



Some Comments on Variational Bayes Block Sparse Modeling with Correlated Entries

Shruti Sharma^(✉) , Santanu Chaudhury^(✉), and Jayadeva^(✉)

Indian Institute of Technology Delhi, New Delhi, India
shruti_sml@yahoo.com, {santanuc, jayadeva}@ee.iitd.ac.in

Abstract. We present some details of Bayesian block sparse modeling using hierarchical prior having deterministic and random parameters when entries within the blocks are correlated. In particular, the effect of the threshold to prune out variance parameters of algorithms corresponding to several choices of marginals, viz. multivariate Jeffery prior, multivariate Laplace distribution and multivariate Student's t distribution, is discussed. We also provide details of experiments with Electroencephalograph (EEG) data which shed some light on the possible applicability of the proposed Sparse Variational Bayes framework.

Keywords: Bayesian block sparse modeling · Gaussian Mixture Model (GSM) · Sparse Variational Bayes (SVB) · Electroencephalograph (EEG) data

1 Introduction to Sparse Variational Bayes Framework

Compressed Sensing problem aims at solving an undetermined system of linear equations:

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{v} \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^{m \times 1}$ is the observation vector, $\mathbf{x} \in \mathbb{R}^{n \times 1}$ is the unknown solution vector with $n \gg m$, \mathbf{v} is the unknown noise vector and $\Phi \in \mathbb{R}^{m \times n}$ is the known *random* matrix with full row rank and satisfies *Restricted Isometry Property*. Infinitely many \mathbf{x} can solve (1) provided solution exists and thus we need to make some assumptions to make the problem well defined [1]. Sparsity is one of the viable assumption which has received a lot of attention in the recent times. In addition to sparsity, sometimes signals exhibit additional structures in the form of blocks and we have block linear sparse model [2]:

$$\mathbf{y} = \sum_{i=1}^g \Phi^i \mathbf{x}_i + \mathbf{v} \quad (2)$$

Supported by grant NPPE-II (RP02874) as per letter no. DEITY/R&D/TDC/13(2)/2013 of Ministry of Communications and Information Technology, India.

where $\Phi^i \in \mathbb{R}^{m \times d_i}$, $\mathbf{x}_i \in \mathbb{R}^{d_i \times 1}$ and $\sum_{i=1}^g d_i = n$, g being the number of non-zero blocks and d_i being the size of i th block.

Generalized Sparse Variational Bayes (cSVB) framework is a three level hierarchical estimation framework [3] which is extension of the work proposed in [4, 5] for block sparse signals with correlated entries. At first level, it assigns heavy tailed sparsity promoting priors (which can also be expressed as Gaussian Scale Mixtures with appropriate mixing density [6]) over each block:

$$\mathbf{x}_i = \frac{1}{\sqrt{\alpha_i}} \mathbf{C}_i \mathbf{g} \quad \forall i = 1, \dots, g \quad (3)$$

where $\mathbf{g} \sim \mathcal{N}(\mathbf{0}_{d_i}, \mathbf{I}_{d_i})$, α_i is the inverse variance random parameter and $\mathbf{B}_i^{-1} \triangleq \mathbf{C}_i \mathbf{C}_i^t \in \mathbb{R}^{d_i \times d_i}$ is the covariance deterministic parameter matrix of the block \mathbf{x}_i . At second level, depending on the choice of prior distribution over parameters α_i , various heavy tailed distributions can be induced over \mathbf{x}_i viz. multivariate Laplace distribution, multivariate Student's t distribution and multivariate Jeffery's prior. At third level, we impose different priors over hyper-parameters. Graphical model representing this framework is shown in Fig. 1.

In this framework, α_i s play an important role in inducing sparsity in the solution vector. When $\alpha_i = \infty$, the corresponding i th block of \mathbf{x} becomes 0. Due to the mechanism of Automatic Relevance Determination (ARD), most of the α_i tend to infinity and thus block sparsity is encouraged. However, in the presence of noise, α_i never becomes ∞ and thus a threshold is used to prune out large α_i . This work aims at addressing the effect of threshold to prune out α_i parameters (Sect. 2) in terms of mean square error, failure rate and speed of the algorithms proposed¹ in our work [3]. For notations and other details, please refer [3]. We also demonstrate the utility of the framework for EEG data reconstruction problem [7] and Steady-State Visual Evoked Potential EEG recognition problem [8, 9].

2 Effect of Threshold to Prune Out Variance Parameters

We randomly generated the unknown solution vector \mathbf{x} of length $n = 480$ with total non-zero coefficients being 24, occurring in blocks at random locations. Coefficients within each blocks were generated as AR(1) process with common AR coefficient ρ . $m = 50$ was kept fixed and block size was varied from 1 to 6. $\Phi \in \mathbb{R}^{m \times n}$ consisted of columns drawn from a standard Gaussian distribution with unit ℓ_2 norm. Zero mean \mathbf{v} was added to measurements $\mathbf{y} = \Phi \mathbf{x} + \mathbf{v}$ with variance depending on the desired SNR. For analysis of algorithms, we carried out simple experiments over synthetic data of 200 independent trials with different realizations of measurement matrix Φ and true signal \mathbf{x} . Correlation coefficient ρ was kept 0.8. We investigated the effect of threshold value to prune out α_i and considered threshold values: 10, 50, 100, 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 .

¹ The codes for [3] can be found at <https://github.com/shruti51/cSVB>.

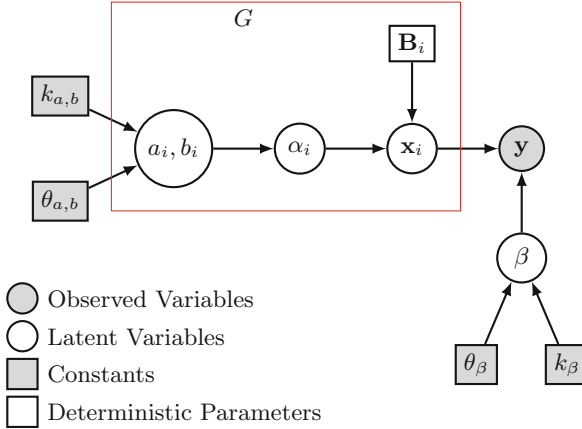


Fig. 1. Graphical Model representing the Bayesian Model. Red plate (the box labeled G) represents G nodes of which only a single node (\mathbf{x}_i and related variables) is shown explicitly (Color figure online)

We measured the algorithm’s performance in terms of failure rate (please refer [3] for definition of failure rate), MSE and speed.

From Figs. 2, 3 and 4, we see that α -pruning threshold plays an important role in determining the performance of the algorithms. Figure 2 shows that while optimal performance, in terms of failure rate, of BSBL variants and SVB variants depends on the threshold, cSVB variants do not depend much on α -pruning threshold. This is desirable in the sense that we don’t want our algorithms to depend much on the parameters of framework. It also shows that cSVB variants have outperformed SVB variants and BSBL-BO. Figure 3 shows that SVB variants have again performed poorly but now BSBL-BO performance is comparable to that of cSVB variants. Finally, we see from Fig. 4 that good performance of cSVB variants has come at a price of their computational complexity where time taken by cSVB variants is high as compared to BSBL-BO. SVB variants offer low complex algorithms as compared to cSVB and BSBL-BO which do not involve extra computational burden of inversion of matrix \mathbf{B} and thus attributing to their fast execution speed at low threshold values.

To summarize, we say that cSVB variants have a potential to recover block sparse signals with high fidelity irrespective of the α_i -pruning threshold. But this comes at a cost of high computational time.

3 Experiments with EEG Data

3.1 Reconstruction Performance of Algorithms with EEG Signals

We have used eeglab_data.set from EEGLAB which has 32 channels. Dataset and related MATLAB codes were downloaded from [10]. Each channel consists

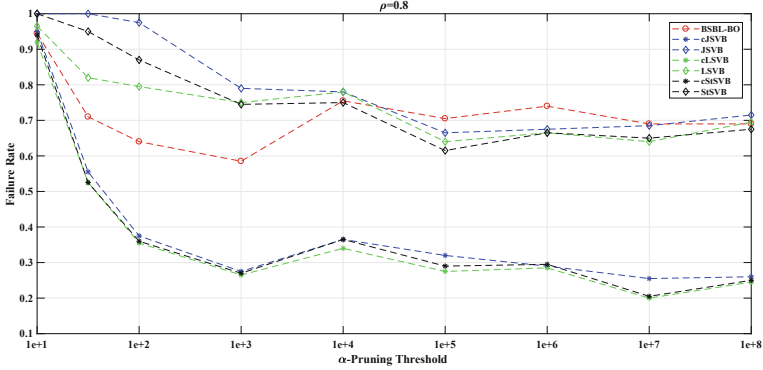


Fig. 2. Failure rate versus α -pruning threshold

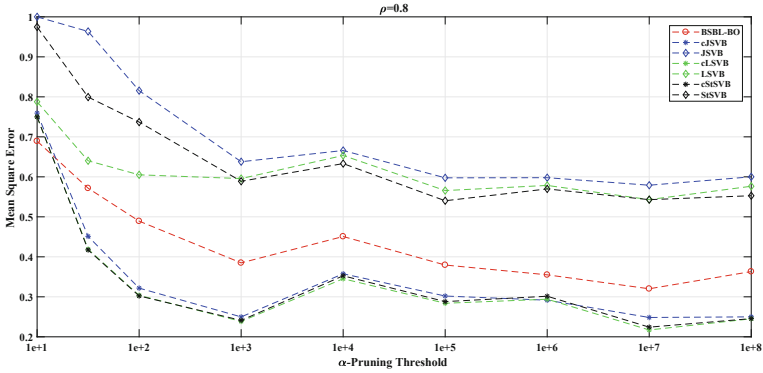


Fig. 3. Mean square error versus α -pruning threshold

of 80 epochs with 384 samples in every channel and epoch was processed independently. The data matrix was firstly transformed using Discrete Cosine Transform (DCT) and sensing matrix Φ was considered to be binary matrix of dimensions 150×384 , each column of which contained 10 ones and rest zeros [7]. This model can be written as:

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \mathbf{D} \mathbf{z} \quad (4)$$

where \mathbf{y} are compressed measurements, \mathbf{x} are original measurements and $\mathbf{z} = \mathbf{D}^{-1} \mathbf{x}$ are DCT coefficients and have few significant entries due to 'energy compaction' property of the transform. Block partitioning was kept equal and block size 24.

The reconstruction performance of all the algorithms is shown in Fig. 5. Due to our inability to interpret EEG signals, it is very difficult to assess the quality of EEG reconstruction by the proposed algorithm. However, it can be seen that at least all the algorithms have managed to capture the trends of original EEG signal. So, in this case, experiments suggest that EEG data does not exhibit

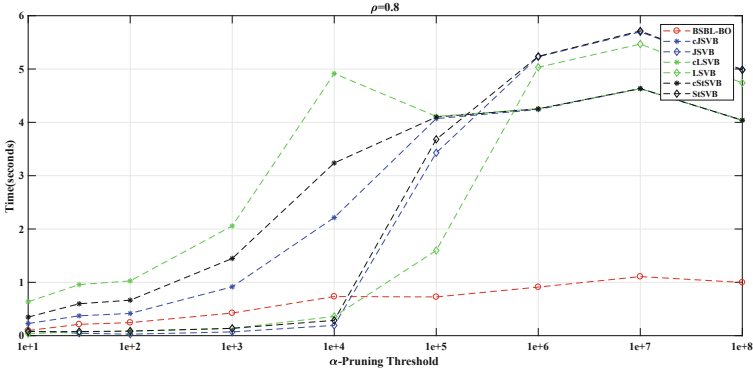


Fig. 4. Time (in seconds) versus α -pruning threshold

strong correlation which is otherwise also true in the sense that EEG data is highly non-stationary data. So, SVB variants can be seen as equally strong candidates for the analysis which do not model any correlation structure of the signal.

3.2 Experimental Results on SSVEP-Recognition

Main aim of this experiment is to demonstrate the power of Sparse Variational Bayesian framework in recognizing Steady-State Visual Evoked Potential (SSVEP).

The benchmark dataset in [8] based on SSVEP-based Brain Computer Interface (BCI) is used for the validation of algorithms. It consists of 64-channel EEG data from 35 healthy subjects (8 experienced and 27 naive) and 40 stimulation frequencies ranging from 8 to 15.8 Hz with an interval of 0.2 Hz. For each subject, the experiment was performed in 6 blocks and each block consisted of 40 trials corresponding to 40 characters (26 English alphabets, 10 digits and 4 other symbols) indicated in random order. Each trial started with a visual cue indicating a target stimulus which appeared for 0.5 s on the screen and then all stimuli started to flicker on the screen concurrently and lasted for 5 s. The screen was kept blank for 0.5 s before the next trial began.

We used the same experimental setup as proposed in [11]. Measurement matrix $\Phi \in \mathbb{R}^{m \times n}$ was sparse binary matrix having each column with two entries of 1 in random locations while rest of the entries are 0. n was kept fixed and m was varied to meet desired Compression Ratio (CR) defined: $CR = \frac{n-m}{n} \times 100$.

For performance evaluation, we used *task-specific* performance evaluation where all the algorithms were evaluated based on their performances on frequency detection of SSVEPs using Canonical Correlation Analysis (CCA) [9]. In particular, at first, SSVEP detection was performed on the original dataset (which also serves as the baseline for algorithms) and then the same task was performed on the recovered dataset from few measurements using the algorithms.

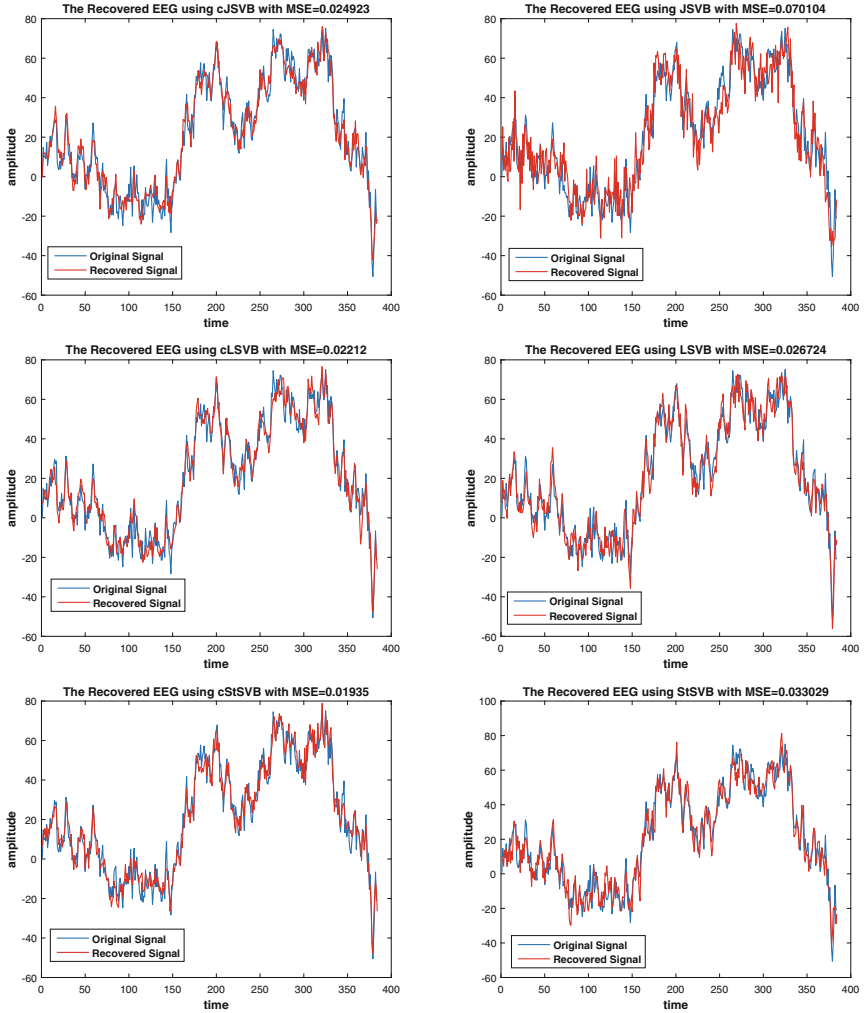


Fig. 5. Performance of Algorithms for EEG Reconstruction using 150 random measurements

For analysis, nine electrodes over the parietal and occipital areas (Pz, PO5, POz, PO4, PO6, O1, Oz and O2) were used. Number of harmonics for reference reconstruction was kept 3.

From Fig. 6, it is clear that cLSVB has outperformed in the experiment. Therefore, it can be seen that for CCA, around 40% (which corresponds to $CR=60$) of the randomly sampled points were sufficient to correctly detect almost 90% (peak) of the letters for cLSVB based recovered EEG signals. For the sake of brevity, we present the result for Subject 2 but similar results were obtained for all the subjects. For more details of this work, please refer [12].

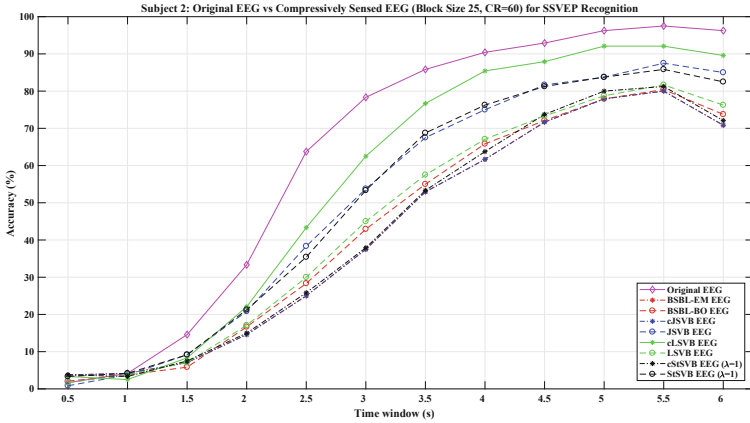


Fig. 6. Classification Rate for Subject 2 using Canonical Correlation Analysis (CCA) of all Algorithms when CR = 60

4 Conclusion

Sparse Variational Bayesian framework offers an alternate to handle block sparse recovery problem. In this paper, we analyzed one of the crucial parameters α_i which ultimately controls the structure of block sparse signals. We also discussed application of the framework in EEG signal processing context. To encourage reproducible research, the codes for [3] can be found at <https://github.com/shruti51/cSVB>.

References

1. Foucart, S., Rauhut, H.: Mathematical Introduction to CS: Applied and Numerical Harmonic Analysis. Springer, New York (2013). <https://doi.org/10.1007/978-0-8176-4948-7>
2. Eldar, Y.C., Kuppinger, P., Bolcskei, H.: Block-sparse signals: uncertainty relations and efficient recovery. *IEEE Trans. Sig. Proc.* **58**(6), 3042–3054 (2010)
3. Sharma, S., Chaudhury, S., Jayadeva: Variational Bayes block sparse modeling with correlated entries. In: 2018 24th International Conference on Pattern Recognition, pp. 1313–1318. ICPR, Beijing (2018)
4. Babacan, S.D., Nakajima, S., Do, M.N.: Bayesian group-sparse modeling and variational inference. *IEEE Trans. Signal Proc.* **62**(11), 2906–2921 (2014)
5. Zhang, Z., Rao, B.D.: Extension of SBL Algo. for the recovery of block sparse signals with intra-block correlation. *IEEE Trans. Sig. Proc.* **1**(8), 2009–2015 (2013)
6. Palamer, J.A., et al.: Variational EM algorithms for non-gaussian latent variable models. In: Proceedings of 18th International Conference on NIPS, pp. 1059–1066 (2005)
7. Zhang, Z., et al.: CS of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware. *IEEE Trans. Bio. Eng.* **60**(1), 221–224 (2013)

8. Wang, Y., et al.: A benchmark dataset for SSVEP-based brain computer interfaces. *IEEE Trans. Neural Syst. Rehab. Eng.* **25**(10), 1746–1752 (2017)
9. Lin, Z., et al.: Frequency recognition based on CCA for SSVEP-based BCIs. *IEEE Trans. Biomed. Eng.* **54**(6), 1172–1176 (2007)
10. <https://sites.google.com/site/researchbyzhang/publications>
11. Zhang, Z., et al.: Spatio-temporal sparse Bayesian learning with applications to CS of multichannel physiological signals. *IEEE Trans. Neural Syst. Rehab. Eng.* **22**(6), 1186–1197 (2014)
12. Sharma S., Chaudhury S., Jayadeva: Temporal modeling of EEG signals using block sparse variational Bayes framework. In: 11th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP), Hyderabad (2018)