

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Geometry of maximum likelihood estimation in Gaussian graphical models

Permalink

<https://escholarship.org/uc/item/6pz47470>

Author

Uhler, Caroline

Publication Date

2011

Peer reviewed|Thesis/dissertation

Geometry of maximum likelihood estimation in Gaussian graphical models

by

Caroline Uhler

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics

and the Designated Emphasis

in

Computational and Genomic Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Bernd Sturmfels, Chair

Professor Steven N. Evans

Professor Lior Pachter

Fall 2011

Geometry of maximum likelihood estimation in Gaussian graphical models

Copyright 2011
by
Caroline Uhler

Abstract

Geometry of maximum likelihood estimation in Gaussian graphical models

by

Caroline Uhler

Doctor of Philosophy in Statistics

University of California, Berkeley

Professor Bernd Sturmfels, Chair

Algebraic statistics exploits the use of algebraic techniques to develop new paradigms and algorithms for data analysis. The development of computational algebra software provides a powerful tool to analyze statistical models. In Part I of this thesis, we use methods from computational algebra and algebraic geometry to study Gaussian graphical models. Algebraic methods have proven to be useful for statistical theory and applications alike. We describe a particular application to computational biology in Part II.

Part I of this thesis investigates geometric aspects of maximum likelihood estimation in Gaussian graphical models. More generally, we study multivariate normal models that are described by linear constraints on the inverse of the covariance matrix. Maximum likelihood estimation for such models leads to the problem of maximizing the determinant function over a spectrahedron, and to the problem of characterizing the image of the positive definite cone under an arbitrary linear projection. In Chapter 2, we examine these problems at the interface of statistics and optimization from the perspective of convex algebraic geometry and characterize the cone of all sufficient statistics for which the maximum likelihood estimator (MLE) exists. In Chapter 3, we develop an algebraic elimination criterion, which allows us to find exact lower bounds on the number of observations needed to ensure that the MLE exists with probability one. This is applied to bipartite graphs, grids and colored graphs. We also present the first instance of a graph for which the MLE exists with probability one even when the number of observations equals the treewidth. Computational algebra software can be used to study graphs with a limited number of vertices and edges. In Chapter 4 we study the problem of existence of the MLE from an asymptotic point of view by fixing a class of graphs and letting the number of vertices grow to infinity. We prove that for very large cycles already two observations are sufficient for the existence of the MLE with probability one.

Part II of this thesis describes an application of algebraic statistics to association studies. Rapid research progress in genotyping techniques have allowed large genome-wide association studies. Existing methods often focus on determining associations between single loci and

a specific phenotype. However, a particular phenotype is usually the result of complex relationships between multiple loci and the environment. We develop a method for finding interacting genes (i.e. epistasis) using Markov bases. We test our method on simulated data and compare it to a two-stage logistic regression method and to a fully Bayesian method, showing that we are able to detect the interacting loci when other methods fail to do so. Finally, we apply our method to a genome-wide dog data set and identify epistasis associated with canine hair length.

Contents

List of Figures	iii
List of Tables	v
I Geometry of ML estimation in Gaussian graphical models	1
1 Introduction	2
1.1 Graphs	2
1.1.1 Colored graphs	3
1.2 Cones	4
1.3 Positive definite matrix completion problem	5
1.3.1 Positive definite matrix completion problem on colored graphs	6
1.4 Multivariate normal models	8
1.4.1 Conditional independence	9
1.5 Gaussian graphical models	9
1.5.1 Maximum likelihood estimation	10
1.5.2 Colored Gaussian graphical models	11
2 Multivariate Gaussians and convex algebraic geometry	13
2.1 Introduction	13
2.2 Linear sections, projections and duality	18
2.2.1 Geometry of maximum likelihood estimation in linear concentration models	18
2.2.2 Derivation of three guiding questions	21
2.2.3 Generic linear concentration models	22
2.3 Diagonal matrices, matroids and polytopes	25
2.4 Gaussian graphical models	29
2.4.1 Chordal graphs	30
2.4.2 The chordless m -cycle	33
2.4.3 Small graphs, suspensions and wheels	37

2.5	Colored Gaussian graphical models	42
3	Minimal number of observations needed for existence of MLE	49
3.1	Introduction	49
3.2	Geometry of maximum likelihood estimation in Gaussian graphical models .	50
3.3	Bipartite graphs	53
3.4	Small graphs	59
3.5	Colored Gaussian graphical models	62
4	Asymptotics of ML estimation in Gaussian cycles	69
4.1	Introduction	69
4.1.1	ML estimation in Gaussian cycles	70
4.2	On the distribution of quotients of normal random variables	71
4.3	Bounds on the MLE existence probability	73
II	Algebraic statistics and disease association studies	79
5	Using Markov bases for disease association studies	80
5.1	Introduction	80
5.2	Method	82
5.2.1	Models of interaction	82
5.2.2	Algorithm	84
5.3	Results	88
5.3.1	Simulation study	89
5.3.2	Comparison to logistic regression	90
5.3.3	Comparison to BEAM	91
5.3.4	Genome-wide association study of hair length in dogs	93
5.4	Discussion	94
	Bibliography	97

List of Figures

1.1	3×3 grid (left) and minimal chordal cover (right).	3
1.2	Colored 4-cycles: middle and right graphs satisfy condition that any pair of edges in the same color class connect the same vertex color classes; right graph is an RCOP graph.	4
2.1	Three figures, taken from [58], illustrate Example 2.1.2. These figures show the spectrahedron $\text{fiber}_{\mathcal{L}}(S)$ (left), a cross section of the spectrahedral cone $\mathcal{K}_{\mathcal{L}}$ (middle), and a cross section of its dual cone $\mathcal{C}_{\mathcal{L}}$ (right).	17
2.2	Geometry of maximum likelihood estimation in Gaussian graphical models. The cone \mathcal{K}_G consists of all concentration matrices in the model and \mathcal{K}_G^{-1} is the corresponding set of covariance matrices. The cone of sufficient statistics \mathcal{C}_G is defined as the projection of $\mathbb{S}_{>0}^m$ onto the edge set of G . It is dual and homeomorphic to \mathcal{K}_G . Given a sample covariance matrix S , $\text{fiber}_G(S)$ consists of all positive definite completions of the G -partial matrix S_G , and it intersects \mathcal{K}_G^{-1} in at most one point, namely the MLE $\hat{\Sigma}$	19
2.3	A Gaussian graphical model on five vertices and seven edges having dimension $d = 12$	39
2.4	Colored Gaussian graphical model for Frets' heads: L_i, B_i denote the length and breadth of the head of son i	43
2.5	The cross section of the cone of sufficient statistics in Example 2.5.3 is the red convex body shown in the left figure. It is dual to Cayley's cubic surface, which is shown in yellow in the right figure and also in Figure 2.1 on the left.	46
3.1	These pictures illustrate the convex geometry of maximum likelihood estimation for Gaussian graphical models. The cross-section of the cone of concentration matrices \mathcal{K}_G with a three-dimensional hyperplane is shown in (a), its algebraic boundary in (b), the dual cone of sufficient statistics in (c), and its algebraic boundary in (d) and (e), where (d) is the transparent version of (e).	51
3.2	Bipartite graph $K_{2,m}$ (left) and minimal chordal cover of $K_{2,m}$ (right).	53

3.3	<i>The MLE on $K_{2,m}$ exists in the following situations. Lines and data vectors corresponding to the variables 1 and 2 are drawn in blue. Lines and data vectors corresponding to the variables 3, \dots, $m + 2$ are drawn in red.</i>	54
3.4	<i>Bipartite graph $K_{3,m}$ (left) and minimal chordal cover of $K_{3,m}$ (middle). The tetrahedron-shaped pillow consisting of all correlation matrices of size 3×3 is shown in the right figure.</i>	58
3.5	<i>3×3 grid \mathcal{H} (left) and grid with additional edge \mathcal{H}' (right).</i>	59
3.6	<i>Graph \mathcal{G} (left) and minimal chordal cover of \mathcal{G} (right).</i>	60
3.7	<i>All possible sufficient statistics from one observation are shown on the left. The cone of sufficient statistics is shown on the right.</i>	64
4.1	<i>Strong correlation between variables 1 and 3 and between variables 2 and 4 leads to lines which are not graph consecutive.</i>	76
4.2	<i>Lines are in graph consecutive order with equal angle between consecutive lines</i>	78
5.1	<i>Nesting relationship of the control model, the additive model, and the multiplicative model. The intersection of the no 3-way interaction model with the multiplicative model corresponds to the additive model. The shading indicates the presence of epistasis.</i>	84
5.2	<i>Rejection rate of the no 3-way interaction test in the two-stage approach on 50 simulated association studies for $MAF=0.1$, $MAF=0.25$, and $MAF=0.4$ (left). Proportion of 50 association studies, in which the two causative SNPs were ranked among the ten SNPs with the lowest p-values by Fisher's exact test (middle). Rejection rate of the no 3-way interaction hypothesis using only the extended version of Fisher's exact test on the 50 causative SNP pairs (right).</i>	90
5.3	<i>ROC curves of the extended version of Fisher's exact test and logistic regression for $MAF=0.1$ (left), $MAF=0.25$ (middle), and $MAF=0.4$ (right) based on the ten filtered SNPs.</i>	91
5.4	<i>Proportion of simulation studies for which the interacting SNP pair belongs to the x SNP pairs with the lowest p-values for $MAF=0.1$ (left), $MAF=0.25$ (middle), and $MAF=0.4$ (right).</i>	92

List of Tables

2.1	<i>ML-degree of generic model for all matrices up to size $m = 6$.</i>	22
2.2	<i>Dimension, degree, ML-degree, and degree and number of minimal generators of the ideal P_{C_m} for $m \leq 8$.</i>	34
2.3	<i>Degree and number of terms of cycle polynomials for $m \leq 11$.</i>	37
2.4	<i>Our three guiding questions for all non-chordal graphs with $m \leq 5$ vertices. Column 4 reports the degrees of the minimal generators together with the number of occurrence (degree:number). The last column lists the degrees of the irreducible factors of the polynomial H_G that define the Zariski closure of the boundary of C_G. For each factor we report in lowercase the rank of the concentration matrices defining its dual irreducible component in the boundary of K_G.</i>	38
2.5	<i>Dimension, degree, ML-degree, and degree and number of minimal generators for W_m with $m \leq 6$.</i>	42
2.6	<i>Results on Questions 1, 2, and 3 for all colored Gaussian graphical models with some symmetry restrictions (namely, edges in the same color class connect the same vertex color classes) on the 4-cycle.</i>	47
2.7	<i>Continuation of Table 2.6.</i>	48
2.8	<i>All RCOP-models (see [39]) when the underlying graph is the 4-cycle.</i>	48
3.1	<i>This table shows the number of observations (obs.) and the corresponding MLE existence probability for all non-chordal graphs on 5 or fewer vertices.</i> . . .	61
3.2	<i>Results on the number of observations and the MLE existence probability for all colored Gaussian graphical models with some symmetry restrictions (namely, edges in the same color class connect the same vertex color classes) on the 4-cycle.</i>	66
3.3	<i>Continuation of Table 2.6.</i>	67
3.4	<i>All RCOP-models (see [39]) on the 4-cycle.</i>	68
5.1	<i>Standard interaction models for three-dimensional contingency tables.</i>	85
5.2	<i>Testing for association between haplotypes and phenotype.</i>	86

5.3	<i>Markov basis for the no 3-way interaction model on a $3 \times 3 \times 2$ table, where the tables are reported as vectors</i>	87
5.4	<i>Pairs of SNPs, which significantly interact with the hair length phenotype for the Canmap dataset. Question marks indicate that we were not able to identify a close-by gene which is functionally related to hair growth.</i>	94

Acknowledgments

First, I would like to express my appreciation and gratitude to my advisor, Bernd, for his endless support and guidance. I feel privileged to have been given the opportunity to learn from his enormous knowledge and grow as a statistician under his influence. He truly believes in each one of his students, and through his constant encouragement, motivation and personal example he inspired me to top performance. I particularly admire Bernd's ability to build bridges between various research areas as well as researchers. His advice and coaching went far beyond research and I am very grateful to him for introducing and integrating me into his personal network of colleagues, which is an invaluable exposure for a young researcher.

Further, I would like to thank Steve for providing me with encouragement and support throughout my whole PhD. He always had an open ear for me be that research questions or career advice. It was a great pleasure to be able to collaborate with Steve. His provision of teaching, time, and patience were an invaluable support on my climb towards graduation.

I would also like to thank Lior, my outside member, for his advice and support. I very much enjoyed working in the Math-Bio Lab, and the many discussions with him and his group provided me with great exposure to many different areas of research at the interface of computational biology, mathematics and statistics.

Since the first semester at Berkeley, I have been engaged in a very fruitful collaboration with Monty's research group, and in particular with Anna. I would like to thank the whole group for letting me spend many days in their lab and for teaching me a great deal about computational biology. Anna has also been a superb co-author, having written two papers with me over the past four years.

I have also received valuable support for my research from members outside of Berkeley. I am especially grateful to Steffen, who introduced me to the problem of existence of the MLE in Gaussian graphical models, which led to Part I of my thesis. I would also like to thank Pablo's group at MIT, in particular Venkat and Pari, for interesting discussions and a productive collaboration. Finally, I would like to thank Don for his valuable perspectives on Gaussian models and helpful career advice.

I am very grateful to Fulbright for supporting me financially during the first three years of my PhD. Besides giving me the opportunity to focus entirely on my research from the beginning of my PhD on, Fulbright greatly facilitated my move to the US by providing me with valuable advice, a network of people, and very useful seminars and fun events in the Bay Area and around the US. I would also like to thank the Janggen-Poehn Stiftung for providing financial support during the last year of my PhD.

Finally, my PhD would not have been as productive and enjoyable without the incredible support of my family and friends. I would like to thank my parents for always being there on skype with support and advice and reminding me that there are more important issues than PhDs. I am really fortunate to have met Nancy, Meromit and Anna during the first semester here at Berkeley. I would like to thank them for going with me through my PhD, for being wonderful house/office mates and for all the fun time we spent together.

Part I

Geometry of ML estimation in Gaussian graphical models

Chapter 1

Introduction

In current statistical applications, we are often faced with problems involving a large number of random variables, but only a small number of observations. This problem arises for example when studying genetic networks: We seek a model potentially involving a vast number of genes, while we are only given gene expression data of a few individuals. Gaussian graphical models have frequently been used to study gene association networks, and the maximum likelihood estimator (MLE) of the covariance matrix is computed to describe the interaction between different genes (e.g. [62, 77]). So the following question is of great interest from an applied as well as a theoretical point of view: What is the minimum number of observations needed to guarantee the existence of the MLE in a Gaussian graphical model? It is a well-known fact that the MLE exists with probability one if the number of observations is at least as large as the number of variables. In the following chapters we examine the case of fewer observations.

The goal of this first chapter is to provide an introduction to maximum likelihood estimation in Gaussian graphical models and to explain the relation to positive definite matrix completion problems. We will also introduce the necessary definitions and notations regarding graphs and multivariate models used throughout Part I of this thesis. The statistical theory of graphical models, which we describe briefly here, is more extensively elaborated upon in the book [49]. More about positive definite matrix completion problems can be found in Johnson's survey article [42].

1.1 Graphs

Throughout Part I, a *graph* is a pair $G = ([m], E)$, where $[m] = \{1, \dots, m\}$ is the set of vertices and $E \subset [m] \times [m]$ is the set of undirected edges. For simplicity of notation we assume, if not otherwise specified, that the edge set E contains all *self-loops*, i.e. $(i, i) \in E$ for all $i \in [m]$.

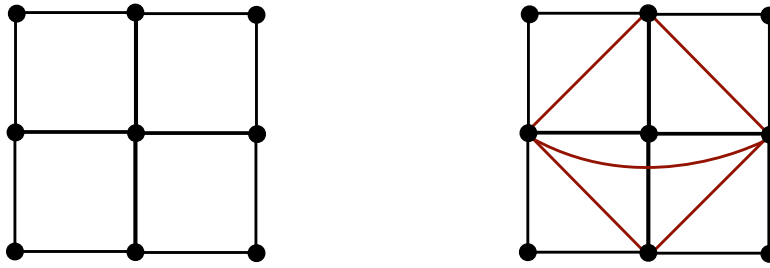


Figure 1.1: 3×3 grid (left) and minimal chordal cover (right).

A graph is *complete* if all pairs of vertices are joined by an edge. A maximal complete subset (with respect to inclusion) is called a *clique* and the maximal clique size of a graph is denoted by q . A graph is *chordal* (or *decomposable*) if every cycle of length $n \geq 4$ possesses a *chord*, i.e. any two non-consecutive vertices of the cycle are connected by an edge. For any graph $G = ([m], E)$ one can define a *chordal cover* $G^+ = ([m], E^+)$, a chordal graph satisfying $E \subset E^+$. We denote its maximal clique size by q^+ . A *minimal chordal cover* of G is a chordal cover with minimal maximal clique size. We denote a minimal chordal cover of G by $G^* = ([m], E^*)$ and its clique size by $q^* = \min(q^+)$. The *treewidth* of a graph is one less than q^* . Trees, for example, have treewidth one.

These graph theoretic concepts are illustrated by the example of the 3×3 grid in Figure 1.1. The maximal clique size of the 3×3 grid is 2. A minimal chordal cover is shown in Figure 1.1 (right). Its maximal clique size is 4. Consequently, the treewidth of the 3×3 grid is 3. Note that more generally, the treewidth of an $m \times m$ grid is m (see e.g. [11]).

1.1.1 Colored graphs

In the following chapters we will also encounter *colored graphs*, which are a generalization of the undirected graphs presented above. Let $G = ([m], E)$ be an undirected graph which does **not** contain any self-loops. Let the vertices of G be colored with $p \leq m$ different colors leading to a partition of the vertices into p non-empty disjoint *vertex color classes*

$$V = V_1 \sqcup V_2 \sqcup \cdots \sqcup V_p,$$

where all vertices in V_i have the same color. Similarly, coloring the edges E with $q \leq |E|$ different colors partitions E into q non-empty disjoint *edge color classes*

$$E = E_1 \sqcup E_2 \sqcup \cdots \sqcup E_q,$$

where all edges in E_j have the same color. A color class with a single element is called *atomic* and a color class which is not atomic is called *composite*. When visualizing a colored

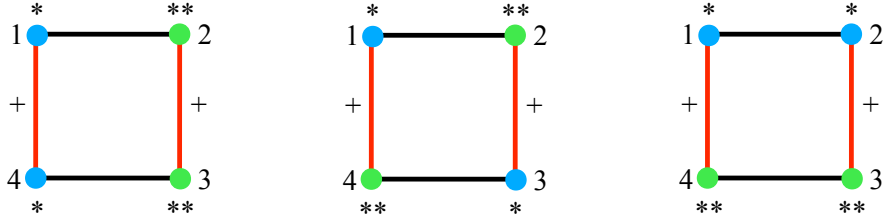


Figure 1.2: *Colored 4-cycles: middle and right graphs satisfy condition that any pair of edges in the same color class connect the same vertex color classes; right graph is an RCOP graph.*

graph, atomic color classes are displayed in black, whereas other colors are used for elements of composite color classes. Examples are given in Figure 1.2.

We will be especially interested in colored graphs, where any pair of edges in the same color class connect the same vertex color classes. The two graphs in Figure 1.2 (middle and right) satisfy this condition, whereas the graph in Figure 1.2 (left) does not.

Let $\text{Aut}(G)$ denote the group of automorphisms of a graph G (i.e. functions that map edges to edges). A special class of colored graphs are generated by permutation symmetry and called *RCOP graphs*. More precisely, a colored graph \mathcal{G} on the underlying undirected graph G is an RCOP graph if there is a subgroup $\Gamma \subset \text{Aut}(G)$, whose vertex orbits are the vertex color classes and whose edge orbits are the edge color classes. Clearly, RCOP graphs satisfy the condition that any pair of edges in the same color class connect the same vertex color classes. The graph in Figure 1.2 (right), for example, is an RCOP graph whose vertex and edge coloring represents the vertex and edge orbits generated by the permutation matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The graph in Figure 1.2 (middle), however, is not an RCOP graph, because the largest subgroup of $\text{Aut}(G)$ preserving the color symmetries consists only of the identity.

1.2 Cones

A (*convex*) *cone* in \mathbb{R}^k is a subset $C \subset \mathbb{R}^k$ such that $\alpha x + \beta y \in C$ for any positive scalars α, β and any x, y in C . We denote by \mathbb{S}^m the vector space of real symmetric $m \times m$ matrices and by $\mathbb{S}_{\geq 0}^m$ the convex cone of positive semidefinite matrices. Its interior is the open cone $\mathbb{S}_{> 0}^m$ of positive definite matrices.

Let $C \subset \mathbb{R}^k$ be a convex cone equipped with a scalar product $\langle \cdot, \cdot \rangle$. The *dual cone* to C is the set

$$C^\vee = \{w \in \mathbb{R}^k \mid \langle v, w \rangle \geq 0 \text{ for all } v \in C\}.$$

This is also a convex cone. If C is equal to its dual cone, C is called *self-dual*.

Let \mathbb{S}^m be a convex cone in a vector space equipped with the *trace inner product* $\langle A, B \rangle := \text{tr}(A \cdot B)$. With respect to this inner product, the cone $\mathbb{S}_{\geq 0}^m$ of positive semidefinite matrices is a full-dimensional self-dual cone in \mathbb{S}^m .

Let \mathcal{L} be a linear subspace of \mathbb{S}^m . We denote by \mathcal{L}^\perp the orthogonal complement of \mathcal{L} in \mathbb{S}^m . Using the fact that the full-dimensional cone $\mathbb{S}_{\geq 0}^m$ is self-dual, general duality theory for convex cones (see e.g. [10]) implies the following useful fact:

$$(\mathbb{S}_{\geq 0}^m \cap \mathcal{L})^\vee = (\mathbb{S}_{\geq 0}^m + \mathcal{L}^\perp) / \mathcal{L}^\perp. \quad (1.1)$$

Finally, a *spectrahedron* is a special convex cone formed by intersecting the positive semidefinite cone with an affine linear subspace.

1.3 Positive definite matrix completion problem

A *partial matrix* is a matrix whose entries are specified only on a subset of its positions, and a *completion* of a partial matrix is gotten by specifying the unspecified entries. Typically, the diagonal entries are assumed to be specified. *Matrix completion problems* are concerned with determining whether or not a completion of a partial matrix exists which satisfies some prescribed properties. Positive definite matrix completion problems are among the most studied matrix completion problems due to their occurrence in many applications.

For positive definite matrix completion problems it can be assumed without loss of generality that all diagonal entries are specified and that the specified entries are symmetric. Hence, the pattern of a partial matrix of size $m \times m$ can be represented by a graph $G = ([m], E)$, where the edges correspond to the specified entries of the partial matrix. The *positive definite matrix completion problem* is then defined as follows: Let $G = ([m], E)$ be the underlying graph. Given a partial symmetric matrix $M_G \in \mathbb{R}^E$ with specified entries at all positions corresponding to edges in the graph G , i.e.

$$M_G = (M_{ij} \mid (i, j) \in E),$$

determine whether there is a positive definite matrix \hat{M} satisfying $\hat{M}_G = M_G$. If this is the case, we say that M_G has a *positive definite completion*.

Clearly, if a partial matrix has a positive definite completion, then every specified principal submatrix is positive definite. Hence, having a positive definite completion imposes some obvious necessary conditions. However, these conditions are in general not sufficient

as seen in the following example where the underlying graph is the 4-cycle:

$$M_G = \begin{pmatrix} 1 & 0.9 & ? & -0.9 \\ 0.9 & 1 & 0.9 & ? \\ ? & 0.9 & 1 & 0.9 \\ -0.9 & ? & 0.9 & 1 \end{pmatrix}.$$

It can easily be checked that this partial matrix does not have a positive definite completion, although it satisfies the obvious necessary conditions. This example leads to asking if there are graphs for which the obvious necessary conditions are also sufficient for the existence of a positive definite matrix completion. The following remarkable theorem proven in [35] answers this question.

Theorem 1.3.1. *For a graph G the following statements are equivalent:*

- i) A G -partial matrix $M_G \in \mathbb{R}^E$ has a positive definite completion if and only if all submatrices corresponding to maximal cliques in M_G are positive definite.*
- ii) G is chordal.*

The proof in [35] is constructive and can be turned into an algorithm for finding a positive definite completion.

Positive definite completions are not unique. However, if a partial matrix has a positive definite completion, then there is a unique completion with maximum determinant. In addition, this unique completion is characterized by having a zero entry in the inverse at all unspecified positions (i.e. positions not corresponding to edges in the underlying graph). The uniqueness is due to the fact that the determinant function is strictly log concave over $\mathbb{S}_{>0}^m$ and the zero pattern in the inverse results from setting the partial derivatives of the determinant function to zero (see [35] for more details). Finding the unique completion which maximizes the determinant is a convex optimization problem and can be solved even for very large instances for example using the `Matlab` software `cvx` [32].

1.3.1 Positive definite matrix completion problem on colored graphs

In the following, we generalize positive definite matrix completion problems to colored graphs. The usual completion problem described above can be seen as a special case where each color class is atomic, i.e. each vertex and each edge is colored with its own color.

We define the *positive definite matrix completion problem for colored graphs* as follows: Let $\mathcal{G} = ([m], \mathcal{E})$ be the underlying colored graph with vertex color classes V_1, \dots, V_p and edge color classes E_1, \dots, E_q . Given a matrix $M \in \mathbb{S}^m$, let $M_G \in \mathbb{R}^{p+q}$ be the partial matrix

$$M_G = \left(\sum_{\alpha \in V_1} M_{\alpha\alpha}, \dots, \sum_{\alpha \in V_p} M_{\alpha\alpha}, \sum_{(\alpha, \beta) \in E_1} M_{\alpha\beta}, \dots, \sum_{(\alpha, \beta) \in E_q} M_{\alpha\beta} \right).$$

We wish to determine whether there is a positive definite matrix \hat{M} satisfying $\hat{M}_{\mathcal{G}} = M_{\mathcal{G}}$. If this is the case, we say that $M_{\mathcal{G}}$ has a *positive definite completion*.

Let G be the graph underlying the colored graph \mathcal{G} including all self-loops. Note that if M_G has a positive definite completion, then $M_{\mathcal{G}}$ also has a positive definite completion. So adding a graph coloring makes the positive definite matrix completion problem easier.

Similar to the result in [35], if $M_{\mathcal{G}}$ has a positive definite completion, then there is a unique positive definite completion with maximal determinant:

Theorem 1.3.2. *Let $\mathcal{G} = ([m], \mathcal{E})$ be a colored graph with vertex color classes V_1, \dots, V_p and edge color classes E_1, \dots, E_q . Let $M_{\mathcal{G}} \in \mathbb{R}^{p+q}$ and suppose $M_{\mathcal{G}}$ has a positive definite completion. Then there is a unique positive definite completion \hat{M} with maximal determinant. Furthermore, the unique completion \hat{M} is characterized by having a zero entry in the inverse at all unspecified positions and constant entries at all positions corresponding to the same color class, i.e.*

$$\begin{aligned} (\hat{M}^{-1})_{\alpha\beta} &= 0 && \text{for all } (\alpha, \beta) \notin E, \\ (\hat{M}^{-1})_{\alpha\alpha} &= (\hat{M}^{-1})_{\beta\beta} && \text{for all } \alpha, \beta \in V_i, \quad \text{for all } i \in \{1, \dots, p\}, \\ (\hat{M}^{-1})_{\alpha\beta} &= (\hat{M}^{-1})_{\gamma\delta} && \text{for all } (\alpha, \beta), (\gamma, \delta) \in E_j, \quad \text{for all } j \in \{1, \dots, q\}. \end{aligned}$$

Proof. Let $M_{\mathcal{G}}$ be a partial matrix which has a positive definite completion. Then the set

$$C := \left\{ Z \in \mathbb{S}_{\geq 0}^m \mid \sum_{\alpha \in V_i} Z_{\alpha\alpha} = \sum_{\alpha \in V_i} M_{\alpha\alpha}, \quad \sum_{(\alpha, \beta) \in E_j} Z_{\alpha\beta} = \sum_{(\alpha, \beta) \in E_j} M_{\alpha\beta}, \quad i = 1, \dots, p, \quad j = 1, \dots, q \right\}$$

is convex (a spectrahedron) and compact. The determinant function is log concave on C and therefore attains a unique maximum in C . Since the maximum is attained in the interior of C , it is uniquely characterized by

$$\det(Z) > 0 \quad \text{and} \quad \frac{\partial}{\partial Z_{kl}} \Lambda(Z, \kappa, \lambda) = 0 \quad \text{for all } (k, l) \in [m] \times [m],$$

where $\Lambda(Z, \kappa, \lambda)$ is the Lagrange function

$$\Lambda(Z, \kappa, \lambda) = \det(Z) - \sum_{i=1}^p \kappa_i \left(\sum_{\alpha \in V_i} Z_{\alpha\alpha} - \sum_{\alpha \in V_i} M_{\alpha\alpha} \right) - \sum_{j=1}^q \lambda_j \left(\sum_{(\alpha, \beta) \in E_j} Z_{\alpha\beta} - \sum_{(\alpha, \beta) \in E_j} M_{\alpha\beta} \right).$$

Computing the partial derivatives of the Lagrange function shows that \hat{M} is characterized by zero entries in the inverse at all unspecified positions and constant entries at all positions corresponding to the same color class. \square

1.4 Multivariate normal models

A random vector $X \in \mathbb{R}^m$ is distributed according to the *multivariate normal distribution* $\mathcal{N}(\mu, \Sigma)$ with parameters $\mu \in \mathbb{R}^m$ and $\Sigma \in \mathbb{S}_{>0}^m$ if it has the density function

$$f_{\mu, \Sigma}(x) = (2\pi)^{-m/2} (\det \Sigma)^{-1/2} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}, \quad x \in \mathbb{R}^m.$$

A *multivariate normal model* is a statistical model consisting of multivariate normal distributions:

$$\mathcal{P}_\Theta = \{\mathcal{N}(\mu, \Sigma) \mid \theta = (\mu, \Sigma) \in \Theta\},$$

where $\Theta \subseteq \mathbb{R}^m \times \mathbb{S}_{>0}^m$.

Ignoring the normalizing constant, the *log-likelihood function* for a Gaussian model on n observations X_1, \dots, X_n is

$$l(\mu, \Sigma) = -\frac{n}{2} \log \det \Sigma - \frac{1}{2} \operatorname{tr} \left(\Sigma^{-1} \sum_{i=1}^n (X_i - \mu)(X_i - \mu)^T \right).$$

In this thesis, we focus on the estimation of the covariance matrix Σ . We make no assumptions on the mean vector μ and always use the *sample mean*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

as estimate for μ . Thus, we are precisely in the situation of [25, Prop. 2.1.12] and the log-likelihood function simplifies to

$$l(\Sigma) = -\frac{n}{2} \log \det \Sigma - \frac{n}{2} \operatorname{tr} (S \Sigma^{-1}), \quad (1.2)$$

where

$$S = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$$

is the *sample covariance matrix*.

The inverse $K = \Sigma^{-1}$ of a covariance matrix Σ is also positive definite and known as the *concentration matrix*. When parametrizing a multivariate normal model, it is often more convenient to use K instead of Σ . The log-likelihood function in terms of K is

$$l(K) = \frac{n}{2} \log \det(K) - \frac{n}{2} \langle S, K \rangle. \quad (1.3)$$

This is a strictly concave function on the cone $\mathbb{S}_{>0}^m$.

1.4.1 Conditional independence

For multivariate normal distributions, independence relations are represented by zeros in the covariance matrix and conditional independence relations by zeros in the concentration matrix (see e.g. [49]). Assume that $Y \in \mathbb{R}^m$ is distributed according to the multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$. Then

$$\begin{aligned} \text{i) } Y_i \perp\!\!\!\perp Y_j & \iff \Sigma_{ij} = 0 \\ \text{ii) } Y_i \perp\!\!\!\perp Y_j \mid Y_{V \setminus \{i,j\}} & \iff K_{ij} = 0 \end{aligned}$$

In the following chapters, we will study multivariate normal models with linear restrictions on the entries of the concentration matrix. In order to interpret the entries of the concentration matrix, the following fact about multivariate normal distributions is particularly helpful: Let Y be distributed as $\mathcal{N}(\mu, \Sigma)$. Dividing the variables into two disjoint sets A, B , the conditional distribution of Y_A given $Y_B = y_B$ is $\mathcal{N}(\mu_{A|B}, \Sigma_{A|B})$, where

$$\mu_{A|B} = \mu_A + \Sigma_{AB} \Sigma_{BB}^{-1} (y_B - \mu_B) \quad \text{and} \quad \Sigma_{A|B} = \Sigma_{AA} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BA}.$$

Note that $\Sigma_{A|B}$ is the Schur complement of Σ_{BB} , and therefore $\Sigma_{A|B} = K_{AA}^{-1}$. Using this fact, it follows easily that the diagonal elements of the concentration matrix K are the reciprocals of the *conditional variances*, i.e.

$$K_{ii} = \frac{1}{\text{Var}(Y_i \mid Y_{V \setminus \{i\}})}.$$

The *conditional covariances* are also represented by the concentration matrix, namely

$$\text{Cov}(Y_i, Y_j \mid Y_{V \setminus \{i,j\}}) = \frac{-k_{ij}}{k_{ii}k_{jj} - k_{ij}^2}.$$

Similarly, we can compute the *partial correlation coefficients*

$$\rho_{ij|V \setminus \{i,j\}} = \frac{\text{Cov}(Y_i, Y_j \mid Y_{V \setminus \{i,j\}})}{\sqrt{\text{Var}(Y_i \mid Y_{V \setminus \{i,j\}}) \text{Var}(Y_j \mid Y_{V \setminus \{i,j\}})}} = \frac{-k_{ij}}{\sqrt{k_{ii}k_{jj}}}.$$

Hence, if we scale the concentration matrix to have all ones on the diagonal, the off-diagonal entries can be interpreted as the negative partial correlation coefficients.

1.5 Gaussian graphical models

Gaussian graphical models were introduced by Dempster [22] under the name of covariance selection models. Subsequently, the graphical representation of these models gained in importance. Lauritzen [49] and Whittaker [74] give introductions to graphical models in

general and discuss the connection between graph and probability distribution for Gaussian graphical models.

A Gaussian graphical model is a graphical interaction model for the multivariate normal distribution. It is determined by assuming conditional independence of selected pairs of variables given all the remaining variables. More precisely, if $G = ([m], E)$ is a graph and Y a random vector taking values in \mathbb{R}^m , the *Gaussian graphical model* for Y on G is given by assuming that Y follows a multivariate normal distribution which satisfies the *pairwise Markov property*, i.e.

$$Y_i \perp\!\!\!\perp Y_j \mid Y_{V \setminus \{i,j\}} \quad \text{for all } (i, j) \notin E.$$

As seen in Subsection 1.4.1, this is equivalent to assuming zeros in the concentration matrix at all entries not corresponding to edges in the graph, i.e.

$$K_{ij} = 0 \quad \text{for all } (i, j) \notin E.$$

Thus, for Gaussian graphical models the log-likelihood function given in (1.3) simplifies to

$$l(K) = \frac{n}{2} \left(\log \det(K) - \sum_{(i,j) \in E} S_{ij} K_{ij} \right).$$

and the *minimal sufficient statistics* are given by the partial sample covariance matrix S_G .

1.5.1 Maximum likelihood estimation

The partial sample covariance matrix S_G plays an important role when studying the existence of the MLE. Dempster [22] first characterized the MLE for Gaussian graphical models:

Theorem 1.5.1. *In the Gaussian graphical model on G , the MLE of the covariance matrix Σ exists if and only if the partial sample covariance matrix S_G can be completed to a positive definite matrix. Then the MLE $\hat{\Sigma}$ is the unique completion satisfying $(\hat{\Sigma}^{-1})_{ij} = 0$ for all $(i, j) \notin E$.*

Subsequently, the theory of exponential families has been developed. Gaussian graphical models are regular exponential families. The statistical theory of exponential families, as presented for example by Brown [13] or Barndorff-Nielsen [7], is a strong tool to establish existence and uniqueness of the MLE and generalizes Theorem 1.5.1. In regular exponential families, the MLE exists if and only if the sufficient statistic lies in the interior of its convex support. In that case the MLE is uniquely determined by the sufficient statistic $t(x)$, or more precisely by the equation

$$t(x) = \mathbb{E}(t(X)). \tag{1.4}$$

From Dempster's result (Theorem 1.5.1) it follows that checking existence of the MLE in Gaussian graphical models is a special matrix completion problem with an additional

rank constraint given by the number of observations. By combining Theorem 1.5.1 and Theorem 1.3.1 we get the following result about the existence of the MLE in Gaussian graphical models (see also [15]).

Corollary 1.5.2. *If $n \geq q^*$ the MLE exists with probability 1. If $n < q$ the MLE does not exist.*

Note that the probability statement in this corollary is with respect to the underlying multivariate normal distribution $\mathcal{N}(0, \Sigma)$, where Σ is unknown. However, because the multivariate normal distribution is continuous, the probability 1 statement does not depend on the value of Σ . The MLE exists with probability 1 means that the MLE exists for all sufficient statistics except possibly on a lower dimensional subspace.

Chordal graphs have $q^* = q$. Therefore, existence of the MLE depends on the number of observations only. For non-chordal graphs, however, there is a gap $q \leq n < q^*$, in which existence of the MLE is not well understood. Cycles are the only non-chordal graphs, which have been studied [8, 9, 15]. In Chapter 2, we will give a geometric description of the convex support of the sufficient statistics. In Chapter 3, we will discuss the connection to the minimum number of samples needed to ensure existence of the MLE and extend the results on cycles to bipartite graphs $K_{2,m}$ and small grids.

1.5.2 Colored Gaussian graphical models

For some applications, symmetries in the underlying Gaussian graphical model can be assumed. These symmetry restrictions can be represented by a graph coloring, where edges, or vertices, respectively, have the same coloring if the corresponding elements of the concentration matrix are equal. Such models have been introduced in [39].

Formally, if $\mathcal{G} = ([m], \mathcal{E})$ is a colored graph with underlying graph G and Y a random vector taking values in \mathbb{R}^m , the *colored Gaussian graphical model* for Y on \mathcal{G} is given by assuming that Y follows a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$ satisfying

$$\begin{aligned} (\Sigma^{-1})_{\alpha\beta} &= 0 && \text{for all } (\alpha, \beta) \notin E, \\ (\Sigma^{-1})_{\alpha\alpha} &= (\Sigma^{-1})_{\beta\beta} && \text{for all } \alpha, \beta \in V_i && \text{for all } i \in \{1, \dots, p\}, \\ (\Sigma^{-1})_{\alpha\beta} &= (\Sigma^{-1})_{\gamma\delta} && \text{for all } (\alpha, \beta), (\gamma, \delta) \in E_j && \text{for all } j \in \{1, \dots, q\}. \end{aligned}$$

If the underlying graph satisfies a permutation symmetry (i.e. it is an RCOP graph), we call the corresponding colored Gaussian graphical model an *RCOP model* (see [39]).

We can easily generalize Dempster's result (Theorem 1.5.1) to colored Gaussian graphical models using regular exponential family theory and Theorem 1.3.2 (see also [39]):

Theorem 1.5.3. *In the colored Gaussian graphical model on \mathcal{G} , the MLE of the covariance matrix Σ exists if and only if the partial sample covariance matrix $S_{\mathcal{G}}$ has a positive definite*

completion. Then the MLE $\hat{\Sigma}$ is the unique completion satisfying

$$\begin{aligned} (\hat{\Sigma}^{-1})_{\alpha\beta} &= 0 && \text{for all } (\alpha, \beta) \notin E, \\ (\hat{\Sigma}^{-1})_{\alpha\alpha} &= (\hat{\Sigma}^{-1})_{\beta\beta} && \text{for all } \alpha, \beta \in V_i, \quad \text{for all } i \in \{1, \dots, p\}, \\ (\hat{\Sigma}^{-1})_{\alpha\beta} &= (\hat{\Sigma}^{-1})_{\gamma\delta} && \text{for all } (\alpha, \beta), (\gamma, \delta) \in E_j, \quad \text{for all } j \in \{1, \dots, q\}. \end{aligned}$$

In Chapters 2 and 3 we will discuss various instances of colored Gaussian graphical models.

Chapter 2

Multivariate Gaussians and convex algebraic geometry

In this chapter, we study multivariate normal models that are described by linear constraints on the inverse of the covariance matrix (i.e. linear concentration models). Gaussian graphical models and colored Gaussian graphical models are examples of such models. In this chapter we give a geometric description of the set of all sufficient statistics for which the MLE exists.

Maximum likelihood estimation for linear concentration models leads to the problem of maximizing the determinant function over a spectrahedron, and to the problem of characterizing the image of the positive definite cone under an arbitrary linear projection. These problems at the interface of statistics and optimization are here examined from the perspective of convex algebraic geometry.

This chapter is joint work with Bernd Sturmfels. The content of this chapter has been published in the *Annals of the Institute of Statistical Mathematics* 62 (2010) 603 - 638 under the title "Multivariate Gaussians, semidefinite matrix completion, and convex algebraic geometry". We made minor changes throughout for consistency with other chapters and added Subsection 2.2.1. Subsection 2.2.1 is part of a paper with title "Geometry of maximum likelihood estimation in Gaussian graphical models", which has been submitted for publication.

2.1 Introduction

Every positive definite $m \times m$ -matrix Σ is the covariance matrix of a multivariate normal distribution on \mathbb{R}^m . Its inverse matrix $K = \Sigma^{-1}$, the concentration matrix, is also positive definite. In this chapter, we study statistical models for multivariate normal distributions

on \mathbb{R}^m , where the concentration matrix can be written as a linear combination

$$K = \lambda_1 K_1 + \lambda_2 K_2 + \cdots + \lambda_d K_d \quad (2.1)$$

of some fixed linearly independent symmetric matrices K_1, K_2, \dots, K_d . Here, $\lambda_1, \lambda_2, \dots, \lambda_d$ are unknown real coefficients. It is assumed that K is positive definite for some choice of $\lambda_1, \lambda_2, \dots, \lambda_d$. Such statistical models, which we call *linear concentration models*, were introduced in [5].

Let \mathcal{L} be a linear subspace of \mathbb{S}^m . Then the set of all concentration matrices in the corresponding linear concentration model is a convex cone

$$\mathcal{K}_{\mathcal{L}} = \mathcal{L} \cap \mathbb{S}_{>0}^m.$$

By taking the inverse of every matrix in $\mathcal{K}_{\mathcal{L}}$, we get the set of all covariance matrices in the model. This set is denoted by $\mathcal{K}_{\mathcal{L}}^{-1}$. This is an algebraic variety intersected with the positive definite cone $\mathbb{S}_{>0}^m$.

Note that Gaussian graphical models and colored Gaussian graphical models are special cases where the linear subspace \mathcal{L} is defined by the underlying graph. Let $G = ([m], E)$ be a graph (resp. $\mathcal{G} = ([m], \mathcal{E})$ a colored graph). The corresponding linear subspaces, which we denote by \mathcal{L}_G (resp. $\mathcal{L}_{\mathcal{G}}$) are defined by

$$K \in \mathcal{L}_G \iff K_{ij} = 0 \quad \text{for all } (i, j) \notin E$$

for Gaussian graphical models, and

$$K \in \mathcal{L}_{\mathcal{G}} \iff \begin{cases} K_{\alpha\beta} = 0 & \text{for all } (\alpha, \beta) \notin E, \\ K_{\alpha\alpha} = K_{\beta\beta} & \text{for all } \alpha, \beta \in V_i, & \text{for all } i \in \{1, \dots, p\}, \\ K_{\alpha\beta} = K_{\gamma\delta} & \text{for all } (\alpha, \beta), (\gamma, \delta) \in E_j, & \text{for all } j \in \{1, \dots, q\} \end{cases}$$

for colored Gaussian graphical models. The corresponding set of concentration matrices is denoted by \mathcal{K}_G (resp. $\mathcal{K}_{\mathcal{G}}$) and the corresponding set of covariance matrices by \mathcal{K}_G^{-1} (resp. $\mathcal{K}_{\mathcal{G}}^{-1}$). See Figure 2.2 later in this chapter for a geometric representation of these sets.

Given a basis K_1, \dots, K_d of the subspace \mathcal{L} as in (2.1), the basic statistical problem is to estimate the parameters $\lambda_1, \dots, \lambda_d$ when n observations X_1, \dots, X_n are drawn from a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$, whose covariance matrix $\Sigma = K^{-1}$ is in the model $\mathcal{K}_{\mathcal{L}}^{-1}$. The n observations X_i and their mean \bar{X} are summarized in the sample covariance matrix

$$S = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T \in \mathbb{S}_{\geq 0}^m.$$

In our model, we make no assumptions on the mean vector μ , and we always use the sample mean \bar{X} as estimate for μ . Thus, we are precisely in the situation (1.2), and the log-likelihood

function for the linear concentration model (2.1) equals

$$\log \det(K) - \langle S, K \rangle = \log \det \left(\sum_{j=1}^d \lambda_j K_j \right) - \sum_{j=1}^d \lambda_j \langle S, K_j \rangle \quad (2.2)$$

times the constant $n/2$. This is a strictly concave function on the relatively open cone $\mathcal{K}_{\mathcal{L}}$. Linear concentration models are regular exponential families, where the scalars $\lambda_1, \dots, \lambda_d$ are the *canonical parameters* and $\langle S, K_1 \rangle, \dots, \langle S, K_d \rangle$ are the *sufficient statistics* of the exponential family (2.1). Therefore, if a maximum (i.e. the MLE) exists, it is attained by a unique matrix \hat{K} in $\mathbb{S}_{>0}^m \cap \mathcal{L}$. Its inverse $\hat{\Sigma} = \hat{K}^{-1}$ is uniquely determined by the equations (1.4), which in this case are

$$\langle \hat{\Sigma}, K_j \rangle = \langle S, K_j \rangle \quad \text{for } j = 1, 2, \dots, d. \quad (2.3)$$

Given a sample covariance matrix S , we define the *spectrahedron*

$$\text{fiber}_{\mathcal{L}}(S) = \left\{ \Sigma \in \mathbb{S}_{>0}^m \mid \langle \Sigma, K \rangle = \langle S, K \rangle \text{ for all } K \in \mathcal{L} \right\}.$$

For a Gaussian graphical model with underlying graph G this spectrahedron consists of all positive definite completions of S_G and is denoted

$$\text{fiber}_G(S) = \left\{ \Sigma \in \mathbb{S}_{>0}^m \mid \Sigma_G = S_G \right\}.$$

For a colored Gaussian graphical model with underlying graph \mathcal{G} the corresponding spectrahedron is defined analogously and denoted by $\text{fiber}_{\mathcal{G}}(S)$. With this notion, Theorem 1.5.1 can easily be generalized to linear concentration models and has the following geometric interpretation (which is also explained later in Figure 2.2).

Theorem 2.1.1. *For a linear concentration model defined by \mathcal{L} the MLEs $\hat{\Sigma}$ and \hat{K} exist for a given sample covariance matrix S if and only if $\text{fiber}_{\mathcal{L}}(S)$ is non-empty, in which case $\text{fiber}_{\mathcal{L}}(S)$ intersects $\mathcal{K}_{\mathcal{L}}^{-1}$ in exactly one point, namely the MLE $\hat{\Sigma}$.*

Note that if $\text{rank}(S) < m$ then it can happen that the fiber is empty, in which case the MLE does not exist for (\mathcal{L}, S) . Work of [15] and [9] addresses this issue for Gaussian graphical models where the underlying graph is a cycle; see Section 2.4 below for a geometric approach.

Our motivating statistical problem is to identify conditions on the pair (\mathcal{L}, S) that ensure the existence of the MLE. This will involve studying the geometry of the semi-algebraic set $\mathcal{K}_{\mathcal{L}}^{-1}$ and of the algebraic function $S \mapsto \hat{\Sigma}$ which takes a sample covariance matrix to its MLE in $\mathcal{K}_{\mathcal{L}}^{-1}$.

Example 2.1.2. We illustrate the concepts introduced so far by way of a small explicit example whose geometry is visualized in Figure 2.1. Let $m = d = 3$ and let \mathcal{L} be the real

vector space spanned by

$$K_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad K_2 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad K_3 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The linear concentration model (2.1) consists of all positive definite matrices of the form

$$K = \begin{pmatrix} \lambda_1 + \lambda_2 + \lambda_3 & \lambda_3 & \lambda_2 \\ \lambda_3 & \lambda_1 + \lambda_2 + \lambda_3 & \lambda_1 \\ \lambda_2 & \lambda_1 & \lambda_1 + \lambda_2 + \lambda_3 \end{pmatrix}. \quad (2.4)$$

Given a sample covariance matrix $S = (s_{ij})$, the sufficient statistics are

$$t_1 = \text{trace}(S) + 2s_{23}, \quad t_2 = \text{trace}(S) + 2s_{13}, \quad t_3 = \text{trace}(S) + 2s_{12}.$$

If $S \in \mathbb{S}_{>0}^3$ then $\text{fiber}_{\mathcal{L}}(S)$ is an open 3-dimensional convex body whose boundary is a cubic surface. This is the *spectrahedron* shown on the left in Figure 2.1. The MLE $\hat{\Sigma}$ is the unique matrix of maximum determinant in the spectrahedron $\text{fiber}_{\mathcal{L}}(S)$. Here is an explicit algebraic formula for the MLE $\hat{\Sigma} = (\hat{s}_{ij})$: First, the matrix entry \hat{s}_{33} is determined (e.g. using *Cardano's formula*) from the equation

$$\begin{aligned} 0 = & 240 \hat{s}_{33}^4 + (-32t_1 - 32t_2 - 192t_3) \hat{s}_{33}^3 + (-8t_1^2 + 16t_1t_2 + 16t_1t_3 - 8t_2^2 + 16t_2t_3 + 32t_3^2) \hat{s}_{33}^2 \\ & + (8t_1^3 - 8t_1^2t_2 - 8t_1t_2^2 + 8t_2^3) \hat{s}_{33} - 4t_1^3t_3 - 6t_1^2t_2^2 + 4t_1^2t_3^2 + 4t_1t_2^3 + 4t_1t_2^2t_3 + 4t_1^2t_2t_3 - t_2^4 \\ & - 4t_2^3t_3 + 4t_2^2t_3^2 - 8t_1t_2t_3^2 - t_1^4 + 4t_1^3t_2. \end{aligned}$$

Next, we read off \hat{s}_{23} from

$$\begin{aligned} -24(t_1^2 - 2t_1t_2 + t_2^2 - t_3^2) \hat{s}_{23} = & 120 \hat{s}_{33}^3 - (16t_1 + 16t_2 + 36t_3) \hat{s}_{33}^2 + (2t_1^2 - 4t_1t_2 + 2t_2^2 - 8t_3^2) \hat{s}_{33} \\ & - 6t_1^3 + 18t_1^2t_2 + t_1^2t_3 - 18t_1t_2^2 - 2t_1t_2t_3 + 10t_1t_3^2 + 6t_2^3 + t_2^2t_3 \\ & - 2t_2t_3^2 - 4t_3^3. \end{aligned}$$

Then we read off \hat{s}_{22} from

$$\begin{aligned} -24(t_1 - t_2) \hat{s}_{22} = & 60 \hat{s}_{33}^2 + (4t_1 - 20t_2 - 24t_3) \hat{s}_{33} + 24(t_1 - t_2 - t_3) \hat{s}_{23} - 11t_1^2 \\ & + 10t_1t_2 + 10t_1t_3 + t_2^2 - 2t_2t_3 - 4t_3^2. \end{aligned}$$

Finally, we obtain the first row of $\hat{\Sigma}$ as follows:

$$\hat{s}_{13} = \hat{s}_{23} - t_1/2 + t_2/2, \quad \hat{s}_{12} = \hat{s}_{23} - t_1/2 + t_3/2, \quad \hat{s}_{11} = t_1 - \hat{s}_{33} - 2\hat{s}_{23} - \hat{s}_{22}.$$

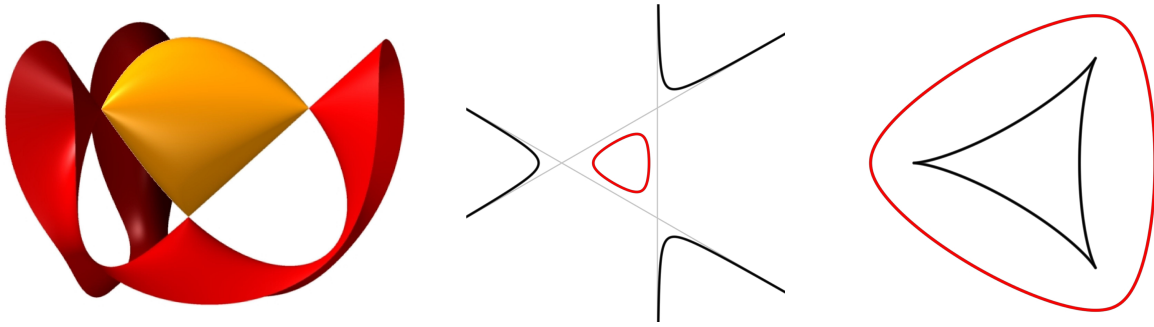


Figure 2.1: Three figures, taken from [58], illustrate Example 2.1.2. These figures show the spectrahedron fiber $\text{fiber}_{\mathcal{L}}(S)$ (left), a cross section of the spectrahedral cone $\mathcal{K}_{\mathcal{L}}$ (middle), and a cross section of its dual cone $\mathcal{C}_{\mathcal{L}}$ (right).

The MLE $\hat{\Sigma} = (\hat{s}_{ij})$ is an algebraic function of degree 4 in the sufficient statistics (t_1, t_2, t_3) . In short, the model (2.4) has ML degree 4. We identify our model with the subvariety \mathcal{L}^{-1} of projective space \mathbb{P}^5 that is parametrized by this algebraic function. The ideal of polynomials vanishing on \mathcal{L}^{-1} equals

$$\begin{aligned} P_{\mathcal{L}} = \langle & s_{13}^2 - s_{23}^2 - s_{11}s_{33} + s_{22}s_{33}, \quad s_{12}^2 - s_{11}s_{22} - s_{23}^2 + s_{22}s_{33}, \\ & s_{12}s_{13} - s_{13}s_{22} - s_{11}s_{23} + s_{12}s_{23} + s_{13}s_{23} + s_{23}^2 - s_{12}s_{33} - s_{22}s_{33}, \\ & s_{11}s_{13} - s_{13}s_{22} - s_{11}s_{23} + s_{22}s_{23} - s_{11}s_{33} - 2s_{12}s_{33} - s_{13}s_{33} - s_{22}s_{33} - s_{23}s_{33} + s_{33}^2, \\ & s_{11}s_{12} - s_{11}s_{22} - s_{12}s_{22} - 2s_{13}s_{22} + s_{22}^2 - s_{11}s_{23} - s_{22}s_{23} - s_{12}s_{33} - s_{22}s_{33} + s_{23}s_{33}, \\ & s_{11}^2 - 2s_{11}s_{22} - 4s_{13}s_{22} + s_{22}^2 - 4s_{11}s_{23} - 2s_{11}s_{33} - 4s_{12}s_{33} - 2s_{22}s_{33} + s_{33}^2 \rangle. \end{aligned}$$

The domain of the maximum likelihood map $(t_1, t_2, t_3) \mapsto \hat{\Sigma}$ is the cone of sufficient statistics $\mathcal{C}_{\mathcal{L}}$ in \mathbb{R}^3 . The polynomial $H_{\mathcal{L}}$ which vanishes on the boundary of this convex cone has degree six. It equals

$$\begin{aligned} H_{\mathcal{L}} = & t_1^6 - 6t_1^5t_2 + 19t_1^4t_2^2 - 28t_1^3t_2^3 + 19t_1^2t_2^4 - 6t_1t_2^5 + t_2^6 - 6t_1^5t_3 + 14t_1^4t_2t_3 - 24t_1^3t_2^2t_3 \\ & - 24t_1^2t_2^3t_3 + 14t_1t_2^4t_3 - 6t_2^5t_3 + 19t_1^4t_2^2t_3 - 24t_1^3t_2t_3^2 + 106t_1^2t_2^2t_3^2 - 24t_1t_2^3t_3^2 + 19t_2^4t_3^2 \\ & - 28t_1^3t_3^3 - 24t_1^2t_2t_3^3 - 24t_1t_2^2t_3^3 - 28t_2^3t_3^3 + 19t_1^2t_3^4 + 14t_1t_2t_3^4 + 19t_2^2t_3^4 - 6t_1t_3^5 - 6t_2t_3^5 + t_3^6. \end{aligned}$$

The sextic curve $\{H_{\mathcal{L}} = 0\}$ in \mathbb{P}^2 is shown on the right in Figure 2.1. This sextic is the dual curve which parametrizes all tangent lines to the cubic curve $\{\det(K) = 0\}$, which is shown in the middle of Figure 2.1. The cone over the convex region enclosed by the red part of that cubic curve is the set $\mathcal{K}_{\mathcal{L}} = \mathbb{S}_{>0}^3 \cap \mathcal{L}$ of concentration matrices in our model (2.1). \square

This chapter is organized as follows. In Section 2.2, we formally define the objects $\mathcal{C}_{\mathcal{L}}$, $P_{\mathcal{L}}$ and $H_{\mathcal{L}}$, which already appeared in Example 2.1.2. We give a geometric description

of the problem and derive three guiding questions that constitute the main thread of this chapter. These questions are answered for generic linear spaces \mathcal{L} in Subsection 2.2.3. In Section 2.3 we answer our three questions for diagonal concentration models, using results from geometric combinatorics. Section 2.4 deals with Gaussian graphical models, which are the most prominent linear concentration models. We resolve our three questions for chordal graphs, then for chordless cycles, and finally for wheels and all graphs with five or less vertices. We conclude this chapter with a study of colored Gaussian graphical models in Section 2.5.

2.2 Linear sections, projections and duality

Convex algebraic geometry is concerned with the geometry of real algebraic varieties and semi-algebraic sets that arise in convex optimization, especially in semidefinite programming. A fundamental problem is to study convex sets that arise as linear sections and projections of the cone of positive definite matrices $\mathbb{S}_{>0}^m$. This problem arises naturally when studying maximum likelihood estimation in linear concentration models for Gaussian random variables. In particular, the issue of estimating a covariance matrix from the sufficient statistics can be seen as an extension of the familiar semidefinite matrix completion problem (see Section 1.5 and [9, 35]). In what follows, we develop an algebraic and geometric framework for systematically addressing such problems.

2.2.1 Geometry of maximum likelihood estimation in linear concentration models

As before, we fix a linear subspace \mathcal{L} in the real vector space \mathbb{S}^m of symmetric $m \times m$ -matrices, and we fix a basis $\{K_1, \dots, K_d\}$ of \mathcal{L} . The *cone of concentration matrices* is the relatively open cone

$$\mathcal{K}_{\mathcal{L}} = \mathcal{L} \cap \mathbb{S}_{>0}^m.$$

We assume throughout that $\mathcal{K}_{\mathcal{L}}$ is non-empty. Using the basis K_1, \dots, K_d of \mathcal{L} , we can identify $\mathcal{K}_{\mathcal{L}}$ with

$$\mathcal{K}_{\mathcal{L}} = \left\{ (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d \mid \sum_{i=1}^d \lambda_i K_i \text{ is positive definite} \right\}. \quad (2.5)$$

This is a non-empty open convex cone in \mathbb{R}^d . The orthogonal complement \mathcal{L}^{\perp} of \mathcal{L} is a subspace of dimension $\binom{m+1}{2} - d$ in \mathbb{S}^m , so that $\mathbb{S}^m / \mathcal{L}^{\perp} \simeq \mathbb{R}^d$, and we can consider the canonical map

$$\pi_{\mathcal{L}} : \mathbb{S}^m \rightarrow \mathbb{S}^m / \mathcal{L}^{\perp}.$$

This is precisely the linear map which takes a sample covariance matrix S to its canonical sufficient statistics. For example, for a Gaussian graphical model with underlying graph G , this map takes a sample covariance matrix S to the partial sample covariance matrix S_G . The chosen basis of \mathcal{L} allows us to identify this map with

$$\pi_{\mathcal{L}} : \mathbb{S}^m \rightarrow \mathbb{R}^d, \quad S \mapsto (\langle S, K_1 \rangle, \dots, \langle S, K_d \rangle). \quad (2.6)$$

Applying Theorem 2.1.1, we can describe the set of all sufficient statistics, for which the MLE exists. We denote this set by $\mathcal{C}_{\mathcal{L}}$. It is the image of the positive-definite cone $\mathbb{S}_{>0}^m$ under the map $\pi_{\mathcal{L}}$. So $\mathcal{C}_{\mathcal{L}}$ is also a convex cone and shown in dark orange in Figure 2.2. We call $\mathcal{C}_{\mathcal{L}}$ the *cone of sufficient statistics*. For Gaussian graphical models (resp. colored Gaussian graphical models) we denote the projection by π_G (resp. π_G) and the image of the positive definite cone by \mathcal{C}_G (resp. \mathcal{C}_G).

The following result explains the duality between the two red curves in Figure 2.1 and is described in Figure 2.2 for Gaussian graphical models.

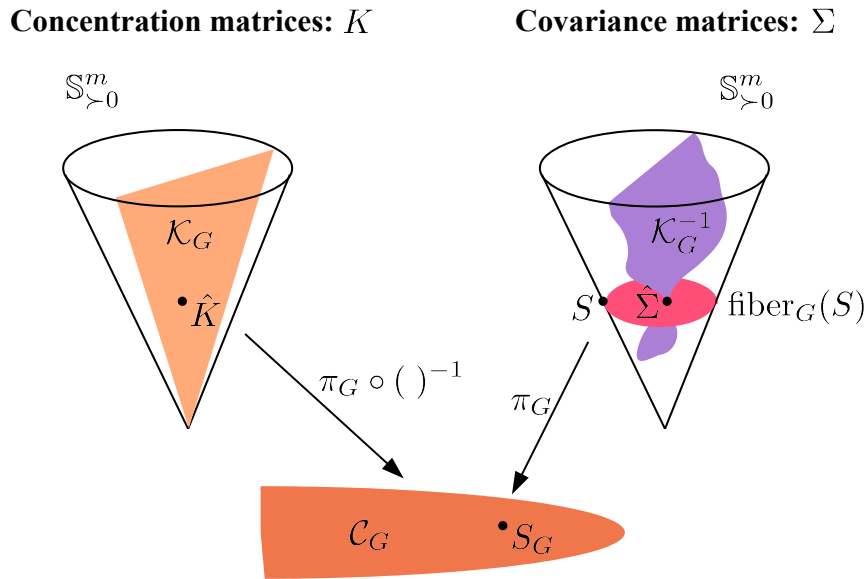


Figure 2.2: *Geometry of maximum likelihood estimation in Gaussian graphical models. The cone \mathcal{K}_G consists of all concentration matrices in the model and \mathcal{K}_G^{-1} is the corresponding set of covariance matrices. The cone of sufficient statistics \mathcal{C}_G is defined as the projection of $\mathbb{S}_{>0}^m$ onto the edge set of G . It is dual and homeomorphic to \mathcal{K}_G . Given a sample covariance matrix S , $\text{fiber}_G(S)$ consists of all positive definite completions of the G -partial matrix S_G , and it intersects \mathcal{K}_G^{-1} in at most one point, namely the MLE $\hat{\Sigma}$.*

Proposition 2.2.1. *The cone of sufficient statistics is the dual cone to the cone of concentration matrices. The basis-free version of this duality states*

$$\mathcal{C}_{\mathcal{L}} = \{S \in \mathbb{S}^m / \mathcal{L}^\perp \mid \langle S, K \rangle > 0 \text{ for all } K \in \mathcal{K}_{\mathcal{L}}\}. \quad (2.7)$$

The basis-dependent version of this duality, in terms of (2.5) and (2.6), states

$$\mathcal{C}_{\mathcal{L}} = \left\{ (t_1, \dots, t_d) \in \mathbb{R}^d \mid \sum_{i=1}^d t_i \lambda_i > 0 \text{ for all } (\lambda_1, \dots, \lambda_d) \in \mathcal{K}_{\mathcal{L}} \right\}. \quad (2.8)$$

Proof. Let $\mathcal{K}_{\mathcal{L}}^\vee$ denote the right-hand side of (2.7) and let $M = \binom{m+1}{2}$. Using the fact that the M -dimensional convex cone $\mathbb{S}_{>0}^m$ is self-dual, equation (1.1) implies

$$\mathcal{K}_{\mathcal{L}}^\vee = (\mathbb{S}_{>0}^m \cap \mathcal{L})^\vee = (\mathbb{S}_{>0}^m + \mathcal{L}^\perp) / \mathcal{L}^\perp = \mathcal{C}_{\mathcal{L}}.$$

To derive (2.8) from (2.7), we pick any basis U_1, \dots, U_M of \mathbb{S}^m whose first d elements serve as the dual basis to K_1, \dots, K_d , and whose last $\binom{m+1}{2} - d$ elements span \mathcal{L}^\perp . Hence $\langle U_i, K_j \rangle = \delta_{ij}$ for all i, j . Every matrix U in \mathbb{S}^m has a unique representation $U = \sum_{i=1}^M t_i U_i$, and its image under the map (2.6) equals $\pi_{\mathcal{L}}(U) = (t_1, \dots, t_d)$. For any matrix $K = \sum_{j=1}^d \lambda_j K_j$ in \mathcal{L} we have $\langle U, K \rangle = \sum_{i=1}^d t_i \lambda_i$, and this expression is positive for all $K \in \mathcal{K}_{\mathcal{L}}$ if and only if (t_1, \dots, t_d) lies in $\mathcal{C}_{\mathcal{L}}$. \square

It can be shown that both the cone $\mathcal{K}_{\mathcal{L}}$ of concentration matrices and its dual, the cone $\mathcal{C}_{\mathcal{L}}$ of sufficient statistics, are well-behaved in the following sense. Their topological closures are closed convex cones that are dual to each other, and they are obtained by respectively intersecting and projecting the closed cone $\mathbb{S}_{\geq 0}^m$ of positive semidefinite matrices. In symbols, these closed semi-algebraic cones satisfy

$$\overline{\mathcal{K}_{\mathcal{L}}} = \mathcal{L} \cap \mathbb{S}_{\geq 0}^m \quad \text{and} \quad \overline{\mathcal{C}_{\mathcal{L}}} = \pi_{\mathcal{L}}(\mathbb{S}_{\geq 0}^m). \quad (2.9)$$

One of our objectives will be to explore the geometry of their boundaries

$$\partial \mathcal{K}_{\mathcal{L}} := \overline{\mathcal{K}_{\mathcal{L}}} \setminus \mathcal{K}_{\mathcal{L}} \quad \text{and} \quad \partial \mathcal{C}_{\mathcal{L}} := \overline{\mathcal{C}_{\mathcal{L}}} \setminus \mathcal{C}_{\mathcal{L}}.$$

These are convex algebraic hypersurfaces in \mathbb{R}^d , as seen in Example 2.1.2. The statistical theory of exponential families implies the following corollary concerning the geometry of their interiors:

Corollary 2.2.2. *The map $K \mapsto T = \pi_{\mathcal{L}}(K^{-1})$ is a homeomorphism between the dual pair of open cones $\mathcal{K}_{\mathcal{L}}$ and $\mathcal{C}_{\mathcal{L}}$. The inverse map $T \mapsto K$ takes the sufficient statistics to the MLE of the concentration matrix. Here, K^{-1} is the unique maximizer of the determinant over the spectrahedron $\pi_{\mathcal{L}}^{-1}(T) \cap \mathbb{S}_{>0}^m$.*

The geometry of maximum likelihood estimation in linear concentration models described so far is summarized in Figure 2.2 for the special case of Gaussian graphical models. The geometric picture for linear concentration models is totally analogous.

2.2.2 Derivation of three guiding questions

One natural first step in studying this picture is to simplify it by passing to the complex numbers \mathbb{C} . This allows us to relax various inequalities over the real numbers \mathbb{R} and to work with varieties over the algebraically closed field \mathbb{C} . We thus identify our model \mathcal{L}^{-1} with its Zariski closure in the $\binom{m+1}{2} - 1$ -dimensional complex projective space $\mathbb{P}(\mathbb{S}^m)$. Let $P_{\mathcal{L}}$ denote the homogeneous prime ideal of all polynomials in $\mathbb{R}[\Sigma] = \mathbb{R}[s_{11}, s_{12}, \dots, s_{mm}]$ that vanish on \mathcal{L}^{-1} . One way to compute $P_{\mathcal{L}}$ is to eliminate the entries of an indeterminate symmetric $m \times m$ -matrix K from the following system of equations:

$$\Sigma \cdot K = \text{Id}_m, \quad K \in \mathcal{L}. \quad (2.10)$$

The fact that the ideal in (2.10) is prime can be seen using Proposition 23 (b) in [29].

Given a sample covariance matrix S , its maximum likelihood estimate $\hat{\Sigma}$ can be computed algebraically, as in Example 2.1.2. We do this by solving the following zero-dimensional system of polynomial equations:

$$\Sigma \cdot K = \text{Id}_m, \quad K \in \mathcal{L}, \quad \Sigma - S \in \mathcal{L}^{\perp}. \quad (2.11)$$

In this chapter we focus on the systems (2.10) and (2.11). Specifically, for various classes of linear concentration models \mathcal{L} , we seek to answer the following three guiding questions. Example 2.1.2 served to introduce these three questions. Many more examples will be featured throughout our discussion.

Question 1. What can be said about the geometry of the $(d-1)$ -dimensional projective variety \mathcal{L}^{-1} ? What is the *degree* of this variety, and what are the minimal generators of its prime ideal $P_{\mathcal{L}}$?

Question 2. The map taking a sample covariance matrix S to its maximum likelihood estimate $\hat{\Sigma}$ is an algebraic function. Its degree is the ML degree of the model \mathcal{L} . See [25, Def. 2.1.4]. The ML degree represents the algebraic complexity of the problem of finding the MLE. Can we find a formula for the ML degree? Which models \mathcal{L} have their ML degree equal to 1?

Question 3. The Zariski closure of the boundary $\partial\mathcal{C}_{\mathcal{L}}$ of the cone of sufficient statistics $\mathcal{C}_{\mathcal{L}}$ is a hypersurface in the complex projective space \mathbb{P}^{d-1} . What is the defining polynomial $H_{\mathcal{L}}$ of this hypersurface?

2.2.3 Generic linear concentration models

In this subsection we examine the case when \mathcal{L} is a generic subspace of dimension d in \mathbb{S}^m . Here “generic” is understood in the sense of algebraic geometry. In terms of the model representation (2.1), this means that the matrices K_1, \dots, K_d were chosen at random. This is precisely the hypothesis made by [58], and one of our goals is to explain the connection of Questions 1, 2 and 3 to that paper.

To begin with, we establish the result that the two notions of degree coincide in the generic case.

Theorem 2.2.3. *The ML degree of the model (2.1) defined by a generic linear subspace \mathcal{L} of dimension d in \mathbb{S}^m equals the degree of the projective variety \mathcal{L}^{-1} . That degree is denoted $\phi(m, d)$ and it satisfies*

$$\phi(m, d) = \phi\left(m, \binom{m+1}{2} + 1 - d\right). \quad (2.12)$$

We calculated the ML degree $\phi(m, d)$ of the generic model \mathcal{L} for all matrix sizes up to $m = 6$ (Table 2.1). This table was computed with the software `Macaulay2` [33] using the commutative algebra techniques discussed in the proof of Theorem 2.2.3. At this point, readers from statistics are advised to skip the algebraic technicalities in the rest of this section and to go straight to Section 2.4 on Gaussian graphical models.

Table 2.1: *ML-degree of generic model for all matrices up to size $m = 6$.*

d	1	2	3	4	5	6	7	8	9	10	11	12	13	...
$\phi(3, d)$	1	2	4	4	2	1								
$\phi(4, d)$	1	3	9	17	21	21	17	9	3	1				
$\phi(5, d)$	1	4	16	44	86	137	188	212	188	137	86	44	16	...
$\phi(6, d)$	1	5	25	90	240	528	1016	1696	2396	2886	3054	2886	2396	...

The last three entries in each row of Table 2.1 follow from Bézout’s Theorem because $P_{\mathcal{L}}$ is a complete intersection when the codimension of \mathcal{L}^{-1} in $\mathbb{P}(\mathbb{S}^m)$ is at most two. Using the duality relation (2.12), we conclude

$$\phi(m, d) = (m - 1)^{d-1} \quad \text{for } d = 1, 2, 3.$$

When \mathcal{L}^{-1} has codimension 3, it is the complete intersection defined by three generic linear combinations of the comaximal minors. From this complete intersection we must remove the variety of $m \times m$ -symmetric matrices of rank $\leq m - 2$, which has also codimension 3 and

has degree $\binom{m+1}{3}$. Hence:

$$\phi(m, 4) = (m-1)^3 - \binom{m+1}{3} = \frac{1}{6}(5m-3)(m-1)(m-2).$$

When d is larger than 4, this approach leads to a problem in *residual intersection theory*. A formula due to [64], rederived in recent work by [17] on this subject, implies that

$$\phi(m, 5) = \frac{1}{12}(m-1)(m-2)(7m^2 - 19m + 6).$$

For any fixed dimension d , our ML degree $\phi(m, d)$ seems to be a polynomial function of degree $d-1$ in m , but it gets progressively more challenging to compute explicit formulas for these polynomials.

Proof of Theorem 2.2.3. Let I be the ideal in the polynomial ring $\mathbb{R}[\Sigma] = \mathbb{R}[s_{11}, s_{12}, \dots, s_{mm}]$ that is generated by the $(m-1) \times (m-1)$ -minors of the symmetric $m \times m$ -matrix $\Sigma = (s_{ij})$. [46] proved that the Rees algebra $\mathcal{R}(I)$ of the ideal I is equal to the symmetric algebra of I . Identifying the generators of I with the entries of another symmetric matrix of unknowns $K = (k_{ij})$, we represent this Rees algebra as $\mathcal{R}(I) = \mathbb{R}[\Sigma, K]/J$ where the ideal J is obtained by eliminating the unknown t from the matrix equation $\Sigma \cdot K = t \cdot \text{Id}_m$. The presentation ideal $J = \langle \Sigma \cdot K - t \text{Id}_m \rangle \cap \mathbb{R}[\Sigma, K]$ is prime and it is homogeneous with respect to the natural \mathbb{N}^2 -grading on the polynomial ring $\mathbb{R}[\Sigma, K]$. Its variety $V(J)$ in $\mathbb{P}^{M-1} \times \mathbb{P}^{M-1}$ is the closure of the set of pairs of symmetric matrices that are inverse to each other. Here $M = \binom{m+1}{2}$. Both the dimension and the codimension of $V(J)$ are equal to $M-1$.

We now make use of the notion of multidegree introduced in the text book [57]. Namely, we consider the *bidegree* of the Rees algebra $\mathcal{R}(I) = \mathbb{R}[\Sigma, K]/J$ with respect to its \mathbb{N}^2 -grading. This bidegree is a homogeneous polynomial in two variables x and y of degree $M-1$. Using notation as in [57, Def. 8.45] and [58, Thm. 10], we claim that

$$\mathcal{C}(\mathcal{R}(I); x, y) = \sum_{d=1}^M \phi(m, d) x^{M-d} y^{d-1}. \quad (2.13)$$

Indeed, the coefficient of $x^{M-d} y^{d-1}$ in the expansion of $\mathcal{C}(\mathcal{R}(I); x, y)$ equals the cardinality of the finite variety $V(J) \cap (\mathcal{M} \times \mathcal{L})$, where \mathcal{L} is a generic plane of dimension $d-1$ in the second factor \mathbb{P}^{M-1} and \mathcal{M} is a generic plane of dimension $M-d$ in the first factor \mathbb{P}^{M-1} . We now take \mathcal{M} to be the specific plane which is spanned by the image of \mathcal{L}^\perp and one extra generic point S , representing a random sample covariance matrix. Thus our finite variety is precisely the same as the one described by the affine equations in (2.11), and we conclude that its cardinality equals the ML degree $\phi(m, d)$.

Note that $V(J) \cap (\mathbb{P}^{M-1} \times \mathcal{L})$ can be identified with the variety $V(P_{\mathcal{L}})$ in \mathbb{P}^{M-1} . The argument in the previous paragraph relied on the fact that $P_{\mathcal{L}}$ is Cohen-Macaulay, which

allowed us to choose any subspace \mathcal{M} for our intersection count provided it is disjoint from $V(P_{\mathcal{L}})$ in \mathbb{P}^{M-1} . This proves that $\phi(m, d)$ coincides with the degree of $V(P_{\mathcal{L}})$. The Cohen-Macaulay property of $P_{\mathcal{L}}$ follows from a result of [38] together with the aforementioned work of [46] which shows that the ideal I has sliding depth. Finally, the duality (2.12) is obvious for the coefficients of the bidegree (2.13) of the Rees algebra $\mathcal{R}(I)$ since its presentation ideal J is symmetric under swapping K and Σ . \square

We now come to our third question, which is to determine the Zariski closure $V(H_{\mathcal{L}})$ of the boundary of the cone $\mathcal{C}_{\mathcal{L}} = \mathcal{K}_{\mathcal{L}}^{\vee}$. Let us assume now that \mathcal{L} is any d -dimensional linear subspace of \mathbb{S}^m , not necessarily generic. The Zariski closure of $\partial\mathcal{K}_{\mathcal{L}}$ is the hypersurface $\{\det(K) = 0\}$ given by the vanishing of the determinant of $K = \sum_{i=1}^d \lambda_i K_i$. This determinant is a polynomial of degree d in $\lambda_1, \dots, \lambda_d$. Our task is to compute the *dual variety* in the sense of projective algebraic geometry of each irreducible component of this hypersurface. See [58, §5] for basics on projective duality. We also need to compute the dual variety for its singular locus, and for the singular locus of the singular locus, etc.

Each singularity stratum encountered along the way needs to be decomposed into irreducible components, whose duals need to be examined. If such a component has a real point that lies in $\partial\mathcal{K}_{\mathcal{L}}$ and if its dual variety is a hypersurface then that hypersurface appears in $H_{\mathcal{L}}$. How to run this procedure in practice is shown in Example 2.4.10. For now, we summarize the construction informally as follows.

Proposition 2.2.4. *Each irreducible hypersurface in the Zariski closure of $\partial\mathcal{C}_{\mathcal{L}}$ is the projectively dual variety to some irreducible component of the hypersurface $\{\det(K) = 0\}$, or it is dual to some irreducible variety further down in the singularity stratification of the hypersurface $\{\det(K) = 0\} \subset \mathbb{P}^{d-1}$.*

The singularity stratification of $\{\det(K) = 0\}$ can be computed by applying primary decomposition to the ideal of $p \times p$ -minors of K for $1 \leq p \leq m$. If I is any minimal prime of such a determinantal ideal then its dual variety is computed as follows. Let $c = \text{codim}(I)$ and consider the Jacobian matrix of I . The rows of the Jacobian matrix are the derivatives of the generators of I with respect to the unknowns $\lambda_1, \dots, \lambda_d$. Let J be the ideal generated by I and the $c \times c$ -minors of the matrix formed by augmenting the Jacobian matrix by the extra row (t_1, t_2, \dots, t_d) . We saturate J by the $c \times c$ -minors of the Jacobian, and thereafter we compute the elimination ideal $J \cap \mathbb{R}[t_1, t_2, \dots, t_d]$. If this elimination ideal is principal, we retain its generator. The desired polynomial $H_{\mathcal{L}}$ is the product of these principal generators, as I runs over all such minimal primes whose variety has a real point on the convex hypersurface $\partial\mathcal{K}_{\mathcal{L}}$.

Proposition 2.2.4 is visualized in Figure 2.5 later in this chapter. Let us now apply this result in the case when the subspace \mathcal{L} is generic of dimension d . The ideal of $p \times p$ -minors of K defines a subvariety of \mathbb{P}^{d-1} , which is irreducible whenever it is positive-dimensional (by Bertini's Theorem, e.g. in [37]). It is known from [58, Prop. 5] that the dual variety to that

determinantal variety is a hypersurface if and only if

$$\binom{m-p+2}{2} \leq d-1 \quad \text{and} \quad \binom{p}{2} \leq \binom{m+1}{2} - d + 1. \quad (2.14)$$

Assuming that these inequalities hold, the dual hypersurface is defined by an irreducible homogeneous polynomial whose degree we denote by $\delta(d-1, m, p-1)$. This notation is consistent with [58] where this number is called the *algebraic degree of semidefinite programming (SDP)*.

Corollary 2.2.5. *For a generic d -dimensional subspace \mathcal{L} of \mathbb{S}^m , the polynomial $H_{\mathcal{L}}$ is the product of irreducible polynomials of degree $\delta(d-1, m, p-1)$. That number is the algebraic degree of semidefinite programming. Here p runs over integers that satisfy (2.14) and $\partial\mathcal{K}_{\mathcal{L}}$ contains a matrix of rank $p-1$.*

2.3 Diagonal matrices, matroids and polytopes

This section concerns the case when \mathcal{L} is a d -dimensional space consisting only of diagonal matrices in \mathbb{S}^m . Here, the set $\mathcal{L}_{>0}^{-1}$ of covariance matrices in the model also consists of diagonal matrices only, and we may restrict our considerations to the space \mathbb{R}^m of diagonal matrices in \mathbb{S}^m . Thus, throughout this section, our ambient space is \mathbb{R}^m , and we identify \mathbb{R}^m with its dual vector space via the standard inner product $\langle u, v \rangle = \sum_{i=1}^m u_i v_i$. We fix any $d \times m$ -matrix A whose row space equals \mathcal{L} , and we assume that $\mathcal{L} \cap \mathbb{R}_{>0}^m \neq \emptyset$. We consider the induced projection of the open positive orthant

$$\pi : \mathbb{R}_{>0}^m \rightarrow \mathbb{R}^d, \quad x \mapsto Ax. \quad (2.15)$$

Since $\mathcal{L} = \text{rowspace}(A)$ contains a strictly positive vector, the image of π is a pointed polyhedral cone, namely $\mathcal{C}_{\mathcal{L}} = \text{pos}(A)$ is the cone spanned by the columns of A . Each fiber of π is a bounded convex polytope, and maximum likelihood estimation amounts to finding a distinguished point \hat{x} in that fiber.

The problem of characterizing the existence of the MLE in this situation amounts to a standard problem of *geometric combinatorics* (see e.g. [80]), namely, to computing the facet description of the convex polyhedral cone spanned by the columns of A . For a given vector $t \in \mathbb{R}^d$ of sufficient statistics, the maximum likelihood estimate exists in this diagonal concentration model if and only if t lies in the interior of the cone $\text{pos}(A)$. This happens if and only if all facet inequalities are strict for t .

This situation is reminiscent of *Birch's Theorem* for toric models in algebraic statistics [59, Theorem 1.10], and, indeed, the combinatorial set-up for deciding the existence of the MLE is identical to that for toric models. For a statistical perspective see [27]. However, the algebraic structure here is not that of toric models, described in [59, §1.2.2], but that of the linear models in [59, §1.2.1].

Our model here is not toric but it is the coordinatewise reciprocal of an open polyhedral cone:

$$\mathcal{L}_{>0}^{-1} = \{ u \in \mathbb{R}_{>0}^m \mid u^{-1} = (u_1^{-1}, u_2^{-1}, \dots, u_m^{-1}) \in \mathcal{L} \}.$$

As in Section 2, we view its Zariski closure \mathcal{L}^{-1} as a subvariety in complex projective space:

$$\mathcal{L}^{-1} = \{ u \in \mathbb{P}^{m-1} \mid u^{-1} = (u_1^{-1}, u_2^{-1}, \dots, u_m^{-1}) \in \mathcal{L} \}.$$

Maximum likelihood estimation means intersecting the variety \mathcal{L}^{-1} with the fibers of π .

Example 2.3.1. Let $m = 4$, $d = 2$ and take \mathcal{L} to be the row space of the matrix

$$A = \begin{pmatrix} 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 \end{pmatrix}.$$

The corresponding statistical model consists of all multivariate normal distributions on \mathbb{R}^4 whose concentration matrix has the diagonal form

$$K = \begin{pmatrix} 3\lambda_1 & 0 & 0 & 0 \\ 0 & 2\lambda_1 + \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_1 + 2\lambda_2 & 0 \\ 0 & 0 & 0 & 3\lambda_2 \end{pmatrix}.$$

Our variety \mathcal{L}^{-1} is the curve in \mathbb{P}^3 parametrized by the inverse diagonal matrices which we write as $K^{-1} = \text{diag}(x_1, x_2, x_3, x_4)$. The prime ideal $P_{\mathcal{L}}$ of this curve is generated by three quadratic equations:

$$x_2x_3 - 2x_2x_4 + x_3x_4 = 2x_1x_3 - 3x_1x_4 + x_3x_4 = x_1x_2 - 3x_1x_4 + 2x_2x_4 = 0.$$

Consider any sample covariance matrix $S = (s_{ij})$, with sufficient statistics

$$t_1 = 3s_{11} + 2s_{22} + s_{33} > 0 \quad \text{and} \quad t_2 = s_{22} + 2s_{33} + 3s_{44} > 0.$$

The MLE for these sufficient statistics is the unique positive solution \hat{x} of the three quadratic equations above, together with the two linear equations

$$3x_1 + 2x_2 + x_3 = t_1 \quad \text{and} \quad x_2 + 2x_3 + 3x_4 = t_2.$$

We find that \hat{x} is an algebraic function of degree 3 in the sufficient statistics (t_1, t_2) , so the ML degree of the model K equals 3. This is consistent with formula (2.16) below, since $\binom{4-1}{2-1} = 3$. \square

We now present the solutions to our three guiding problems for arbitrary d -dimensional subspaces \mathcal{L} of the space \mathbb{R}^m of $m \times m$ -diagonal matrices. The degree of the projective

variety \mathcal{L}^{-1} and its prime ideal $P_{\mathcal{L}}$ are known from work of [69] and its refinements due to [61]. Namely, the degree of \mathcal{L}^{-1} equals the *beta-invariant*¹, of the rank $m - d$ matroid on $[m] = \{1, 2, \dots, m\}$ associated with \mathcal{L} . We denote this beta-invariant by $\beta(\mathcal{L})$. For matroid basics see [73].

The beta-invariant $\beta(\mathcal{L})$ is known to equal the number of bounded regions in the $(m - d)$ -dimensional hyperplane arrangement (cf. [78]) obtained by intersecting the affine space $u + \mathcal{L}^{\perp}$ with the m coordinate hyperplanes $\{x_i = 0\}$. Here u can be any generic vector in $\mathbb{R}_{>0}^m$. One of these regions, namely the one containing u , is precisely the fiber of π . If \mathcal{L} is a generic d -dimensional linear subspace of \mathbb{R}^m , meaning that the above matroid is the uniform matroid, then the beta-invariant equals

$$\beta(\mathcal{L}) = \binom{m-1}{d-1}. \quad (2.16)$$

For non-generic subspaces \mathcal{L} , this binomial coefficient is always an upper bound for $\beta(\mathcal{L})$.

Theorem 2.3.2. ([61, 69]) *The degree of the projective variety \mathcal{L}^{-1} equals the beta-invariant $\beta(\mathcal{L})$. Its prime ideal $P_{\mathcal{L}}$ is generated by the homogeneous polynomials*

$$\sum_{i \in \text{supp}(v)} v_i \cdot \prod_{j \neq i} x_j \quad (2.17)$$

where v runs over all non-zero vectors of minimal support in \mathcal{L}^{\perp} .

For experts in combinatorial commutative algebra, we note that [61] actually proves the following stronger results. The homogeneous polynomials (2.17) form a *universal Gröbner basis* of $P_{\mathcal{L}}$. The initial monomial ideal of $P_{\mathcal{L}}$ with respect to any term order is the Stanley-Reisner ideal of the corresponding *broken circuit complex* of the matroid of \mathcal{L} . Hence the Hilbert series of $P_{\mathcal{L}}$ is the rational function obtained by dividing the h -polynomial $h(t)$ of the broken circuit complex by $(1 - t)^d$. In particular, the degree of $P_{\mathcal{L}}$ is the number $h(1) = \beta(\mathcal{L})$ of broken circuit bases [73, §7].

We next consider Question 2 in the diagonal case. The maximum likelihood map takes each vector t in the cone of sufficient statistics $\mathcal{C}_{\mathcal{L}} = \text{pos}(A)$ to a point of its fiber, namely:

$$\hat{x} = \operatorname{argmax} \left\{ \sum_{i=1}^m \log(x_i) \mid x \in \mathbb{R}_{>0}^m \text{ and } Ax = t \right\}. \quad (2.18)$$

This is the unique point in the polytope $\pi^{-1}(t) = \{x \in \mathbb{R}_{>0}^m \text{ and } Ax = t\}$ which maximizes the product $x_1 x_2 \cdots x_n$ of the coordinates. It is also the unique point in $\pi^{-1}(t)$ that lies in the reciprocal linear variety \mathcal{L}^{-1} . In the linear programming literature, the point \hat{x} is known as the *analytic center* of the polytope $\pi^{-1}(t)$. In Section 2 we discussed the extension of this concept from linear programming to semidefinite programming: the *analytic center of a*

¹Note that the beta-invariant is called Moebius invariant in [20].

spectrahedron is the unique point $\hat{\Sigma}$ at which the determinant function attains its maximum. For an applied perspective see [70].

For any linear subspace \mathcal{L} in \mathbb{S}^m , the algebraic degree of the maximum likelihood map $t \mapsto \hat{\Sigma}$ is always less than or equal to the degree of the projective variety \mathcal{L}^{-1} . We saw in Theorem 2.2.3 that these degrees are equal for generic \mathcal{L} . We next show that the same conclusion holds for diagonal subspaces \mathcal{L} .

Corollary 2.3.3. *The ML degree of any diagonal linear concentration model $\mathcal{L} \subset \mathbb{R}^m \subset \mathbb{S}^m$ is equal to the beta-invariant $\beta(\mathcal{L})$ of the corresponding matroid of rank $m - d$ on $\{1, 2, \dots, m\}$.*

Proof. The beta-invariant $\beta(\mathcal{L})$ counts the bounded regions in the arrangement of hyperplanes arising from the given facet description of the polytope $\pi^{-1}(t)$. *Varchenko's Formula* for linear models, derived in [59, Theorem 1.5], states that the optimization problem (2.18) has precisely one real critical point in each bounded region, and that there are no other complex critical points. \square

A fundamental question regarding the ML degree of any class of algebraic statistical models is to characterize those models which have ML degree one. These are the models whose maximum likelihood estimator is a rational function in the sufficient statistics [25, §2.1]. In the context here, we have the following characterization of matroids whose beta-invariant $\beta(\mathcal{L})$ equals one.

Corollary 2.3.4. *The ML degree $\beta(\mathcal{L})$ of a diagonal linear concentration model \mathcal{L} is equal to one if and only if the matroid of \mathcal{L} is the graphic matroid of a series-parallel graph.*

Proof. The equivalence of series-parallel and $\beta = 1$ first appeared in [14, Theorem 7.6]. \square

We now come to Question 3 which concerns the duality of convex cones in Proposition 2.2.1. In the diagonal case, the geometric view on this duality is as follows. The cone of sufficient statistics equals $\mathcal{C}_{\mathcal{L}} = \text{pos}(A)$ and its convex dual is the cone $\mathcal{K}_{\mathcal{L}} = \text{rowspan}(A) \cap \mathcal{L}$. Both cones are convex, polyhedral, pointed, and have dimension d . By passing to their cross sections with suitable affine hyperplanes, we can regard the two cones $\mathcal{C}_{\mathcal{L}}$ and $\mathcal{K}_{\mathcal{L}}$ as a dual pair of $(d - 1)$ -dimensional convex polytopes.

The hypersurface $\{\det(K) = 0\}$ is a union of m hyperplanes. The strata in its singularity stratification, discussed towards the end of Section 2, correspond to the various faces F of the polytope $\mathcal{K}_{\mathcal{L}}$. The dual variety to a face F is the complementary face of the dual polytope $\mathcal{C}_{\mathcal{L}}$, and hence the codimension of that dual variety equals one if and only if F is a vertex (= 0-dimensional face) of $\mathcal{K}_{\mathcal{L}}$. This confirms that the polynomial $H_{\mathcal{L}}$ sought in Question 3 is the product of all facet-defining linear forms of $\mathcal{C}_{\mathcal{L}}$.

Corollary 2.2.2 furnishes a homeomorphism $u \mapsto Au^{-1}$ from the interior of the polytope $\mathcal{K}_{\mathcal{L}}$ onto the interior of its dual polytope $\mathcal{C}_{\mathcal{L}}$. The inverse to the rational function $u \mapsto Au^{-1}$ is an algebraic function whose degree is the beta-invariant $\beta(\mathcal{L})$. This homeomorphism is

the natural generalization, from simplices to arbitrary polytopes, of the classical *Cremona transformation* of projective geometry. We close this section with a nice 3-dimensional example which illustrates this homeomorphism.

Example 2.3.5 (How to morph a cube into an octahedron). Fix $m = 8$, $d = 4$, and \mathcal{L} the row space of

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

We identify the cone $\mathcal{K}_{\mathcal{L}} = \text{rowspace}(A) \cap \mathbb{R}_{>0}^6$ with $\{\lambda \in \mathbb{R}^4 : \lambda \cdot A > 0\}$. This is the cone over the 3-cube, which is obtained by setting $\lambda_4 = 1$. The dual cone $\mathcal{C}_{\mathcal{L}} = \text{pos}(A)$ is spanned by the six columns of the matrix A . It is the cone over the octahedron, which is likewise obtained by setting $t_4 = 1$.

We write the homeomorphism $u \mapsto Au^{-1}$ between these two four-dimensional cones in terms of the coordinates of λ and t . Explicitly, the equation $t = A \cdot (\lambda A)^{-1}$ translates into the scalar equations:

$$\begin{aligned} t_1 &= \frac{1}{\lambda_4 + \lambda_1} - \frac{1}{\lambda_4 - \lambda_1}, \\ t_2 &= \frac{1}{\lambda_4 + \lambda_2} - \frac{1}{\lambda_4 - \lambda_2}, \\ t_3 &= \frac{1}{\lambda_4 + \lambda_3} - \frac{1}{\lambda_4 - \lambda_3}, \\ t_4 &= \frac{1}{\lambda_4 + \lambda_1} + \frac{1}{\lambda_4 - \lambda_1} + \frac{1}{\lambda_4 + \lambda_2} + \frac{1}{\lambda_4 - \lambda_2} + \frac{1}{\lambda_4 + \lambda_3} + \frac{1}{\lambda_4 - \lambda_3}. \end{aligned}$$

Substituting $\lambda_4 = 1$, we get the bijection $(\lambda_1, \lambda_2, \lambda_3) \mapsto (t_1/t_4, t_2/t_4, t_3/t_4)$ between the open cube $(-1, +1)^3$ and the open octahedron $\{t \in \mathbb{R}^3 : |t_1| + |t_2| + |t_3| < 1\}$. The inverse map $t \mapsto \lambda$ is an algebraic function of degree $\beta(\mathcal{L}) = 7$. That the ML degree of this model is 7 can be seen as follows. The fibers $\pi^{-1}(t)$ are the convex polygons which can be obtained from a regular hexagon by parallel displacement of its six edges. The corresponding arrangement of six lines has 7 bounded regions. \square

2.4 Gaussian graphical models

A Gaussian graphical model arises when the subspace \mathcal{L} of \mathbb{S}^m is defined by the vanishing of some off-diagonal entries of the concentration matrix K . We fix a graph $G = ([m], E)$ with vertex set $[m] = \{1, 2, \dots, m\}$ and whose edge set E is assumed to contain all self-loops. A basis for \mathcal{L} is the set $\{K_{ij} \mid (i, j) \in E\}$ of matrices K_{ij} with a single 1-entry in position (i, j) and 0-entries in all other positions. As noted earlier, we shall use the notation $\mathcal{K}_G, \mathcal{C}_G, P_G$ for

the objects $\mathcal{K}_{\mathcal{L}}, \mathcal{C}_{\mathcal{L}}, \mathcal{P}_{\mathcal{L}}$, respectively. Given a sample covariance matrix S , the set $\text{fiber}_G(S)$ consists of all positive definite matrices $\Sigma \in \mathbb{S}_{>0}^m$ with

$$\Sigma_{ij} = S_{ij} \quad \text{for all } (i, j) \in E.$$

The cone of concentration matrices \mathcal{K}_G is important for semidefinite matrix completion problems. Its closure was denoted \mathcal{P}_G by [47, 48]. The dual cone \mathcal{C}_G consists of all partial matrices $T \in \mathbb{R}^E$ with entries in positions $(i, j) \in E$, which can be completed to a positive definite matrix. So as noted already in Subsection 1.5.1, maximum likelihood estimation in Gaussian graphical models corresponds to the classical positive definite matrix completion problem [8, 9, 35, 47], and the ML degree of a Gaussian graphical model is the algebraic complexity of the corresponding positive definite matrix completion problem. In this section we investigate our three guiding questions, first for chordal graphs, next for the chordless m -cycle C_m , then for all graphs with five or less vertices, and finally for the m -wheel W_m .

2.4.1 Chordal graphs

A graph G is *chordal* if every induced m -cycle in G for $m \geq 4$ has a chord. Theorem 1.3.1 due to [35] fully resolves Question 3 when G is chordal. Namely, a partial matrix $T \in \mathbb{R}^E$ lies in the cone \mathcal{C}_G if and only if all principal minors T_{CC} indexed by cliques C in G are positive definite. The “only if” direction in this statement is true for all graphs G , but the “if” direction holds only when G is chordal. This result is equivalent to the characterization of chordal graphs as those that have *sparsity order* equal to one, i.e., all extreme rays of \mathcal{K}_G are matrices of rank one. We refer to [3] and [48] for details. From this characterization of chordal graphs in terms of sparsity order, we infer the following description of the Zariski closure of the boundary of \mathcal{C}_G .

Proposition 2.4.1. *For a chordal graph G , the defining polynomial H_G of $\partial\mathcal{C}_G$ is equal to*

$$H_G = \prod_{\substack{C \text{ maximal} \\ \text{clique of } G}} \det(T_{CC}).$$

We now turn to Question 2 regarding the ML degree of a Gaussian graphical model G . This number is here simply denoted by $\text{ML-degree}(G)$. Every chordal graph is a clique sum of complete graphs. We shall prove that the ML degree is multiplicative with respect to taking clique sums.

Lemma 2.4.2. *Let G be a clique sum of n graphs G_1, \dots, G_n . Then the following equality holds:*

$$\text{ML-degree}(G) = \prod_{i=1}^n \text{ML-degree}(G_i).$$

Proof. We first prove this statement for $n = 2$. Let G be a graph which can be decomposed in disjoint subsets (A, B, C) of the vertex set V , such that C is a clique and separates A from B . Let $G_{[W]}$ denote the induced subgraph on a vertex subset $W \subset V$. So, we wish to prove:

$$\text{ML-degree}(G) = \text{ML-degree}(G_{[A \cup C]}) \cdot \text{ML-degree}(G_{[B \cup C]}). \quad (2.19)$$

Given a generic matrix $S \in \mathbb{S}^m$, we fix $\Sigma \in \mathbb{S}^m$ with entries $\Sigma_{ij} = S_{ij}$ for $(i, j) \in E$ and unknowns $\Sigma_{ij} = z_{ij}$ for $(i, j) \notin E$. The ML degree of G is the number of complex solutions to the equations

$$(\Sigma^{-1})_{ij} = 0 \quad \text{for all } (i, j) \notin E. \quad (2.20)$$

Let $K = \Sigma^{-1}$ and denote by $K^1 = (\Sigma_{[A \cup C]})^{-1}$ (respectively, $K^2 = (\Sigma_{[B \cup C]})^{-1}$) the inverse of the submatrix of Σ corresponding to the induced subgraph on $A \cup C$ (respectively, $B \cup C$). Using Schur complements, we can see that these matrices are related by the following block structure:

$$K = \begin{pmatrix} K_{AA}^1 & K_{AC}^1 & 0 \\ K_{CA}^1 & K_{CC}^1 & K_{CB}^2 \\ 0 & K_{BC}^2 & K_{BB}^2 \end{pmatrix}, \quad K^1 = \begin{pmatrix} K_{AA}^1 & K_{AC}^1 \\ K_{CA}^1 & K_{CC}^1 \end{pmatrix}, \quad K^2 = \begin{pmatrix} K_{CC}^2 & K_{CB}^2 \\ K_{BC}^2 & K_{BB}^2 \end{pmatrix}.$$

This block structure reveals that, when solving the system (2.20), one can solve for the variables z_{ij} corresponding to missing edges in the subgraph $A \cup C$ independently from the variables over $B \cup C$ and $A \cup B$. This implies the equation (2.19). Induction yields the theorem for $n \geq 3$. \square

The following theorem characterizes chordal graphs in terms of their ML degree. It extends the equivalence of parts (iii) and (iv) in [25, Thm. 3.3.5] from discrete to Gaussian models.

Theorem 2.4.3. *A graph G is chordal if and only if $\text{ML-degree}(G) = 1$.*

Proof. The only-if-direction follows from Lemma 2.4.2 since every chordal graph is a clique sum of complete graphs, and a complete graph trivially has ML degree one. For the if-direction suppose that G is a graph that is not chordal. Then G contains the chordless cycle C_m as an induced subgraph for some $m \geq 4$. It is easy to see that the ML degree of any graph is bounded below by that of any induced subgraph. Hence what we must prove is that the chordless cycle C_m has strictly positive ML degree. This is precisely the content of Lemma 2.4.7 below. \square

We now come to Question 1 which concerns the homogeneous prime ideal P_G that defines the Gaussian graphical model as a subvariety of $\mathbb{P}(\mathbb{S}^m)$. Fix a symmetric $m \times m$ -matrix of unknowns $\Sigma = (s_{ij})$ and let Σ_{ij} denote the comaximal minor obtained by deleting the i th row and the j th column from Σ . We shall define several ideals in $\mathbb{R}[\Sigma]$ that approximate P_G .

The first is the saturation

$$P'_G = (I : J^\infty) \quad \text{where} \quad I = \langle \det(\Sigma_{ij}) \mid (i, j) \in E \rangle \quad \text{and} \quad J = \langle \det(\Sigma) \rangle. \quad (2.21)$$

This ideal is contained in the desired prime ideal, i.e. $P'_G \subseteq P_G$. The two ideals have the same radical, but it might happen that they are not equal. One disadvantage of the ideal P'_G is that the saturation step (2.21) is computationally expensive and terminates only for very small graphs.

A natural question is whether the prime ideal P_G can be constructed easily from the prime ideals P_{G_1} and P_{G_2} when G is a clique sum of two smaller graphs G_1 and G_2 . As in the proof of Lemma 2.4.2, we partition $[m] = A \cup B \cup C$, where G_1 is the induced subgraph on $A \cup C$, and G_2 is the induced subgraph on $B \cup C$. If $|C| = c$ then we say that G is a *c-clique sum* of G_1 and G_2 .

The following ideal is contained in P_G and defines the same algebraic variety in the open cone $\mathbb{S}_{>0}^m$:

$$P_{G_1} + P_{G_2} + \langle (c+1) \times (c+1)\text{-minors of } \Sigma_{A \cup C, B \cup C} \rangle. \quad (2.22)$$

One might guess that (2.22) is equal to P_G , at least up to radical, but this fails for $c \geq 2$. Indeed we shall see in Example 2.4.5 that the variety of (2.22) can have extraneous components on the boundary $\mathbb{S}_{\geq 0}^m \setminus \mathbb{S}_{> 0}^m$ of the semidefinite cone. We do conjecture, however, that this equality holds for $c \leq 1$. This is easy to prove for $c = 0$ when G is disconnected and is the disjoint union of G_1 and G_2 . The case $c = 1$ is considerably more delicate. At present, we do not have a proof that (2.22) is prime for $c = 1$, but we believe that even a lexicographic Gröbner basis for P_G can be built by taking the union of such Gröbner basis for P_{G_1} and P_{G_2} with the 2×2 -minors of $\Sigma_{A \cup C, B \cup C}$. This conjecture would imply the following.

Conjecture 2.4.4. *The prime ideal P_G of an undirected Gaussian graphical model is generated in degree ≤ 2 if and only if each connected component of the graph G is a 1-clique sum of complete graphs. In this case, P_G has a Gröbner basis consisting of entries of Σ and 2×2 -minors of Σ .*

This conjecture is an extension of the results and conjectures for (directed) trees in [66, §5]. Formulas for the degree of P_G when G is a tree are found in [66, Corollaries 5.5 and 5.6]. The “only if” direction in the first sentence of Conjecture 2.4.4 can be shown as follows. If G is not chordal then it contains an m -cycle ($m \geq 4$) as an induced subgraph, and, this gives rise to cubic generators for P_G , as seen in Subsection 4.2 below. If G is chordal but is not a 1-clique sum of complete graphs, then its decomposition involves a c -clique sum for some $c \geq 2$, and the right hand side of (2.22) contributes a minor of size $c + 1 \geq 3$ to the minimal generators of P_G . The algebraic structure of chordal graphical models is more delicate in the Gaussian case than in the discrete case, and there is no Gaussian analogue to the characterizations of chordality in (i) and (ii) of [25, Theorem 3.3.5]. This is highlighted by the following example which was suggested to us by Seth Sullivant.

Example 2.4.5. Let G be the graph on $m = 7$ vertices consisting of the triangles $\{i, 6, 7\}$ for $i = 1, 2, 3, 4, 5$. Then G is chordal because it is the 2-clique sum of these five triangles. The ideal P_G is minimally generated by 105 cubics and one quintic. The cubics are spanned by the 3×3 -minors of the matrices $\Sigma_{A \cup C, B \cup C}$ where $C = \{6, 7\}$ and $\{A, B\}$ runs over all unordered partitions of $\{1, 2, 3, 4, 5\}$. These minors do not suffice to define the variety $V(P_G)$ set-theoretically. For instance, they vanish whenever the last two rows and columns of Σ are zero. The additional quintic generator of P_G equals

$$\begin{aligned} & s_{12}s_{13}s_{24}s_{35}s_{45} - s_{12}s_{13}s_{25}s_{34}s_{45} - s_{12}s_{14}s_{23}s_{35}s_{45} + s_{12}s_{14}s_{25}s_{34}s_{35} \\ & + s_{12}s_{15}s_{23}s_{34}s_{45} - s_{12}s_{15}s_{24}s_{34}s_{35} + s_{13}s_{14}s_{23}s_{25}s_{45} - s_{13}s_{14}s_{24}s_{25}s_{35} \\ & - s_{13}s_{15}s_{23}s_{24}s_{45} + s_{13}s_{15}s_{24}s_{25}s_{34} + s_{14}s_{15}s_{23}s_{24}s_{35} - s_{14}s_{15}s_{23}s_{25}s_{34}. \end{aligned}$$

This polynomial is the *pentad* which is relevant for factor analysis [25, Example 4.2.8]. \square

Given an undirected graph G on $[m]$, we define its *Sullivant-Talaska ideal* ST_G to be the ideal in $\mathbb{R}[\Sigma]$ that is generated by the following collection of minors of Σ . For any submatrix $\Sigma_{A,B}$ we include in ST_G all $c \times c$ -minors of $\Sigma_{A,B}$ provided c is the smallest cardinality of a set C of vertices that separates A from B in G . Here, A, B and C need not be disjoint, and separation means that any path from a node in A to a node in B must pass through a node in C . In [67] it is shown that the generators of ST_G are precisely those subdeterminants of Σ that lie in P_G , and both ideals cut out the same variety in the positive definite cone $\mathbb{S}_{>0}^m$. However, generally their varieties differ on the boundary of that cone, even for chordal graphs G , as seen in Example 2.4.5. In our experiments, we found that ST_G can often be computed quite fast, and it frequently coincides with the desired prime ideal P_G .

2.4.2 The chordless m -cycle

We next discuss Questions 1, 2, and 3 for the simplest non-chordal graph, namely, the m -cycle C_m . Its Sullivant-Talaska ideal ST_{C_m} is generated by the 3×3 -minors of the submatrices $\Sigma_{A,B}$ where $A = \{i, i+1, \dots, j-1, j\}$, $B = \{j, j+1, \dots, i-1, i\}$, and $|i - j| \geq 2$. Here $\{A, B\}$ runs over all diagonals in the m -gon, and indices are understood modulo m . We conjecture that

$$P_{C_m} = \text{ST}_{C_m}. \quad (2.23)$$

We computed the ideal P_{C_m} in **Singular** for small m . Table 2.2 lists the results. In all cases in this table, the minimal generators consist of cubics only, which is consistent with the conjecture (2.23). For the degree of the Gaussian m -cycle we conjecture the following formula.

Table 2.2: *Dimension, degree, ML-degree, and degree and number of minimal generators of the ideal P_{C_m} for $m \leq 8$.*

m	3	4	5	6	7	8
dimension d	6	8	10	12	14	16
degree	1	9	57	312	1578	7599
ML-degree	1	5	17	49	129	321
minimal generators (degree:number)	0	3:2	3:15	3:63	3:196	3:504

Conjecture 2.4.6. *The degree of the projective variety $V(P_{C_m})$ associated with the m -cycle equals*

$$\frac{m+2}{4} \binom{2m}{m} - 3 \cdot 2^{2m-3}.$$

Regarding Question 2, the following formula was conjectured in [25, §7.4]:

$$\text{ML-degree}(C_m) = (m-3) \cdot 2^{m-2} + 1, \quad \text{for } m \geq 3. \quad (2.24)$$

This quantity is an algebraic complexity measure for the following matrix completion problem. Given real numbers x_i between -1 and $+1$, fill up the partially specified symmetric $m \times m$ -matrix

$$\begin{pmatrix} 1 & x_1 & ? & ? & \cdots & ? & x_m \\ x_1 & 1 & x_2 & ? & \cdots & ? & ? \\ ? & x_2 & 1 & x_3 & ? & \ddots & ? \\ ? & ? & x_3 & 1 & x_4 & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \ddots & ? \\ ? & ? & ? & ? & x_{m-2} & 1 & x_{m-1} \\ x_m & ? & ? & ? & ? & x_{m-1} & 1 \end{pmatrix} \quad (2.25)$$

to make it positive definite. We seek the unique fill-up that maximizes the determinant. The solution to this convex optimization problem is an algebraic function of x_1, x_2, \dots, x_m whose degree equals $\text{ML-degree}(C_m)$. We do not know how to prove (2.24) for $m \geq 9$. Even the following lemma is not easy.

Lemma 2.4.7. *The ML-degree of the cycle C_m is strictly larger than 1 for $m \geq 4$.*

Sketch of Proof. We consider the special case of (2.25) when all of the parameters are equal:

$$x := x_1 = x_2 = \cdots = x_m. \quad (2.26)$$

Since the logarithm of the determinant is a concave function, the solution to our optimization problem is fixed under the symmetry group of the m -gon, i.e., it is a symmetric *circulant*

matrix Σ_m . Hence there are only $\lfloor \frac{m-2}{2} \rfloor$ distinct values for the question marks in (2.25), one for each of the symmetry class of long diagonals in the m -gon. We denote these unknowns by $s_1, s_2, \dots, s_{\lfloor \frac{m-2}{2} \rfloor}$ where s_i is the unknown on the i -th circular off-diagonal. For instance, for $m = 7$, the circulant matrix we seek has two unknown entries s_1 and s_2 and is of the following form:

$$\Sigma_7 = \begin{pmatrix} 1 & x & s_1 & s_2 & s_2 & s_1 & x \\ x & 1 & x & s_1 & s_2 & s_2 & s_1 \\ s_1 & x & 1 & x & s_1 & s_2 & s_2 \\ s_2 & s_1 & x & 1 & x & s_1 & s_2 \\ s_2 & s_2 & s_1 & x & 1 & x & s_1 \\ s_1 & s_2 & s_2 & s_1 & x & 1 & x \\ x & s_1 & s_2 & s_2 & s_1 & x & 1 \end{pmatrix}$$

The key observation is that the determinant of the circular symmetric matrix Σ_m factors into a product of m linear factors with real coefficients, one for each m th root of unity. For example,

$$\det(\Sigma_7) = \prod_{w:w^7=1} (1 + (w + w^6) \cdot x + (w^2 + w^5) \cdot s_1 + (w^3 + w^4) \cdot s_2).$$

Thus, for fixed x , our problem is to maximize a product of linear forms. By analyzing the critical equations, obtained by taking logarithmic derivatives of $\det(\Sigma_m)$, we can show that the optimal solution $(\hat{s}_1, \hat{s}_2, \dots, \hat{s}_{\lfloor \frac{m-2}{2} \rfloor})$ is not a rational function in x . For example, when $m = 7$, the solution (\hat{s}_1, \hat{s}_2) is an algebraic function of degree 3 in x . Its explicit representation is

$$\hat{s}_1 = \frac{x^2 + \hat{s}_2 x - \hat{s}_2^2 - \hat{s}_2}{1 - x} \quad \text{and} \quad \hat{s}_2^3 + (1 - 2x)\hat{s}_2^2 + (-x^2 + x - 1)\hat{s}_2 + x^3 = 0$$

So the ML degree for this special symmetric case is 3, giving a lower bound for the ML degree of C_7 . A similar computation using **Singular** [21] shows that the ML degree for the symmetric case is $\lfloor \frac{m}{2} \rfloor$ for $4 \leq m \leq 12$ giving a lower bound on the ML degree for m -cycles with $4 \leq m \leq 12$. A general proof for $m > 13$ is still open. \square

We now come to our third problem, namely to giving an algebraic description of the cone of sufficient statistics, denoted $\mathcal{C}_m := \mathcal{C}_{C_m}$. This is a full-dimensional open convex cone in \mathbb{R}^{2m} . The coordinates on \mathbb{R}^{2m} are $s_{11}, s_{22}, \dots, s_{mm}$ and $x_1 = s_{12}, x_2 = s_{23}, \dots, x_m = s_{m1}$. We consider

$$\mathcal{C}'_m := \mathcal{C}_m \cap \{s_{11} = s_{22} = \dots = s_{mm} = 1\}.$$

This is a full-dimensional open bounded spectrahedron in \mathbb{R}^m . It consists of all (x_1, \dots, x_m) such that (2.25) can be filled up to a positive definite matrix. The 2×2 -minors of (2.25) imply that \mathcal{C}'_m lies in the cube $(-1, 1)^m = \{|x_i| < 1\}$. The issue is to identify further constraints.

We note that any description of the m -dimensional spectrahedron \mathcal{C}'_m leads to a description of the $2m$ -dimensional cone \mathcal{C}_m because a vector $s \in \mathbb{R}^{2m}$ lies in \mathcal{C}_m if and only if the vector $x \in \mathbb{R}^m$ with the following coordinates lies in \mathcal{C}'_m :

$$x_i = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} \quad \text{for } i = 1, 2, \dots, m \quad (2.27)$$

The article [9] gave a beautiful polyhedral description of the spectrahedron \mathcal{C}'_m . The idea is to replace each x_i by its arc-cosine, that is, to substitute $x_i = \cos(\phi_i)$ into (2.25). Remarkably, the image of the spectrahedron \mathcal{C}'_m under this transformation is a convex polytope. Explicit linear inequalities in the angle coordinates ϕ_i describing the facets of this polytope are given in [9].

To answer Question 3, we take the cosine-image of any of these facets and compute its Zariski closure. This leads to the following trigonometry problem. Determine the unique (up to scaling) irreducible polynomial Γ'_m which is obtained by rationalizing the equation

$$x_1 = \cos\left(\sum_{i=2}^m \arccos(x_i)\right). \quad (2.28)$$

We call Γ'_m the m -th *cycle polynomial*. Interestingly, Γ'_m is invariant under all permutations of the m variables x_1, x_2, \dots, x_m . We also define the *homogeneous m -th cycle polynomial* Γ_m to be the numerator of the image of Γ'_m under the substitution (2.27). The first cycle polynomials arise for $m = 3$:

$$\Gamma'_3 = \det \begin{pmatrix} 1 & x_1 & x_3 \\ x_1 & 1 & x_2 \\ x_3 & x_2 & 1 \end{pmatrix} \quad \text{and} \quad \Gamma_3 = \det \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{pmatrix}.$$

The polyhedral characterization of \mathcal{C}_m given in [9] translates into the following theorem.

Theorem 2.4.8. *The Zariski closure of the boundary of the cone \mathcal{C}_m , $m \geq 4$, is defined by the polynomial*

$$H_{\mathcal{C}_m}(s_{ij}) = \Gamma_m(s_{ij}) \cdot (s_{11}s_{22} - s_{12}^2) \cdot (s_{22}s_{33} - s_{23}^2) \cdots (s_{mm}s_{11} - s_{1m}^2).$$

To compute the cycle polynomial Γ'_m , we iteratively apply the sum formula for the cosine,

$$\cos(a + b) = \cos(a) \cdot \cos(b) - \sin(a) \cdot \sin(b),$$

and we then use the following relation to write (2.28) as an algebraic expression in x_1, \dots, x_n :

$$\sin(\arccos(x_i)) = \sqrt{1 - x_i^2}.$$

Finally, we eliminate the square roots (e.g. by using resultants) to get the polynomial Γ'_m .

For example, the cycle polynomial for the square ($m = 4$) has degree 6 and has 19 terms:

$$\Gamma'_4 = 4 \sum_{i < j < k} x_i^2 x_j^2 x_k^2 - 4x_1 x_2 x_3 x_4 \sum_i x_i^2 + \sum_i x_i^4 - 2 \sum_{i < j} x_i^2 x_j^2 + 8x_1 x_2 x_3 x_4.$$

By substituting (2.27) into this expression and taking the numerator, we obtain the homogeneous cycle polynomial Γ_4 which has degree 8. Table 2.3 summarizes what we know about the expansions of these cycle polynomials. Note that Γ'_m and Γ_m have different degrees but the same number of terms. The degree of the m -th cycle polynomial Γ'_m grows roughly like 2^m , but we do not know an exact formula. However, for the homogeneous cycle polynomial Γ_m we predict the following behavior.

Conjecture 2.4.9. *The degree of the homogeneous m -th cycle polynomial Γ_m equals $m \cdot 2^{m-3}$.*

There is another way of defining and computing the cycle polynomial Γ_m , without any reference to trigonometry or semidefinite programming. Consider the prime ideal generated by the 3×3 -minors of the generic symmetric $m \times m$ -matrix $\Sigma = (s_{ij})$. Then $\langle \Gamma_m \rangle$ is the principal ideal obtained by eliminating all unknowns s_{ij} with $|i - j| \geq 2$. Thus, geometrically, vanishing of the homogeneous polynomial Γ_m characterizes partial matrices on the m -cycle C_m that can be completed to a matrix of rank ≤ 2 . Similarly, vanishing of Γ'_m characterizes partial matrices (2.25) that can be completed to rank ≤ 2 .

Independently of the work of [9], a solution to the problem of characterizing the cone \mathcal{C}_m appeared in the same year in the statistics literature, namely by [15]. The cone \mathcal{C}_m is the set of partial sample covariance matrices on the m -cycle for which the MLE exists.

Table 2.3: *Degree and number of terms of cycle polynomials for $m \leq 11$.*

m	3	4	5	6	7	8	9	10	11
degree(Γ'_m)	3	6	15	30	70	140	315	630	1260
degree(Γ_m)	3	8	20	48	112	256	576	1280	2816
# of terms	5	19	339	19449	?	?	?	?	?

2.4.3 Small graphs, suspensions and wheels



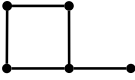
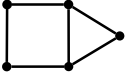
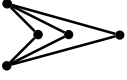
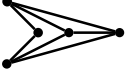
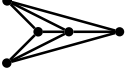
We next examine Questions 1, 2 and 3 for all graphs with at most five vertices. In this analysis we can restrict ourselves to connected graphs only. Indeed, if G is the disjoint union of two graphs G_1 and G_2 , then the prime ideal P_G is obtained from P_{G_1} and P_{G_2} as in (2.22) with $c = 0$, the ML-degrees multiply by Lemma 2.4.2, and the two dual cones both

decompose as direct products:

$$\mathcal{C}_G = \mathcal{C}_{G_1} \times \mathcal{C}_{G_2} \quad \text{and} \quad \mathcal{K}_G = \mathcal{K}_{G_1} \times \mathcal{K}_{G_2}.$$

Chordal graphs were dealt with in Section 2.4.1. We now consider connected non-chordal graphs with $m \leq 5$ vertices. There are seven such graphs, and in Table 2.4 we summarize our findings for these seven graphs. In the first two rows of Table 2.4 we find the 4-cycle and the 5-cycle which were discussed in Subsection 2.4.2. As an illustration we examine in detail the graph in the second-to-last row of Table 2.4.

Table 2.4: *Our three guiding questions for all non-chordal graphs with $m \leq 5$ vertices. Column 4 reports the degrees of the minimal generators together with the number of occurrence (degree:number). The last column lists the degrees of the irreducible factors of the polynomial H_G that define the Zariski closure of the boundary of \mathcal{C}_G . For each factor we report in lowercase the rank of the concentration matrices defining its dual irreducible component in the boundary of \mathcal{K}_G .*

Graph G	dim d	deg P_G	mingens P_G	ML-deg	deg H_G
	8	9	3:2	5	$4 \cdot 2_1 + 8_2$
	10	57	3:15	17	$5 \cdot 2_1 + 20_3$
	10	30	2:6, 3:4	5	$5 \cdot 2_1 + 8_2$
	11	31	3:10	5	$3 \cdot 2_1 + 3_1 + 8_2$
	11	56	3:7, 4:1	7	$6 \cdot 2_1 + 3 \cdot 8_2$
	12	24	3:4, 4:1	5	$2 \cdot 2_1 + 2 \cdot 3_1 + 10_2$
	13	16	4:2	5	$4 \cdot 3_1 + 12_2$

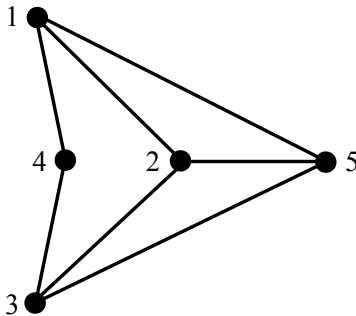


Figure 2.3: A Gaussian graphical model on five vertices and seven edges having dimension $d = 12$.

Example 2.4.10. The graph in Figure 2.3 defines the Gaussian graphical model with concentration matrix

$$K = \begin{pmatrix} \lambda_1 & \lambda_6 & 0 & \lambda_9 & \lambda_{10} \\ \lambda_6 & \lambda_2 & \lambda_7 & 0 & \lambda_{11} \\ 0 & \lambda_7 & \lambda_3 & \lambda_8 & \lambda_{12} \\ \lambda_9 & 0 & \lambda_8 & \lambda_4 & 0 \\ \lambda_{10} & \lambda_{11} & \lambda_{12} & 0 & \lambda_5 \end{pmatrix}.$$

We wish to describe the boundary of the cone \mathcal{C}_G by identifying the irreducible factors in its defining polynomial H_G . We first use the `Matlab` software `cvx` [32], which is specialized in convex optimization, to find the ranks of all concentration matrices K that are extreme rays in the boundary of \mathcal{K}_G . Using `cvx`, we maximize random linear functions over the compact spectrahedron $\overline{\mathcal{K}_G} \cap \{\text{trace}(K) = 1\}$, and we record the ranks of the optimal matrices. We found the possible matrix ranks to be 1 and 2, which agrees with the constraints $2 \leq p \leq 3$ seen in (2.14) for generic subspaces \mathcal{L} with $m = 5$ and $d = 12$.

We next ran the software `Singular` [21] to compute the minimal primes of the ideals of $p \times p$ -minors of K for $p = 2$ and $p = 3$, and thereafter we computed their dual ideals in $\mathbb{R}[t_1, t_2, \dots, t_{12}]$ using `Macaulay2` [33]. The latter step was done using the procedure with Jacobian matrices described in Subsection 2.2.3. We only retained dual ideals that are principal. Their generators are the candidates for factors of H_G .

The variety of rank one matrices K has four irreducible components. Two of those components correspond to the edges $(3, 4)$ and $(1, 4)$ in Figure 2.3. Their dual ideals are generated by the quadrics

$$p_1 = 4t_3t_4 - t_8^2 \quad \text{and} \quad p_2 = 4t_1t_4 - t_9^2.$$

The other two irreducible components of the variety of rank one concentration matrices correspond to the 3-cycles $(1, 2, 5)$ and $(2, 3, 5)$ in the graph. Their dual ideals are generated

by the cubics

$$p_3 = 4t_1t_2t_5 - t_5t_6^2 - t_2t_{10}^2 + t_6t_{10}t_{11} - t_1t_{11}^2 \quad \text{and} \quad p_4 = 4t_2t_3t_5 - t_5t_7^2 - t_3t_{11}^2 + t_7t_{11}t_{12} - t_2t_{12}^2.$$

The variety of rank two matrices K has two irreducible components. One corresponds to the chordless 4-cycle $(1, 2, 3, 4)$ in the graph and its dual ideal is generated by $p_5 = \Gamma_4$, which is of degree 8. The other component consists of rank two matrices K for which rows 2 and 5 are linearly dependent. The polynomial p_6 that defines the dual ideal consists of 175 terms and has degree 10.

The polynomial H_G is the product of those principal generators p_i whose hypersurface meets $\partial\mathcal{C}_G$. We again used `cvx` to check which of the six components actually contribute extreme rays in $\partial\mathcal{K}_G$. We found only one of the six components to be missing, namely the one corresponding to the chordless 4-cycle $(1, 2, 3, 4)$. This means that p_5 is not a factor of H_G , and we conclude that

$$H_G = p_1p_2p_3p_4p_6 \quad \text{and} \quad \deg(H_G) = 2 \cdot 2_1 + 2 \cdot 3_1 + 10_2. \quad (2.29)$$

Concerning Question 1 we note that the ideal P_G is minimally generated by the four 3×3 -minors of $\Sigma_{1235,134}$ and the determinant of $\Sigma_{1245,2345}$, and for Question 2 we note that the ML degree is five because the MLE can be derived from the MLE of the 4-cycle obtained by contracting the edge $(2, 5)$. \square

The graph in the last row of Table 2.4 is the wheel W_4 . It is obtained from the cycle C_4 in the first row by connecting all four vertices to a new fifth vertex. We see in Table 2.4 that the ML degree 5 is the same for both graphs, the two cubic generators of P_{C_4} correspond to the two quartic generators of P_{W_4} , and there is a similar correspondence between the irreducible factors of the dual polynomials H_{C_4} and H_{W_4} . In the remainder of this section we shall offer an explanation for these observations.

Let $G = (V, E)$ be an undirected graph and $G^* = (V^*, E^*)$ its *suspension graph* with an additional completely connected vertex 0. The graph G^* has vertex set $V^* = V \cup \{0\}$ and edge set $E^* = E \cup \{(0, v) \mid v \in V\}$. The m -wheel W_m is the suspension graph of the m -cycle C_m ; in symbols, $W_m = (C_m)^*$. We shall compare the Gaussian graphical models for the graph G and its suspension graph G^* .

Theorem 2.4.11. *The ML degree of a Gaussian graphical model with underlying graph G equals the ML degree of a Gaussian graphical model whose underlying graph is the suspension graph G^* . In symbols,*

$$\text{ML-degree}(G) = \text{ML-degree}(G^*).$$

Proof. Let $V = [m]$ and let $S^* \in \mathbb{S}_{>0}^{m+1}$ be a sample covariance matrix on G^* , where the first row and column correspond to the additional vertex 0. We denote by S' the lower right $m \times m$ submatrix of S^* corresponding to the vertex set V and by S the Schur complement

of S^* at S_{00}^* :

$$S = S' - \frac{1}{S_{00}^*} (S_{01}^*, \dots, S_{0m}^*)^T (S_{01}^*, \dots, S_{0m}^*). \quad (2.30)$$

Then $S \in \mathbb{S}_{>0}^m$ is a sample covariance matrix on G . Let $\hat{\Sigma}$ be the MLE for S on the graph G . We claim that the MLE $\hat{\Sigma}^*$ for S^* on the suspension graph G^* is given by

$$\hat{\Sigma}^* = \left[\begin{array}{c|ccc} S_{00}^* & S_{01}^* & \cdots & S_{0m}^* \\ \hline S_{01}^* & & & \\ \vdots & & \hat{\Sigma} + S' - S & \\ S_{0m}^* & & & \end{array} \right].$$

Clearly, $\hat{\Sigma}^*$ is positive definite and satisfies $\hat{\Sigma}_{ij}^* = S_{ij}^*$ for all $(i, j) \in E^*$. The inverse of the covariance matrix $\hat{\Sigma}^*$ can be computed by using the inversion formula based on Schur complements:

$$(\hat{\Sigma}^*)^{-1} = \left[\begin{array}{c|c} \frac{1}{S_{00}^*} + (S_{01}^*, \dots, S_{0m}^*)(\hat{\Sigma})^{-1}(S_{01}^*, \dots, S_{0m}^*)^T & \frac{1}{S_{00}^*} (S_{01}^*, \dots, S_{0m}^*)(\hat{\Sigma})^{-1} \\ \hline \frac{1}{S_{00}^*} (\hat{\Sigma})^{-1} (S_{01}^*, \dots, S_{0m}^*)^T & (\hat{\Sigma})^{-1} \end{array} \right].$$

Since the lower right block equals $(\hat{\Sigma})^{-1}$, its entries are indeed zero in all positions $(i, j) \notin E^*$.

We have shown that the MLE $\hat{\Sigma}^*$ is a rational function of the MLE $\hat{\Sigma}$. This shows

$$\text{ML-degree}(G^*) \leq \text{ML-degree}(G).$$

The reverse inequality is also true since we can compute the MLE on G for any $S \in \mathbb{S}_{>0}^m$ by computing the MLE on G^* for its extension $S^* \in \mathbb{S}_{>0}^{m+1}$ with $S_{00}^* = 1$ and $S_{0j}^* = 0$ for $j \in [m]$. \square

We next address the question of how the boundary of the cone \mathcal{C}_{G^*} can be expressed in terms of the boundary of \mathcal{C}_G . We use coordinates t_{ij} for both \mathbb{S}^m and its subspace \mathbb{R}^E , and we use the coordinates u_{ij} for both \mathbb{S}^{m+1} and its subspace \mathbb{R}^{E^*} . The Schur complement (2.30) defines a rational map from \mathbb{S}^{m+1} to \mathbb{S}^m which restricts to a rational map from \mathbb{R}^{E^*} to \mathbb{R}^E . The formula is

$$t_{ij} = u_{ij} - \frac{u_{0i}u_{0j}}{u_{00}} \quad \text{for } 1 \leq i \leq j \leq m. \quad (2.31)$$

A partial matrix (u_{ij}) on G^* can be completed to a positive definite $(m+1) \times (m+1)$ -matrix if and only if the partial matrix (t_{ij}) on G given by this formula can be completed to a positive definite $m \times m$ -matrix. The rational map (2.31) takes the boundary of the cone \mathcal{C}_{G^*} onto the boundary of the cone \mathcal{C}_G . For our algebraic question, we can derive the following conclusion:

Proposition 2.4.12. *The polynomial $H_{G^*}(u_{ij})$ equals the numerator of the Laurent polynomial obtained from $H_G(t_{ij})$ by the substitution (2.31), and the same holds for each irreducible factor.*

Example 2.4.13. The polynomial $H_{W_4}(u_{00}, u_{01}, u_{02}, u_{03}, u_{04}, u_{11}, u_{22}, u_{33}, u_{44}, u_{12}, u_{23}, u_{34}, u_{14})$ for the 4-wheel W_4 has as its main factor an irreducible polynomial of degree 12 which is the sum of 813 terms. It is obtained from the homogeneous cycle polynomial Γ_4 by the substitution (2.31). Recall that the homogeneous cycle polynomial $\Gamma_4(t_{11}, t_{22}, t_{33}, t_{44}, t_{12}, t_{23}, t_{34}, t_{14})$ has only degree 8 and is the sum of 19 terms. \square

We briefly discuss an issue raised by Question 1, namely, how to construct the prime ideal P_{G^*} from the prime ideal P_G . Again, we can use the transformation (2.31) to turn every generator of P_G into a Laurent polynomial whose numerator lies in P_{G^*} . However, the resulting polynomials will usually not suffice to generate P_{G^*} . This happens already for the 5-cycle $G = C_5$ and the 5-wheel $G^* = W_5$. The ideal P_{C_5} is generated by 15 linearly independent cubics arising as 3×3 -minors of the matrices $\Sigma_{132,1345}$, $\Sigma_{243,2451}$, $\Sigma_{354,3512}$, $\Sigma_{415,4123}$ and $\Sigma_{521,5234}$, while P_{W_5} is generated by 20 linearly independent quartics arising as 4×4 -minors of $\Sigma_{0132,01345}$, $\Sigma_{0243,02451}$, $\Sigma_{0354,03512}$, $\Sigma_{0415,04123}$ and $\Sigma_{0521,05234}$. Table 2.5 summarizes what we know about the Gaussian wheels W_m .

Table 2.5: *Dimension, degree, ML-degree, and degree and number of minimal generators for W_m with $m \leq 6$.*

m	3	4	5	6
dimension d	10	13	16	19
degree	1	16	198	2264
ML-degree	1	5	17	49
minimal generators (degree:number)	0	4:2	4:20	4:108

2.5 Colored Gaussian graphical models

We now add a graph coloring to the setup and study colored Gaussian graphical models. These were briefly described in Subsection 1.5.2 and introduced by [39], who called them *RCON-models*. In the underlying graph G , the vertices are colored with p different colors and the edges are colored with q different colors:

$$\begin{aligned} V &= V_1 \sqcup V_2 \sqcup \cdots \sqcup V_p, & p &\leq |V|, \\ E &= E_1 \sqcup E_2 \sqcup \cdots \sqcup E_q, & q &\leq |E|. \end{aligned}$$

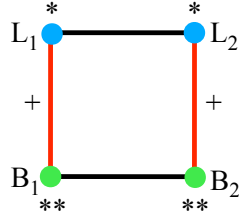


Figure 2.4: *Colored Gaussian graphical model for Frets' heads*: L_i , B_i denote the length and breadth of the head of son i .

We denote the uncolored graph by G and the colored graph by \mathcal{G} . Note that in this section the graph G does not contain any self-loops. The vertices play the same role as the self-loops in Section 2.4. In addition to the restrictions given by the missing edges in the graph, the entries of the concentration matrix K are now also restricted by equating entries in K according to the edge and vertex colorings. To be precise, the linear space \mathcal{L} of \mathbb{S}^m associated with a colored graph \mathcal{G} on $m = |V|$ nodes is defined by the following linear equations:

- For any pair of nodes α, β that do not form an edge we set $k_{\alpha\beta} = 0$ as before.
- For any pair of nodes α, β in a common color class V_i we set $k_{\alpha\alpha} = k_{\beta\beta}$.
- For any pair of edges $(\alpha, \beta), (\gamma, \delta)$ in a common color class E_j we set $k_{\alpha\beta} = k_{\gamma\delta}$.

The dimension of the model \mathcal{G} is $d = p + q$. Note again that, for any sample covariance matrix S ,

$$\pi_G(S) \in \mathcal{C}_G \quad \text{implies} \quad \pi_{\mathcal{G}}(S) \in \mathcal{C}_{\mathcal{G}}.$$

Thus, introducing a graph coloring on G relaxes the question of existence of the MLE.

In this section we shall examine Questions 1, 2 and 3 for various colorings \mathcal{G} of the 4-cycle $G = C_4$. We begin with a statistical application of colored Gaussian graphical models.

Example 2.5.1 (Frets' heads). We revisit the heredity study of head dimensions reported in [53] and known as *Frets' heads*. The data reported in this study consists of the length and breadth of the heads of 25 pairs of first and second sons. Because of the symmetry between the two sons, it makes sense to try to fit the colored Gaussian graphical model given in Figure 2.4.

This model has $d = 5$ degrees of freedom and it consists of all concentration matrices of the form

$$K = \begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}.$$

In Figure 2.4, the first random variable is denoted L_1 , the second L_2 , the third B_2 , and the fourth B_1 . Given a sample covariance matrix $S = (s_{ij})$, the five sufficient statistics for this model are

$$t_1 = s_{11} + s_{22}, \quad t_2 = s_{33} + s_{44}, \quad t_3 = 2s_{12}, \quad t_4 = 2(s_{23} + s_{14}), \quad t_5 = 2s_{34}. \quad (2.32)$$

The ideal of polynomials vanishing on K^{-1} is generated by four linear forms and one cubic in the s_{ij} :

$$P_G = \langle s_{11} - s_{22}, s_{33} - s_{44}, s_{23} - s_{14}, s_{13} - s_{24}, s_{23}^2 s_{24} - s_{24}^3 - s_{22} s_{23} s_{34} + s_{12} s_{24} s_{34} - s_{12} s_{23} s_{44} + s_{22} s_{24} s_{44} \rangle.$$

Note that the four linear constraints on the sample covariance matrix seen in P_G are also valid constraints on the concentration matrix. Models with this property are called *RCOP-models* and were studied by [39, 41].

The data reported in the Frets' heads study results in the following sufficient statistics:

$$t_1 = 188.256, \quad t_2 = 95.408, \quad t_3 = 133.750, \quad t_4 = 210.062, \quad t_5 = 67.302.$$

Substituting these values into (2.32) and solving the equations on $V(P_G)$, we find the MLE for this data:

$$\hat{\Sigma} = \begin{pmatrix} 94.1280 & 66.8750 & 44.3082 & 52.5155 \\ 66.8750 & 94.1280 & 52.5155 & 44.3082 \\ 44.3082 & 52.5155 & 47.7040 & 33.6510 \\ 52.5155 & 44.3082 & 33.6510 & 47.7040 \end{pmatrix}.$$

Both the degree and the ML-degree of this colored Gaussian graphical model is 3, which answers Questions 1 and 2. It remains to describe the boundary of the cone \mathcal{C}_G and to determine its defining polynomial H_G . The variety of rank one concentration matrices has four irreducible components:

$$\langle k_2, k_4, k_5, k_1 + k_3 \rangle, \langle k_2, k_4, k_5, k_1 - k_3 \rangle, \langle k_1, k_3, k_4, k_2 + k_5 \rangle, \langle k_1, k_3, k_4, k_2 - k_5 \rangle.$$

These are points in \mathbb{P}^4 and the ideals of their dual hyperplanes are $\langle t_1 - t_3 \rangle$, $\langle t_1 + t_3 \rangle$, $\langle t_2 - t_5 \rangle$, $\langle t_2 + t_5 \rangle$. The variety of rank two concentration matrices is irreducible. Its prime ideal and the dual thereof are

$$\langle k_2 k_3 + k_1 k_5, k_1 k_2 - k_4^2 + k_3 k_5, k_3 k_4^2 + k_1^2 k_5 - k_3^2 k_5 \rangle \\ \langle 4t_2^2 t_3^2 - 4t_1 t_2 t_4^2 + t_4^4 + 8t_1 t_2 t_3 t_5 - 4t_3 t_4^2 t_5 + 4t_1^2 t_5^2 \rangle.$$

This suggests that the hypersurface $\partial\mathcal{C}_G$ is given by the polynomial

$$H_G = (t_1 - t_3)(t_1 + t_3)(t_2 - t_5)(t_2 + t_5)(4t_2^2 t_3^2 - 4t_1 t_2 t_4^2 + t_4^4 + 8t_1 t_2 t_3 t_5 - 4t_3 t_4^2 t_5 + 4t_1^2 t_5^2). \quad (2.33)$$

Using `cvx` as in Example 2.4.10 we checked that all five factors meet $\partial\mathcal{C}_{\mathcal{G}}$, so (2.33) is indeed correct. \square

We performed a similar analysis for all colored Gaussian graphical models on the 4-cycle C_4 , which have the property that edges in the same color class connect the same vertex color classes. The results are presented in Table 2.6, 2.7 and 2.8. These models are of special interest because they are invariant under rescaling of variables in the same vertex color class. Such models were introduced and studied by [39]. For models with an additional permutation property (these are the *RCOP-models*), we explicitly list the polynomial $H_{\mathcal{G}}$. A census of these models appears in Table 2.8.

Example 2.5.2. We can gain a different perspective on the proof of Lemma 2.4.7 by considering colored Gaussian graphical models. Under the assumption (2.26) that all parameters in the partial matrix (2.25) are equal to some fixed value x , the MLE \hat{K} for the concentration matrix has the same structure. Namely, all diagonal entries of \hat{K} are equal, and all non-zero off-diagonal entries of \hat{K} are equal. This means that we can perform our MLE computation for the colored Gaussian graphical model with the chordless m -cycle as the underlying graph, where all vertices and all edges have the same color:

$$K = \begin{pmatrix} \lambda_1 & \lambda_2 & 0 & 0 & \cdots & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_2 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \lambda_1 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \lambda_2 & \lambda_1 & \lambda_2 \\ \lambda_2 & 0 & 0 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}. \quad (2.34)$$

In contrast to the approach in the proof of Lemma 2.4.7, in this representation we only need to solve a system of two polynomial equations in two unknowns, regardless of the cycle size m . The equations are

$$(K^{-1})_{11} = 1 \quad \text{and} \quad (K^{-1})_{12} = x.$$

By clearing denominators we obtain two polynomial equations in the unknowns λ_1 and λ_2 . We need to express these in terms of the parameter x , but there are many extraneous solutions. The ML degree is the algebraic degree of the special solution $(\hat{\lambda}_1(x), \hat{\lambda}_2(x))$ which makes (2.34) positive definite. \square

Example 2.5.3. Let \mathcal{G} be the colored triangle with the same color for all three vertices and three distinct colors for the edges. This is an RCOP model with $m = 3$ and $d = 4$. The corresponding subspace \mathcal{L} of \mathbb{S}^3 consists of all concentration matrices

$$K = \begin{pmatrix} \lambda_4 & \lambda_1 & \lambda_2 \\ \lambda_1 & \lambda_4 & \lambda_3 \\ \lambda_2 & \lambda_3 & \lambda_4 \end{pmatrix}.$$

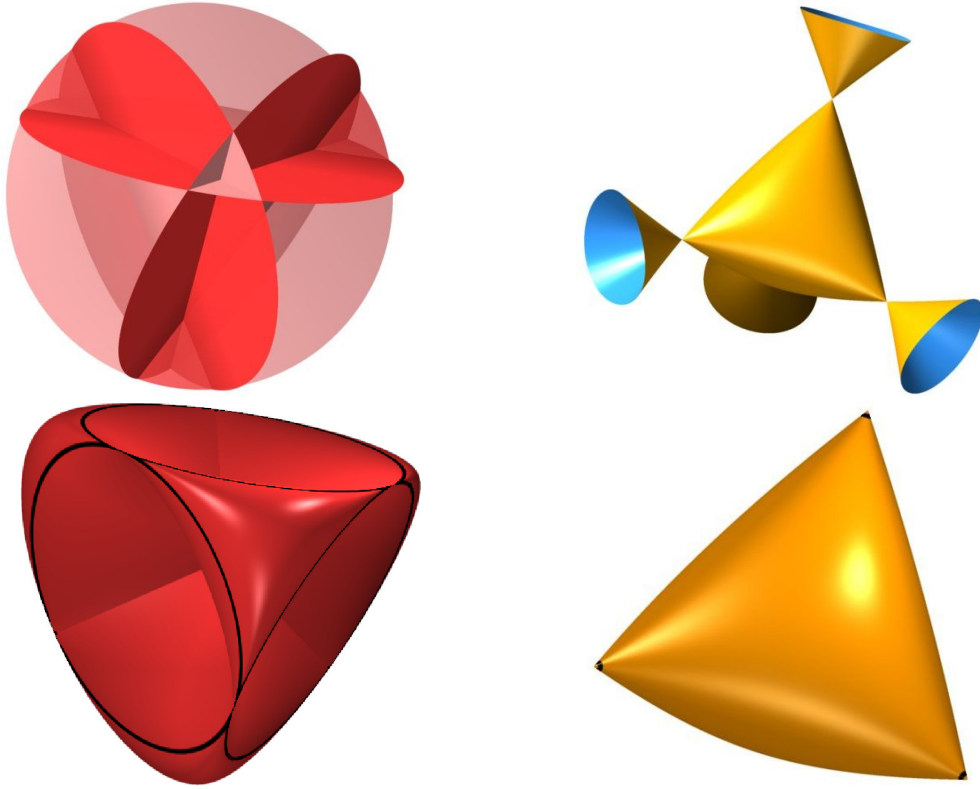


Figure 2.5: The cross section of the cone of sufficient statistics in Example 2.5.3 is the red convex body shown in the left figure. It is dual to Cayley's cubic surface, which is shown in yellow in the right figure and also in Figure 2.1 on the left.

This linear space \mathcal{L} is generic enough so as to exhibit the geometric behavior described in Subsection 2.2. The four-dimensional cone $\mathcal{K}_{\mathcal{L}}$ is the cone over the 3-dimensional spectrahedron bounded by *Cayley's cubic surface* as shown on the right in Figure 2.5. Its dual $\mathcal{C}_{\mathcal{L}}$ is the cone over the 3-dimensional convex body shown on the left in Figure 2.5. The boundary of this convex body consists of four flat 2-dimensional circular faces (shown in black) and four curved surfaces whose common Zariski closure is a *quartic Steiner surface*. Figure 2.5 was made with `surfex` [40], a software package for visualizing algebraic surfaces.

Here, the inequalities (2.14) state $2 \leq p \leq 3$, and the algebraic degree of semidefinite programming is $\delta(3, 3, 2) = \delta(3, 3, 1) = 4$. We find that $H_{\mathcal{L}}$ is a polynomial of degree 8 which factors into four linear forms and one quartic:

$$H_{\mathcal{L}} = (t_1 - t_2 + t_3 - t_4)(t_1 + t_2 - t_3 - t_4)(t_1 - t_2 - t_3 + t_4)(t_1 + t_2 + t_3 + t_4)(t_1^2 t_2^2 + t_1^2 t_3^2 + t_2^2 t_3^2 - 2t_1 t_2 t_3 t_4).$$

By Theorem 2.2.3, both the degree and the ML degree of this model are also equal to $\phi(3, 4) = 4$. \square

Table 2.6: Results on Questions 1, 2, and 3 for all colored Gaussian graphical models with some symmetry restrictions (namely, edges in the same color class connect the same vertex color classes) on the 4-cycle.

Graph	K	dim d	degree	mingens P_G	ML-degree	deg H_C
1)	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_2 \\ \lambda_2 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}$	3	5	1:4, 2:5	5	$2_2 + 2_3 + 2_3$
2)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_3 \\ \lambda_3 & 0 & \lambda_3 & \lambda_2 \end{pmatrix}$	4	11	1:1, 2:10	5	$2_2 + 4_3$
3)	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_3 \\ \lambda_2 & 0 & \lambda_3 & \lambda_1 \end{pmatrix}$	3	4	1:4, 2:6	2	2_3
4)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_4 \\ \lambda_3 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	4	6	1:3, 2:4	3	$2_2 + 4_3$
5)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_4 \\ \lambda_3 & 0 & \lambda_4 & \lambda_2 \end{pmatrix}$	4	8	1:3, 2:2, 3:4	2	2_3
6)	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_4 \\ \lambda_2 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	4	11	1:1, 2:10, 3:1	6	$2_2 + 2_2 + 4_3$
7)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_3 & 0 & \lambda_5 & \lambda_1 \end{pmatrix}$	5	13	2:8, 3:3	3	$4_2 + 4_3$
8)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_3 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	5	21	2:5, 3:10	6	$2_2 + 2_2 + 4_3$
9)	$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_4 \\ \lambda_4 & \lambda_2 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_3 & \lambda_6 \\ \lambda_4 & 0 & \lambda_6 & \lambda_2 \end{pmatrix}$	6	15	2:5, 3:1	3	$2_2 + 3_2$
10)	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_3 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_4 \\ \lambda_3 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	4	5	1:4, 2:1, 3:2	3	$1_2 + 1_2 + 2_2$
11)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	5	11	1:1, 2:5, 3:4	3	$2_2 + 2_2$
12)	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_5 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_4 \\ \lambda_5 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	5	11	1:1, 2:5, 3:4	3	$2_2 + 2_2$

Table 2.7: Continuation of Table 2.6.

13)		$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_6 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_6 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	6	17	2:3, 3:4	5	$1_1 + 1_1 + 1_1 + 1_1 + 4_2 + 4_2$
14)		$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_6 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_6 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	6	21	3:10, 4:12	3	$2_2 + 2_2$
15)		$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_6 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_6 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	6	17	2:2, 3:8, 4:1	4	10_2
16)		$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_7 \\ \lambda_4 & \lambda_1 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_2 & \lambda_6 \\ \lambda_7 & 0 & \lambda_6 & \lambda_3 \end{pmatrix}$	7	13	2:1, 3:3	5	$1_1 + 1_1 + 2_1 + 12_2$
17)		$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_7 \\ \lambda_4 & \lambda_2 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_1 & \lambda_6 \\ \lambda_7 & 0 & \lambda_6 & \lambda_3 \end{pmatrix}$	7	17	3:3, 4:6	3	4_2
18)		$\begin{pmatrix} \lambda_1 & \lambda_5 & 0 & \lambda_8 \\ \lambda_5 & \lambda_2 & \lambda_6 & 0 \\ 0 & \lambda_6 & \lambda_3 & \lambda_7 \\ \lambda_8 & 0 & \lambda_7 & \lambda_4 \end{pmatrix}$	8	9	3:2	5	$2_1 + 2_1 + 2_1 + 2_1 + 8_2$

Table 2.8: All RCOP-models (see [39]) when the underlying graph is the 4-cycle.

Graph	K	dim d	degree	mingens P_G	ML-degree	$H_{\mathcal{L}}$
1)	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_2 & 0 \\ 0 & \lambda_2 & \lambda_1 & \lambda_2 \\ \lambda_2 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}$	2	2	1:7, 2:1	2	$(2t_1 - t_2)(2t_1 + t_2)$
2)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_3 \\ \lambda_3 & 0 & \lambda_3 & \lambda_2 \end{pmatrix}$	3	4	1:5, 2:2	2	$16t_1t_2 - t_3^2$
3)	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_3 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_2 \\ \lambda_3 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}$	3	3	1:6, 3:1	3	$(t_1 - t_2)(t_1 + t_2)(t_1 - t_3)(t_1 + t_3)$
4)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_3 \\ \lambda_4 & 0 & \lambda_3 & \lambda_2 \end{pmatrix}$	4	5	1:4, 2:1, 3:2	3	$(4t_1t_2 - t_3^2)(4t_1t_2 - t_4^2)$
5)	$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_4 \\ \lambda_4 & \lambda_2 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_3 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	5	6	1:3, 2:1, 3:1	3	$(8t_1t_2 - t_4^2)(8t_2t_3 - t_5^2)$
6)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	5	3	1:4, 3:1	3	(2.33) in Example 2.5.1

Chapter 3

Minimal number of observations needed for existence of MLE

After having studied the problem of existence of the MLE at the level of sufficient statistics in the previous chapter, we now want to find the minimum number of observations needed for the existence of the MLE in a Gaussian graphical model. Using the geometric description of the cone of sufficient statistics in Chapter 2, we give an algebraic elimination criterion, which allows us to find exact lower bounds on the number of observations needed to ensure that the MLE exists with probability one. This is applied to bipartite graphs, grids and colored graphs. We also study the ML degree, and we present the first instance of a graph for which the MLE exists with probability one even when the number of observations equals the treewidth.

The material of this chapter was submitted for publication in a paper with title "Geometry of maximum likelihood estimation in Gaussian graphical models".

3.1 Introduction

Gaussian graphical models are regular exponential families. The statistical theory of exponential families, as presented for example by Brown [13] or Barndorff-Nielsen [7], is a strong tool to establish existence and uniqueness of the MLE. The MLE exists and is unique if and only if the sufficient statistic lies in the interior of its convex support. We gave a geometric description of the convex support of the sufficient statistics in Chapter 2, and will now discuss the connection to the minimum number of samples needed to guarantee that the corresponding sufficient statistic lies in the interior of its convex support.

This chapter is organized as follows. In Section 3.2, we briefly review the geometry of ML estimation in Gaussian graphical models, and we develop an exact algebraic algorithm to determine lower bounds on the number of observations needed to ensure existence of the MLE with probability one. In Section 3.3, we discuss the existence of the MLE for bipartite

graphs. Section 3.4 deals with small graphs. The 3×3 grid motivated this chapter and is the original problem Steffen Lauritzen posed during his talk about the existence of the MLE in Gaussian graphical models at the "Durham Symposium on Mathematical Aspects of Graphical Models" on July 8, 2008. The 3×3 grid is also the first example of a graph, for which the MLE exists with probability one even when the number of observations equals the treewidth of the underlying graph. We conclude this chapter with a characterization of Gaussian models on colored 4-cycles in Section 3.5.

3.2 Geometry of maximum likelihood estimation in Gaussian graphical models

In this chapter we study Gaussian graphical models with mean 0 and covariance matrix Σ . However, the results can easily be generalized to a model $\mathcal{N}(\mu, \Sigma)$ with given mean μ , as n i.i.d. samples from $\mathcal{N}(\mu, \Sigma)$ can be transformed into $n - 1$ i.i.d. samples from $\mathcal{N}(0, \Sigma)$.

The set of all concentration matrices in a Gaussian graphical model with underlying graph G define the convex cone

$$\mathcal{K}_G = \{K \in \mathbb{S}_{>0}^m \mid K_{ij} = 0, \quad \forall (i, j) \notin E\}.$$

Note that the edge set of G contains all self-loops, i.e. $(i, i) \in E$ for all $i \in [m]$. By taking the inverse of every matrix in \mathcal{K}_G , we get the set of all covariance matrices in the model denoted by \mathcal{K}_G^{-1} .

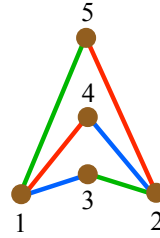
The *minimal sufficient statistic* of a sample covariance matrix S in a Gaussian graphical model is the G -partial matrix S_G . The set of all sufficient statistics, for which the MLE exists, is denoted by \mathcal{C}_G . It is given by the projection of the positive definite cone $\mathbb{S}_{>0}^m$ onto the edge set of the graph G :

$$\mathcal{C}_G = \pi_G(\mathbb{S}_{>0}^m).$$

So \mathcal{C}_G is also a convex cone. Moreover, we proved in Proposition 2.2.1 that the cone of sufficient statistics \mathcal{C}_G is the convex dual to the cone of concentration matrices \mathcal{K}_G . See Figure 2.2 for a geometric interpretation of the various cones involved in ML estimation in Gaussian graphical models. In the following, we also visualize the cones by a small example.

Example 3.2.1. For small dimensional problems we are able to give a graphical representation of the cone of sufficient statistics \mathcal{C}_G . For example, consider the Gaussian graphical model on the bipartite graph $K_{2,3}$ with concentration matrices of the form

$$K = \begin{pmatrix} \lambda_1 & 0 & \lambda_2 & \lambda_3 & \lambda_4 \\ 0 & \lambda_1 & \lambda_4 & \lambda_2 & \lambda_3 \\ \lambda_2 & \lambda_4 & \lambda_1 & 0 & 0 \\ \lambda_3 & \lambda_2 & 0 & \lambda_1 & 0 \\ \lambda_4 & \lambda_3 & 0 & 0 & \lambda_1 \end{pmatrix}$$



Note that in order to reduce the number of parameters and be able to draw \mathcal{C}_G in 3-dimensional space, we assume additional equality constraints on the non-zero entries of the concentration matrix represented by the graph coloring above. Such colored Gaussian graphical models, where the coloring represents equality constraints on the concentration matrix, are called RCON-models and have been introduced in [39].

Without loss of generality we can rescale K and assume that all diagonal entries are one. The cone of concentration matrices \mathcal{K}_G for this model is shown in Figure 3.1(a). Its algebraic boundary is described by $\{\det(K) = 0\}$ and shown in Figure 3.1(b). In this example, the determinant factors into two components, a cylinder and an ellipsoid. Dualizing the boundary of \mathcal{K}_G by the algorithm described in Proposition 2.2.4 results in the hypersurface shown in Figure 3.1(e). The double cone is dual to the cylinder in Figure 3.1(b). By making

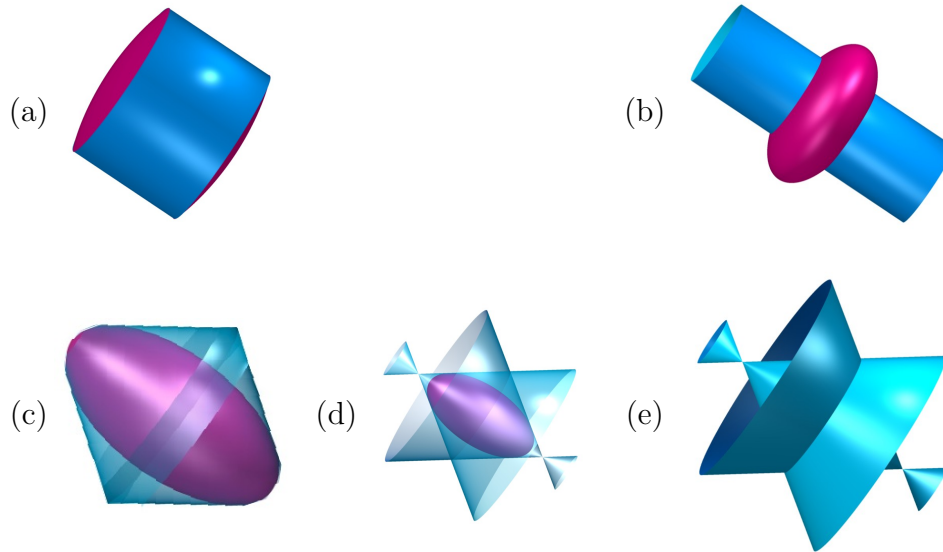


Figure 3.1: *These pictures illustrate the convex geometry of maximum likelihood estimation for Gaussian graphical models. The cross-section of the cone of concentration matrices \mathcal{K}_G with a three-dimensional hyperplane is shown in (a), its algebraic boundary in (b), the dual cone of sufficient statistics in (c), and its algebraic boundary in (d) and (e), where (d) is the transparent version of (e).*

the double cone transparent as shown in Figure 3.1(d), we see the enclosed ellipsoid, which is dual to the ellipsoid in Figure 3.1(b). The cone of sufficient statistics \mathcal{C}_G is shown in Figure 3.1(c). The MLE exists if and only if the sufficient statistic lies in the interior of this convex body. Using the elimination criterion of Theorem 3.2.2, we can show that the MLE exists with probability one already for one observation. \square

In this chapter, we examine the existence of the MLE for n observations in the range $q \leq n < q^*$, for which the existence of the MLE is not well understood (see Corollary 1.5.2). Cycles are the only non-chordal graphs, which have been studied [8, 9, 15]. We will extend the results on cycles to bipartite graphs $K_{2,m}$ and small grids.

Geometrically, we look at the manifold of rank n matrices on the boundary of the cone $\mathbb{S}_{\geq 0}^m$. In general, its projection

$$\pi_G(\{M \in \mathbb{S}_{\geq 0}^m \mid \text{rk}(M) = n\}) \quad (3.1)$$

lies in the topological closure of the cone \mathcal{C}_G . The MLE exists with probability one for n observations if and only if the projection (3.1) lies in the interior of \mathcal{C}_G .

Based on the geometric interpretation of maximum likelihood estimation in Gaussian graphical models, we can derive a sufficient condition for the existence of the MLE. The following algebraic elimination criterion can be used as an algorithm to establish existence of the MLE with probability one for n observation.

Theorem 3.2.2 (Elimination criterion). *Let $I_{G,n}$ be the elimination ideal obtained from the ideal of $(n+1) \times (n+1)$ -minors of a symmetric $m \times m$ matrix S of unknowns by eliminating all unknowns corresponding to non-edges of the graph G . If $I_{G,n}$ is the zero ideal, then the MLE exists with probability one for n observations.*

Proof. The variety corresponding to the ideal of $(n+1) \times (n+1)$ -minors of a symmetric $m \times m$ matrix S of unknowns consists of all $m \times m$ matrices of rank at most n . Eliminating all unknowns corresponding to non-edges of the graph G results in the elimination ideal $I_{G,n}$ (see e.g. [19]) and is geometrically equivalent to a projection onto the cone of sufficient statistics \mathcal{C}_G . Let V be the variety corresponding to the elimination ideal $I_{G,n}$. We denote by k its dimension and by μ the k -dimensional Lebesgue measure. The MLE exists with probability one for n observations if

$$\mu(V \cap \partial\mathcal{C}_G) = 0,$$

where $\partial\mathcal{C}_G$ denotes the boundary of the cone of sufficient statistics \mathcal{C}_G .

If $I_{G,n}$ is the zero ideal, then the variety V is full-dimensional and its dimension satisfies $\dim(V) = k = \dim(\mathcal{C}_G)$. So if we assume that $\mu(V \cap \partial\mathcal{C}_G) > 0$, then $\mu(\partial\mathcal{C}_G) > 0$, which is a contradiction to $\dim(\partial\mathcal{C}_G) < k$. \square

For small examples, the elimination ideal $I_{G,n}$ can be computed e.g. using `Macaulay2` [33], a software system for research in algebraic geometry. If $I_{G,n}$ is not the zero ideal, then

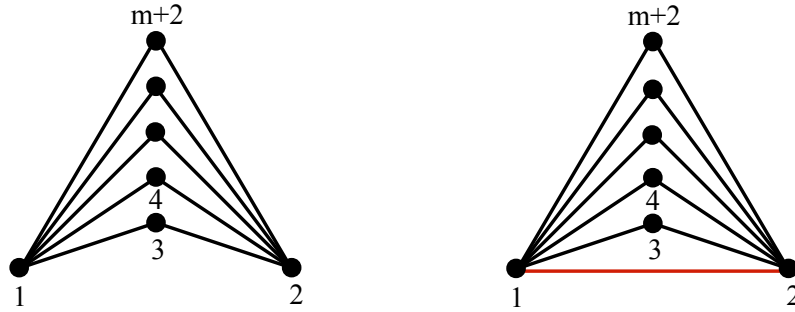


Figure 3.2: *Bipartite graph $K_{2,m}$ (left) and minimal chordal cover of $K_{2,m}$ (right).*

an analysis of polynomial inequalities is required. One needs to carefully examine how the components of V are located. The argument is subtle because the algebraic boundary of \mathcal{C}_G may in fact intersect the interior of \mathcal{C}_G . So even if the projection V is a component of the algebraic boundary of \mathcal{C}_G , the MLE might still exist with positive probability. We will encounter and describe such an example in detail in Section 3.5.

3.3 Bipartite graphs

In this section, we first derive the MLE existence results for bipartite graphs $K_{2,m}$ paralleling the results on cycles proven by Buhl [15]. Let the graph $K_{2,m}$ be labelled as shown in Figure 3.2. A minimal chordal cover is given in Figure 3.2 (right). As for cycles, for bipartite graphs $K_{2,m}$ we have $q = 2$ and $q^* = 3$. Therefore only the case of $n = 2$ observations is interesting.

Let X_1 and X_2 denote two independent samples from the distribution $\mathcal{N}_{m+2}(0, \Sigma)$, which obeys the undirected pairwise Markov property on $K_{2,m}$. We denote by X the $(m+2) \times 2$ data matrix consisting of the two samples X_1 and X_2 as columns. The rows of X are denoted by x_1, \dots, x_{m+2} . Similarly as for cycles in [15], we will describe a criterion on the configuration of data vectors x_1, \dots, x_{m+2} ensuring the existence of the MLE. The following characterization of positive definite matrices of size 3×3 proven in [9] will be helpful in this context.

Lemma 3.3.1. *The matrix*

$$\begin{pmatrix} 1 & \cos(\alpha) & \cos(\beta) \\ \cos(\alpha) & 1 & \cos(\gamma) \\ \cos(\beta) & \cos(\gamma) & 1 \end{pmatrix}$$

with $0 < \alpha, \beta, \gamma < \pi$ is positive definite if and only if

$$\alpha < \beta + \gamma, \quad \beta < \alpha + \gamma, \quad \gamma < \alpha + \beta, \quad \alpha + \beta + \gamma < 2\pi.$$

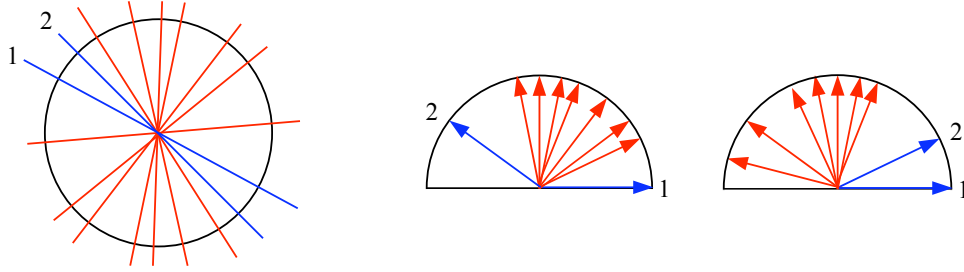


Figure 3.3: The MLE on $K_{2,m}$ exists in the following situations. Lines and data vectors corresponding to the variables 1 and 2 are drawn in blue. Lines and data vectors corresponding to the variables 3, \dots , $m + 2$ are drawn in red.

Proposition 3.3.2. *The MLE on the graph $K_{2,m}$ exists with probability one for $n \geq 3$ observations, and the MLE does not exist for $n < 2$ observations. For $n = 2$ observations the MLE exists if and only if the lines generated by x_1 and x_2 are direct neighbors (see Figure 3.3 left).*

Proof. Because the problem of existence of the MLE is a positive definite matrix completion problem, we can rescale and rotate the data vectors x_1, \dots, x_{m+2} (i.e. perform an orthogonal transformation) without changing the problem. So without loss of generality we can assume that the vectors $x_1, \dots, x_{m+2} \in \mathbb{R}^2$ have length one, lie in the upper unit half circle, and $x_1 = (1, 0)$. We need to prove that the MLE exists if and only if the data configuration is as shown in Figure 3.3 (middle) or (right).

Let θ_{ij} denote the angle between vector x_i and x_j . Then the $K_{2,m}$ -partial sample covariance matrix $S_{K_{2,m}}$ is of the form

$$\left(\begin{array}{cc|cccc} 1 & ? & \cos(\theta_{13}) & \cos(\theta_{14}) & \cdots & \cos(\theta_{1,m+2}) \\ ? & 1 & \cos(\theta_{23}) & \cos(\theta_{24}) & \cdots & \cos(\theta_{2,m+2}) \\ \hline \cos(\theta_{13}) & \cos(\theta_{23}) & 1 & ? & \cdots & ? \\ \cos(\theta_{14}) & \cos(\theta_{24}) & ? & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & ? \\ \cos(\theta_{1,m+2}) & \cos(\theta_{2,m+2}) & ? & \cdots & ? & 1 \end{array} \right).$$

The graph $K_{2,m}$ can be extended to a chordal graph by adding one edge as shown in Figure 3.2 (right). So by Theorem 1.3.1, $S_{K_{2,m}}$ can be extended to a positive definite matrix if and only if the (1,2)-entry of $S_{K_{2,m}}$ can be completed in such a way that all the submatrices corresponding to maximal cliques are positive definite. This is equivalent to the existence of

$\rho \in \mathbb{R}$ with $0 < \rho < \pi$ such that

$$\begin{pmatrix} 1 & \cos(\rho) & \cos(\theta_{1i}) \\ \cos(\rho) & 1 & \cos(\theta_{2i}) \\ \cos(\theta_{1i}) & \cos(\theta_{2i}) & 1 \end{pmatrix} \succ 0 \quad \text{for all } i \in \{3, 4, \dots, m+2\}.$$

By Lemma 3.3.1 this occurs if and only if

$$\left. \begin{matrix} \theta_{1i} - \theta_{2i} \\ \theta_{2i} - \theta_{1i} \end{matrix} \right\} < \rho < \left\{ \begin{matrix} \theta_{1j} + \theta_{2j} \\ 2\pi - \theta_{1j} - \theta_{2j} \end{matrix} \right. \quad \text{for all } i, j \in \{3, 4, \dots, m+2\},$$

which is equivalent to

$$2\theta_{ai} < \theta_{1i} + \theta_{2i} + \theta_{1j} + \theta_{2j} < 2\pi + 2\theta_{ai} \quad (3.2)$$

for all $a \in \{1, 2\}$, $i, j \in \{3, 4, \dots, m+2\}$. We distinguish two cases.

Case 1 There is a vector x_j lying between x_1 and x_2 , which implies that $\theta_{1j} + \theta_{2j} = \theta_{12}$. If there was a vector x_i , $i \neq j$, which does not lie between x_1 and x_2 , then

$$\theta_{1j} + \theta_{2j} + \theta_{1i} + \theta_{2i} = 2\theta_{1i},$$

which is a contradiction to (3.2). Hence all vectors x_3, \dots, x_{m+2} lie between x_1 and x_2 , in which case

$$\theta_{1i} + \theta_{2i} + \theta_{1j} + \theta_{2j} = 2\theta_{12}$$

and Inequality (3.2) is satisfied.

Case 2 The vectors x_1 and x_2 are direct neighbors, which implies that $\theta_{1i} + \theta_{2i} = \theta_{12} + 2\theta_{2i}$ for all $i \in \{3, \dots, m+2\}$, in which case Inequality (3.2) is satisfied.

This proves that for two observations the MLE exists if and only if the data configuration is as shown in Figure 3.3 (middle) or (right). □

The geometric explanation of what is happening in this example, is that the projection of the positive definite matrices of rank 2 intersects the interior and the boundary of the cone of sufficient statistics \mathcal{C}_G with positive measure. The sufficient statistics originating from data vectors, where lines 1 and 2 are neighbors, lie in the interior of \mathcal{C}_G . If lines 1 and 2 are not neighbors, the corresponding sufficient statistics lie on the boundary of the cone \mathcal{C}_G and the MLE does not exist. A similar situation is encountered in Example 2.5.1 and depicted in Figure 3.7.

A different approach to gaining a better understanding of maximum likelihood estimation in Gaussian graphical models is to study the ML degree of the underlying graph. The map

taking a sample covariance matrix S to its maximum likelihood estimate $\hat{\Sigma}$ is an algebraic function and its degree is the ML degree of the model. See [25, Def. 2.1.4]. The ML degree represents the algebraic complexity of the problem of finding the MLE. This suggests that a larger ML degree results in a more difficult MLE existence problem. We proved in Theorem 2.4.3 that the ML degree is one if and only if the underlying graph is chordal. It is conjectured in [25, Section 7.4] that the ML degree of the cycle grows exponentially in the cycle length. An interesting contrast to the cycle conjecture is the following theorem, where we prove that the ML degree for bipartite graphs $K_{2,m}$ grows linearly in the number of variables.

Theorem 3.3.3. *In a Gaussian graphical model with underlying graph $K_{2,m}$ the ML degree is $2m + 1$.*

Proof. Given a generic matrix $S \in \mathbb{S}^{m+2}$, we fix $\Sigma \in \mathbb{S}^{m+2}$ with entries $\Sigma_{ij} = S_{ij}$ for $(i, j) \in E$ and unknowns $\Sigma_{12} = \Sigma_{21} = y$ and $\Sigma_{ij} = z_{ij}$ for all other $(i, j) \notin E$. We denote by $K = \Sigma^{-1}$ the corresponding concentration matrix. The ML degree of $K_{2,m}$ is the number of complex solutions to the equations

$$(\Sigma^{-1})_{ij} = 0 \quad \text{for all } (i, j) \notin E.$$

Let A denote the set consisting of the two distinguished vertices $\{1, 2\}$ and let $B = V \setminus A$. In the following we will use the block structure

$$\Sigma = \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix}, \quad K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{BA} & K_{BB} \end{pmatrix}.$$

For example for the graph $K_{2,5}$ the corresponding covariance matrix Σ and concentration matrix K are of the form

$$\Sigma = \left(\begin{array}{cc|ccc} 1 & y & S_{13} & S_{14} & S_{15} \\ y & 1 & S_{23} & S_{24} & S_{25} \\ \hline S_{13} & S_{23} & 1 & z_{34} & z_{35} \\ S_{14} & S_{24} & z_{34} & 1 & z_{45} \\ S_{15} & S_{25} & z_{35} & z_{45} & 1 \end{array} \right), \quad K = \left(\begin{array}{cc|ccc} K_{11} & 0 & K_{13} & K_{14} & K_{15} \\ 0 & K_{22} & K_{23} & K_{24} & K_{25} \\ \hline K_{13} & K_{23} & K_{33} & 0 & 0 \\ K_{14} & K_{24} & 0 & K_{44} & 0 \\ K_{15} & K_{25} & 0 & 0 & K_{55} \end{array} \right).$$

Note that in general the block K_{BB} is a diagonal matrix. Hence the Schur complement

$$\Sigma_{BB} - \Sigma_{BA} \Sigma_{AA}^{-1} \Sigma_{AB}$$

is also a diagonal matrix. Writing out the off-diagonal entries of this matrix results in the following expression for the variables z in terms of the variable y :

$$z_{ij} = -\frac{1}{1-y^2} (y(S_{1i}S_{2j} + S_{1j}S_{2i}) - S_{1i}S_{1j} - S_{2i}S_{2j}).$$

Setting the minor M_{12} of Σ to zero results in the last equation of the form

$$y \det(\Sigma_{BB}) + (\text{polynomial in } z \text{ of degree } m - 1) = 0. \quad (3.3)$$

Note that $\det(\Sigma_{BB})$ is a polynomial in z of degree m , where the degree 0 term is 1. So by multiplying equation (3.3) with $(1 - y^2)^m$, we get a degree $2m + 1$ equation in y and therefore $2m + 1$ complex solutions for y . For each solution of y we get one solution for the variables z , which proves that the ML degree of $K_{2,m}$ is $2m + 1$. \square

Both bipartite graphs and cycles are classes of graphs with $q = 2$ and $q^* = 3$. What can we say about such graphs in general regarding the existence of the MLE for two observations? In [8] a related question has been studied from a purely algebraic point of view. A cycle-completable graph is defined to be a graph such that every partial matrix M_G has a positive definite completion if and only if M_G is positive definite on all submatrices corresponding to maximal cliques in the graph and all submatrices corresponding to cycles in the graph can be completed to a positive definite matrix. It is shown in [8] that a graph is cycle-completable if and only if there is a chordal cover with no new 4-clique.

Buhl [15] studied cycles from a more statistical point of view and described a criterion on the data vectors for the existence of the MLE for two observations. Combining the results of [8] and [15], we get the following result:

Corollary 3.3.4. *Let G be a graph with $q = 2$ and $q^* \geq 3$. Then the following statements are equivalent:*

- i) For $n = 2$ observations the MLE exists if and only if Buhl's cycle condition is satisfied on every induced cycle.*
- ii) $q^* = 3$.*

This result solves the problem of existence of the MLE for all graphs with $q = 2$ and $q^* = 3$. Note that Corollary 3.3.4 is more general than Proposition 3.3.2. The proof, however, is more involved and less constructive.

For bipartite graphs $K_{3,m}$ the situation is more complicated. We do not yet have statements as in Proposition 3.3.2 and Theorem 3.3.3. We here describe some preliminary results.

Let the graph $K_{3,m}$ be labelled as shown in Figure 3.4. A minimal chordal cover is given in Figure 3.4 (middle). Hence, $q = 2$ and $q^* = 4$. The convex body shown in Figure 3.4 (right) consists of all positive semidefinite 3×3 matrices with ones on the diagonal. We call this the tetrahedron-shaped pillow, and we will prove that the existence of the MLE is equivalent to a non-empty intersection of such inflated and shifted tetrahedron-shaped pillows.

Remark 3.3.5. The MLE on the graph $K_{3,m}$ exists if and only if the m inflated and shifted tetrahedron-shaped pillows corresponding to the maximal cliques in the minimal chordal cover of $K_{3,m}$ shown in Figure 3.4 (middle) have non-empty intersection.

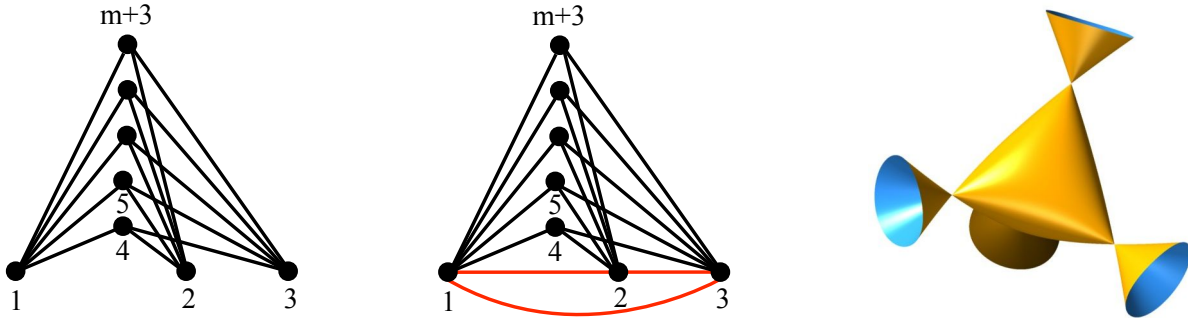


Figure 3.4: Bipartite graph $K_{3,m}$ (left) and minimal chordal cover of $K_{3,m}$ (middle). The tetrahedron-shaped pillow consisting of all correlation matrices of size 3×3 is shown in the right figure.

Proof. Applying Theorem 1.3.1 in a similar way as in the proof of Theorem 3.3.2, the partial covariance matrix $S_{K_{3,m}}$ can be extended to a positive definite matrix if and only if the entries corresponding to the missing edges (1, 2), (1, 3), and (2, 3) can be completed in such a way that all the submatrices corresponding to maximal cliques in the minimal chordal cover (Figure 3.4, middle) are positive definite. This is equivalent to the existence of $x, y, z \in \mathbb{R}$ with $-1 < x, y, z < 1$ such that

$$\begin{pmatrix} 1 & s_{1i} & s_{2i} & s_{3i} \\ s_{1i} & 1 & x & y \\ s_{2i} & x & 1 & z \\ s_{3i} & y & z & 1 \end{pmatrix} \succ 0 \quad \text{for all } i \in \{4, 5, \dots, m+3\}, \quad (3.4)$$

where s_{ai} , $a \in \{1, 2, 3\}$, $i \in \{4, 5, \dots, m+3\}$ are the sufficient statistics corresponding to edges in $K_{3,m}$. Using Schur complements and rescaling, (3.4) holds if and only if

$$\begin{pmatrix} 1 & x_i & y_i \\ x_i & 1 & z_i \\ y_i & z_i & 1 \end{pmatrix} \succ 0 \quad \text{for all } i \in \{4, 5, \dots, m+3\}, \quad (3.5)$$

where

$$x_i = \frac{x - s_{1i}s_{2i}}{\sqrt{1 - s_{1i}^2}\sqrt{1 - s_{2i}^2}}, \quad y_i = \frac{y - s_{1i}s_{3i}}{\sqrt{1 - s_{1i}^2}\sqrt{1 - s_{3i}^2}}, \quad z_i = \frac{z - s_{2i}s_{3i}}{\sqrt{1 - s_{2i}^2}\sqrt{1 - s_{3i}^2}}.$$

So the MLE exists if and only if the inflated and shifted tetrahedron-shaped pillows corresponding to the inequalities in (3.5) have non-empty intersection. \square

We used the software `Macaulay2` [33] to compute the ML degree of $K_{3,m}$ for $m \leq 4$. It is an open problem to find the general formula of the ML degree for $K_{l,m}$, where $l \geq 3$.

m	1	2	3	4
ML degree	1	7	57	131

3.4 Small graphs

In this section we analyze the 3×3 grid in particular and complete the discussion of Subsection 2.4.3 with the number of observations and the corresponding existence probability of the MLE for all graphs with 5 or less vertices.

The 3×3 grid is shown in Figure 3.5 (left) and has $q = 2$ and $q^* = 4$. This graph is of particular interest, because characterizing the existence of the MLE for this graph constitutes the original problem posed by Steffen Lauritzen at the "Durham Symposium on Mathematical Aspects of Graphical Models" in 2008. As a preparation, we first discuss the existence of the MLE for the graph \mathcal{G} on six vertices shown in Figure 3.6. The graph \mathcal{G} also has $q = 2$ and $q^* = 4$, and is the first example, for which we can prove that the bound $n \geq q^*$ for the existence of the MLE with probability one is not tight and that the MLE can exist with probability one even when the number of observations equals the treewidth.

Theorem 3.4.1. *The MLE on the graph \mathcal{G} (Figure 3.6, left) exists with probability one for $n = 3$ observations.*

Proof. We compute the ideal $I_{\mathcal{G},3}$ by eliminating the variables $s_{13}, s_{15}, s_{16}, s_{24}, s_{26}, s_{34}, s_{35}$ from the ideal of 4×4 minors of the matrix S . This results in the zero ideal, which by Theorem 3.2.2 completes the proof. \square

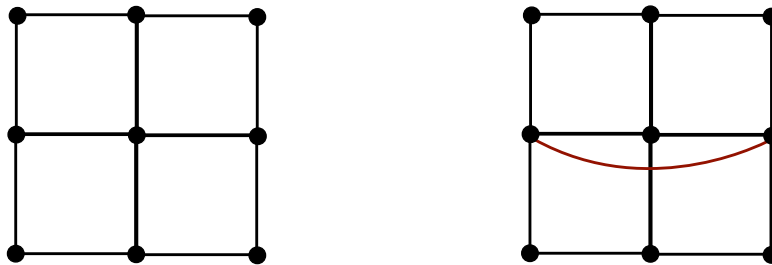
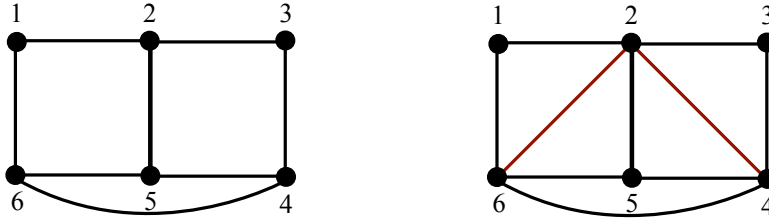


Figure 3.5: 3×3 grid \mathcal{H} (left) and grid with additional edge \mathcal{H}' (right).

Figure 3.6: Graph \mathcal{G} (left) and minimal chordal cover of \mathcal{G} (right).

Remark 3.4.2. Theorem 3.4.1 is equivalent to the following purely algebraic statement. Let

$$S = \begin{pmatrix} 1 & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ s_{12} & 1 & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{13} & s_{23} & 1 & s_{34} & s_{35} & s_{36} \\ s_{14} & s_{24} & s_{34} & 1 & s_{45} & s_{46} \\ s_{15} & s_{25} & s_{35} & s_{45} & 1 & s_{56} \\ s_{16} & s_{26} & s_{36} & s_{46} & s_{56} & 1 \end{pmatrix} \in \mathbb{S}_{\geq 0}^6$$

with $\text{rank}(S) = 3$. Then there exist $x, y, a, b, c, d, e \in \mathbb{R}$ such that

$$S' = \begin{pmatrix} 1 & s_{12} & a & s_{14} & b & c \\ s_{12} & 1 & s_{23} & x & s_{25} & y \\ a & s_{23} & 1 & d & e & s_{36} \\ s_{14} & x & d & 1 & s_{45} & s_{46} \\ b & s_{25} & e & s_{45} & 1 & s_{56} \\ c & y & s_{36} & s_{46} & s_{56} & 1 \end{pmatrix} \in \mathbb{S}_{> 0}^6.$$

So any partial matrix of rank 3 with specified entries at all positions corresponding to edges in \mathcal{G} can be completed to a positive definite matrix.

Corollary 3.4.3. *Let \mathcal{H} be the 3×3 -grid shown in Figure 3.5. Then the MLE on \mathcal{H} exists with probability one for $n \geq 3$ observations, and the MLE does not exist for $n < 2$ observations.*

Proof. First note that Gröbner bases computations are extremely memory intensive and the elimination ideal $I_{\mathcal{H},3}$ cannot be computed directly due to insufficient memory. We solve this problem by glueing together smaller graphs. The probability of existence of the MLE for the 3×3 grid \mathcal{H} is at least as large as the existence probability when the underlying graph is \mathcal{H}' . The graph \mathcal{H}' is a clique sum of two graphs of the form \mathcal{G} , for which the MLE existence probability is one for $n \geq 3$. \square



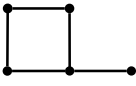
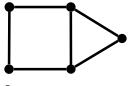
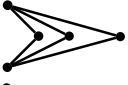
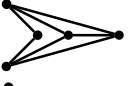
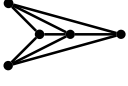
This example shows that although we are not able to compute the elimination ideal for large graphs directly, the algebraic elimination criterion (Theorem 3.2.2) is still useful also

in this situation. We can study small graphs with the elimination criterion and glue them together using clique sums to build larger graphs.

For two observations on the 3×3 grid, the cycle conditions are necessary but not sufficient for the existence of the MLE (Corollary 3.3.4). Unlike for bipartite graphs $K_{2,m}$, the existence of the MLE does not only depend on the ordering of the lines corresponding to the data vectors in \mathbb{R}^2 . By simulations with the `Matlab` software `cvx` [32], one can easily find orderings for which the MLE sometimes exists and sometimes doesn't. Finding a necessary and sufficient criterion for the existence of the MLE for two observations remains an open problem.

We now complete the discussion of Subsection 2.4.3 with the number of observations and the corresponding existence probability of the MLE for all graphs with 5 or less vertices. All non-chordal graphs with 5 or less vertices are shown in Table 3.1. The 4-cycle and 5-cycle in a) and b) are covered by Buhl's results [15]. The graphs in c) and d) are clique sums of two graphs and therefore completable if and only if the submatrices corresponding to the two subgraphs are completable. Graph e) is the bipartite graph $K_{2,3}$ and covered by Theorem 3.3.2. For the graph in f) $q = 3$ and $q^* = 4$. Applying the elimination criterion from Theorem 3.2.2 shows that three observations are sufficient for the existence of the MLE. Finally, for 3 observations on the graph g) we can find by simulation configurations of data vectors for which the MLE exists and others for which the MLE does not exist. Because

Table 3.1: *This table shows the number of observations (obs.) and the corresponding MLE existence probability for all non-chordal graphs on 5 or fewer vertices.*

Graph G	1 obs.	2 obs.	3 obs.	≥ 4 obs.
a) 	no	$p \in (0, 1)$	$p = 1$	$p = 1$
b) 	no	$p \in (0, 1)$	$p = 1$	$p = 1$
c) 	no	$p \in (0, 1)$	$p = 1$	$p = 1$
d) 	no	no	$p = 1$	$p = 1$
e) 	no	$p \in (0, 1)$	$p = 1$	$p = 1$
f) 	no	no	$p = 1$	$p = 1$
g) 	no	no	$p \in (0, 1)$	$p = 1$

every configuration has a positive probability under the multivariate normal distribution, the MLE exists with probability strictly between 0 and 1 in this case.

3.5 Colored Gaussian graphical models

For some applications, symmetries in the underlying Gaussian graphical model can be assumed. Adding symmetry to the conditional independence restrictions of a graphical model reduces the number of parameters and in some cases also the number of observations needed for the existence of the MLE. The symmetry restrictions can be represented by a graph coloring, where edges, or vertices, respectively, have the same coloring if the corresponding elements of the concentration matrix are equal. Such colored Gaussian graphical models have been introduced in Subsection 1.5.2 and studied in Section 2.5. We also discussed such a model in Example 3.2.1.

We denote the uncolored graph by G and the colored graph by \mathcal{G} . Note that in this section the graph G does not contain any self-loops. Let the vertices be colored with p different colors and the edges with q different colors:

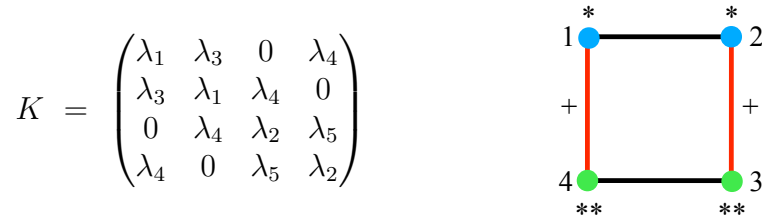
$$\begin{aligned} V &= V_1 \sqcup V_2 \sqcup \cdots \sqcup V_p, & p &\leq |V| \\ E &= E_1 \sqcup E_2 \sqcup \cdots \sqcup E_q, & q &\leq |E|. \end{aligned}$$

Then the set of all concentration matrices $\mathcal{K}_{\mathcal{G}}$ consists of all positive definite matrices satisfying

- $K_{\alpha\beta} = 0$ for any pair of vertices α, β that do not form an edge in G .
- $K_{\alpha\alpha} = K_{\beta\beta}$ for any pair of vertices α, β in a common vertex color class V_i .
- $K_{\alpha\beta} = K_{\gamma\delta}$ for any pair of edges $(\alpha, \beta), (\gamma, \delta)$ in a common edge color class E_j .

Example 3.5.1 (Frets' heads). We revisit the heredity study of head dimensions known as *Frets' heads* reported in [53] and also discussed in Example 2.5.1. We now want to find the minimum number of observations needed for the existence of the MLE.

The data set consists of the length and breadth of the heads of 25 pairs of first and second sons. The data supports the following colored Gaussian graphical model, where the joint distribution remains the same when the two sons are exchanged.



In this graph variable 1 corresponds to the length of the first son's head, variable 2 to the length of the second son's head, variable 3 to the breadth of the second son's head, and variable 4 to the breadth of the first son's head.

Given a sample covariance matrix $S = (s_{ij})$, the five sufficient statistics for this model according to the graph coloring are

$$t_1 = s_{11} + s_{22}, \quad t_2 = s_{33} + s_{44}, \quad t_3 = 2s_{12}, \quad t_4 = 2(s_{23} + s_{14}), \quad t_5 = 2s_{34}.$$

The algebraic boundary of the cone of sufficient statistics $\mathcal{C}_{\mathcal{G}}$ is computed in Example 2.5.1 and given by the polynomial

$$\begin{aligned} H_{\mathcal{G}} = & (t_1 - t_3) \cdot (t_1 + t_3) \cdot (t_2 - t_5) \cdot (t_2 + t_5) \\ & \cdot (4t_2^2t_3^2 - 4t_1t_2t_4^2 + t_4^4 + 8t_1t_2t_3t_5 - 4t_3t_4^2t_5 + 4t_1^2t_5^2). \end{aligned}$$

For two observations the elimination ideal $I_{\mathcal{G},2}$ is the zero ideal. Therefore, the MLE exists with probability 1 for two or more observations in this model. For one observation we get

$$I_{\mathcal{G},1} = \langle 4t_2^2t_3^2 - 4t_1t_2t_4^2 + t_4^4 + 8t_1t_2t_3t_5 - 4t_3t_4^2t_5 + 4t_1^2t_5^2 \rangle,$$

which corresponds to one of the components of the algebraic boundary of the cone of sufficient statistics. In this example, the algebraic boundary of the cone of sufficient statistics intersects its interior. This is illustrated in Figure 3.7. In order to get a graphical representation in 3-dimensional space, we fixed t_3 and t_5 . The variety corresponding to $I_{\mathcal{G},1}$ is shown on the left. We call this hypersurface the bow tie. The cone of sufficient statistics $\mathcal{C}_{\mathcal{G}}$ is the convex hull of the bow tie and shown in Figure 3.7 (right). Its boundary consists of four planes corresponding to the components $t_1 - t_3$, $t_1 + t_3$, $t_2 - t_5$, and $t_2 + t_5$ shown in blue, and the bows of the bow tie shown in yellow. The black curves show where the planes touch the bow tie. Note that the upper and lower two triangles of the bow tie lie in the interior of $\mathcal{C}_{\mathcal{G}}$. Only the two bows are part of the boundary of $\mathcal{C}_{\mathcal{G}}$. So the MLE exists if the sufficient statistic lies on one of the triangles of the bow tie, and it does not exist if the sufficient statistic lies on one of the bows of the bow tie. Consequently, for one observation the MLE exists with probability strictly between 0 and 1.

A different approach is to run simulations for example using `cvx` [32]. We can generate vectors of length four and compute the MLE by solving a convex optimization problem. If `cvx` finds a solution, the MLE exists. This also shows that the MLE exists with probability strictly between 0 and 1 for one observation.

For this example, we can characterize not just the sufficient statistics, but also the observations, for which the MLE exists. In other words, we can characterize the observations whose sufficient statistics lie on the triangles of the bow tie. First, note that by exchanging variables 1 and 2 and simultaneously exchanging variables 3 and 4, we get the same model. This means that from one observation $X_1 = (x_1, x_2, x_3, x_4)$ we can generate a second

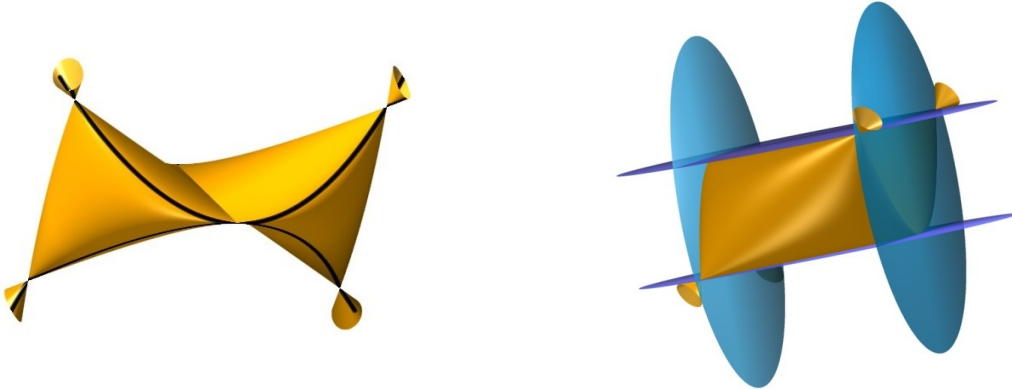


Figure 3.7: All possible sufficient statistics from one observation are shown on the left. The cone of sufficient statistics is shown on the right.

observation $X_2 = (x_2, x_1, x_4, x_3)$. So the resulting data matrix is given by

$$X = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_1 \\ x_3 & x_4 \\ x_4 & x_3 \end{pmatrix}.$$

Applying Buhl's result about two observations on a Gaussian cycle [15], the MLE exists if and only if the lines corresponding to the vectors

$$y_1 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad y_2 = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}, \quad y_3 = \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}, \quad y_4 = \begin{pmatrix} x_4 \\ x_3 \end{pmatrix}$$

are not graph consecutive. This is the case if and only if

$$|x_1| > |x_2| \text{ and } |x_3| > |x_4| \quad \text{or} \quad |x_1| < |x_2| \text{ and } |x_3| < |x_4|.$$

Hence, the MLE for one observation exists if and only if the data is inconsistent, meaning that the head of the first (second) son is longer than the head of the second (first) son, but the breadth is smaller. In this situation the corresponding sufficient statistics lie on the triangles of the bow tie in Figure 3.7. Otherwise the corresponding sufficient statistics lie on the bows of the bow tie. \square

In Section 2.5 we found the defining polynomial $\mathcal{H}_{\mathcal{G}}$ of the cone of sufficient statistics for all colored Gaussian graphical models on the 4-cycle, which have the property that edges in the same color class connect the same vertex color classes. Such models have been studied in [39] and are of special interest, because they are invariant under rescaling of variables in the

same vertex color class. In Table 3.2, 3.3 and 3.4, we complete the discussion of Section 2.5 with the number of observations and the corresponding existence probability of the MLE.

For every colored 4-cycle we computed the elimination ideal $I_{\mathcal{G},n}$ for $n = 1, 2, 3$. If it is the zero ideal, we know from Theorem 3.2.2 that the MLE exists with probability one. If $I_{\mathcal{G},n}$ is non-zero, we run simulations using `cvx`. The MLE exists with probability strictly between 0 and 1 for n observations, if we find examples for which the MLE exists and other examples for which the MLE does not exist. In cases where simulations don't yield any counterexamples, we need to prove that the MLE does indeed not exist by carefully analyzing the components corresponding to the ideal $I_{\mathcal{G},n}$. This is the case for one observation on the RCON graphs 9), 11), 14) and 17). Note that the graphical models 9) and 11) are sub-models of 14) and 17). So if we prove that the MLE does not exist for one observation on the graphs 9) and 11), this follows also for the graphs 14) and 17).

If the cone $\mathcal{C}_{\mathcal{G}}$ for the RCON graphs 9) and 11) is a basic open semialgebraic set (see e.g. [2]), then $\mathcal{C}_{\mathcal{G}}$ does not meet its algebraic boundary and the MLE does not exist for one observation. We end this chapter with two conjectures which would answer the question marks in Tables 3.2 and 3.3:

Conjecture 3.5.2. *The cones $\mathcal{C}_{\mathcal{G}}$ corresponding to the RCON graphs 9) and 11) are basic open semialgebraic sets.*

The question marks could also be resolved by solving a non-commutative sum of squares problem. Let $x_1, \dots, x_4 \in \mathbb{R}^n$ be the observations on each vertex of the 4-cycle. Note that the dimension n corresponds to the number of observations and is not fixed. For the RCON graph 9) the sufficient statistics are

$$t_1 = x_1^T x_1, \quad t_2 = x_2^T x_2 + x_4^T x_4, \quad t_3 = x_3^T x_3, \quad t_4 = 2x_1^T x_2 + 2x_1^T x_4, \quad t_5 = 2x_2^T x_3, \quad t_6 = 2x_3^T x_4.$$

The two components defining the boundary of the cone of sufficient statistics for this example are given by

$$p_1 = 4t_2 t_3 - t_5^2 - t_6^2 \quad \text{and} \quad p_2 = 8t_1 t_2 t_3 - t_3 t_4^2 - t_1 t_5^2 + 2t_1 t_5 t_6 - t_1 t_6^2.$$

It is straightforward to prove that $p_1 \geq 0$ for any $x_1, \dots, x_4 \in \mathbb{R}^n$ for **any** dimension $n \geq 1$. If we can prove that p_2 is non-negative as well, then the cone $\mathcal{C}_{\mathcal{G}}$ does not meet its algebraic boundary and the MLE does not exist for one observation.

Similarly, for the RCON graph 9) the sufficient statistics are given by

$$t_1 = x_1^T x_1 + x_3^T x_3, \quad t_2 = x_2^T x_2 + x_4^T x_4, \quad t_3 = 2x_1^T x_2, \quad t_4 = 2x_2^T x_3 + 2x_1^T x_4, \quad t_5 = 2x_3^T x_4.$$

The two components defining the boundary of the cone of sufficient statistics for this example are given by

$$q_1 = 4t_1 t_2 - t_3^2 - 2t_3 t_5 - t_5^2 \quad \text{and} \quad q_2 = 4t_1 t_2 - t_3^2 - t_4^2 + 2t_3 t_5 - t_5^2.$$

It is straightforward to prove that $q_1 \geq 0$ for any $x_1, \dots, x_4 \in \mathbb{R}^n$ with $n \geq 1$. The following conjecture would prove that the MLE does not exist for one observation on the RCON graphs 9) and 11):

Conjecture 3.5.3. *The polynomials p_2 and q_2 can be written as sums of squares in the variables $x_1, \dots, x_4 \in \mathbb{R}^n$ for **any** dimension $n \geq 1$.*

This is a partially non-commutative sum of squares problem, since

$$(x_1^T x_2)(x_3^T x_4) = (x_4^T x_3)(x_1^T x_2), \quad \text{but} \quad (x_1^T x_2)(x_3^T x_4) \neq (x_1^T x_3)(x_2^T x_4).$$

Table 3.2: *Results on the number of observations and the MLE existence probability for all colored Gaussian graphical models with some symmetry restrictions (namely, edges in the same color class connect the same vertex color classes) on the 4-cycle.*

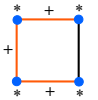
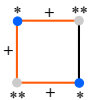
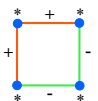

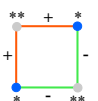
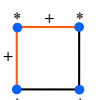
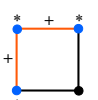
Graph	K	1 obs.	2 obs.	≥ 3 obs.
1) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_2 \\ \lambda_2 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
2) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_3 \\ \lambda_3 & 0 & \lambda_3 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
3) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_3 \\ \lambda_2 & 0 & \lambda_3 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
4) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_4 \\ \lambda_3 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
5) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_4 \\ \lambda_3 & 0 & \lambda_4 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
6) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_4 \\ \lambda_2 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
7) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_3 & 0 & \lambda_5 & \lambda_1 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$

Table 3.3: Continuation of Table 2.6.

8)		$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_3 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$
9)		$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_4 \\ \lambda_4 & \lambda_2 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_3 & \lambda_6 \\ \lambda_4 & 0 & \lambda_6 & \lambda_2 \end{pmatrix}$	no?	$p = 1$	$p = 1$
10)		$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_3 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_4 \\ \lambda_3 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
11)		$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	no?	$p = 1$	$p = 1$
12)		$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_5 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_4 \\ \lambda_5 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
13)		$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_6 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_6 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$
14)		$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_6 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_6 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	no?	$p = 1$	$p = 1$
15)		$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_6 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_6 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$
16)		$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_7 \\ \lambda_4 & \lambda_1 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_2 & \lambda_6 \\ \lambda_7 & 0 & \lambda_6 & \lambda_3 \end{pmatrix}$	no	$p = 1$	$p = 1$
17)		$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_7 \\ \lambda_4 & \lambda_2 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_1 & \lambda_6 \\ \lambda_7 & 0 & \lambda_6 & \lambda_3 \end{pmatrix}$	no?	$p = 1$	$p = 1$
18)		$\begin{pmatrix} \lambda_1 & \lambda_5 & 0 & \lambda_8 \\ \lambda_5 & \lambda_2 & \lambda_6 & 0 \\ 0 & \lambda_6 & \lambda_3 & \lambda_7 \\ \lambda_8 & 0 & \lambda_7 & \lambda_4 \end{pmatrix}$	no	$p \in (0, 1)$	$p = 1$

Table 3.4: All RCOP-models (see [39]) on the 4-cycle.

Graph	K	1 obs.	2 obs.	≥ 3 obs.
1)	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_2 & 0 \\ 0 & \lambda_2 & \lambda_1 & \lambda_2 \\ \lambda_2 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
2)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_3 \\ \lambda_3 & 0 & \lambda_3 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
3)	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_3 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_2 \\ \lambda_3 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
4)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_3 \\ \lambda_4 & 0 & \lambda_3 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
5)	$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_4 \\ \lambda_4 & \lambda_2 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_3 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
6)	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$

Chapter 4

Asymptotics of ML estimation in Gaussian cycles

In the previous chapters, we described the geometry of maximum likelihood estimation in Gaussian graphical models and gave an algebraic criterion to find the minimum number of observations needed for the existence of the MLE. This criterion can in theory be applied to any graph. However, due to the computational limitations of Gröbner bases techniques, we are in practice only able to apply this criterion to small graphs. In this chapter, we describe a possible future direction on how our results could be used for studying very large graphs. We prove that for very large cycles already two observations are sufficient for the existence of the MLE with probability one.

The content of this chapter is based on various discussions (mainly during the fall semester of 2009) with Venkat Chandrasekaran and Parikshit Shah, two PhD students in the Department of Electrical Engineering and Computer Science at MIT.

4.1 Introduction

In this chapter, we study the problem of existence of the MLE in Gaussian graphical models from an asymptotic point of view. Given a class of graphs (e.g. cycles), we let the graph size grow to infinity and we want to determine how fast the sample size has to grow in order for the MLE existence probability to go to one. In this chapter, we show how the results about the existence of the MLE for (small) cycles can be used to find the corresponding asymptotic results. We believe that our results about small graphs in the previous chapters (in particular the 3×3 grid) can be extended in a similar way to large graphs.

In this chapter, we denote a graph with m vertices by G_m . We distinguish between the underlying true Gaussian graphical model and the assumed model used for covariance estimation. We denote the family of graphs underlying the Gaussian graphical model from which the n samples are drawn from by G_m^{true} and the corresponding covariance matrix by

Σ_m^{true} . The family of graphs underlying the assumed Gaussian graphical model with which we analyze the samples is denoted by G_m^{assumed} . This graph defines the structure of the partial sample covariance matrix. Those two families of graphs need not necessarily be the same. Note that the question of existence of the MLE is only interesting for non-chordal families of graphs G_m^{assumed} . If G_m^{assumed} is chordal, existence of the MLE depends on the sample size only (see Corollary 1.5.2).

With this notation we can rephrase the main question as follows: What is the asymptotic behavior of $\mathbb{P}(\text{MLE exists}, n \text{ observations}, \Sigma_m^{\text{true}}, G_m^{\text{assumed}})$ as m goes to infinity?

In this chapter, we will be working with the following family of graphs:

- i) Trivial graphs: $G_m = ([m], E)$ with $E = \emptyset$.
- ii) Cycles: $G_m = ([m], E)$ with $E = \{(i, i + 1) \mid i = 1, \dots, m - 1\} \cup \{(1, m)\}$.
- iii) Complete graphs: $G_m = ([m], E)$ with $E = \{(i, j) \mid i, j \in [m]\}$.

4.1.1 ML estimation in Gaussian cycles

Cycles have been the first family of non-chordal graphs G_m^{assumed} for which the probability $\mathbb{P}(\text{MLE exists}, n, \Sigma_m^{\text{true}}, G_m^{\text{assumed}})$ for a *fixed* number of vertices m has been studied [9, 15]. For a fixed number of vertices m , $G_m^{\text{assumed}} = \text{cycle}$, and $G_m^{\text{true}} = \text{complete graph}$ (no restrictions on the entries of the concentration matrix), Buhl [15] showed the following result.

Theorem 4.1.1. *The MLE exists with probability 1 for $n \geq 3$ and does not exist for $n \leq 1$. For $n = 2$ let*

$$X = (X_1, \dots, X_m), Y = (Y_1, \dots, Y_m) \sim \mathcal{N}(0, \Sigma) \text{ independently with } \Sigma \text{ positive definite.}$$

Then the MLE exists if and only if the lines defined by the vectors $(X_1, Y_1), \dots, (X_m, Y_m)$ are not graph consecutive, i.e. when starting with an arbitrary line and turning half a revolution, the lines do not occur in one of the two sequences conforming with the graph cycle.

As a direct consequence one gets an asymptotic result for $G_m^{\text{true}} = \text{trivial}$. Under independence, every configuration of lines has the same probability. Fixing one line, there are $(m - 1)!$ different cyclic arrangements of which 2 are graph consecutive, which explains the following Corollary:

Corollary 4.1.2.

$$\mathbb{P}(\text{MLE exists}, n = 2, \Sigma_m^{\text{true}} = \text{Id}_m, G_m^{\text{assumed}} = \text{cycle}) = 1 - \frac{2}{(m - 1)!}.$$

So in the most restrictive case, where G_m^{true} is the trivial family of graphs, the probability that the MLE exists for two observations goes to 1 exponentially in the cycle length. In the following, we study the asymptotic behavior of

$$\mathbb{P}(\text{MLE exists}, n = 2, \Sigma_m^{\text{true}}, G_m^{\text{assumed}} = \text{cycle})$$

as m goes to infinity and we will give a criterion on Σ_m^{true} which ensures the existence of the MLE for two observations.

4.2 On the distribution of quotients of normal random variables

Let (X_1, \dots, X_m) and (Y_1, \dots, Y_m) be two independent samples from $\mathcal{N}(0, \Sigma_m)$, where Σ_m is a positive definite matrix. We denote the quotients $\frac{X_i}{Y_i}$ by Z_i . Understanding the joint distribution of these Gaussian quotients is important, because the sum over all $2m$ probabilities of the form $\mathbb{P}(Z_1 < Z_2 < \dots < Z_m)$ yields the MLE existence probability in the case $G_m^{\text{assumed}} = \text{cycle}$.

Note that the joint distribution of (Z_1, \dots, Z_m) is scale invariant. So without loss of generality we can consider correlation matrices Σ_m only. For $m = 1$ we get the distribution of the quotient of two independent Gaussian variables, which is known to have a standard Cauchy distribution. We can also consider the angle $\Theta = \arctan(Z_1)$, which has a uniform distribution on the unit half circle. For $m = 2$ it can be seen that

$$\mathbb{P}(Z_1 < Z_2) = \frac{1}{2}$$

without computing the joint density by applying a symmetry argument. The joint density for $m = 2$ is given in the following lemma.

Lemma 4.2.1. *Let (X_1, X_2) and (Y_1, Y_2) be two independent samples from $\mathcal{N}(0, \Sigma)$, where*

$$\Sigma = \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}$$

is a positive definite matrix. The joint density of $(Z_1 = \frac{X_1}{Y_1}, Z_2 = \frac{X_2}{Y_2})$ is given by

$$f(z_1, z_2) = \frac{1 - \sigma^2}{(1 + z_1^2)(1 + z_2^2)\pi^2(1 - \rho^2)^{\frac{3}{2}}} \left(\sqrt{1 - \rho^2} + \rho \sin^{-1} \rho \right),$$

where

$$\rho = \frac{(1 + z_1 z_2)}{\sqrt{(1 + z_1^2)(1 + z_2^2)}} \sigma.$$

Proof. The cumulative probability distribution function equals

$$\begin{aligned}
F(z_1, z_2) &= \mathbb{P}(Z_1 \leq z_1, Z_2 \leq z_2) \\
&= \int_0^\infty \int_0^\infty \mathbb{P}(X_1 \leq z_1 y_1, X_2 \leq z_2 y_2 \mid Y_1 = y_1, Y_2 = y_2) f(y_1, y_2) dy_2 dy_1 \\
&\quad + \int_{-\infty}^0 \int_0^\infty \mathbb{P}(X_1 \geq z_1 y_1, X_2 \leq z_2 y_2 \mid Y_1 = y_1, Y_2 = y_2) f(y_1, y_2) dy_2 dy_1 \\
&\quad + \int_0^\infty \int_{-\infty}^0 \mathbb{P}(X_1 \leq z_1 y_1, X_2 \geq z_2 y_2 \mid Y_1 = y_1, Y_2 = y_2) f(y_1, y_2) dy_2 dy_1 \\
&\quad + \int_{-\infty}^0 \int_{-\infty}^0 \mathbb{P}(X_1 \geq z_1 y_1, X_2 \geq z_2 y_2 \mid Y_1 = y_1, Y_2 = y_2) f(y_1, y_2) dy_2 dy_1 \\
&= \int_0^\infty \int_0^\infty \int_{-\infty}^{z_1 y_1} \int_{-\infty}^{z_2 y_2} f(x_1, x_2) f(y_1, y_2) dx_2 dx_1 dy_2 dy_1 \\
&\quad + \int_{-\infty}^0 \int_0^\infty \int_{z_1 y_1}^\infty \int_{-\infty}^{z_2 y_2} f(x_1, x_2) f(y_1, y_2) dx_2 dx_1 dy_2 dy_1 \\
&\quad + \int_0^\infty \int_{-\infty}^0 \int_{-\infty}^{z_1 y_1} \int_{z_2 y_2}^\infty f(x_1, x_2) f(y_1, y_2) dx_2 dx_1 dy_2 dy_1 \\
&\quad + \int_{-\infty}^0 \int_{-\infty}^0 \int_{z_1 y_1}^\infty \int_{z_2 y_2}^\infty f(x_1, x_2) f(y_1, y_2) dx_2 dx_1 dy_2 dy_1
\end{aligned}$$

By taking derivatives with respect to z_1 and z_2 we get the joint density.

$$\begin{aligned}
f(z_1, z_2) &= \int_0^\infty \int_0^\infty y_1 y_2 f(z_1 y_1, z_2 y_2) f(y_1, y_2) dy_2 dy_1 \\
&\quad - \int_{-\infty}^0 \int_0^\infty y_1 y_2 f(z_1 y_1, z_2 y_2) f(y_1, y_2) dy_2 dy_1 \\
&\quad - \int_0^\infty \int_{-\infty}^0 y_1 y_2 f(z_1 y_1, z_2 y_2) f(y_1, y_2) dy_2 dy_1 \\
&\quad + \int_{-\infty}^0 \int_{-\infty}^0 y_1 y_2 f(z_1 y_1, z_2 y_2) f(y_1, y_2) dy_2 dy_1 \\
&= \int_{-\infty}^\infty \int_{-\infty}^\infty |y_1 y_2| f(z_1 y_1, z_2 y_2) f(y_1, y_2) dy_2 dy_1 \\
&= \frac{1}{4\pi^2} \frac{1}{1 - \sigma^2} \int_{-\infty}^\infty \int_{-\infty}^\infty |y_1 y_2| \exp\left(-\frac{1}{2(1 - \sigma^2)} [a_1 y_1^2 - 2a_{12} \sigma y_1 y_2 + a_2 y_2^2]\right) dy_2 dy_1,
\end{aligned}$$

where $a_1 = 1 + z_1^2$, $a_{12} = 1 + z_1 z_2$, and $a_2 = 1 + z_2^2$. We make the following change of variables

$$w_i = \sqrt{a_i} \sqrt{\frac{1 - \sigma^2 \frac{a_{12}^2}{a_1 a_2}}{1 - \sigma^2}} y_i, \quad i = 1, 2$$

and define a random vector (W_1, W_2) with distribution

$$\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where

$$\rho = \frac{a_{12}}{\sqrt{a_1 a_2}} \sigma.$$

With these transformations we get

$$\begin{aligned} f(z_1, z_2) &= \frac{1}{2\pi} \frac{1 - \sigma^2}{(1 - \rho^2)^{\frac{3}{2}}} \frac{1}{a_1 a_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |w_1 w_2| f(w_1, w_2) dw_2 dw_1 \\ &= \frac{1}{2\pi} \frac{1 - \sigma^2}{(1 - \rho^2)^{\frac{3}{2}}} \frac{1}{a_1 a_2} E(|W_1 W_2|) \\ &= \frac{1}{2\pi} \frac{1 - \sigma^2}{(1 - \rho^2)^{\frac{3}{2}}} \frac{1}{a_1 a_2} \frac{2}{\pi} \left(\sqrt{1 - \rho^2} + \rho \sin^{-1} \rho \right) \\ &= \frac{1 - \sigma^2}{a_1 a_2 \pi^2 (1 - \rho^2)^{\frac{3}{2}}} \left(\sqrt{1 - \rho^2} + \rho \sin^{-1} \rho \right) \end{aligned}$$

For the second to last equality we used the formula about the absolute moments of a product of two Gaussians given in [43, Chapter 46, Section 3]. \square

4.3 Bounds on the MLE existence probability

In this section, we compute bounds on $\mathbb{P}(\text{MLE exists}, n = 2, \Sigma_m^{\text{true}}, G_m^{\text{assumed}} = \text{cycle})$. We will prove that under some regularity conditions on the condition number $\mathcal{K}(\Sigma_m^{\text{true}})$, the MLE existence probability goes to 1 for two observations and a growing cycle length. The condition number of a matrix is defined as the ratio of its largest eigenvalue to its smallest eigenvalue. A problem with a low condition number is said to be well-conditioned.

Theorem 4.3.1. *Denote by k_m the minimal condition number of Σ_m^{true} after scaling, i.e.*

$$k_m = \min_{D \text{ diagonal}} \mathcal{K}(D \Sigma_m^{\text{true}} D).$$

Then

$$1 - k_m^m \frac{2}{(m-1)!} \leq \mathbb{P}(\text{MLE exists}, n=2, \Sigma_m^{\text{true}}, G_m^{\text{assumed}} = \text{cycle}) \leq 1 - \frac{1}{k_m^m} \frac{2}{(m-1)!},$$

and

$$\mathbb{P}(\text{MLE exists}, n=2, \Sigma_m^{\text{true}}, G_m^{\text{assumed}} = \text{cycle}) \xrightarrow{m \rightarrow \infty} 1 \quad \text{if} \quad k_m < \frac{m^{1-\frac{1}{2m}}}{e} \quad \text{for all } m.$$

Proof. We denote by A the event that the MLE does not exist for $n=2$ observations and $G_m^{\text{assumed}} = \text{cycle}$. We write \mathbb{Q} for the multivariate probability measure with underlying covariance matrix

$$\Sigma_{\mathbb{Q}} = \begin{pmatrix} \Sigma_m^{\text{true}} & 0 \\ 0 & \Sigma_m^{\text{true}} \end{pmatrix},$$

and \mathbb{P} for the multivariate probability measure with underlying covariance matrix

$$\Sigma_{\mathbb{P}} = \begin{pmatrix} \alpha \text{Id}_m & 0 \\ 0 & \alpha \text{Id}_m \end{pmatrix},$$

where Id_m denotes the $m \times m$ identity matrix. Note that we listed the $2m$ random variables as one vector for simpler notation. We are interested in bounds on $\mathbb{Q}(A)$ and will compute these by using the formula for $\mathbb{P}(A)$ given in Corollary 4.1.2. Note that $\mathbb{P}(A)$ is independent of the value α .

$$\begin{aligned} \mathbb{Q}(A) &= \int_A \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P} \\ &= \frac{\sqrt{|\Sigma_{\mathbb{P}}|}}{\sqrt{|\Sigma_{\mathbb{Q}}|}} \int_A \exp\left(-\frac{1}{2}x^T (\Sigma_{\mathbb{Q}}^{-1} - \Sigma_{\mathbb{P}}^{-1})x\right) d\mathbb{P} \\ &= \frac{\alpha^m}{|\Sigma_m^{\text{true}}|} \int_A \exp\left(-\frac{1}{2}x^T \left(\Sigma_{\mathbb{Q}}^{-1} - \frac{1}{\alpha} \text{Id}_{2m}\right)x\right) d\mathbb{P} \end{aligned}$$

We choose α equal to the largest eigenvalue of Σ_m^{true} . Then the matrix

$$\Sigma_{\mathbb{Q}}^{-1} - \frac{1}{\alpha} \text{Id}_{2m}$$

is positive semidefinite and

$$\exp\left(-\frac{1}{2}x^T \left(\Sigma_{\mathbb{Q}}^{-1} - \frac{1}{\alpha} \text{Id}_{2m}\right)x\right)$$

is maximized by $x = 0$. With this choice of α and by applying Corollary 4.1.2 we get

$$\begin{aligned} \mathbb{Q}(A) &\leq \frac{\alpha^m}{|\Sigma_m^{\text{true}}|} \int_A d\mathbb{P} \\ &= \frac{\alpha^m}{|\Sigma_m^{\text{true}}|} \frac{2}{(m-1)!} \\ &\leq \mathcal{K}(\Sigma_m^{\text{true}})^m \frac{2}{(m-1)!} \end{aligned}$$

In order to get a lower bound on $\mathbb{Q}(A)$ we choose α equal to the smallest eigenvalue of Σ_m^{true} . Then the matrix

$$\Sigma_{\mathbb{Q}}^{-1} - \frac{1}{\alpha} Id_{2m}$$

is negative semidefinite and

$$\exp\left(-\frac{1}{2}x^T \left(\Sigma_{\mathbb{Q}}^{-1} - \frac{1}{\alpha} Id_{2m}\right) x\right)$$

is minimized by $x = 0$. Similarly to the previous argument we get

$$\mathbb{Q}(A) \geq \frac{1}{\mathcal{K}(\Sigma_m^{\text{true}})^m} \frac{2}{(m-1)!}$$

Because the MLE existence probability is invariant under rescaling of the covariance matrix, we get

$$1 - k_m^m \frac{2}{(m-1)!} \leq \mathbb{P}(\text{MLE exists, } n = 2, \Sigma_m^{\text{true}}, G_m^{\text{assumed}} = \text{cycle}) \leq 1 - \frac{1}{k_m^m} \frac{2}{(m-1)!}.$$

We use the following lower bound on $m!$ resulting from Stirling's approximation

$$m! \geq \sqrt{2\pi m} m^{m+\frac{1}{2}} e^{-m}$$

to conclude that

$$\begin{aligned} k_m^m \frac{2}{(m-1)!} &\leq k_m^m \frac{2m}{\sqrt{2\pi m} m^{m+\frac{1}{2}} e^{-m}} \\ &= \sqrt{\frac{2}{\pi}} \left(k_m m^{\frac{1}{2m}-1} e\right)^m, \end{aligned}$$

which converges to zero for $m \rightarrow \infty$ if and only if

$$k_m < \frac{m^{1-\frac{1}{2m}}}{e} \quad \text{for all } m.$$

□

It is interesting that the upper bound on the MLE existence probability found in this theorem increases with increasing condition number, while the lower bound on the MLE existence probability decreases with increasing condition number. Thus, ill-conditioned covariance matrices represent the worst and the best examples with respect to MLE existence. In the following, we give two examples explaining this phenomena.

Example 4.3.2. Let (X_1, X_2, X_3, X_4) and (Y_1, Y_2, Y_3, Y_4) be two independent samples from $\mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} 1 & 0 & \sigma & 0 \\ 0 & 1 & 0 & \sigma \\ \sigma & 0 & 1 & 0 \\ 0 & \sigma & 0 & 1 \end{pmatrix}.$$

For a large value of σ , we expect the resulting lines given by (X_i, Y_i) to lie as shown in Figure 4.1. The lines are not graph consecutive, meaning that the MLE exists. More precisely, let $Z_i = \frac{X_i}{Y_i}$, $i = 1, \dots, 4$. Using Lemma 4.2.1 we can prove that

$$\mathbb{P}(Z_1 \leq Z_2 \leq Z_3 \leq Z_4) \xrightarrow{\sigma \rightarrow 1} 0.$$

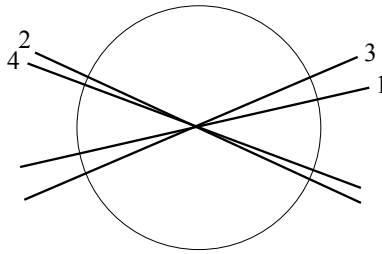


Figure 4.1: *Strong correlation between variables 1 and 3 and between variables 2 and 4 leads to lines which are not graph consecutive.*

Proof.

$$\begin{aligned}
\mathbb{P}(Z_1 \leq Z_2 \leq Z_3 \leq Z_4) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{P}(Z_1 \leq z_2 \leq Z_3 \leq z_4 \mid Z_2 = z_2, Z_4 = z_4) f(z_2, z_4) dz_4 dz_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{P}(Z_1 \leq z_2 \leq Z_3 \leq z_4) f(z_2, z_4) dz_4 dz_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{z_2} \int_{z_2}^{z_4} f(z_1, z_3) f(z_2, z_4) dz_3 dz_1 dz_4 dz_2 \xrightarrow{\sigma \rightarrow 1} 0,
\end{aligned}$$

because $f(z_1, z_3)$ and $f(z_2, z_4)$ go to zero as σ approaches 1 (Lemma 4.2.1). \square

This example can easily be extended to $m \geq 4$. The above result holds for $m \geq 4$ when there exist indices $1 \leq i < j < k < l \leq m$ such that the submatrix of the covariance matrix corresponding to the indices i, j, k, l is of the form

$$\begin{pmatrix} 1 & 0 & \sigma & 0 \\ 0 & 1 & 0 & \sigma \\ \sigma & 0 & 1 & 0 \\ 0 & \sigma & 0 & 1 \end{pmatrix}.$$

This example shows that for fixed m and for any $\epsilon > 0$ one can find a positive definite matrix Σ_m^{true} such that

$$\mathbb{P}(\text{MLE exists}, n = 2, \Sigma_m^{\text{true}}, G_m^{\text{assumed}} = \text{cycle}) > 1 - \epsilon.$$

Example 4.3.3. In the following, we describe a class of covariance matrices, for which the converse happens, namely, for fixed m and for any $\epsilon > 0$ one can find a positive definite matrix Σ_m^{true} such that

$$\mathbb{P}(\text{MLE exists}, n = 2, \Sigma_m^{\text{true}}, G_m^{\text{assumed}} = \text{cycle}) < \epsilon.$$

Intuitively, we would like to find a class of covariance matrices, which make the resulting data vectors as spread out as possible (see Figure 4.2 for $m = 7$). Simulations suggest that for odd m a nearly singular circulant concentration matrix of the form

$$(\Sigma_m^{\text{true}})^{-1} = \begin{pmatrix} 1 & \rho & 0 & \cdots & 0 & \rho \\ \rho & 1 & \rho & \ddots & \ddots & 0 \\ 0 & \rho & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \rho & 0 \\ 0 & \ddots & \ddots & \rho & 1 & \rho \\ \rho & 0 & \cdots & 0 & \rho & 1 \end{pmatrix}$$

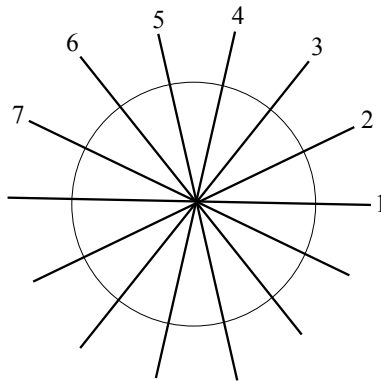


Figure 4.2: Lines are in graph consecutive order with equal angle between consecutive lines

leads to this phenomena. It is an open problem to prove that, indeed, for any $\epsilon > 0$ one can find a $\rho \in \mathbb{R}$ such that

$$\mathbb{P}(\text{MLE exists, } n = 2, \Sigma_m^{\text{true}}, G_m^{\text{assumed}} = \text{cycle}) < \epsilon,$$

when m is odd.

In this chapter we studied Gaussian cycles from an asymptotic point of view. We believe that our results in Chapter 3, in particular the results about bipartite graphs and the 3×3 grid, can be used in a similar way to study these models for a growing graph size and answer the question of existence of the MLE for very large bipartite graphs and grids.

Part II

Algebraic statistics and disease association studies

Chapter 5

Using Markov bases for disease association studies

Rapid research progress in genotyping techniques have allowed large genome-wide association studies. Existing methods often focus on determining associations between single loci and a specific phenotype. However, a particular phenotype is usually the result of complex relationships between multiple loci and the environment. In this chapter, we describe a two-stage method for detecting epistasis by combining the traditionally used single-locus search with a search for multiway interactions. Our method is based on an extended version of Fisher's exact test. To perform this test, a Markov chain is constructed on the space of multidimensional contingency tables using the elements of a Markov basis as moves. We test our method on simulated data and compare it to a two-stage logistic regression method and to a fully Bayesian method, showing that we are able to detect the interacting loci when other methods fail to do so. Finally, we apply our method to a genome-wide data set consisting of 685 dogs and identify epistasis associated with canine hair length for four pairs of single nucleotide polymorphisms (SNPs).

The material of this chapter is joint work with Anna-Sapfo Malaspinas. It will be published in the Journal of Algebraic Statistics in a paper of the same title.

5.1 Introduction

Conditions with genetic components such as cancer, heart disease, and diabetes, are the most common causes of mortality in developed countries. Therefore, the mapping of genes involved in such complex diseases represents a major goal of human genetics. However, genetic variants associated with complex diseases are hard to detect. Indeed, only a small portion of the heritability of complex diseases can be explained by the variants identified so far. This led to several hypotheses (see e.g. [51]). One of them is that most common diseases are caused by several rare variants with low effects, rather than a few common variants with

large effects [60]. Another hypothesis is that the variants interact in order to produce the disease phenotype and independently only explain a small fraction of the genetic variance. In this chapter, we mainly focus on the interaction hypothesis, but we will also discuss the relevance of our method to the rare variant hypothesis along the way.

Recent development of methods to screen hundreds of thousands of SNPs has allowed the discovery of over 50 disease susceptibility loci with marginal effects [56]. Genome-wide association studies have hence proven to be fruitful in understanding complex multifactorial traits. The absence of reports of interacting loci, however, shows the need for better methods for detecting not only marginal effects of specific loci, but also interactions of loci. Although some progress in detecting interactions has been achieved in the last few years using simple log-linear models, these methods remain inefficient to detect interactions for large-scale data [4].

Many models of interaction have been presented in the past, as for example the additive model and the multiplicative model. The former model assumes that the SNPs act independently, and a single marker approach seems to perform well. In the multiplicative model, SNPs interact in the sense that the presence of two (or more) variants have a stronger effect than the sum of the effects of each single SNP. We will discuss such models in more detail in Section 5.2.1. A complete classification of two-locus interaction models has been given in [36].

In the method described in this chapter, we first reduce the potential interacting SNPs to a small number by filtering all SNPs genome-wide with a single locus approach. The loci achieving some threshold are then further examined for interactions. Such a two-stage approach has been suggested in [52]. For some models of interaction, they show that the two-stage approach outperforms the single-locus search and performs at least as well as when testing for interaction within all subsets of k SNPs.

Single locus methods consider each SNP individually and test for association based on differences in genotypic frequencies between case and control individuals. Widely used methods for the single-locus search are the χ^2 goodness-of-fit test or Fisher's exact test together with a Bonferroni correction of the p-values to account for the large number of tests performed. We suggest using Fisher's exact test as a first stage to rank the SNPs by their p-value and select a subset of SNPs, which is then further analyzed. Under the rare variant hypothesis the resulting contingency tables are sparse and it is desirable to test for interactions within the selected subset using an exact test. We suggest using Markov bases for this purpose.

In Section 5.2, we define three models of interaction and present our algorithm for detecting epistasis using Markov bases in hypothesis testing. In Section 5.3, we test our method on simulated data and make a comparison to logistic regression and BEAM, a Bayesian approach [79]. Finally, we run our algorithm on a genome-wide dataset from dogs [16] to test for epistasis related to canine hair length.

5.2 Method

5.2.1 Models of interaction

In this chapter, we mainly study the interaction between two SNPs and a binary phenotype, as for example the disease status of an individual. However, our method can be easily generalized for studying interaction between three or more SNPs and a phenotype with three or more states. We show a generalization in Section 5.3.4, where we analyze a genome-wide dataset from dogs and, *inter alia*, test for interaction between three SNPs and a binary hair length phenotype (short hair versus long hair).

The binary phenotype is denoted by D , taking values 0 and 1. We assume that the SNPs are polymorphic with only two possible nucleotides. The two SNPs are denoted by X and Y , each with genotypes taking values 0, 1 and 2 representing the number of minor alleles. We investigate three different models of interaction: a control model, an additive model, and a multiplicative model. The parameterization is given in the following tables showing the odds of having a specific phenotype

$$\frac{\mathbb{P}(D = 1|\text{genotype})}{\mathbb{P}(D = 0|\text{genotype})}.$$

		Y		
		0	1	2
• Control model:	0	ϵ	ϵ	ϵ
	X 1	ϵ	ϵ	ϵ
	2	ϵ	ϵ	ϵ

		Y		
		0	1	2
• Additive model:	0	ϵ	$\epsilon\beta$	$\epsilon\beta^2$
	X 1	$\epsilon\alpha$	$\epsilon\alpha\beta$	$\epsilon\alpha\beta^2$
	2	$\epsilon\alpha^2$	$\epsilon\alpha^2\beta$	$\epsilon\alpha^2\beta^2$

		Y		
		0	1	2
• Multiplicative model:	0	ϵ	$\epsilon\beta$	$\epsilon\beta^2$
	X 1	$\epsilon\alpha$	$\epsilon\alpha\beta\delta$	$\epsilon\alpha\beta^2\delta^2$
	2	$\epsilon\alpha^2$	$\epsilon\alpha^2\beta\delta^2$	$\epsilon\alpha^2\beta^2\delta^4$

These three models can also be expressed as log-linear models. We denote the state of X by i , the state of Y by j , and the state of D by k . If n_{ijk} describes the expected cell counts in a $3 \times 3 \times 2$ contingency table, then the three models can be expressed in the following way, where the γ terms represent the effects the variables have on the cell counts (e.g. γ_i^X represents the main effect for X), and α, β, δ , and ϵ are defined by the odds of having a specific phenotype shown in the above tables:

$$\text{Control model:} \quad \log(n_{ijk}) = \gamma + \gamma_i^X + \gamma_j^Y + \gamma_{ij}^{XY} + k \log(\epsilon)$$

$$\text{Additive model:} \quad \log(n_{ijk}) = \gamma + \gamma_i^X + \gamma_j^Y + \gamma_{ij}^{XY} + k \log(\epsilon) + ik \log \alpha + jk \log \beta$$

$$\text{Multiplicative model:} \quad \log(n_{ijk}) = \gamma + \gamma_i^X + \gamma_j^Y + \gamma_{ij}^{XY} + k \log(\epsilon) + ik \log \alpha + jk \log \beta + ijk \log \delta$$

Note that in the additive model the interaction effect for SNP X (SNP Y) and the disease status is additive with respect to the number of causative SNPs i (j), whereas in the multiplicative model there is an additional 3-way interaction effect between SNPs X, Y , and the disease status, which is multiplicative in the number of causative SNPs i, j . From the representation as log-linear models we can deduce the nesting relationship shown on the Venn diagram in Figure 5.1. Note that the additive model corresponds to the intersection of the no 3-way interaction model ($\log(n_{ijk}) = \gamma + \gamma_i^X + \gamma_j^Y + \gamma_k^D + \gamma_{ij}^{XY} + \gamma_{ik}^{XD} + \gamma_{jk}^{YD}$) with the multiplicative model, and the control model is nested within the additive model.

In a biological context, interaction between markers (or SNPs) is usually used as a synonym for *epistasis*. Cordell [18] gives a broad definition: “Epistasis refers to departure from ‘independence’ of the effects of different genetic loci in the way they combine to cause disease”. Epistasis is for example the result of a multiplicative effect between two markers (i.e. $\log(\delta) \neq 0$ in the multiplicative model).

In contrast, in a mathematical context interaction is used as synonym for *dependence*. Two markers are said to be interacting if they are dependent, i.e.

$$\mathbb{P}(\text{marker 1} = i, \text{marker 2} = j) \neq \mathbb{P}(\text{marker 1} = i)\mathbb{P}(\text{marker 2} = j).$$

In general, in association studies the goal is to find a set of markers that are associated with a specific phenotype. In what follows, we will use the term interaction as synonym for dependence and the term epistasis with respect to a specific phenotype synonymously to the presence of a k -way interaction ($k \geq 3$) between $k - 1$ SNPs and a specific phenotype. The epistatic models are indicated by the shading in Figure 5.1.

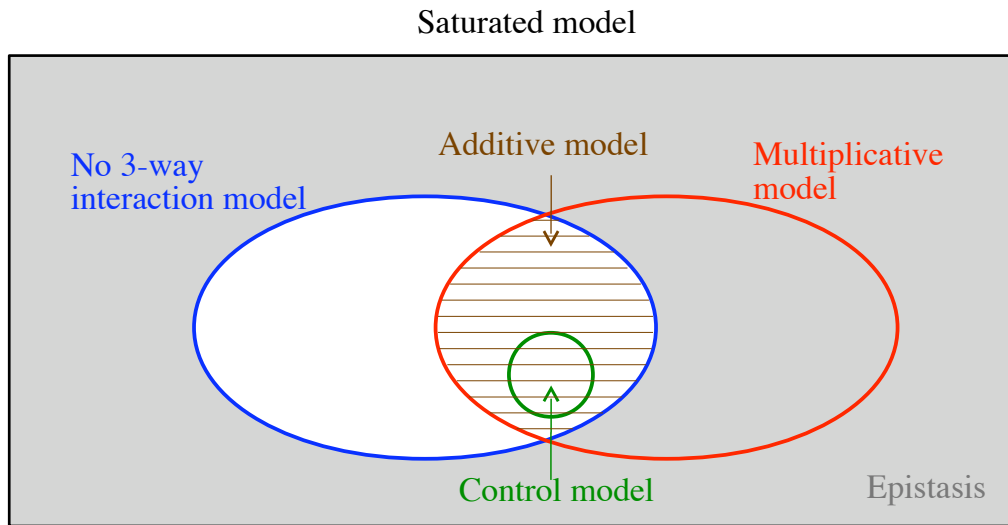


Figure 5.1: *Nesting relationship of the control model, the additive model, and the multiplicative model. The intersection of the no 3-way interaction model with the multiplicative model corresponds to the additive model. The shading indicates the presence of epistasis.*

5.2.2 Algorithm

The χ^2 goodness-of-fit-test is the most widely used test for detecting interaction within contingency tables. Under independence the χ^2 statistic is asymptotically χ^2 distributed. However, this approximation is problematic when some cell counts are small, which is often the case in contingency tables resulting from association studies and particularly problematic under the rare variant hypothesis. The other widely used test is Fisher's exact test. As its name suggests, it has the advantage of being exact. But it is a permutation test and therefore computationally more intensive. For tables with large total counts or tables of higher dimension enumerating all possible tables with given margins is not feasible.

Diaconis and Sturmfels [23] describe an extended version of Fisher's exact test using Markov bases. A Markov basis for testing a specific interaction model is a set of moves connecting all contingency tables with the same sufficient statistics. So a Markov basis allows constructing a Markov chain on the set of contingency tables with given margins and computing the p -value of a given contingency table using the resulting posterior distribution. Such a test can be used for analyzing multidimensional tables with large total counts. In addition, it has been shown in [23] that the resulting posterior distribution is a good approximation of the exact distribution of the χ^2 -statistic even for very sparse contingency tables, leading to a substantially more accurate interaction test than the χ^2 -test for sparse tables. Useful properties of Markov bases can be found in [25].

Table 5.1: *Standard interaction models for three-dimensional contingency tables.*

Model	Minimal sufficient statistics	Expected counts
(X, Y, D)	$(n_{i..}), (n_{.j.}), (n_{..k})$	$\hat{n}_{ijk} = \frac{n_{i..}n_{.j.}n_{..k}}{(n_{...})^2}$
(XY, D)	$(n_{ij.}), (n_{..k})$	$\hat{n}_{ijk} = \frac{n_{ij.}n_{..k}}{(n_{...})}$
(XD, Y)	$(n_{i.k}), (n_{.j.})$	$\hat{n}_{ijk} = \frac{n_{i.k}n_{.j.}}{(n_{...})}$
(X, YD)	$(n_{i..}), (n_{.jk})$	$\hat{n}_{ijk} = \frac{n_{.jk}n_{i..}}{(n_{...})}$
(XY, YD)	$(n_{ij.}), (n_{.jk})$	$\hat{n}_{ijk} = \frac{n_{ij.}n_{.jk}}{(n_{.j.})}$
(XY, XD)	$(n_{ij.}), (n_{i.k})$	$\hat{n}_{ijk} = \frac{n_{ij.}n_{i.k}}{(n_{i..})}$
(XD, YD)	$(n_{i.k}), (n_{.jk})$	$\hat{n}_{ijk} = \frac{n_{i.k}n_{.jk}}{(n_{..k})}$
(XY, XD, YD)	$(n_{ij.}), (n_{i.k}), (n_{.jk})$	Iterative proportional fitting

The Markov basis of the null model can be computed using the software `4ti2` [1]. An example is given in Table 5.3. Then a Markov chain is started in the observed $3 \times 3 \times 2$ data table using the elements of the Markov basis as moves in the Metropolis-Hastings steps. At each step the χ^2 statistic is computed. Its posterior distribution is an approximation of the exact distribution of the χ^2 statistic.

Interaction tests with the extended version of Fisher's exact test

We present various hypotheses that can easily be tested using Markov bases and discuss a hypothesis that is particularly interesting for association studies. The corresponding Markov basis can be found in Table 5.3. For simplicity we constrain this discussion to the case of two SNPs and a binary phenotype.

Table 5.1 consists of the standard log-linear models on three variables. Their fit to a given data table can be computed using the extended version of Fisher's exact test. We use the notation presented in [12] to denote the different models. Interaction is assumed between the variables not separated by commas in the model. So the model (X, Y, D) in Table 5.1 represents the independence model, the model (XY, XD, YD) the no 3-way interaction model and the other models are intermediate models. For association studies the no 3-way interaction model (XY, XD, YD) is particularly interesting and will be used as null model in our testing procedure.

Performing the extended version of Fisher's exact test involves sampling from the space of contingency tables with fixed minimal sufficient statistics and computing the χ^2 statistic. So the minimal sufficient statistics and the expected counts for each cell of the table need to be calculated. These are given in Table 5.1. If a loop is present in the model configuration as for example in the no 3-way interaction model (this model can be rewritten as (XY, YD, DX)), then there is no closed-form estimator for the cell counts (see [12]). But in this case, estimates can be achieved by iterative proportional fitting (e.g. [28]).

Table 5.2: *Testing for association between haplotypes and phenotype.*

		Phenotype status:		Total:
		0	1	
Haplotype:	00	n_{000}	n_{001}	$n_{00.}$
	01	n_{010}	n_{011}	$n_{01.}$
	10	n_{100}	n_{101}	$n_{10.}$
	11	n_{110}	n_{111}	$n_{11.}$
Total:		$n_{..0}$	$n_{..1}$	$n_{..}$

It is important to note that testing for epistasis necessarily implies working with multidimensional contingency tables and is not possible in the collapsed two-dimensional table shown above. In this table, the two SNPs are treated like a single variable and we consider their haplotype. The sufficient statistics for the model described in Table 5.2 are the row and column sums ($n_{i.j.}$) and ($n_{..k}$). So testing for association in this collapsed table is the same as using (XY, D) as null model. In this case, the null hypothesis would be rejected even in the presence of marginal effects only, showing that testing for epistasis in Table 5.2 is impossible.

Hypothesis testing with the extended version of Fisher's exact test

Our goal is to detect epistasis when present. According to the definition of epistasis in Section 5.2.1 and as shown in Figure 5.1, epistasis is present with regard to two SNPs and a specific phenotype, when a 3-way interaction is found. So we suggest using as null hypothesis the no 3-way interaction model and testing this hypothesis with the extended version of Fisher's exact test. The corresponding Markov basis consists of 15 moves and is given in Table 5.3. It can be used to compute the posterior distribution of the χ^2 statistic and approximate the exact p-value of the data table. If the p-value is lower than some threshold, we reject the null hypothesis of no epistasis.

Although in this chapter we focus merely on epistasis, it is worth noting that one can easily build tests for different types of interaction using Markov bases. If one is interested in detecting whether the epistatic effect is of multiplicative nature, one can perform the extended version of Fisher's exact test on the contingency tables, which have been classified as epistatic, using the multiplicative model as null hypothesis. In this case, the corresponding Markov basis consists of 49 moves. Similarly, if one is interested in detecting additive effects, one can use the additive model as null hypothesis and test the contingency tables, which have been classified as non-epistatic. In this case, the corresponding Markov basis consists of 156 moves. The Markov bases for these tests can be found on our website¹.

¹<http://www.carolineuhler.com/epistasis.htm>

Table 5.3: *Markov basis for the no 3-way interaction model on a $3 \times 3 \times 2$ table, where the tables are reported as vectors*

$(n_{000}, n_{100}, n_{200}, n_{010}, n_{110}, n_{210}, n_{020}, n_{120}, n_{220}, n_{001}, n_{101}, n_{201}, n_{011}, n_{111}, n_{211}, n_{021}, n_{121}, n_{221})$.

$$\begin{aligned}
f_1 &= (0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & -1) \\
f_2 &= (0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 & -1) \\
f_3 &= (1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 & -1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & -1) \\
f_4 &= (0 & 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & -1) \\
f_5 &= (0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 & -1 & 0) \\
f_6 &= (1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0) \\
f_7 &= (1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0) \\
f_8 &= (1 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0) \\
f_9 &= (0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0) \\
f_{10} &= (0 & 1 & -1 & -1 & 0 & 1 & 1 & -1 & 0 & 0 & -1 & 1 & 1 & 0 & -1 & -1 & 1 & 0) \\
f_{11} &= (1 & 0 & -1 & 0 & -1 & 1 & -1 & 1 & 0 & -1 & 0 & 1 & 0 & 1 & -1 & 1 & -1 & 0) \\
f_{12} &= (-1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 1 & 1 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & -1) \\
f_{13} &= (1 & -1 & 0 & 0 & 1 & -1 & -1 & 0 & 1 & -1 & 1 & 0 & 0 & -1 & 1 & 1 & 0 & -1) \\
f_{14} &= (1 & 0 & -1 & -1 & 1 & 0 & 0 & -1 & 1 & -1 & 0 & 1 & 1 & -1 & 0 & 0 & 1 & -1) \\
f_{15} &= (0 & 1 & -1 & 1 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 1 & -1 & 1 & 0 & 1 & 0 & -1)
\end{aligned}$$

The elements of the Markov basis can be viewed as binomials in the variables n_{ijk} . We write each element of the Markov basis as a long vector $f_i \in \mathbb{Z}^{18}$ and decompose it into two parts

$$f_i = f_i^+ - f_i^-,$$

where f_i^+ and f_i^- are non-negative and have disjoint support. Then the elements of the Markov basis can be interpreted as binomials

$$n^{f_i^+} - n^{f_i^-},$$

where n is the vector of variables n_{ijk} . For example, the first element of the Markov basis in Table 5.3 can be interpreted as the binomial

$$n_{010}n_{220}n_{211}n_{021} - n_{210}n_{020}n_{011}n_{221}.$$

These binomials generate the toric ideal corresponding to the underlying model. Under the no 3-way interaction model the Markov basis consists of 9 elements of degree 4 and 6 elements of degree 6. The 49 moves under the multiplicative model consist of 7 moves of degree 4, 14 moves of degree 6, 24 moves of degree 8, and 4 moves of degree 10. Finally, the Markov basis corresponding to the additive model consists of 30 moves of degree 4, 78 moves of degree 6, 40 moves of degree 8, and 8 moves of degree 10.

The no 3-way interaction model, the multiplicative model and the additive model are log-linear models, which can be transformed into logistic models. We represent the interaction effects of the covariates by a matrix \mathcal{A} , where the columns of \mathcal{A} correspond to the SNP values 00, 10, 20, 01, 11, 21, 02, 12, 22. Then

$$\mathcal{A} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

for the no 3-way interaction model,

$$\mathcal{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & 2 & 4 \end{pmatrix}$$

for the multiplicative model, and

$$\mathcal{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \end{pmatrix}$$

for the additive model.

Let Id denote the identity matrix. The matrix

$$\Lambda(\mathcal{A}) = \begin{pmatrix} \mathcal{A} & 0 \\ \text{Id} & \text{Id} \end{pmatrix}$$

is called the Lawrence lifting of \mathcal{A} . The Lawrence liftings of the above matrices \mathcal{A} define the toric ideals of the three models under consideration.

It is interesting that the correspondence between the log-linear models and the logistic models translates into the fact that the Graver basis corresponding to \mathcal{A} coincides with the unique minimal Markov basis corresponding to the Lawrence lifting $\Lambda(\mathcal{A})$ ([65, Theorem 7.1]).

5.3 Results

In this section, we first conduct a simulation study to evaluate the performance of the suggested method. We then compare our method to a two-stage logistic regression approach and to BEAM ([79]). Logistic regression is a widely used method for detecting epistasis

within a selection of SNPs. BEAM is a purely Bayesian method for detecting epistatic interactions on a genome-wide scale. We end this section by applying our method to a genome-wide data set consisting of 685 dogs with the goal of finding epistasis associated with canine hair length.

5.3.1 Simulation study

We simulated a total of 50 potential association studies with 400 cases and 400 controls for three different minor allele frequencies of the causative SNPs and the three models of interaction presented in Section 5.2.1. We chose as minor allele frequencies (MAF) 0.1, 0.25 and 0.4. The parameters for the three models of interaction were determined numerically by fixing the marginal effect measured by the effect size

$$\lambda_i := \frac{p(D = 1|g_i = 1) p(D = 0|g_i = 0)}{p(D = 0|g_i = 1) p(D = 1|g_i = 0)} - 1$$

and the prevalence

$$\pi := \sum_{g_1, g_2} p(D|g_1, g_2)p(g_1, g_2).$$

For our simulations, we used an effect size of $\lambda_1 = \lambda_2 = 1$ and a sample prevalence of $\pi = 0.5$. Choosing in addition $\alpha = \beta$ in the additive model, and $\alpha = \beta$ and $\delta = 3\alpha$ in the multiplicative model determines all parameters of the interaction models and one can solve for α, β, δ and ϵ numerically.

The simulations were performed using HAP-SAMPLE ([76]) and were restricted to the SNPs typed with the Affy CHIP on chromosome 9 and chromosome 13 of the Phase I/II HapMap data², resulting in about 10,000 SNPs per individual. On each of the two chromosomes we selected one SNP to be causative. The causative SNPs were chosen consistent with the minor allele frequencies and far apart from any other marker (at least 20,000bp apart). Note that HAP-SAMPLE generates the cases and controls by resampling from HapMap. This means that the simulated data show linkage disequilibrium and allele frequencies similar to real data.

As suggested in [52], we took a two-stage approach for finding interacting SNPs. In the first step, we ranked all SNPs according to their p-value in Fisher's exact test on the 2x3 genotype table and selected the ten SNPs with the lowest marginal p-values. Within this subset, we then tested for interaction using the extended version of Fisher's exact test with the no 3-way interaction model as null hypothesis. We generated three Markov chains with 40,000 iterations each and different starting values, and used the tools described in [30] to assess convergence of the chains. This included analyzing the Gelman-Rubin statistic and the autocorrelations. After discarding an initial burn-in of 10,000 iterations, we combined

²<http://hapmap.ncbi.nlm.nih.gov/>

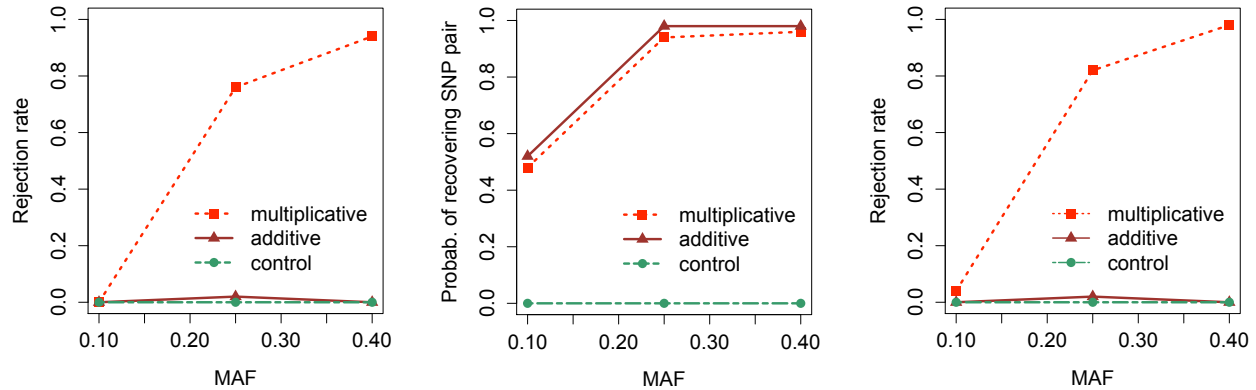


Figure 5.2: *Rejection rate of the no 3-way interaction test in the two-stage approach on 50 simulated association studies for $MAF=0.1$, $MAF=0.25$, and $MAF=0.4$ (left). Proportion of 50 association studies, in which the two causative SNPs were ranked among the ten SNPs with the lowest p -values by Fisher's exact test (middle). Rejection rate of the no 3-way interaction hypothesis using only the extended version of Fisher's exact test on the 50 causative SNP pairs (right).*

the remaining samples of the three chains to generate the posterior distribution of the χ^2 statistic.

In Figure 5.2 (left), we report the rejection rate of the no 3-way interaction hypothesis for each of the three minor allele frequencies. Per point in the figure we simulated 50 potential association studies. The power of our two-stage testing procedure corresponds to the curve under the multiplicative model. The higher the minor allele frequency, the more accurately we can detect epistasis. Under the additive model and the control model, no epistasis is present. We never rejected the null hypothesis under the control model and only once under the additive model, resulting in a high specificity of the testing procedure.

We also analyze the performance of each step separately. Figure 5.2 (middle) shows the performance of the first step and reports the proportion of 50 association studies, in which the two causative SNPs were ranked among the ten SNPs with the lowest p -values. Because Fisher's exact test measures marginal association, the curves under the additive model and the multiplicative model are similar.

Figure 5.2 (right) shows the performance of the second step in our method and reports the proportion of 50 association studies, in which the null hypothesis of no 3-way interaction was rejected using only the extended version of Fisher's exact test on the 50 causative SNP pairs.

5.3.2 Comparison to logistic regression

For validation, we compare the performance of our method to logistic regression via ROC curves. Logistic regression is probably the most widely used method for detecting epistasis

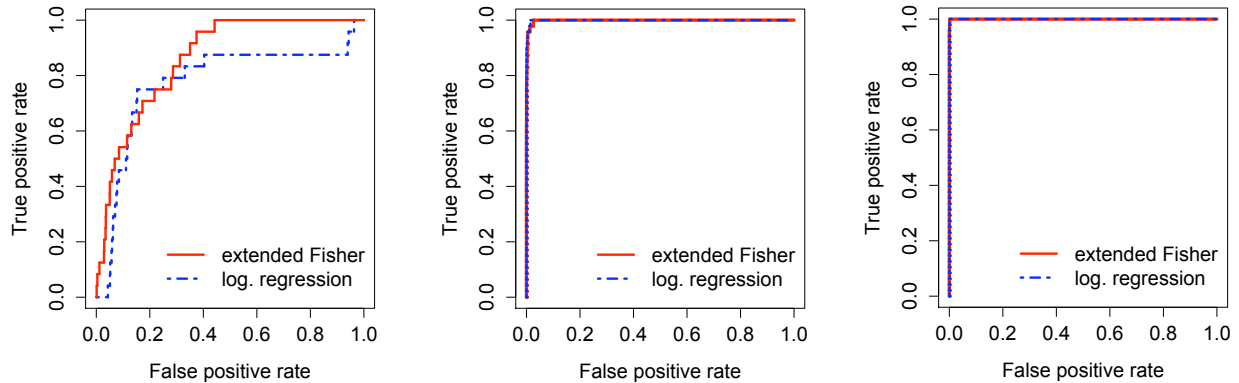


Figure 5.3: *ROC curves of the extended version of Fisher’s exact test and logistic regression for $MAF=0.1$ (left), $MAF=0.25$ (middle), and $MAF=0.4$ (right) based on the ten filtered SNPs.*

within a selection of SNPs nowadays. We base the comparison on the simulated association studies presented in the previous section using only the simulations under the multiplicative model. The structure of interaction within this model should favor logistic regression as logistic regression tests for exactly this kind of interaction.

As before, for each minor allele frequency and each of the 50 simulation studies we first filtered all SNPs with Fisher’s exact test and chose the ten SNPs with the lowest p-values for further analysis. Both causative SNPs are within the ten filtered SNPs for 19 (46) [45] out of the 50 simulation studies for $MAF=0.1$ ($MAF=0.25$) [$MAF=0.4$]. We then ran the extended version of Fisher’s exact test and logistic regression on all possible pairs of SNPs in the subsets consisting of the ten filtered SNPs. This results in $50 \cdot \binom{10}{2}$ tests per minor allele frequency with 19 (46) [45] true positives for $MAF=0.1$ ($MAF=0.25$) [$MAF=0.4$].

Because both methods, logistic regression and our method, require filtering all SNPs first, we compare the methods only based on the ten filtered SNPs. The ROC curves comparing the second stage of our method to logistic regression are plotted in Figure 5.3 showing that our method performs substantially better than logistic regression for $MAF=0.1$ with an area under the ROC curve of 0.861 compared to 0.773 for logistic regression. For $MAF=0.25$ and $MAF=0.4$ both methods have nearly perfect ROC curves with areas 0.9986 [0.99994] for our method compared to 0.9993 [0.99997] for logistic regression for $MAF=0.25$ [$MAF=0.4$].

5.3.3 Comparison to BEAM

We also compare our method to BEAM, a Bayesian approach for detecting epistatic interactions in association studies ([79]). We chose to compare our method to BEAM, because the authors show it is more powerful than a variety of other approaches including the step-

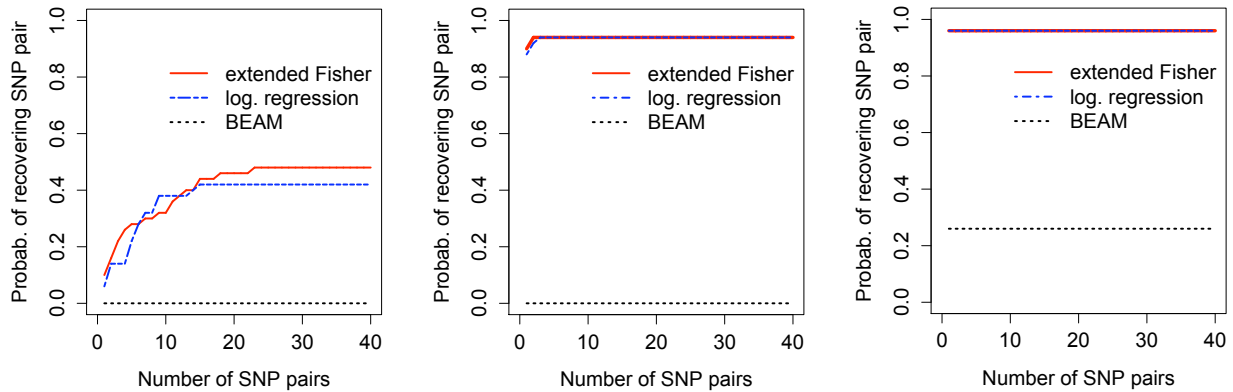


Figure 5.4: *Proportion of simulation studies for which the interacting SNP pair belongs to the x SNP pairs with the lowest p -values for $MAF=0.1$ (left), $MAF=0.25$ (middle), and $MAF=0.4$ (right).*

wise logistic regression approach, and it is one of the few recent methods that can handle genome-wide data.

In this method, all SNPs are divided into three groups, namely, SNPs that are not associated with the disease, SNPs that contribute to the disease risk only through main effects, and SNPs that interact to cause the disease. BEAM outputs the posterior probabilities for each SNP to belong to these three groups. The authors propose to use the results in a frequentist hypothesis-testing framework calculating the so called B-statistic and testing for association between each SNP or set of SNPs and the disease phenotype. BEAM was designed to increase the power to detect any association with the disease, and not to separate main effects from epistasis. Therefore, BEAM outputs SNPs that interact marginally OR through a k -way interaction with the disease. This does not match our definition of epistasis since the presence of marginal effects only already gives rise to a significant result using BEAM.

We compare our method to BEAM using the B-statistic. BEAM reports this statistic only for the pairs of SNPs which have a non-zero posterior probability of belonging to the third group. In addition, the B-statistic is automatically set to zero for the SNP pairs where any of the SNPs is found to be interacting marginally with the disease. We force BEAM to include the marginal effects into the B-statistic by choosing a significance level of zero for marginal effects. This should favor BEAM in terms of sensitivity.

We ran BEAM with the default parameters on our simulated datasets for the multiplicative model. Due to the long running time of BEAM, we based the comparison only on 1,000 SNPs out of the 10,000 SNPs simulated for the analysis in Section 5.3.1. BEAM takes about 10.6 hours for the analysis of one dataset with 10,000 SNPs and 400 cases and controls, whereas the same analysis with our method takes about 0.7 hours on an Intel Core 2.2 GHz laptop with 2 Gb memory.

In contrast to BEAM, our method is a stepwise approach, which makes a comparison via ROC curves difficult. We therefore compare the performance of all three tests by plotting for a fixed number x of SNP pairs the proportion of simulation studies for which the interacting SNP pair belongs to the x SNP pairs with the lowest p-values. The resulting curves are shown in Figure 5.4. Although the marginal effects were not extracted, BEAM has a very high false negative rate, attributing a p-value of 1 to the majority of SNPs, interacting and not interacting SNPs.

5.3.4 Genome-wide association study of hair length in dogs

We demonstrate the potential of our Markov basis method in genome-wide association studies by analyzing a hair length dataset consisting of 685 dogs from 65 breeds and containing 40,842 SNPs [16].

The individuals in [16] were divided into two groups for the hair length phenotype: 319 dogs from 31 breeds with long hair as cases and 364 from 34 breeds with short hair as controls. In the original study, it is shown that the long versus short hair phenotype is associated with a mutation (Cys95Phe) that changes exon one in the *fibroblast growth factor-5* (*FGF5* gene). Indeed, the SNP with the lowest p-value using Fisher’s exact test is located on chromosome 32 at position 7,100,913 for the Canmap dataset, i.e. about 300Kb apart from *FGF5*.

We ranked the 40,842 SNPs by their p-value using Fisher’s exact test and selected the 20 lowest ranked SNPs (about 0.05%) to test for 3-way interaction. Note that all 20 SNPs are significantly correlated (p-value < 0.05) with the phenotype. We found a significant p-value (< 0.05) for four out of the $\binom{20}{2}$ pairs. These pairs together with their p-values are listed in Table 5.3.4.

The pairs include six distinct SNPs located on five different chromosomes and the two SNPs lying on the same chromosome are not significantly interacting (p-value of 0.54). This means that a false positive correlation due to hitchhiking effects can likely be avoided. Hitchhiking effects are known to extend across long stretches of chromosomes in particular in domesticated species [54, 68, 71] consistent with the prediction of [63].

In order to identify potential pathways we first considered genes, which are close to the six SNPs we identified as interacting. To do so, we used the dog genome available through the ncbi website³. Most of the genes we report here have been annotated automatically. Our strategy was to consider the gene containing the candidate SNP (if any) and the immediate left and right neighboring gene, resulting in a total of two or three genes per SNP.

Among the six significantly interacting SNPs, four are located close to genes that have been shown to be linked to hair growth in other organisms. This is not surprising, since these SNPs also have a significant marginal association with hair growth. We here report the function of these candidate genes. The two other SNPs are located close to genes that we were not able to identify as functionally related to hair growth.

³<http://www.ncbi.nlm.nih.gov/genome/guide/dog/>, build 2.1

Table 5.4: Pairs of SNPs, which significantly interact with the hair length phenotype for the Canmap dataset. Question marks indicate that we were not able to identify a close-by gene which is functionally related to hair growth.

chromosome and location of SNPs	p-value	potential relevant genes
chr30.18465869, chr26.6171079	0	<i>FGF7-?</i>
chr15.44092912, chr23.49871523	0	<i>IGF1-P2RY1</i>
chr24.26359293, chr15.43667654	2e-04	<i>ASIP-?</i>
chr15.43667654, chr23.49871523	1e-04	<i>?-P2RY1</i>

First, the SNP chr30.18465869 is located close to (about 80Kb) *fibroblast growth factor 7* (*FGF7* also called *keratinocyte growth factor*, *KGF*), i.e. it belongs to the same family as the gene reported in the original study (but on a different chromosome). The FGF family members are involved in a variety of biological processes including hair development reported in human, mouse, rat and chicken (GO:0031069, [6]).

Secondly, chr15.44092912 is located between two genes, and about 200Kb from the *insulin-like growth factor 1* gene (*IGF1*). *IGF1* has been reported to be associated with the hair growth cycle and the differentiation of the hair shaft in mice [72].

Thirdly, chr23.49871523 is located about 430Kb from the *purinergic receptor P2Y1* (*P2RY1*). The purinergic receptors have been shown to be part of a signaling system for proliferation and differentiation in human anagen hair follicles [34].

Finally, chr24.26359293 is located inside the agouti-signaling protein (gene *ASIP*), a gene known to affect coat color in dogs and other mammals. The link to hair growth is not obvious but this gene is expressed during four to seven days of hair growth in mice [75].

According to our analysis, *IGF1* and *P2RY1* are significantly interacting. All other pairs of interacting SNPs involve at least one SNP for which we were not able to identify a close-by candidate gene related to hair growth (see Table 5.3.4). *IGF1* has a tyrosine kinase receptor and *P2RY1* is a G-protein coupled receptor. One possibility is that these receptors cross-talk as has been shown previously for these types of receptors in order to control mitogenic signals [24]. However, a functional assay would be necessary to establish that any of the statistical interactions we found are also biologically meaningful.

We also considered all triplets of SNPs among the 20 preselected SNPs and tested for 4-way interaction. However, we did not find any evidence for interaction among the $\binom{20}{3}$ triplets.

5.4 Discussion

In this chapter, we proposed a Markov basis approach for detecting epistasis in genome-wide association studies. The use of different Markov bases allows to easily test for different types of interaction and epistasis involving two or more SNPs. These Markov bases need to be

computed only once and can be downloaded from our website⁴ for the tests presented in this chapter.

The use of an exact test is of particular relevance for disease mapping studies where the contingency tables are often sparse. One example where there has been also functional validation, is a deletion associated with Crohn's disease [55] This deletion was found to have a population frequency of 0.07, and a frequency of 0.11 in the cases [51]. So within 400 controls and under Hardy-Weinberg equilibrium, we would expect only 2 individuals to be homozygote for this deletion. This shows that also for a moderate number of cases and controls the resulting tables for disease association studies are likely to be sparse. The sparsity is even more pronounced for rare variants, defined as variants with a MAF smaller than 0.005. Current genome wide association studies are still missing these rare variants, but advances in sequencing technologies should allow to sequence these variants and appropriate statistical methods will then be necessary.

We tested our method in simulation studies and showed that it outperforms a stepwise logistic regression approach and BEAM for the multiplicative interaction model. Logistic regression has the advantage of a very short running time (3 seconds compared to 39 minutes using our method for the analysis of one dataset with 10,000 SNPs and 400 cases and controls not including the filtering step, which takes about 1 minute for both methods on an Intel Core 2.2 GHz laptop with 2 Gb memory). However, especially for a minor allele frequency of 0.1 logistic regression performs worse than our method, even when simulating epistasis under a multiplicative model, which should favor logistic regression. This difference arises because our method approximates the exact p-value well for all sample sizes while the performance of logistic regression increases with larger sample size. 400 cases and 400 controls are not sufficient to get a good performance using logistic regression for a minor allele frequency of 0.1 and it is expected to do even worse for rare variants. Another advantage of our method compared to logistic regression is that it is not geared towards testing for multiplicative interaction only, but should be able to detect epistasis regardless of the interaction model chosen. It would be interesting to compare these two methods on data sets generated by other interaction models.

BEAM on the other hand, has the advantage of not needing to filter the large number of SNPs first. However, it runs about 15 times slower than our method for our simulations and has a very high false negative rate. The difference between our results and what the authors of BEAM have found might be due to linkage disequilibrium in our data. BEAM handles linkage disequilibrium with a first order Markov chain, which will be improved in future versions (Yu Zhang, personal communication). But as of today, we conclude that this method is impractical for whole genome association studies, since linkage disequilibrium is present in most real datasets.

The limitation of our method is the need for a filtering step to reduce the number of SNPs to a small subset. Especially if the marginal association of the interacting SNPs

⁴<http://www.carolineuhler.com/epistasis.htm>

with the disease is small, these SNPs might not be caught by the filter. However, in our simulations using Fisher's exact test as a filter seems to perform well. Another possibility is to incorporate biological information such as existing pathways ([26]) to choose a subset of possibly interacting SNPs.

We demonstrated the potential of the proposed two-stage method in genome-wide association studies by analyzing a hair length dataset consisting of 685 dogs and containing 40,842 SNPs using the extended version of Fisher's exact test. In this dataset, we found a significant epistatic effect for four SNP pairs. These SNPs lie on different chromosomes, reducing the risk of a false positive correlation due to linkage effects. The dataset includes dogs from 65 distinct breeds. Although linkage disequilibrium has been shown to extend over several megabases within breeds, linkage disequilibrium extends only over tens of kilobases between breeds and drops faster than in human populations ([45, 50, 68]), suggesting that it is possible to do fine-mapping between breeds. These observations are consistent with two bottlenecks, the first associated with the domestication from wolves and the second associated with the intense selection to create the breeds. Other studies have successfully employed the extensive variation between breeds to map genes affecting size and behavior ([16, 44]). The validity of this approach rests on the assumption that the breeds used are random samples of unrelated breeds or that related breeds make up a small part of our sample ([31, 44]). This is rarely the case and false positive results may therefore have arisen from population structure. A second independent dataset would be useful to confirm our findings. Finally, a functional assay would be necessary to establish if the interactions we found are also biologically meaningful.

Bibliography

- [1] 4ti2 team. 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de.
- [2] F. Acquistapace, F. Broglia, and M. P. Vélez. Basicness of semialgebraic sets. *Geometriae Dedicata*, 78:229–240, 1999.
- [3] J. Agler, J. W. Helton, S. McCullough, and S. Rodman. Positive semidefinite matrices with a given sparsity pattern. *Linear Algebra and its Applications*, 107:101–149, 1988.
- [4] A. Albrechtsen, S. Castella, G. Andersen, T. Hansen, O. Pedersen, and R. Nielsen. A Bayesian multilocus association method: allowing for higher-order interaction in association studies. *Genetics*, 176(2):1197–1208, 2007.
- [5] T. W. Anderson. Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In R. C. Bose, I. M. Chakravati, P. C. Mahalanobis, C. R. Rao, and K. J. C. Smith, editors, *Essays in Probability and Statistics*, pages 1–24. University of North Carolina Press, Chapel Hill, 1970.
- [6] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1):25–29, 2000.
- [7] O. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. John Wiley & Sons, Chichester, 1978.
- [8] W. Barrett, C. Johnson, and R. Loewy. The real positive definite completion problem: cycle completability. *Memoirs of the American Mathematical Society*, 584:69, 1996.
- [9] W. Barrett, C. Johnson, and P. Tarazaga. The real positive definite completion problem for a simple cycle. *Linear Algebra and its Applications*, 192:3–31, 1993.
- [10] A. I. Barvinok. *A Course in Convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2002.

- [11] C. Biró. Treewidth and grids. Technical report, School of Mathematics, Georgia Institute of Technology, 2005.
- [12] Y. Bishop, S. Fienberg, and P. Holland. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, 1975.
- [13] L. D. Brown. *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory*, volume IX of *Institute of Mathematical Statistics Lecture Notes-Monograph*. Institute of Mathematical Statistics., Hayward, CA, 1986.
- [14] T. Brylawski. A combinatorial model for series-parallel networks. *Transactions of the American Mathematical Society*, 154:1–22, 1971.
- [15] S. L. Buhl. On the existence of maximum likelihood estimators for graphical Gaussian models. *Scandinavian Journal of Statistics*, 20:263–270, 1993.
- [16] E. Cadieu, M. W. Neff, P. Quignon, K. Walsh, K. Chase, H. G. Parker, B. M. Vonholdt, A. Rhue, A. Boyko, A. Byers, A. Wong, D. S. Mosher, A. G. Elkahloun, T. C. Spady, C. Andre, K. G. Lark, M. Cargill, C. D. Bustamante, R. K. Wayne, and E. A. Ostrander. Coat variation in the domestic dog is governed by variants in three genes. *Science*, 326(5949):150–153, 2009.
- [17] M. Chardin, D. Eisenbud, and B. Ulrich. Hilbert series of residual intersections. Manuscript, 2009.
- [18] H. J. Cordell. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics*, 11:2463–2468, 2002.
- [19] D. Cox, J. Little, and D. O'Shea. *Ideals, varieties, and algorithms: An introduction to computational algebraic geometry and commutative algebra*. Springer-Verlag, New York, 1997.
- [20] J. A. De Loera, B. Sturmfels, and C. Vinzant. The central curve in linear programming. Preprint available at <http://front.math.ucdavis.edu/1012.3978>.
- [21] W. Decker, G.-M. Greuel, G. Pfister, and H. Schönemann. SINGULAR 3-1-3 – A computer algebra system for polynomial computations. <http://www.singular.uni-kl.de>.
- [22] A. P. Dempster. Covariance selection. *Biometrics*, 28:157–175, 1972.
- [23] P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, 26:363–397, 1998.
- [24] I. Dikic and A. Blaukat. Protein tyrosine kinase-mediated pathways in G protein-coupled receptor signaling. *Cell Biochemistry and Biophysics*, 30(3):369–387, 1999.

- [25] M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on Algebraic Statistics*, volume 40 of *Oberwolfach Seminars*. Birkhäuser, Basel, 2009.
- [26] M. Emily, T. Mailund, J. Hein, L. Schauer, and M. H. Schierup. Using biological networks to search for interacting loci in genome-wide association studies. *European Journal of Human Genetics*, 17(10):1231–1240, 2009.
- [27] N. Eriksson, S. Fienberg, A. Rinaldo, and S. Sullivant. Polyhedral conditions for the nonexistence of the mle for hierarchical log-linear models. *Journal of Symbolic Computation*, 41:222–233, 2006.
- [28] S. E. Fienberg. An iterative procedure for estimation in contingency tables. *Annals of Mathematical Statistics*, 41:907–917, 1970.
- [29] L. Garcia, M. Stillman, and B. Sturmfels. Algebraic geometry of bayesian networks. *Journal of Symbolic Computation*, 39:331–355, 2005.
- [30] W. R. Gilks, S. Richardson, and D. J. (eds.) Spiegelhalter. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall, 1995.
- [31] M. E. Goddard and B. J. Hayes. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6):381–91, 2009.
- [32] M. Grant and S. Boyd. Cvx, a matlab software for disciplined convex programming. Available at <http://stanford.edu/~boyd/cvx>.
- [33] D. R. Grayson and M. E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at <http://www.math.uiuc.edu/Macaulay2/>.
- [34] A. V. Greig, C. Linge, and G. Burnstock. Purinergic receptors are part of a signalling system for proliferation and differentiation in distinct cell lineages in human anagen hair follicles. *Purinergic Signalling*, 4(4):331–338, 2008.
- [35] B. Grone, C. R. Johnson, E. M. Sa, and H. Wolkowicz. Positive definite completions of partial Hermitian matrices. *Linear Algebra and its Applications*, 58:109–124, 1984.
- [36] I. B. Hallgrimsdottir and D. S. Yuster. A complete classification of epistatic two-locus models. *BMC Genetics*, 9:17, 2008.
- [37] R. Hartshorne. *Algebraic geometry*. Springer-Verlag, New York, 1977. Graduate Texts in Mathematics, No. 52.
- [38] J. Herzog, W. Vasconcelos, and R. Villareal. Ideals with sliding depth. *Nagoya Mathematical Journal*, 99:159–172, 1985.

- [39] S. Højsgaard and S. L. Lauritzen. Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society*, 70:1005–1027, 2008.
- [40] S. Holzer and O. Labs. SURFEX 0.90. Technical report, University of Mainz, University of Saarbrücken, 2008. www.surfex.AlgebraicSurface.net.
- [41] S. T. Jensen. Covariance hypothesis which are linear in both the covariance and the inverse covariance. *Annals of Statistics*, 16:302–322, 1988.
- [42] C. R. Johnson. Matrix completion problems: a survey. In *Matrix Theory and Applications*, volume 40 of *Proceedings of Symposia in Applied Mathematics*, pages 171–198. American Mathematical Society, Providence, Rhode Island, 1990.
- [43] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Multivariate Distributions*, volume 1 (second edition). Wiley & Sons, New York, 2000.
- [44] P. Jones, K. Chase, A. Martin, P. Davern, E. A. Ostrander, and K. G. Lark. Single-nucleotide-polymorphism-based association mapping of dog stereotypes. *Genetics*, 179(2):1033–44, 2008.
- [45] E. K. Karlsson, I. Baranowska, C. M. Wade, N. H. Salmon Hillbertz, M. C. Zody, N. Anderson, T. M. Biagi, N. Patterson, G. R. Pielberg, E. J. Kulbokas, 3rd, K. E. Comstock, E. T. Keller, J. P. Mesirov, H. von Euler, O. Kampe, A. Hedhammar, E. S. Lander, G. Andersson, L. Andersson, and K. Lindblad-Toh. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature Genetics*, 39(11):1321–8, 2007.
- [46] B. Kotzev. Determinantal ideals of linear type of a generic symmetric matrix. *Journal of Algebra*, 139:484–504, 1991.
- [47] M. Laurent. Matrix completion problems. In *The Encyclopedia of Optimization*, pages 221–229. Kluwer, 2001.
- [48] M. Laurent. On the sparsity order of a graph and its deficiency in chordality. *Combinatorica*, 21:543–570, 2001.
- [49] S. L. Lauritzen. *Graphical models*. Clarendon Press, Oxford, 1996.
- [50] K. Lindblad-Toh, C. M. Wade, T. S. Mikkelsen, and E. K. Karlsson. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438(7069):803–19, 2005.
- [51] T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines,

- T. F. Mackay, S. A. McCarroll, and P. M. Visscher. Finding the missing heritability of complex diseases. *Nature*, 461:747–753, 2009.
- [52] J. Marchini, P. Donnelly, and L. R. Cardon. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature Genetics*, 37:413–417, 2005.
- [53] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.
- [54] K. A. Mather, A. L. Caicedo, N. R. Polato, K. M. Olsen, S. McCouch, and M. D. Purugganan. The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics*, 177(4):2223–2232, 2007.
- [55] S. A. McCarroll, A. Huett, P. Kuballa, S. D. Cholewicki, A. Landry, P. Goyette, M. C. Zody, J. L. Hall, S. R. Brant, J. H. Cho, R. H. Duerr, M. S. Silverberg, K. D. Taylor, J. D. Rioux, D. Altshuler, M. J. Daly, and R. J. Xavier. Deletion polymorphism upstream of *irgm* associated with altered *irgm* expression and crohn’s disease. *Nature Genetics*, 40:1107–1112, 2008.
- [56] M. I. McCarthy, G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. A. Ioannidis, and J. N. Hirschhorn. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, 9:356–369, 2008.
- [57] E. Miller and B. Sturmfels. *Combinatorial Commutative Algebra*. Graduate Texts in Math. Springer, New York, 2004.
- [58] J. Nie, K. Ranestad, and B. Sturmfels. The algebraic degree of semidefinite programming. *Mathematical Programming*, 122:379–405, 2010.
- [59] L. Pachter and B. Sturmfels. *Algebraic Statistics for Computational Biology*. Cambridge University Press, 2005.
- [60] J. K. Pritchard. Are rare variants responsible for susceptibility to complex diseases? *American Journal of Human Genetics*, 69:124–137, 2001.
- [61] N. Proudfoot and D. Speyer. A broken circuit ring. *Beiträge zur Algebra und Geometrie*, 47:161–166, 2006.
- [62] J. Schäfer and K. Strimmer. Learning large-scale graphical Gaussian models from genomic data. In *In Science of Complex Networks: From Biology to the Internet and WWW*, 2005.
- [63] J. M. Smith and J. Haigh. The hitchhiking effect of a favourable gene. *Genetical Research*, 23:23–35, 1974.

- [64] J. Stückrad. On quasi-complete intersections. *Archiv der Mathematik*, 58:529–538, 1992.
- [65] B. Sturmfels. *Gröbner Bases and Convex Polytopes*, volume 8 of *University Lecture Series*. American Mathematical Society, Providence, 1996.
- [66] S. Sullivant. Algebraic geometry of gaussian bayesian networks. *Advances in Applied Mathematics*, 40:482–513, 2008.
- [67] S. Sullivant and K. Talaska. Trek separation for gaussian graphical models. *Annals of Statistics*, 38:1665–1685, 2010.
- [68] N. B. Sutter, M. A. Eberle, H. G. Parker, B. J. Pullar, E. F. Kirkness, L. Kruglyak, and E. A. Ostrander. Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Research*, 14(12):2388–96, 2004.
- [69] H. Terao. Algebras generated by reciprocals of linear forms. *Journal of Algebra*, 250:549–558, 2002.
- [70] L. Vandenberghe, S. Boyd, and S-P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19:499–533, 1996.
- [71] R. K. Wayne and E. A. Ostrander. Lessons learned from the dog genome. *Trends Genet*, 23(11):557–67, 2007.
- [72] N. Weger and T. Schlake. Igf-I signalling controls the hair growth cycle and the differentiation of hair shafts. *Journal of Investigative Dermatology*, 125(5):873–882, 2005.
- [73] N. White, editor. *Matroid Applications*. Cambridge University Press, 1992.
- [74] J. Whittaker. *Graphical models in applied multivariate statistics*. John Wiley & Sons, Chichester, 1990.
- [75] G. L. Wolff, J. S. Stanley, M. E. Ferguson, P. M. Simpson, M. J. Ronis, and T. M. Badger. Agouti signaling protein stimulates cell division in "viable yellow" (A^{vy}/a) mouse liver. *Experimental Biology and Medicine (Maywood)*, 232(10):1326–1329, 2007.
- [76] F. A. Wright, H. Huang, X. Guan, K. Gamiel, C. Jeffries, W. T. Barry, F. Pardo-Manuel de Villena, P. F. Sullivan, K. C. Wilhelmsen, and F. Zou. Simulating association studies: a data-based resampling method for candidate regions or whole genome scans. *Bioinformatics*, 23:2581–2588, 2007.
- [77] X. Wu, Y. Ye, and K. R. Subramanian. Interactive analysis of gene interactions using graphical Gaussian model. *ACM SIGKDD Workshop on Data Mining in Bioinformatics*, 3:63–69, 2003.

- [78] T. Zaslavsky. *Facing up to arrangements: Face-count formulas for partitions of space by hyperplanes*, volume 154 of *Memoirs of the American Mathematical Society*. American Mathematical Society, Providence, R.I., 1975.
- [79] Y. Zhang and J. S. Liu. Bayesian inference of epistatic interactions in case-control studies. *Nature Genetics*, 39:1167–1173, 2007.
- [80] G. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer, New York, 1995.