

Towards Unsupervised Online Band Selection in Hyperspectral Imaging

by

Gautam Balasubramanian

B.E. (Computer-Systems),
University of Technology, Sydney, Australia, 2003.

MSc. (Elec. Eng),
Monash University, Melbourne, Australia, 2005.

Thesis submitted for the degree of

Doctor of Philosophy

in

School of Electrical and Electronic Engineering
The University of Adelaide, Australia

December 2013

© 2013

Gautam Balasubramanian

All Rights Reserved



Typeset in L^AT_EX 2_ε

Gautam Balasubramanian

To my Beautiful Wife, Loving Parents and Omnisicent Teachers

Contents

Contents	v
Abstract	ix
Statement of Originality	xi
Acknowledgments	xiii
List of Figures	xvii
List of Tables	xxiii
Chapter 1. Introduction	1
1.1 Overview	2
1.2 Hyperspectral Imaging (HSI)	4
1.3 Hyperspectral Band Selection	8
1.4 Problems Addressed	9
1.4.1 Unsupervised Band Selection to improve Model Estimation Accuracy of a Scene	9
1.4.2 Inferring Appropriate Bands to find Anomalies in the Scene	10
1.4.3 Pixel-by-Pixel Online Band Selection for Band Cueing using Sub-Pixel Mixing Criteria	11
1.5 Contributions and Publications	12
1.6 Thesis Structure	13
Chapter 2. Preliminaries	15
2.1 Introduction	16

- 2.2 Gaussian Mixture Models 16
 - 2.2.1 Parameter Estimation - Maximum Likelihood 16
 - 2.2.2 Maximum Likelihood Estimation of Gaussian Mixtures via Expectation-Maximisation 18
 - 2.2.3 Space Alternating Generalised Expectation (SAGE) 19
- 2.3 Compositional Models 21
 - 2.3.1 Bayesian Parameter Estimation 23
 - 2.3.2 Inference using Gibbs Sampling 23
- 2.4 Band Selection 24
 - 2.4.1 Convex Optimisation 25
 - 2.4.2 Stochastic Beta Processes 26

Chapter 3. Unsupervised Band Selection using Gaussian Mixtures and Maximum

Likelihood Criteria 29

- 3.1 Introduction 30
- 3.2 Background 31
 - 3.2.1 Gaussian Mixtures for Hyperspectral Data 31
 - 3.2.2 Nonlinear vs Linear Band Scoring 31
- 3.3 Existing Work 32
- 3.4 Maximum Likelihood Criteria for Band Scoring 34
 - 3.4.1 Motivation 34
 - 3.4.2 Proposed Model 35
 - 3.4.3 EM Algorithm 37
 - 3.4.4 Non-linear Band Scoring using Convex Optimisation 40
 - 3.4.5 EM CVX Algorithm 41
 - 3.4.6 Proof of Concavity for the Band Selection Objective 42
- 3.5 Experiment A 43
- 3.6 Conclusions and Limitations 44

Chapter 4. Inferring Appropriate Bands To Find True Anomalies	49
4.1 Introduction	52
4.2 Existing Work	53
4.2.1 Band Selection Criteria	53
4.2.2 Band Selection or Reduction Process	55
4.2.3 Summary of Work and Contributions	56
4.3 Methodology	58
4.3.1 Problem Formulation	58
4.3.2 Labelling Outliers and Partial Backgrounds using Convex Relaxation	61
4.3.3 Maximum Likelihood Estimation of Gaussian Mixture Band-Subsets	64
4.3.4 A Kullback-Leibler Divergence for Maximising Partially Labelled Gaussian Mixtures	67
4.3.5 Anomaly Detection and Band Ranking Using Convex Relaxation	69
4.4 Experiments	72
4.4.1 Experiment A: Simulated Gaussian Mixture data	72
4.4.2 Experiment B: Real Hyperspectral Data	73
4.5 Discussion	76
4.6 Conclusion	77
Chapter 5. Band Sparsity for Compositional Models in Hyperspectral Imaging	81
5.1 Introduction	84
5.1.1 Motivation and Significance	86
5.1.2 Summary of Work and Contributions	87
5.2 Existing Work	89
5.3 Problem Formulation	91
5.4 Background	94
5.4.1 Representing End-members using Gaussian Processes	94

- 5.4.2 Using Gamma and Dirichelet Distribution to represent Abundance 94
- 5.5 Posterior Probability Estimates for a Naive Gibbs Sampler 96
 - 5.5.1 Estimating the Endmember Posterior 96
 - 5.5.2 Estimating the Abundance Posterior 97
 - 5.5.3 Abundance Sampling - Technique A: Gamma Dirichelet Relation 97
 - 5.5.4 Abundance Sampling - Technique B: Non-Linear Transformation 99
 - 5.5.5 Gibbs Sampler using Abundance Sampling Technique A 100
 - 5.5.6 Gibbs Sampler using Abundance Sampling Technique B 102
- 5.6 Recursive Band Selection using Beta Processes 102
 - 5.6.1 Estimating Base Measure using Convex Relaxation 103
 - 5.6.2 Beta and Bernoulli Processes 106
 - 5.6.3 RSBS Algorithm Summary 108
- 5.7 Experiments 108
 - 5.7.1 Experiment A 109
 - 5.7.2 Experiment B 110
- 5.8 Discussion 112
- 5.9 Conclusion 119
- 5.10 Appendix 120

Chapter 6. Concluding Remarks 123

- 6.1 Conclusion 124
- 6.2 Recommendations on Future Work 125

References 127

Abstract

This thesis explores the problem of unsupervised selection of a set of spectral wavebands in a hyperspectral sensor for a surveillance task. Selecting a subset of wavebands for surveillance has the advantage of reducing data throughput and hence network bandwidth requirements, computational complexity for processing the data and storage requirements in a ground-station. For the sensor designer, Signal-To-Noise Ratio and other sensor-band improvements can be made on those bands deemed critical for the surveillance task. In chapters 3 and 4, we propose the use of locally correlated high-dimensional Gaussian Mixture models to account for band overlap where maximum likelihood estimates of the parameters of such a model are provided using the SAGE-EM (Space Alternating Generalised Expectation Maximisation) algorithm. In both these chapters convex-relaxation strategies are proposed to handle the combinatorial complexity of selecting a subset-of bands that are locally correlated and contain non-Gaussian measurements. However, in chapter 4, we select bands according to anomaly detection criteria as opposed to modelling estimation accuracy (likelihood) as done in chapter 3. We breakdown the problem such that any pixel contains band measurements that belong to either an outlier or partial background distribution, where the distributions diverge across band-subsets in a Kullback-Leibler (KL) divergence sense. A pixel is deemed as an anomaly if it contains a certain number of outliers. We identify the bands that contain the most number of contiguous outlier measurements and also subsequently reveal the presence of anomalies. Finally, in the last chapter we solve the problem of online band selection for sub-pixel compositional hyperspectral models using a Bayesian approach. Online band-selection enables spectral-band cueing and automation for adaptive focal plane arrays where not all bands are used to measure each pixel. We apply beta process models to provide a recursive strategy to select bands based on prior knowledge of their utility as well as bands used in neighbouring pixels. Band utility is measured through convex-relaxation as the subset of bands that provides the best abundance estimation accuracy of training data. The

combination of a Gaussian process prior for possible end-members (pure materials) as well as a Gamma distributions for the abundance, enables efficient posterior sampling from a joint Normal-Gamma distribution. Furthermore, natural spectral band variations are retained making the model suitable for band selection, where approximate sum-to-one constraints are enforced through an intelligent update of the Gamma hyperparameters, based on the Dirichlet-Gamma relation. Experiments are conducted on synthetic Gaussian Mixture data with additive noise (Chapters 3, 4), Rochester Institute of Technology (RIT) Target Detection Test using the HyMAP sensor, (Chapter 4), synthetic sub-pixel data created using USGS spectral database [1] (Chapter 4) and AVIRIS-Cuprite dataset used by Mittelman et.al. in [2] (Chapter 4).

Statement of Originality

This work contains no material that has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of the thesis, when deposited in the University Library, being available for loan, photocopying and dissemination through the library digital thesis collection.

The author of this thesis acknowledges that copyright of published work contained within this thesis (as listed in the publications page) resides with the copyright holder(s) of that work.

Signed

Date

Acknowledgments

It is said that success has many fathers and failures but one. I measure success by personal growth rather than gain and in this sense I consider the journey thus far, a success. There are many contributors that have enabled this growth, one not greater or lesser in importance than the other both personally and academically.

Personally, my gorgeous wife, Prabha had to put up with me rushing to submit a conference paper on the first night of our marriage, alas it was the first of many more nights where her husband was missing, lost in equations within the confines of the study. Her enthusiasm for my work, belief in my abilities and passion for her husband's success made me work harder, gave me the constancy of effort required, she was always an un-ending source of encouragement and inspiration. She also instilled a sense of discipline I previously did not possess, an attribute of hers which I always hope to emulate. Thank You for the Umpteen Sacrifices, Love You heaps gorgeous. Next, my parents, who always gave me the freedom to pursue my own dreams throughout my Masters and now my PhD, selflessly, even though they knew it would keep me away from home, it must have been very difficult. Their presence always gave me perspective and reminded me to appreciate all facets of life and to never forgo responsibilities to wife, family and society although I am sure I did plenty of times. I'll always be eternally grateful for their presence in my life. I hope I have made you proud Mum n Dad. Finally, my mother-in-law and grandma who gave me constant nourishment, never shy from cooking up a storm with or without my behest also showing interest and enthusiasm on the outcomes of my work, where would I be without your kindness. I'll be eternally grateful for the sacrifice and faith of all my family members, this PhD is as much theirs as it is mine.

Academically, my principal Supervisor Professor Lang White, he's truly my first academic teacher from whom I have learned the most. His level of intelligence, comfort,

Acknowledgments

enthusiasm and care with subjects of Maths and Signal Processing was a sight to behold, a memory to treasure and I hope to emulate that throughout my career. Through many informal chats, emails and presentation of his own work he also gave me a great deal of structure to my thinking and improved my ability to frame problems which opened up creative avenues to solve problems and translated to a higher standard of work. Finally, his general niceness, fairness as a human being and compassion to myself and fellow students always made him a pleasure to be around. Secondly, Dr Jason Williams who is a Senior Research Scientist at DSTO: Jason was a force to behold. I tried picking-up many tid-bits, from the way he expressed himself mathematically on the whiteboard and paper, the way he attacked problems, his vast knowledge-base about a variety of different fields and unending thirst to gain knowledge. Jason also introduced the Machine Learning and Bayesian aspects to my work bringing to my attention the latest and greatest techniques relevant to problems pursued. His humility and work habits is something I always hope to emulate. My secondary Supervisors Dr Vittala Shettigara and Dr Tim Payne, bestowed the faith in me to pursue a PhD in spite of our research group being short-staffed, remained steadfast with their commitment for which I'll always be grateful. Vittal, introduced me to the Band Selection problem and brought to my attention the significance of almost all the sub-problems pursued in this thesis always imparting the notion of presentation and completeness. Tim's timely feedback during the initial stages of my PhD also made me think deeper about framing Maximum Likelihood Estimation as a Convex Optimisation problem. Dr Sanjeev Arulampalam a Senior Research Scientist at DSTO and Professor Ian Fuss gave me the self-belief and faith to pursue the PhD and beyond, with vigour and to receive their encouraging words from people of stature made a striking difference to my work. Gratitude must also be extended to my peers at Adelaide Uni mainly Jiang He and Yassir. You made the journey very pleasant with your cheerful demeanour and inquisitiveness in my work. To the staff in the HSI group at DSTO, thank you for your patience and sacrifice in taking on a greater workload due to my constant absence.

Finally, to my Teachers at Chinmaya Mission, you set me off on this journey in 2003 and this work is but my humble offering to You. I hope it is worthy.

Gautam Balasuramanian (August 2013)

List of Figures

1.1	a) A slab of measurements are recorded by multiple frequency band snapshots as the sensor platform moves along track. The X and Y axes correspond to the 2D detector array. The multi-dimensional radiance signal recorded across multiple bands, for a single pixel is also shown. The acronyms VIS refers to Visible, NIR refers to the near-infrared region of the electro-magnetic spectrum and SWIR refers to the short-wave-infrared region.	6
<hr/>		
3.1	Exp. A1: Combined Band MSE; EM-CVX Algorithm; Selected 5 out of 10 frequency bands, 0 dB SNR across all frequency bands	45
3.2	Exp. A2: Combined Band MSE; EM-CVX; Select 20 out of 100 frequency bands, 0 dB SNR across all channels	46
3.3	Exp. A3: Combined Band MSE; EM-CVX; Select 25 out of 200 frequency bands, 20 dB SNR across all frequency bands	47
3.4	Combined Band MSE; EM-CVX; Select 10 frequency bands, no removal of frequency bands	48
3.5	Exp. A5: Combined Band MSE; EM-CVX; Select 10 out of 50 frequency bands, 20, 10 ,5 dB SNR evenly distributed across all frequency bands for two arbitrary initialisation points	48

4.1 a) Proposed algorithm simultaneously identifies critical band-subsets which reveal the presence of anomalies. Outlier (O) and partial backgrounds (PB) measurements are first identified across R band-subsets. The diagram on the right shows $P < N$ anomaly pixels (A) that produce the greatest KL divergence between PB and O distributions using a subset of bands from a critical band rank. b) The graphical model describes the generative process for each r -th band-subset which contains $Q(r)$ bands. Indicator variable $T(r)$ indicates the membership of a spectral sample from the r -th band-subset to outlier, $Z(r)$ and partial background $Z(r)^c$ subsets. The full circles are random variables whilst the square plates around the circles indicate number of measurements, bands or components of the variable in the circle whilst dotted circles represent parameters that are non-random variables. A measurement window contain a maximum of P anomalies where anomalous pixels exceed $P + 1$ thresholds. Both anomalies and critical bands are derived from convex matrix Φ which is restricted by inequality constraints λ_1, λ_2 . Z_n^A, Z_n^B are binary matrices that indicate membership of the n th pixel to anomaly and background groups. 62

4.2 Experiment A: a) Subset of backgrounds (blue) and all true anomalies (red) detected for the least difficult scenario considered where there are 7/10 band-subsets that contain outliers. b)Arbitrary backgrounds (blue) and all true anomalies (red) detected for the tougher case where there are 4/10 band-subsets that contain outliers 73

4.3	<p>Experiment A: a) P_D vs. FAR for $M = 7, 6, 5, 4, 3$ anomalous bands-subsets out of $R = 10$ total band-subsets. They are represented by red, blue, purple, black and green curves respectively. Each band-subset consists of $Q(r) = 3$ bands hence making the total number of bands equal to 30. Error bars (referred to as SD in the legend) indicate the accuracy range for $E = 1000$ simulations. b) Band Ranking performance is measured according to how many times critical band-subsets that contain unique outliers were actually chosen. It is represented as a function of the cumulative Cauchy-Schwarz distances between partial background and outlier distributions and as a function of the anomalous band subsets is indicated in the legend.</p>	78
4.4	<p>Spectral measurements of anomaly vs background in local window consisting grass, tree and soil. Green asterix indicates critical bands inferred for each material. Note how the locations vary. For anomaly, F_4, which produced the worst result in terms of P_D, 34 bands are required to obtain the result. Critical bands identified for F_1 can be validated visually, whereas the inferred critical bands are not so obvious for the others. . .</p>	79
4.5	<p>Experiment B: P_d vs FAR for finding 54 anomalies out of 10000 pixels: F_1, F_2, F_3, F_4.</p>	80

5.1 This directed graphical model represents the generative model used to capture linear sub-pixel mixing phenomena described in equation (5.1). Random variables (circles) and hyperparameters (smooth boxes) are unknown and inferred using a Gibbs Sampler. The lower-case symbols used inside the circles represents samples of those random variables. Arrows indicate the dependencies between random variables. The exception to this rule is the pixel y_n which is an observation. In this model the k th abundance of the n th pixel is represented by $x_{k,n}$ and $g_{k,n}$ is the k th endmember that is present in the n th pixel and sampled from posterior probabilities of random variables X_k, G_k . The endmember and abundance are conditionally dependant given the measurement at the n th pixel y_n . Hyperparameters $\alpha_{k,n}$ (shape) and $\beta_{k,n}$ (scale) vary for each n th pixel. In this model, the measurement at each n th pixel is assumed to be independent of remaining $N - 1$ pixels. Indicator matrix $T_{d,n}$ is not a random variable and is iteratively inferred to determine whether M_n bands are sufficient to describe the pixel. y_n . The band selection aspect of the model is specific to training data and is used to estimate the base distribution of the Beta process B_0 92

5.2 The following graphical model applies to test data where prior band utility is described by a base measure B_0 obtained from training data. Each n th pixel is represented by no more than $Q_n \subset D$ total number of bands in the sensor array, where value Q_n is drawn from a Poisson random variable, $\pi_n(d)$ is the posterior band utility which forms the posterior Beta process that depends on base distribution B_0 derived from training data and prior band labels. Hyperparameters for the Beta process are indicated by ρ_1, c . Both sets of weights are combined in estimating binary random matrix $Z = z_1(d), \dots, z_N(d)$ which is a result of successive draws from a Bernoulli random variable and forms the posterior Bernoulli process. 103

5.3 The figure displays the base probability measure used for the Beta process, B_0 in experiment A and experiment B, where the prior probabilities indicate band utility. For both experiments, bands between 170 – 200 are more useful than the remainder. 112

5.4 Experiment A a), c): Posterior Beta process measured after 20, 100 pixels, where the concentration parameter $c = 1$ and $\gamma = 15$. Experiment B b), d): Posterior Beta process measured after 20, 100 pixels, with the same hyper-parameters. Prior band utility is captured from the discrete base measures B_0 as is evident from the number of bands chosen after band 100. The small size of c ensures that bands used to describe previous pixels is captured. The size of the γ value ensures that a sufficient number of new bands are sampled as evident from bands with a small number of counts. 114

5.5 Experiment A a), c): Posterior Beta process measured after 1000, 10000 pixels, where the concentration parameter $c = 1$ and $\gamma = 15$. Experiment B b), d): Posterior Beta process measured after 1000, 6400 pixels, with the same hyper-parameters. Prior band utility is captured from the discrete base measures B_0 as is evident from the number of bands chosen after band 100. The small size of c ensures that bands used to describe previous pixels is captured. The size of the γ value ensures that a sufficient number of new bands are sampled as evident from bands with a small number of counts. 115

5.6 a) Experiment A: shows 20, 22, 23, 27 bands used at pixel locations 20, 100 and 1000, 10000. Bands greater than 140 or 1500nm are used consistently across all 4 locations. This is agreeable with the base measure which contains larger probabilities at these locations. b) Experiment B: Only 10 – 14 bands are used at the four locations 20, 100, 1000, 6400 pixels. Bands greater than 1700 nm are used consistently across these locations. 117

5.7 Experiment B: Abundance Map of AVIRIS-Cuprite image subset. a) Original Image (courtesy Mittleman *et. al.* [2]). b) Kaolinite 1 c) Kaolinite 2 d) Alunite e) Montmorollinite f) Sphene 118

5.8 Experiment B: a) Posterior endmember estimates at the first pixel of Kaolinite 1 (blue), Kaolinite 2 (green), Alunite (red), Montmorillonite (cyan) and Sphene (magenta). The signatures marked with crosses represent the estimate whilst those without any markers represent the true value. 119

List of Tables

3.1	Band Selection Experiment Summary	44
4.1	Experiment B: Anomaly Details	74
4.2	Experiments A,B: Parameter Summary (* Parameter setting for Atmospheric Bands)	75
4.3	Critical Bands, True Anomaly Detection Summary	75
5.1	Experiment A: Best-Case Abundance SSE for Synthetic Data: RSBS - Tech. A, noBandSel - Tech. A, B vs SCU, VCA, BLU	112
5.2	Experiment B: Endmember SSE against USGS ground-truth. Mean and Standard Deviation with added Gaussian Noise at 10dB SNR: RSBS vs SCU, VCA, BLU	113
5.3	Experiment A: Bands Used at different Pixel Locations	116