A Novel Compressive Sensing Scheme under the Variational Bayesian Framework

Vangelis P. Oikonomou[*,1], Spiros Nikolopoulos[1] and Ioannis Kompatsiaris[1]

***Abstract*— In this work we provide a novel algorithm for Bayesian Compressive Sensing. The proposed algorithm is considered for signals that features two properties: grouping structure and sparsity between groups. The Compressive Sensing problem is formulated using the Bayesian linear model. Furthermore, the sparsity of the unknown signal is modeled by a parameterized sparse prior while the inference procedure is conducted using the Variational Bayesian framework. Experimental results, using 1D and 2D signals, demonstrate that the proposed algorithm provides superior performance compared to state-of-the-art Compressive Sensing reconstruction algorithms.**

compressed sensing, group sparsity, parameterized prior, variational bayesian

I. Introduction

The notion of sparsity has influenced computer science in a fundamental way, causing scientists to challenge well established ideas. One such case can be considered the framework of Compressive Sensing (CS) [1], [2]. Compressive Sensing states that sparse signals can be recovered by using much less measurements than dictated by Shannon/Nyquist Theorem. This is achieved by "transferring" information from the sampling procedure (i.e. Shannon/Nyquist Theorem) to the underlying structure of the signal (i.e. sparsity). More precisely, within this framework we use prior information about the signal to constraint the sampling procedure. However, CS do not rely only on sparsity but also to the incoherence and the restricted isometry property (RIP) [2]. The last two properties discriminate CS from other similar linear inverse problems such as denconvolution, superresolution, Sparse Bayesian Learning and the inverse EEG problem.

In the CS framework, we do not measure directly the N-dimensional signal of interest $\mathbf{x}$, but we obtain M linear measurements $\mathbf{y}$. The CS measurements, $\mathbf{y}$, can be represented as $\mathbf{y} = \mathbf{\Phi}\mathbf{w}$, where $\mathbf{w}$ represents the transform coefficients associated with a basis, $\mathbf{\Psi}$, in which the signal of interest is (or assumed to be) sparse, $\mathbf{x} = \mathbf{\Psi}\mathbf{w}$, and $\mathbf{\Phi} \in \Re^{M\times N}$ is a random projection matrix with $M < N$ [1], [2]. Finding $\mathbf{w}$ (and hence $\mathbf{x}$) represents an ill-posed problem, however, if we exploit the fact that $\mathbf{w}$ is sparse with respect to a known orthonormal basis (for example in wavelet basis), then we can find an approximation with high accuracy.

A typical approach to solve the above linear inverse problem, by taking into account the sparsity of $\mathbf{w}$, is the ℓ_1-regularized formulation [2], [3]

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}}\{\|\mathbf{y} - \mathbf{\Phi}\mathbf{w}\|_2^2 + \rho\|\mathbf{w}\|_1\} \tag{1}$$

The above CS formulation can be considered as the application of a deterministic regularization approach to signal reconstruction. However, the problem can also be formulated using the Bayesian framework [4], [5], which provides certain distinct advantages over other formulations such as probabilistic predictions, automatic estimation of model parameters, and estimation of the reconstruction uncertainty [4], [6].

Recently, group-sparse models have been proposed for CS [7]–[10] where the sparsity is assumed wth respect to groups of coefficients. By taking into account the grouping structure of the signal, results into superior reconstruction performance [7]. Guided from these works, we proposed a new model-based CS reconstruction algorithm. The novelty of our work relays on a new parameterized group-sparse prior and the development of a novel BCS inversion algorithm. Experimental results have shown that the proposed BCS inversion algorithm provides superior reconstruction performance compared to other similar state-of-the-art reconstruction algorithms.

Next sections are organized as follows. First, we present the proposed algorithm for CS reconstruction. In this section, the Bayesian Linear Model is introduced for Bayesian CS reconstruction. More specifically, we describe the group-sparse prior, the corresponding likelihood of the data, and then, the inference procedure by using the Variational Bayesian (VB) framework. After that, we present the experiments based on synthetic 1D signals and 2D real data (i.e.images). Also, we provide a comparison with well known CS reconstructions algorithms. Finally, we conclude this work by a short discussion and future extensions of our work.

II. Proposed Method

The basic model for Bayesian Compressive Sensing, according to [4], is described by:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{w} + \mathbf{e}, \tag{2}$$

where the matrix $\mathbf{\Phi} \in \Re^{M\times N}$ is the random projection matrix and $\mathbf{e} \in \Re^{M}$ denotes the noise of the model, which follows a gaussian distribution with zero mean and precision (inverse variance) β. Finally, $\mathbf{w} \in \Re^{N}$ is a vector containing the coefficients.

[*] Corresponding author.
[1]V. P. Oikonomou, Spiros Nikolopoulos and I. Kompatsiaris are with the Information Technologies Institute, Centre for Research and Technology Hellas, CERTH-ITI, 6th km Charilaou-Thermi Road, 57001 Thermi-Thessaloniki, Greece. {viknmu, nikolopo, ikom}@iti.gr

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

A. Group - Sparse Priors

In our analysis, we assume that the coefficients $\mathbf{w}$ have a group structure. More specifically, we define G groups of coefficients such that the vector $\mathbf{w}_g$ contains d_g coefficients assigned to group g. Sparsity between groups can be achieved by selecting carefully the prior distribution over them. Assuming a priori independence between groups, and, that each group follows a Gaussian distribution with zero mean and covariance matrix $a_g^{-1}\mathbf{I}_{d_g}$, the prior over coefficients is given by:

$$p(\mathbf{w}|\mathbf{a}) = \mathcal{N}(\mathbf{w}|\mathbf{0}_N, \mathbf{C}_0) = \prod_{g=1}^{G} \mathcal{N}(\mathbf{w}_g|\mathbf{0}_{d_g}, a_g^{-1}\mathbf{I}_{d_g}), \quad (3)$$

where $\mathcal{N}$ is the symbol for Gaussian distribution. Furthermore, we assume that each parameter a_g, which controls the group sparsity of the parameters $\mathbf{w}$, follows a Gamma distribution, so the overall prior over all a_g is a product of Gamma distributions given by: $p(\mathbf{a}) = \prod_{g=1}^{G} Gamma(a_g; b_a, c_a)$. The above hierarchical prior is well known for its sparse properties [11], [12]. In our study we change the above prior by introducing one more parameter. More specifically, we assume that the covariance matrix $\mathbf{C}_0$ is a diagonal matrix with elements $a_g^{-1}\lambda_g^{-1}$. In our analysis, parameters λ_g are assumed known and deterministic quantities. Now the prior distribution of coefficients is given by:

$$p(\mathbf{w}|\mathbf{a};\boldsymbol{\lambda}) = \mathcal{N}(\mathbf{w}|\mathbf{0}_N, \mathbf{C}_0) = \prod_{g=1}^{G} \mathcal{N}(\mathbf{w}_g|0, a_g^{-1}\lambda_g^{-1}\mathbf{I}_{d_g}), \quad (4)$$

This prior has been used to extract features in a classification problem when no group structure is assumed ($G = 1$) [13]. Finally, the overall precision (inverse variance) β of the noise follows a Gamma distribution: $p(\beta) = Gamma(\beta; b, c) = \frac{1}{\Gamma(c)} \frac{\beta^{(c-1)}}{b^c} \exp\Big\{ - \frac{\beta}{b} \Big\}$, where b and c is the scale and the shape of the Gamma distribution, respectively. We use the Gamma distribution for two reasons: First, this distribution is conjugate to the Gaussian distribution, which helps us in the derivation of closed form solutions, and second it places the positivity restriction on the overall variance and scaling parameters.

B. VB Inference

The overall prior over model parameters $\{\mathbf{w}, \mathbf{a}, \beta\}$ is given by: $p(\mathbf{w}, \mathbf{a}, \beta; \boldsymbol{\lambda}) = \prod_{g=1}^{G} p(\mathbf{w}_g|a_g; \lambda_g) \prod_{g=1}^{G} p(a_g)p(\beta)$. The likelihood of the data is given by:

$$p(\mathbf{y}|\mathbf{w}, \beta; \boldsymbol{\lambda}) = \frac{\beta^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \exp\Big\{ - \frac{\beta}{2}(\mathbf{y} - \boldsymbol{\Phi}\mathbf{w})^T(\mathbf{y} - \boldsymbol{\Phi}\mathbf{w}) \Big\} \quad (5)$$

At this point it is worth to note that two general bayesian approaches could be used to estimated the interested quantities: the Bayesian Evidence Framework and the VB Methodology [14]. While the two approaches have significant similarities, there also have a significant difference. Bayesian Evidence, since it is based on the classical EM algorithm, computes point estimates of posterior distribution. From the other side, the VB computes estimates of the actual posterior distribution. In our study we adopt the VB methodology. To apply the VB methodology [14] we need to define an approximate posterior based on one factorization over the parameters $\{\mathbf{w}, \mathbf{a}, \beta\}$. In our study we choose the following factorization: $q(\mathbf{w}, \mathbf{a}, \beta; \boldsymbol{\lambda}) = q(\mathbf{w}; \boldsymbol{\lambda}) \prod_{g=1}^{G} q(a_g)q(\beta)$ [14]. Notice that this factorization makes the groups dependent a posteriori. Applying the VB methodology, and taking into account the above factorization, the following posteriors are obtained:

$$\begin{aligned} q(\mathbf{w}) &= \mathcal{N}(\langle\mathbf{w}\rangle, \mathbf{C}_{\mathbf{w}}), & (6)\\ q(\beta) &= Gamma(\beta; b', c'), & (7)\\ q(\mathbf{a}) &= \prod_{g=1}^{G} Gamma(a_g; b'_{a_g}, c'_{a_g}), & (8) \end{aligned}$$

The moments of each distribution are calculated by applying iteratively the following equations until convergence [13], [14]:

$$\begin{aligned} \mathbf{C}_{\mathbf{w}} &= (\langle\beta\rangle\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \langle\mathbf{C}_0\rangle)^{-1}, & (9)\\ &= \langle\mathbf{C}_0\rangle^{-1} - \langle\mathbf{C}_0\rangle^{-1}\boldsymbol{\Phi}^T\Big(\langle\beta\rangle^{-1}\mathbf{I} + \\ &\quad \boldsymbol{\Phi}\langle\mathbf{C}_0\rangle^{-1}\boldsymbol{\Phi}^T\Big)^{-1}\boldsymbol{\Phi}\langle\mathbf{C}_0\rangle^{-1} \\ \langle\mathbf{w}\rangle &= \mathbf{C}_{\mathbf{w}}\langle\beta\rangle\boldsymbol{\Phi}^T\mathbf{y}, & (10)\\ \frac{1}{b'_{a_g}} &= \frac{\lambda_g}{2}(\langle\mathbf{w}_g\rangle^T\langle\mathbf{w}_g\rangle + Tr\{\mathbf{C}_g\}) + \frac{1}{b_a}, & (11)\\ c'_{a_g} &= \frac{d_g}{2} + c_a, & (12)\\ \langle a_g\rangle &= b'_{a_g}c'_{a_g}, & (13)\\ \frac{1}{b'_{\beta}} &= \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\langle\mathbf{w}\rangle\|_2^2 + Tr(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{C}_{\mathbf{w}}) + \frac{1}{b}, & (14)\\ c'_{\beta} &= \frac{N}{2} + c, & (15)\\ \langle\beta\rangle &= b'_{\beta}c'_{\beta}, & (16) \end{aligned}$$

In the above equations the matrix $\langle\mathbf{C}_0\rangle$ is a diagonal matrix with $\langle a_g\rangle \cdot \lambda_g$ in its main diagonal, where each element $\langle a_g\rangle \cdot \lambda_g$ is repeated d_g times. Each parameter λ_g is set to $\frac{1}{\|\langle\mathbf{w}_g\rangle\|_1}$ at the end of each iteration. Also, $\mathbf{C}_g$ denotes the submatrix of $\mathbf{C}_{\mathbf{w}}$ corresponding to the g-th group.

III. Experimental Results

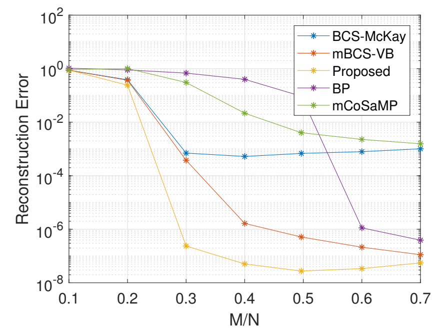
In this section we present experimental results demonstrating the performance of the proposed CS reconstruction algorithm. Our experiments are divided into two parts. In the first part, we perform experiments using 1D synthetic signals, where we can control the group-sparsity and various other experimental settings such as group size, etc. In the second part, 2D images are used to demonstrate the performance of the proposed method in a more practical setting.

We compare our algorithm with five well - known CS reconstruction algorithms: the Basis Pursuit (BP) [15], the model-based CoSaMP recovery (mCoSaMP) [7], the modified Bayesian Compressive Sensing (BCS) [4] based on VB (mBCS-VB), the group-sparse modeling using the McKay distribution (BCS-McKay) [9] and the wavelet based tree-structured reconstruction algorithm (TS-BCS-VB) [8]. For the BP implementation, we use the solver *SolveBP* from the *SparseLab* toolbox[1]; for the mCoSaMP algorithm we use the package *modelcs_v1.1*[2]; for the BCS-McKay algorithm we use the package *VBGS_v02*[3]; for the TS-BCS-VB algorithm we use the package *tsbcs_vb*[4]. Finally, in our implementation for mBCS-VB, we have extended the BCS-VB algorithm [4] in order to include the case of group sparsity under the bayesian framework. As a measure of performance comparison between the methods, we use the reconstruction error between the true signal and the estimated signal. In the case of synthetic signals the reconstruction error is given by $\|\mathbf{w}_{est} - \mathbf{w}_{true}\|_2^2 / \|\mathbf{w}_{true}\|_2^2$, where $\mathbf{w}_{true}$ is the true coefficients and $\mathbf{w}_{est}$ is the estimated coefficients. In the case of images the reconstruction error is given by $\|\mathbf{f}_{est} - \mathbf{f}_{true}\|_2^2 / \|\mathbf{f}_{true}\|_2^2$, where $\mathbf{f}_{true}$ is the true image and $\mathbf{f}_{est}$ is the estimated image. Finally. each sample point in the curves, shown below, is generated by performing 100 trials of the corresponding algorithm and averaging the results.

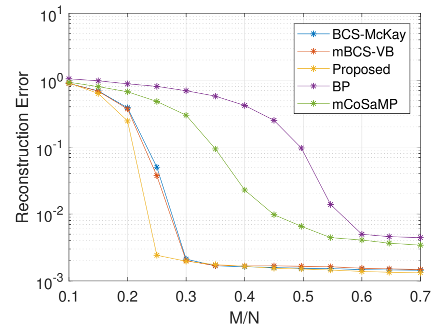
A. Synthetic Signals

To compare the estimation performance of different reconstruction algorithms, we generated a collection of signals with length $N = 300$. For the construction of coefficients we use sparse signals with 60 coefficients Gaussian-distributed with variance 1, and set the remaining coefficients to zero. Also, we fix the group size to 20 ($d_g = 20$, $G = 15$). The $M \times N$ $\mathbf{\Phi}$ matrix is generated by drawing its entries from a standard Gaussian distribution and normalizing the columns to have unit -norm. Finally, with respect to the CS measurements $\mathbf{y}$ we examine two cases, noise-free measurements and noisy measurements. In the noisy case, we have added noise from a Gaussian distribution with zero mean and standard deviation 0.001. In this series of experiments we exclude the TS-BCS-VB, since this algorithm is specifically design to work by taking into account the grouping structure of wavelet transform that, at this point, is not suitable for our analysis.

Fig. 1 indicates the reconstruction error as a function of the number of measurements. When we have noiseless measurements, the proposed algorithm outperforms all algorithms in terms of reconstruction error. Furthermore, our algorithm needs less measurements to reach his optimal reconstruction error. In the case of noisy measurements we can see that all algorithm achieves similar reconstruction error having adequate number of measurements (M/N>0.4). However, the proposed algorithm achieves this performance by using less measurements. Clearly, the above results demonstrate the ability of the proposed algorithm to provide accurate reconstruction. Furthermore, we can observe that the proposed algorithm needs less measurements to achieve the same level of reconstruction, compared to the others



(a) Noiseless case



(b) Noisy case

Fig. 1. Comparison of the reported methods using (a) noiseless CS measurements and (b) noisy CS measurements - $\{d_g = 20,\ G = 15\}$

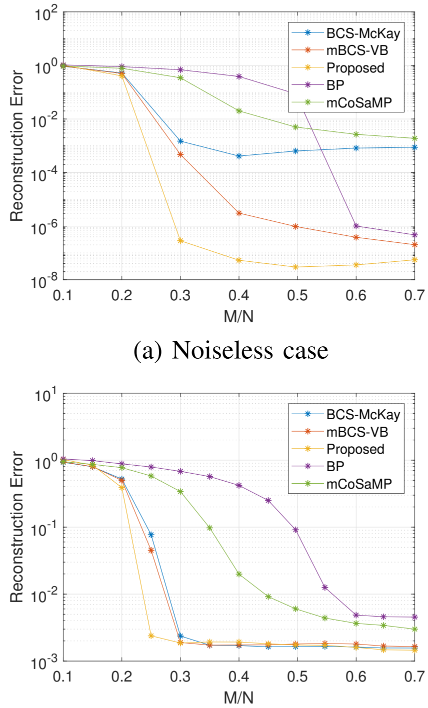
methods. Experimental results with respect to the group size ($\{d_g = 10,\ G = 30\}$,$\{d_g = 5,\ G = 60\}$) results in similar conclusions, see Figs. (2) and (3), respectively.

B. Images

In this section, we present a comparison between the proposed method and a number of existing methods on three image-guided example problems. All examples considered below are for 64×64 images. Our objective is to estimate $N(= 4096)$ scaling and wavelet coefficients. The $M \times N$ $\mathbf{\Phi}$ matrix is constructed in a similar way reported in the previous section. For each algorithm we produce a curve of reconstruction error as a function of the number of CS measurements. We examine three cases which include images with different properties. The images are shown in Fig. 4. In all cases, the grouping structure has been constructed by adopting the tree-like structure reported in [8]. More specifically, wavelet coefficients having the same parent create a group. Also, if a coefficient does not have a parent then it creates its own group. In this series of experiments we exclude the mCoSaMP algorithm (this algorithm works strictly for block sparse signals knowing a priori the number of active blocks) and include the TS-BCS-VB algorithm which is designed specifically for this kind of problems.

Fig. 5 shows the reconstruction results for the three images. In all cases, the decomposition level of wavelet transform was set to 6. Furthermore, for the analysis of the Mondrian image, we used the "symmelet8" wavelet, for the

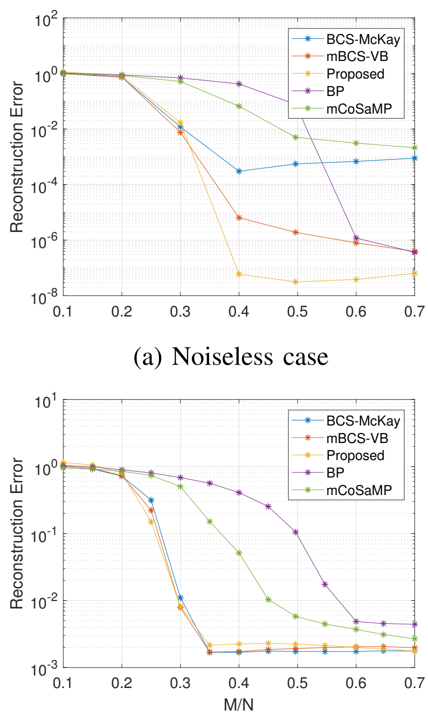
[1]https://sparselab.stanford.edu/
[2]http://dsp.rice.edu/software/
[3]http://dbabacan.info/software.html
[4]http://people.ee.duke.edu/ lcarin/BCS.html



(a) Noiseless case

(b) Noisy case

Fig. 2. Comparison of the reported methods using (a) noiseless CS measurements and (b) noisy CS measurements - $\{d_g = 10, G = 30\}$



(a) Noiseless case

(b) Noisy case

Fig. 3. Comparison of the reported methods using (a) noiseless CS measurements and (b) noisy CS measurements - $\{d_g = 5, G = 60\}$

MRI image, we used the Haar wavelet and for the Peppers image, we used the 'db4' wavelet. We can see that the proposed method clearly outperforms all methods in all cases besides the TS-BCS-VB. By comparing the proposed method with the TS-BCS-VB we can see that when we have very small number of measurements (M/N=0.1) the TS-BCS-VB provides slightly better reconstruction ability, however, as the number of measurements is increased (M/N>0.1), the situation is reversed, and the proposed method provides better performance. It is worth to point out that in the case of Peppers image, when we have adequate number of measurements (M/N≥0.4), the BCS-McKay and mBCS-VB have better performance than the TS-BCS-VB and similar performance to our algorithm. Form the above results it is evident that the proposed algorithm provides more stable behaviour compared to other reconstruction algorithms since, having sufficient number of measurements (M/N>0.1), presents better performance in a number of images with different properties ranging from medical images to paintings to natural images. Finally, in Fig. 6, we provide comparisons on the average computational time of each method. As a test image in this case we have used the Peppers image. We can see that the proposed method need less time than BCS-McKay, and similar to mBCS-VB. The method that is the most quickly is TS-BCS-VB.

IV. Conclusions

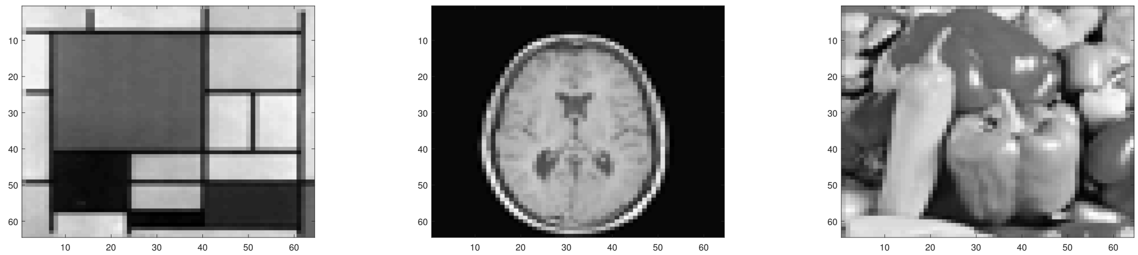
In this work, we proposed a new algorithm for CS reconstruction when an underlying group-sparsity structure exists on the signal of interest. Also, a comparison of our algorithm with well - known reconstruction algorithms is provided. Experimental results using synthetic 1D signals and 2D real data (i.e.images) have shown the usefulness of our algorithm. More specifically, in the case of 1D synthetic signals, where the various experimental settings can be controlled efficiently, the proposed algorithm has shown superior reconstruction ability. Furthermore, our algorithm needs less measurements to achieve his optimal performance. Additional experiments with images, have shown that the proposed algorithm provides superior performance than the mBCS-VB and BCS-McKay, and, slightly better performance than the TS-BCS-VB. Future extensions of our algorithm could include overlapping grouping structure as well as multi-CS schemes, such as those in [16], [17], and CS schemes based on multichannel recordings [18]. Also, approximation schemes of our algorithm could be devised in order to apply it into higher dimensional problems.

V. Acknowledgements

This work is part of projects MAMEM and RADAR-AD that have received funding from the European Unions Horizon 2020 research and innovation programme under grant agreements No 644780 and No 806999, respectively.

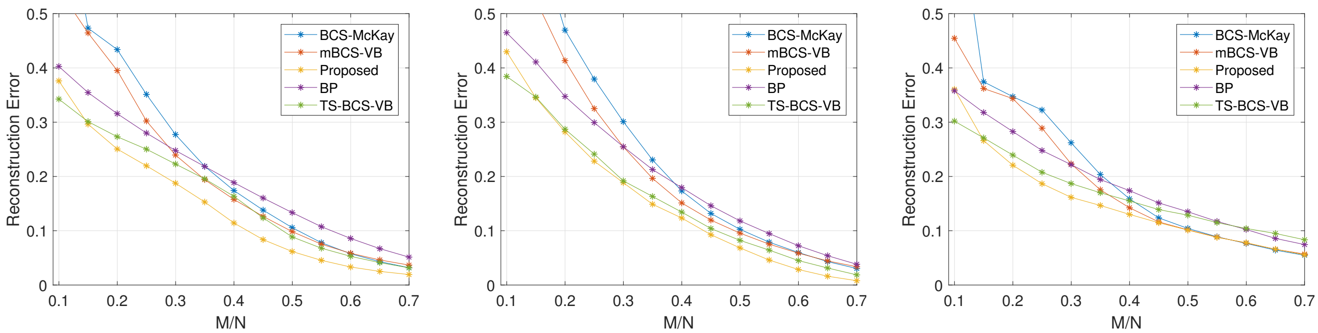
References

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, April 2006.



(a) (b) (c)

Fig. 4. Example Images: (a) Mondrian image (b) MRI image and (c) Peppers image



(a) (b) (c)

Fig. 5. Example Images: (a) Mondrian image (b) MRI image and (c) Peppers image

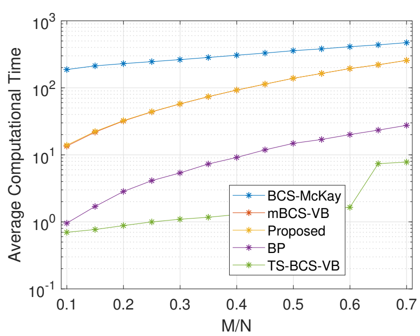


Fig. 6. Average Computational Time (in secs)

[2] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, “Introduction to compressed sensing,” in *Compressed Sensing: Theory and Applications* (Y. C. Eldar and G. Kutyniok, eds.), pp. 1–64, Cambridge University Press, 2012.

[3] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Boston: Birkhaüser, 2013.

[4] S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” *IEEE Transactions on Signal Processing*, vol. 56, pp. 2346–2356, June 2008.

[5] S. Limmer and S. Staczak, “A neural architecture for bayesian compressive sensing over the simplex via laplace techniques,” *IEEE Transactions on Signal Processing*, vol. 66, pp. 6002–6015, Nov 2018.

[6] S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Bayesian compressive sensing using laplace priors,” *IEEE Transactions on Image Processing*, vol. 19, pp. 53–63, Jan 2010.

[7] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE Transactions on Information Theory*, vol. 56, pp. 1982–2001, April 2010.

[8] L. He, H. Chen, and L. Carin, “Tree-structured compressive sensing with variational bayesian analysis,” *IEEE Signal Processing Letters*, vol. 17, pp. 233–236, March 2010.

[9] S. D. Babacan, S. Nakajima, and M. N. Do, “Bayesian group-sparse modeling and variational inference,” *IEEE Transactions on Signal Processing*, vol. 62, pp. 2906–2921, June 2014.

[10] L. Qin, J. Tan, Z. Wang, G. Wang, and X. Guo, “Exploiting the tree-structured compressive sensing of wavelet coefficients via block sparse bayesian learning,” *Electronics Letters*, vol. 54, no. 16, pp. 975–976, 2018.

[11] M. E. Tipping, “Sparse Bayesian Learning and the Relevance Vector Machine,” *Journal of Mach. Learn. Research*, vol. 1, pp. 211–244, 2001.

[12] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[13] V. P. Oikonomou, S. Nikolopoulos, P. Petrantonakis, and I. Kompatsiaris, “Sparse kernel machines for motor imagery eeg classification,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 207–210, July 2018.

[14] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, October 2007.

[15] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[16] S. Ji, D. Dunson, and L. Carin, “Multitask compressive sensing,” *IEEE Transactions on Signal Processing*, vol. 57, pp. 92–106, Jan 2009.

[17] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, “Multi-task bayesian compressive sensing exploiting intra-task dependency,” *IEEE Signal Processing Letters*, vol. 22, pp. 430–434, April 2015.

[18] Z. Zhang, T. Jung, S. Makeig, Z. Pi, and B. D. Rao, “Spatiotemporal sparse bayesian learning with applications to compressed sensing of multichannel physiological signals,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, pp. 1186–1197, Nov 2014.