

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Computationally tractable fitting of
graphical models:
the cost and benefits of decomposable Bayesian
and penalized likelihood approaches.**

A thesis presented in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

at Massey University, Albany,

New Zealand.

Anne Marie Fitch

2012

Abstract

Gaussian graphical models are a useful tool for eliciting information about relationships in data with a multivariate normal distribution. In the first part of this thesis we demonstrate that partial correlation graphs facilitate different and better insight into high-dimensional data than sample correlations. This raises the question of which method one should use to model and estimate the parameters. In the second, and major part, we take a more theoretical focus examining the costs and benefits of two popular approaches to model selection and parameter estimation (penalized likelihood and decomposable Bayesian) when the true graph is non-decomposable.

We first consider the effect a restriction to decomposable models has on the estimation of both the inverse covariance matrix and the covariance matrix. Using the variance as a measure of variability we compare non-decomposable and decomposable models. Here we find that, if the true model is non-decomposable, the variance of estimates is demonstrably larger when a decomposable model is used. Although the cost in terms of accuracy is fairly small when estimating the inverse covariance matrix, this is not the case when estimation of the covariance matrix is the goal. In this case using a decomposable model caused up to 200-fold increases in the variance of estimates.

Finally we compare the latest decomposable Bayesian method (the feature-inclusion stochastic search) with penalized likelihood methods (graphical lasso and adaptive graphical lasso) on measures of model selection and prediction performance. Here we find that graphical lasso is clearly outclassed on all measures by both adaptive graphical lasso and feature-inclusion stochastic search. The sample size and the ultimate goal of the estimation will determine whether adaptive graphical lasso or feature-inclusion stochastic search is better.

Acknowledgements

First and foremost thanks to my supervisor Beatrix Jones . I have greatly benefited from her encouragement and help along the way. Without her I would never have started let alone finished! Thanks also to my co-supervisor James Curran, especially for his help with R-graphics and \LaTeX code.

Three authors made their code available to us. Thanks to Xianghong Zhou for the shortest path analysis code used by the authors of Zhou *et al.* (2002); James Scott for his C++ code for feature-inclusion stochastic search and Yang Feng for his R-code for adaptive graphical lasso.

Thanks to Robert McKibbin and Tony Norris (Heads of Institute) and Jeff Hunter, Howard Edwards and Marti Anderson (Statistics discipline leaders) who in these roles have both facilitated the opportunity for me to combine study and teaching and also gave much encouragement along the way. Thanks to fellow student Insha Ullah for sharing his R-code for finding a perfect elimination ordering. Thanks also to the rest of the Massey Statistics and Mathematics teams for your encouragement, especially over the write-up months.

Last, but certainly not least, a big thank you to my family. Thanks to Harmen for his help in getting Figure 3.3 into a printable form. Thanks to Nigel, Miriam and Harmen, Joanna, and Andrew, without their support and encouragement I could not have done this.

Whaia te iti kahurangi ki te tuohu koe me he maunga teitei.

Contents

1	Introduction	1
1.1	Research objectives	1
1.2	Gaussian graphical models	2
1.3	Overview	3
1.4	Publications	6
2	Background	7
2.1	Gaussian graphical models	7
2.1.1	Overview	7
2.1.2	Penalized likelihood methods: graphical lasso and adaptive graphical lasso	9
2.2	Graph theory	12
2.2.1	Decomposability and decomposable graphs	12
2.2.2	Gaussian graphical models	16
2.2.3	Acyclic directed graphs	18
2.3	Bayesian model selection	19
2.4	Predictions	22
3	Shortest paths with partial correlations	23
3.1	Introduction	23
3.2	Data and methods	25
3.2.1	Data	25
3.2.2	Shortest path analysis	26
3.2.3	Graphical lasso	28

3.2.4	Using partial correlations in shortest path analysis	29
3.3	Results	29
3.4	Discussion	35
4	The cost of using a decomposable model	39
4.1	Introduction	39
4.2	Background	42
4.2.1	General properties of graphs	42
4.2.2	Parameter and variance estimation	43
4.3	Theory for the four variable case	44
4.4	Simulation study methods	46
4.4.1	The four variable case	46
4.4.2	20 and 50 variable cases	48
4.5	Simulation study results	49
4.5.1	Estimating Ω - the four variable case	49
4.5.2	Estimating Ω - 20 and 50 variable cases	52
4.5.3	Estimating the covariance matrix (Σ)	56
4.6	Case studies: Fisher's iris data and 12 node data	58
4.6.1	Fisher's iris data	58
4.6.2	12 node case	60
4.7	Discussion	62
5	Decomposable covariance selection	65
5.1	Introduction	65
5.2	Background	68
5.2.1	General properties of graphs	68
5.2.2	Feature-inclusion stochastic search	69
5.2.3	Kullback-Leibler divergence	71
5.2.4	Graphical lasso and adaptive graphical lasso	72
5.3	Data and methods	74
5.3.1	Data	74
5.3.2	Model selection	75

5.3.3	Prediction	76
5.4	Results	77
5.4.1	Feature-inclusion stochastic search (FINCS) treatment of non-decomposable graphs	77
5.4.2	Model selection comparison of FINCS with graphical lasso methods.	79
5.4.3	Comparison of predictions using FINCS and graphical lasso derived graphs	83
5.5	Discussion	87
6	Concluding discussion	89
6.1	Estimating the inverse covariance matrix	89
6.2	Regularizing the estimate of the covariance matrix	91
6.3	Concluding remarks	92
 Appendices		
A	Supplementary Tables and Figures for Chapter 4	95
A.1	Tables and Figures for the four variable case	95
A.2	Tables and Figures for 20 and 50 variable cases	101
A.3	Tables and Figures for estimations of Σ	110
A.4	Tables and Figures for 12 node example	113
B	Supplementary Tables and Figures for Chapter 5	117
B.1	Ω matrices	117
B.2	Graphs and Tables for Section 5.4.1	118
B.3	Relative inclusion probabilities	126
B.4	Maximum likelihood estimate (MLE) results	131
C	DRC forms	133
References		137

List of Figures

1.1	Example of a graphical model: Fisher’s <i>Iris virginica</i> dataset.	3
2.1	Graphs of five variables illustrating the idea of decomposition and decomposable versus non-decomposable.	13
3.1	Proportion of categorised paths less than upper bounds.	30
3.2	Percentage of transitive genes with annotations matching the annotations of the terminal genes on the same path.	31
3.3	Subgraphs of cytoplasm categorised path graphs.	33
4.1	A 4 variable non-decomposable graph(a) and a 4 variable decomposable graph (b).	45
4.2	Two different decomposable models when $p = 6$	48
4.3	OFI variances for $\hat{\omega}_{1,2}$ in Ω_{same} with sample sizes 10, 100 and 1000.	50
4.4	Percentage change in expected (EFI) and empirical variances when a decomposable model is fitted.	54
4.5	Empirical variances for elements of Σ when a decomposable model (Type A) vs when the true (non-decomposable) model is fitted.	57
4.6	The 12 variable models.	60
5.1	The 12 variable model.	74
5.2	Model score vs Kullback-Leibler divergence.	78
5.3	Model comparison measures: $p = 4, n = 50$ and $n = 1000$	80
5.4	Model comparison measures: $p = 20$	81
5.5	Model comparison measures: $p=12$	82
5.6	Mutual-funds data:edges vs total sum of squared errors.	83

5.7	Total sum of squared errors for 5 simulated samples of each n and p as specified.	85
5.8	Total sum of squared errors for 5 simulated samples: $n = 51$ and $p = 50$	86
A.1	OFI variances for $\widehat{\Omega}_{same}$ with sample sizes 10, 100 and 1000.	98
A.2	OFI variances for $\widehat{\Omega}_{big}$ with sample sizes 10, 100 and 1000.	99
A.3	OFI variances for $\widehat{\Omega}_{small}$ with sample sizes 10, 100 and 1000.	100
A.4	Percentage change in expected (EFI) and empirical variances when a decomposable model is fitted and relative standard deviation (RSD) when fitting true (cycle) and decomposable models.	105
A.5	Percentage change in expected (EFI) and empirical variances when a decomposable model is fitted and relative standard deviation (RSD) when fitting true and decomposable models.	106
A.6	Range of OFI variances when a cycle and two different decomposable models are fitted for $p=20, n=21$	107
A.7	Range of OFI variances when a cycle and two different decomposable models are fitted for $p=50, n=51$	107
A.8	Percentage of true non-zero elements declared zero when a cycle and two different decomposable models are fitted for $p=20, n=21$	108
A.9	Percentage of elements corresponding to ‘extra edges’ which are declared non-zero for $p=20, n=21$	108
A.10	Percentage of true non-zero elements declared zero when a cycle and two different decomposable models are fitted for $p=50, n=51$	109
A.11	Percentage of elements corresponding to ‘extra edges’ which are declared non-zero for $p=50, n=51$	109
A.12	Empirical variances for elements of Σ when a decomposable model (type B) vs when the true (non-decomposable) model is fitted $p=20$ and $n=21$	111
A.13	Empirical variances for elements of Σ when a decomposable model (type A) vs when the true (non-decomposable) model is fitted for $p=50$ and $n=51$	111

A.14	Empirical variances for elements of Σ when a decomposable model (type B) vs when the true (non-decomposable) model is fitted for $p=50$ and $n=51$	112
A.15	Elements of Σ when a decomposable model vs when the true (non-decomposable) model is fitted.	115
B.1	Top 10 graphs found by FINCS for $p=4$, samples of $n=1000$ for different $ \tilde{\rho}_{ij} $	118
B.2	Top 50 graphs found by FINCS for $p=20$, samples of $n=1000$ for different $ \tilde{\rho}_{ij} $	119
B.3	Top 10 graphs found by FINCS for $p=4$, samples of $n=1000$ for Ω_{small} and Ω_{big}	120
B.4	Top 50 graphs found by FINCS for samples of $n=1000$ for for Ω_{small} and Ω_{big} (both $p=20$) and for Ω_{twelve} ($p=12$).	122

List of Tables

4.1	Ω matrices used for simulating data.	47
4.2	$n = 1000$ empirical variances for cycle and percentage increase to decomposable.	49
4.3	$n = 1000$ Relative Standard Deviations for non-decomposable model and increase for decomposable model.	51
4.4	Percentage of times an element is declared zero ($ \text{estimate} < 2 \times$ standard error) for 1000 simulations ($n = 10$).	53
4.5	Estimated Ω and Σ matrices for <i>Iris virginica</i> dataset.	59
4.6	Estimates and standard error ($\sqrt{\text{OFI variance}}$) for <i>Iris virginica</i> dataset. 59	
4.7	Estimates and standard error ($\sqrt{\text{OFI variance}}$) for $\hat{\omega}_{1,2}$, $\hat{\omega}_{10,11}$ and $\hat{\omega}_{1,8}$. 61	
A.1	$n = 1000$ Relative Standard Deviations for non-decomposable model and increase for decomposable model.	96
A.2	Four variable empirical and EFI variances for elements of $\hat{\Omega}_{\text{same}}$ for cycle and decomposable, when $n=10$	96
A.3	Four variable empirical and EFI variances for elements of $\hat{\Omega}_{\text{big}}$ for cycle and decomposable, when $n=10$	96
A.4	Four variable empirical and EFI variances for elements of $\hat{\Omega}_{\text{small}}$ for cycle and decomposable, when $n=10$	97
A.5	Percentage of times an element is declared zero ($ \text{estimate} < 2 \times$ standard error) for 1000 simulations (Ω_{small}).	97
A.6	Non-zero elements of Ω matrices for $p=20$	102
A.7	Non-zero elements of Ω matrices for $p=50$	103

A.8	Four variable empirical variances for elements of $\widehat{\Sigma}$ for cycle and percentage increase to decomposable, when $n=10$	110
A.9	Estimates and OFI standard deviations for elements of $\widehat{\Omega}$ for cycle and three decomposable models , when $p=12$ and $n=250$	113
B.1	Ω_{twelve}	117
B.2	True edges missing in top 10 graphs found by FINCS for Ω_{small} when $p=4$	121
B.3	True edges missing in top 10 graphs found by FINCS for Ω_{big} when $p=4$	121
B.4	True edges missing in top 10 graphs found by FINCS for Ω_{small} when $p=20$	123
B.5	True edges missing in top 10 graphs found by FINCS for Ω_{big} when $p=20$	124
B.6	True edges missing in top 10 graphs found by FINCS for Ω_{twelve}	125
B.7	Relative inclusion probability matrices for $n=1000$ and $\tilde{\rho}_{ij}=-0.45$	126
B.8	Relative inclusion probability matrix for 12 node case when $n=1000$ and $\tilde{\rho}_{ij}=-0.4$ and for 20-node cycle when $n=1000$ and $\tilde{\rho}_{ij}=-0.45$	127
B.9	Relative inclusion probability matrices when $n=1000$ and top graphs are supersets.	128
B.10	Relative inclusion probability matrices when $n=50$	129
B.11	Relative inclusion probability matrices when $n=1000$ and partial correlations are small.	130
B.12	Model comparison measures for MLE results.	131
B.13	Sum of squared errors using the MLE.	131