Clustered Pattern Sparse Signal Recovery Using Hierarchical Bayesian Learning

Mohammad Shekaramiz, *Student Member, IEEE,*
Advisor: Todd K. Moon, *Senior Member, IEEE,*
Co-Advisor: Jacob H. Gunther, *Senior Member, IEEE*
ECE Department and Information Dynamics Laboratory, Utah State University

Abstract— **Recently, we proposed a novel hierarchical Bayesian learning algorithm for the recovery of sparse signals with unknown clustered pattern for the general framework of multiple measurement vectors (MMVs). In order to recover the unknown clustered pattern we incorporated a parameter to learn the number of transitions over the support set of the solution. This parameter does not exist in other algorithms, and it is learned via our hierarchical Bayesian algorithm.**

I. Introduction

The problem that we address in this paper is for the recovery of sparse signals via either Single- and multiple measurement vector (SMV or MMVs) when the sparse solution exhibits an unknown clustered pattern. In this case, we provide a new hierarchical sparse Bayesian learning model. SMV and MMV are computational inverse problems in the compressive sensing area. Compressive sensing (CS) enables us to represent a sparse or compressible signal via a small set of linear measurements [1]. The SMV seeks the sparsest solution $\mathbf{x}_s$ in $\mathbf{y} = A\mathbf{x}_s + \mathbf{e}$, where A is wide and known and $\mathbf{e}$ is the measurement noise. MMV deals with the same problem as the SMV but for the case where Y and X are matrices i.e., $Y = AX_s + E$. In MMV, it is usually assumed that the non-zero elements in the columns of X_s occur at the same rows. This has been referred to as joint sparsity in the literature.

In some applications such as neuromagnetic imaging [2] and direction of arrival (DOA) estimation [3], non-zero components of the sparse signal appear in clusters. Therefore, in addition to joint sparsity, a clustered pattern also appears in the columns of X_s. There exist many greedy-based algorithms for the recovery of clustered pattern sparse signals via the SMV [4], [5]. However, it turns out that such algorithms need some prior knowledge on the cluster pattern. Other than the greedy algorithms, there also exist some sparse Bayesian learning (SBL) algorithms devised for the SMV and MMVs, categorized as follows. The first approach uses a zero-mean Gaussian prior with some precision accompanied with a Gamma hyper prior on the precision [6]. Promoting the clustered pattern solution is then accomplished by defining different set of hyper parameters on the Gamma hyper prior depending on the active or inactive status of neighbor supports [7]. The second SBL approach uses a Bernoulli-Gaussian or equivalently spike-and-slab prior [8]. Our proposed model falls within the second sparse Bayesian modeling category i.e., using Bernoulli-Gaussian prior. The proposed model can be applied to either SMV or MMVs. Our main contribution in this area is incorporating an additional parameter to measure the contiguity in the supports of the solution. One can think of such parameter as a knob that determines the overall clumpiness of the supports of the solution. This parameter is learned via our hierarchical SBL algorithm. Previously developed algorithms do not have this control parameter for learning the pattern via the measure of overall clumpiness over the solution.

This work is supported by the Grant NASA NNX13 AD 39 A.

II. Clustered-Sparse Bayesian Learning

The basic model for the MMV problem is defined as follows

$$Y = A(\mathbf{s} \circ X) + E, \tag{1}$$

where $Y \in \mathbb{R}^{M\times N}$, $A \in \mathbb{R}^{M\times P}$, $\mathbf{s} \in \{0,1\}^{P\times 1}$, $X \in \mathbb{R}^{P\times N}$, and "$\circ$" denotes the Hadamard product. In (1), Y is the observed noisy data, A is a known sensing matrix, $\mathbf{s}$ is an unknown binary support-learning vector, X is an unknown solution-values matrix, and E denotes the measurement noise. Notice that we assumed the same precision on the components of X for the simplification purposes. The columns of X in (1) are assumed to be drawn i.i.d. as follows

$$\mathbf{x}_n \sim \mathcal{N}(0, \tau^{-1} I_P), \tau \sim \text{Gamma}(a_0, b_0), n = 1, \dots, N. \tag{2}$$

The hyper-parameters are set to $a_0 = b_0 = 10^{-3}$, where a_0 and b_0 denote the shape and rate of the Gamma distribution, respectively. We model the noise components in E as follows

$$e_{mn} \sim \mathcal{N}(0, \varepsilon^{-1}),\ \forall m, n, \quad \varepsilon \sim \text{Gamma}(\theta_0, \theta_1). \tag{3}$$

The hyper-parameters in (3) are set to $\theta_0 = \theta_1 = 10^{-3}$.

We assume that the columns of the solution matrix are jointly sparse and each $\mathbf{x}_n$ might have groups of adjacent non-zeros. In this case, we measure the amount of clumpiness in the support-learning vector $\mathbf{s}$ as follows [9]

$$\Sigma\Delta(\mathbf{s}) = \sum_{p=2}^{P} |s_p - s_{p-1}|. \tag{4}$$

There exist fewer transitions in $\mathbf{s}$ for the case where the supports of the solution have a clustered pattern compared to random distribution of supports, corresponding to a smaller $(\Sigma\Delta)$ measure computed according to (4). We model the prior on the elements of $\mathbf{s}$ as follows

$$\begin{aligned}(s_p; \omega_{0,p}, \omega_{1,p}) &\sim \text{Bernoulli}(\frac{\omega_{1,p}}{\omega_{0,p} + \omega_{1,p}}), \forall p = 1, 2, ..., P,\\ \omega_{k,p} &= e^{-\alpha(\Sigma\Delta)_{k,p}} \text{Binomial}(\Sigma_{k,p}, P, \gamma_p), \forall k = 0, 1,\end{aligned} \tag{5}$$

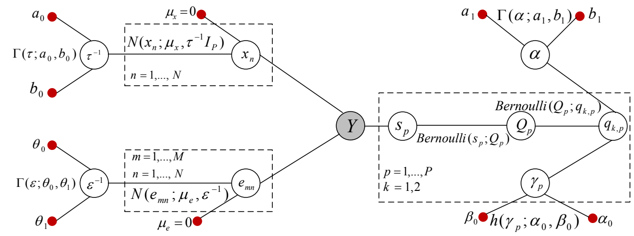

Fig. 1. Graphical model of the proposed model.

where α specifies the significance of the amount of clumpiness in the support vector $\mathbf{s}$. As was mentioned earlier, α is also to be learned in a Bayesian fashion. For this purpose, we employ the prior $\alpha \sim \text{Gamma}(a_1, b_1)$ on α, where $a_1 = 2 \times 10^{-3}$ and $b_1 = 10^{-3}$. The proposed graphical model is shown in Fig. 1. In Fig. 1, the large shaded node shows the observations and small solid nodes represent the hyper-parameters. Each unshaded node denotes a random variable(s) [10].

Below, we provide a pseudocode description of our algorithm (C-SBL) for the clustered pattern SMV/MMVs using the obtained posterior inferences on the variables.

C-SBL Algorithm:

$\{\Theta^{(i)}\}_{i=1 \text{ to } N_{collect}} = \textbf{C-SBL}(Y, A, \Theta_0, N_{burn-in}, N_{collect})$

For $Iter = 1$ to $N_{burn-in} + N_{collect}$

For $p = 1$ to P

$\tilde{y}_{mn}^{-p} = y_{mn} - \sum_{l \neq p}^{P} a_{ml} s_l x_{ln}, \quad \forall m = 1 \text{ to } M, \forall n = 1 \text{ to } N$

$c_p = \frac{1-\gamma_p}{\gamma_p} \frac{\Sigma_{1,p}}{P+1-\Sigma_{1,p}}$

$k_p = e^{\frac{\varepsilon}{2}\left((\|\mathbf{a}_p\|_2^2 \sum_{n=1}^{N} x_{pn}^2) - 2\mathbf{a}_p^T(\sum_{n=1}^{N} x_{pn}\tilde{\mathbf{y}}_n^{-p})\right)}$

$(s_p|-) \sim \text{Bernoulli}\left(\frac{1}{1+c_p k_p e^{-\alpha\left((\Sigma\Delta)_{0,p} - (\Sigma\Delta)_{1,p}\right)}}\right)$

For $l = 1$ to P

$\Sigma_x = (\tau + \varepsilon s_l^2 \, \|\mathbf{a}_l\|_2^2)^{-1}$, $\bar{\boldsymbol{\mu}} = \varepsilon s_l \Sigma_x \mathbf{a}_l$

$\tilde{\mathbf{y}}_n^{-l} = \mathbf{y}_n - A(\mathbf{s} \circ \mathbf{x}_n) + s_l x_{l,n} \mathbf{a}_l, \ \forall n = 1, \ldots, N$

$(x_{l,n}|-) \sim \mathcal{N}(\bar{\boldsymbol{\mu}}^T \tilde{\mathbf{y}}_n^{-l}, \Sigma_x), \ \forall n = 1, \ldots, N$

End For $\{l\}$

$(\gamma_p|-) \sim \text{Beta}\left(\alpha_0 + 1 + 2\sum_{k \neq p}^{P} s_k\ ,\ \beta_0 - 1 + 2(P - \sum_{k \neq p}^{P} s_k)\right)$

End For $\{p\}$

$(\tau|-) \sim \text{Gamma}(a_0 + \frac{NP}{2}, b_0 + \frac{1}{2}\|X\|_F^2)$

$(\varepsilon|-) \sim \text{Gamma}(\theta_0 + \frac{MN}{2}, \theta_1 + \frac{1}{2}\|Y - A(\mathbf{s} \circ X)\|_F^2)$

α obtains from soving for $\alpha^{[t+1]}$ in

$$\sum_{p=1}^{P} \frac{(\bar{\Sigma\Delta})_p^{[t]}}{1 + \frac{1}{c_p^{[t]}} e^{\alpha^{[t+1]}(\bar{\Sigma\Delta})_p^{[t]}}} - \sum_{p=1}^{P}(1 - s_p^{[t]})(\bar{\Sigma\Delta})_p^{[t]} + \frac{a_1 - 1}{\alpha^{[t+1]}} - b_1 = 0$$

$\Theta^{(Iter - N_{burn-in})} \leftarrow \Theta, \quad \forall Iter > N_{burn-in}$

End For {Iter}

In the above inferences, the terms $(\Sigma\Delta)_{k,p}$ and $\Sigma_{k,p}$ denote the $(\Sigma\Delta)$ value and the sum over all the elements of $\mathbf{s}$ for the case where $s_p = k$, respectively. In our implementation we approximate the posterior densities via Markov chain Monte Carlo (MCMC) method and using Gibbs sampling. We run C-SBL for $N_{burn-in}$ iterations, where $N_{burn-in}$ is set experimentally until reaching stable Markov chain. Then, we perform $N_{collect}$ more iterations and start collecting samples.

III. Simulation Results

We show the performace of our proposed algorithm compared to MFOCUSS [2], MSBL [11], and T-MSBL [12], [13] algorithms for the MMV problem. The number of columns in the solution X is set to $N = 2$. In Fig. 2, we illustrate the obtained experimental results in terms of normalized mean-square error (NMSE) for the slution when the ratio M/P varies. In Fig. 2, the term ρ denotes the correlation factor that has been considered between the columns of the true solution.

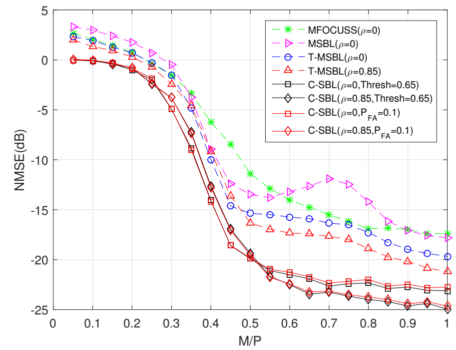

Fig. 2. Comparison of the error between the true and the estimated solution.

IV. Conclusion and Future Work

We proposed a new sparse Bayesian learning algorithm for the recovery of sparse signals with unknown clustered pattern via MMVs. Based on the simulations, we showed that our algorithm provides encouraging performance. As a future work, we will modify our model to account for the possible existance of correlation between the columns of the solution.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Th.*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] S. F. Cotter, B. D. Rao, K. Engan, and K. K. Delgado, "Sparse solutions to linear inverse problem with multiple measurement vectors," *IEEE Trans. Sig. Proc.*, vol. 53, no. 7, pp. 2477–2488, 2005.

[3] J. A. Luo, X. P. Zhang, and Z. Wang, "Direction-of-arrival estimation using sparse variable projection optimization," in *Proc. Int. Symposium on Circs. and Syst. (ISCAS)*, pp. 3106–3109, May 2012.

[4] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Sig. Proc.*, vol. 56, no. 10, pp. 4692–4702, 2008.

[5] Y. C. Eldar, P. Kuppinger, and H. Bőlcskei, "Block-sparse signals: Uncertainty relations and effiecient recovery," *IEEE Trans. Sig. Proc.*, vol. 58, no. 6, pp. 3042–3054, 2010.

[6] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis pursuit selection," *IEEE Trans. Sig. Proc.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[7] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Sig. Proc.*, vol. 63, no. 2, pp. 360–372, 2015.

[8] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Process.*, vol. 92, no. 1, pp. 259–269, 2012.

[9] M. Shekaramiz, T. K. Moon, and J. H. Gunther, "Hierarchical Bayesian approach for jointly-sparse solution of multiple-measurement vectors," in *48th Asilomar Conf. of Signals, Systems, and Computers*, pp. 1962–1966, Nov. 2014.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2009.

[11] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Sig. Proc.*, vol. 55, no. 7, pp. 3704–3716, 2007.

[12] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. of Selected Topics in Sig. Proc.*, vol. 5, no. 5, pp. 912–926, 2011.

[13] Z. Zhang and B. D. Rao, "Clarify some issues on the sparse Bayesian learning for sparse signal recovery." Univ. of Cal., San Diego, Technical Report, Sep. 2011.