

GENERALIZED MODEL BASED FRAMEWORK AND APPLICATIONS

SHRUTI SHARMA



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
OCTOBER 2021

GENERALIZED MODEL BASED FRAMEWORK AND APPLICATIONS

by

SHRUTI SHARMA

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy
to the

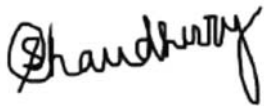


INDIAN INSTITUTE OF TECHNOLOGY DELHI
OCTOBER 2021

I want to dedicate this endeavour to my parents, my grandparents and my teachers—my mother for teaching us how to work hard, my father for teaching us how to fly, my grandparents for giving us the roots, and my teachers for giving us the direction in life. Our education has been their only worldly wealth.

THESIS CERTIFICATE

This is to certify that the thesis titled **Generalized Model Based Framework and Applications**, submitted by **Shruti Sharma**, to the Indian Institute of Technology, Delhi, for the award of the degree of **Doctor of Philosophy**, is a bonafide record of the research work done by her under our supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.



Prof. Santanu Chaudhury
Professor
Dept. of Electrical Engg.
Indian Institute of Technology
Delhi, 110016

Prof. Jayadeva
Professor
Dept. of Electrical Engg.
Indian Institute of Technology
Delhi, 110016

ACKNOWLEDGEMENTS

First and foremost, I wish to express my sincere gratitude to my supervisors, Professor Santanu Chaudhury and Professor Jayadeva, for being a great source of inspiration and guiding me with patience. I sincerely thank them for bringing me into the exciting field of model-based frameworks and encouraging me to do multi-disciplinary research. Their keen interest and constant support gave me the confidence to complete my work.

I want to take the opportunity to thank Professor Surendra Prasad for his course, ‘Compressed Sensing’, which helped me to build a basis for the study of sparse signals. I am very grateful to the members of my SRC committee. I am deeply thankful for Professor S.D. Joshi for teaching ‘Detection and Estimation Theory’, which eventually laid the subject’s strong foundation. I shall always remain indebted to him for motivating me to keep going when I wanted to give up. I wish to thank Professor Brejesh Lall for his consistent support and words of encouragement. I want to express my sincere regards to Professor Amit Kumar for his insightful suggestions. I also wish to take the opportunity to thank Professor I.N. Kar, for supporting me throughout the PhD.

I want to express my heartiest gratitude towards Mr Shailesh Kumar, a source of inspiration for me to learn Mathematics. I especially want to mention Mr Shubham and Mr Chetan for always being with me, helping me to learn Python and discuss various research problems. I also wish to thank Professor Santanu Chaudhury for providing me with an excellent working environment at IIT Jodhpur and widening my research from multiple perspectives. I want to extend my sincere thanks to all the research scholars Ms Shilpi, Ms Shefali, Ms Snigdha, Mr Ronak, Dr Deepak, and Ms Swati, for their constant encouragement. I am grateful to the non-technical staff of the multimedia lab; mainly, Mr Amit Kumar, (late) Ms Sushila Devi, Ms Usha Bhola and Ms Manrika Lochan, who consistently helped me all these years to deal with administrative works.

I am particularly indebted to Lord Shiva, my parents, brother, and sister for their eternal love and constant support during my PhD journey. The successful completion of this venture has been the reflection of their blessings.



Shruti Sharma

ABSTRACT

Real-world data involving multivariate observations exhibit skewness, correlated entries across various dimensions and non-Gaussian characteristics such as low-dimensional structures with densities that are sharply peaked and possess heavy tails. As a result, the normality assumption is violated and it is desirable to have methods to model a broader class of signals exploiting domain specific knowledge, while offering simple solutions such as available with Gaussian models.

This thesis, *Generalized Model-Based Frameworks and Applications*, attempts to address the problem of data modeling under model-based framework, when data exhibit correlation structure with heavy tails and skewness. The basic idea is to assume specific models for data generation, so that we can have specialized algorithms, that require less data for inference. In this work, we rely on the generalized scale mixture model, which comprises of various random and deterministic parameters. By varying distribution over random parameter, we derive a family of generalized multivariate heavy tailed distributions, which encompasses several choice of classical distributions such as Laplace, Student's t and Jefferrys prior. For inference, we develop a hybrid framework, where random parameters are estimated using Variational Bayes (VB) and deterministic parameters using Expectation Maximization (EM). To show the applicability of the proposed model, we present its utility in two different application areas.

Firstly, we consider the model-based compressed sensing framework associated with Bayesian linear model. Specifically, we address the Bayesian group sparse modeling with correlated entries and derive cSVB (correlated Sparse Variational Bayes) framework. We also propose two frameworks, *viz.* TSVB (Temporal Sparse Variational Bayes) and TMSVB (low complexity version of TSVB), to solve related problems on sparse signal recovery in the multiple measurement vector (MMV) context. Next, we consider its extension to spatio-temporal data and present the Spatio-Temporal Sparse Variational Bayes (STSVB) framework. The proposed algorithm processes multi-dimensional data, and is particularly helpful where we have a limited access to computational resources. We briefly discuss applications in the domains of fetal electrocardiogram (FECGs), electroencephalogram (EEG) data processing, and in the modeling of a dynamic hand dataset. We also present the utility of cSVB and STSVB frameworks for Steady-State Visual Evoked Potential (SSVEP) based EEG signals, where the task is to detect frequency components present in the data.

Next, we discuss a novel model-based clustering/classification framework using a gener-

alized class of skewed distribution. Specifically, we present an alternative to the Gaussian Mixture Model (GMM), based on the EM framework, that is capable of handling more complex structures present in data, while offering advantages of almost similar computational complexity as that of GMM. Experimental results on five real-world datasets demonstrate the effectiveness of the model.

Finally, we explore the possible future directions in the context of model-based methods.

सार

बहुभिन्नरूपी प्रेक्षणों से जुड़े वास्तविक-विश्व डेटा में विषमता, विभिन्न आयामों में सहसंबद्ध प्रविष्टियाँ और गैर-गॉसियन विशेषताओं जैसे निम्न-आयामी संरचनाएँ घनत्व के साथ प्रदर्शित होती हैं जो तेजी से चरम पर होती हैं और जिनमें भारी पूंछ होती है। नतीजतन, सामान्यता धारणा का उल्लंघन किया जाता है और यह वांछनीय है कि गाऊसी मॉडल के साथ उपलब्ध सरल समाधानों की पेशकश करते हुए डोमेन विशिष्ट ज्ञान का शोषण करने वाले संकेतों के व्यापक वर्ग को मॉडल करने के तरीके हैं।

यह थीसिस, सामान्यीकृत मॉडल-आधारित ढांचे और अनुप्रयोग, मॉडल-आधारित ढांचे के तहत डेटा मॉडलिंग की समस्या को हल करने का प्रयास करते हैं, जब डेटा भारी पूंछ और तिरछीपन के साथ सहसंबंध संरचना प्रदर्शित करता है। मूल विचार डेटा पीढ़ी के लिए विशिष्ट मॉडल ग्रहण करना है, ताकि हमारे पास विशेष एल्गोरिदम हो सकें, जिसके लिए अनुमान के लिए कम डेटा की आवश्यकता होती है। इस काम में, हम सामान्यीकृत पैमाने के मिश्रण मॉडल पर भरोसा करते हैं, जिसमें विभिन्न यादृच्छिक और नियतात्मक पैरामीटर शामिल हैं। यादृच्छिक पैरामीटर पर वितरण अलग-अलग करके, हम सामान्यीकृत बहुभिन्नरूपी भारी पूंछ वाले वितरण का एक परिवार प्राप्त करते हैं, जिसमें शास्त्रीय वितरण के कई विकल्प शामिल हैं जैसे कि लैपलेस, स्टूडेंट का टी और जेफरी का पूर्वी अनुमान के लिए, हम एक हाइब्रिड फ्रेमवर्क विकसित करते हैं, जहां वैरिएशनल बेयस (VB) और नियतात्मक मापदंडों का उपयोग करके एक्सपेक्टेड मैक्सिमाइजेशन (EM) का उपयोग करके यादृच्छिक मापदंडों का अनुमान लगाया जाता है। प्रस्तावित मॉडल की प्रयोज्यता दिखाने के लिए, हम दो अलग-अलग अनुप्रयोग क्षेत्रों में इसकी उपयोगिता प्रस्तुत करते हैं।

सबसे पहले, हम बायेसियन रैखिक मॉडल से जुड़े मॉडल-आधारित संपीड़ित संवेदन ढांचे पर विचार करते हैं। विशेष रूप से, हम सहसंबद्ध प्रविष्टियों के साथ बायेसियन समूह विरल मॉडलिंग को संबोधित करते हैं और cSVB (correlated Sparse Variational Bayes) ढांचे को प्राप्त करते हैं। हम दो रूपरेखाओं का भी प्रस्ताव करते हैं, अर्थात् TSVB (Temporal Sparse Variational Bayes) और TMSVB (TSVB का एक कम जटिलता संस्करण), मल्टीपल मेजरमेंट वेक्टर (MMV) संदर्भ में विरल सिमल रिकवरी पर संबंधित समस्याओं को हल करने के लिए। इसके बाद, हम spatiotemporal डेटा के लिए इसके विस्तार पर विचार करते हैं और Spatio-temporal Sparse Variational Bayes (STSVB) फ्रेमवर्क प्रस्तुत करते हैं। प्रस्तावित एल्गोरिदम बहु-आयामी डेटा को संसाधित करता है और विशेष रूप से सहायक होता है जहां हमारे पास कम्प्यूटेशनल संसाधनों तक सीमित पहुंच होती है। हम संक्षेप में भ्रूण इलेक्ट्रोकार्डियोग्राम (FECGs), इलेक्ट्रोएन्सेफेलोग्राम (EEG) डेटा प्रोसेसिंग के डोमेन में और एक गतिशील हाथ डेटासेट के मॉडलिंग में अनुप्रयोगों पर चर्चा करते हैं। हम की उपयोगिता भी प्रस्तुत करते हैं। स्थिर-राज्य दृश्य विकसित संभावित (SSVEP) आधारित EEG संकेतों के लिए cSVB और STSVB ढांचे, जहां कार्य डेटा में मौजूद आवृत्ति घटकों का पता लगाना है।

इसके बाद, हम विषम वितरण के एक सामान्यीकृत वर्ग का उपयोग करते हुए एक उपन्यास मॉडल-आधारित क्लस्टरिंग/वर्गीकरण ढांचे पर चर्चा करते हैं। विशेष रूप से, हम ईएम ढांचे के आधार पर गाऊसी मिश्रण मॉडल (GMM) का एक विकल्प प्रस्तुत करते हैं, जो डेटा में मौजूद अधिक जटिल संरचनाओं को संभालने में सक्षम है, जबकि जीएमएम के लगभग समान कम्प्यूटेशनल जटिलता के फायदे पेश करते हैं। पांच वास्तविक दुनिया के डेटासेट पर प्रायोगिक परिणाम मॉडल की प्रभावशीलता को प्रदर्शित करते हैं।

अंत में, हम मॉडल-आधारित विधियों के संदर्भ में भविष्य की संभावित दिशाओं का पता लगाते हैं।

Contents

ACKNOWLEDGEMENTS	ii
ABSTRACT	iii
LIST OF FIGURES	xvii
LIST OF TABLES	xix
ABBREVIATIONS	xx
NOTATION	xxiii
1 Introduction	1
1.1 Model-Based Approach	1
1.1.1 Data Models	2
1.2 Data-driven/Deep Learning methods	2
1.3 Why model-based methods over Deep Learning?	2
1.4 Thesis Scope	3
1.5 Thesis Objectives	5
1.6 Contributions and Layout of Thesis	6
1.6.1 Contributions	6
1.6.2 Layout of Thesis	6
2 Background	8
2.1 Compressed Sensing, Model and Algorithms: A review	8
2.1.1 Single Measurement Vector Model	8

2.1.2	Multiple Measurement Vector Model	11
2.2	Compressed Sensing for EEG data processing	13
2.2.1	Steady-State Visual Evoked Potential (SSVEP) based Brain Computer Interface (BCI)	14
2.2.2	CS for BCI	15
2.3	Finite Mixture Modeling	16
3	Variational Bayes Block Sparse Modeling with Correlated Entries	18
3.1	Problem formulation and Bayesian Modeling	18
3.1.1	Observation Model	19
3.1.2	Signal Model	19
3.1.3	Complete Model	22
3.2	Inference	23
3.2.1	Signal Estimate	24
3.2.2	Noise Parameter Estimation	25
3.2.3	Inverse Variance Parameter Update	25
3.2.4	Estimation of hyper-parameters a_i 's and b_i 's	26
3.2.5	Estimation of deterministic parameter $\mathbb{B}$	27
3.2.6	Computational Complexity of cSVB	28
3.3	Synthetic Data Experiments	29
3.3.1	Benefits of correlation information at different group sizes	30
3.3.2	Varying the number of measurements at different correlation levels	32
3.3.3	Effect of Threshold to prune out variance parameters	36
3.4	Experiments with Fetal Electrocardiograph (FECG) Data	36
3.5	Experiments with Electroencephalograph (EEG) data	38
3.6	Experiments with Steady State Visual Evoked Potential (SSVEP) based EEG data	41
3.7	Conclusion and Future Scope	47

3.8	Publications	48
3.9	Appendix	48
3.9.1	Calculation of $\mathbb{E}[\alpha_i]$ for cStSVB	48
3.9.2	Calculation of $\mathbb{E}[\alpha_i]$ and $\mathbb{E}[\alpha_i^{-1}]$ for cLSVB	49
4	Sparse Signal Recovery for Multiple Measurement Vectors with Temporally Correlated Entries	50
4.1	Problem formulation and Bayesian Modeling	50
4.1.1	Relation to Block Sparse Modeling	52
4.1.2	Observation Model	52
4.1.3	Signal Model	53
4.2	Inference	54
4.2.1	Signal Estimate	54
4.2.2	Noise Parameter Estimation	55
4.2.3	Inverse Variance Parameter Update	55
4.2.4	Estimation of hyper-parameters	56
4.2.5	Estimation of deterministic parameter $\mathbf{B}$	56
4.2.6	Computational Complexity of TSVB	56
4.3	Temporal Multiple Sparse Variational Bayes	57
4.3.1	Signal Estimate	58
4.3.2	Noise Parameter Estimate	58
4.3.3	Inverse Variance Parameter Estimates	58
4.3.4	Deterministic parameter estimates	59
4.3.5	Computational Complexity of TMSVB	60
4.4	Experimental Results	60
4.4.1	Benefit at Different Temporal Correlation Levels in MMV	61
4.4.2	Quality of Reconstruction	62

4.4.3	Ability to Handle Highly Underdetermined Problem	63
4.4.4	Varying number of sources K at different correlation levels	63
4.4.5	Application to Dynamic Hand Posture Dataset	64
4.5	Conclusion and Future Scope	66
4.6	Publications	66
5	Block Sparse Variational Bayes Regression using Matrix Variate Distributions	67
5.1	Problem Formulation and Bayesian modeling	67
5.1.1	Hierarchical Signal Model	68
5.1.2	Observation Model	71
5.2	Comparative Analysis	73
5.2.1	Comparison to BSBL, cSVB, TSBL, TSVB, STSBL	73
5.2.2	Comparison to Kronecker Product Principal Component Analysis (PCA)	73
5.2.3	Comparison to Regression methods using Matrix Distributions . . .	74
5.2.4	Low Rank Inducing Prior on $\tilde{\mathbf{X}}_{[i]}$	74
5.3	Inference	76
5.3.1	Learning in Spatially Whitened Model	76
5.3.2	Learning in Temporally Whitened Model	79
5.3.3	Computational Complexity of STSVB	80
5.3.4	$\{\Pi_i \triangleq \alpha_i \Psi_i\}_{i=1}^G$ as random parameters and $\mathbf{B}$ as deterministic parameter	80
5.3.5	Variational Update of Π_i for low-rank inducing prior over $\mathbf{X}_{[i]}$	81
5.4	Experimental Results with SSVEP based EEG data	82
5.4.1	Performance Evaluation Measure	86
5.5	Discussions	89
5.5.1	Time Complexity	89

5.5.2	Convergence Analysis	93
5.5.3	Effect of Harmonics and Filter Bank size	93
5.5.4	Performance across all Subjects	94
5.5.5	Effectiveness of STSVB-L in SSVEP detection task	94
5.6	Conclusion and Future Scope	95
5.7	Publications	96
6	Model based Clustering Framework using a Generalized Class of Skewed Distributions	97
6.1	Normal Variance Mean Mixture (NVMM)	97
6.2	Application to Finite Mixture Models	104
6.2.1	Expectation Maximization for Parameter Estimation	105
6.3	Discussion	107
6.3.1	Dealing with Infinite Log-Likelihood Values	107
6.3.2	Comparison to the Classical GMM	107
6.3.3	Computational Complexity	108
6.3.4	Extension to the fully Bayesian model	108
6.4	Experimental Results	109
6.4.1	Old Faithful Geyser dataset	110
6.4.2	Wind Data	113
6.4.3	Yeast Dataset	113
6.4.4	Australian Institute of Sports (AIS) Data	114
6.4.5	Seeds Dataset	114
6.5	Conclusion and Future Scope	117
6.6	Publications	119
6.7	Appendix	119
6.7.1	Marginal Density Calculation for fully Bayesian Case	119

6.7.2	Details of Deterministic Parameter Updates	120
6.7.3	Extension to Dirichlet Process Mixture Modeling	121
7	Finite Mixture Modeling using Matrix Variate Distributions	125
7.1	Matrix Variate Normal Variance Mean Mixture Model	125
7.2	Finite Mixture Modeling	129
7.2.1	Expectation Maximization for Parameter Estimation	130
7.3	Conclusion and Future Scope	132
7.4	Publications	132
7.5	Appendix	133
7.5.1	Details of deterministic parameter updates	133
8	Conclusions and Future Work	134
8.1	Summary of Contributions	134
8.2	Future Work	135
8.3	Appendix	136
8.3.1	Hierarchical model over latent space	136
8.3.2	Experimental Results	146
A	APPENDIX	147
A.1	Modified Bessel Function of second kind	147
A.2	Distributions	147
A.3	Kronecker Product And Its Properties	150
A.3.1	Properties	150
A.4	Matrix Derivatives	151
A.5	Parameter Estimation	151
A.5.1	Expectation Maximization for Maximum likelihood estimation . . .	151
A.5.2	Variational Bayesian Inference	153

A.6	SSVEP Training Methods	155
A.6.1	Canonical Correlation Analysis (CCA)	155
A.6.2	L1-Regularized Multiway Canonical Correlation Analysis (L1-MCCA)	155
A.6.3	Filter-Bank Canonical Correlation Analysis (FBCCA)	156
A.6.4	Task Related Component Analysis (TRCA)	156
A.7	Multivariate Analysis of Variance (MANOVA)	157
A.7.1	Descriptive Discriminant Analysis (DDA)	161
A.7.2	Graphing the Groups in the Discriminant Plane	161
A.8	Variational Autoencoders (VAEs)	162
	Bibliography	162
	Publications	174
	Biography	175

List of Figures

2.1	Single Measurement Vector (SMV) based Compressed Sensing (CS) model in its standard form, where the solution vector $\mathbf{x}$ exhibits sparsity and every coefficient is treated as an independent entity.	9
2.2	Block sparse modeling of the solution vector $\mathbf{x}$, where $\mathbf{x}$ is viewed as a concatenation of blocks and out of G blocks only few are non-zero.	10
2.3	Multiple Measurement Vector based Compressed Sensing model with the Common Sparsity Assumption accross time. In this formulation, the solution matrix $\mathbf{X}$ exhibit row sparsity and out of N rows only few are active. .	12
2.4	Spatio-temporal block sparse modeling of the solution matrix $\mathbf{X}$ in Eq. (4.1), where $\mathbf{X}$ is viewed as a concatenation of spatio-temporal sub-blocks, exhibiting spatio-temporal characteristics, and out of G blocks only few are non-zero.	13
2.5	(a) A system diagram of SSVEP based BCI 40-target speller, (b) Frequency and phase values used for encoding each character. The frequencies range from 8.0 to 15.8 Hz with an interval of 0.2 Hz. (Figure taken from [1]) . . .	15
3.1	Graphical Model representing the proposed 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.	22
3.2	Comparison of the performance of all algorithms in terms of the failure rate at different level of correlated entries in blocks when the block size was varied from 1 to 6 in noiseless case.	31
3.3	Comparison of the performance of all algorithms in terms of the failure rate at different level of correlated entries in blocks when the block size was varied from 1 to 6 with $SNR = 25dB$	33
3.4	Comparison of the relative reconstruction error corresponding to the proposed block sparse modeling frameworks based recovered signals with block size = 4 and $SNR = 50dB$	34
3.5	Comparison of the performance in terms of the failure rate for the proposed block sparse modeling frameworks with block size = 3 and $SNR = 25dB$. .	35
3.6	Comparison of the performance in terms of the failure rate for the proposed algorithms with respect to α -pruning threshold, which is varied from 10 to 10^8	37

3.7	Comparison of the performance in terms of the mean square error for the proposed algorithms with respect to α -pruning threshold, which is varied from 10 to 10^8	37
3.8	Comparison of the time taken (in seconds) by the proposed algorithms with respect to α -pruning threshold, which is varied from 10 to 10^8	38
3.9	Qualitative comparison of the reconstruction performance of algorithms for telemonitoring of FECG data with the number of compressively sensed measurements to be 200 and block size = 16.	39
3.10	Qualitative comparison of the ICA decomposition of compressively sensed measurements based recovered data from the algorithms.	40
3.11	Qualitative comparison of the reconstruction performance of algorithms with different blocks sizes from compressively sensed observations, when the number of measurements are 150.	42
3.12	Qualitative comparison of reconstruction performance of algorithms with different block sizes from compressively sensed observations, when the number of measurements are 100.	43
3.13	Comparison of the frequency detection accuracy for Subject 27 using Canonical Correlation Analysis (CCA) of all algorithms, when CR is varied from 40 to 90.	45
3.14	Comparison of the frequency detection accuracy for Subject 27 using L1-Multiway Canonical Correlation Analysis (L1MCCA) of all algorithms, when CR varied from 40 to 90.	46
4.1	Multiple measurement vector (MMV) model to single measurement vector (SMV) model transformation.	51
4.2	Comparison of the failure rate based performance of the algorithms at different correlation levels in multiple measurement vector model, when $SNR = 25dB$	61
4.3	Comparison of the failure rate based performance of the algorithms at different correlation levels in multiple measurement vector model, when $SNR = 25dB$ and the row sparsity index K was varied from 10 to 16.	61
4.4	Comparison of the quality of reconstructed rows with respected to time of all the algorithms for a typical successful trial, when $SNR = 25dB$, $L = 3$ and $\rho = 0.5$	62
4.5	Comparison of all algorithms in terms of the failure rate with $\rho = 0.8$ and $SNR = 25dB$ by varying N/M ratio from 1 to 30 to demonstrate the ability of the proposed framework to handle highly underdetermined problems. . .	64

4.6	Qualitative comparison of the reconstruction performance of all algorithms corresponding to the number of compressively sensed measurements to be 750, with Dynamic Hand Gesture Dataset [2].	65
5.1	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.	72
5.2	Descriptive Discriminant Analysis (DDA) to show the statistical significance of the classification results of CS based Bayesian recovery algorithms for Subject 27 at CR = 70 when frequency detection method is (a) CCA (b) L1MCCA (c) FBCCA (d) TRCA. The one-way MANOVA results in Table 5.3 show that the null hypothesis should be rejected. DDA is a post-hoc procedure to determine why the null hypothesis was rejected. Here, c_1 is the linear combination of the columns/groups of the accuracy matrix X (each group corresponds to the recovery algorithm used), that has the maximum separation between the groups. c_2 has the maximum separation subject to it being orthogonal to c_1 . The scatter plot of the first two canonical variables of X are chosen to maximize the separation between the groups. It can be seen that the cloud of points (encircled in red) corresponding to STSVB-L have a distinct center as compared to other Bayesian algorithms irrespective of frequency detection method used, and thus shows its performance is statistically significant. . .	88
5.3	Comparison of the reconstruction performance (in terms of MSE) of all Bayesian algorithms at CR = 70 for Subject 27. Note that though STSVB-L has outperformed in the frequency detection task; however, its reconstruction performance is inferior to that of STSVB-J, STSVB-St and BSBL variants, and converges slowly to the solution matrix.	90
5.4	Descriptive Discriminant Analysis (DDA) to show the statistical significance of the classification results of CS based STSVB-L for Subject 27 at CR = 70 when the number of harmonics H in the reference signals is varied from 1 to 6 in (a) CCA (b) FBCCA. The scatter plot in the discriminant plane (c_1 - c_2 plane) shows the cloud of points (encircled in red) with a distinct center, illustrating that the STSVB-L performance using CCA and FBCCA is significantly different for $H = 2$ or 3 as compared to the other values of H . It should also be noted that the performance is insignificant for $H = 2$ and 3 and in this work we have used $H = 2$ for analysis purpose.	91
5.5	Descriptive Discriminant Analysis (DDA) to show the statistical significance of the classification results of CS based STSVB-L for Subject 27 at CR = 70 when the number of sub-bands in filter bank Fb is varied from 2 to 7 in (a) FBCCA (b) TRCA. The scatter plot in the discriminant plane for FBCCA shows that the cloud of points (encircled in red) corresponding to different Fb form clusters with the respective distinct centers, illustrating that the STSVB-L performance is significantly different for all Fbs. However, for TRCA, the scatter plot appears to be random and thus depicts that the result is statistically insignificant.	92

5.6	Comparison of the Mean Frequency Detection Accuracy of CS based Bayesian recovery algorithms at CR = 70 using CCA. This result, averaged across all 35 subjects, is the mean of 25 independent experiments with different realizations of Φ and the vertical bars represent the variance across the experiments.	95
6.1	Contour Plots of Normal Inverse Gaussian (NIG) density function Eq. (6.7) for different values of a and b (a) $a = 1, b = 1$, (b) $a = 1, b = 0.01$, (c) $a = 0.01, b = 1$ and (d) $a = 10, b = 10$. Plots are generated by $N = 100$ random samples using Eq. (6.2), with α following inverse Gaussian distribution with parameters a and b , mean value $\mu = [1, 0]^T$ (marked with green cross), skewness parameter $\beta = [1, 1]^T$ and correlation matrix parameter $\Sigma = \text{matrix}[1, 0.5, 0.5, 1]$	99
6.2	Contour Plots of Variance Gamma (VG) density function Eq. (6.8) for different values of a and λ (a) $a = 2, \lambda = 1$, (b) $a = 2, \lambda = (D+1)/2$, (c) $a = 2, \lambda = 0.1$, and (d) $a = 0.1, \lambda = 1$. Plots are generated by $N = 100$ random samples using Eq. (6.2), with α following gamma distribution with parameters a and λ , mean value $\mu = [1, 0]^T$ (marked with green cross), skewness parameter $\beta = [1, 1]^T$ and correlation matrix parameter $\Sigma = \text{matrix}[1, 0.5, 0.5, 1]$	100
6.3	Contour Plots of Student's t density function Eq. (6.10) for different values of b and ν (a) $\nu = 1, b = 1$, (b) $\nu = 1, b = 10$, (c) $\nu = 0.5, b = 5$, and (d) $\nu = 10, b = 0.1$. Plots are generated by $N = 100$ random samples using Eq. (6.2), with α following inverse gamma distribution with parameters ν and b , mean value $\mu = [1, 0]^T$ (marked with green cross), skewness parameter $\beta = [1, 1]^T$ and correlation matrix parameter $\Sigma = \text{matrix}[1, 0.5, 0.5, 1]$	101
6.4	Contour plots of Student's t density function Eq. (6.10) for $\nu = 2, b = 10$. Plot is generated by $N = 500$ random samples using Eq. (6.2) with $\mu = [1, 0]^T$ (marked with green cross), $\beta = [1, 1]^T$ and $\Sigma = \text{matrix}[1, 0.5, 0.5, 1]$	103
6.5	The graphical Model representing the proposed mixture model. Red plate (the box labeled N) represents N nodes of which only a single node ($\mathbf{x}_n, \mathbf{z}_n$ and α_n) is shown explicitly. Similarly, green plate represents the k th mixing node and associated parameters.	105
6.6	Model-based clustering results on Old-faithful dataset using (a) MSNIG-EM algorithm with $a = 1, b = 1$, (b) MSSSt-EM algorithm with $\nu = 1, b = 1$, (c) SAL (MSVG-EM with $\lambda = 1, a = 2$), and (d) MSVG-EM with $\lambda = 10, a = 2$	111
6.7	Model-based clustering Results on Wind dataset using (a) MSNIG-EM with $a = 1, b = 1$, (b) MSSSt-EM with $\nu = 10, b = 0.1$, (c) SAL [3] (MSVG-EM with $\lambda = 1, a = 2$), and (d) MSVG-EM with $\lambda = \frac{D+1}{2}, a = 2$	112
6.8	Model-based clustering Results on AIS dataset using (a) MSNIG-EM with $a = 10, b = 10$, (b) MSSSt-EM with $\nu = 10, b = 10$, (c) SAL (MSVG-EM with $\lambda = 1, a = 2$), and (d) MSVG-EM with $\lambda = 0.1, a = 2$. $\circ$ denotes data points corresponding to Female athletes and $\triangle$ for Male athletes.	115

6.9	Contour plots of the fitted multivariate skewed Student's t distribution on AIS dataset corresponding to female atheletes for $b = 1$ and $\nu = 1$, with parameters being estimated by MSSSt-EM using $K = 1$, (a) without outliers, (b) with outliers (marked in blue).	116
6.10	Model-based Clustering Results on Seeds dataset using (a) NIG based model with $a = 1, b = 1$, (b) Student's t based model with $\nu = 1, b = 1$, (c) SAL [3] (MSVG-EM with $\lambda = 1, a = 2$), and (d) VG based model with $\lambda = \frac{D+1}{2}, a = 2$.	118
6.11	The graphical Model representing the proposed mixture model. Red plate (the box labeled N) represents N 'observation' nodes of which only a single node ($\mathbf{x}_n, \mathbf{z}_n$ and α_n) is shown explicitly. Similarly, green plate represents the k th mixing node and associated parameters.	124
8.1	Density function plots corresponding to (a) Laplace density $Q_L(z X)$, (b) $Q(z X)$ (c) Gaussian density.	138
8.2	Contour Plots where red lines correspond to the contour plots of $Q_L(z X)$, green lines correspond to the contour plots of $Q(z X)$ and black lines correspond to the contour plots of $\mathcal{N}(\mu, \Sigma^2 I)$.	140
8.3	Images generated when the network was trained using 20,000 random images of MNIST dataset and the latent space dimension was kept to be 20 using (a) the proposed Laplace based VAE, and (b) the traditional VAE.	142
8.4	Images generated when the network was trained using 20,000 random images of Fashion MNIST dataset and the latent space dimension was kept to be 20 using (a) the proposed Laplace based VAE, and (b) the traditional VAE.	143
8.5	Images generated when the network was trained using 20,000 random images of MNIST dataset and the latent space variable $\mathbf{z}$ was sampled uniformly from the Gaussian percentiles in 2D (a) the proposed Laplace based VAE, and (b) the traditional VAE.	144
8.6	Images generated when the network was trained using 20,000 random images of Fashion MNIST dataset and the latent space variable $\mathbf{z}$ was sampled uniformly from the Gaussian percentiles in the 2D (a) the proposed Laplace based VAE, and (b) the traditional VAE.	145

List of Tables

3.1	Summary of the parameter estimates corresponding to the variants of cSVB	26
4.1	Summary of the parameter estimates corresponding to the variants of T-SVB and T-MSVB	57
4.2	Quantitative Performance Comparison of Algorithms on Dynamic Hand Gesture Dataset [2] in terms of PSNR (in dB) for different M	63
5.1	Summary of the parameter estimates of STSVB Variants.	78
5.2	Comparison of the Classification Performance of CS based Bayesian recovery algorithms using the frequency detection methods for Subject 27 with Compression Ratio = 80.	83
5.3	Comparison of the Classification Performance of CS based Bayesian recovery algorithms using the frequency detection methods for Subject 27 with Compression Ratio = 70.	84
5.4	Comparison of the Classification Performance of CS based Bayesian recovery algorithms using the frequency detection methods for Subject 27 with Compression Ratio = 60.	85
5.5	Time Complexity (in seconds Per Block Per Trial) when CR = 70	86
5.6	Effect of number of sub-bands of Filter Bank when CR = 70. STSVB-L + FBCCA (Wilk's λ : 0.0089, $F(50,1303.16) = 46.49$, p -value < 0.0001 , $d = 4$)	89
5.7	Effect of number of sub-bands of Filter Bank when CR = 70. STSVB-L + TRCA (Wilk's λ : 0.4140, $F(50, 1303.16) = 4.99$, p -value < 0.0001 , $d = 1$).	89
5.8	Effect of Harmonics when CR = 70. STSVB-L + CCA (Wilk's λ : 0.00631, $F(60, 1329)=17.81$, p -value < 0.0001 , $d = 2$)	89
5.9	Effect of Harmonics when CR = 70. STSVB-L + L1MCCA (Wilk's λ : 0.173, $F(60, 1329)=10.02$, p -value < 0.0001 , $d = 1$)	90
5.10	Effect of Harmonics when CR = 70. STSVB-L+FBCCA (Wilk's λ value: 0.0229, $F(60, 1329)=27.46$, p -value < 0.0001 , $d = 2$)	90
6.1	BIC values for Old-Faithful dataset for different values of K	111

6.2	Classification performance of the proposed model on Yeast dataset.	113
6.3	Quantitative Comparison of the Clustering Performance of the Proposed Model on Australia Institute of Sports (AIS) dataset.	116
6.4	Quantitative Comparison of the Clustering Performance of the Proposed Model on Seeds dataset.	117
6.5	The Cross-tabulation of True and Predicted Class Memberships of Seeds dataset.	117

ABBREVIATIONS

AIS	Australian Institute of Sports
AR	Autoregressive
ARD	Automatic Relevance Determination
ARI	Adjusted Random Index
BCI	Brain Computer Interface
BIC	Bayesian Information Criterion
BSBL	Block Sparse Bayesian Learning (Block SBL)
CAR	Conditional Autoregressive
CCA	Canonical Correlation Analysis
CLT	Central Limit Theorem
CR	Compression Ratio
CS	Compressed Sensing
cSVB	correlated Sparse Variational Bayes
cJSVB	cSVB based on Jeffreys prior
cLSVB	cSVB based on Laplace distribution
cStSVB	cSVB based on Student distribution
DCT	Discrete Cosine Transform
DDA	Descriptive Discriminant Analysis
DOA	Direction of Arrival
EEG	Electroencephalogram
EPGMM	Expanded Parsimonious Gaussian mixture model
EM	Expectation Maximization
FBCCA	Filter Bank Canonical Correlation Analysis
FECG	Fetal electrocardiogram
FM	Finite Mixture Model
FM-HMT	FM with Hierarchical Multivariate Student
FM-MGH	FM with Multivariate Generalized Hyperbolic
FM-MN	FM with Multivariate Normal
FM-MT	FM with Multivariate Student
FM-MNIG	FM with Multivariate Normal Inverse Gaussian
FM-rMSN	FM with restricted Multivariate Skewed Normal
FM-rMST	FM with restricted Multivariate Skewed Student
FM-rMSTN	FM with restricted Multivariate Skewed Student Normal
FM-uMST	FM with un-restricted Multivariate Skewed Student

FOCUSS	FOCal Underdetermined System Solution
GH	Generalized Hyperbolic
GIG	Generalized Inverse Gaussian
GMM	Gaussian Mixture Model
GSM	Gaussian Scale Mixture
ICA	Independent Component Analysis
L1MCCA	L1-regularized Multiway CCA
MANOVA	Multivariate Analysis of Variance
MCGFA	Mixture of Contaminated Gaussian Factor Analyzers
MCMC	Markov Chain Monte Carlo
MCR	Misclassification Rate
MEG	Magnetoencephalography
ML	Machine Learning
MMtFA	Mixtures of Modified t-factor Analyzers
MMV	Multiple Measurement Vector
MSE	Mean Square Error
MSNIG	Multivariate Skewed Normal Inverse Gaussian
MSSt	Multivariate Skewed Student
NIG	Normal Inverse Gaussian
NVMM	Normal Variance Mean Mixture
PCA	Principal Component Analysis
PSNR	Peak Signal to Noise Ratio
OMP	Orthogonal Matching Pursuit
RIP	Restricted Isometry Property
SAL	Shifted Asymmetric Laplace
SBL	Sparse Bayesian Learning
SMV	Single Measurement Vector
SSR	Sparse Signal Recovery
SSVEP	Steady State Visual Evoked Potential
STSBL	Spatio-Temporal SBL
STSVB	Spatio-Temporal Sparse Variational Bayes
STSVB-J	STSVB based on Jeffreys as marginal
STSVB-L	STSVB based on Laplace as marginal
STSVB-St	STSVB based on Student's t as marginal
SVB	Sparse Variational Bayes
TRCA	Task Related Component analysis
TMSVB	Low Complexity version of Temporal (Multiple) Sparse Variational Bayes
TSVB	Temporal Sparse Variational Bayes
JTMSVB	TMSVB based on Jeffreys as Marginal
JTSVB	TSVB based on Jeffreys as Marginal

LTMSVB	TMSVB based on Laplace as Marginal
LTSVB	TSVB based on Laplace as Marginal
StTMSVB	TMSVB based on Student as Marginal
StTTSVB	TSVB based on Student as Marginal
TSBL	Temporal SBL
VAE	Variational Autoencoders
VB	Variational Bayes
VG	Variance Gamma

NOTATION

Small letters	Scalars
Small bold letters	Vectors
Capital bold letters	Matrices
$\text{diag}(a_1, a_2, \dots, a_n)$	Diagonal matrix with principal diagonal $a_1, a_2, \dots, a_n$
$\text{diag}(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n)$	Block diagonal matrix with principal diagonal $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$
$\mathbf{A}_{.j}$	j^{th} column of matrix $\mathbf{A}$
$\mathbf{A}_i$	i^{th} row of matrix $\mathbf{A}$
$\mathbf{A}_{[i]}$	i th block of all columns
$\mathbf{A}_{[i]}$	i th diagonal block
$\mathbf{A}_{[i]\ell}$	ℓ th column of $\mathbf{A}$ in i th block
a_{ij}	ij^{th} element of matrix $\mathbf{A}$
$\det(\mathbf{A})$ or $ \mathbf{A} $	Determinant of $\mathbf{A}$
$\mathbf{A}^t$	Transpose of $\mathbf{A}$
$\ \mathbf{x}\ _2$	l_2 norm of $\mathbf{x}$ ($= \sqrt{\sum_i x_i^2}$)
$\ \mathbf{x}\ _\infty$	l_∞ norm of $\mathbf{x}$ ($= \max_i x_i$)
$\ \mathbf{A}\ _{\mathcal{F}}$	Frobenius norm of $\mathbf{A}$ ($= \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^t)} = \sqrt{\text{tr}(\mathbf{A}^t\mathbf{A})}$)
$\text{tr}(\mathbf{A})$	Trace of matrix $\mathbf{A}$
$\text{vec}(\mathbf{A})$	Vectorization of $\mathbf{A}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker Product of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathbb{E}[\cdot]$	Expected value of random parameter
$\mathcal{N}(\cdot)$	Gaussian distribution
$\mathcal{G}(\cdot)$	Gamma distribution
$\mathcal{IG}(\cdot)$	Inverse Gamma Distribution
$\mathcal{GIG}(\cdot)$	Generalized Inverse Gaussian distribution
$\Gamma(\cdot)$	Gamma function