big], \forall i \tag{12}
\end{aligned}
$$

---

[4]We denote the $L \times L$ identity matrix by $\mathbf{I}_L$. When the dimension is evident from the context, for simplicity we use $\mathbf{I}$. $\otimes$ is the Kronecker product.

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, and $\mathbf{D}_i$ consists of columns of $\mathbf{D}$ from the $((i-1)L+1)$-th to the (iL)-th. From (10) the optimal γ_i for fixed $\mathbf{x}, \mathbf{z}, \mathbf{B}$ is given by $\gamma_i = \frac{1}{L}[\mathbf{x}_i^T \mathbf{B}^{-1}\mathbf{x}_i + z_i]$. Substituting Eq.(12) into it, we have

$$
\begin{aligned}
\gamma_i^{(k+1)} &= \frac{\mathbf{x}_i^T \mathbf{B}^{-1}\mathbf{x}_i}{L} + \gamma_i \\
&\quad - \frac{\gamma_i^2}{L}\mathrm{Tr}\Big[\mathbf{B}\mathbf{D}_i^T\big(\lambda\mathbf{I} + \mathbf{D}\boldsymbol{\Sigma}_0\mathbf{D}^T\big)^{-1}\mathbf{D}_i\Big], \ \forall i \ (13)
\end{aligned}
$$

By minimizing (10) over $\mathbf{B}$, the updating rule for $\mathbf{B}$ is given by

$$
\mathbf{B}^{(k+1)} = \overline{\mathbf{B}}/\|\overline{\mathbf{B}}\|_{\mathcal{F}}, \quad \text{with} \quad \overline{\mathbf{B}} = \sum_{i=1}^{M}\frac{\mathbf{x}_i\mathbf{x}_i^T}{\gamma_i}. \quad (14)
$$

The updating rules (11) (13) and (14) are our reweighted algorithm minimizing the penalty based on quadratic Mahalanobis distance of $\mathbf{x}_i$. Since for a given i, the weighting $1/\gamma_i$ depends on the whole estimated source matrix in the previous iteration (via $\mathbf{B}$ and $\boldsymbol{\Sigma}_0$), the algorithm is a *nonseparable* reweighted algorithm.

The complexity of this algorithm is high because it learns the parameters in a higher dimensional space than the original problem space. For example, the dictionary matrix in the bSBL framework is of size $NL \times ML$, while in the original MMV model is of size $N \times M$. We now simplify the algorithm and develop an efficient variant. Using the approximation:

$$
\big(\lambda\mathbf{I}_{NL} + \mathbf{D}\boldsymbol{\Sigma}_0\mathbf{D}^T\big)^{-1} \approx \big(\lambda\mathbf{I}_N + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T\big)^{-1} \otimes \mathbf{B}^{-1}, \quad (15)
$$

which takes the equal sign when $\lambda = 0$ or $\mathbf{B} = \mathbf{I}$, the updating rule (11) can be transformed to

$$
\mathbf{X}^{(k+1)} = \mathbf{W}\boldsymbol{\Phi}^T\big(\lambda\mathbf{I} + \boldsymbol{\Phi}\mathbf{W}\boldsymbol{\Phi}^T\big)^{-1}\mathbf{Y}, \quad (16)
$$

where $\mathbf{W} \triangleq \mathrm{diag}([1/w_1, \cdots, 1/w_M])$ with $w_i \triangleq 1/\gamma_i$. Using the same approximation, the last term in (13) becomes

$$
\begin{aligned}
&\mathrm{Tr}\Big[\mathbf{B}\mathbf{D}_i^T\big(\lambda\mathbf{I}_{NL} + \mathbf{D}\boldsymbol{\Sigma}_0\mathbf{D}^T\big)^{-1}\mathbf{D}_i\Big] \\
&\approx \mathrm{Tr}\Big[\mathbf{B}(\boldsymbol{\Phi}_i^T \otimes \mathbf{I})\big[(\lambda\mathbf{I}_N + \boldsymbol{\Phi}\mathbf{W}\boldsymbol{\Phi}^T)^{-1} \otimes \mathbf{B}^{-1}\big](\boldsymbol{\Phi}_i \otimes \mathbf{I})\Big] \\
&= L\boldsymbol{\Phi}_i^T(\lambda\mathbf{I}_N + \boldsymbol{\Phi}\mathbf{W}\boldsymbol{\Phi}^T)^{-1}\boldsymbol{\Phi}_i.
\end{aligned}
$$

Therefore, from the updating rule of γ_i in (13) we have

$$
w_i^{(k+1)} = \Big[\frac{1}{L}\mathbf{X}_{i\cdot}\mathbf{B}^{-1}\mathbf{X}_{i\cdot}^T + \{(\mathbf{W}^{-1} + \frac{1}{\lambda}\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\}_{ii}\Big]^{-1}. \quad (17)
$$

Accordingly, the updating rule for $\mathbf{B}$ becomes

$$
\mathbf{B}^{(k+1)} = \overline{\mathbf{B}}/\|\overline{\mathbf{B}}\|_{\mathcal{F}}, \quad \text{with} \quad \overline{\mathbf{B}} = \sum_{i=1}^{M} w_i \mathbf{X}_{i\cdot}^T\mathbf{X}_{i\cdot}. \quad (18)
$$

We denote the updating rules (16) (17) and (18) by ***ReSBL-QM***.

3.2. Estimate of the Regularization Parameter λ

To estimate the regularization parameter λ, many methods have been proposed, such as the modified L-curve method [1]. Here, straightforwardly following the Expectation-Maximization method in [5] and using the approximation (15), we derive a learning rule for λ, given by:

$$
\lambda^{(k+1)} = \frac{1}{NL}\|\mathbf{Y} - \boldsymbol{\Phi}\mathbf{X}\|_{\mathcal{F}}^2 + \frac{\lambda}{N}\mathrm{Tr}\big[\mathbf{G}(\lambda\mathbf{I} + \mathbf{G})^{-1}\big].
$$

where $\mathbf{G} \triangleq \boldsymbol{\Phi}\mathbf{W}\boldsymbol{\Phi}^T$.

3.3. Theoretical Analysis in the Noiseless Case

For the noiseless inverse problem $\mathbf{Y} = \boldsymbol{\Phi}\mathbf{X}$, denote the generating sources by $\mathbf{X}_{\mathrm{gen}}$, which is the sparsest solution among all the possible solutions. Assume $\mathbf{X}_{\mathrm{gen}}$ is full column-rank. Denote the true nonzero source number (i.e. the number of nonzero rows in $\mathbf{X}_{\mathrm{gen}}$) by K_0. Now we have the following result on the global minimum of the cost function (5):

Theorem 1 *In the noiseless case, assuming $K_0 < (N + L)/2$, for the cost function (5) the unique global minimum $\widehat{\boldsymbol{\gamma}} = [\widehat{\gamma}_1, \cdots, \widehat{\gamma}_M]$ produces a source estimate $\widehat{\mathbf{X}}$ that equals to $\mathbf{X}_{\mathrm{gen}}$ irrespective of the estimated $\widehat{\mathbf{B}}_i, \forall i$, where $\widehat{\mathbf{X}}$ is obtained from $\mathrm{vec}(\widehat{\mathbf{X}}^T) = \widehat{\mathbf{x}}$ and $\widehat{\mathbf{x}}$ is computed using Eq.(11).*

The proof is given in [6]. The theorem implies that even if the estimated $\widehat{\mathbf{B}}_i$ is different from the true $\mathbf{B}_i$, the estimated sources are the true sources at the global minimum of the cost function. As a reminder, in deriving our algorithm, we assumed $\mathbf{B}_i = \mathbf{B} \ (\forall i)$ to avoid overfitting. The theorem ensures that this strategy does not harm the global minimum property; in other words, once at the global convergence, our algorithm returns the true source matrix.

In our work [6] we have shown that $\mathbf{B}$ plays a role of whitening sources in the SBL procedure, which can be seen in our algorithm as well. This gives us a motivation to improve some state-of-the-art reweighted ℓ_2 algorithms by whitening the estimated sources in their weighting rules and penalties, detailed in the next section.

4. MODIFIED EXISTING REWEIGHTED ℓ_2 METHODS

Motivated by the above results and our analysis in [6], we can modify many reweighted ℓ_2 algorithms via replacing the ℓ_2 norm of $\mathbf{X}_{i\cdot}$ by some suitable function of its Mahalanobis distance. Note that similar modifications can be applied on reweighted ℓ_1 algorithms.

The regularized M-FOCUSS [1] is a typical reweighted ℓ_2 algorithm, which solves a reweighted ℓ_2 minimization with weights $w_i^{(k)} = (\|\mathbf{X}_{i\cdot}^{(k)}\|_2^2)^{p/2-1}$ in each iteration. It is given by

$$
\begin{aligned}
\mathbf{X}^{(k+1)} &= \mathbf{W}^{(k)}\boldsymbol{\Phi}^T\big(\lambda\mathbf{I} + \boldsymbol{\Phi}\mathbf{W}^{(k)}\boldsymbol{\Phi}^T\big)^{-1}\mathbf{Y} \quad (19) \\
\mathbf{W}^{(k)} &= \mathrm{diag}\{[1/w_1^{(k)}, \cdots, 1/w_M^{(k)}]\} \\
w_i^{(k)} &= \big(\|\mathbf{X}_{i\cdot}^{(k)}\|_2^2\big)^{p/2-1}, \ p \in [0,2], \forall i \quad (20)
\end{aligned}
$$

We can modify the algorithm by changing (20) to the following one:

$$
w_i^{(k)} = \big(\mathbf{X}_{i\cdot}^{(k)}(\mathbf{B}^{(k)})^{-1}(\mathbf{X}_{i\cdot}^{(k)})^T\big)^{p/2-1}, \ p \in [0,2], \forall i \ (21)
$$

The matrix $\mathbf{B}$ can be calculated using the learning rule (18). We denote the modified algorithm by ***tMFOCUSS***.

In [4] Chartrand and Yin proposed an iterative reweighted ℓ_2 algorithm based on the classic FOCUSS algorithm. Its MMV extension (denoted by ***Iter-L2***) changed (20) to:

$$
w_i^{(k)} = \big(\|\mathbf{X}_{i\cdot}^{(k)}\|_2^2 + \epsilon\big)^{p/2-1}, \ p \in [0,2], \forall i \quad (22)
$$

Their algorithm adopts the strategy: initially use a relatively large ϵ, and then repeat the process of decreasing ϵ after convergence and repeating the iteration (19). This dramatically improves the recovery ability. Similarly, we can modify the weighting (22) to the following rule incorporating the temporal correlations of sources:

$$
w_i^{(k)} = \big(\mathbf{X}_{i\cdot}^{(k)}(\mathbf{B}^{(k)})^{-1}(\mathbf{X}_{i\cdot}^{(k)})^T + \epsilon\big)^{p/2-1}, \ p \in [0,2], \forall i (23)
$$

and adopts the same ϵ-decreasing strategy. $\mathbf{B}$ is also given by (18). We denote the modified algorithm by ***tIter-L2***.

The proposed tMFOCUSS and tIter-L2 have convergence properties similar to M-FOCUSS and Iter-L2, respectively. Due to space limit we omit theoretical analysis, and instead, provide some representative simulation results in the next section.

5. EXPERIMENTS

In our experiments, a dictionary matrix $\mathbf{\Phi} \in \mathbb{R}^{N \times M}$ was created with columns uniformly drawing from the surface of a unit hypersphere. The source matrix $\mathbf{X}_{\text{gen}} \in \mathbb{R}^{M \times L}$ was randomly generated with K nonzero rows of unit norms, whose row locations were randomly chosen. Amplitudes of the i-th nonzero row were generated as an AR(1) process whose AR coefficient was denoted by β_i [5]. Thus β_i indicates the temporal correlation of the i-th source. The measurement matrix was constructed by $\mathbf{Y} = \mathbf{\Phi}\mathbf{X}_{\text{gen}} + \mathbf{V}$, where $\mathbf{V}$ was a zero-mean homoscedastic Gaussian noise matrix with variance adjusted to have a desired value of SNR. For each different experiment setting, we repeated 500 trials and averaged results. The performance measurement was the *Failure Rate* defined in [7], which indicated the percentage of failed trials in the 500 trials. When noise was present, since we could not expect any algorithm to recover $\mathbf{X}_{\text{gen}}$ exactly, we classified a trial as a failure trial if the K largest estimated row-norms did not align with the support of $\mathbf{X}_{\text{gen}}$.

The compared algorithms included our proposed ReSBL-QM, tMFOCUSS, tIter-L2, the reweighted ℓ_2 SBL in [2] (denoted by ReSBL-L2), M-FOCUSS [1], Iter-L2 presented in Section 4, and Candes' reweighted ℓ_1 algorithm [3] (extended to the MMV case as suggested by [2], denoted by Iter-L1). For tMFOCUSS, M-FOCUSS, and Iter-L2, we set $p = 0.8$, which gave the best performance in our simulations. For Iter-L1, we used 5 iterations.

In the first experiment we fixed $N = 25$, $M = 100$, $L = 3$ and SNR $= 25$dB. The number of nonzero sources K varied from 10 to 16. Fig.1 (a) shows the results when each β_i was uniformly chosen from $[0, 0.5)$ at random. Fig.1 (b) shows the results when each β_i was uniformly chosen from $[0.5, 1)$ at random.

In the second experiment we fixed $N = 25$, $L = 4$, $K = 12$, and SNR $= 25$dB, while M/N varied from 1 to 25. β_i $(\forall i)$ in Fig.2 (a) and (b) were generated as in Fig.1 (a) and (b), respectively. This experiment aims to see algorithms' performance in highly underdetermined inverse problems, which met in some applications such as neuroelectromagnetic source localization.

From the two experiments we can see that: (a) in all cases, the proposed ReSBL-QM has superior performance to other algorithms,

---

[5]Since in our experiments the measurement vector number is very small ($L = 3$ or 4), generating sources as AR(1) with various AR coefficient values is sufficient.
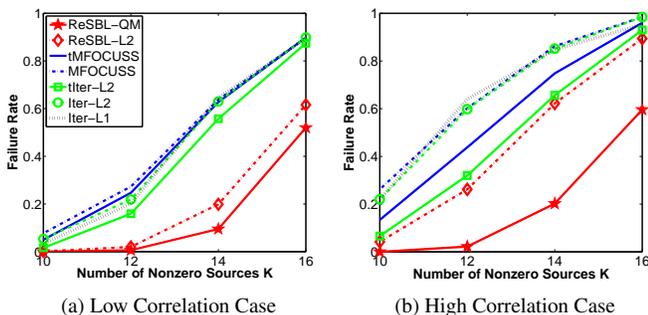

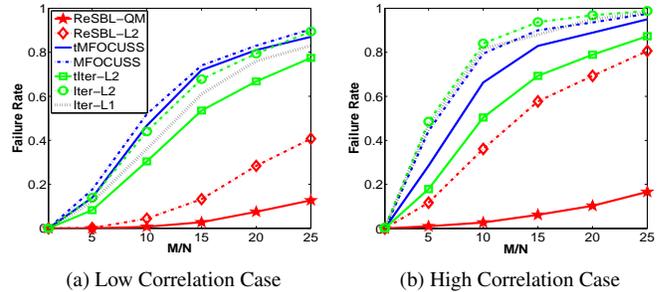(a) Low Correlation Case   (b) High Correlation Case

Fig. 2. Performance when M/N changes.

capable to recover more sources and solve more highly underdetermined inverse problems; (b) without considering temporal correlations of sources, existing algorithms' performance significantly degrades with increasing correlations; (c) after incorporating the temporal correlations, the modified algorithms, i.e. tMFOCUSS and tIter-L2, have better performance than the original M-FOCUSS and Iter-L2, respectively. Also, we note that our proposed algorithms are more effective when the norms of nonzero sources have no large difference (results are not shown here due to space limit).

6. CONCLUSIONS

In this paper, we derived an iterative reweighted sparse Bayesian algorithm exploiting the temporal correlations among source vectors. Its simplified variant was also obtained, which has less computational load. Motivated by our analysis we modified some state-of-the-art reweighted ℓ_2 algorithms achieving improved performance. This work not only provides some effective reweighted algorithms, but also provides a strategy to design effective reweighted algorithms enriching current algorithms on this topic.

7. REFERENCES

[1] S. F. Cotter and B. D. Rao, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. on Signal Processing*, vol. 53, no. 7, 2005.

[2] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, 2010.

[3] E. J. Candes and et al, "Enhancing sparsity by reweighted ℓ_1 minimization," *J Fourier Anal Appl*, vol. 14, 2008.

[4] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *ICASSP 2008*.

[5] Z. Zhang and B. D. Rao, "Sparse signal recovery in the presence of correlated multiple measurement vectors," in *ICASSP 2010*.

[6] ——, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE Journal of Selected Topics in Signal Processing*, submitted.

[7] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. on Signal Processing*, vol. 55, no. 7, 2007.

[8] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

(a) Low Correlation Case   (b) High Correlation Case

Fig. 1. Performance when the nonzero source number changes.