Exploiting Correlation in Sparse Signal Recovery Problems: Multiple Measurement Vectors, Block Sparsity, and Time-Varying Sparsity

Zhilin Zhang   Z4ZHANG@UCSD.EDU
Bhaskar D. Rao   BRAO@UCSD.EDU
ECE Department,University of California at San Diego, La Jolla, CA 92093-0407, USA

1. Introduction

A trend in compressed sensing (CS) is to exploit structure for improved reconstruction performance. In the basic CS model (i.e. the single measurement vector model), exploiting the clustering structure among nonzero elements in the solution vector has drawn much attention, and many algorithms have been proposed such as group Lasso (Yuan & Lin, 2006). However, few algorithms explicitly consider correlation within a cluster. Meanwhile, in the multiple measurement vector (MMV) model (Cotter et al., 2005) correlation among multiple solution vectors is largely ignored. Although several recently developed algorithms consider the exploitation of the correlation, such as the Kalman Filtered Compressed Sensing (KF-CS) (Vaswani, 2008), these algorithms need to know a priori the correlation structure, thus limiting their effectiveness in practical problems.

Recently, we developed a sparse Bayesian learning (SBL) algorithm, namely **T-SBL**, and its variants (Zhang & Rao, 2011a;b; 2010), which adaptively learn the correlation structure and exploit such correlation information to significantly improve reconstruction performance. Here we establish their connections to other popular algorithms, such as the group Lasso, iterative reweighted ℓ_1 and ℓ_2 algorithms, and algorithms for time-varying sparsity. We also provide strategies to improve these existing algorithms.

2. T-SBL: Exploiting Correlation in the MMV Model

The MMV model is expressed as:

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

Here $\mathbf{Y} \triangleq [\mathbf{Y}_{\cdot 1}, \cdots, \mathbf{Y}_{\cdot L}] \in \mathbb{R}^{N \times L}$ is an available measurement matrix consisting of L measurement vectors. $\mathbf{\Phi} \in \mathbb{R}^{N \times M}(N \ll M)$ is a known dictionary matrix,

and any N columns of $\mathbf{\Phi}$ are linearly independent. $\mathbf{X} \triangleq [\mathbf{X}_{\cdot 1}, \cdots, \mathbf{X}_{\cdot L}] \in \mathbb{R}^{M \times L}$ is an unknown and full column-rank solution matrix. A key assumption here is that $\mathbf{X}$ has only a few nonzero rows (i.e. the common sparsity assumption (Cotter et al., 2005)). $\mathbf{V}$ is an unknown noise matrix.

Most existing algorithms ignore the correlation structure in each row of $\mathbf{X}$. In contrast, T-SBL considers such correlation by assuming the joint density of each row vector of $\mathbf{X}$ to be

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

where γ_i is a nonnegative hyperparameter determining whether the i-th row $\mathbf{X}_{i\cdot}$ is zero or not. $\mathbf{B}_i$ is a positive definite matrix that captures the correlation structure of $\mathbf{X}_{i\cdot}$.

By letting $\mathbf{y} = \text{vec}(\mathbf{Y}^T) \in \mathbb{R}^{NL \times 1}$, $\mathbf{D} = \mathbf{\Phi} \otimes \mathbf{I}_L$, $\mathbf{x} = \text{vec}(\mathbf{X}^T) \in \mathbb{R}^{ML \times 1}$, and $\mathbf{v} = \text{vec}(\mathbf{V}^T)$, we can transform the MMV model (1) to the following one

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{v}, \tag{2}$$

where $\mathbf{x}$ is block-sparse with each block being $\mathbf{x}_i \in \mathbb{R}^{L \times 1}$, i.e, $\mathbf{x} = [\mathbf{x}_1^T, \cdots, \mathbf{x}_M^T]^T$. Here $\otimes$ indicates the Kronecker product, and $\text{vec}(\cdot)$ is the vectorization operator.

In the SBL framework (Tipping, 2001), the T-SBL algorithm was derived as follows (Zhang & Rao, 2011a):

$$
\begin{aligned}
\mathbf{x} &= (\lambda \mathbf{\Sigma}_0^{-1} + \mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{y} \\
\mathbf{\Sigma}_x &= \mathbf{\Sigma}_0 - \mathbf{\Sigma}_0 \mathbf{D}^T (\lambda \mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0 \mathbf{D}^T)^{-1} \mathbf{D}\mathbf{\Sigma}_0 \\
\gamma_i &= \frac{1}{L}\text{Tr}[\mathbf{B}^{-1}\mathbf{\Sigma}_x^i] + \frac{1}{L}\text{Tr}[\mathbf{x}_i^T \mathbf{B}^{-1}\mathbf{x}_i], \quad \forall i \ (3) \\
\mathbf{B} &= \frac{1}{M}\sum_{i=1}^{M} \frac{\mathbf{\Sigma}_x^i + \mathbf{x}_i(\mathbf{x}_i)^T}{\gamma_i}
\end{aligned}
$$

where $\mathbf{\Sigma}_x^i$ is the i-th principal diagonal block of size $L \times L$ in $\mathbf{\Sigma}_x$. $\mathbf{\Sigma}_0$ is a block diagonal matrix with

each block given by $\gamma_i \mathbf{B}$. In this algorithm we assume $\mathbf{B}_i = \mathbf{B}$ ($\forall i$) to avoid overfitting. λ is the noise variance, which is also estimated in T-SBL; for clarity we omit its learning rule here (and we also omit such learning rules in the following algorithms). A simplified version, which has much less computational load, is also derived in (Zhang & Rao, 2011a).

We now describe an experiment (Zhang & Rao, 2011a) showing that the proposed T-SBL and its simplified version T-MSBL have superior performance when correlation exists among the solution vectors. In the experiment the Gaussian random dictionary matrix $\mathbf{\Phi}$ had the size of 25×125, the number of nonzero rows of $\mathbf{X}$ was $K = 12$, and L varied from 1 to 4. The correlation among solution vectors was 0 and 0.9 in two cases, respectively. Five algorithms were compared (for details see (Zhang & Rao, 2011a)) to T-SBL and T-MSBL. To avoid the disturbance of the regularization parameters of all the algorithms, we considered a noiseless case. Results (Fig.1) show that when the solution vectors are highly correlated, all the compared algorithms have very poor performance, due to their inability to exploit such correlation.

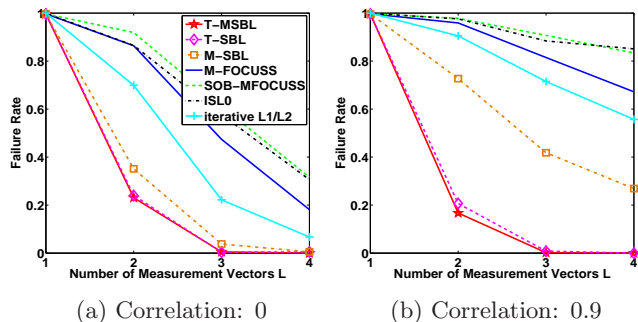In the following we connect T-SBL to other related models.

(a) Correlation: 0   (b) Correlation: 0.9

Figure 1. Failure rates of various algorithms.

3. Connection to Iterative Reweighted ℓ_2 Framework in the MMV Model

The iterative reweighted ℓ_2 minimization framework extended for the MMV problem (in noisy case) computes the solution at the $(k+1)$-th iteration as follows (Wipf & Nagarajan, 2010):

$$\mathbf{X}^{(k+1)} = \arg\min_{\mathbf{x}} \|\mathbf{Y} - \mathbf{\Phi X}\|_{\mathcal{F}}^2 + \lambda \sum_i w_i^{(k)} (\|\mathbf{X}_{i\cdot}\|_q)^2 \quad (4)$$

where $w_i^{(k)}$ is the weight depending on the previous estimate of $\mathbf{X}$. Typically $q = 2$ or $q = \infty$. In (Zhang & Rao, 2011b) we have shown that T-SBL can

be interpreted as an iterative reweighted ℓ_2 algorithm:

$$\begin{aligned} \mathbf{X}^{(k+1)} &= \arg\min_{\mathbf{X}} \Big\{ \|\mathbf{Y} - \mathbf{\Phi X}\|_{\mathcal{F}}^2 + \\ &\quad \lambda \sum_{i=1}^{M} (\gamma_i^{(k)})^{-1} \mathbf{X}_{i\cdot} (\mathbf{B}^{(k)})^{-1} \mathbf{X}_{i\cdot}^T \Big\}. \end{aligned}$$

The learning rules for $\gamma_i^{(k)}$ and $\mathbf{B}^{(k)}$ are given in (Zhang & Rao, 2011b). Note that $\mathbf{X}_{i\cdot} \mathbf{B}^{-1} \mathbf{X}_{i\cdot}^T$ is the quadratic Mahalanobis distance (MD) measure of $\mathbf{X}_{i\cdot}$.

This interpretation reveals the potential advantage of T-SBL is due to using the MD measure of $\mathbf{X}_{i\cdot}$ in the penalty, instead of using typical ℓ_q ($q = 2, \infty$) norms of $\mathbf{X}_{i\cdot}$ (Negahban & Wainwright, 2011). By comparing it to M-SBL (Wipf & Rao, 2007), another SBL algorithm ignoring the correlation in each $\mathbf{X}_{i\cdot}$, we found that T-SBL applies the MD measure also on the weights $(\gamma_i^{(k)})^{-1}$. These observations motivated us to modify existing iterative reweighted ℓ_2 algorithms for better performance, as shown in (Zhang & Rao, 2011b).

Although a strict mathematical proof is missing, these empirical results suggest that the mixed norm based penalties as shown in (4) are not very effective for solving the MMV problem in practice, since the unknown solution vectors are often correlated.

4. Connection to Iterative Reweighted ℓ_1 Framework and Block Sparsity Model

The iterative reweighted ℓ_1 minimization framework (Candes & et al, 2008) extended for the MMV problem is given by (Wipf & Nagarajan, 2010)

$$\mathbf{X}^{(k+1)} = \arg\min_{\mathbf{x}} \|\mathbf{Y} - \mathbf{\Phi X}\|_{\mathcal{F}}^2 + \lambda \sum_i w_i^{(k)} \|\mathbf{X}_{i\cdot}\|_q. \quad (5)$$

We now connect T-SBL to this framework.

For the model (2) the cost function to estimate all the hyperparameters $\Theta \triangleq \{\mathbf{B}, \gamma_i, \forall i\}$ is:

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

Using the identity $\mathbf{y}^T(\lambda \mathbf{I} + \mathbf{D}\mathbf{\Sigma}_0\mathbf{D}^T)^{-1}\mathbf{y} \equiv \min_{\mathbf{x}}\left[\frac{1}{\lambda}\|\mathbf{y} - \mathbf{Dx}\|_2^2 + \mathbf{x}^T\mathbf{\Sigma}_0^{-1}\mathbf{x}\right]$, we can upper-bound the above cost function as follows:

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

By first minimizing over each member of Θ and then minimizing over $\mathbf{x}$, we can get the solution:

$$\mathbf{x} = \arg\min_{\mathbf{x}}\left\{\|\mathbf{y} - \mathbf{Dx}\|_2^2 + \lambda g_{\text{TC}}(\mathbf{x})\right\}, \quad (6)$$

with the penalty defined by $g_{\text{TC}}(\mathbf{x}) \triangleq \min_{\mathbf{B} \succ \mathbf{0}, \gamma_i \geq 0, \forall i} \left\{ \mathbf{x}^T \mathbf{\Sigma}_0^{-1} \mathbf{x} + \log |\lambda \mathbf{I} + \mathbf{D} \mathbf{\Sigma}_0 \mathbf{D}^T| \right\}$. Using the duality theory (Boyd & Vandenberghe, 2004) as in (Wipf & Nagarajan, 2010), we can re-express the optimization problem (6) as follows:

$$
\begin{aligned}
\mathbf{x}^{(k+1)} &= \arg\min_{\mathbf{x}} \| \mathbf{y} - \mathbf{D}\mathbf{x} \|_2^2 \\
&\quad + \lambda \sum_i w_i^{(k)} \sqrt{\mathbf{x}_i^T (\mathbf{B}^{(k)})^{-1} \mathbf{x}_i}. \quad (7)
\end{aligned}
$$

The learning rules for $w_i^{(k)}$ and $\mathbf{B}^{(k)}$ can be derived using the duality theory and the gradient method. Further, using the approximation in (Zhang & Rao, 2011a) we have:

$$
\begin{aligned}
\mathbf{X}^{(k+1)} &= \arg\min_{\mathbf{X}} \| \mathbf{Y} - \mathbf{\Phi}\mathbf{X} \|_{\mathcal{F}}^2 \\
&\quad + \lambda \sum_i w_i^{(k)} \sqrt{\mathbf{X}_{i\cdot} (\mathbf{B}^{(k)})^{-1} \mathbf{X}_{i\cdot}^T}. \quad (8)
\end{aligned}
$$

The learning rules for $w_i^{(k)}$ and $\mathbf{B}^{(k)}$ are given by

$$
\begin{aligned}
w_i &\leftarrow 2 \left(L \mathbf{\Phi}_i^T (\lambda \mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T)^{-1} \mathbf{\Phi}_i \right)^{\frac{1}{2}} \\
\mathbf{B} &\leftarrow \frac{1}{C} \sum_{i=1}^M \frac{\mathbf{X}_{i\cdot}^T \mathbf{X}_{i\cdot}}{\gamma_i}, \text{ with } \gamma_i \triangleq 2 \frac{\sqrt{\mathbf{X}_{i\cdot} \mathbf{B}^{-1} \mathbf{X}_{i\cdot}^T}}{w_i} (9)
\end{aligned}
$$

where $C \triangleq \sum_{i=1}^M \gamma_i \mathbf{\Phi}_i^T (\lambda \mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T)^{-1} \mathbf{\Phi}_i$ and $\mathbf{\Gamma} \triangleq \text{diag}(\gamma_1, \cdots, \gamma_M)$. Note that in each iteration k we need an inner loop to iteratively compute w_i, γ_i and $\mathbf{B}$ until convergence for a better estimate of $\mathbf{B}$. The inner loop generally takes several iterations, and the whole algorithm needs very few outer-loop iterations to achieve its best performance (see Fig.2). In fact, each iteration of (8) yields a sparse solution.

When $\mathbf{B}^{(k)} = \mathbf{I}$ and no iteration was performed, the problem (8) reduces to the group Lasso (for the MMV model). When $\mathbf{B}^{(k)} = \mathbf{I}$ ($\forall k$) and iterative reweighting was performed, the problem (8) is a typical iterative reweighted ℓ_1 algorithm. Thus, T-SBL can be viewed as a variant of iterative reweighted ℓ_1 algorithms. Similar to the ℓ_2 interpretation in the previous section, this interpretation also suggests replacing ℓ_q norms imposed on $\mathbf{X}_{i\cdot}$ by the MD measure in both the penalty and the weights.

To clearly see the advantage of our suggestion, we conduct the same simulation as in Fig.1 (b) when $L = 4$. We used the reweighted ℓ_1 version of T-SBL, the reweighted ℓ_1 version of M-SBL (which corresponds to the ℓ_1 version of T-SBL with $\mathbf{B} = \mathbf{I}$) (Wipf & Nagarajan, 2010), and the original reweighted ℓ_1 algorithm (5) with $q = 2$ and $w_i^{(k)} = (\|\mathbf{X}_{i\cdot}^{(k)}\|_2 + \epsilon)^{-1}$. We also modified this original
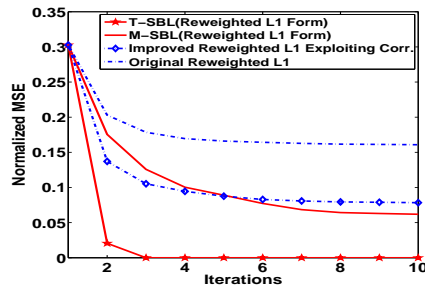

Figure 2. Performance improved when exploiting the correlation.

reweighted ℓ_1 algorithm to exploit the correlation by changing the weights to :

$$
w_i^{(k)} = \left(\sqrt{\mathbf{X}_{i\cdot}^{(k)} (\mathbf{B}^{(k)})^{-1} (\mathbf{X}_{i\cdot}^{(k)})^T} + \epsilon \right)^{-1},
$$

where $\mathbf{B}^{(k)}$ can be estimated by the learning rule (9). But here we set $\mathbf{B}^{(k)}$ ($\forall k$) to be the true value. The result (Fig.2) shows the algorithms are improved when exploiting the correlation. It is worthwhile to notice that the original iterative reweighted ℓ_1 algorithm is greatly improved after we replace the ℓ_2 norm by the MD measure in its weights.

Note that the model (2) is essentially the same as the block sparsity model (Yuan & Lin, 2006; Eldar & Mishali, 2009) [1], a variant of the basic CS model. Thus T-SBL can be applied to this model.

5. Connection to the Time-Varying Sparsity Model

The time-varying sparsity model is a natural extension of the MMV model. It considers the case when the support of each column of $\mathbf{X}$ is time-varying. Several algorithms have been proposed, such as the Kalman Filtered Based Compressed Sensing (KF-CS) (Vaswani, 2008) and Least-Square Compressed Sensing (LS-CS) (Vaswani, 2010). Since this model generally assumes the support is changing slowly, we can view such a time-varying sparsity model as concatenation of several MMV models, where in each MMV model the support does not change. Therefore, T-SBL can be used in this model. Note that in this model exploiting the multiple measurement vectors is important because of the enhanced support-recovery ability afforded by the MMV model, but unfortunately this strategy is missing in current approaches.

To verify this strategy, we conduct an experiment using KF-CS, LS-CS, T-SBL and M-SBL. The Gaussian

---

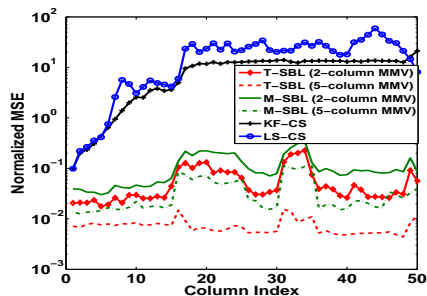[1] Now $\mathbf{D}$ is the original dictionary matrix.

Figure 3. Performance in a time-varying sparsity case.

dictionary matrix was of the size 60×256. The column number of $\mathbf{X}$ was 50. The number of nonzero rows, K, during the first 15 columns of $\mathbf{X}$ was 15. K was increased by 10 since the 16-th and the 31-th column of $\mathbf{X}$, respectively. But since the 26-th column 5 nonzero rows were set to zeros. Each nonzero row had temporal correlation varying from 0.7 to 0.99, and had a duration of 20 columns (if was not set to zeros). SNR was about 20 dB. KF-CS and LS-CS were fed with the true noise variance and the true correlation information. However, both T-SBL and M-SBL learned the noise variance. T-SBL also learned the correlation structures. When performing T-SBL and M-SBL, we approximated the time-varying sparsity model using two methods. One was using the concatenation of 25 MMV models, each MMV model containing 2 columns. The second was using 10 MMV models, each containing 5 columns. The experiment was repeated 100 times. Figure 3 shows that the two MMV algorithms have better performance than KF-CS and LS-CS. Furthermore, T-SBL is super to M-SBL. The experiment code can be downloaded from the first author's website.

6. Conclusions

A general methodology to capture sparsity structure of signals is to use combinations/hierarchy of various norms (Zhao et al., 2009). However, our work showed that another effective way is to use covariance estimation methods to learn the sparsity structures in the framework of SBL. Besides, we showed that iterative reweighted ℓ_1 and ℓ_2 algorithms for the MMV model and the block sparsity model can be greatly improved through replacing their ℓ_q norms imposed on the blocks/groups by the Mahalanobis distance measure, whose covariance matrix is learned data-adaptively.

Acknowledgement

References

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

Candes, E. J. and et al. Enhancing sparsity by reweighted ℓ_1 minimization. *J Fourier Anal Appl,* 14:877–905, 2008.

Cotter, S. F., Rao, B. D., and et al. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE TSP,* 53(7):2477–2488, 2005.

Eldar, Y. C. and Mishali, M. Robust recovery of signals from a structured union of subspaces. *IEEE Trans on Information Theory,* 55(11), 2009.

Negahban, S. and Wainwright, M. J. Simultaneous support recovery in high dimensions: benefits and perils of block ℓ_1/ℓ_∞-regularization. *IEEE Trans. on Information Theory,* 57(6):3841–3863, 2011.

Tipping, M. E. Sparse Bayesian learning and the relevance vector machine. *JMLR,* 1:211–244, 2001.

Vaswani, N. Kalman filtered compressed sensing. In *Proc. of Int. Conf. on Image Process.,* 2008.

Vaswani, N. LS-CS-residual (LS-CS): Compressive sensing on the least squares residual. *IEEE TSP,* 58(8), 2010.

Wipf, D. and Nagarajan, S. Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions. *IEEE Journal of Selected Topics in Signal Processing,* 4 (2), 2010.

Wipf, D. P. and Rao, B. D. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE TSP,* 55(7):3704–3716, 2007.

Yuan, M. and Lin, Y. Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc. B,* 68, 2006.

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

Zhang, Z. and Rao, B. D. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE Journal of Selected Topics in Signal Processing,* 2011a.

Zhang, Z. and Rao, B. D. Iterative reweighted algorithms for sparse signal recovery with temporally correlated source vectors. In *ICASSP,* 2011b.

Zhao, P., Rocha, G., and Yu, B. The composite absolute penalties family for grouped and hierarchical variable selection. *The Annals of Statistics,* 37, 2009.