Reconstruction of a Generalized Joint Sparsity Model Using Principal Component Analysis

Alireza Makhzani
Department of Electrical and Computer Engineering
University of Toronto
makhzani@comm.utoronto.ca

Shahrokh Valaee
Department of Electrical and Computer Engineering
University of Toronto
valaee@comm.utoronto.ca

***Abstract*—In this paper, we define a new Joint Sparsity Model (JSM) and use Principal Component Analysis followed by Minimum Description Length and Compressive Sensing to reconstruct spatially and temporally correlated signals in a sensor network. The proposed model decomposes each sparse signal into two sparse components. The first component has a common support across all sensed signals. The second component is an innovation part that is specific to each sensor and might have a support that is different from the support of the other innovation signals. We use the fact that the common component generates a common subspace that can be found using the principal component analysis and the minimum description length. We show that with this general model, we can reconstruct the signal with smaller samples that are needed by the direct application of the compressive sensing on each sensor.**

I. Introduction

In compressive sensing [1, 2] a relatively small number of random incoherent measurements can be used to uniquely decode a signal that is sparse in a fixed basis. Compressive sensing has many applications including the sensing of spatially and temporally correlated signals by a set of spatially distributed sensors. One of the earliest studies that exploits CS in distributed communication scenarios is the work of Nowak [3, 4]. It has been shown that the spatial correlation can be exploited to reduce the total number of samples to reconstruct the signal. In [5, 6], the signals are assumed to be temporally and spatially correlated, and three models for joint sparsity have been proposed, which are represented by JSM-1, JSM-2 and JSM-3. The three Joint Sparsity Models (JSM) impose certain structure for the sparsity that might be too restrictive in many cases. In JSM-1 and JSM-3, it is assumed that the sensed signals have a common component, which is identical for all sensors. This common component is assumed to be sparse in JSM-1 and non-sparse in JSM-3. However the assumption of having an exact common component across all signals might not be satisfied in practical applications. JSM-2 assumes that the support set of all signals are exactly the same. Algorithms such as SOMP [7], Mixed norm approach [8], M-FOCUSS [9], M-SBL [10], ReMBo [11] and model-based compressed sensing [12] have been proposed for reconstructing JSM-2 signals under the concept of Multiple Measurement Vectors (MMV). However, This model does not allow any variations among the support sets and the assumption of fixed support set is not valid if we are interested in the fine features of the signals. Therefore, despite much improvement obtained in the JSM models, their restrictive assumptions limit their application in practice.

In this work, motivated by the signal models proposed in [5], we propose a more general model called *Generalized Joint Sparsity Model* (G-JSM) that can model more practical cases. We show that JSM-1 and JSM-2 are especial cases of this model and then propose a sensing and reconstruction algorithm for G-JSM. We will see that even in this general model we can still have a better performance compared to separate compressed sensing. Our reconstruction algorithm is based on Principal Component Analysis and Minimum Description Length (MDL) and we will show that this algorithm outperforms SOMP [7] in the G-JSM model. We show that with this general model, we can reconstruct the signal with smaller samples that are needed by the direct application of the compressive sensing on each sensor.

This paper is structured as follows. Our Generalized Joint Sparsity Model is introduced in Section II. In Section III we will propose our reconstruction algorithm. We will compare the performance of our algorithm with that of SOMP and separate compressed sensing using Orthogonal Matching Pursuit [13] in Section IV. Section V concludes the paper.

II. Generalized Joint Sparsity Model

Assume J sensors measuring signals that might be correlated both in space and time. Each sensor node transmits its data to a fusion center. At the fusion center, the data of all sensor nodes are decoded jointly and the spatial correlation is exploited using the Joint Sparsity Models (JSM). In this scheme measurements are taken independently and there is no need for collaboration among sensors.

Let $\mathbf{x}_j, j = 1, \ldots, J$ represents the $N \times 1$ signal of the jth sensor node. We assume that each of the signals consists of two components. The first component ($\mathbf{w}_j$) has exactly the same support set across all sensors, with the sparsity level denoted by K_C, but maybe with different coefficients and the second component is called the *innovation component* ($\mathbf{z}_j$) whose sparsity level is K_j. Therefore, for $j = 1, \ldots, J$,

$$\mathbf{x}_j = \mathbf{w}_j + \mathbf{z}_j$$

$$\|\mathbf{w}_j\|_0 = K_C \ , \ \|\mathbf{z}_j\|_0 = K_j$$

Due to the universality of compressive sensing, without loss of generality, we can assume that the sparsity basis matrix is the identity matrix, and hence $\mathbf{x}_j$, $\mathbf{w}_j$ and $\mathbf{z}_j$ are the signals in the sparse domain.

We call the above model the *Generalized Joint Sparsity Model* (G-JSM). The G-JSM model is less restrictive than JSM-1 [5] since it does not assume that the first components are exactly the same among all signals. It only assumes that the support set of the first components are the same. Furthermore, it is more general than JSM-2 since it assumes that each signal has an innovation component and thus the support set of the signals could be different. We will show that even in this more general model we can achieve perfect reconstruction with fewer measurements than separate compressed sensing.

A practical situation that could be well modeled by G-JSM is where several acoustic sensors are listening to a speech signal. Each signal experiences different attenuation and multi-path effect which cause different amplitudes and phases. However, we expect that the location of the excited coefficients be roughly the same. So we can model the intersection of the support sets by $\mathbf{w}_j$ and the variations by $\mathbf{z}_j$.

At each sensor node, we use the same $M \times N$ random Gaussian measurement matrix $\mathbf{\Phi}$ to get the measurement vector $\mathbf{y}_j = \mathbf{\Phi}\mathbf{x}_j$ and transmit the sampled vector to the fusion center. At the fusion center, we form an $N \times J$ matrix $\mathbf{X}$ whose columns are the J signals and an $M \times J$ matrix $\mathbf{Y}$ whose columns are the measurements. Thus,

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{X}$$

Let $C \subset \{1, ..., N\}$ denotes the common sparse support set of the signals (the support of $\mathbf{w}_j$) and $\mathbf{\Phi}_C$ represents the matrix that is formed by the columns of $\mathbf{\Phi}$ indexed by C. Each measurement vector $\mathbf{y}_j$ is a linear combination of the columns of $\mathbf{\Phi}_C$ plus a linear combination of the columns of $\mathbf{\Phi}$ that correspond to the sparse support set of $\mathbf{z}_j$. If we ignore the innovation components $\mathbf{z}_j$, all of the measurements will lie in a K_C-dimensional subspace which is the span of the columns of $\mathbf{\Phi}_C$ and is equal to the span of the columns of $\mathbf{Y}$ if and only if $J > K_C$. But if we have innovation components, the dimension of the span of the columns of $\mathbf{Y}$ could be as large as M. However we can interpret the departure of measurements from the span of the columns of $\mathbf{\Phi}_C$ as noise and conclude that in the case of G-JSM, the intrinsic dimensionality of the span of the columns of $\mathbf{Y}$ is still K_C.

To estimate the shared support set of the signals, we first estimate the span of the columns of $\mathbf{\Phi}_C$ as the principal subspace of the span of the columns of $\mathbf{Y}$ using the dominant eigenvalues of the covariance matrix of $\mathbf{Y}$. The principal subspace would be the span of the eigenvectors corresponding to the dominant eigenvalues. A model order selection method such as the *Minimum Description Length* (MDL) can be used to find the size of the shared support set (K_C). Afterward, we project all of the columns of $\mathbf{\Phi}$ onto the principal subspace and choose the K_C columns that have the minimum projection error and introduce them as the indices of the shared support set.

The above discussion motivates the application of *principal component analysis* (PCA) [14], which is a method to project signals onto a lower dimensional subspace given that the signals lie close to a manifold of lower dimensionality than that of the original data space.

The sample mean $\bar{\mathbf{y}}$ and the sample covariance matrix $\bar{\mathbf{S}}$ of the vectors $\mathbf{y}_j, j = 1, \ldots, J$ are given as:

$$\bar{\mathbf{y}} = \frac{1}{J}\sum_{j=1}^{J}\mathbf{y}_j \ , \ \bar{\mathbf{S}} = \frac{1}{J}\sum_{j=1}^{J}(\mathbf{y}_j - \bar{\mathbf{y}})(\mathbf{y}_j - \bar{\mathbf{y}})^T$$

Now consider the projection of signals onto a K_C-dimensional space whose direction is defined by orthonormal vectors $\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_{K_C}$. Maximizing the variance of the projected data is equivalent to maximizing $\sum_{j=1}^{K_C} {\mathbf{b}_j}^T\bar{\mathbf{S}}\mathbf{b}_j$ and according to the theory of PCA the maximum is $\sum_{j=1}^{K_C} \lambda_j$ which is achieved by the eigenvectors $\mathbf{u}_1, ..., \mathbf{u}_{K_C}$ of the sample covariance matrix $\bar{\mathbf{S}}$ corresponding to the K_C largest eigenvalues $\lambda_1, \lambda_2, ..., \lambda_{K_C}$.

III. Reconstruction of Correlated Signals

Our reconstruction algorithm has two stages. In the first stage, we estimate the common sparse support set of the first component $\mathbf{w}_j$ using PCA followed by MDL. In the second stage, we subtract the contribution of the estimated common sparse support set and use Orthogonal Matching Pursuit [13] to reconstruct the innovation components of the signals.

A. Recovery of the Shared Support Set

As discussed earlier, the common components of the measurement vectors span a subspace with dimension K_C. We can estimate this subspace by computing the largest eigenvalues of the sample correlation matrix $\bar{\mathbf{S}}$. A simple thresholding can detect the largest eigenvalues. However, thresholding is only useful if the K_Cth largest eigenvalue is much larger than the $(K_C + 1)$st eigenvalue, which is the case if K_C is very small compared to M. In this paper we use MDL, which is a method used to find the similarity of the smallest eigenvalue. Although the equal eigenvalue assumption may not hold in our example, it is nevertheless a proper approximation. This is in particular a good approximation when the innovative components of the signal have support sets that are random across all available dimensions. We are motivated to use PCA followed by MDL since even for a perfect model with spherical white noise, the so-called *signal subspace* remains unaffected by noise. Hence, MDL can successfully identify the largest eigenvalues of the sample covariance matrix.

The proposed algorithm is as follows.

1) Apply PCA on the subspace created by the span of the columns of Y and find the eigenvalues and corresponding eigenvectors:

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_M$$

2) Use MDL to find K_C as the point that minimizes the MDL cost function:

$$MDL(i) = -\log\left[\frac{\prod_{k=i+1}^{M}\lambda_k}{\left(\frac{1}{M-i}\sum_{k=i+1}^{M}\lambda_k\right)^{M-i}}\right]^N + \frac{1}{2}i(2M-i)\log N$$

3) Pick the first K_C eigenvalues and form a subspace by the span of their corresponding eigenvectors.
4) Find the error of the projection of all of the columns of $\mathbf{\Phi}$ onto the subspace found in the previous step.
5) The shared support set is the indices of the columns that have the minimum projection error.

B. Recovery of the Innovations

After finding the indices of the shared support set, we are essentially dealing with a compressed sensing problem with partially known support set in which the size of the known support set is K_C and the size of the unknown support set is K_j. In [15] , a modified-CS algorithm is proposed which uses convex relaxation to find the sparsest signal outside of the known part of the support set. However, in many of applications, ℓ_1 minimization approaches are not fast enough to meet the needs. Here, we use the approach proposed in [16]. In this algorithm, we can treat $\mathbf{w}_j$ as the sparse noise and subtract its contribution from the measurements $\mathbf{y}_j$.

Thus we just need to find a matrix $\mathbf{P}$ such that when we multiply both sides of equation (1) by $\mathbf{P}$, the contribution of $\mathbf{w}_j$ is removed.

$$\begin{aligned}\mathbf{x}_j &= \mathbf{w}_j + \mathbf{z}_j\\ \mathbf{y}_j &= \mathbf{\Phi}(\mathbf{w}_j + \mathbf{z}_j)\\ \mathbf{P}\mathbf{y}_j &= \mathbf{P}\mathbf{\Phi}\mathbf{w}_j + \mathbf{P}\mathbf{\Phi}\mathbf{z}_j\end{aligned} \tag{1}$$

We know that $\mathbf{\Phi}\mathbf{w}_j$ lies in the subspace created by the span of the columns of $\mathbf{\Phi}_C$. So by choosing matrix $\mathbf{P}$ as the projection matrix that projects the vectors of $\mathbb{R}^M$ onto the orthogonal complement subspace of the span of the columns of $\mathbf{\Phi}_C$, we can null out all of the possible vectors of the form $\mathbf{\Phi}\mathbf{w}_j$. More formally if we choose $\mathbf{P}$ as

$$\mathbf{P} = \mathbf{I} - \mathbf{\Phi}_C(\mathbf{\Phi}_C^T\mathbf{\Phi}_C)^{-1}\mathbf{\Phi}_C^T$$

we will have

$$\begin{aligned}\forall \mathbf{w}_j,\ \mathbf{P}\mathbf{\Phi}\mathbf{w}_j &= (\mathbf{I} - \mathbf{\Phi}_C(\mathbf{\Phi}_C^T\mathbf{\Phi}_C)^{-1}\mathbf{\Phi}_C^T)\mathbf{\Phi}\mathbf{w}_j = 0\\ \Rightarrow \mathbf{P}\mathbf{y}_j &= \mathbf{P}\mathbf{\Phi}(\mathbf{w}_j + \mathbf{z}_j) = \mathbf{P}\mathbf{\Phi}\mathbf{z}_j\end{aligned} \tag{2}$$

It can be shown that $\mathbf{P}\mathbf{\Phi}$ satisfies the RIP condition and thus we can use standard compressed sensing algorithms such as Orthogonal Matching Pursuit [13] to reconstruct each $\mathbf{z}_j$ separately.

The intuition behind our algorithm is that we are essentially reducing the sparsity level of the signals by subtracting the contribution of the common support set, thus we can achieve better performance compared to separate compressed sensing.

IV. Numerical Results

In this section we investigate the performance of the proposed algorithm numerically. We run the algorithm on $J \in \{75, 100\}$ signals of length $N = 50$ and different sparsity levels and number of measurements. We then find the probability of exact reconstruction and average the results over 1000 simulation runs.

A. Comparison of PCA and SOMP in Reconstructing the Shared Support Set of the Signals

In this part, we compare the performance of PCA with $J \in \{75, 100\}$ and that of SOMP with $J = 100$, which is proposed in [7] in terms of finding the shared support set of signals with sparsity level $K_j = 5$, $K_C = 5$. One obvious advantage of PCA is the ability to find K_C using statistical techniques such as MDL, while in SOMP we would need to know K_C in advance to determine the number of iterations. Moreover, as Figure 1 indicates, PCA outperforms SOMP in terms of the probability of exact reconstruction of the shared support set in G-JSM. This is mainly due to the fact that SOMP is tailor-made for the case that support set of the signals are exactly the same, while PCA considers the existence of noise and tries to minimize it.

B. Comparison of Joint Reconstruction with Separate Reconstruction of the Signals

In this part, we use $J \in \{75, 100\}$ signals with sparsity level $K_j = 3$, $K_C = 3$ to compare the performance of the proposed algorithm with separate compressed sensing in terms of probability of exact reconstruction of the signals. This comparison is shown in Figure 2. We first use PCA to find the common sparse support, subtract its contribution from the measurements and then use OMP to reconstruct the innovation components. The second plot corresponds to applying OMP directly on each signal and computing the probability of exact reconstruction. As the Figure illustrates, we can exploit the inter-sensor correlation and substantially decrease the number of measurements.

C. Performance of PCA Followed by MDL in an Imperfect G-JSM

Up to now, we have assumed that all signals have the shared support set ($\mathbf{w}_j$), However in practical situations, it is likely that a few number of sensors do not share the common support set. Here, we investigate the performance of applying PCA followed by MDL on the measurements in the scenario where some of the signals are not obeying the G-JSM model. Assume that we have $J = 100$ signals with $N = 50$, $M = 25$ and α percent of them share a common support set with $K_C = 3$, $K_j = 3$ and the other signals do not share the common support set and the sparsity level of them is $K_C + K_j = 6$. Figure 3 shows the probability of reconstruction of the common support set versus α using MDL and PCA. As this Figure indicates, even in the case where $\%10$ of the signals do not obey the G-JSM model, our reconstruction algorithm achieves perfect reconstruction of the shared support set. In this case,

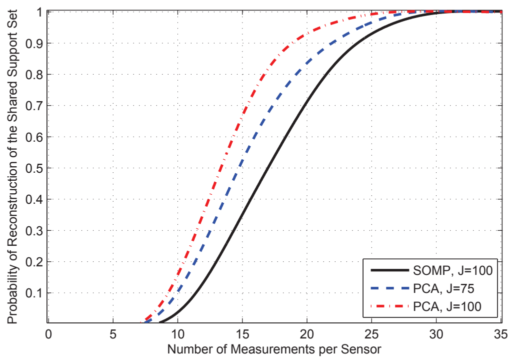

Fig. 1. Reconstruction of the common support set in G-JSM using PCA and SOMP for $J \in \{75, 100\}$, $N = 50$, $K_C = 5$ and $K_j = 5$.

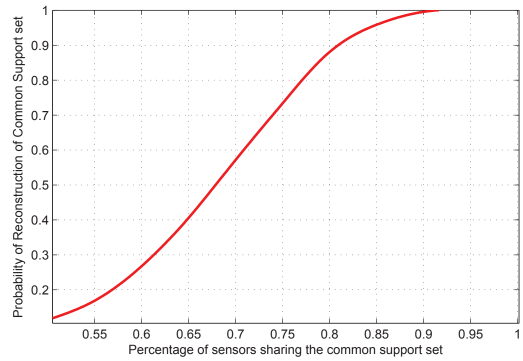

Fig. 3. Reconstruction of the common support set in an imperfect G-JSM model for $J = 100$, $N = 50$, $M = 25$, $K_C = 3$ and $K_j = 3$.

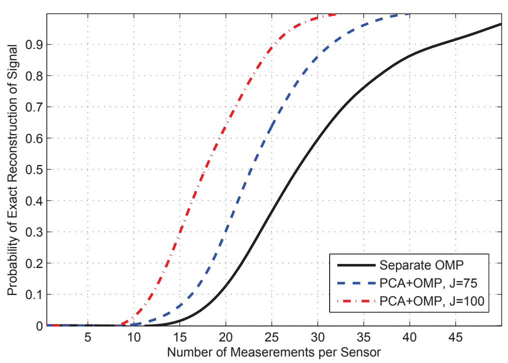

Fig. 2. Reconstruction of signals in G-JSM using PCA-OMP and Separate OMP for $J \in \{75, 100\}$, $N = 50$, $K_C = 3$ and $K_j = 3$.

subtracting the contribution of the common support set does not affect $\%10$ of the signals but improves the probability of reconstruction of $\%90$ of the signals as it does in a perfect G-JSM model. Thus we can conclude that the performance of the whole algorithm in this scenario is the same as G-JSM model for $\%90$ percent of the signals and the same as separate CS for $\%10$ of them.

V. Conclusion

In this work, we proposed the G-JSM model which is more general than the joint sparsity models in the literature and is suitable for practical applications. We applied PCA on the measurements to determine the common component of the signals and then used MDL to find the sparsity level of the common component. After finding the common component, we subtracted the contribution of it from the measurements and used standard compressive sensing on each individual signal to find the innovation components. We further showed that our proposed sampling and reconstruction algorithm outperforms the conventional SOMP algorithm for the reconstruction of the common support set of the signals in this particular joint sparsity model (G-JSM). Our simulation results show that even in this general model for sensor networks, we can achieve a much better performance than separate compressed sensing reconstruction.

References

[1] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3, pp. 1433–1452, Citeseer, 2006.

[3] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proceedings of the 5th international conference on Information processing in sensor networks*, IPSN '06, (New York, NY, USA), pp. 134–142, ACM, 2006.

[4] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 92–101, 2008.

[5] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *In Asilomar Conf. Signals, Sys., Comput*, pp. 1537–1541, 2005.

[6] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and R. Baraniuk, "Distributed compressed sensing," 2005.

[7] J. Tropp, A. Gilbert, and M. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 5, pp. v/721 – v/724 Vol. 5, 2005.

[8] J. Tropp, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[9] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2477–2488, 2005.

[10] D. Wipf, *Bayesian methods for finding sparse representations*. PhD thesis, University of California, San Diego, 2006.

[11] M. Mishali and Y. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *Signal Processing, IEEE Transactions on*, vol. 56, no. 10, pp. 4692–4702, 2008.

[12] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982–2001, 2010.

[13] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655 –4666, 2007.

[14] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[15] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4595–4607, 2010.

[16] M. Davenport, P. Boufounos, and R. Baraniuk, "Compressive domain interference cancellation," *Structure et parcimonie pour la représentation adaptative de signaux (SPARS), Saint-Malo, France*, 2009.