Geometry of conditional independence

by

Jason Ryder Morton

A.B. (Harvard University) 1998
M.A. (University of Michigan, Ann Arbor) 2003

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

in

Mathematics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Bernd Sturmfels, Chair
Professor Lior Pachter
Professor Xin Guo

Fall 2007

The dissertation of Jason Ryder Morton is approved:

Chair Date

Date

Date

University of California, Berkeley

Fall 2007

Geometry of conditional independence

Abstract

Geometry of conditional independence

by

Jason Ryder Morton

Doctor of Philosophy in Mathematics

University of California, Berkeley

Professor Bernd Sturmfels, Chair

This thesis investigates geometric aspects of the notions of conditional independence and conditional probability. In Chapter 2, we develop the connection between conditional independence models and polyhedral fans. The main result uses algebraic techniques and the *permutohedron*, a polytope that plays an important role in the geometry of conditional independence. The results are applied to define a class of rank tests useful for exploratory data analysis. In Chapter 3, we further develop this class of rank tests, called topographical models, for use in analyzing microarray data. We find the necessary algorithms and counting theorems required to make this test practical. We apply these methods to two data sets. In Chapter 4, we use the machinery of Chapter 2 to settle three open theoretical questions about conditional independence models. In doing so we explore a more algebraic perspective on semigraphoids. In Chapter 5, we answer a question raised by the work of Besag on the relations among conditional probabilities. This is accomplished via toric geometry. We also show that the moment map relates the space of conditional probability

distributions to generalized permutohedra.

Professor Bernd Sturmfels
Dissertation Committee Chair

To,

Sarah.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my gratitude to my advisor, Bernd Sturmfels, for the great deal he has taught me, and his insights and encouragement along the way. He has generously provided a great deal of personal attention and thoughtful mentorship that takes a long-term view of my career. I would like to thank my co-authors Raymond Hemmecke, Lior Pachter, Anne Shiu, Oliver Wienand, and Bernd Sturmfels for interesting discussions, productive collaborations, and permission to include some co-authored material. Milan Studený, Ferro Matúš, and Seth Sullivant have also provided helpful comments. I would like to thank my committee members, Bernd, Lior, Xin Guo, and (honorary member) Steve Evans, for their generous provision of teaching, time, and patience in helping me on the climb toward graduation. I appreciate the editors of *Proceedings of the Third European Workshop on Probabalistic Graphical Models*, *Statistical Applications in Genetics and Molecular Biology* and *Combinatorics, Probability and Computing* providing permission to include published and in-press material. I would like to thank my department for supporting me financially in the early years of graduate school and Bernd and DARPA for supporting me during the research years. I would like to thank my parents Mel and Ginny for their support and encouragement throughout my educational career. Finally, I thank my wife Sarah for her much needed love, encouragement, and support.

Chapter 1

Introduction

In this thesis we study the geometry of conditional independence and conditional probability. In this Chapter we introduce some of the concepts required for our study. In Section 1.1, we cover polyhedral geometry; in Section 1.2, toric geometry; and in Section 1.3, we introduce some concepts from algebraic statistics.

1.1 Polyhedral geometry

In this section, we define essential concepts in polyhedral geometry, based on the exposition in [69] and [73]. A point set $K \subset \mathbb{R}^d$ is *convex* if given $u, v \in K$, K contains the segment $\{\lambda u + (1-\lambda)v : \lambda \in [0,1]\}$. For a set $S \subset \mathbb{R}^d$, the *convex hull* $\mathrm{conv}(S)$ is the intersection of all convex sets containing S:

$$\mathrm{conv}(S) = \bigcap \{K \subset \mathbb{R}^d : S \subset K,\ K \text{ convex}\}.$$

A polyhedron is a finite intersection of closed halfspaces in $\mathbb{R}^d$, and accordingly may be written as $P = \{x \in \mathbb{R}^d : x^\top A \le b^\top\}$ for a $d \times n$ matrix A. When $b = 0$, there are points

$a_1, \ldots, a_m$ in $\mathbb{R}^d$ such that $P = \text{pos}(a_1, \ldots, a_m) := \{\lambda_1 a_1 + \cdots + \lambda_m a_m : \lambda_j \in \mathbb{R}_{\geq 0}\}$. This is a polyhedron by the Weyl-Minkowski Theorem [73]. Such a polyhedron is a *polyhedral cone* . A cone is *pointed* if it contains no line. The *lineality space* of a polyhedron P is the inclusion-maximal linear space contained in P. A bounded polyhedron (i.e. one containing no ray) is a *polytope* . Every polytope may be written as the convex hull of a finite set of points in $\mathbb{R}^d$, and in particular is the convex hull of its vertices.

Let $P \subset \mathbb{R}^d$ be a polyhedron and $w \in (\mathbb{R}^d)^*$ a linear functional. Define $\text{face}_w(P) := \{u \in P : w \cdot u \geq w \cdot v \text{ for all } v \in P\}$. A *face* of P is a subset of this form. The dimension of a face is the dimension of its affine span. Vertices are 0-dimensional faces, edges 1-dimensional. Faces other than the polytope itself are *proper* faces, and *facets* are maximal proper faces. The *face lattice* of a polytope is the poset of all faces of P, partially ordered by inclusion. A d-dimensional polytope is *simplicial* if all its proper faces are simplices, i.e. if every facet has d vertices. A polytope is *simple* if every vertex is contained in only d facets.

Definition 1.1.1. A *fan* in $\mathbb{R}^n$ is a finite collection $\mathcal{F}$ of polyhedral cones which satisfies the following properties:

(i) if $C \in \mathcal{F}$ and C' is a face of C, then $C' \in \mathcal{F}$,

(ii) if $C, C' \in \mathcal{F}$, then $C \cap C'$ is a face of C.

Definition 1.1.2. [32] Let P be a d-polytope. Its *normal fan* is the collection of, for each nonempty face $F \subset P$, the set C_F of all vectors $w \in \mathbb{R}^d$ such that the linear function $v \mapsto \langle w, v \rangle$ is maximized by all points in F.

There is an inclusion-reversing bijection between the faces of a polytope and the faces of its normal fan. A fan can be described by the sets of vectors that generate its cones. In Chapter 2 we will define another notion of normal fans, the *vector normal fan* that retains more information about these generating vectors. A fan is *simplicial* if all its cones are simplex cones and *complete* if its cones cover the ambient $\mathbb{R}^d$.

Finally, we need the notion of *Minkowski sum.* Given two polytopes $P, Q \subset \mathbb{R}^d$, their *Minkowski sum* is

$$P + Q = \{x + y : x \in P, y \in Q\}.$$

P is a *Minkowski summand* of Q if there exists a third polytope $P' \subset \mathbb{R}^d$ and $\lambda > 0$ such that $\lambda Q = P + P'$. By [73], Proposition 7.12, the normal fan of a Minkowski sum is the common refinement of the individual fans, where the *common refinement* of two fans $\mathcal{F}, \mathcal{G}$ is the fan $\{C \cap C' : C \in \mathcal{F}, \mathcal{G}\}$. Every polyhedron P can be written as the Minkowski sum $P = Q + C$ of a polytope Q and a polyhedral cone C. The cone C is unique and is called the recession cone of P [69, 60].

Let $P = \text{conv}(\mathcal{A})$ where the columns of $\mathcal{A}$ lie in a common hyperplane in $\mathbb{R}^d$. As in [69], define the *Ehrhart polynomial* $E_P : \mathbb{N} \to \mathbb{N}, r \mapsto |\mathbb{Z}\mathcal{A} \cap rP|$. Here $\mathbb{Z}\mathcal{A}$ denotes the abelian group generated by $\mathcal{A}$. It can be shown that E_P is a polynomial of degree $\dim P \leq d-1$. The leading coefficient of this polynomial is denoted $\text{Vol}(P)$, the *normalized volume* of P. When $\dim(P) = d - 1$, and $vol(P)$ is the Euclidean volume of P, $\text{Vol}(P) = vol(P) \cdot \dim(P)! \cdot |\mathbb{Z}^d/\mathbb{Z}\mathcal{A}|$.

1.2 Toric ideals and toric varieties

Here we collect several useful facts about toric ideals and toric varieties based primarily on Sturmfels' book [69], also referring to [18, 24, 26, 48, 52].

1.2.1 Affine toric varieties

Let $\mathcal{A}$ be a $d \times m$ integer matrix, with columns $a_{\cdot,1}, \ldots, a_{\cdot m}$. Let $\mathbb{C}[x_1, \ldots, x_m]$ be a polynomial ring in m variables, and for $u \in \mathbb{Z}^m$ let $x^u = \prod_{j=1}^m x_j^{u_j}$. The matrix $\mathcal{A}$ defines a *toric ideal*

$$I_{\mathcal{A}} = \langle x^{u^+} - x^{u^-} : u \in \ker \mathcal{A} \cap \mathbb{Z}^m \rangle,$$

where u^+ is the positive part of u and u^- the negative. The toric ideal $I_{\mathcal{A}}$ is a prime ideal. A minimal set of binomials which generates $I_{\mathcal{A}}$ is said to be a *Markov basis* for the matrix $\mathcal{A}$. A *Gröbner basis* $\{f_1, \ldots, f_k\}$ for an ideal I with respect to a monomial term order $\succ$ has $\mathrm{in}_{\succ}(I) = \langle \mathrm{in}_{\succ}(f_1), \ldots \mathrm{in}_{\succ}(f_k) \rangle$. A Gröbner basis is *universal* if it is a Gröbner basis for all term orders $\succ$.

In the affine space $\mathbb{C}^m$ with coordinates $x_1, \ldots, x_m$, the ideal $I_{\mathcal{A}}$ cuts out the affine toric variety $X_{\mathcal{A}}$. The $\mathbb{R}_{\geq 0}$-span of the columns of $\mathcal{A}$ define a cone $\mathrm{pos}(\mathcal{A})$, and the $\mathbb{N}$-span defines a semigroup $\mathbb{N}\mathcal{A}$. The corresponding semigroup ring $\mathbb{C}[\mathbb{N}\mathcal{A}]$ is isomorphic to the affine coordinate ring $\mathbb{C}[x_1, \ldots, x_m]/I_{\mathcal{A}}$, i.e.$X_{\mathcal{A}} \cong \mathrm{Spec}(\mathbb{C}[x_1, \ldots x_m]/I_{\mathcal{A}}) \cong \mathrm{Spec}\, \mathbb{C}[\mathbb{N}\mathcal{A}]$. Such varieties are not always normal.

Proposition 1.2.1. *For a $d \times m$ integer matrix $\mathcal{A}$, the following are equivalent*

(i) *the affine coordinate ring $\mathbb{C}[\mathbb{N}\mathcal{A}] \cong [x_1, \ldots, x_m]/I_{\mathcal{A}}$ is normal,*

(ii) *the affine toric variety $X_{\mathcal{A}}$ is normal,*

(iii) *the semigroup $\mathbb{N}\mathcal{A}$ is normal,*

(iv) $\mathbb{N}\mathcal{A} = \mathrm{pos}(\mathcal{A}) \cap \mathbb{Z}^d$.

Proposition 1.2.1 also tells us one way to construct the normalization of a toric variety, namely by normalizing its semigroup [13]. Let σ be the cone $\{\eta \in \mathbb{R}^d : \eta^\top \mathcal{A} \geq 0\}$; then $\sigma^\vee = \text{pos}(\mathcal{A})$ and the normalization of $X_\mathcal{A}$ is X_σ as defined in [26].

The matrix $\mathcal{A}$ defines a map $\mathbf{f}_\mathcal{A} : \theta \mapsto \theta^\mathcal{A}$ from the d-dimensional torus $\mathbb{T}_d$ to the toric variety $X_\mathcal{A}$. This gives an explicit torus action and torus embedding. The closure of the image of $\mathbf{f}$ is $X_\mathcal{A}$.

1.2.2 Polytopes and projective toric varieties

Let conv($\mathcal{A}$) be the convex hull of the columns of $\mathcal{A}$. This is a polytope. Let $Y_\mathcal{A}$ be the projective toric variety defined by taking the closure of the image of $\mathbf{f}_\mathcal{A}$, and viewing $x_1, \ldots, x_m$ as homogeneous coordinates. The corresponding homogeneous toric ideal is the ideal

$$J_\mathcal{A} = \langle x^{u^+} - x^{u^-} : u \in \ker \mathcal{A} \cap \mathbb{Z}^m,\ \|u^+\|_1 = \|u^-\|_1 \rangle. \tag{1.1}$$

The affine cone over $Y_\mathcal{A}$ is the toric variety $X_{\mathcal{A}'}$, where $\mathcal{A}'$ is $\mathcal{A}$ with a row of ones added at the bottom unless the vector of all ones already lies in rowspan($\mathcal{A}$). This induces homogeneity with respect to the $\mathbb{Z}$-grading. When $\mathcal{A}$ has $(1, 1, \ldots, 1)$ in its row span (e.g. by having equal column sums or $(1, 1, \ldots, 1)$ as a row), we say it is $\mathbb{Z}$-graded and the norm restriction in (1.1) is not required. Instead of $(1, 1, \ldots, 1)$, we can use another grading of the columns of $\mathcal{A}$ to obtain multihomogeneous ideals. We will make use of such a grading in Chapter 5.

The following definition and lemma are given in Geiger, Meek, and Sturmfels [29] and will be used in Chapter 4. A subset F of $[m] = \{1, \ldots, m\}$ is *facial* if there is a vector $c \in \mathbb{R}^d$ such that $c^\top a_j = 0$ for $i \in F$ and $c^\top a_i \geq 1$ for $i \in [m] \setminus F$. Thus the vector c is orthogonal to the columns whose index is in F and not orthogonal to the other columns of $\mathcal{A}$. The *characteristic vector* of F is $(z_1, \ldots, z_m)$ with $z_i = 1$ if $i \in F$ and $z_i = 0$ if $i \notin F$.

Proposition 1.2.2 ([29]). *For a subset F of $[m]$ and a matrix $\mathcal{A}$, the following are equivalent:*

1. *F is facial for $\mathcal{A}$,*

2. *The characteristic vector of F lies in the nonnegative toric variety $X_{\mathcal{A}}$,*

3. *There exists a vector with support F in the nonnegative toric variety $X_{\mathcal{A}}$.*

1.2.3 The moment map

The moment map sends a projective toric variety $Y_{\mathcal{A}}$ onto its polytope $\operatorname{conv}(\mathcal{A})$, bijectively on the nonnegative part of the variety. In Chapter 5, we will use a version of this theorem for toric varieties in a product of projective spaces.

Theorem 1.2.3. *Let $\mathcal{A}$ be a $d \times m$, $\mathbb{Z}$-graded matrix, and $Y_{\mathcal{A}}$ the corresponding projective toric variety. Then the map*

$$\mu : Y_{\mathcal{A}} \to \operatorname{conv}(\mathcal{A}), \quad \textit{given by}$$

$$\mu(y) = \frac{1}{Z(y)} \sum_j |y_j| a_{\cdot j},$$

where $Z(y) = \sum_j |y_j|$, is a bijection from $Y_{\mathcal{A},\geq 0}$ onto $\operatorname{conv}(\mathcal{A})$. If further $\operatorname{rank}(\mathcal{A}) = d$, with $\mathbf{f}_{\mathcal{A}}$ the torus embedding, then $\mu \circ \mathbf{f}_{\mathcal{A}}$ is homeomorphism $\mathbb{R}^d_{>0} \to \operatorname{conv}(\mathcal{A})^\circ$.

Proof. The map μ is the composition of two maps: $\mu_1 : Y_{\mathcal{A}} \to \Delta_{m-1}$ sending $(y_1 : \cdots : y_m)$ to $\mathbf{p} = (\frac{1}{Z(y)}|y_1|, \ldots, \frac{1}{Z(y)}|y_m|)$, and $\mu_2 : \Delta_{m-1} \to \operatorname{conv}(\mathcal{A})$ sending $\mathbf{p}$ to $\mathcal{A}\mathbf{p}$. For a representative of a nonnegative point in $Y_{\mathcal{A}}$ which is scaled so its coordinates sum to one, μ_1 is the identity; μ_2 is also an identity in the special case where $\mathcal{A} = I_d$, the identity matrix. In general, given a point $b \in \operatorname{conv}(\mathcal{A})$, there is ambiguity in that there is a polytope

$P_{\mathcal{A}}(b) = \{\mathbf{p} : \mathcal{A}\mathbf{p} = b\} \cap \Delta_{m-1}$ of barycentric coordinates $\mathbf{p} \in \Delta_{m-1}$ mapping to b under μ_2, and correspondingly there could be a space of toroidal fibers in $Y_{\mathcal{A}}$. As it turns out, only one point in $Y_{\mathcal{A},\geq 0}$ maps to b under μ. This point is the maximum entropy point in the space $P_{\mathcal{A}}(b)$ of feasible points of Δ_{m-1}, or equivalently the point which minimizes the KL-divergence to the uniform distribution $p_j = 1/m$ for all j.

Let $D(p) = D(p||p^{uniform}) = \sum_j p_j \log p_j - \log \frac{1}{m} \sum_j p_j$. Then the Hessian is $\frac{1}{p_j}$ on the diagonal and zero otherwise. This Hessian is positive definite on the interior Δ°_{m-1}, and on points of the relative interior of the faces of Δ_{m-1} after restricting to nonzero coordinates. Thus D has a unique minimum on Δ_{m-1}: were there another, the (possibly restricted) Hessian would be positive definite on the open segment connecting the two points, contradicting their minimality.

We argue that this minimum $\mathbf{p}^*$ lies in $Y_{\mathcal{A}}$. First suppose $\mathbf{p}^*$ is a point in the interior and $u \in \ker \mathcal{A}$. Then since $\mathbf{p}^*$ lies in the interior, for small t so does $p^* + tu$ and

$$D(p^* + tu) = \sum_j (p_j + tu_j) \log(p_j + tu_j) - \log(\frac{1}{m}) \sum_j (p_j + tu_j).$$

Then $\frac{dD}{dt} = \sum_j u_j \log(p_j + tu_j)$ where the other terms vanish because the $\mathbb{Z}$-grading of $\mathcal{A}$ means that $(1, 1, \ldots, 1)$ lies in its rowspan and $\sum_j u_j = 0$. Setting $\frac{dD}{dt} = 0$ at $t = 0$ to satisfy the first order condition, we have $\sum_j u_j p_j = 0$. Grouping by the sign of the u_j and exponentiating, we have $\mathbf{p}^{u^+} = \mathbf{p}^{u^-}$ so $\mathbf{p} \in Y_{\mathcal{A}}$.

Now suppose $\mathbf{p}^*$ lies on the boundary of Δ_{m-1}, where one or more coordinates p_j are zero, and again let $u \in \ker(\mathcal{A})$. If the zeros of $\mathbf{p}$ are outside $\text{supp}(u)$, the interior argument holds after extending D with the limit $p \log(p) \to 0$ as $p \to 0$. If a p_j is zero in both $\text{supp}(u^+)$ and $\text{supp}(u^-)$, the relation holds as $0 = 0$. If $p_j = 1$ for some $j \in \text{supp}(u)$ and the rest are 0, we will again obtain $0 = 0$ for any nontrivial relation. Suppose then that we have the remaining case: $p_i = 0$ for $i \in \text{supp}(u^+)$ but $0 < p_j < 1$ for all $j \in \text{supp}(u^-)$. Then since $u_i > 0$ for $i \in \text{supp}(u^+)$, for small t we have $p^* + tu \in \Delta_{m-1}$. Now

$$\frac{dD}{dt} = \sum_{i:p_i=0} u_i(p_i + tu_i) + \sum_{j:p_j\neq 0} u_j(p_j + tu_j)$$

which goes to $-\infty$ as $t \to 0$, contradicting minimality of $\mathbf{p}^*$, so this case cannot arise. □

1.3 Algebraic statistics and conditional independence

1.3.1 A projective view of probability

Consider a probability space with m disjoint atomic events $([m], 2^{[m]}, P)$. The space of probability distributions P on them is typically represented as a *probability simplex* as in Figure 1.1.

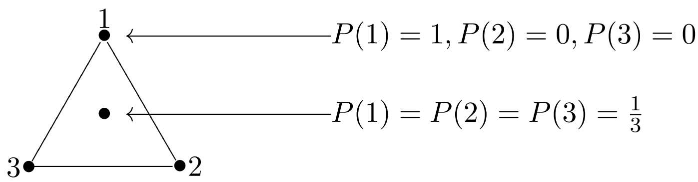


Figure 1.1: Probability simplex

Where each $P(i)$ is a coordinate p_i such that $p_i \geq 0$ and $\sum_i p_i = 1$. We will be describing families of probability distributions in terms of *algebraic varieties*, and we prefer to think of points $(p_1 : \cdots : p_m)$ as lying in complex projective space because it is easier to work with. There are two ways to match up the notion of the probability simplex with that of complex projective space. One way to do so is to identify the probability simplex Δ_{m-1} with the real, positive part of the affine open $\sum_i y_i \neq 0$ of the $\mathbb{P}^{m-1}$ with homogeneous coordinates $(y_1 : y_2 : \cdots : y_m)$ as illustrated in Figure 1.2.

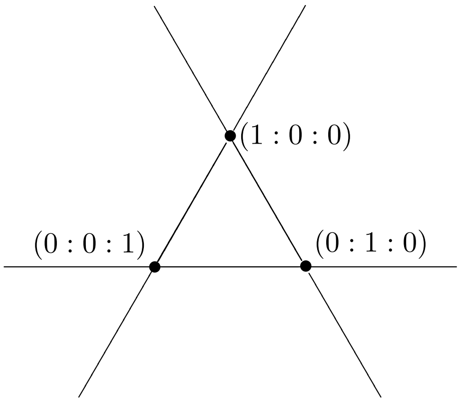


Figure 1.2: Probability simplex in the projective plane

Another representation, which is equivalent in the special case that $(y_1 : \cdots : y_m) \in \Delta_{m-1}$, is to use the moment map (Theorem 1.2.3). The identity matrix $\mathcal{A} = I_m$ comprised of standard unit vectors e_i defines the probability simplex $\Delta_{m-1} = \text{conv}(\mathcal{A})$. The toric variety $Y_{\mathcal{A}}$ is then the projective space $\mathbb{P}^{m-1}$ and the moment map is:

$$\mu : \mathbb{P}^{m-1} \to \Delta_{m-1}$$

$$\mu((y_1 : \cdots : y_m)) : \frac{1}{\sum_i |y_i|} |y_i| e_i$$

The moment map μ is the identity map on the probability simplex, but allows us to define a point on the probability simplex for more general points in complex projective space. The fiber over any of these points is the torus $(\mathbb{S}^1)^n$, a product of m unit circles, since $\mu(y_1 : \ldots, : y_m) = \mu(e^{i\theta_1} y_1 : \cdots : e^{i\theta_m} y_m)$. This point of view is also used in quantum physics and elsewhere with the modified moment map $\mu'(y) : \frac{1}{\sum_i |y_i|^2} |y_i|^2 e_i$.

1.3.2 Algebraic statistical models

A (discrete) random variable X is a set of states $\{x^1, \ldots, x^m\}$ together with a function $P : \{x^1, \ldots, x^m\} \to [0, 1]$ such that $\sum_i P(x^i) = 1$. The number $P(x^i)$ is the probability of X taking on the value x^i. This function P is called a *probability distribution*. A *statistical model* is family (a set) of probability distributions, i.e. a set of points in the simplex $\Delta_{m-1} \subset \mathbb{R}^m$. In *algebraic statistics* [52], we consider statistical models that are defined by the vanishing of systems of polynomial equations and possibly inequalities in the indeterminates $p_1, \ldots, p_m$, together with the constraints $p_i \geq 0 \; \forall i$, and $\sum_i p_i = 1$; these are semi-algebraic sets.

Typically, algebraic statistical models are defined parametrically as the image of some parameter space $\Theta \subset \mathbb{R}^d$ under a polynomial function $\mathbf{f} = (f_1, \ldots, f_m)$. Then $P(x^j) = \frac{1}{Z(\theta)} f_j(\theta), \theta \in \Theta$ where $Z(\theta) = \sum_j f_j(\theta)$ is a normalizing factor. Algebraic statistical models are extremely common in applications, and include log-linear exponential families, hierarchical models, and graphical models such as Bayesian networks, hidden Markov models, and Markov random fields [52].

Theorem 1.2.3 is also called Birch's theorem; in Pachter and Sturmfels [52], Birch's theorem has the following form. Given the matrix $\mathcal{A} \in \mathbb{N}^{d \times m}$ and any vector $b \in \mathbb{R}^d$, we consider the relatively open polytope

$$\tilde{P}_{\mathcal{A}(b)} \quad = \quad \left\{ p \in \mathbb{R}^m \; : \; \mathcal{A} \cdot p = \frac{1}{N} \cdot b \; \text{ and } \; p_j > 0 \; \text{ for all } j \; \right\}.$$

Theorem 1.3.1 (Birch's Theorem)**.** *Fix a toric model $\mathcal{A}$ and let $u \in \mathbb{N}^m_{>0}$ be a strictly positive data vector with sufficient statistic $b = \mathcal{A}u$. The intersection of the polytope $\tilde{P}_{\mathcal{A}}(b)$ with the toric model $\mathbf{f}(\mathbb{R}^d_{>0})$ consists of precisely one point. That point is the maximum likelihood estimate $\widehat{p}$ for the data u.*

1.3.3 Conditional independence

Conditional independence statements [19] describe the dependence relationship among random variables. A *conditional independence (CI) statement* on a finite set of random variables, indexed by $[n] = \{1, 2, \ldots, n\}$, is a formal symbol $I \perp\!\!\!\perp J \mid K$ where I, J, K are disjoint subsets of $[n]$. The symbol $I \perp\!\!\!\perp J \mid K$ represents the statement that the joint random variables I and J are conditionally independent given the joint random variable K. For $K = \emptyset$, we abbreviate $I \perp\!\!\!\perp J|K$ by $I \perp\!\!\!\perp J$. A CI statement is called *elementary* if it is of the form $i \perp\!\!\!\perp j \mid K$ where i and j are singletons. A *conditional independence model* is a set $\mathcal{M}$ of CI statements, typically subject to certain consistency axioms or other inference rules.

For any joint probability distribution on n random variables, the set $\mathcal{M}$ of all CI statements that are valid for the given distribution is a *semigraphoid*. However, it was shown by Studený [65] that semigraphoids do not provide a complete set of axioms for probabilistic conditional independence, and in fact no finite set of axioms could do so.

Definition 1.3.2. A *semigraphoid* [53] is a set $\mathcal{M}$ of conditional independence statements satisfying, for X, Y, Z pairwise disjoint subsets of $[n]$,

(SG1) $$X \perp\!\!\!\perp Y \mid Z \in \mathcal{M} \implies Y \perp\!\!\!\perp X \mid Z \in \mathcal{M}$$

(SG2) $$X \perp\!\!\!\perp Y \mid Z \in \mathcal{M} \text{ and } U \subset X \implies U \perp\!\!\!\perp Y \mid Z \in \mathcal{M}$$

(SG3) $$X \perp\!\!\!\perp Y \mid Z \in \mathcal{M} \text{ and } U \subset X \implies X \perp\!\!\!\perp Y \mid (U \cup Z) \in \mathcal{M}$$

(SG4) $$X \perp\!\!\!\perp Y \mid Z \in \mathcal{M} \text{ and } X \perp\!\!\!\perp W \mid (Y \cup Z) \implies X \perp\!\!\!\perp (W \cup Y) \mid Z \in \mathcal{M}.$$

Taken together, we refer to these conditions as the semigraphoid axiom **(SG)**. A semigraphoid $\mathcal{N}$ is a *coarsening* of a semigraphoid $\mathcal{M}$ if every CI statement in $\mathcal{M}$ is also in

$\mathcal{N}$.

In later chapters, we will develop three more versions of **(SG)**: an elementary CI statement version ($\mathbf{SG}'$), and versions in terms of additive ($\mathbf{SG}''$) and multiplicative ($\mathbf{SG}'''$) relations. These will also have geometric interpretations. Studený's book [67] gives an introduction to semigraphoids and their role in statistical learning theory. For further details and references see also Matúš [44, 46].

A CI statement indicates that the combined random variable X_I is independent of the combined random variable X_J given knowledge of the state of the combined random variable X_K. In terms of probability this means

$$P(x_{I\cup J}^{i_{I\cup J}}|x_K^{i_K}) = P(x_I^{i_I}|x_K^{i_K})P(x_I^{i_I}|x_K^{i_K}).$$

for the details, including the continuous case, see the book by Lauritzen [39].

1.3.4 Graphical models

Hierarchical models on n random variables describe dependence relationships topologically by a simplicial complex Δ on $[n] = \{1, \ldots, n\}$. A hierarchical model for discrete random variables $X_1, X_2, \ldots, X_n$ where X_i has the finite state space $\{1, \ldots, d_i\}$ is a model in the exponential family which takes the form

$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{1}{Z} \prod_{C\in\Delta} \phi_C(x_C). \tag{1.2}$$

For each $C \in \Delta$, ϕ_C is a potential function depending only on the random variables indexed by C; Z is a normalizing constant. These are *algebraic statistical models* [52]: they are given by polynomial parameterizations. *Undirected graphical models* are defined by a graph G on n vertices, with Δ the set of cliques of G.

The Hammersley-Clifford Theorem [29] states that under suitable assumptions, graphical models with no hidden variables are probabilistic CI models (statistical models completely described by the CI statements they satisfy). In other words, graphical models may be described either as a factorization such as (1.2) of probability distributions or in terms of conditional independence statements (Markov properties) [27, 29, 39]. This duality between the factorization and CI representation is one of the facts about graphical models that make them so useful in practice [52].

We would like to study CI models algebraically. Key to this effort is the observation that a CI statement translates into an ideal of polynomial relations on the joint probabilities. This ideal is generated by the *cross product differences*; for I, J, K a disjoint cover of $1, \ldots, n$,

$$I_{I \perp\!\!\!\perp J|K} = \langle p_{x_I^{i_I} x_J^{i_J} x_K^{i_K}} p_{x_I^{i'_I} x_J^{i'_J} x_K^{i_K}} - p_{x_I^{i'_I} x_J^{i_J} x_K^{i_K}} p_{x_I^{i_I} x_J^{i'_J} x_K^{i_K}} : \forall i_I \neq i'_I, i_J \neq i'_J, i_K \rangle$$

Thus given a CI model $\mathcal{M}$ (a set of CI statements), we define the *conditional independence ideal*

$$I_{\mathcal{M}} = \sum_{I \perp\!\!\!\perp J|K \in \mathcal{M}} I_{I \perp\!\!\!\perp J|K}.$$

Let $I_{pairwise(G)}$ be the ideal generated by all conditional independence statements $i \perp\!\!\!\perp j|K$ where i, j are single variables and $K = [n] \setminus \{i, j\}$. The following version of the Hammersley-Clifford theorem is stated by Geiger, Meek, and Sturmfels [29].

Theorem 1.3.3 (Hammersley-Clifford Theorem)**.** *Let G be an undirected graphical model. A strictly positive probability distribution P factors according to $A(G)$ if and only if P is in the variety $X_{pairwise(G),>0}$; that is, $X_{A(G),>0} = X_{pairwise(G),>0}$.*

Chapter 2

Convex rank tests and semigraphoids

2.1 Introduction

The material in this chapter is based on the papers "Convex Rank Tests and Semigraphoids" [51] and "Geometry of Rank Tests" [50], which were authored jointly with L. Pachter, A. Shiu, B. Sturmfels and O. Wienand. Algorithm 2.6.1 and various improvements to the proofs of Theorem 2.3.3 and 2.4.1 are new. In this chapter, we develop the theory of *convex rank tests* and their connection to statistical learning theory. In Chapter 3, we will detail how to apply convex rank tests in practice.

The non-parametric approach to statistics was introduced by [54] via the method of permutation testing. Subsequent development of these ideas revealed a close connection between non-parametric tests and *rank tests*, which are statistical tests suitable for ordinal data. Beginning in the 1950s, many rank tests were developed for specific applications, such as the comparison of populations or testing hypotheses for determining the location of a population. The polyhedral geometry of these tests was first considered in in [17]. More recently, the search for patterns in large datasets has spurred the development and exploration of new tests. For instance, the emergence of microarray data in molecular biology has led to tests for identifying significant patterns in gene expression time series; see e.g. [1, 16, 23, 72]. This application motivated us to develop a mathematical theory of rank tests. We propose that a *rank test* is a partition of S_n induced by a map $\tau : S_n \to T$ from the symmetric group S_n of all permutations of $[n] = \{1, \ldots, n\}$ onto a set T of statistics. The statistic $\tau(\pi)$ is the *signature* of the permutation $\pi \in S_n$. Each rank test defines a partition of S_n into classes, where π and π' are in the same class if and only if $\tau(\pi) = \tau(\pi')$. We identify $T = \text{image}(\tau)$ with the set of all classes in this partition of S_n. Assuming the uniform distribution on S_n, the probability of seeing a particular signature $t \in T$ is $1/n!$ times $|\tau^{-1}(t)|$. The computation of a p-value for a given permutation $\pi \in S_n$ leads to the problem of summing

$$\Pr(\pi') \quad = \quad \frac{1}{n!} \cdot |\, \tau^{-1}\big(\tau(\pi')\big)\,| \tag{2.1}$$

over permutations π' with $\Pr(\pi') \leq \Pr(\pi)$, a computational task to be addressed in Section 2.6.

This chapter is organized as follows. In Section 2.2 we explain how existing rank tests in non-parametric statistics can be understood from our geometric point of view, and how they are described in the language of algebraic combinatorics [63]. In Section 2.3 we define the class of *convex rank tests.* These tests are most natural from both the statistical and the combinatorial point of view. Convex rank tests can be defined as polyhedral fans that coarsen the hyperplane arrangement of S_n. Our main result (Theorem 2.3.3) states that convex rank tests are in bijection with conditional independence structures known as semigraphoids [67, 19, 53].

Section 2.4 is devoted to convex rank tests that are induced by submodular functions. These *submodular rank tests* are in bijection with Minkowski summands of the $(n-1)$-dimensional permutohedron and with *structural imset models* [67]. These tests are at a suitable level of generality for the biological applications [72, 49] that motivated us. The connection between polytopes and conditional independence models is made concrete in the classification of small models in Remarks 2.4.6–2.4.8.

In Section 2.5 we study the subclass of *graphical tests* or *topographical models*. In combinatorics, these correspond to graph associahedra, and in statistics to undirected graphical models. The equivalence of these two structures is shown in Theorem 2.5.2. As with graphical models, topographical models provide a large class of easily-interpretable models that interface well with domain knowledge. In doing so, they extend the graphical model formalism to new types of data. The implementation of convex rank tests in applications (Chapter 3) requires the efficient enumeration of linear extensions of partially ordered sets. Our algorithms and software are discussed in Section 2.6. A key ingredient is the efficient computation of distributive lattices.

2.2 Rank tests and posets

A permutation π in S_n is a total order on the set $[n] := \{1, \ldots, n\}$. This means that π is a set of $\binom{n}{2}$ ordered pairs of elements in $[n]$. For example, $\pi = \{(1,2),(2,3),(1,3)\}$ represents the total order $1 > 2 > 3$. If π and π' are permutations then $\pi \cap \pi'$ is a partial order.

In the applications we have in mind, the data are vectors $u \in \mathbb{R}^n$ with distinct coordinates. The permutation associated with u is the total order $\pi = \{\, (i,j) \in [n] \times [n] : u_i < u_j \,\}$. We shall employ two other ways of writing a permutation. The first is the *rank vector* $\rho = (\rho_1, \ldots, \rho_n)$, whose defining properties are $\{\rho_1, \ldots, \rho_n\} = [n]$ and $\rho_i < \rho_j$ if and only if $u_i < u_j$. That is, the coordinate of the rank vector with value i is at the same position as the ith smallest coordinate of u. The second is the *descent vector* $\delta = (\delta_1|\delta_2|\ldots|\delta_n)$. The descent vector is defined by $u_{\delta_i} > u_{\delta_{i+1}}$ for $i = 1, 2, \ldots, n{-}1$. Thus the ith coordinate of the descent vector is the position of the ith largest value of the data vector u. For example, if $u = (11, 7, 13)$ then its permutation is represented by $\pi = \{(2,1), (1,3), (2,3)\}$, by $\rho = (2,1,3)$, or by $\delta = (3|1|2)$.

A permutation π is a *linear extension* of a partial order P on $[n]$ if $P \subseteq \pi$, i.e. π is a total order that refines the partial order P. We write $\mathcal{L}(P) \subseteq S_n$ for the set of linear extensions of P. A partition τ of the symmetric group S_n is a *pre-convex rank test* if the following axiom holds:

(PC) If $\tau(\pi) = \tau(\pi')$ and $\pi'' \in \mathcal{L}(\pi \cap \pi')$ then $\tau(\pi){=}\tau(\pi'){=}\tau(\pi'')$.

Note that $\pi'' \in \mathcal{L}(\pi \cap \pi')$ means $\pi \cap \pi' \subseteq \pi''$. The number of all rank tests τ on $[n]$ is the *Bell number* $B_{n!}$, which is the number of set partitions of a set of cardinality $n!$.

Example 2.2.1. For $n = 3$ there are $B_6 = 203$ rank tests, or partitions of the symmetric group S_3, which consists of six permutations. Of these 203 rank tests, only 40 satisfy the axiom (PC). One example is the pre-convex rank test in Figure 2.1. Here the symmetric group S_3 is partitioned into the four classes $\big\{(1|2|3)\big\}$, $\big\{(2|1|3)\big\}$, $\big\{(2|3|1)\big\}$, and $\big\{(1|3|2), (3|1|2), (3|2|1)\big\}$.

Each class C of a pre-convex rank test τ corresponds to a poset P on the ground set $[n]$; namely, the partial order P is the intersection of all total orders in that class: $P = \bigcap_{\pi \in C} \pi$. The axiom (PC) ensures that C coincides with the set $\mathcal{L}(P)$ of all linear extensions of P. The inclusion $C \subseteq \mathcal{L}(P)$ is clear. The proof of the reverse inclusion $\mathcal{L}(P) \subseteq C$ is based on the fact that, from any permutation π in $\mathcal{L}(P)$, we can obtain any other π' in $\mathcal{L}(P)$ by a sequence of reversals $(a, b) \mapsto (b, a)$, where each intermediate $\hat{\pi}$ is also in $\mathcal{L}(P)$. Consider any $\pi_0 \in \mathcal{L}(P)$ and suppose that $\pi_1 \in C$ differs by only one reversal $(a, b) \in \pi_0$, $(b, a) \in \pi_1$. Then $(b, a) \notin P$, so there is some $\pi_2 \in C$ such that $(a, b) \in \pi_2$; thus, $\pi_0 \in \mathcal{L}(\pi_1 \cap \pi_2)$ by (PC). This shows $\pi_0 \in C$.

A pre-convex rank test therefore can be characterized by an unordered collection of posets $P_1, P_2, \ldots, P_k$ on $[n]$ that satisfies the property that the symmetric group S_n is the disjoint union of the subsets $\mathcal{L}(P_1), \mathcal{L}(P_2), \ldots, \mathcal{L}(P_k)$. This structure was discovered independently and studied by Postnikov, Reiner and Williams [56, §3] who used the term *complete fan of posets* for what we shall call a convex rank test in Section 2.3. The posets $P_1, P_2, \ldots, P_k$ that represent the classes in a pre-convex rank test capture the shapes of data vectors. In graphical rank tests (Section 2.5), this shape can be interpreted as a smoothed topographic map of the data vector.

Example 2.2.2 (The sign test for paired data)**.** The *sign test* is performed on data that are paired as two vectors $u = (u_1, u_2, \ldots, u_m)$ and $v = (v_1, v_2, \ldots, v_m)$. The null hypothesis is that the median of the differences $u_i - v_i$ is 0. The test statistic is the number of differences that are positive. This test is a rank test, because u and v can be transformed into the overall ranks of the $n = 2m$ values, and the rank vector entries can then be compared. This test coarsens the convex rank test which is the MSS test of Section 2.4 with $\mathcal{K} = \{\{1, m+1\}, \{2, m+2\}, \ldots\}$.

Example 2.2.3 (Runs tests)**.** A *runs test* can be used when there is a natural ordering on the data points, such as in a time series. The data are transformed into a sequence of 'pluses' and 'minuses,' and the null hypothesis is that the number of observed runs is no more than that expected by chance. Common types of runs tests include the sequential runs test ('plus' if consecutive data points increase, 'minus' if they decrease), and the runs test to check randomness of residuals, i.e. deviation from a curve fit to the data. A runs test is a coarsening of a convex rank test, known as *up-down analysis* [72, §6.1.1], which is described in Example 2.3.4 below.

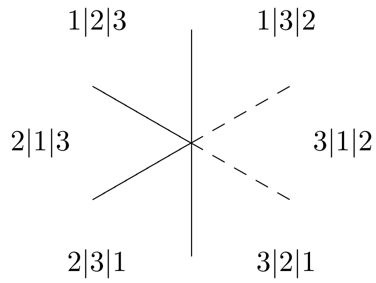


Figure 2.1: Illustration of a pre-convex rank test that is not convex. Cones are labeled by descent vectors, so 1|2|3 indicates the cone $u_1 > u_2 > u_3$. This rank test is specified by the four posets $P_1 = \{3{<}1, 2{<}1, 3{<}2\}$, $P_2 = \{1{<}2, 3{<}2, 3{<}1\}$, $P_3 = \{3{<}2, 1{<}3, 1{<}2\}$ and $P_4 = \{2{<}3\}$.

These two examples suggest that many rank tests from classical non-parametric statistics have a natural refinement by a pre-convex rank test. However, not all tests have this property. Because many classical rank tests apply to loosely grouped data (e.g. data which are divided into two samples), the axiom (PC) is not always satisfied. In such cases, the pre-convex rank test is a first step, after which permutations are grouped together under additional symmetries, e.g., the permutations $\delta = (1|2|3|4|5)$ and $\delta' = (5|4|3|2|1)$ might be identified.

The adjective "pre-convex" refers to the following interpretation of the axiom (PC).

Consider any two data vectors u and u' in $\mathbb{R}^n$, and a convex combination $u'' = \lambda u + (1-\lambda)u'$, with $0 < \lambda < 1$. If π, π', π'' are the permutations of u, u', u'' then $\pi'' \in \mathcal{L}(\pi \cap \pi')$. Thus the equivalence classes in $\mathbb{R}^n$ specified by a pre-convex rank test are convex cones. In the next section, we shall remove the prefix from "pre-convex" if the faces of these cones fit together well.

2.3 Convex rank tests

Two vectors u and v in $\mathbb{R}^n$ are *permutation equivalent* when $u_i < u_j$ if and only if $v_i < v_j$, and $u_i = u_j$ if and only if $v_i = v_j$ for all $i, j \in [n]$. Note that for two data vectors, each with distinct coordinates, they are permutation equivalent if and only if they have the same rank vector. The permutation equivalence classes (of which there are 13 for $n = 3$) induce a fan (see Definition 1.1.1) called the *S_n-fan.* The set of sets of vectors generating this fan defines a vector fan (§2.4) as well. The arrangement of hyperplanes $\{x_i = x_j\}$ that defines these classes is also known as the *braid arrangement* , and its regions as the *Weyl chambers* of the Lie algebra $\mathfrak{sl}(n)$. The maximal cones in the S_n-fan, which are the closures of the permutation equivalence classes, are indexed by descent permutations δ in S_n. A *coarsening* of the S_n-fan is a fan $\mathcal{F}$ such that each permutation equivalence class of $\mathbb{R}^n$ is fully contained in a cone C of $\mathcal{F}$. Such a fan $\mathcal{F}$ defines a partition of S_n because each maximal cone of the S_n-fan is contained in some cone $C \in \mathcal{F}$.

Definition 2.3.1. A *convex rank test* is a partition of the symmetric group S_n which is induced by a coarsening of the S_n-fan. We identify the fan with that rank test.

We say that two maximal cones, indexed by δ and δ', of the S_n-fan *share a wall* if there exists an index k such that $\delta_k = \delta'_{k+1}$, $\delta_{k+1} = \delta'_k$, and $\delta_i = \delta'_i$ for $i \notin \{k, k+1\}$.

This condition means that the corresponding permutations δ and δ' differ by an adjacent transposition. To such an unordered pair $\{\delta, \delta'\}$, we associate the following (elementary) conditional independence (CI) statement:

$$\delta_k \perp\!\!\!\perp \delta_{k+1} \,|\, \{\delta_1, \ldots, \delta_{k-1}\}. \tag{2.2}$$

Example 2.3.2. For $n = 3$ there are 40 pre-convex rank tests (Example 2.2.1), but only 22 of them are convex rank tests. The corresponding CI models are shown in Figure 5.6 on page 108 in [67].

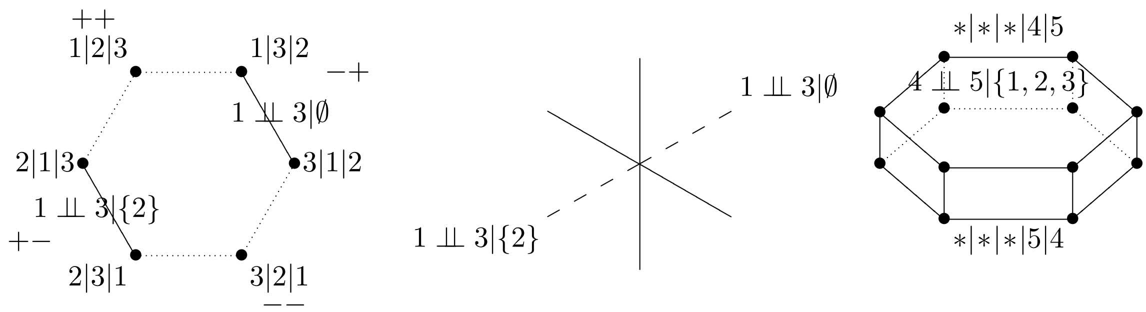
The formula (2.2) defines a map from the set of walls of the S_n-fan onto the set

$$\mathcal{T}_n \;:=\; \big\{\, i \perp\!\!\!\perp j \,|\, K \;:\; K \subseteq [n]\backslash\{i,j\} \big\}.$$

of all elementary CI statements. In this manner, each wall of the S_n-fan is labeled by a CI statement. The map from walls to CI statements is not injective; there are $(n-k-1)!(k-1)!$ walls which are labeled by (2.2). See Figure 2.2(c) for an illustration.

The S_n-fan is the normal fan of the *permutohedron* $\mathbf{P}_n$, which is the $(n-1)$-dimensional convex hull of the vectors $(\rho_1, \ldots, \rho_n) \in \mathbb{R}^n$, where ρ runs over all rank vectors of permutations in S_n. Each edge of $\mathbf{P}_n$ joins two permutations if they differ by an adjacent transposition. In other words, each edge corresponds to a wall and is thus labeled by a CI statement. A collection of parallel edges of $\mathbf{P}_n$ that are perpendicular to a given hyperplane $\{x_i = x_j\}$ corresponds to the set of CI statements $i \perp\!\!\!\perp j|K$, where K ranges over all subsets of $[n]\backslash\{i,j\}$.

The two-dimensional faces of $\mathbf{P}_n$ are squares and regular hexagons, and two edges of $\mathbf{P}_n$ have the same label in $\mathcal{T}_n$ if, but not only if, they are opposite edges of a square. Figure 2.2(c) depicts the subset of $\mathbf{P}_5$ in which the last two coordinates of $u \in \mathbb{R}^n$ are less than or equal to all other coordinates. It consists of two copies of the hexagon in 2(a), with the final two entries of the descent vector either 4|5 (in the top hexagon) or 5|4 (in the bottom hexagon). All vertical edges are labeled by the CI statement $4 \perp\!\!\!\perp 5|\{1,2,3\}$.



(a) **(b)** **(c)**

Figure 2.2: **(a)** The permutohedron $\mathbf{P}_3$ and **(b)** the S_3-fan projected to the plane. The indicated rank test is up-down analysis. Each permutation is represented by its descent vector $\delta = \delta_1|\delta_2|\delta_3$. Missing walls of the S_n-fan, or solid edges of $\mathbf{P}_n$, are labeled by CI statements. **(c)** Edges of the permutohedron on opposite sides of a square (here, all vertical edges) are labeled by the same CI statement; hexagonal prisms such as the one pictured here appear in $\mathbf{P}_n$ for $n \geq 5$.

Any convex rank test $\mathcal{F}$ is characterized by the collection of walls $\{\delta, \delta'\}$ that are removed when passing from the S_n-fan to $\mathcal{F}$. So, from (2.2), any convex rank test $\mathcal{F}$ maps to a set $\mathcal{M}_\mathcal{F}$ of CI statements corresponding to missing walls, or a set $\mathbf{M}_\mathcal{F}$ of edges of the permutohedron. For example, if $\mathcal{F}$ is the fan obtained by removing the two dashed rays in Figure 2.2 (b) then the corresponding set of CI statements is $\mathcal{M}_\mathcal{F} = \big\{1 \perp\!\!\!\perp 3|\emptyset,\ 1 \perp\!\!\!\perp 3|\{2\}\big\}$.

Recall from Section 1.3.3 that a semigraphoid is a set of CI statements subject to certain axioms. In fact, a semigraphoid is determined by its *trace* among elementary CI statements of the form $i \perp\!\!\!\perp j \,|\, K$ where i and j are singletons. Namely, $I \perp\!\!\!\perp J|K$ holds if and only if $i \perp\!\!\!\perp j|L$ for all $i \in I, j \in J$ and L such that $K \subseteq L \subseteq (I \cup J \cup K) \setminus ij$; see [43]. Casting the semigraphoid axiom in terms of the trace, we say that a subset $\mathcal{M}$ of $\mathcal{T}_n$ is a semigraphoid if $i \perp\!\!\!\perp j \,|\, K \in \mathcal{M}$ implies $j \perp\!\!\!\perp i \,|\, K \in \mathcal{M}$ and the following axiom holds:

$$(\mathbf{SG'}) \qquad i \perp\!\!\!\perp j \,|\, K \cup \ell \in \mathcal{M} \quad \text{and} \quad i \perp\!\!\!\perp \ell \,|\, K \in \mathcal{M}$$
$$\text{implies} \quad i \perp\!\!\!\perp j \,|\, K \in \mathcal{M} \quad \text{and} \quad i \perp\!\!\!\perp \ell \,|\, K \cup j \in \mathcal{M}.$$

This axiom is stated in [67, 46]. Our first result is that semigraphoids and convex rank tests are the same combinatorial object:

Theorem 2.3.3. *The map $\mathcal{F} \mapsto \mathcal{M}_{\mathcal{F}}$ is a bijection between convex rank tests and semigraphoids.*

Before presenting the proof of this theorem, we shall discuss an important example.

Example 2.3.4 (Up-down analysis)**.** Let $\mathcal{F}$ denote the convex rank test called up-down analysis [72]. In this test, each permutation $\pi \in S_n$ is mapped to the sign vector of its first differences, or, equivalently, its descent set. Thus this test is the natural map $\tau : S_n \to \{-,+\}^{n-1}$. The corresponding semigraphoid $\mathcal{M}_{\mathcal{F}}$ consists of all CI statements $i \perp\!\!\!\perp j \,|\, K$ where $|i-j| \geq 2$.

This convex rank test is visualized in Figure 2.2(a,b) for $n = 3$. Permutations are in the same class (have the same sign pattern) if they are connected by a solid edge; there are four classes. In the S_3-fan, the two missing walls are labeled by conditional independence statements as defined in (2.2). For $n = 4$ the up-down analysis test $\mathcal{F}$ is depicted in Figure 2.3. The double edges correspond to the twelve CI statements in $\mathcal{M}_{\mathcal{F}}$. There are eight classes; e.g., the class $\{3|4|1|2, 3|1|4|2, 1|3|4|2, 1|3|2|4, 3|1|2|4\}$ consists of the five permutations in S_4 which have the up-down pattern $(-,+,-)$.

Our proof of Theorem 2.3.3 rests on translating the semigraphoid axiom $(\mathbf{SG'})$ into geometric statements about edges of the permutohedron. Recall that a semigraphoid

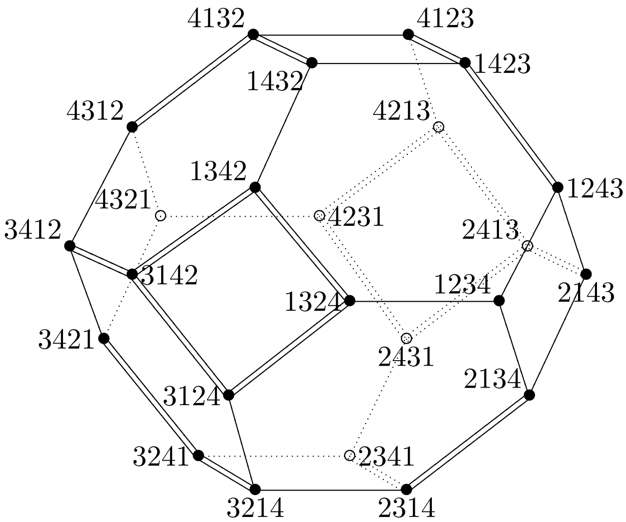
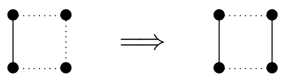


Figure 2.3: The permutohedron $\mathbf{P}_4$ with vertices marked by descent vectors δ (bars | omitted). The convex rank test indicated by the double edges is up-down analysis.

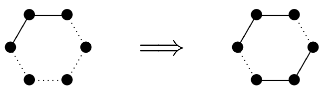
$\mathcal{M}$ can be identified with the set $\mathbf{M}$ of edges of the permutohedron whose CI statement labels are those of $\mathcal{M}$.

Observation 2.3.5. *A set* $\mathbf{M}$ *of edges of the permutohedron* $\mathbf{P}_n$ *is a semigraphoid if and only if the set* $\mathbf{M}$ *satisfies the following two geometric axioms:*

Square axiom: *Whenever an edge of a square is in* $\mathbf{M}$*, then the opposite edge is also in* $\mathbf{M}$.



Hexagon axiom: *Whenever two adjacent edges of a hexagon are in* $\mathbf{M}$*, then the two opposite edges of that hexagon are also in* $\mathbf{M}$.



Let $\mathbf{M}$ be the subgraph of the edge graph of $\mathbf{P}_n$ defined by the statements in $\mathcal{M}$; that is, $\mathbf{M}$ consists of edges whose labels are in $\mathcal{M}$. Each class of the rank test defined by

$\mathcal{M}$ consists of the permutations in some connected component of $\mathbf{M}$. We regard a path from δ to δ' on $\mathbf{P}_n$ as a word $\sigma^{(1)} \cdots \sigma^{(l)}$ in the free associative algebra $\mathcal{A}$ generated by the adjacent transpositions of $[n]$. For example, the transposition $\sigma_{23} := (23)$ gives the path from δ to $\delta' = \sigma_{23}\delta = \delta_1|\delta_3|\delta_2|\delta_4|\ldots|\delta_n$. The following relations in $\mathcal{A}$ define a presentation of the group algebra of S_n as a quotient of $\mathcal{A}$:

(BS) $\quad \sigma_{i,i+1} \cdot \sigma_{i+k+1,i+k+2} - \sigma_{i+k+1,i+k+2} \cdot \sigma_{i,i+1},$

(BH) $\quad \sigma_{i,i+1} \cdot \sigma_{i+1,i+2} \cdot \sigma_{i,i+1} - \sigma_{i+1,i+2} \cdot \sigma_{i,i+1}\sigma_{i+1,i+2},$ and

(BN) $\quad \sigma_{i,i+1}^2 - 1,$

where suitable i and k vary over $[n]$. The first two are the *braid relations* , and the third represents the idempotency of each transposition.

Now, we regard these relations as properties of a set of edges of $\mathbf{P}_n$, by identifying a word and a permutation δ with the set of edges that comprise the corresponding path in $\mathbf{P}_n$. For example, a set satisfying (BS) is one such that, starting from any δ, the edges of the path $\sigma_{i,i+1}\sigma_{i+k+1,i+k+2}$ are in the set if and only if the edges of the path $\sigma_{i+k+1,i+k+2}\sigma_{i,i+1}$ are in the set. Note then, that (BS) is the square axiom, and (BH) is a weaker version of the hexagon axiom of semigraphoids. That is, implications in either direction hold in a semigraphoid. However, (BN) holds only directionally in a semigraphoid: if an edge lies in the semigraphoid, then its two vertices are in the same class; but the empty path at some vertex δ certainly does not imply the presence of all incident edges in the semigraphoid. Thus, for a semigraphoid, (BS) and (BH) hold, but (BN) must be replaced with the directional version

(BN′) $\quad \sigma_{i,i+1}^2 \to 1.$

We now consider a path p from δ to δ' in a semigraphoid. Here is a crucial lemma for our proof:

Lemma 2.3.6. *Suppose that $\mathcal{M}$ is a semigraphoid. If δ and δ' lie in the same class of $\mathcal{M}$, then so do all shortest paths on $\mathbf{P}_n$ between them.*

The lemma in turn depends on the following version of a classical result due to Jacques Tits. This result, which can be found in [12, p. 49-51], essentially states that the relations (BS),(BH),(BN) form a Gröbner basis for the two-sided ideals they generate in $\mathcal{A}$.

Theorem 2.3.7 (Tits [70]). *Let p and q be words representing paths on $\mathbf{P}_n$.*

(1) A word p is (BS),(BH),(BN)-reduced if and only if it is (BS),(BH),(BN')-reduced.

(2) If p and q are reduced, then they represent the same element of the symmetric group S_n if and only if p can be transformed to q by the application of (BS) and (BH) only.

Proof of Lemma 2.3.6. Theorem 2.3.7 (1) says that if there is any path connecting δ and δ', then there is a shortest path connecting them. Thus if δ and δ' lie in the same class of $\mathcal{M}$, some shortest path $\delta \to \delta'$ also lies in that class. Now (2) says that if p and q are both shortest paths, then q can be obtained from p by application of only the square and hexagon axioms, (BS) and (BH). Thus if any shortest path $\delta \to \delta'$ lies in the class of $\mathcal{M}$ containing them both, so do all other shortest paths connecting them. □

Proof of Theorem 2.3.3. Both semigraphoids and convex rank tests can be regarded as sets of edges of $\mathbf{P}_n$. We first show that a semigraphoid satisfies (PC). Consider δ, δ' in the same class C of a semigraphoid, and let $\delta'' \in \mathcal{L}(\delta \cap \delta')$. Further, let p be a shortest path from δ to δ'' (so, $p\delta = \delta''$), and let q be a shortest path from δ'' to δ'. We claim that qp is a shortest path from δ to δ', and thus $\delta'' \in C$ by Lemma 2.3.6. Suppose qp is not a shortest path.

Then, we can obtain a shorter path in the semigraphoid by some sequence of substitutions according to (BS), (BH), and (BN'). Only (BN') decreases the length of a path, so the sequence must involve (BN'). Therefore, there is some i, j in $[n]$, such that their positions relative to each other are reversed twice in qp. But p and q are shortest paths, hence one reversal occurs in each of p and q. Then δ and δ' agree on whether $i > j$ or $j > i$, but the reverse holds in δ'', contradicting $\delta'' \in \mathcal{L}(\delta \cap \delta')$. Thus every semigraphoid is a pre-convex rank test.

Now, we show that a semigraphoid corresponds to a fan. Consider the cone corresponding to a class C. We need only show that it meets any other cone in a shared face. Since C is a cone of a coarsening of the S_n-fan, each nonmaximal face of C lies in a hyperplane $H = \{x_i = x_j\}$. Suppose a face of C coincides with the hyperplane H and that $i > j$ in C. A vertex δ borders H if i and j are adjacent in δ. We will show that if $\delta, \delta' \in C$ border H, then their reflections $\widehat{\delta} = \delta_1|\ldots|j|i|\ldots|\delta_n$ and $\widehat{\delta'} = \delta'_1|\ldots|j|i|\ldots|\delta'_n$ both lie in some class C'. Consider a 'great circle' path between δ and δ' which stays closest to H: all vertices in the path have i and j separated by at most one position, and no two consecutive vertices have i and j nonadjacent. This is a shortest path, so it lies in C, by Lemma 2.3.6. Using the square and hexagon axioms (Observation 2.3.5), we see that the reflection of the path across H is a path in the semigraphoid that connects $\widehat{\delta}$ to $\widehat{\delta'}$ (Figure 2.4). Thus a semigraphoid is a convex rank test.

Finally, if $\mathbf{M}$ is a set of edges of $\mathbf{P}_n$ representing a convex rank test, then it is easy to show that $\mathbf{M}$ satisfies the square and hexagon axioms. $\square$

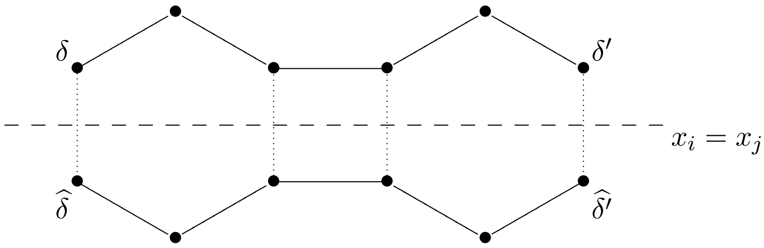


Figure 2.4: Reflecting a path across a hyperplane.

2.4 The submodular cone

In this section we focus on a subclass of the convex rank tests. Let $2^{[n]}$ denote the collection of all subsets of $[n] = \{1, 2, \ldots, n\}$. Any real-valued function $w : 2^{[n]} \to \mathbb{R}$ defines a convex polytope Q_w of dimension $\leq n-1$ as follows:

$$Q_w \;:=\; \big\{\, x \in \mathbb{R}^n \,:\, x_1 + x_2 + \cdots + x_n = w([n]) \\ \text{and } \textstyle\sum_{i \in I} x_i \leq w(I) \text{ for all } \emptyset \neq I \subseteq [n] \,\big\}.$$

A function $w : 2^{[n]} \to \mathbb{R}$ is called *submodular* if $w(I) + w(J) \;\geq\; w(I \cap J) + w(I \cup J)$ for $I, J \subseteq [n]$. The *submodular cone* is the cone $\mathbf{C}_n$ of all submodular functions $w : 2^{[n]} \to \mathbb{R}$. Working modulo its lineality space $\mathbf{C}_n \cap (-\mathbf{C}_n)$, we regard $\mathbf{C}_n$ as a pointed cone of dimension $2^n - n - 1$.

Studying functions w means that in considering the normal fan of a polytope Q_w, we want to retain information about non-binding inequalities that are just barely so, i.e. that hold with equality. For this reason we define the *vector (normal) fan* [6]. The indicator function of each $I \in 2^{[n]}$ defines a vector e_I in the 1-skeleton of the S_n-fan, understood modulo $e_{[n]}$; for example, these vectors for $n = 3$ are $e_{001}, e_{010}, e_{100}, e_{011}, \ldots, e_{111}$. A *vector fan* $\mathfrak{F}$ with base $\{e_I : I \in 2^{[n]}\}$ is a collection of subsets of $2^{[n]}$ such that $e_I, e_J \in \mathfrak{F}$ implies $e_{I \cap J} \in \mathfrak{F}$. A vector fan defines a usual fan by taking the maximal cones of the fan to be the cones generated by the vector sets in the vector fan. We say that a vector fan is *complete* if its fan is. A vector fan $\mathfrak{F}$ *coarsens* another vector fan $\mathfrak{G}$ if for all $e_I \in \mathfrak{G}$, there exists $e_J \in \mathfrak{F}$ with $I \subset J$.

Given a function $w : 2^{[n]} \to \mathbb{R}$, Each I also defines an inequality $\sum_{i \in I} x_i \le w_I$ appearing in the definition of Q_w; the vector normal fan tells us which of these inequalities holds with equality on some face of Q_w. We define the *vector normal fan* of a function $w : 2^{[n]} \to \mathbb{R}$ is the set of subsets of $\{e_I : I \in 2^{[n]}, \sum_{i \in I} x_i = w_I$ for all x in some face of $Q_w\}$. The vector normal fan of w defines a fan which is the normal fan of Q_w; however the vector fan retains additional information about w.

Proposition 2.4.1. *A function $w : 2^{[n]} \to \mathbb{R}$ is submodular if and only if the vector normal fan of w is a coarsening of the vector S_n-fan.*

Example 2.4.2. Let $w_1 = w_2 = w_3 = 1, w_{12} = w_{13} = w_{23} = w_{123} = 3$. The polytope Q_w is the point $(1, 1, 1)$ but the function w is not submodular. The vector normal fan $\mathfrak{F}$ of w is $\{\{e_{001}, e_{010}, e_{100}\}\}$ and the normal fan is all of $\mathbb{R}^3/(1, 1, 1)$. $\mathfrak{F}$ does not coarsen the S_n-fan since, for example, e_{110} is not contained in any set in $\mathfrak{F}$.

However, if we change w slightly to define the same Q_w but with the inequalities corresponding to $011, 101$, and 110 also holding with equality, e.g. $w_1 = w_2 = w_3 = 1, w_{12} = w_{13} = w_{23} = 2$, and $w_{123} = 3$, the resulting vector normal fan of w is a coarsening of the (vector) S_n-fan.

Proof. We show only the if direction. Suppose w is not submodular. Then there exist $I, J \subset 2^{[n]}$ such that

$$w_I + w_J < w_{I \cap J} + w_{I \cup J}$$

We also have that

$$\begin{aligned}\sum_{i\in I\cup J} x_i + \sum_{i\in I\cap J} x_i &= \sum_{i\in I} x_i + \sum_{i\in J} x_i \\ &\leq w_I + w_J < w_{I\cap J} + w_{I\cup J}\end{aligned}$$

So $\sum_{i\in I\cup J} x_i < w_{I\cup J} + (w_{I\cap J} - \sum_{i\in I\cap J} x_i)$ and similarly $\sum_{i\in I\cap J} x_i < w_{I\cap J} + (w_{I\cup J} - \sum_{i\in I\cup J} x_i)$, so that at most one of the inequalities corresponding to $I \cup J$ and $I \cap J$ can hold with equality at any point of Q_w. Then either $e_{I\cap J}$ or $e_{I\cup J}$ is not contained in any of the sets in the vector normal fan of w. □

Proposition 2.4.1 can be paraphrased as follows: the function w is submodular if and only if the optimal solution of

$$\text{maximize } u \cdot x \text{ subject to } x \in Q_w$$

depends only on the permutation equivalence class of u. Thus, solving this linear programming problem constitutes a convex rank test. Any such test is called a *submodular rank test*

.

Recall that a convex polytope is a Minkowski summand of another polytope if the normal fan of the latter refines the normal fan of the former. The polytope Q_w that represents a submodular rank test is a summand of the permutohedron $\mathbf{P}_n$.

Theorem 2.4.3. *The following combinatorial objects are equivalent for any positive integer n:*

1. *submodular rank tests,*
2. *summands of the permutohedron $\mathbf{P}_n$,*
3. *structural conditional independence models [67],*
4. *faces of the submodular cone $\mathbf{C}_n$ in $\mathbb{R}^{2^n}$.*

Proof. We have $1 \iff 2$ from Proposition 2.4.1, and $1 \iff 3$ follows from [67]. Further, $1 \iff 4$ is a direct consequence of our definition of submodular rank tests. □

Remark 2.4.4. All 22 convex rank tests for $n = 3$ are submodular. The submodular cone $\mathbf{C}_3$ is a 4-dimensional cone whose base is a bipyramid. Its f-vector is $(1, 5, 9, 6, 1)$. The polytopes Q_w, as w ranges over representatives of the faces of $\mathbf{C}_3$, are all the Minkowski summands of $\mathbf{P}_3$.

Proposition 2.4.5. *For $n \geq 4$, there exist convex rank tests that are not submodular rank tests. Equivalently, there are fans that coarsen the S_n-fan but are not the normal fan of any polytope.*

Proof. This result is well-known. It is stated in Section 2.2.4 of [67] in the following form: "There exist semigraphoids that are not structural." □

An interesting example which also proves Proposition 2.4.5 is the following semigraphoid:

$$\mathcal{M} \quad = \quad \big\{ 2 \perp\!\!\!\perp 3 | \{1,4\},\, 1 \perp\!\!\!\perp 4 | \{2,3\},\, 1 \perp\!\!\!\perp 2 | \emptyset,\, 3 \perp\!\!\!\perp 4 | \emptyset \,\big\}.$$

The corresponding fan consists of unimodular cones, or, equivalently, the posets P_i representing this non-submodular convex rank test are all trees. This example answers a question posed in the first version of [56]. In Chapter 4, we will develop a systematic method for showing that a semigraphoid is not submodular.

Remark 2.4.6. For $n = 4$ there are 22108 submodular rank tests, one for each face of the 11-dimensional cone $\mathbf{C}_4$. The base of this submodular cone is a 10-dimensional polytope with f-vector $(1, 37, 356,\ 1596, 3985, 5980, 5560, 3212, 1128, 228, 24, 1)$. The 37 vertices of this polytope correspond to the maximal semigraphoids. These come in seven symmetry

Type	Symmetry	Count	$i \perp\!\!\!\perp j$
1	$1\times$ and $*$	2	all
2	$4\times$ and $*$	8	all
3	$6\times$ incl. $*$	6	all but $1 \perp\!\!\!\perp 2$
4	$4\times$ and $*$	8	all
5	$1\times$, self-$*$	1	all
6	$6\times$ incl. $*$	6	all but $1 \perp\!\!\!\perp 2$
7	$6\times$ incl. $*$	6	$3 \perp\!\!\!\perp 4$

Type	$i \perp\!\!\!\perp j \mid k$	$i \perp\!\!\!\perp j \mid \{k,l\}$
1	all	none
2	all but $2 \perp\!\!\!\perp 3\mid 1, 1 \perp\!\!\!\perp 3\mid 2, 1 \perp\!\!\!\perp 2\mid 3$	$3 \perp\!\!\!\perp 4\mid 12, 2 \perp\!\!\!\perp 4\mid 13, 1 \perp\!\!\!\perp 4\mid 23$
3	all but $1 \perp\!\!\!\perp 2\mid 3, 1 \perp\!\!\!\perp 2\mid 4$	all but $1 \perp\!\!\!\perp 2\mid 34$
4	$2 \perp\!\!\!\perp 3\mid 4, 2 \perp\!\!\!\perp 4\mid 3, 3 \perp\!\!\!\perp 4\mid 2$	$3 \perp\!\!\!\perp 4\mid 12, 2 \perp\!\!\!\perp 4\mid 13, 2 \perp\!\!\!\perp 3\mid 14$
5	none	all
6	$2 \perp\!\!\!\perp 3\mid 1, 2 \perp\!\!\!\perp 4\mid 1, 1 \perp\!\!\!\perp 3\mid 2, 1 \perp\!\!\!\perp 4\mid 2$	all but $3 \perp\!\!\!\perp 4\mid 12$
7	all but $2 \perp\!\!\!\perp 3\mid 4, 2 \perp\!\!\!\perp 4\mid 3, 1 \perp\!\!\!\perp 4\mid 3, 1 \perp\!\!\!\perp 3\mid 4$	$1 \perp\!\!\!\perp 2\mid 34$

Table 2.1: Maximal semigraphoids for $n = 4$.

classes up to the $*$ involution and the S_4-action. The types of maximal semigraphoids for $n = 4$ are displayed in Table 2.1.

Remark 2.4.7. For $n = 5$ there are 117978 coarsest submodular rank tests, in 1319 S_5 symmetry classes. We confirmed this result of [68] with `POLYMAKE` [28].

We now define a class of submodular rank tests, which we call *Minkowski sum of simplices (MSS) tests* . Note that each subset K of $[n]$ defines a submodular function w_K by setting $w_K(I) = 1$ if $K \cap I$ is non-empty and $w_K(I) = 0$ if $K \cap I$ is empty. The corresponding polytope Q_{w_K} is the simplex $\Delta_K = \text{conv}\{e_k : k \in K\}$.

Now consider an arbitrary subset $\mathcal{K} = \{K_1, K_2, \ldots, K_r\}$ of $2^{[n]}$. It defines the submodular function $w_{\mathcal{K}} = w_{K_1} + w_{K_2} + \cdots + w_{K_r}$. The corresponding polytope is the Minkowski sum

$$\Delta_{\mathcal{K}} \quad = \quad \Delta_{K_1} + \Delta_{K_2} + \cdots + \Delta_{K_r}. \tag{2.3}$$

The associated MSS test $\tau_{\mathcal{K}}$ is defined as follows. Given $\rho \in S_n$, we compute the number of indices $j \in [r]$ such that $\max\{\rho_k : k \in K_j\} = \rho_i$, for each $i \in [n]$. The signature $\tau_{\mathcal{K}}(\rho)$ is the vector in $\mathbb{N}^n$ whose ith coordinate is that number. Few submodular rank tests are MSS tests:

Remark 2.4.8. For $n = 3$, there are 22 submodular rank tests, but only 15 of them are MSS tests. For $n = 4$, there are 22108 submodular rank tests, but only 1218 of them are MSS tests.

In light of Theorem 2.3.3, it is natural to ask which semigraphoids correspond to an MSS test. Geometrically, we wish to know which edges of the permutohedron $\mathbf{P}_n$ are contracted when passing to the polytope $Q_{w_{\mathcal{K}}}$. To be precise, let $\mathcal{M}_{\mathcal{K}}$ denote the semigraphoid derived from $\mathcal{F}_{w_{\mathcal{K}}}$ using the bijection in Theorem 2.3.3. We then have the following result:

Proposition 2.4.9. *The semigraphoid $\mathcal{M}_{\mathcal{K}}$ is the set of CI statements of the form $i \perp\!\!\!\perp j \,|\, K$ where all sets containing $\{i,j\}$ and contained in $\{i,j\} \cup [n] \backslash K$ are not in $\mathcal{K}$.*

Proof. Consider two permutations δ and δ' which are adjacent on the permutohedron $\mathbf{P}_n$, and let $i \perp\!\!\!\perp j \,|\, K$ be the label of the edge that connects δ and δ'. That CI statement is in $\mathcal{M}_{\mathcal{K}}$ if and only if δ and δ' are mapped to the same vertex in $\Delta_{\mathcal{K}}$ if and only if δ and δ' are mapped to the same vertex in each simplex Δ_{K_l} for $l = 1, 2, \ldots, r$. For each l, this means that the leftmost entry of the descent vector δ that lies in K_l agrees with the leftmost entry of the other descent vector δ' that lies in K_l. This condition is equivalent to

$$K_l \cap (K \cup \{i,j\}) \neq \{i,j\} \qquad \text{for } l = 1, 2, \ldots, r.$$

Thus $i \perp\!\!\!\perp j \,|\, K$ is in the semigraphoid $\mathcal{M}_{\mathcal{K}}$ associated with the set family $\mathcal{K}$ if and only if $\mathcal{K}$ contains no set whose intersection with $K \cup \{i, j\}$ equals $\{i, j\}$. This is precisely our claim. □

There is a natural involution $*$ on the set of all CI statements which is defined as follows:

$$(i \perp\!\!\!\perp j \,|\, C)^* \quad := \quad i \perp\!\!\!\perp j \,|\, [n] \backslash (C \cup \{i, j\}).$$

If $\mathcal{M}$ is any semigraphoid, then the semigraphoid $\mathcal{M}^*$ is obtained by applying the involution $*$ to all the CI statements in the model $\mathcal{M}$. This involution is referred to as *duality* in [42]. In the *boolean lattice*, whose elements are the subsets of $[n]$, the involution corresponds to switching the role of set intersection and set union.

The MSS test $\tau_{\mathcal{K}}$ was defined above in terms of weight functions w. What follows is a similar construction for the duals of MSS tests. Let $z_{\mathcal{K}}(J) = 1$ for $J \in \mathcal{K}$ and $z_{\mathcal{K}}(J) = 0$ otherwise. Then the function $w^* : 2^{[n]} \to \mathbb{R}$ defined by $w^*_{\mathcal{K}}(I) := \sum_{J \subset I} z_{\mathcal{K}}(J)$ is supermodular. We set

$$Q^*_w \;:=\; \big\{ \, x \in \mathbb{R}^n \,:\, x_1 + x_2 + \cdots + x_n = w([n]) \\ \text{and } \textstyle\sum_{i \in I} x_i \geq w(I) \text{ for all } \emptyset \neq I \subseteq [n] \, \big\}.$$

Then the equality $Q^*_{w_{\mathcal{K}}} = \Delta_K$ holds for $\Delta_{\mathcal{K}} = \Delta_{\mathcal{K}_1} + \Delta_{K_2} + \cdots + \Delta_{K_r}$. This equality is precisely the statement in Proposition 6.3 of Postnikov's "Permutohedra, Associahedra, and Beyond" [55].

2.5 Graphical tests

We have seen that semigraphoids are equivalent to convex rank tests. We now explore the connection to graphical models. Let G be a graph with vertex set $[n]$ and $\mathcal{K}(G)$

the collection of all subsets $K \subseteq [n]$ such that the induced subgraph of $G|_K$ is connected. Recall that the *undirected graphical model* (or *Markov random field*) derived from the graph G is the set $\mathcal{M}^G$ of CI statements:

$$\mathcal{M}^G \;=\; \big\{\, i \perp\!\!\!\perp j \,|\, C \;:\; \text{the restriction of } G \text{ to } [n]\backslash C \text{ contains no path from } i \text{ to } j \big\}. \tag{2.4}$$

Theorem 2.5.1. *The set $\mathcal{M}^G$ of CI statements in the graphical model G is equal to the semigraphoid $\mathcal{M}_{\mathcal{K}(G)}$ associated with the family $\mathcal{K}(G)$ of connected induced subgraphs of G.*

Proof. The defining condition in (2.4) is equivalent to saying that the restriction of G to any node set containing $\{i, j\}$ and contained in $\{i, j\} \cup ([n]\backslash C)$ is disconnected. With this observation, Theorem 2.5.1 follows directly from Proposition 2.4.9. □

The polytope $\Delta_G = \Delta_{\mathcal{K}(G)}$ associated with the graph G is the *graph associahedron*. This is a well-studied object in combinatorics [55, 14]. Carr and Devadoss [14] showed that Δ_G is a simple polytope whose faces are in bijection with the tubings of the graph G. Tubings are defined as follows. Two subsets $A, B \subset [n]$ are *compatible* for G if one of the following conditions holds: $A \subset B$, $B \subset A$, or $A \cap B = \emptyset$, and there is no edge between any node in A and B. A *tubing* of the graph G is a subset $\mathbf{T}$ of $2^{[n]}$ such that any two elements of $\mathbf{T}$ are compatible. The set of all tubings on G is a simplicial complex; it is dual to the face lattice of the simple polytope Δ_G.

For any graph G on $[n]$ we now have two convex rank tests. First, there is the *graphical model rank test* $\tau_{\mathcal{K}(G)}$, which is the MSS test of the set family $\mathcal{K}(G)$. Second, we have the *graphical tubing rank test* $\tau^*_{\mathcal{K}(G)}$, which is the convex rank test associated with the semigraphoid $(\mathcal{M}^G)^*$ dual to $\mathcal{M}^G$. Explicitly, that dual semigraphoid is given by

$$(\mathcal{M}^G)^* \;=\; \big\{\, i \perp\!\!\!\perp j \,|\, C \;:\text{ the restriction of } G \text{ to } C\cup\{i,j\} \text{ contains no path from } i \text{ to } j \big\}. \tag{2.5}$$

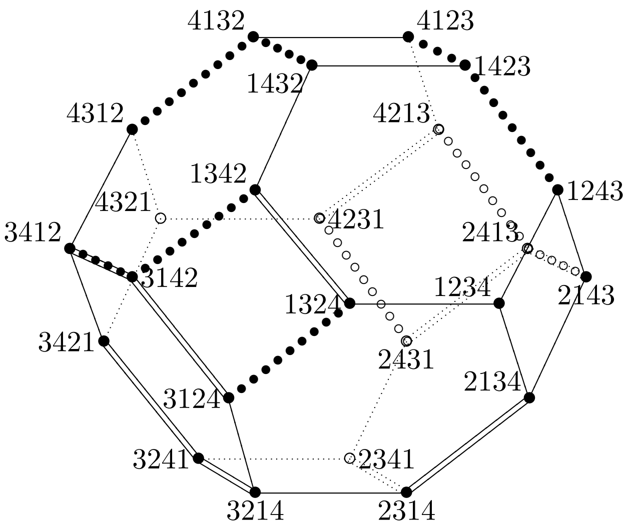


Figure 2.5: The permutohedron $\mathbf{P}_4$. Double edges indicate the MSS test $\tau_{\mathcal{K}(G)}$ where G is the 4-chain. Edges with large dots indicate the dual tubing test $\tau^*_{\mathcal{K}(G)}$.

We summarize our discussion in the following theorem:

Theorem 2.5.2. *The following four combinatorial objects are isomorphic for any graph G on $[n]$:*

- *the graphical model rank test $\tau_{\mathcal{K}(G)}$,*
- *the graphical tubing rank test $\tau^*_{\mathcal{K}(G)}$,*
- *the fan of the graph associahedron Δ_G,*
- *the simplicial complex of all tubings on G.*

We note that when the graph G is a path of length n, then Δ_G is the *associahedron*, and when it is an n-cycle, then Δ_G is the *cyclohedron.* The number of classes in either the MSS test $\tau_{\mathcal{K}(G)}$ or the tubing test $\tau^*_{\mathcal{K}(G)}$ is the *G-Catalan number* of [55]. This number equals the classical Catalan number $\frac{1}{n+1}\binom{2n}{n}$ for the associahedron test and it equals $\binom{2n-2}{n-1}$ for the cyclohedron test.

Example 2.5.3. Let $n = 4$ and let G be the 4-chain 1—2—3—4. Then

$$\begin{aligned}
\mathcal{M}^G &= \{1 \perp\!\!\!\perp 3|24,\ 1 \perp\!\!\!\perp 4|23,\ 2 \perp\!\!\!\perp 4|13,\ 1 \perp\!\!\!\perp 3|2,\ 1 \perp\!\!\!\perp 4|2,\ 1 \perp\!\!\!\perp 4|3,\ 2 \perp\!\!\!\perp 4|3\}, \\
(\mathcal{M}^G)^* &= \{1 \perp\!\!\!\perp 3,\quad 1 \perp\!\!\!\perp 4,\quad 2 \perp\!\!\!\perp 4,\quad 1 \perp\!\!\!\perp 3|4,\ 1 \perp\!\!\!\perp 4|3,\ 1 \perp\!\!\!\perp 4|2,\ 2 \perp\!\!\!\perp 4|1\}.
\end{aligned}$$

The corresponding tests $\tau_{\mathcal{K}(G)}$ and $\tau^*_{\mathcal{K}(G)}$ are depicted in Figure 2.5. Note that contracting either class of marked edges on the permutohedron in Figure 2.5 leads to the 3-dimensional associahedron Δ_G. The associahedron Δ_G is the Minkowski sum of the simplices Δ_K where K runs over

$$\mathcal{K}(G) \quad = \quad \big\{\{1\},\{2\},\{3\},\{4\},\{1,2\},\{2,3\},\{3,4\},\{1,2,3\},\{2,3,4\},\{1,2,3,4\}\big\}.$$

The 3-dimensional simple polytope Δ_4 has 14 vertices, one for each of the 14 tubings of G. □

In our application of graphical rank tests, we found it more natural to work with the tubing test $\tau^*_{\mathcal{K}(G)}$ instead of the MSS test $\tau_{\mathcal{K}(G)}$. Chapter 3 gives a detailed discussion of the cyclohedron test and its applications. By the cyclohedron test we mean the tubing test $\tau^*_{\mathcal{K}(G)}$ where the graph G is a cycle of length n.

Applying the tubing test to a data vector $u \in \mathbb{R}^n$ can be viewed as an iterative procedure for drawing a topographic map on the graph G. Namely, we encircle the vertices of G by sets $U_1, \ldots, U_n$ in the order $\delta_1, \delta_2, \ldots, \delta_{n-1}$, with the following provision: if δ_i is next to be encircled and shares an edge with some vertex j which has already been encircled by some U_j, then U_i must also contain the circle U_j. The result is a collection U of $n-1$ encircled sets $U_1, U_2, \ldots, U_{n-1}$, and this unordered collection of sets is the signature of v.

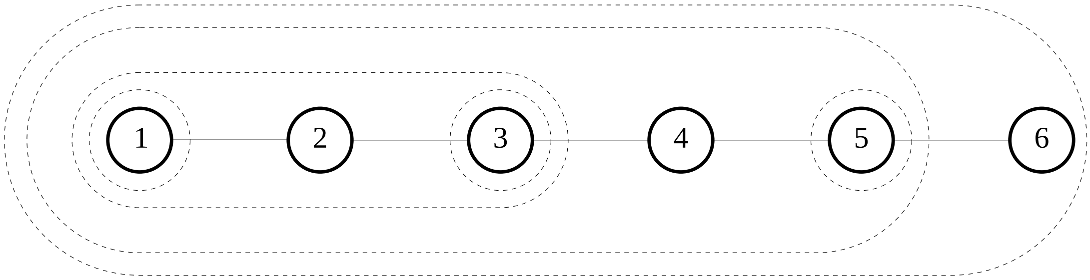


Figure 2.6: Tubing of the 6-chain. Encircled regions indicate the sets U_j.

The height h_i of the i-th node in the topographic map for v is the number of sets U_j which contain i. We can identify the signature U with the *height vector* $h = (h_1, h_2, \ldots, h_n)$, since U can be recovered uniquely from the vector h. The map $u \mapsto h(u)$ can be interpreted as a *smoothing of the data*. Figure 6 displays the topographic map when the data vector is $u = (2.1, 0.3, 1.8, , 2.0, 1.1, 0.1)$. Here G is the 6-chain 1—2—3—4—5—6. and the descent vector of u equals $\delta = (1|5|3|2|4|6)$. Because of the interpretation of the tubing as a topographic map, we call the statistical models derived from this procedure *topographical models* . We will have more to say about these in the next chapter.

2.6 On counting linear extensions

In this chapter, we have introduced a hierarchy of rank tests, which range from pre-convex to graphical. Convex rank tests are applied to data vectors $u \in \mathbb{R}^n$, or permutations $\pi \in S_n$, and determine their cones in a fan $\mathcal{F}$ which coarsens the S_n-fan. The significance of a data vector in such a test is measured by a certain p-value, whose precise derivation is described in Chapter 3. Computation of that p-value rests on our ability to compute the quantity $|\tau^{-1}(\tau(\pi))|$, which is the number of permutations in the maximal cone of $\mathcal{F}$ corresponding to π. Recall that the cones of a convex rank test are indexed by posets

$P_1, P_2, \ldots, P_k$ on $[n]$, and our computations amount to finding the cardinality of the set $\mathcal{L}(P_i)$ of linear extensions of P_i.

The problem of computing linear extensions of general posets is #P-complete [10], so our task is an intractable problem when n grows large. However, for special classes of posets, and for moderate values of n, the situation is not so bad. For example, in the up-down analysis of Willbrand *et.al* (see Example 2.3.4), we need to count all permutations with a fixed descent set, a task for which an explicit determinantal formula appears in Stanley [63, page 69]. We refer to [11] for a detailed study of the combinatorics of these *up-down numbers*.

Likewise, there is an efficient method for computing the quantities $|\,\tau^{-1}\big(\tau(\pi)\big)\,|$ for any graphical graphical tubing test $\tau^*_{\mathcal{K}(G)}$, as defined in Section 2.5. Indeed, here the fan $\mathcal{F}$ is unimodular, and hence the posets P_i are all trees. The special trees arising from a graph G in this manner are known as *G-trees* [55, 14]. The G-tree of a permutation π is a representation of the poset P_i as a tree $T = \tau^*_{\mathcal{K}(G)}(\pi)$ with the minimum value as the root and maximal values as the leaves. Suppose the root of the tree T has k children, each of which is a root of a subtree T^i for $i = 1, \ldots, k$. Writing $|T^i|$ for the number of nodes in T^i, we have

$$|\,\tau^{-1}(T)\,| \quad = \quad \binom{\sum_{i=1}^{k}|T^i|}{|T^1|, \ldots, |T^k|}\left(\prod_{i=1}^{k}|\tau^{-1}(T^i)|\right).$$

This recursive formula translates into an efficient iterative algorithm. Our implementation of this algorithm, when G is the n-cycle, is the workhorse behind our computations in Chapter 3. For a graph G, let nbhd(i) be the set of vertices j such that there is an edge (i, j) in G.

Algorithm 2.6.1. (Permutation Counting)

Input: A data point u as a descent permutation δ and a graph G.

Output: The number of permutations with the same signature as δ, $|\,\tau^{-1}\tau(\pi(u))\,|$.

> **Initialize:**
> An indexed set of largest enclosing sets $LE_1 = \cdots = LE_n = \emptyset$, and counter $c = 1$
> **for** δ_i in δ:
>   Initialize ℓ an empty list of enclosed tree lengths
> $LE_{\delta_i} = \{\delta_i\}$
>   **for** j in nbhd(δ_i):
>   **if** $LE_j \neq \emptyset$ and $j \notin LE_{\delta_i}$:
> $LE_{\delta_i} = LE_{\delta_i} \sqcup LE_j$
>   append $|LE_j|$ to ℓ
> $c = c \cdot \binom{\sum_i (\ell_i)}{\ell}$
>   **for** j in LE_{δ_i}:
> $LE_j = LE_{\delta_i}$
> **Return** the permutation count c

Actual `Python` code implementing this algorithm is (assuming a function `multichoose`, and indexing from zero):

```
def count(delta, nbhd):
    n = len(delta);  LE=[[]]*n; c=1
    for delta_i in delta[:n-1]:
        li=[]
        LE[delta_i] = [delta_i]
        for j in nbhd[delta_i]:
            if j not in LE[delta_i] and LE[j]:
                LE[delta_i] = LE[delta_i]+LE[j]
                li.append(len(LE[j]))
        c = c * multichoose(sum(li),li)
        for j in LE[delta_i]:
            LE[j]=LE[delta_i]
    return c
```

In the remainder of this section we discuss our method for performing these computations for an arbitrary convex rank test. The test is specified (implicitly or explicitly) by a collection of posets $P_1, \ldots, P_k$ on $[n]$. From the given permutation, we identify the unique poset P_i of which that permutation is a linear extension, and we construct the *distributive lattice* $L(P_i)$ whose elements are the order ideals of P_i. Recall that an *order ideal* of P_i is a subset O of $[n]$ such that if $l \in O$ and $(k, l) \in P_i$ then $k \in O$. The set of all order ideals is a distributive lattice with meet and join operations given by set intersection $O \cap O'$ and set union $O \cup O'$.

The distributive lattice $L(P_i)$ is a sublattice of the Boolean lattice $2^{[n]}$, whose nodes are the 2^n subsets of $[n] = \{1, 2, \ldots, n\}$, and we represent $L(P_i)$ by its nodes and edges (cover relations) in $2^{[n]}$. We write each edge in $2^{[n]}$ as a pair (K, l) where $K \subset [n]$ and $l \in [n] \backslash K$. The edge in the Boolean lattice $2^{[n]}$ represented by the pair (K, l) is the cover relation $K \subset K \cup \{l\}$.

Permutations in S_n are in natural bijection with maximal chains in the Boolean lattice $2^{[n]}$. For example, the descent permutation $\delta = (4|2|1|3)$ corresponds to the maximal chain $(\emptyset, \{4\}, \{2, 4\}, \{1, 2, 4\}, \{1, 2, 3, 4\})$ in the Boolean lattice $2^{[4]}$. If the poset P_i is the linear order δ then $L(P_i)$ is the subgraph of $2^{[4]}$ consisting of the five nodes in the chain and the four edges $(\emptyset, 4)$, $(\{4\}, 2)$, $(\{2, 4\}, 1)$ and $(\{1, 2, 4\}, 3)$ which connect them. The maximal chains in $2^{[n]}$ that lie in the sublattice $L(P_i)$ are precisely the permutations that are linear extensions of P_i. Therefore our task is to construct $L(P_i)$ and then count its maximal chains.

Remark 2.6.2. The linear extensions of the poset P_i are in bijection with the maximal chains in the distributive lattice $L(P_i)$. See [63, Section 3.5] for further information on this bijection.

In general, $L(P_i)$ is the graph whose nodes are those subsets of $[n]$ which are order ideals in P_i, and the edges are (K, l) where both K and $K \cup \{l\}$ are order ideals in P_i.

Our strategy in computing the graph which represents $L(P_i)$ is as follows. We start with a given permutation δ which lies in the class indexed by P_i. That permutation determines a maximal chain in $2^{[n]}$ which must lie in $L(P_i)$. We then compute a certain closure of that subgraph in $2^{[n]}$ with respect to the semigraphoid $\mathcal{M}$ under consideration. This is precisely what is done in Algorithm 2.6.3 below. Knowledge of the distributive lattice $L(P_i)$ solves our problem since the number of maximal chains of $L(P_i)$ can be read off easily from the representation of $L(P_i)$ in terms of nodes and edges.

Algorithm 2.6.3. (Building the Distributive Lattice)

Input: A data point as a descent permutation δ and a semigraphoid $\mathcal{M}$.

Output: A distributive lattice $L(P_i)$ representing the class of δ in the convex rank test $\mathcal{M}$.

Initialize:
A set of confirmed lattice nodes, $\mathbb{H} = \big\{\emptyset, \{\delta_1\}, \{\delta_1, \delta_2\}, \ldots, \{\delta_1, \ldots, \delta_n\}\big\}$
A set of checked lattice edges, $E = \big\{(\{\delta_1, \ldots, \delta_{n-1}\}, \delta_n)\big\}$,
where each pair has the form (history, next position).
A stack of edges waiting to be checked:
$W = \big[(\emptyset, \delta_1), (\{\delta_1\}, \delta_2), (\{\delta_1, \delta_2\}, \delta_3), \ldots, (\{\delta_1, \ldots, \delta_{n-2}\}, \delta_{n-1})\big]$

While $W \neq \emptyset$:
  Pop (H, i) from the stack W
  Add (H, i) to E
  for j such that $(H \cup \{i\}, j) \in E$:
    if $i \perp\!\!\!\perp j | H \in \mathcal{M}$:
      Add (H, j) to E
      if $H \cup \{j\} \notin \mathbb{H}$:
        Add $H \cup \{j\}$ to $\mathbb{H}$
        Push $(H \cup \{j\}, i)$ onto W
Return the distributive lattice $L(P_i) = \big(\mathbb{H}, E\big)$

Using a program based on Algorithm 2.6.3, computing the number of linear extensions of the Boolean poset $P = 2^{[5]}$ (consisting of all subsets of $\{1, 2, 3, 4, 5\}$) took less than one second on a laptop and found that

$$|L(2^{[5]})| \quad = \quad 14,807,804,035,657,359,360.$$

This computation was inspired by work in population genetics by Weinreich [71] who reports the same calculation for $P = 2^{[4]}$.

Chapter 3

The cyclohedron test

3.1 Introduction

The material in this chapter is based on the paper "The cyclohedron test for finding periodic genes in time course expression studies" [49] which was authored jointly with L. Pachter, A. Shiu, and B. Sturmfels. The implementation described in Section 3.7 and the application in Section 3.8 are new. In this chapter, we will apply one of the topographical model tests first introduced in Section 2.5 to biologically-derived time series. The search for the molecular components of biological clocks is an important first step towards understanding the regulatory mechanisms underlying periodic behavior at the molecular level. Examples of clocks that have been studied include the circadian clock [47], the respiratory cycle clock in yeast [38, 62] and the segmentation clock in vertebrates [57]. In order to find clock-related genes in a high-throughput fashion, time course array experiments are performed to measure the expression levels of genes on a genome-wide scale. This is followed by a statistical analysis to find periodically expressed genes. The analysis is non-trivial for reasons that include noisy measurements, variable times between experiments, vague notions of periodicity, and loss of power due to multiple testing.

The question of how best to analyze cyclic time series is a topic of extensive research in statistics [15]. Recent approaches, proposed in the context of microarray analysis include splines and other curve approximations [40, 64], methods based on signal processing techniques such as the Lomb-Scargle test [30], and non-parametric rank tests [72]. Several of these approaches to pattern detection, including the cyclohedron test, are compared in Dequéant et al. [20].

The cyclohedron C_n is the graph associahedron when the graph G is the n-cycle, and the cyclohedron test is the greedy method for linear programming on C_n. It is equivalent to the test denoted by $\tau^*_{\mathcal{K}(G)}$ in Chapter 2. Cyclohedra are also known as *Bott-Taubes polytopes*, and they play an important role in representation theory [25, Section 3.2], combinatorics [37, 59], and homotopy theory [41].

When using rank tests, an expression time-course is represented by a permutation. This has the advantage of providing robustness to noise, monotonic transformations, and uncertainty with respect to the underlying probability distributions, and the disadvantage of precluding a parametric analysis of the untransformed time courses.

The cyclohedron test is explained in detail in Section 3.2. In Section 3.3 we present a method for assigning p-values to top-ranked groups of genes. This is done within a multiple hypothesis testing framework, which is compatible with any rank test for permutation data, including up-down analysis (Example 2.3.4). In Sections 3.5 and 3.6 we develop the combinatorial details and efficient algorithms for the cyclohedron test. Our `R` and `Python` code is available online, and its use is described in Section 3.7.

We apply the cyclohedron test to data reported in [21], consisting of 17 distinct expression array experiments from the presomitic mesoderm tissue of mouse embryos. These data were chosen because of the analyses already undertaken and the possibility for biological validation. Results are discussed in Section 3.4. We find that although the high-throughput array experiments are effective for finding groups of genes likely to be involved with clock regulation, multiple testing issues preclude the assignment of significance to any individual gene on the basis of periodic-looking patterns alone.

3.2 The cyclohedron test

The cyclohedron test is appropriate when seeking to determine whether a time course expression is periodic. Within a single hypothesis setting, the null hypothesis states that a gene or other unit of interest does not exhibit cyclic expression. As with other convex rank tests, the cyclohedron test provides the *permutation count* as a test statistic, replacing this vague null hypothesis. The test applies to data vectors $v = (v_1, \ldots, v_n)$ whose coordinates are distinct real numbers. The coordinates v_i are measurements of the same quantity at distinct points. In our applications, the ordering of each vector should be with respect to some 'cyclic' time, so that any $v' = (v_i, v_{i+1}, \ldots, v_n, v_1, \ldots, v_{i-1})$ is an equally meaningful ordering. For example, the data vectors v we analyze in Section 3.4 are ordered within a somite-formation cycle; so v_j is a measurement taken before v_{j+1} in the cycle, where $j + 1$ is understood $\mod n$.

The following specialization of Algorithm 2.6.1 computes, for any given data vector v, its permutation signature (maximal tubing) $\sigma(v)$ and its *permutation count* $\mathbf{c}(v)$. The signature is an unordered set $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_{n-1}\}$ of subsets of $\{1, 2, \ldots, n\}$ and the permutation count is a positive integer.

Algorithm 3.2.1. (Cyclohedron test)

Input: A vector $v = (v_1, \ldots, v_n)$ of distinct real numbers.

Output: The signature $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_{n-1}\}$ and the permutation count $\mathbf{c}$ for v.

Initialize $\mathbf{c} := 1$.

For i from 1 to $n-1$, do

Initialize $\sigma_i = \emptyset$, the empty set.

Let δ_i be the unique index such that v_{δ_i} is the i-th largest coordinate of v.

Initialize Left $:= \emptyset$ and Right $:= \emptyset$.

For k from 1 to $i-1$, do

if σ_k contains $\delta_i - 1$ (modulo n) then set Left $:= \sigma_k$,

if σ_k contains $\delta_i + 1$ (modulo n) then set Right $:= \sigma_k$.

Set $\sigma_i := \{\delta_i\} \cup \text{Right} \cup \text{Left}$ and $\mathbf{c} := \mathbf{c} \cdot \binom{|\text{Right}|+|\text{Left}|}{|\text{Right}|}$.

Let $\mathcal{C}_n$ denote the set of all signatures $\sigma(v)$ as v runs over $\mathbb{R}^n$. Algorithm 3.2.1 constructs not only the signature σ and the permutation count $\mathbf{c}$ but also the descent vector $\delta = (\delta_1, \delta_2, \ldots, \delta_n)$ of the data vector $v = (v_1, v_2, \ldots, v_n)$. Since $\sigma(v)$ depends only on the descent vector δ, our algorithm specifies a map $\delta \mapsto \sigma$ from the symmetric group S_n onto the set $\mathcal{C}_n$. For $n \geq 4$, this map is not injective, and we are interested in the cardinalities of the preimages. For instance, the permutations $\delta = (1, 3, 2, 4)$ and $\delta' = (3, 1, 2, 4)$ have the same signature $\sigma(\delta) = \sigma(\delta') = \big\{\{1\}, \{3\}, \{1, 2, 3\}\big\}$.

The test statistic, the *permutation count* $\mathbf{c}$, is the number of permutations having the same signature as the permutation of interest. Significant data vectors have small test statistics, because it is unlikely that a random permutation will have a topographical map shared by few permutations. The permutation count $\mathbf{c} = \mathbf{c}(v)$ has the following interpretation. Suppose that under the null hypothesis, the data generating distribution for each data vector v induces the uniform distribution on all descent order permutations δ in the symmetric group S_n. Note that this assumption is valid if the coordinates of the data vector are independent and identically distributed under the null distribution, so our test is therefore broadly applicable. For each signature $\sigma \in \mathcal{C}_n$, let $p(\sigma)$ denote the probability that the signature σ would be observed under such a null distribution. The following proposition states that $p(\sigma)$ is the fraction of permutations δ that map to σ.

Proposition 3.2.2. *The permutation count* $\mathbf{c}$ *computed by Algorithm 3.2.1 depends only on the signature* σ. *It equals the number of permutations* δ *that are mapped to* σ, *and hence*

$$\mathbf{c} \; = \; \mathbf{c}(\sigma) \quad = \quad p(\sigma) \cdot n!.$$

Proof. For each σ_i in the signature σ, at most two other sets σ_j and σ_k are contained in σ_i and are maximal with this property. Here σ_j and σ_k are necessarily disjoint. The permutation count $\mathbf{c}$ is the product of the corresponding binomial coefficients $\binom{|\sigma_j \cup \sigma_k|}{|\sigma_j|}$. It depends only on σ. The second statement is proved by induction on n, using the fact that any valid permutation of σ_j can be shuffled with any valid permutation of σ_k, and augmented by δ_i, to get a valid permutation for σ_i. Carrying out this process until $i = n$, with $\sigma_n = \{1, 2, \ldots, n\}$, yields precisely all permutations δ that have signature σ. □

As discussed in §2.5 for general graphs G, the signature $\sigma(v)$, can be viewed as a topographic map on the n-cycle that captures the shape of the data v. Algorithm 3.2.1 is an iterative procedure for drawing this topographic map in the cyclic case. Namely, we encircle the vertices of the n-cycle in decreasing order of their corresponding data vector coordinates, that is, in the order $\delta_1, \delta_2, \ldots, \delta_{n-1}$. (The first circle is the set σ_1, the second is σ_2, and so on.) We do this according to the following provision: in order to encircle δ_i, if it is adjacent to some vertex j which has already been encircled by some σ_k, then σ_i must contain the σ_k circle. Accordingly, the sets "Left" and "Right" keep track of how far to the left and right σ_i must extend for the result to form a valid tubing. The resulting tubing is an unordered set σ of $n-1$ encircled sets $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$. Figure 3.1 displays the beginning of an example of this encircling process for $n = 11$.

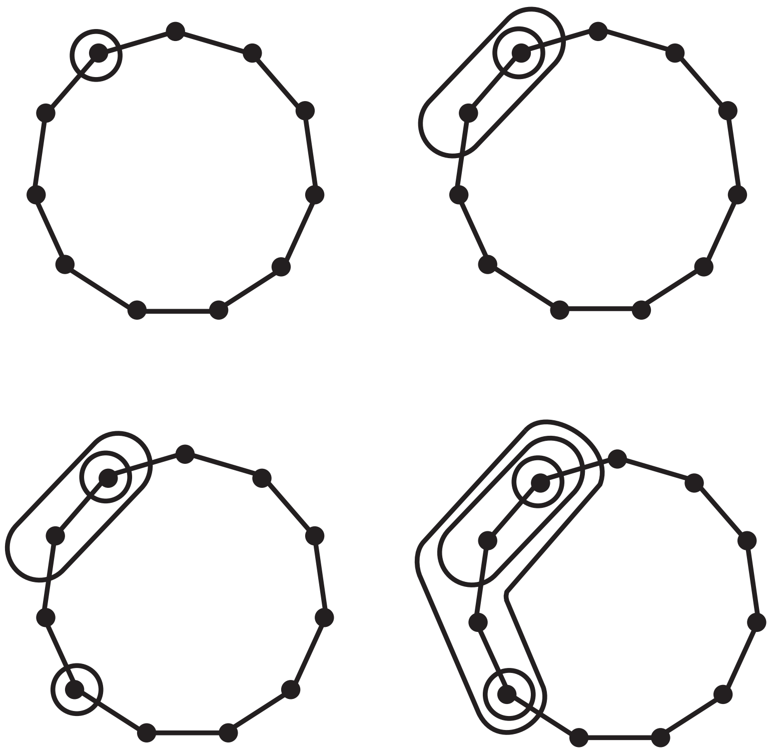


Figure 3.1: Algorithm 3.2.1 constructs a topographic map on the n-cycle by subsequently encircling vertices in order of decreasing size of the corresponding component of a data vector. Displayed at the top are the formations of the first two components σ_i, and at the bottom are the third and fourth, of the signature for an example with $n = 11$.

We say that the height h_i of the i-th vertex in the topographic map for v is the number of sets σ_j which contain i. We can identify the signature σ with the *height vector* $h = (h_1, h_2, \ldots, h_n)$, because σ can be recovered uniquely from the vector h. The map $v \mapsto h(v)$ can be viewed as a *smoothing of the data*; see Figure 3.2.

Remark 3.2.3. The cyclohedron test applies when there are no ties $v_i = v_j$ in the data.

When ties occur, we examine all possible permutations δ arising from small perturbations.

Example 3.2.4. In our analysis in Section 3.4, the number of microarray experiments is $n = 17$, and the number of probesets (labels of the data vectors) is $N = 13,873$. The probeset ranked first in Table 3.1 represents a gene named `Obox`. Its data vector equals

$$\begin{aligned} v \;=\; & \big(0.738, 0.996, 0.705, 0.150, -0.566, -0.673, 0.774, -0.736, -0.788, \\ & -0.802, -1.276, -0.521, 0.238, -0.258, -0.249, -0.084, -0.117\big). \end{aligned}$$

The descent order permutation for this vector v equals

$$\delta \;=\; \big(2, 7, 1, 3, 13, 4, 16, 17, 15, 14, 12, 5, 6, 8, 9, 10, 11\big).$$

The signature σ is given by the unordered set $\sigma_1 = \{2\}$, $\sigma_2 = \{7\}$, $\sigma_3 = \{1, 2\}$, $\sigma_4 = \{1, 2, 3\}$, etc. The permutation count $\mathbf{c} = 480$ is the product of the three contributions made by 5, 8 and 12, respectively, when constructing $\sigma_8 = \{1, 2, 3, 4, 16, 17\}$, $\sigma_{10} = \{1, 2, 3, 4, 13, 14, 15, 16, 17\}$, and $\sigma_{13} = \{1, 2, 3, 4, 5, 6, 7, 12, 13, 14, 15, 16, 17\}$. When viewing σ as a topographic map for the data v, we obtain the height vector

$$h(v) \;=\; \big(12, 13, 11, 10, 5, 4, 5, 3, 2, 1, 0, 6, 8, 7, 8, 10, 9\big).$$

Figure 3.2 displays the data v and the height vector $h(v)$ plotted around the circle. $\square$

3.3 Significance testing

Multiple hypothesis testing is of concern in microarray experiments, because the number of hypotheses that are tested simultaneously is large. In our application, there are $N = 13,873$ null hypotheses. The hypotheses take the form "the r genes with the smallest

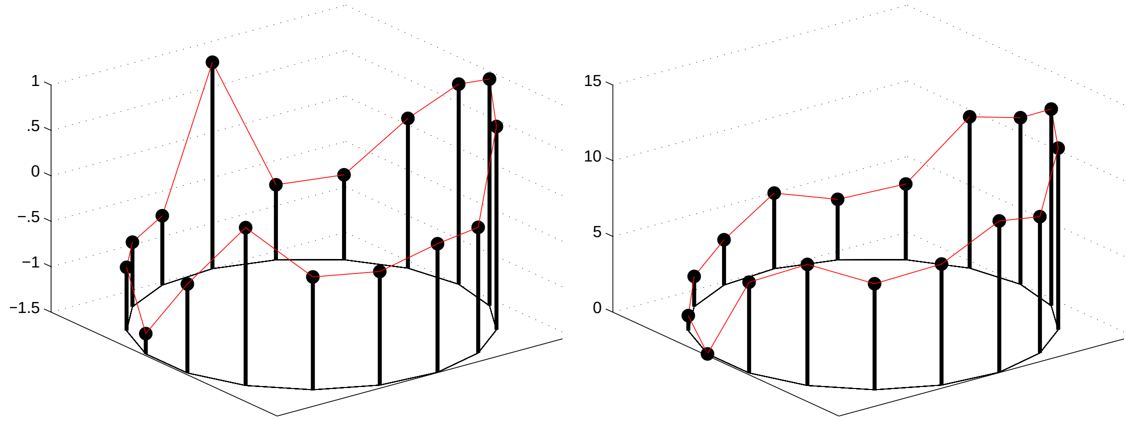


Figure 3.2: The data v (left) and the height vector $h(v)$ (right) for the gene `Obox`.

counts $\mathbf{c}$ arose by chance", for $r = 1, 2, \ldots, N$. In this section, we explain how to assign p-values to these groups, leading to a criterion for determining which hypotheses to reject.

Applying the cyclohedron test to N data vectors $v^{(1)}, \ldots, v^{(N)}$ in $\mathbb{R}^n$ means computing their permutation counts $\mathbf{c}(v^{(1)}), \ldots, \mathbf{c}(v^{(N)})$. The highest ranked data are those for which $\mathbf{c}(v^{(i)})$ is smallest. Under the null hypothesis, the probability distribution on $\mathbb{R}^n$ of each data vector $v^{(i)}$ induces the uniform distribution U on the $n!$ permutations δ. Viewed as a random variable, the permutation count $\mathbf{c}$ has probability distribution function

$$P_{\mathbf{c}} : \mathrm{im}(\mathbf{c}) \to [0,1], \quad \gamma \mapsto \Pr_U(\mathbf{c}(\delta) = \gamma). \tag{3.1}$$

Here $\mathrm{im}(\mathbf{c}) = \{\gamma_1 < \gamma_2 < \cdots < \gamma_{s_n}\}$ is the set of all positive integers that arise as permutation counts $\mathbf{c}(\sigma)$ for some $\sigma \in \mathcal{C}_n$. The probability distribution function $P_{\mathbf{c}}$ is displayed in Figure 3.3 for $n = 17$, which will be the number of time points in Section 3.4.

Figure 3.3: The probability distribution function of $\mathbf{c}$ for $n = 17$.

We now fix two integers $1 \le r \le N$. The *order statistic* $\mathbf{C}_{(r)}$ is the function $\mathbb{R}^{N\times n} \to \mathrm{im}(\mathbf{c})$ which takes any list of N data vectors $V = (v^{(1)}, \ldots, v^{(N)})$ and returns the r^{th} smallest value among the permutation counts $\mathbf{c}(v^{(1)}), \ldots, \mathbf{c}(v^{(N)})$. Recall that under the null hypothesis, each $v^{(i)}$ has a distribution on $\mathbb{R}^n$ which induces the uniform distribution on permutations δ. Further let us assume that the data vectors $v^{(i)}$ are independent. This induces a joint distribution Q_0 on the vector $\big(\,\mathbf{c}(v^{(1)}), \ldots, \mathbf{c}(v^{(N)})\,\big) \in \mathbb{R}^{N\times n}$ of counts. In this framework, we view the order statistic $\mathbf{C}_{(r)}$ as a random variable with distribution

$$F_{(r)} : \mathrm{im}(\mathbf{c}) \to [0,1],\, \gamma \mapsto \Pr_{Q_0}(\mathbf{C}_{(r)} = \gamma).$$

In other words, $F_{(r)}(\gamma)$ is the probability that the r-th smallest value among the permutation counts $\mathbf{c}(v^{(1)}), \ldots, \mathbf{c}(v^{(N)})$ of N random data vectors equals γ. The function $F_{(r)}(\gamma)$ depends only on n, N and r. Its efficient computation is explained in Section 3.6.

Definition 3.3.1. (p-value) Suppose we apply the cyclohedron test to N data vectors in $\mathbb{R}^n$, and the data vector whose permutation count is the r-th smallest has permutation count γ_k. Then the collective *p-value* of the group of r highest ranked data vectors is

$$\Pr_{Q_0}(\mathbf{C}_{(r)} \le \gamma_k) \quad = \quad F_{(r)}(\gamma_1) + F_{(r)}(\gamma_2) + \cdots + F_{(r)}(\gamma_k). \tag{3.2}$$

The p-value (3.2) is the probability that the r-th order statistic for random data under the null would be less or equal to the value of the r-th order statistic for the observed data.

3.4 Application to mouse microarray data

We applied the cyclohedron test to microarray data from recent work that investigated the mouse segmentation clock [21]. Dequéant *et al.* took 17 expression measurements from mouse presomitic mesoderm on Affymetrix MOE430A arrays. By independently measuring the expression of the gene Lunatic Fringe (`Lfng`) which is known to be periodic within the somitogenesis cycle of embryonic development, Dequéant *et al.* ordered the 17 experiments within the cycle. Each array consisted of over $22,000$ probesets, however we restrict the analysis to a subset of $13,873$ probesets by removing genes whose expressions are deemed "absent" across the experiments by Affymetrix standards (which also employ rank tests). In other words, the data consisted of 13,873 data vectors v, each of which was the expression level of one gene (divided by the mean across experiments and transformed to $\log_2$). We then applied the cyclohedron test to these data. We were interested in those genes whose counts $\mathbf{c}(v)$ were small. Accordingly, we ranked the genes by their counts; Table 3.1 presents the first 32 genes. Table 3.2 lists the significance of top groups of genes. For example, the first 32 genes collectively have a p-value of 0.081, which suggests that these 32 genes are of interest. At this point we recall the definition of a p-value which was given in equation (3.2). The p-value of the rank-1 gene `Obox1` is the probability under the null hypothesis that the top-ranked permutation count is less than or equal to 480, while the p-value of the first 16 genes (the number 0.008 in Table 2) is the probability that the gene ranked 16 has permutation count less than or equal to 4928. It is important to emphasize that the p-values do not reveal the significance of any individual gene, but rather of a collection of genes. For example, the top 19 genes having a collective p-value of 0.046 means this: the probability that the first 19 genes would collectively all have permutation count at most 6825 under the null distribution is 0.046. In other words, the group as a whole is significant. However, we determine whether any individual gene in that group is significant. For example, there is no significance to the fact that `Obox1` is ranked first. While it appears to be the most periodic pattern in the data by our analysis, that could have happened

Rank	ProbeSet	Gene Name	Gene Description	Count
1	1456017_x	Obox1	similar to oocyte specific gene	480
2	1452041	Klhl26	kelch-like 26 (Drosophila)	1440
3	1418593	Taf6	TAF6 RNA polymerase II	1560
4	1417985	Nrarp	Notch-regulated ankyrin repeat protein	1950
5	1436845	Axin2	axin2	2240
5	1436343	Chd4	chromodomain helicase DNA binding prot.	2240
7	1426267	Zbtb8os	zinc finger and BTB domain	2310
8	1420360	Dkk1	dickkopf homolog 1 (Xenopus laevis)	2520
9	1449643_s	Btf3	basic transcription factor 3	2772
10	1417399	Gas6	growth arrest specific 6	2800
11	1418102	Hes1	hairy and enhancer of split 1 (Drosophila)	3120
12	1448799_s	Mrps12	mitochondrial ribosomal protein S12	3150
13	1418729	Star	steroidogenic acute regulatory protein	3600
14	1425424	MGC7817	hypothetical protein LOC620031	3850
15	1455740	Hnrpa1	heterogeneous nuclear ribonucleoprotein	4004
16	1450204_a	Mynn	myoneurin	4928
17	1449120_a	Pcm1	pericentriolar material 1	6006
18	1423106	Ube2b	ubiquitin-conjugating enzyme E2B	6720
19	1420386	Seh1l	SEH1-like (S. cerevisiae)	6825
20	1456380_x	Cnn3	calponin 3, acidic	8008
21	1419438	Sim2	single-minded homolog 2 (Drosophila)	8640
22	1426524	Gnpda2	glucosamine-6-phosphate deaminase 2	9009
23	1438557_x	Dnpep	aspartyl aminopeptidase	9450
24	1454904	Mtm1	X-linked myotubular myopathy gene 1	10500
25	1448951	Tnfrsf1b	tumor necrosis factor receptor superfamily	10530
25	1433952	Tufm	Tu translation elongation factor	10530
27	1422327_s	*G6pd2/ G6pdx*	glucose-6-phosphate dehydrogenase 2	10725
28	1416295_a	Il2rg	interleukin 2 receptor, gamma chain	10920
29	1417316	Them2	thioesterase superfamily member 2	11025
30	1450242	Tlr5	toll-like receptor 5	11232
31	1449164	Cd68	CD68 antigen	11340
32	1418337	Rpia	ribose 5-phosphate isomerase A	11760
⋮	⋮	⋮	⋮	⋮

Table 3.1: The 32 genes ranked highest by the cyclohedron test. Gene descriptions are derived from those provided by Affymetrix. The suffix "_at" was removed from each ProbeSet ID.

by chance (p-value 0.279). A natural cutoff value is to look at the first 32 genes because collectively they have a p-value of 0.081 (the next ten p-values are between 0.13 and 0.30). Note that analyses of microarray data have this property, that Type-1 errors are all but guaranteed due to the large number of genes (and thus the large number of hypotheses) that are tested. Our computations were performed with the statistical software `R` [58] and the language `Python`, using the implementation described in Section 3.7.

Dequéant *et al.* performed significance testing according to a Lomb-Scargle analysis, and then based on gene expression profile clustering, they identified genes belonging to three pathways Notch/FGF and Wnt that are involved with somitogenesis. There are genes that are deemed interesting by both the analysis of Dequéant *et al.* and the cyclohedron test. For example, `Axin2` is ranked highly by the Lomb-Scargle (rank 6) and the cyclohedron test (rank 5). In addition, `nrarp` (rank 4 according to the cyclohedron test) is ranked poorly by Lomb-Scargle (rank 482), although it belongs to the Notch pathway and its gene expression clusters accordingly. Finally, there are novel genes such as `Obox` (rank 1 by the cyclohedron test, but not known to be related to somitogenesis) that require further investigation. This suggests that to find periodic gene expression, it is beneficial to apply many methods, including Lomb-Scargle, clustering, and the cyclohedron test. Doing so enables us to find genes overlooked by each method, as well as to confirm findings of other tests. In other words, the findings of each method complement those of others by identifying candidate genes for knockout experiments. In Dequéant et al. [20], we compare various methods, including Lomb-Scargle, up-down analysis, and the cyclohedron test, for identifying cyclic genes from this data set.

In conclusion, we remark that, although microarray expression analyses are frequently criticized due to the noise in individual measurements, the massively parallel nature

Group	1..1	1..2	1..3	1..4	1..5	1..6	1..7	1..8
p-value	0.279	0.458	0.244	0.204	0.064	0.064	0.031	0.020
Group	1..9	1..10	1..11	1..12	1..13	1..14	1..15	1..16
p-value	0.014	0.005	0.005	0.002	0.003	0.003	0.002	0.008
Group	1..17	1..18	1..19	1..20	1..21	1..22	1..23	1..24
p-value	0.047	0.069	0.046	0.139	0.165	0.173	0.195	0.312
Group	1..25	1..26	1..27	1..28	1..29	1..30	1..31	1..32
p-value	0.192	0.192	0.168	0.159	0.118	0.096	0.075	0.081

Table 3.2: Significance of top-ranked groups of genes. For example, the first 32 genes have a collective p-value of 0.081.

of the experiments provide the possibility for finding groups of significant genes. Indeed, we confirm this in our analysis of the Dequéant *et al.* experiments [21], in which we are unable to confirm whether any individual gene is statistically significant, yet we can identify a group of genes that collectively are significant. The biological significance of individual genes can be determined by further targeted experimental validation.

3.5 Combinatorics of the cyclohedron test

We now describe the combinatorics and geometry particular to the cyclohedron test. First, the set $\mathcal{C}_n$ of cyclic signatures is in natural bijection with the vertices of a certain convex polytope. The n-cycle has $n(n-1)$ connected induced proper subgraphs, namely, the *cyclic segments* of the form $S = \{i, i+1, \ldots, i+k\}$. Here $k < n-1$, and the indices are understood modulo n. The *cyclohedron vertex* of a data vector $v \in \mathbb{R}^n$ is the vector $\tau(v) \in \mathbb{N}^n$ whose i-th coordinate $\tau(v)_i$ is the number of cyclic segments S containing i such that $v_i = \min\{v_s : s \in S\}$. The *cyclohedron* C_n is the convex hull in $\mathbb{R}^n$ of all the cyclohedron vertices $\tau(v)$ where v ranges over $\mathbb{R}^n$. For $n = 4$ and the data vector $v = (0.49, 5.73, 4.01, 2.67)$, we have $\tau(v) = (6, 1, 2, 3)$, while for $v' = (0.49, 5.73, 2.67, 4.01)$

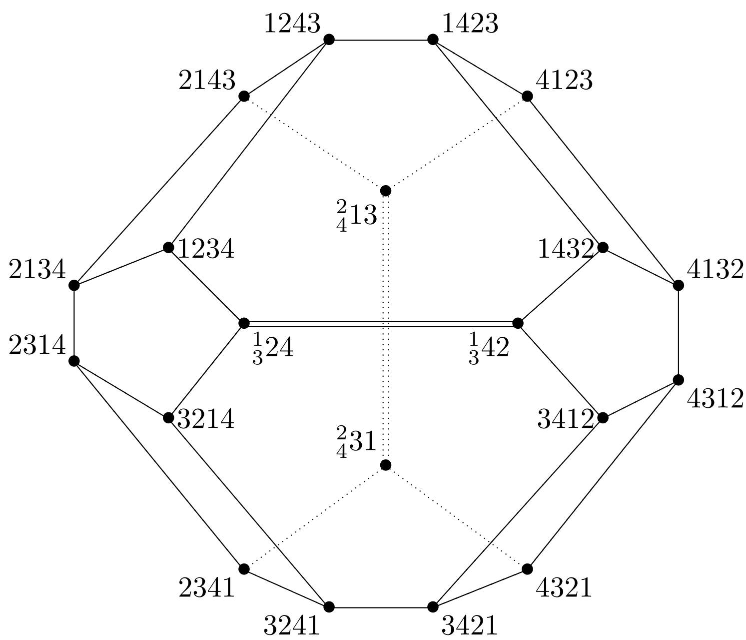


Figure 3.4: The cyclohedron C_4; its vertices correspond to the distinct signatures for $n = 4$. The string $^1_3 24$ labels the cyclohedron vertex of data vectors whose descent permutation is 1324 or 3124.

we have $\tau(v') = (6, 1, 4, 1)$. For example, $\tau(v)_3 = 2$ because $v_3 = 4.01$ is minimal in $S^1 = \{3\}$ and $S^2 = \{2, 3\}$.

Two vectors in $\mathbb{R}^4$ share the same signature $\sigma = \{\sigma_1, \sigma_2, \sigma_3\}$ if and only if they are mapped to the same cyclohedron vertex τ. The convex hull of all cyclohedron vertices $\tau(v)$ is the 3-dimensional cyclohedron C_4. This is a simple polytope with 20 vertices, 30 edges and 12 facets (for the 12 cyclic segments). It is depicted in Figure 3.4. Vertices in the figure, incident to a 'double' edge indicate signatures σ with $\mathbf{c}(\sigma) = 2$. Thus the set $\mathcal{C}_4$ of all signatures has 20 elements, one for each vertex of C_4.

The following theorem summarizes what is known about the cyclohedron. It is extracted from [25, 37, 41].

Theorem 3.5.1. *The cyclohedron C_n is an $(n-1)$-dimensional polytope. It is the solution* *set in $\mathbb{R}^n$ of the following system of one linear equation and $n(n-1)$ linear inequalities:*

$$x_1 + x_2 + \cdots + x_n \;=\; n(n-1), \tag{3.3}$$

$$\sum_{s\in S} x_s \;\geq\; \binom{|S|+1}{2} \quad \text{for each cyclic segment } S. \tag{3.4}$$

The cyclohedron C_n is simple, i.e. each vertex lies on precisely $n-1$ facets. Each inequality (3.4) defines a facet. The total number of vertices equals $\binom{2n-2}{n-1}$. More generally, the number f_i of i-dimensional faces of the cyclohedron is given by the generating function

$$\sum_{i=0}^{n-1} f_i \cdot z^i \;=\; \sum_{k=0}^{n-1} \binom{n-1}{k}^2 \cdot (z+1)^k. \tag{3.5}$$

Algorithm 3.2.1 is a greedy method for linear programming on the cyclohedron C_n. Indeed, computing the cyclohedron vertex $\tau(v)$ of a data vector $v = (v_1, \ldots, v_n)$ is equivalent to the linear program of minimizing $\sum_{i=1}^n v_i x_i$ subject to the constraints (3.3) and (3.4). The optimal vertex of that linear program on C_n is precisely the vector $x = \tau(v)$.

Given the linear functional $\sum_{i=1}^n v_i x_i$ to be minimized, Algorithm 3.2.1 generates a collection $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_{n-1}\}$ of subsets of $\{1, 2, \ldots, n\}$. These sets $S = \sigma_i$ are cyclic segments, and they indicate which $n-1$ inequalities (3.4) are tight at the optimal vertex $x = \tau(v)$ of C_n. This implies that $\tau(v)$ can be recovered from $\sigma(v)$ and vice versa:

Corollary 3.5.2. *The cyclohedron vertex $\tau(v)$ of any data vector $v \in \mathbb{R}^n$ can be obtained from the signature $\sigma(v) = \{\sigma_1, \sigma_2, \ldots, \sigma_{n-1}\}$ by solving the linear system of equations*

$$(3.3) \quad \text{and} \quad \sum_{s\in\sigma_i} x_s \;=\; \binom{|\sigma_i|+1}{2} \quad \text{for } i = 1, 2, \ldots, n-1. \tag{3.6}$$

Conversely, the signature $\sigma(v)$ is recovered from the vertex $\tau(v)$ by substituting $x = \tau(v)$ into the inequalities (3.4) and collecting all index sets S for which equality holds.

In light of Corollary 3.5.2, we henceforth shall identify signatures $\sigma \in \mathcal{C}_n$ with their corresponding vertices τ of the cyclohedron C_n. We note that the solution τ to (3.6) can be read off easily within Algorithm 3.2.1. It always holds that $\tau_{\delta_n} := \binom{n}{2}$, and the other $n-1$ coordinates are obtained by adding one line at the end of the main i loop:

$$\textbf{Output} \;\; \tau_{\delta_i} \; = \; (|\text{Left}| + 1) \cdot (|\text{Right}| + 1).$$

Two data vectors v and v' are *cyclically equivalent* if and only if $\sigma(v) = \sigma(v')$, i.e., if and only if the linear functionals corresponding to v and v' are minimized at the same vertex $\tau(v) = \tau(v')$ of the cyclohedron C_n. The cyclic equivalence classes are the normal cones at the vertices of C_n. They are specified by the inequalities $v_{\delta_i} < v_{\delta_k}$ for all inclusions $\sigma_k \subset \sigma_j$ in $\sigma(v)$. Since C_n is simple, $n-1$ inequalities suffice, and these can be generated by augmenting Algorithm 3.2.1, again at the end of the main i loop, as follows:

$$\text{if Right} \neq \emptyset \;\text{ or }\; \text{Left} \neq \emptyset \;\text{ then } \textbf{output} \;\; v_{\delta_i} < v_{\delta_k}.$$

The generated inequalities permit the study of confidence regions for the cyclohedron test.

Example 3.5.3. Fix $n = 17$ and let v be the data vector for the `Obox` gene in Example 3.2.4. The augmented Algorithm 3.2.1 reveals that the cyclic equivalence class of v is given by

$$v_{11} < v_{10} < v_9 < v_8 < v_6 < v_5 < v_{12} < v_{14} < v_{15} < v_{17} < v_4 < v_3 < v_1 < v_2$$

$$\text{and} \quad v_6 < v_7 \quad \text{and} \quad v_{17} < v_{16} \quad \text{and} \quad v_{14} < v_{13}.$$

These inequalities specify the normal cone at the vertex

$$\tau(v) \; = \; \big(2, 1, 3, 4, 11, 24, 1, 14, 15, 16, 136, 10, 1, 16, 7, 1, 10\big)$$

of the 16-dimensional cyclohedron C_{17}. Recall that the possible signatures for data with

$n = 17$ are (in bijection with) the vertices of C_{17}, and their total number equals

$$|\mathcal{C}_{17}| \;=\; \binom{2\cdot 17-2}{17-1} \;=\; \binom{32}{16} \;=\; 601,080,390.$$

Among all these signatures, the vertex $\tau(v)$ is of interest because the probability that a random linear functional attains its minimum over C_{17} at that vertex is rather small:

$$p(v) \;=\; \mathbf{c}(v)/n! \;=\; 480/17! \;=\; 1.35\cdot 10^{-12}.$$

The results of our analysis for the full data set were presented in Section 3.4. □

3.6 Null distribution of the counts and order statistics

We next compute two probability distribution functions, that of the random variable $\mathbf{c}$ and of its order statistics. In the first part of this section we introduce a generating function that represents the distribution $P_{\mathbf{c}}$ of $\mathbf{c}$ under the null distribution. This is applied in the second part to derive the order statistics of $P_{\mathbf{c}}$ and a formula for computing the collective p-values (3.2) *exactly*. Recall that the set $\mathcal{C}_n$ of signatures equals the set of maximal tubings or vertices of the cyclohedron C_n. For each $\sigma \in \mathcal{C}_n$, the quantity $\mathbf{c}(\sigma) = p(\sigma)\cdot n!$ is the number of permutations δ which map to τ. See Algorithm 3.2.1 and Proposition 3.2.2.

We define the *count generating function* for the cyclohedron test to be the polynomial

$$\Gamma_n(t) \quad := \sum_{\sigma\in\mathcal{C}_n} t^{\mathbf{c}(\sigma)}.$$

By Theorem 3.5.1, this polynomial gives a refinement of the central binomial coefficient:

$$\Gamma_n(1) \;=\; |\mathrm{Vert}(C_n)| \;=\; \binom{2n-2}{n-1}.$$

Similarly, the first derivative $\Gamma_n'(t) = \frac{d}{dt}\Gamma_n(t)$ gives a refined count of the permutations:

$$\Gamma_n'(1) \;=\; |S_n| \;=\; n!.$$

We list the first few non-trivial instances of the count generating function:

$$\begin{aligned}
\Gamma_4(t) &= 4t^2 + 16t, \\
\Gamma_5(t) &= 20t^3 + 10t^2 + 40t, \\
\Gamma_6(t) &= 12t^8 + 24t^6 + 48t^4 + 48t^3 + 24t^2 + 96t, \\
\Gamma_7(t) &= 28t^{20} + 56t^{15} + 140t^{10} + 28t^8 + 56t^6 + 112t^5 + 112t^4 + 112t^3 + 56t^2 + 224t, \\
\Gamma_8(t) &= 8t^{80} + 32t^{48} + 128t^{45} + 64t^{40} + 64t^{36} + 64t^{30} + \cdots + 256t^3 + 128t^2 + 512t, \\
\Gamma_9(t) &= 72t^{210} + 72t^{168} + 108t^{140} + 144t^{126} + 432t^{105} + \cdots + 576t^3 + 288t^2 + 1152t.
\end{aligned}$$

The count generation function encodes the probability distribution function of $\mathbf{c}$:

Remark 3.6.1. The probability $P_{\mathbf{c}}(\gamma)$ is the coefficient of t^γ in the polynomial $(t/n!)\cdot\Gamma'_n(t)$.

Example 3.6.2. Consider the case $n = 7$. The $s_7 = 10$ possible permutation counts are

$$\mathrm{im}(\mathbf{c}) \;=\; \{1, 2, 3, 4, 5, 6, 8, 10, 15, 20\}.$$

The probability for each of these counts to be observed is the corresponding coefficient in

$$\sum_{\gamma \in \mathrm{im}(\mathbf{c})} P_{\mathbf{c}}(\gamma) \cdot t^\gamma \quad = \quad \frac{t}{5040} \cdot \Gamma'_7(t) \quad = \quad \frac{1}{9}t^{20} + \frac{1}{6}t^{15} + \frac{5}{18}t^{10} + \cdots + \frac{1}{45}t^2 + \frac{2}{45}t.$$

For instance, the cyclohedron C_6 has 56 vertices σ with $\mathbf{c}(\sigma) = 15$, and this accounts for $56 \cdot 15 = 840$ of the 5040 permutations δ in S_7. Thus the probability that a random data vector $v \in \mathbb{R}^7$ has permutation count $\mathbf{c}(v) = 15$ is equal to $P_{\mathbf{c}}(15) = 840/5040 = 1/6$. □

We now describe a formula for computing the count generating function. Let $\mathcal{T}_m$ denote the set of unlabeled rooted trees with m nodes, where each node has at most two children. The number of these trees is the *Wedderburn-Etherington number* , denoted by $t_m := |\mathcal{T}_m|$. Starting with $t_0 = 1$, the Wedderburn-Etherington numbers are

$$1, 1, 2, 3, 6, 11, 23, 46, 98, 207, 451, 983, 2179, 4850, 10905, 24631, 56011, 127912, \ldots$$

and they can be computed by the following recursion:

$$t_m = \sum_{i=0}^{\lfloor m/2 \rfloor - 1} t_i \cdot t_{m-i-1} \qquad \text{if } m \text{ is even,}$$
$$t_m = \binom{t_{(m-1)/2}}{2} + \sum_{i=0}^{\lfloor m/2 \rfloor - 1} t_i \cdot t_{m-i-1} \quad \text{if } m \text{ is odd.}$$

This holds because each tree T in $\mathcal{T}_m$ is constructed uniquely by taking an unordered pair consisting of a tree T_1 in $\mathcal{T}_i$ and a tree T_2 in $\mathcal{T}_{m-i-1}$ and attaching them to a new root. Note that $t_0 = 1$ corresponds to the case when the new root has outdegree one. We call $\binom{m-1}{i}$ the *order* of the root. The node is called *balanced* if $i = (m-1)/2$ and the two subtrees T_1 and T_2 are isomorphic. In this manner, each node of a tree $T \in \mathcal{T}_m$ has an order, and it is either balanced or unbalanced. For instance, all leaves are balanced of order 1, all nodes with one child are unbalanced of order 1, and nodes with two children have order ≥ 2. For a tree $T \in \mathcal{T}_m$ let $\mathrm{unbal}(T)$ denote the number of unbalanced nodes in the tree T, and let $\mathrm{order}(T)$ denote the product of the orders of all nodes in T.

Theorem 3.6.3. *The count generating function for the cyclohedron test equals*

$$\Gamma_n(t) \quad = \quad n \cdot \sum_{T \in \mathcal{T}_{n-1}} 2^{\mathrm{unbal}(T)} \cdot t^{\mathrm{order}(T)}.$$

Proof. Every signature $\sigma = \{\sigma_1, \ldots, \sigma_{n-1}\}$ in $\mathcal{C}_n$ maps to an unordered tree $T = T(\sigma)$ in $\mathcal{T}_{n-1}$. If $n = 2$ then T is the tree with one node. For $n \geq 3$ we construct T iteratively as in Algorithm 3.2.1: by induction, the sets Left and Right correspond to two subtrees T_1 and T_2, and a new root is attached to form the tree corresponding to $\{\delta_i\} \cup \mathrm{Right} \cup \mathrm{Left}$. The order of the resulting tree $T(\sigma)$ equals the permutation count $\mathbf{c}(\sigma)$ computed. It remains to be shown that the set of all signatures σ which are mapped to the same tree $T \in \mathcal{T}_{n-1}$ has precisely $n \cdot 2^{\mathrm{unbal}(T)}$ elements. The factor n comes from the fact that the last element δ_n can be chosen arbitrarily. So, let us suppose $\delta_n = n$. Then the indices appearing in

σ are precisely $1, 2, \ldots, n{-}1$. Let T_1 and T_2 be the two subtrees of the root of T, and suppose they have i and $n-2-i$ nodes respectively. If $i \neq n/2$ then either $\delta_{n-1} = i+1$ and both $\{1, 2, \ldots, i\}$ and $\{n{-}2{-}i, \ldots, n{-}2, n{-}1\}$ are in σ, or $\delta_{n-1} = n{-}1{-}i$ and both $\{1, 2, \ldots, n{-}2{-}i\}$ and $\{n-i, \ldots, n{-}2, n{-}1\}$ are in σ. If $i = n/2$ then $\delta_{n-1} = n/2$ and both $\{1, 2, \ldots, n/2-1\}$ and $\{n/2+1, \ldots, n-1\}$ are in σ. The choices for the remaining elements of σ are constructed inductively by identifying the nodes of the two subtrees with these two sets. If the two subtrees are identical (i.e. the root is balanced) then there is only one identification to be considered, otherwise we must consider two cases. Proceeding in this manner along the tree, we see that there are $2^{\mathrm{order}(T)}$ many choices of signatures σ on $\{1, 2, \ldots, n-1\}$ which map to T. □

We next present a recursive method for computing the count generating function $\Gamma_n(t)$. Let $f = \sum_i a_i t^i$ and $g = \sum_j b_j t^j$ be any two generating functions and M any positive integer. Then we define the *$*$-product* of f and g with respect to M as follows:

$$f *_M g \quad := \quad \sum_{i,j} a_i \cdot b_j \cdot t^{i \cdot j \cdot M}. \tag{3.7}$$

Corollary 3.6.4. *Let $\Omega_n(t)$ be the polynomial defined recursively by*

$$\Omega_0(t) \;=\; \Omega_1(t) \;=\; t \qquad \textit{and} \qquad \Omega_m(t) \;=\; \sum_{i=0}^{m-1} \Omega_i(t) *_{\binom{m-1}{i}} \Omega_{m-1-i}(t).$$

Then $\Gamma_n(t) \;=\; n \cdot \Omega_{n-1}(t)$ is the count generating function for the cyclohedron test.

Proof. This follows from the recursive tree construction in the proof of Theorem 3.6.3. □

Example 3.6.5. Corollary 3.6.4 easily yields the full expansion of $\Omega_n(t)$ for small values of n. For $n = 17$, the case of interest in Section 3.4 (see also Examples 3.2.4 and 3.5.3), we find

$$\Gamma_{17}(t) \;=\; 272t^{108108000} + 544t^{89689600} + 272t^{86486400} + 544t^{80720640} + \cdots\cdots + 348160t^{8}$$
$$+278528t^{7} + 417792t^{6} + 278528t^{5} + 278528t^{4} + 278528t^{3} + 139264t^{2} + 557056t.$$

The number of terms in this polynomial equals $|\operatorname{im}(\mathbf{c})| = 2438$. The 2438 values of the probability distribution function $P_{\mathbf{c}}$ are plotted on a logarithmic scale in Figure 3.3. For larger values of n, say $n \geq 30$, it becomes infeasible to compute the expansion of $\Gamma_n(t)$, but Corollary 3.6.4 can still be used to design efficient methods for sampling from $P_{\mathbf{c}}$. □

The distribution function $F_{(r)}(\gamma)$ of the order statistic $\mathbf{C}_{(r)}$ is now computed. Defining $p_i = P_{\mathbf{c}}(\gamma_i)$ to be the probability under the null hypothesis that the count is equal to γ_i, Remark 3.6.1 tells us that

$$(t/n!) \cdot \Gamma_n'(t) \;=\; p_1 t^{\gamma_1} + p_2 t^{\gamma_2} + \cdots + p_{s_n} t^{\gamma_{s_n}}, \qquad \text{where} \quad \gamma_1 < \gamma_2 < \cdots < \gamma_{s_n}.$$

Consider the identity

$$(p_1 + p_2 + \cdots + p_{s_n})^N \;=\; \sum_{i_1+i_2+\cdots+i_{s_n}=N} \binom{N}{i_1\, i_2\, \cdots\, i_{s_n}} \cdot p_1^{i_1} p_2^{i_2} \cdots p_{s_n}^{i_{s_n}} \;=\; 1.$$

By definition, $F_{(r)}(\gamma_k)$ is the sub-sum of all terms in this sum whose indices satisfy

$$i_1 + \cdots + i_{k-1} \;<\; r \;\leq\; N - i_{k+1} - \cdots - i_{s_n}.$$

For the purpose of computational efficiency we rewrite this sub-sum as follows. The formula below furnishes us with an efficient method for computing the collective p-values.

Lemma 3.6.6. *The probability distribution function under the null distribution Q_0 of the order statistic $\mathbf{C}_{(r)}$ is given by*

$$F_{(r)}(\gamma_k) \;=\; \sum_{i=1}^{N} \sum_{j=\max(0,r-i)}^{\min(r-1,N-i)} \binom{N}{i,\, j,\, N-i-j} (p_1 + \cdots + p_{k-1})^j \cdot p_k^i \cdot (p_{k+1} + \cdots + p_{s_n})^{N-i-j}$$

Proof. The first sum is over the number of data points that have permutation count γ_k. The second sum is over j, the number of data points whose permutation count is less than γ_k. Then, the multinomial coefficient gives the possible ways to partition $\{1, 2, \ldots, N\}$ into sets of size i, j, and $N - i - j$; that is, it accounts for possible rearrangements among the permutation counts equal to, less than, and greater than γ_k. The probability that such rearrangement occurs is the product $(p_1 + \cdots + p_{k-1})^j \cdot p_k^i \cdot (p_{k+1} + \cdots + p_{s_n})^{N-i-j}$. □

3.7 R and Python code for the cyclohedron test

The `R` source code `topoGraph.R` is available for the cyclohedron test. The software can be downloaded from

`http://bio.math.berkeley.edu/ranktests/index.html`

Our code requires the free statistical software package `R` [58]. Here we describe how to perform basic tasks related to the cyclohedron test. The data file must be a CSV (comma-separated values) file, where the first column consists of identifying labels (such as gene names), and the first row labels the time points (all other rows are the corresponding data vectors). We illustrate the use of the basic functions with the data file (named '13873.csv') that we described in Section 3.4. The first column consists of the ProbeSet IDs. The source code containing the `R` functions is `topoGraph.R`. First, we call the source code and load the data file from an `R` command line (here, we assume that both files are in the current working directory):

```
source("topoGraph.R")
dataset<-loaddata("13873.csv")
```

Next, we calculate the count of each data vector, which is done by the following command:

```
counts<-cycleCounts(dataset)
```

This defines "counts," a vector which lists the counts **c** of the data vectors in the order given by the data file. To list the genes according to their count ranking, as shown in Table 1, we call the function `rankby` which outputs the labels (here, the ProbeSet IDs) of the genes. The following command outputs the ten highest ranked ProbeSets.

```
rankby(row.names(dataset),counts)[1:10]
 [1] "1456017_x_at" "1452041_at"   "1418593_at"   "1417985_at"
 [5] "1436845_at"   "1436343_at"   "1426267_at"   "1420360_at"
 [9] "1449643_s_at" "1417399_at"
```

More extensive documentation is available online. The `Python` code implementing the cyclohedron test and other graphical tests is available at `http://math.berkeley.edu/ ~ mortonj/code.html`.

3.8 Roots

We also applied this method to root expression data generated by Siobhan Brady and Philip Benfey [9]. This data was far less noisy and resulted in a 17 "perfectly cyclical" gene expression time series: topographical maps with preimage size 1. This remained highly significant after correcting for multiple hypothesis testing as above. The top 17 genes are shown in Figure 3.5, and correspond to the following probesets.

```
255502_at, 248530_at, 248526_at, 254363_at, 257697_at, 247024_at,
249971_at, 262236_at, 262371_at, 256237_at, 259195_at, 266135_at,
257823_at, 256723_at, 261144_s_at, 261772_at
```

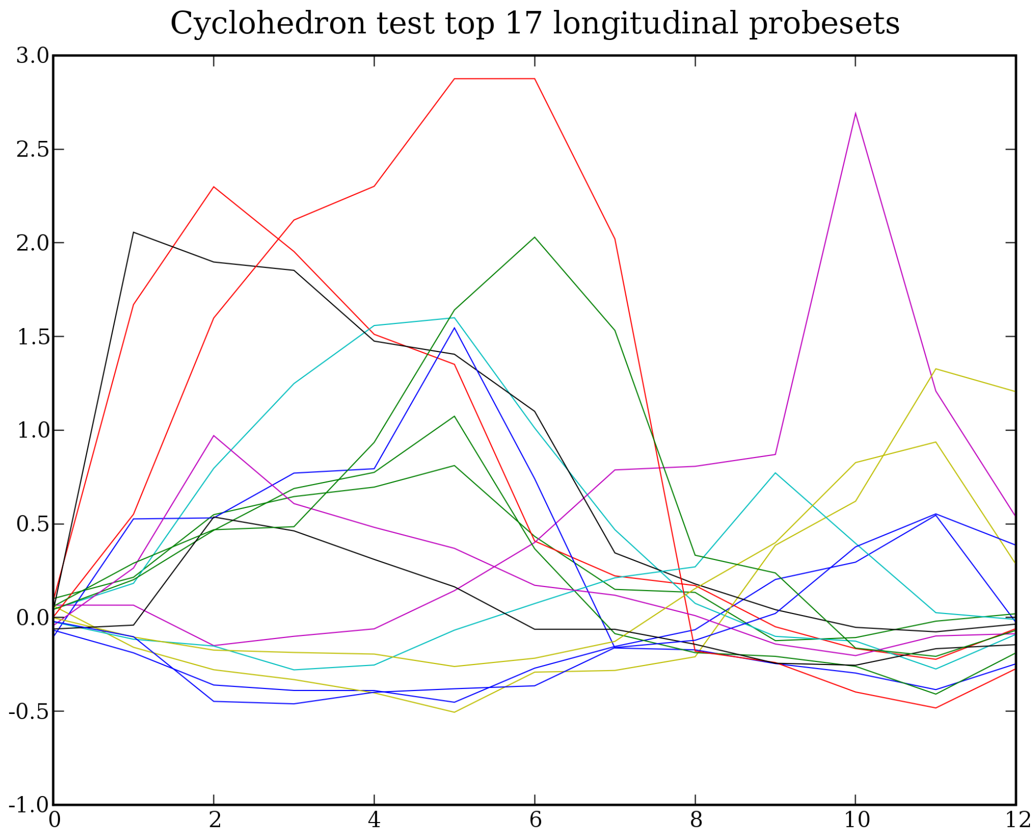


Figure 3.5: Top 17 longitudinal probesets in the *Arabidopsis* data.

Chapter 4

Three counterexamples on semigraphoids

4.1 Introduction

The material in this chapter is based on the paper "Three counterexamples in Semigraphoids," [34] which was authored jointly with R. Hemmecke, A. Shiu, B. Sturmfels and O. Wienand. While Chapter 3 focused on applications, in this chapter we exploit Theorem 2.3.3 to answer three purely theoretical questions about conditional independence. The following two problems about semigraphoids were stated by Studený:

(**Q1**) *Is it true that every coatom of the lattice of (disjoint) semigraphoids over* $[n]$ *is a structural independence model over* $[n]$*?* [67, Question 4, page 194]

(**Q2**) *Is every structural imset over* $[n]$ *already a combinatorial imset over* $[n]$*?* [67, Question 7, page 207]

Let $\mathbf{P}_{n-1}$ denote the $(n{-}1)$-dimensional *permutohedron* and let $C_n = [0,1]^n$ denote the standard n-dimensional cube. Recall that the vertices of $\mathbf{P}_{n-1}$ are in bijection with the elements of the symmetric group S_n; they are also in bijection with the monotone edge paths from $(0, 0, \ldots, 0)$ to $(1, 1, \ldots, 1)$ on the cube C_n. The 2-dimensional faces of C_n are in bijection with the CI statements on $[n]$. Namely, $i \perp\!\!\!\perp j \,|\, K = j \perp\!\!\!\perp i \,|\, K$ represents the 2-face of C_n with $x_k = 1$ for $k \in K$ and $x_l = 0$ for $l \in [n] \backslash (K \cup \{i, j\})$. The number of these 2-cubes equals $\gamma_n := \binom{n}{2} 2^{n-2}$. There is a natural surjection from the edges of $\mathbf{P}_{n-1}$ onto the 2-faces of C_n. Namely, an edge of $\mathbf{P}_{n-1}$ corresponds to a pair of adjacent monotone edge paths on C_n. These adjacent paths differ only along a 2-cube $[i \perp\!\!\!\perp j \,|\, K]$. In this manner, we identify any set $\mathcal{M}$ of CI statements on $[n]$ with a set of 2-cubes on the boundary of C_n. As in (2.2), we also identify $\mathcal{M}$ with a set of edges of the permutohedron $\mathbf{P}_{n-1}$, bearing in mind that opposite edges of a square have the same CI statement as their label.

A basic question about any semigraphoid $\mathcal{M}$ is whether its corresponding convex rank test is *submodular* (§2.4), in other words, whether it is the normal fan of a convex polytope. That polytope would then be a Minkowski summand of $\mathbf{P}_{n-1}$, i.e. a generalized permutohedra [55, 56].

Studený's first question has the following geometric translations:

(**Q1**) *Is every coarsest convex rank test submodular?*

(**Q1**) *Is every fan which maximally coarsens the S_n-arrangement the normal fan of a generalized permutohedron?*

In the first version of [56], Postnikov, Reiner and Williams asked a similar question:

(**Q3**) *Is every simplicial fan which coarsens the S_n-arrangement the normal fan of a simple generalized permutohedron?*

This chapter answers all three questions. In Section 4.2 we derive and explain our counterexample for Question (**Q3**). That example is discussed in [56, Example 3.8]. By Studený's classification of the 26424 semigraphoids for $n = 4$, it had been known that the answers to Questions (**Q1**) and (**Q2**) are affirmative for $n \leq 4$. In Sections 4.3 and 4.4 we construct counterexamples for (**Q1**) and (**Q2**) with $n = 5$.

Question (**Q2**) has the following reformulation in the setting of toric algebra [48, §7]. We now represent the elementary semigraphoid axiom $(\mathbf{SG})'$ of §2.3 as an additive equation among CI statements as formal symbols (indeterminates):

$$(\mathbf{SG}'') \qquad [i \perp\!\!\!\perp j \,|K \cup \ell] + [i \perp\!\!\!\perp \ell \,|K] \quad = \quad [i \perp\!\!\!\perp j \,|K] + [i \perp\!\!\!\perp \ell \,|K \cup j]$$

for all i, j, l, K. We enclose the CI statements in brackets to emphasize that they are now indeterminants. These relations span the kernel of the linear map

$$\mathcal{A} : \mathbb{Z}^{\gamma_n} \to \mathbb{Z}^{2^n}, \ [i \perp\!\!\!\perp j \,|\, K] \ \mapsto \ e_{iK} + e_{jK} - e_K - e_{ijK}. \tag{4.1}$$

A semigraphoid is a solution to the equations $(\mathbf{SG}'')$ in the semiring $\{0, +\}$, representing "zero" and "positive". A semigraphoid is submodular if it is the set of zero coordinates of a solution to $(\mathbf{SG}'')$ in the non-negative real numbers. These definitions furnish us with an algebraic representation of a semigraphoid $\mathcal{M}$ and a systematic method for testing submodularity of $\mathcal{M}$ by linear programming. Studený's question (**Q2**) concerns the $\mathbb{N}$-linear span of the columns of the matrix $\mathcal{A}$:

(**Q2**) *Is the semigroup $\mathcal{A}(\mathbb{N}^{\gamma_n})$ normal, i.e., does it coincide with $\mathcal{A}(\mathbb{R}^{\gamma_n}_{\geq 0}) \cap \mathbb{Z}^{2^n}$?*

4.2 A non-submodular simplicial semigraphoid

Let $n = 4$ and consider the 4-dimensional cube C_4 and the 3-dimensional permutohedron $\mathbf{P}_3$. Each hexagon on $\mathbf{P}_3$ corresponds to one of the eight facets of C_4. Each facet specifies three semigraphoid axioms, written additively as in ($\mathbf{SG}''$):

$$
\begin{array}{llcll}
 & [\![1 \perp\!\!\!\perp 2|\emptyset]\!] + [2 \perp\!\!\!\perp 3|1] & = & [2 \perp\!\!\!\perp 3|\emptyset] + [1 \perp\!\!\!\perp 2|3] & \Longleftarrow \\
(*,*,*,0) & [1 \perp\!\!\!\perp 3|\emptyset] + [1 \perp\!\!\!\perp 2|3] & = & [\![1 \perp\!\!\!\perp 2|\emptyset]\!] + [1 \perp\!\!\!\perp 3|2] & \\
 & [1 \perp\!\!\!\perp 3|\emptyset] + [2 \perp\!\!\!\perp 3|1] & = & [2 \perp\!\!\!\perp 3|\emptyset] + [1 \perp\!\!\!\perp 3|2] & \\
 & & & & \\
 & [\![1 \perp\!\!\!\perp 2|\emptyset]\!] + [2 \perp\!\!\!\perp 4|1] & = & [2 \perp\!\!\!\perp 4|\emptyset] + [1 \perp\!\!\!\perp 2|4] & \\
(*,*,0,*) & [\![1 \perp\!\!\!\perp 2|\emptyset]\!] + [1 \perp\!\!\!\perp 4|2] & = & [1 \perp\!\!\!\perp 4|\emptyset] + [1 \perp\!\!\!\perp 2|4] & \\
 & [1 \perp\!\!\!\perp 4|\emptyset] + [2 \perp\!\!\!\perp 4|1] & = & [2 \perp\!\!\!\perp 4|\emptyset] + [1 \perp\!\!\!\perp 4|2] & \\
 & & & & \\
 & [1 \perp\!\!\!\perp 3|\emptyset] + [1 \perp\!\!\!\perp 4|3] & = & [1 \perp\!\!\!\perp 4|\emptyset] + [1 \perp\!\!\!\perp 3|4] & \\
(*,0,*,*) & [\![3 \perp\!\!\!\perp 4|\emptyset]\!] + [1 \perp\!\!\!\perp 3|4] & = & [1 \perp\!\!\!\perp 3|\emptyset] + [3 \perp\!\!\!\perp 4|1] & \\
 & [\![3 \perp\!\!\!\perp 4|\emptyset]\!] + [1 \perp\!\!\!\perp 4|3] & = & [1 \perp\!\!\!\perp 4|\emptyset] + [3 \perp\!\!\!\perp 4|1] & \Longleftarrow \\
 & & & & \\
 & [2 \perp\!\!\!\perp 3|\emptyset] + [3 \perp\!\!\!\perp 4|2] & = & [\![3 \perp\!\!\!\perp 4|\emptyset]\!] + [2 \perp\!\!\!\perp 3|4] & \\
(0,*,*,*) & [2 \perp\!\!\!\perp 4|\emptyset] + [2 \perp\!\!\!\perp 3|4] & = & [2 \perp\!\!\!\perp 3|\emptyset] + [2 \perp\!\!\!\perp 4|3] & \\
 & [\![3 \perp\!\!\!\perp 4|\emptyset]\!] + [2 \perp\!\!\!\perp 4|3] & = & [2 \perp\!\!\!\perp 4|\emptyset] + [3 \perp\!\!\!\perp 4|2] & \\
 & & & & \\
 & [3 \perp\!\!\!\perp 4|1] + [\![2 \perp\!\!\!\perp 3|14]\!] & = & [2 \perp\!\!\!\perp 3|1] + [3 \perp\!\!\!\perp 4|12] & \Longleftarrow \\
(*,*,*,1) & [2 \perp\!\!\!\perp 4|1] + [\![2 \perp\!\!\!\perp 3|14]\!] & = & [2 \perp\!\!\!\perp 3|1] + [2 \perp\!\!\!\perp 4|13] & \\
 & [2 \perp\!\!\!\perp 4|1] + [3 \perp\!\!\!\perp 4|12] & = & [3 \perp\!\!\!\perp 4|1] + [2 \perp\!\!\!\perp 4|13] & \\
 & & & & \\
 & [1 \perp\!\!\!\perp 3|2] + [3 \perp\!\!\!\perp 4|12] & = & [3 \perp\!\!\!\perp 4|2] + [1 \perp\!\!\!\perp 3|24] & \\
(*,*,1,*) & [1 \perp\!\!\!\perp 3|2] + [\![1 \perp\!\!\!\perp 4|23]\!] & = & [1 \perp\!\!\!\perp 4|2] + [1 \perp\!\!\!\perp 3|24] & \\
 & [3 \perp\!\!\!\perp 4|2] + [\![1 \perp\!\!\!\perp 4|23]\!] & = & [1 \perp\!\!\!\perp 4|2] + [3 \perp\!\!\!\perp 4|12] & \\
 & & & & \\
 & [1 \perp\!\!\!\perp 2|3] + [\![1 \perp\!\!\!\perp 4|23]\!] & = & [1 \perp\!\!\!\perp 4|3] + [1 \perp\!\!\!\perp 2|34] & \Longleftarrow \\
(*,1,*,*) & [1 \perp\!\!\!\perp 4|3] + [2 \perp\!\!\!\perp 4|13] & = & [2 \perp\!\!\!\perp 4|3] + [\![1 \perp\!\!\!\perp 4|23]\!] & \\
 & [1 \perp\!\!\!\perp 2|3] + [2 \perp\!\!\!\perp 4|13] & = & [2 \perp\!\!\!\perp 4|3] + [1 \perp\!\!\!\perp 2|34] &
\end{array}
$$

$$\begin{array}{lrcl} & [1 \perp\!\!\!\perp 3|4] + [\![2 \perp\!\!\!\perp 3|14]\!] & = & [2 \perp\!\!\!\perp 3|4] + [1 \perp\!\!\!\perp 3|24] \\ (1, *, *, *) & [1 \perp\!\!\!\perp 2|4] + [1 \perp\!\!\!\perp 3|24] & = & [1 \perp\!\!\!\perp 3|4] + [1 \perp\!\!\!\perp 2|34] \\ & [1 \perp\!\!\!\perp 2|4] + [\![2 \perp\!\!\!\perp 3|14]\!] & = & [2 \perp\!\!\!\perp 3|4] + [1 \perp\!\!\!\perp 2|34]. \end{array}$$

This is a system of 24 equations in $\gamma_4 = 24$ formal symbols $[i \perp\!\!\!\perp j \,|\, K]$.

A semigraphoid is a solution to these equations over the semiring $\{0, +\}$. More precisely, given such a solution vector in $\{0, +\}^{24}$, the semigraphoid $\mathcal{M}$ consists of all coordinates $[i \perp\!\!\!\perp j \,|\, K]$ that have the value 0. There are 26424 such semigraphoids. They form a sublattice of the Boolean lattice $\{0, +\}^{24}$, with $+ < 0$. Question (**Q1**) concerns the coatoms of this lattice. But let us first resolve Question (**Q3**).

We consider the following collection of CI statements:

$$\mathcal{M} \;=\; \big\{\, [\![2 \perp\!\!\!\perp 3 \,|\, 14]\!],\, [\![1 \perp\!\!\!\perp 4 \,|\, 23]\!],\, [\![1 \perp\!\!\!\perp 2 \,|\, \emptyset]\!],\, [\![3 \perp\!\!\!\perp 4 \,|\, \emptyset]\!] \,\big\}. \qquad (4.2)$$

These four symbols are highlighted in the 24 equations above by the use of double brackets $[\![\cdots]\!]$. Each equation (individually) can be solved among the positive reals after these four symbols have been set to zero, or equivalently they can be solved as a system over $\{0, +\}$. This shows that $\mathcal{M}$ is a semigraphoid.

The semigraphoid $\mathcal{M}$ is represented geometrically by the three-dimensional polytope in Figure 4.1. This polytope is *simple*, i.e., each of the 16 vertices is adjacent to three other vertices. The eight vertices whose labels include three bars (such as 4|2|1|3) correspond to unique permutations in S_4 (namely the permutation 4213), while the eight vertices whose labels have two bars (such as 4|1|23) correspond to pairs of permutations in S_4 (namely 4123 and 4132). This partition of S_4 into eight singletons and eight pairs is the convex rank test of $\mathcal{M}$. The normal fan of the polytope in Figure 4.1 is a simplicial

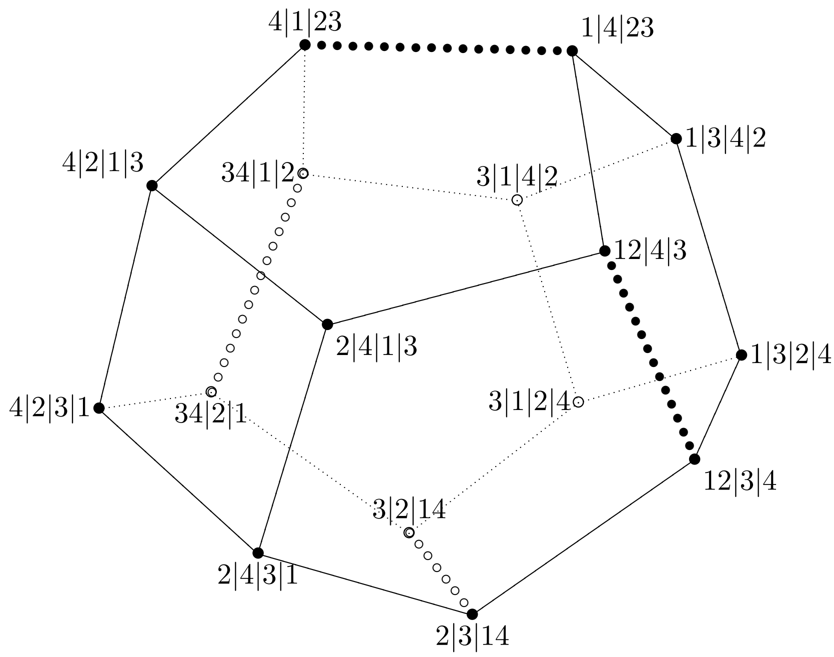


Figure 4.1: A simple 3-dimensional polytope with 16 vertices and 10 facets

fan which is combinatorially (but not geometrically) isomorphic to a fan that coarsens the hyperplane arrangement of S_4.

Proposition 4.2.1. *The simplicial semigraphoid $\mathcal{M}$ is not submodular.*

Proof. Suppose that $\mathcal{M}$ were submodular. Then the above equations have a solution in $(\mathbb{R}_{\geq 0})^{24}$ whose coordinates in $\mathcal{M}$ are zero and whose other 20 coordinates are positive. The four equations marked by an "$\Longleftarrow$" give the following four equations:

$$\begin{aligned}
[2 \perp\!\!\!\perp 3|1] &= [2 \perp\!\!\!\perp 3|\emptyset] + [1 \perp\!\!\!\perp 2|3] \\
[1 \perp\!\!\!\perp 4|3] &= [1 \perp\!\!\!\perp 4|\emptyset] + [3 \perp\!\!\!\perp 4|1] \\
[3 \perp\!\!\!\perp 4|1] &= [2 \perp\!\!\!\perp 3|1] + [3 \perp\!\!\!\perp 4|12] \\
[1 \perp\!\!\!\perp 2|3] &= [1 \perp\!\!\!\perp 4|3] + [1 \perp\!\!\!\perp 2|34].
\end{aligned}$$

Adding the left hand sides and the right hand sides of the four equations yields

$$[2 \perp\!\!\!\perp 3|\emptyset] + [1 \perp\!\!\!\perp 4|\emptyset] + [3 \perp\!\!\!\perp 4|12] + [1 \perp\!\!\!\perp 2|34] = 0.$$

This contradicts the assumption that these four values are strictly positive. □

The set of all non-negative solutions to the 24 equations is an 11-dimensional cone in $(\mathbb{R}_{\geq 0})^{24}$. This cone is isomorphic to the 16-dimensional cone of submodular functions on $2^{[4]}$, modulo its 5-dimensional lineality space. Its 22108 faces are in bijection with the submodular semigraphoids, or, equivalently, with the generalized permutohedra for $n = 4$. In addition to these, there are 4316 semigraphoids that are not submodular. Each of the latter can be represented by a polytope of dimension ≤ 3 as in Figure 1. These polytopes have the combinatorial properties of generalized permutohedra, but they cannot be realized as Minkowski summands of $\mathbf{P}_3$. For example, see [36, Figure 5] for a polytope that depicts Studený's example of a semigraphoid that is not submodular.

We now give a classification of non-submodular semigraphoids for $n = 4$ and $|\mathcal{M}|$ small. All simplicial examples are coarsenings (up to relabeling) of the particular semigraphoid $\mathcal{M}$ in Proposition 4.2.1. The following table lists the number of semigraphoids classified by number of CI statements, their type, and whether they are simplicial. Here, the *type* of a semigraphoid is the triple (m_0, m_1, m_2) where m_t is the number of CI statements $[i \perp\!\!\!\perp j \,|\, K]$ in $\mathcal{M}$ such that $|K| = m_t$.

$\|\mathcal{M}\|$	type	non-simplicial	simplicial	total
3	(0 , 3 , 0)	8	0	8
4	(0 , 4 , 0)	78	0	78
4	(1 , 2 , 1)	30	0	30
4	(2 , 0 , 2)	0	6	6
5	(0 , 5 , 0)	300	0	300
5	(1 , 2 , 2)	30	0	30
5	(1 , 3 , 1)	84	0	84
5	(2 , 0 , 3)	12	12	24
5	(2 , 2 , 1)	30	0	30
5	(3 , 0 , 2)	24	0	24
6	(0 , 6 , 0)	604	0	604
6	(1 , 3 , 2)	84	0	84
6	(1 , 4 , 1)	78	0	78
6	(2 , 0 , 4)	30	3	33
6	(2 , 2 , 2)	30	0	30
6	(2 , 3 , 1)	84	0	84
6	(3 , 0 , 3)	74	12	96
6	(4 , 0 , 2)	30	3	33
7	(0 , 7 , 0)	684	0	684
7	(1 , 4 , 2)	78	0	78
7	(1 , 5 , 1)	24	0	24
7	(2 , 0 , 5)	18	0	18
7	(2 , 3 , 2)	84	0	84
7	(2 , 4 , 1)	78	0	78
7	(3 , 0 , 4)	132	0	132
7	(4 , 0 , 3)	132	0	132
7	(5 , 0 , 2)	18	0	18

$\lvert\mathcal{M}\rvert$	type	non-simplicial	simplicial	total
8	(0 , 8 , 0)	450	0	450
8	(1 , 5 , 2)	24	0	24
8	(2 , 0 , 6)	3	0	3
8	(2 , 4 , 2)	48	0	48
8	(2 , 5 , 1)	24	0	24
8	(3 , 0 , 5)	72	0	72
8	(4 , 0 , 4)	174	0	174
8	(5 , 0 , 3)	72	0	72
8	(6 , 0 , 2)	3	0	3
9	(0 , 9 , 0)	212	0	212
9	(3 , 0 , 6)	12	0	12
9	(4 , 0 , 5)	84	0	84
9	(5 , 0 , 4)	84	0	84
9	(6 , 0 , 3)	12	0	12
10	(0 , 10 , 0)	60	0	60
10	(4 , 0 , 6)	15	0	15
10	(5 , 0 , 5)	24	0	24
10	(6 , 0 , 4)	15	0	15
11	(0 , 11 , 0)	12	0	12
11	(5 , 0 , 6)	6	0	6
11	(6 , 0 , 5)	6	0	6

4.3 A non-submodular coarsest semigraphoid

We now consider the case $n = 5$. There are $\gamma_5 = 80$ CI statements, one for each two-dimensional face of the 5-cube C_5. There are 120 semigraphoid axioms (**SG**$''$), three for each of the 40 three-dimensional faces of C_5, listed as additive equations in Section 4.6. The semigraphoids are the solutions of these equations over $\{0, +\}^{80}$. These solutions include the all-zero vector $\mathbf{0}$ which represents the semigraphoid that consists of all 80 CI statements, and which is the maximal element in the lattice of semigraphoids. A semigraphoid is said to be *coarsest* if it is maximal among non-$\mathbf{0}$ semigraphoids. Geometrically, such a semigraphoid corresponds to a fan which coarsens the S_5-arrangement but cannot be coarsened to a non-trivial fan.

We now present the counterexample which answers question (**Q1**). Our constructions make use of the identification of semigraphoids with convex rank tests that was derived in Chapter 2. Let Γ denote the partition of the symmetric group S_5 into fourteen classes as follows. There are eight classes containing 12 permutations each:

$$15|234 \quad 234|15 \quad 123|45 \quad 235|14$$
$$124|35 \quad 245|13 \quad 134|25 \quad 345|12.$$

And there are six classes containing four permutations each:

$$12|5|34 \quad 25|1|34 \quad 13|5|24$$
$$35|1|24 \quad 14|5|23 \quad 45|1|23.$$

Here $15|234$ denotes the class of all permutations $ijklm$ with $\{i,j\} = \{1,5\}$ and $\{k,l,m\} = \{2,3,4\}$. Similarly, $45|1|23$ denotes the class of all permutations $ijklm$ with $\{i,j\} = \{4,5\}$, $k = 1$, and $\{l,m\} = \{2,3\}$. Clearly, Γ is a pre-convex rank test, as each of the 14 classes is the set of all linear extensions of a poset on $[5] = \{1,2,3,4,5\}$. Note that the stabilizer of the pre-convex rank test Γ in S_5 has order 12, because Γ is fixed under permutations of $\{1,5\}$ and permutations of $\{2,3,4\}$. The 14 classes of Γ are represented by the 14 vertices of the polytope in Figure 4.2.

Each pair of adjacent permutations in a given class of Γ specifies a CI statement. For instance, the four-element class $45|1|23$ specifies the two CI statements $[\![4 \perp\!\!\!\perp 5|\emptyset]\!]$ and $[\![2 \perp\!\!\!\perp 3|145]\!]$, while the 12-element class $15|234$ specifies the seven CI statements

$$[\![1 \perp\!\!\!\perp 5|\emptyset]\!], [\![2 \perp\!\!\!\perp 3|15]\!], [\![2 \perp\!\!\!\perp 3|145]\!], [\![2 \perp\!\!\!\perp 4|15]\!], [\![2 \perp\!\!\!\perp 4|135]\!], [\![3 \perp\!\!\!\perp 4|15]\!], [\![3 \perp\!\!\!\perp 4|125]\!].$$

Altogether, we obtain 44 CI statements $[\![\cdot \,|\, \cdot]\!]$ from the 14 classes, and we identify the pre-convex rank test Γ with this set of 44 CI statements. We now prove:

Theorem 4.3.1. Γ *is a coarsest convex rank test which is not submodular.*

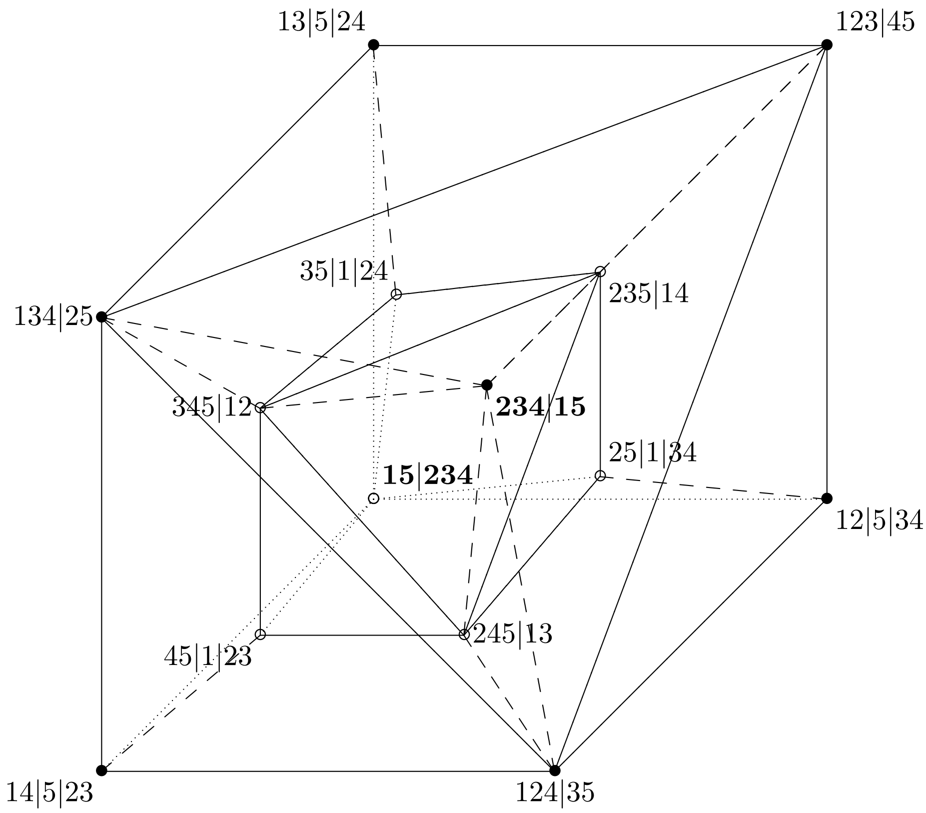


Figure 4.2: Schlegel diagram of a 4-dimensional polytope with 10 facets

Proof. To establish this theorem, we must prove the following three claims:

- Γ is a convex rank test, i.e. it satisfies the semigraphoid axioms (**SG**).
- There is no proper convex rank test which is coarser than Γ.
- The convex rank test Γ is not submodular.

We shall prove all three statements at once, by examining the semigraphoid equations (**SG**$''$). As in Section 2, the 44 symbols in Γ are denoted with double brackets $[\![\cdot|\cdot]\!]$, while the 36 symbols not in Γ are denoted with brackets $[\cdot|\cdot]$. With this distinction between brackets, there are four symmetry types of semigraphoid equations that involve the 36 positive unknowns $[\cdot|\cdot]$. The full list is given in Section 4.6:

$$\begin{array}{lccc} \text{Type I} & [3 \perp\!\!\!\perp 5|12] + [\![3 \perp\!\!\!\perp 4|125]\!] & = & [3 \perp\!\!\!\perp 4|12] + [\![3 \perp\!\!\!\perp 5|124]\!] \\ \text{Type II} & [1 \perp\!\!\!\perp 5|2] + [1 \perp\!\!\!\perp 3|25] & = & [\![1 \perp\!\!\!\perp 3|2]\!] + [1 \perp\!\!\!\perp 5|23] \\ \text{Type III} & [4 \perp\!\!\!\perp 5|1] + [2 \perp\!\!\!\perp 5|14] & = & [2 \perp\!\!\!\perp 5|1] + [4 \perp\!\!\!\perp 5|12] \\ \text{Type IV} & [1 \perp\!\!\!\perp 2|5] + [\![2 \perp\!\!\!\perp 3|15]\!] & = & [\![2 \perp\!\!\!\perp 3|5]\!] + [1 \perp\!\!\!\perp 2|35] \end{array}$$

After setting the 44 unknowns $[\![\cdot|\cdot]\!]$ to zero, we are left with 120 equations in the 36 strictly positive unknowns. For instance, the first three types give

$$\begin{array}{lccc} \text{Type I} & [3 \perp\!\!\!\perp 5|12] & = & [3 \perp\!\!\!\perp 4|12] \\ \text{Type II} & [1 \perp\!\!\!\perp 5|2] + [1 \perp\!\!\!\perp 3|25] & = & [1 \perp\!\!\!\perp 5|23] \\ \text{Type III} & [4 \perp\!\!\!\perp 5|1] + [2 \perp\!\!\!\perp 5|14] & = & [2 \perp\!\!\!\perp 5|1] + [4 \perp\!\!\!\perp 5|12] \end{array}$$

The axiom (**SG**$''$) merely requires that each of these equations is *individually* solvable. This is obviously the case. Hence Γ is a semigraphoid.

The 78 equations of Type I listed in Section 4.6 imply that all 36 positive unknowns must be equal. So, if another CI statement is added to the semigraphoid Γ, then all others must be added in order for (**SG**) to remain valid. This proves our second claim that Γ is a coarsest convex rank test.

Given that the 36 unknowns $[\cdot|\cdot]$ must be equal, the 12 Type II equations imply that their common value is zero, contradicting the requirement that they be positive. Hence the 120 original equations *altogether* have no non-negative real solution that is consistent with Γ. This proves our third claim that Γ is not submodular. □

Every semigraphoid for $n = 5$ corresponds to a 4-dimensional fan. Intersecting this fan with a sphere around the origin, we obtain a polyhedral cell decomposition of the 3-dimensional sphere. We do not know whether each of these 3-spheres can be realized as the boundary of a 4-dimensional polytope. However, using [73, §5], every semigraphoid can be represented by a 3-dimensional diagram as in Figure 4.2.

For the specific semigraphoid Γ of Theorem 4.3.1, the diagram in Figure 4.2 is indeed the boundary of a 4-polytope with f-vector $(14, 36, 32, 10)$. The following coordinates for this polytope were found by a direct calculation, using the techniques described in [8]. Each of the following ten row vectors represents a facet of our polytope:

```
POINTS
 1    1/4        0      0     0      0
 1     0         1      0     0      0
 1     0         0      1     0      0
 1     0         0      0     1      0
 1     0         0      0     0      1
 1   -1/4       1/4    1/4   5/4    1/4
 1  280/893  -280/893 25/893  0    28/893
 1    1/57      1/57  -1/57 17/19   2/57
 1     1         1      0    -5      1
 1    2/37     20/37   1/37 10/37  -2/37
```

For instance, the last row represents the facet-defining inequality

$$\frac{2}{37} \cdot x_1 + \frac{20}{37} \cdot x_2 + \frac{1}{37} \cdot x_3 + \frac{10}{37} \cdot x_4 - \frac{2}{37} \cdot x_5 \;\leq\; 1.$$

Here, we are considering the vectors $(x_1, x_2, x_3, x_4, x_5)$ to be elements in the quotient of $\mathbb{R}^5$ modulo the one-dimensional linear subspace spanned by $(4, 1, 1, 1, 1)$. Our format is that of the software `Polymake` [28]. If the above eleven lines are put in a file named `mypolytope` then the following command in `Polymake` will verify that this polytope does indeed have the combinatorial structure displayed in Figure 4.2:

```
polymake mypolytope F_VECTOR VERTICES_IN_FACETS
```

The 10 facets of our 4-polytope correspond to the facets of the 5-cube, and they comprise all classes of permutations in S_5 in which the first or last coordinate is fixed.

The facets corresponding to permutations with 1 *or* 5 in the *first* coordinate have seven vertices, twelve edges, and eight 2-faces. The facets corresponding to permutations with 2, 3 *or* 4 *first* have seven vertices, 13 edges, and eight 2-faces. The facets for 1 *or* 5 *last* are tetrahedra. The facets for 2, 3 *or* 4 *last* are cubes in which one edge has been contracted; they have seven vertices and 11 edges.

4.4 The semigraphoid semigroup is not normal

Continuing to assume $n = 5$, we now consider the linear map $\mathcal{A}$ in Section 4.1. It maps the free abelian group $\mathbb{Z}^{80}$ spanned by the CI statements to the free abelian group $\mathbb{Z}^{32}$ with basis $\{e_K : K \subseteq [5]\}$ as specified in (4.1). The matrix representing $\mathcal{A}$ has 32 rows and 80 columns; each column has four non-zero entries: two $+1$'s and two -1's. The rank of $\mathcal{A}$ is 26. The *semigraphoid semigroup* is $\mathcal{A}(\mathbb{N}^{80})$, the non-negative integer span of the columns of this 32×80-matrix. This is a subsemigroup of $\mathbb{Z}^{32}$. Equivalently, the semigraphoid semigroup is the affine semigroup with 80 generators and 120 relations (given in the Section 4.6). Note that the polyhedral cone dual to the semigraphoid semigroup is the cone of submodular functions.

In the language of [67], the vectors in $\mathbb{Z}^{32}$ are called *imsets*, the columns of $\mathcal{A}$ are *elementary imsets*, and the elements of $\mathcal{A}(\mathbb{N}^{80})$ are *combinatorial imsets.* A *structural imset* is a lattice point which lies in the polyhedral cone spanned by the elementary imsets. Studený's question (**Q2**) whether each structural imset is combinatorial translates into the question whether the semigroup $\mathcal{A}(\mathbb{N}^{80})$ is normal.

Theorem 4.4.1. *The semigraphoid semigroup is not normal for $n = 5$.*

Proof. Consider the following element in the free abelian group $\mathbb{Z}^{80}$:

$$\begin{aligned} &[1 \perp\!\!\!\perp 5|2] + [1 \perp\!\!\!\perp 4|3] + [2 \perp\!\!\!\perp 3|4] + [2 \perp\!\!\!\perp 3|5] + [3 \perp\!\!\!\perp 4|12] \\ &+[2 \perp\!\!\!\perp 5|13] + [1 \perp\!\!\!\perp 2|45] + [1 \perp\!\!\!\perp 3|45] + [4 \perp\!\!\!\perp 5|23] - [2 \perp\!\!\!\perp 3|45]. \end{aligned} \tag{4.3}$$

The image of this element under the map $\mathcal{A} : \mathbb{Z}^{80} \to \mathbb{Z}^{32}$ is the imset

$$\begin{aligned} \mathbf{b} \ :=& \ \ -e_2 - e_3 - e_4 - e_5 - e_{23} + e_{24} + 2e_{25} + 2e_{34} + e_{35} - e_{45} + 2e_{123} \\ &+e_{124} - e_{125} - e_{134} + e_{135} + 2e_{145} - e_{1234} - e_{1235} - e_{1245} - e_{1345}. \end{aligned} \tag{4.4}$$

The imset $\mathbf{b}$ is structural because $2 \cdot \mathbf{b}$ is a combinatorial imset. It is the image of

$$\begin{aligned} &[4 \perp\!\!\!\perp 5|2] + [4 \perp\!\!\!\perp 5|3] + [1 \perp\!\!\!\perp 3|4] + [1 \perp\!\!\!\perp 2|5] + [2 \perp\!\!\!\perp 5|14] + [3 \perp\!\!\!\perp 4|15] \\ &\quad +[1 \perp\!\!\!\perp 4|23] + [1 \perp\!\!\!\perp 5|23] + [1 \perp\!\!\!\perp 5|2] + [1 \perp\!\!\!\perp 4|3] + [2 \perp\!\!\!\perp 3|4] \\ &\quad +[2 \perp\!\!\!\perp 3|5] + [3 \perp\!\!\!\perp 4|12] + [2 \perp\!\!\!\perp 5|13] + [1 \perp\!\!\!\perp 2|45] + [1 \perp\!\!\!\perp 3|45] \qquad \in \ \mathbb{N}^{80} \end{aligned} \tag{4.5}$$

under the linear map $\mathcal{A}$.

Suppose that $\mathbf{b}$ were a combinatorial imset. Then there exists $\mathbf{x} \in \mathbb{N}^{80}$ such that $\mathcal{A} \cdot \mathbf{x} = \mathbf{b}$. We write $\mathbf{x} = \sum_i [a_i \perp\!\!\!\perp b_i | K_i]$, where we allow repetition in the sum. In any elementary imset, the basis vector $e_\emptyset$ occurs with coefficient -1 or 0, and the basis vector e_{12345} occurs with coefficient -1 or 0. However, neither $e_\emptyset$ nor e_{12345} appears in the imset $\mathbf{b}$, so we conclude that $|K_i| = 1$ or $|K_i| = 2$ for all terms $[a_i \perp\!\!\!\perp b_i | K_i]$ in the representation of $\mathbf{x}$. The first four terms $-e_2 - e_3 - e_4 - e_5$ in $\mathbf{b}$ imply that $\mathbf{x}$ has precisely four terms $[a_i \perp\!\!\!\perp b_i | K_i]$ with $|K_i| = 1$, and the terms $-e_{1234} - e_{1235} - e_{1245} - e_{1345}$ imply that $\mathbf{x}$ has precisely four terms with $|K_i| = 2$.

Each of the eight terms in $\mathbf{x}$ evaluates to an alternating sum of 4 terms under the map $\mathcal{A}$. Some cancellation occurs among the resulting 32 terms. Prior to that cancellation, our imset had been written as the sum of two subsums, $\mathbf{b} = \mathcal{A} \cdot \mathbf{x} =$

$$-e_2 - e_3 - e_4 - e_5 + e_{24} + 2e_{25} + 2e_{34} + e_{35} + e_{A_1} + e_{A_2} - e_{125} - e_{134} - e_{B_1} - e_{B_2}$$

$$-e_{23} - e_{45} - e_{A_1} - e_{A_2} + 2e_{123} + e_{124} + e_{135} + 2e_{145} + e_{B_1} + e_{B_2} - e_{1234} - e_{1235} - e_{1245} - e_{1345},$$

where $|A_1| = |A_2| = 2$ and $|B_1| = |B_2| = 3$. The first line is the sum of the four elementary imsets $\mathcal{A}([a_i \perp\!\!\!\perp b_i|K_i])$ with $|K_i| = 1$, and the second line is the sum of the four elementary imsets with $|K_i| = 2$. A contradiction will arise when we try to determine the unknown pairs A_1 and A_2. The term $-e_{125}$ in the first line must come from $K_i = \{2\}$ or $K_i = \{5\}$. This implies that either $\{1,2\}$ or $\{1,5\}$ is in $A_* = \{A_1, A_2\}$. Similarly, the term $-e_{134}$ shows that either $\{1,3\}$ or $\{1,4\}$ is in A_*. Now consider the second line. The presence of the term $2e_{123}$ implies that $\{1,2\}$ or $\{1,3\}$ is in A_*, and the term $2e_{145}$ implies that $\{1,4\}$ or $\{1,5\}$ is in A_*. The term e_{124} shows that $\{1,2\}$, $\{1,4\}$, or $\{2,4\}$ is in A_*, and, finally, the term e_{135} shows that $\{1,3\}$, $\{1,5\}$, or $\{3,5\}$ is in A_*. However, no such pair of pairs A_* satisfies these six restrictions. This proves that $\mathbf{b}$ is not a combinatorial imset. □

The main point of the above proof was to show that the linear system $\mathcal{A}\cdot\mathbf{x} = \mathbf{b}$ has no solution with non-negative *integer* coordinates. This can also be verified automatically using integer programming software. In fact, using such software we found that $\mathcal{A} \cdot \mathbf{x} = \mathbf{b}$ has only one solution with non-negative *real* coordinates, namely, that unique solution $\mathbf{x} \in (\mathbb{R}_{\geq 0})^{80}$ is the expression in (4.5) scaled by 1/2.

The reader might now inquire how the imset $\mathbf{b}$ was found. There are several algorithms that test whether a given affine semigroup is normal, including one recently proposed by Hemmecke, Takemura, and Yoshida [35], and the method of Bruns and Koch [13] which is implemented in their software `normaliz`.

Our original attempts to apply these methods directly to the 32×80-matrix $\mathcal{A}$ were unsuccessful. Instead we succeeded by partially computing a *Markov basis* for the matrix $\mathcal{A}$ using the software `4ti2` [33]. The imset $\mathbf{b}$ was found by inspecting the partial results produced by `4ti2`. We explain the details in the next section.

4.5 Computations in toric algebra

Let $\mathbb{Q}[\mathrm{CI}_n]$ denote the polynomial ring over the field of rational numbers $\mathbb{Q}$ generated by the symbols $[i \perp\!\!\!\perp j \,|\, K]$. Thus $\mathbb{Q}[\mathrm{CI}_n]$ is a polynomial ring in γ_n unknowns, one for each 2-face of the n-cube C_n. We write $\prod \mathrm{CI}_n$ for the product of all the unknowns. We define the *semigraphoid ideal* to be the ideal $I_{\mathbf{SG}}$ generated by the binomials in $(\mathbf{SG}''')$. Thus the generators of $I_{\mathbf{SG}}$ represent the semigraphoid axioms. Following [48, §7], we introduce the *toric ideal* $I_{\mathcal{A}}$ which is a prime ideal obtained from $I_{\mathbf{SG}}$ by saturation:

$$I_{\mathcal{A}} \quad := \quad \big(I_{\mathbf{SG}} \,:\, (\textstyle\prod \mathrm{CI}_n)^{\infty}\big). \tag{4.6}$$

The binomials in $I_{\mathcal{A}}$ represent the vectors in the kernel of the linear map $\mathcal{A} : \mathbb{Z}^{\gamma_n} \to \mathbb{Z}^{2^n}$. A minimal set of binomials which generates $I_{\mathcal{A}}$ is said to be a *Markov basis* for the matrix $\mathcal{A}$. See [22] for a discussion of Markov bases in the context of statistics.

Let us illustrate these concepts for $n = 3$. The polynomial ring $\mathbb{Q}[\mathrm{CI}_3]$ has six unknowns, one for each facet of the 3-cube. They are the entries of the 2×3-matrix

$$\begin{pmatrix} [1 \perp\!\!\!\perp 2|\emptyset] & [1 \perp\!\!\!\perp 3|\emptyset] & [2 \perp\!\!\!\perp 3|\emptyset] \\ [1 \perp\!\!\!\perp 2|3] & [1 \perp\!\!\!\perp 3|2] & [2 \perp\!\!\!\perp 3|1] \end{pmatrix}. \tag{4.7}$$

The semigraphoid ideal $I_{\mathbf{SG}}$ is generated by the three 2×2-minors of the matrix (4.7). This is a prime ideal of codimension 2 and degree 3, and hence we have $I_{\mathbf{SG}} = I_{\mathcal{A}}$. Here the Markov basis for $\mathcal{A}$ consists precisely of the three semigraphoid axiom instances.

We next consider the case $n = 4$. The polynomial ring $\mathbb{Q}[\mathrm{CI}_4]$ has 24 unknowns, one for each 2-face of the 4-cube. They are the entries of eight 2×3-matrices as in (4.7), one for each of the eight facets of the 4-cube. Thus the semigraphoid ideal $I_{\mathbf{SG}}$ is generated by 24 quadrics, one for each of the 24 axiom instances ($\mathbf{SG}''$) in the list given in Section 2. For instance, the last instance in that list translates into the quadratic binomial $[1{\perp\!\!\!\perp}2|4] \cdot [2{\perp\!\!\!\perp}3|14] - [2{\perp\!\!\!\perp}3|4] \cdot [1{\perp\!\!\!\perp}2|34]$, which is one of the 24 generators of $I_{\mathbf{SG}}$. Using the software `Macaulay2` [31] we derived the following result:

Proposition 4.5.1. *The semigraphoid ideal $I_{\mathbf{SG}}$ for $n = 4$ is a radical ideal which is the intersection of the toric ideal $I_{\mathcal{A}}$ and 17 additional associated monomial prime ideals.*

Before discussing this prime decomposition in detail, let us make a few general remarks. We wish to argue that toric algebra and algebraic geometry provide useful algorithmic tools for the research directions presented in [67]. For any ideal I of $\mathbb{Q}[\mathrm{CI}_n]$ and any subset Ω of the complex affine space $\mathbb{C}^{\gamma_n}$, the *variety* $V_\Omega(I)$ is defined as the set of all vectors in Ω which are common zeros of all the polynomials in I. Then $V_{\mathbb{C}}(I_{\mathbf{SG}})$ is a complex variety, reducible for $n \geq 4$, one of whose irreducible components is the complex toric variety $V_{\mathbb{C}}(I_{\mathcal{A}})$. Inside this toric variety are the real toric variety $V_{\mathbb{R}}(I_{\mathcal{A}})$. Its non-negative part $V_{\mathbb{R}_{\geq 0}}(I_{\mathcal{A}})$ is homeomorphic to the cone spanned by the elementary imsets. Our next result shows that the semigraphoids are precisely the points on these varieties whose coordinates are 0 or 1.

Theorem 4.5.2. *The semigraphoids on the set $[n]$ are in bijection with the points in $V_{\{0,1\}}(I_{\mathbf{SG}})$. The submodular semigraphoids on $[n]$ are in bijection with the points in $V_{\{0,1\}}(I_{\mathcal{A}})$.*

Proof. We replace the additive semiring $\{0, +\}$ with the multiplicative semiring $\{1, 0\}$. This translates from the additive notation ($\mathbf{SG}''$) to the multiplicative notation ($\mathbf{SG}'''$). With this translation, the first statement in Theorem 4.5.2 is obvious.

The second statement is less obvious and is based on the geometry of toric varieties. Specifically, we shall use the characterization of *facial* index sets which is developed in [29]. If we consider our specific $2^n \times \gamma_n$-matrix $\mathcal{A}$ then the role of the set $\{1, \ldots, m\}$ in [29] is played by the set of CI statements, and a subset of CI statements is facial for $\mathcal{A}$ if and only if it is submodular semigraphoid. With this observation, our second assertion follows from Lemma A.2 in the Appendix of [29]. □

Using Theorem 4.5.2, we can study semigraphoids by studying the zero-dimensional ideals obtained by adding $\langle\, x^2 - x \,:\, x \in \mathrm{CI}_n \,\rangle$ to the ideal $I_{\mathbf{SG}}$ or $I_{\mathcal{A}}$. For instance, with the command `degree` in `Macaulay2` [31], it takes only a few seconds to compute

$$\#V_{\{0,1\}}(I_{\mathbf{SG}}) \;=\; 26424 \qquad \text{and} \qquad \#V_{\{0,1\}}(I_{\mathcal{A}}) \;=\; 22108. \tag{4.8}$$

The difference between these numbers is explained geometrically by the prime decomposition in Proposition 4.5.1, which we shall now describe in explicit terms.

The 17 associated monomial primes of $I_{\mathbf{SG}}$ come in three symmetry classes. First there are two primes of codimension 12. A representative is the ideal

$$\left\langle \begin{array}{c} [1\perp\!\!\!\perp 2|\emptyset],\ [1\perp\!\!\!\perp 3|\emptyset],\ [1\perp\!\!\!\perp 4|\emptyset],\ [2\perp\!\!\!\perp 3|\emptyset],\ [2\perp\!\!\!\perp 4|\emptyset],\ [3\perp\!\!\!\perp 4|\emptyset], \\ [3\perp\!\!\!\perp 4|12], [2\perp\!\!\!\perp 4|13], [2\perp\!\!\!\perp 3|14], [1\perp\!\!\!\perp 4|23], [1\perp\!\!\!\perp 3|24], [1\perp\!\!\!\perp 2|34] \end{array} \right\rangle.$$

The semigraphoid ideal $I_{\mathbf{SG}}$ has 12 associated primes of codimension 15, such as

$$\left\langle \begin{array}{c} [1\perp\!\!\!\perp 2|\emptyset], [1\perp\!\!\!\perp 3|\emptyset], [1\perp\!\!\!\perp 4|\emptyset], [3\perp\!\!\!\perp 4|\emptyset], [1\perp\!\!\!\perp 3|2], [1\perp\!\!\!\perp 4|2], [3\perp\!\!\!\perp 4|2], [1\perp\!\!\!\perp 2|3], \\ [2\perp\!\!\!\perp 4|3], [1\perp\!\!\!\perp 2|4], [2\perp\!\!\!\perp 3|4], [3\perp\!\!\!\perp 4|12], [2\perp\!\!\!\perp 4|13], [2\perp\!\!\!\perp 3|14], [1\perp\!\!\!\perp 2|34] \end{array} \right\rangle.$$

Next, $I_{\mathbf{SG}}$ has three associated primes of codimension 16. A representative is

$$\left\langle \begin{array}{c} [1\perp\!\!\!\perp 2|\emptyset], [1\perp\!\!\!\perp 3|\emptyset], [2\perp\!\!\!\perp 4|\emptyset], [3\perp\!\!\!\perp 4|\emptyset], [2\perp\!\!\!\perp 4|1], [3\perp\!\!\!\perp 4|1], [1\perp\!\!\!\perp 3|2], [3\perp\!\!\!\perp 4|2], \\ [1\perp\!\!\!\perp 2|3], [2\perp\!\!\!\perp 4|3], [1\perp\!\!\!\perp 2|4], [1\perp\!\!\!\perp 3|4], [3\perp\!\!\!\perp 4|12], [2\perp\!\!\!\perp 4|13], [1\perp\!\!\!\perp 3|24], [1\perp\!\!\!\perp 2|34] \end{array} \right\rangle.$$

Each of the 4316 non-submodular semigraphoids is a $\{0,1\}$-valued point not in $V(I_{\mathcal{A}})$ but in one of the 17 coordinate subspaces corresponding to these primes.

Finally, the last associated prime of $I_{\mathbf{SG}}$ is the toric ideal $I_{\mathcal{A}}$. This ideal has codimension 13 and degree 396. Its minimal generating set consists of 52 binomials. Besides the 24 quadrics (the axiom instances), the Markov basis of $\mathcal{A}$ contains four cubics

$$[2 \perp\!\!\!\perp 3|1] \cdot [3 \perp\!\!\!\perp 4|2] \cdot [1 \perp\!\!\!\perp 3|4] - [3 \perp\!\!\!\perp 4|1] \cdot [1 \perp\!\!\!\perp 3|2] \cdot [2 \perp\!\!\!\perp 3|4],$$

$$[2 \perp\!\!\!\perp 3|1] \cdot [2 \perp\!\!\!\perp 4|3] \cdot [1 \perp\!\!\!\perp 2|4] - [2 \perp\!\!\!\perp 4|1] \cdot [1 \perp\!\!\!\perp 2|3] \cdot [2 \perp\!\!\!\perp 3|4],$$

$$[1 \perp\!\!\!\perp 3|2] \cdot [1 \perp\!\!\!\perp 4|3] \cdot [1 \perp\!\!\!\perp 2|4] - [1 \perp\!\!\!\perp 4|2] \cdot [1 \perp\!\!\!\perp 2|3] \cdot [1 \perp\!\!\!\perp 3|4],$$

$$[2 \perp\!\!\!\perp 4|1] \cdot [3 \perp\!\!\!\perp 4|2] \cdot [1 \perp\!\!\!\perp 4|3] - [3 \perp\!\!\!\perp 4|1] \cdot [1 \perp\!\!\!\perp 4|2] \cdot [2 \perp\!\!\!\perp 4|3],$$

and 24 quartics such as

$$[1 \perp\!\!\!\perp 2|\emptyset] \cdot [3 \perp\!\!\!\perp 4|\emptyset] \cdot [2 \perp\!\!\!\perp 4|13] \cdot [1 \perp\!\!\!\perp 3|24] \; - \; [1 \perp\!\!\!\perp 3|\emptyset] \cdot [2 \perp\!\!\!\perp 4|\emptyset] \cdot [3 \perp\!\!\!\perp 4|12] \cdot [1 \perp\!\!\!\perp 2|34].$$

These cubics and quartics correspond to the non-semigraphoid properties (A.6) and (A.7) in [66].

We now come to the case $n = 5$. It will be a challenge for future commutative algebra software to compute a primary decomposition of the semigraphoid ideal $I_{\mathbf{SG}}$ for $n = 5$. At present we do not know even whether $I_{\mathbf{SG}}$ is radical. Let us therefore focus on the main component of this ideal, namely, the toric ideal $I_{\mathcal{A}}$.

We succeeded in computing its minimal generators, that is, the Markov basis of A. This was accomplished using the software `4ti2` [33], and we now state the result:

Theorem 4.5.3. *The toric ideal $I_{\mathcal{A}}$ representing semigraphoids on $n = 5$ has codimension 54. The reduced Gröbner basis of $I_{\mathcal{A}}$ in the reverse lexicographic term order consists of 958,202 binomials of degree up to 30.*

The minimal Markov basis we found consists of 75, 889 *binomials of degrees ranging from* 2 *to* 12. *They come in* 613 *symmetry classes which are explained in Table 4.1 below. The minimal Markov basis of $I_{\mathcal{A}}$ is not unique. Moreover, the semigroup of $\mathcal{A}$ is not normal.*

degree	Markov basis elements	symmetry classes
2	120	2
3	40	1
4	270	6
5	984	9
6	12,630	126
7	26,280	195
8	26,925	193
9	3,600	33
10	3,420	35
11	240	2
12	1,380	11

Table 4.1: Degrees of Markov basis elements of $I_{\mathcal{A}}$ for $n = 5$; representatives of the 613 symmetry classes are available from `http://www.4ti2.de/articles/data/`

Proof. The Markov basis and the Gröbner basis of $I_{\mathcal{A}}$ were computed with `4ti2` and the above numbers of elements correspond to the output sizes when invoking the functions `markov` and `groebner` in `4ti2`. The Markov basis of $I_{\mathcal{A}}$ is not unique, since the Markov move

$$[2 \perp\!\!\!\perp 3|1]+[1 \perp\!\!\!\perp 3|5]+[3 \perp\!\!\!\perp 4|12]+[3 \perp\!\!\!\perp 5|24]-[3 \perp\!\!\!\perp 5|1]-[2 \perp\!\!\!\perp 3|5]-[1 \perp\!\!\!\perp 3|24]-[3 \perp\!\!\!\perp 4|25] \quad (4.9)$$

is not indispensable. Recall (e.g. from [3]) that a binomial $\mathbf{x}^{\mathbf{g}^+} - \mathbf{x}^{\mathbf{g}^-}$ in the toric ideal $I_{\mathcal{A}}$ is called *indispensable* if

$$\{\mathbf{z} \in \mathbb{N}^{80} : \mathcal{A} \cdot \mathbf{z} = \mathcal{A} \cdot \mathbf{g}^+\} \quad = \quad \{\mathbf{g}^+, \mathbf{g}^-\}.$$

This means that the Markov move $\mathbf{g} = \mathbf{g}^+ - \mathbf{g}^-$ corresponds to a 2-element fiber given by the right-hand side $\mathcal{A} \cdot \mathbf{g}^+$ and consequently, $\mathbf{g}$ must belong to *every* Markov basis of $I_{\mathcal{A}}$.

To see that the Markov move from (4.9) is not indispensable, we enumerate the full fiber of $[2 \perp\!\!\!\perp 3|1] + [1 \perp\!\!\!\perp 3|5] + [3 \perp\!\!\!\perp 4|12] + [3 \perp\!\!\!\perp 5|24]$ using the moves from the computed Markov basis. In addition to the two known elements, we obtain exactly two more elements, namely $[3 \perp\!\!\!\perp 4|1] + [1 \perp\!\!\!\perp 3|5] + [2 \perp\!\!\!\perp 3|14] + [3 \perp\!\!\!\perp 5|24]$ and $[3 \perp\!\!\!\perp 5|1] + [3 \perp\!\!\!\perp 4|5] + [1 \perp\!\!\!\perp 3|24] + [2 \perp\!\!\!\perp 3|45]$. In fact, we obtain a different minimal Markov basis of $I_{\mathcal{A}}$ by replacing the Markov move from (4.9) with the move

$$[3 \perp\!\!\!\perp 4|1] + [1 \perp\!\!\!\perp 3|5] + [2 \perp\!\!\!\perp 3|14] + [3 \perp\!\!\!\perp 5|24] - [3 \perp\!\!\!\perp 5|1] - [3 \perp\!\!\!\perp 4|5] - [1 \perp\!\!\!\perp 3|24] - [2 \perp\!\!\!\perp 3|45].$$

In order to show that the semigroup of $\mathcal{A}$ is not normal, consider the Markov basis move

$$\mathbf{g} \ := \ \big(\alpha + 2 \cdot [2 \perp\!\!\!\perp 3|45]\big) \ - \ \big(\beta + 2 \cdot [4 \perp\!\!\!\perp 5|23]\big) \ \in \ \mathbb{Z}^{80}, \qquad \text{where}$$

$$\alpha = [4 \perp\!\!\!\perp 5|2] + [4 \perp\!\!\!\perp 5|3] + [1 \perp\!\!\!\perp 3|4] + [1 \perp\!\!\!\perp 2|5] + [2 \perp\!\!\!\perp 5|14] + [3 \perp\!\!\!\perp 4|15] + [1 \perp\!\!\!\perp 4|23] + [1 \perp\!\!\!\perp 5|23],$$

$$\beta = [1 \perp\!\!\!\perp 5|2] + [1 \perp\!\!\!\perp 4|3] + [2 \perp\!\!\!\perp 3|4] + [2 \perp\!\!\!\perp 3|5] + [3 \perp\!\!\!\perp 4|12] + [2 \perp\!\!\!\perp 5|13] + [1 \perp\!\!\!\perp 2|45] + [1 \perp\!\!\!\perp 3|45].$$

This lattice vector corresponds to a binomial $\mathbf{x}^{\mathbf{g}^+} - \mathbf{x}^{\mathbf{g}^-}$ which is in the toric ideal $I_{\mathcal{A}}$ and has the property that both of its monomials are not square-free. We verified that $\mathbf{x}^{\mathbf{g}^+} - \mathbf{x}^{\mathbf{g}^-}$ is indispensable for $I_{\mathcal{A}}$ by computing the minimal Hilbert basis (that is, the $\leq$-minimal integer solutions) of the cone

$$\{(\mathbf{z}, u) \in \ \mathbb{R}^{81} : \mathcal{A} \cdot \mathbf{z} - (\mathcal{A} \cdot \mathbf{g}^+) \cdot u = 0, (\mathbf{z}, u) \geq 0\}.$$

This was done using the function `hilbert` of `4ti2` which produced precisely the two expected elements $(\mathbf{g}^+, 1)$ and $(\mathbf{g}^-, 1)$ within a few seconds. Again, one could also use $\mathbf{g}^+$ and the computed Markov basis to enumerate all (two) elements in this fiber.

From our indispensable Markov move $\mathbf{g} = (\alpha + 2 \cdot [2 \perp\!\!\!\perp 3|45]) - (\beta + 2 \cdot [4 \perp\!\!\!\perp 5|23])$, we now construct the imset $\mathbf{b}$ presented in Section 4. We first check that $\mathbf{b}$ was not a combinatorial imset by showing that $\mathcal{A}\mathbf{x} = \mathbf{b}$ has no solutions with non-negative integer coordinates. Using the functions `hilbert` and `rays` of the program `4ti2`, we computed the Hilbert basis and the extreme rays of the cone

$$\left\{(\mathbf{z}, u) \in \mathbb{R}^{81} : \mathcal{A} \cdot \mathbf{z} = \mathbf{b} \cdot u \text{ and } (\mathbf{z}, u) \geq 0\right\}.$$

Both computations quickly finished. They show that this cone has dimension one and is generated by the single vector $(\alpha+\beta, 2)$. Consequently, the only non-negative real solution to $\mathcal{A} \cdot \mathbf{x} = \mathbf{b}$ is $(\alpha+\beta)/2$, which is not an integer solution.

We complete this discussion by presenting a degree 12 Markov basis element, which can easily be proved to be indispensable as well:

$$\begin{gathered}[1 \perp\!\!\!\perp 5|2] + [2 \perp\!\!\!\perp 5|3] + 2 \cdot [2 \perp\!\!\!\perp 3|4] + [1 \perp\!\!\!\perp 2|5] + [3 \perp\!\!\!\perp 4|5] + [3 \perp\!\!\!\perp 4|12] + [3 \perp\!\!\!\perp 4|15] + \\ 2 \cdot [2 \perp\!\!\!\perp 5|134] + [3 \perp\!\!\!\perp 4|5] + [3 \perp\!\!\!\perp 4|12] - [3 \perp\!\!\!\perp 4|2] - [1 \perp\!\!\!\perp 4|3] - [1 \perp\!\!\!\perp 3|4] - [2 \perp\!\!\!\perp 5|4] - \\ 2 \cdot [2 \perp\!\!\!\perp 3|5] - [2 \perp\!\!\!\perp 5|13] - [2 \perp\!\!\!\perp 5|14] - 2 \cdot [3 \perp\!\!\!\perp 4|125] - [1 \perp\!\!\!\perp 5|234] - [1 \perp\!\!\!\perp 2|345].\end{gathered}$$

□

Finally, let us remark that the computations of the 75,889 Markov basis elements and of the 959,202 Gröbner basis elements were both started independently from the given matrix using the `markov` and the `groebner` commands of `4ti2` version 1.3.1. They took about 20 and 17.75 days, respectively, on an AMD Opteron 2.4 GHz CPU running SuSE Linux 10.0. Input and output files are available at `http://www.4ti2.de/articles/data/`.

4.6 The 120 semigraphoid axioms

Here is the list of all 120 semigraphoid axiom for $n = 5$, grouped into triples according to which 3-face of the 5-cube they come from. The two types of brackets specify the non-submodular coarsest semigraphoid Γ which was discussed in Section 4.3.

$$\begin{aligned}
[3 \perp\!\!\!\perp 5|12] + [\![3 \perp\!\!\!\perp 4|125]\!] &= [3 \perp\!\!\!\perp 4|12] + [\![3 \perp\!\!\!\perp 5|124]\!] \\
[2 \perp\!\!\!\perp 5|13] + [\![2 \perp\!\!\!\perp 4|135]\!] &= [2 \perp\!\!\!\perp 4|13] + [\![2 \perp\!\!\!\perp 5|134]\!] \\
[4 \perp\!\!\!\perp 5|12] + [\![3 \perp\!\!\!\perp 4|125]\!] &= [3 \perp\!\!\!\perp 4|12] + [\![4 \perp\!\!\!\perp 5|123]\!] \\
[4 \perp\!\!\!\perp 5|13] + [\![2 \perp\!\!\!\perp 4|135]\!] &= [2 \perp\!\!\!\perp 4|13] + [\![4 \perp\!\!\!\perp 5|123]\!] \\
[4 \perp\!\!\!\perp 5|12] + [\![3 \perp\!\!\!\perp 5|124]\!] &= [3 \perp\!\!\!\perp 5|12] + [\![4 \perp\!\!\!\perp 5|123]\!] \\
[4 \perp\!\!\!\perp 5|13] + [\![2 \perp\!\!\!\perp 5|134]\!] &= [2 \perp\!\!\!\perp 5|13] + [\![4 \perp\!\!\!\perp 5|123]\!]
\end{aligned}$$

$$\begin{aligned}
[2 \perp\!\!\!\perp 3|14] + [\![2 \perp\!\!\!\perp 5|134]\!] &= [2 \perp\!\!\!\perp 5|14] + [\![2 \perp\!\!\!\perp 3|145]\!] \\
[\![2 \perp\!\!\!\perp 4|15]\!] + [\![2 \perp\!\!\!\perp 3|145]\!] &= [\![2 \perp\!\!\!\perp 3|15]\!] + [\![2 \perp\!\!\!\perp 4|135]\!] \\
[2 \perp\!\!\!\perp 5|14] + [\![3 \perp\!\!\!\perp 5|124]\!] &= [3 \perp\!\!\!\perp 5|14] + [\![2 \perp\!\!\!\perp 5|134]\!] \\
[\![3 \perp\!\!\!\perp 4|15]\!] + [\![2 \perp\!\!\!\perp 3|145]\!] &= [\![2 \perp\!\!\!\perp 3|15]\!] + [\![3 \perp\!\!\!\perp 4|125]\!] \\
[3 \perp\!\!\!\perp 5|14] + [\![2 \perp\!\!\!\perp 3|145]\!] &= [2 \perp\!\!\!\perp 3|14] + [\![3 \perp\!\!\!\perp 5|124]\!] \\
[\![3 \perp\!\!\!\perp 4|15]\!] + [\![2 \perp\!\!\!\perp 4|135]\!] &= [\![2 \perp\!\!\!\perp 4|15]\!] + [\![3 \perp\!\!\!\perp 4|125]\!]
\end{aligned}$$

$$\begin{aligned}
[1 \perp\!\!\!\perp 5|23] + [\![1 \perp\!\!\!\perp 4|235]\!] &= [1 \perp\!\!\!\perp 4|23] + [\![1 \perp\!\!\!\perp 5|234]\!] \\
[1 \perp\!\!\!\perp 3|24] + [\![3 \perp\!\!\!\perp 5|124]\!] &= [3 \perp\!\!\!\perp 5|24] + [\![1 \perp\!\!\!\perp 3|245]\!] \\
[4 \perp\!\!\!\perp 5|23] + [\![1 \perp\!\!\!\perp 4|235]\!] &= [1 \perp\!\!\!\perp 4|23] + [\![4 \perp\!\!\!\perp 5|123]\!] \\
[1 \perp\!\!\!\perp 5|24] + [\![1 \perp\!\!\!\perp 3|245]\!] &= [1 \perp\!\!\!\perp 3|24] + [\![1 \perp\!\!\!\perp 5|234]\!] \\
[4 \perp\!\!\!\perp 5|23] + [\![1 \perp\!\!\!\perp 5|234]\!] &= [1 \perp\!\!\!\perp 5|23] + [\![4 \perp\!\!\!\perp 5|123]\!] \\
[1 \perp\!\!\!\perp 5|24] + [\![3 \perp\!\!\!\perp 5|124]\!] &= [3 \perp\!\!\!\perp 5|24] + [\![1 \perp\!\!\!\perp 5|234]\!]
\end{aligned}$$

$$
\begin{aligned}
[1 \perp\!\!\!\perp 3|25] + [\![1 \perp\!\!\!\perp 4|235]\!] &= [1 \perp\!\!\!\perp 4|25] + [\![1 \perp\!\!\!\perp 3|245]\!] \\
[1 \perp\!\!\!\perp 2|34] + [\![2 \perp\!\!\!\perp 5|134]\!] &= [2 \perp\!\!\!\perp 5|34] + [\![1 \perp\!\!\!\perp 2|345]\!] \\
[1 \perp\!\!\!\perp 3|25] + [\![3 \perp\!\!\!\perp 4|125]\!] &= [3 \perp\!\!\!\perp 4|25] + [\![1 \perp\!\!\!\perp 3|245]\!] \\
[1 \perp\!\!\!\perp 5|34] + [\![1 \perp\!\!\!\perp 2|345]\!] &= [1 \perp\!\!\!\perp 2|34] + [\![1 \perp\!\!\!\perp 5|234]\!] \\
[3 \perp\!\!\!\perp 4|25] + [\![1 \perp\!\!\!\perp 4|235]\!] &= [1 \perp\!\!\!\perp 4|25] + [\![3 \perp\!\!\!\perp 4|125]\!] \\
[1 \perp\!\!\!\perp 5|34] + [\![2 \perp\!\!\!\perp 5|134]\!] &= [2 \perp\!\!\!\perp 5|34] + [\![1 \perp\!\!\!\perp 5|234]\!]
\end{aligned}
$$

$$
\begin{aligned}
[1 \perp\!\!\!\perp 2|35] + [\![1 \perp\!\!\!\perp 4|235]\!] &= [1 \perp\!\!\!\perp 4|35] + [\![1 \perp\!\!\!\perp 2|345]\!] \\
[1 \perp\!\!\!\perp 2|45] + [\![1 \perp\!\!\!\perp 3|245]\!] &= [1 \perp\!\!\!\perp 3|45] + [\![1 \perp\!\!\!\perp 2|345]\!] \\
[1 \perp\!\!\!\perp 2|35] + [\![2 \perp\!\!\!\perp 4|135]\!] &= [2 \perp\!\!\!\perp 4|35] + [\![1 \perp\!\!\!\perp 2|345]\!] \\
[1 \perp\!\!\!\perp 3|45] + [\![2 \perp\!\!\!\perp 3|145]\!] &= [2 \perp\!\!\!\perp 3|45] + [\![1 \perp\!\!\!\perp 3|245]\!] \\
[1 \perp\!\!\!\perp 4|35] + [\![2 \perp\!\!\!\perp 4|135]\!] &= [2 \perp\!\!\!\perp 4|35] + [\![1 \perp\!\!\!\perp 4|235]\!] \\
[2 \perp\!\!\!\perp 3|45] + [\![1 \perp\!\!\!\perp 2|345]\!] &= [1 \perp\!\!\!\perp 2|45] + [\![2 \perp\!\!\!\perp 3|145]\!]
\end{aligned}
$$

$$
\begin{aligned}
[\![2 \perp\!\!\!\perp 4|1]\!] + [3 \perp\!\!\!\perp 4|12] &= [\![3 \perp\!\!\!\perp 4|1]\!] + [2 \perp\!\!\!\perp 4|13] \\
[\![2 \perp\!\!\!\perp 3|1]\!] + [3 \perp\!\!\!\perp 5|12] &= [3 \perp\!\!\!\perp 5|1] + [\![2 \perp\!\!\!\perp 3|15]\!] \\
[\![2 \perp\!\!\!\perp 4|1]\!] + [2 \perp\!\!\!\perp 3|14] &= [\![2 \perp\!\!\!\perp 3|1]\!] + [2 \perp\!\!\!\perp 4|13] \\
[\![2 \perp\!\!\!\perp 3|1]\!] + [2 \perp\!\!\!\perp 5|13] &= [2 \perp\!\!\!\perp 5|1] + [\![2 \perp\!\!\!\perp 3|15]\!] \\
[\![3 \perp\!\!\!\perp 4|1]\!] + [2 \perp\!\!\!\perp 3|14] &= [\![2 \perp\!\!\!\perp 3|1]\!] + [3 \perp\!\!\!\perp 4|12] \\
[3 \perp\!\!\!\perp 5|1] + [2 \perp\!\!\!\perp 5|13] &= [2 \perp\!\!\!\perp 5|1] + [3 \perp\!\!\!\perp 5|12]
\end{aligned}
$$

$$\begin{aligned}
[\![2 \perp\!\!\!\perp 4|1]\!] + [4 \perp\!\!\!\perp 5|12] &= [4 \perp\!\!\!\perp 5|1] + [\![2 \perp\!\!\!\perp 4|15]\!] \\
[3 \perp\!\!\!\perp 5|1] + [4 \perp\!\!\!\perp 5|13] &= [4 \perp\!\!\!\perp 5|1] + [3 \perp\!\!\!\perp 5|14] \\
[2 \perp\!\!\!\perp 5|1] + [\![2 \perp\!\!\!\perp 4|15]\!] &= [\![2 \perp\!\!\!\perp 4|1]\!] + [2 \perp\!\!\!\perp 5|14] \\
[3 \perp\!\!\!\perp 5|1] + [\![3 \perp\!\!\!\perp 4|15]\!] &= [\![3 \perp\!\!\!\perp 4|1]\!] + [3 \perp\!\!\!\perp 5|14] \\
[4 \perp\!\!\!\perp 5|1] + [2 \perp\!\!\!\perp 5|14] &= [2 \perp\!\!\!\perp 5|1] + [4 \perp\!\!\!\perp 5|12] \\
[4 \perp\!\!\!\perp 5|1] + [\![3 \perp\!\!\!\perp 4|15]\!] &= [\![3 \perp\!\!\!\perp 4|1]\!] + [4 \perp\!\!\!\perp 5|13]
\end{aligned}$$

$$\begin{aligned}
[\![1 \perp\!\!\!\perp 3|2]\!] + [1 \perp\!\!\!\perp 4|23] &= [\![1 \perp\!\!\!\perp 4|2]\!] + [1 \perp\!\!\!\perp 3|24] \\
[1 \perp\!\!\!\perp 5|2] + [1 \perp\!\!\!\perp 3|25] &= [\![1 \perp\!\!\!\perp 3|2]\!] + [1 \perp\!\!\!\perp 5|23] \\
[\![1 \perp\!\!\!\perp 4|2]\!] + [3 \perp\!\!\!\perp 4|12] &= [\![3 \perp\!\!\!\perp 4|2]\!] + [1 \perp\!\!\!\perp 4|23] \\
[\![3 \perp\!\!\!\perp 5|2]\!] + [1 \perp\!\!\!\perp 5|23] &= [1 \perp\!\!\!\perp 5|2] + [3 \perp\!\!\!\perp 5|12] \\
[\![3 \perp\!\!\!\perp 4|2]\!] + [1 \perp\!\!\!\perp 3|24] &= [\![1 \perp\!\!\!\perp 3|2]\!] + [3 \perp\!\!\!\perp 4|12] \\
[\![3 \perp\!\!\!\perp 5|2]\!] + [1 \perp\!\!\!\perp 3|25] &= [\![1 \perp\!\!\!\perp 3|2]\!] + [3 \perp\!\!\!\perp 5|12]
\end{aligned}$$

$$\begin{aligned}
[\![1 \perp\!\!\!\perp 4|2]\!] + [4 \perp\!\!\!\perp 5|12] &= [\![4 \perp\!\!\!\perp 5|2]\!] + [1 \perp\!\!\!\perp 4|25] \\
[\![3 \perp\!\!\!\perp 5|2]\!] + [4 \perp\!\!\!\perp 5|23] &= [\![4 \perp\!\!\!\perp 5|2]\!] + [3 \perp\!\!\!\perp 5|24] \\
[\![1 \perp\!\!\!\perp 4|2]\!] + [1 \perp\!\!\!\perp 5|24] &= [1 \perp\!\!\!\perp 5|2] + [1 \perp\!\!\!\perp 4|25] \\
[\![3 \perp\!\!\!\perp 5|2]\!] + [3 \perp\!\!\!\perp 4|25] &= [\![3 \perp\!\!\!\perp 4|2]\!] + [3 \perp\!\!\!\perp 5|24] \\
[1 \perp\!\!\!\perp 5|2] + [4 \perp\!\!\!\perp 5|12] &= [\![4 \perp\!\!\!\perp 5|2]\!] + [1 \perp\!\!\!\perp 5|24] \\
[\![4 \perp\!\!\!\perp 5|2]\!] + [3 \perp\!\!\!\perp 4|25] &= [\![3 \perp\!\!\!\perp 4|2]\!] + [4 \perp\!\!\!\perp 5|23]
\end{aligned}$$

$$
\begin{aligned}
\llbracket 1 \perp\!\!\!\perp 4|3 \rrbracket + [2 \perp\!\!\!\perp 4|13] &= \llbracket 2 \perp\!\!\!\perp 4|3 \rrbracket + [1 \perp\!\!\!\perp 4|23] \\
\llbracket 1 \perp\!\!\!\perp 2|3 \rrbracket + [2 \perp\!\!\!\perp 5|13] &= \llbracket 2 \perp\!\!\!\perp 5|3 \rrbracket + [1 \perp\!\!\!\perp 2|35] \\
\llbracket 1 \perp\!\!\!\perp 4|3 \rrbracket + [1 \perp\!\!\!\perp 2|34] &= \llbracket 1 \perp\!\!\!\perp 2|3 \rrbracket + [1 \perp\!\!\!\perp 4|23] \\
[1 \perp\!\!\!\perp 5|3] + [1 \perp\!\!\!\perp 2|35] &= \llbracket 1 \perp\!\!\!\perp 2|3 \rrbracket + [1 \perp\!\!\!\perp 5|23] \\
\llbracket 2 \perp\!\!\!\perp 4|3 \rrbracket + [1 \perp\!\!\!\perp 2|34] &= \llbracket 1 \perp\!\!\!\perp 2|3 \rrbracket + [2 \perp\!\!\!\perp 4|13] \\
\llbracket 2 \perp\!\!\!\perp 5|3 \rrbracket + [1 \perp\!\!\!\perp 5|23] &= [1 \perp\!\!\!\perp 5|3] + [2 \perp\!\!\!\perp 5|13]
\end{aligned}
$$

$$
\begin{aligned}
\llbracket 1 \perp\!\!\!\perp 4|3 \rrbracket + [4 \perp\!\!\!\perp 5|13] &= \llbracket 4 \perp\!\!\!\perp 5|3 \rrbracket + [1 \perp\!\!\!\perp 4|35] \\
\llbracket 2 \perp\!\!\!\perp 4|3 \rrbracket + [4 \perp\!\!\!\perp 5|23] &= \llbracket 4 \perp\!\!\!\perp 5|3 \rrbracket + [2 \perp\!\!\!\perp 4|35] \\
\llbracket 1 \perp\!\!\!\perp 4|3 \rrbracket + [1 \perp\!\!\!\perp 5|34] &= [1 \perp\!\!\!\perp 5|3] + [1 \perp\!\!\!\perp 4|35] \\
\llbracket 2 \perp\!\!\!\perp 5|3 \rrbracket + [2 \perp\!\!\!\perp 4|35] &= \llbracket 2 \perp\!\!\!\perp 4|3 \rrbracket + [2 \perp\!\!\!\perp 5|34] \\
\llbracket 4 \perp\!\!\!\perp 5|3 \rrbracket + [1 \perp\!\!\!\perp 5|34] &= [1 \perp\!\!\!\perp 5|3] + [4 \perp\!\!\!\perp 5|13] \\
\llbracket 4 \perp\!\!\!\perp 5|3 \rrbracket + [2 \perp\!\!\!\perp 5|34] &= \llbracket 2 \perp\!\!\!\perp 5|3 \rrbracket + [4 \perp\!\!\!\perp 5|23]
\end{aligned}
$$

$$
\begin{aligned}
\llbracket 1 \perp\!\!\!\perp 2|4 \rrbracket + [2 \perp\!\!\!\perp 3|14] &= \llbracket 2 \perp\!\!\!\perp 3|4 \rrbracket + [1 \perp\!\!\!\perp 2|34] \\
\llbracket 1 \perp\!\!\!\perp 2|4 \rrbracket + [2 \perp\!\!\!\perp 5|14] &= \llbracket 2 \perp\!\!\!\perp 5|4 \rrbracket + [1 \perp\!\!\!\perp 2|45] \\
\llbracket 1 \perp\!\!\!\perp 2|4 \rrbracket + [1 \perp\!\!\!\perp 3|24] &= \llbracket 1 \perp\!\!\!\perp 3|4 \rrbracket + [1 \perp\!\!\!\perp 2|34] \\
\llbracket 1 \perp\!\!\!\perp 2|4 \rrbracket + [1 \perp\!\!\!\perp 5|24] &= [1 \perp\!\!\!\perp 5|4] + [1 \perp\!\!\!\perp 2|45] \\
\llbracket 1 \perp\!\!\!\perp 3|4 \rrbracket + [2 \perp\!\!\!\perp 3|14] &= \llbracket 2 \perp\!\!\!\perp 3|4 \rrbracket + [1 \perp\!\!\!\perp 3|24] \\
[1 \perp\!\!\!\perp 5|4] + [2 \perp\!\!\!\perp 5|14] &= \llbracket 2 \perp\!\!\!\perp 5|4 \rrbracket + [1 \perp\!\!\!\perp 5|24]
\end{aligned}
$$

$$
\begin{aligned}
[\![1 \perp\!\!\!\perp 3|4]\!] + [1 \perp\!\!\!\perp 5|34] &= [1 \perp\!\!\!\perp 5|4] + [1 \perp\!\!\!\perp 3|45] \\
[\![2 \perp\!\!\!\perp 3|4]\!] + [3 \perp\!\!\!\perp 5|24] &= [\![3 \perp\!\!\!\perp 5|4]\!] + [2 \perp\!\!\!\perp 3|45] \\
[\![3 \perp\!\!\!\perp 5|4]\!] + [1 \perp\!\!\!\perp 5|34] &= [1 \perp\!\!\!\perp 5|4] + [3 \perp\!\!\!\perp 5|14] \\
[\![2 \perp\!\!\!\perp 3|4]\!] + [2 \perp\!\!\!\perp 5|34] &= [\![2 \perp\!\!\!\perp 5|4]\!] + [2 \perp\!\!\!\perp 3|45] \\
[\![3 \perp\!\!\!\perp 5|4]\!] + [1 \perp\!\!\!\perp 3|45] &= [\![1 \perp\!\!\!\perp 3|4]\!] + [3 \perp\!\!\!\perp 5|14] \\
[\![2 \perp\!\!\!\perp 5|4]\!] + [3 \perp\!\!\!\perp 5|24] &= [\![3 \perp\!\!\!\perp 5|4]\!] + [2 \perp\!\!\!\perp 5|34]
\end{aligned}
$$

$$
\begin{aligned}
[1 \perp\!\!\!\perp 2|5] + [\![2 \perp\!\!\!\perp 3|15]\!] &= [\![2 \perp\!\!\!\perp 3|5]\!] + [1 \perp\!\!\!\perp 2|35] \\
[1 \perp\!\!\!\perp 2|5] + [\![2 \perp\!\!\!\perp 4|15]\!] &= [\![2 \perp\!\!\!\perp 4|5]\!] + [1 \perp\!\!\!\perp 2|45] \\
[1 \perp\!\!\!\perp 2|5] + [1 \perp\!\!\!\perp 3|25] &= [1 \perp\!\!\!\perp 3|5] + [1 \perp\!\!\!\perp 2|35] \\
[1 \perp\!\!\!\perp 2|5] + [1 \perp\!\!\!\perp 4|25] &= [1 \perp\!\!\!\perp 4|5] + [1 \perp\!\!\!\perp 2|45] \\
[1 \perp\!\!\!\perp 3|5] + [\![2 \perp\!\!\!\perp 3|15]\!] &= [\![2 \perp\!\!\!\perp 3|5]\!] + [1 \perp\!\!\!\perp 3|25] \\
[1 \perp\!\!\!\perp 4|5] + [\![2 \perp\!\!\!\perp 4|15]\!] &= [\![2 \perp\!\!\!\perp 4|5]\!] + [1 \perp\!\!\!\perp 4|25]
\end{aligned}
$$

$$
\begin{aligned}
[1 \perp\!\!\!\perp 3|5] + [\![3 \perp\!\!\!\perp 4|15]\!] &= [\![3 \perp\!\!\!\perp 4|5]\!] + [1 \perp\!\!\!\perp 3|45] \\
[\![2 \perp\!\!\!\perp 3|5]\!] + [2 \perp\!\!\!\perp 4|35] &= [\![2 \perp\!\!\!\perp 4|5]\!] + [2 \perp\!\!\!\perp 3|45] \\
[1 \perp\!\!\!\perp 3|5] + [1 \perp\!\!\!\perp 4|35] &= [1 \perp\!\!\!\perp 4|5] + [1 \perp\!\!\!\perp 3|45] \\
[\![2 \perp\!\!\!\perp 4|5]\!] + [3 \perp\!\!\!\perp 4|25] &= [\![3 \perp\!\!\!\perp 4|5]\!] + [2 \perp\!\!\!\perp 4|35] \\
[\![3 \perp\!\!\!\perp 4|5]\!] + [1 \perp\!\!\!\perp 4|35] &= [1 \perp\!\!\!\perp 4|5] + [\![3 \perp\!\!\!\perp 4|15]\!] \\
[\![3 \perp\!\!\!\perp 4|5]\!] + [2 \perp\!\!\!\perp 3|45] &= [\![2 \perp\!\!\!\perp 3|5]\!] + [3 \perp\!\!\!\perp 4|25]
\end{aligned}
$$

$$\begin{aligned}
\llbracket 1 \perp\!\!\!\perp 2 | \rrbracket + \llbracket 2 \perp\!\!\!\perp 3 | 1 \rrbracket &= \llbracket 2 \perp\!\!\!\perp 3 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 2 | 3 \rrbracket \\
\llbracket 1 \perp\!\!\!\perp 2 | \rrbracket + \llbracket 2 \perp\!\!\!\perp 4 | 1 \rrbracket &= \llbracket 2 \perp\!\!\!\perp 4 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 2 | 4 \rrbracket \\
\llbracket 1 \perp\!\!\!\perp 3 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 2 | 3 \rrbracket &= \llbracket 1 \perp\!\!\!\perp 2 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 3 | 2 \rrbracket \\
\llbracket 1 \perp\!\!\!\perp 2 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 4 | 2 \rrbracket &= \llbracket 1 \perp\!\!\!\perp 4 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 2 | 4 \rrbracket \\
\llbracket 2 \perp\!\!\!\perp 3 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 3 | 2 \rrbracket &= \llbracket 1 \perp\!\!\!\perp 3 | \rrbracket + \llbracket 2 \perp\!\!\!\perp 3 | 1 \rrbracket \\
\llbracket 1 \perp\!\!\!\perp 4 | \rrbracket + \llbracket 2 \perp\!\!\!\perp 4 | 1 \rrbracket &= \llbracket 2 \perp\!\!\!\perp 4 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 4 | 2 \rrbracket
\end{aligned}$$

$$\begin{aligned}
\llbracket 1 \perp\!\!\!\perp 2 | \rrbracket + [2 \perp\!\!\!\perp 5 | 1] &= \llbracket 2 \perp\!\!\!\perp 5 | \rrbracket + [1 \perp\!\!\!\perp 2 | 5] \\
\llbracket 1 \perp\!\!\!\perp 4 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 3 | 4 \rrbracket &= \llbracket 1 \perp\!\!\!\perp 3 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 4 | 3 \rrbracket \\
\llbracket 1 \perp\!\!\!\perp 2 | \rrbracket + [1 \perp\!\!\!\perp 5 | 2] &= \llbracket 1 \perp\!\!\!\perp 5 | \rrbracket + [1 \perp\!\!\!\perp 2 | 5] \\
\llbracket 3 \perp\!\!\!\perp 4 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 4 | 3 \rrbracket &= \llbracket 1 \perp\!\!\!\perp 4 | \rrbracket + \llbracket 3 \perp\!\!\!\perp 4 | 1 \rrbracket \\
\llbracket 1 \perp\!\!\!\perp 5 | \rrbracket + [2 \perp\!\!\!\perp 5 | 1] &= \llbracket 2 \perp\!\!\!\perp 5 | \rrbracket + [1 \perp\!\!\!\perp 5 | 2] \\
\llbracket 3 \perp\!\!\!\perp 4 | \rrbracket + \llbracket 1 \perp\!\!\!\perp 3 | 4 \rrbracket &= \llbracket 1 \perp\!\!\!\perp 3 | \rrbracket + \llbracket 3 \perp\!\!\!\perp 4 | 1 \rrbracket
\end{aligned}$$

$$\begin{aligned}
\llbracket 1 \perp\!\!\!\perp 3 | \rrbracket + [3 \perp\!\!\!\perp 5 | 1] &= \llbracket 3 \perp\!\!\!\perp 5 | \rrbracket + [1 \perp\!\!\!\perp 3 | 5] \\
\llbracket 1 \perp\!\!\!\perp 4 | \rrbracket + [1 \perp\!\!\!\perp 5 | 4] &= \llbracket 1 \perp\!\!\!\perp 5 | \rrbracket + [1 \perp\!\!\!\perp 4 | 5] \\
\llbracket 1 \perp\!\!\!\perp 5 | \rrbracket + [3 \perp\!\!\!\perp 5 | 1] &= \llbracket 3 \perp\!\!\!\perp 5 | \rrbracket + [1 \perp\!\!\!\perp 5 | 3] \\
\llbracket 4 \perp\!\!\!\perp 5 | \rrbracket + [1 \perp\!\!\!\perp 5 | 4] &= \llbracket 1 \perp\!\!\!\perp 5 | \rrbracket + [4 \perp\!\!\!\perp 5 | 1] \\
\llbracket 1 \perp\!\!\!\perp 5 | \rrbracket + [1 \perp\!\!\!\perp 3 | 5] &= \llbracket 1 \perp\!\!\!\perp 3 | \rrbracket + [1 \perp\!\!\!\perp 5 | 3] \\
\llbracket 4 \perp\!\!\!\perp 5 | \rrbracket + [1 \perp\!\!\!\perp 4 | 5] &= \llbracket 1 \perp\!\!\!\perp 4 | \rrbracket + [4 \perp\!\!\!\perp 5 | 1]
\end{aligned}$$

$$\begin{aligned}
[\![2 \perp\!\!\!\perp 4|]\!] + [\![2 \perp\!\!\!\perp 3|4]\!] &= [\![2 \perp\!\!\!\perp 3|]\!] + [\![2 \perp\!\!\!\perp 4|3]\!] \\
[\![2 \perp\!\!\!\perp 3|]\!] + [\![2 \perp\!\!\!\perp 5|3]\!] &= [\![2 \perp\!\!\!\perp 5|]\!] + [\![2 \perp\!\!\!\perp 3|5]\!] \\
[\![3 \perp\!\!\!\perp 4|]\!] + [\![2 \perp\!\!\!\perp 4|3]\!] &= [\![2 \perp\!\!\!\perp 4|]\!] + [\![3 \perp\!\!\!\perp 4|2]\!] \\
[\![2 \perp\!\!\!\perp 5|]\!] + [\![3 \perp\!\!\!\perp 5|2]\!] &= [\![3 \perp\!\!\!\perp 5|]\!] + [\![2 \perp\!\!\!\perp 5|3]\!] \\
[\![3 \perp\!\!\!\perp 4|]\!] + [\![2 \perp\!\!\!\perp 3|4]\!] &= [\![2 \perp\!\!\!\perp 3|]\!] + [\![3 \perp\!\!\!\perp 4|2]\!] \\
[\![3 \perp\!\!\!\perp 5|]\!] + [\![2 \perp\!\!\!\perp 3|5]\!] &= [\![2 \perp\!\!\!\perp 3|]\!] + [\![3 \perp\!\!\!\perp 5|2]\!]
\end{aligned}$$

$$\begin{aligned}
[\![2 \perp\!\!\!\perp 4|]\!] + [\![4 \perp\!\!\!\perp 5|2]\!] &= [\![4 \perp\!\!\!\perp 5|]\!] + [\![2 \perp\!\!\!\perp 4|5]\!] \\
[\![3 \perp\!\!\!\perp 4|]\!] + [\![4 \perp\!\!\!\perp 5|3]\!] &= [\![4 \perp\!\!\!\perp 5|]\!] + [\![3 \perp\!\!\!\perp 4|5]\!] \\
[\![2 \perp\!\!\!\perp 5|]\!] + [\![4 \perp\!\!\!\perp 5|2]\!] &= [\![4 \perp\!\!\!\perp 5|]\!] + [\![2 \perp\!\!\!\perp 5|4]\!] \\
[\![3 \perp\!\!\!\perp 4|]\!] + [\![3 \perp\!\!\!\perp 5|4]\!] &= [\![3 \perp\!\!\!\perp 5|]\!] + [\![3 \perp\!\!\!\perp 4|5]\!] \\
[\![2 \perp\!\!\!\perp 5|]\!] + [\![2 \perp\!\!\!\perp 4|5]\!] &= [\![2 \perp\!\!\!\perp 4|]\!] + [\![2 \perp\!\!\!\perp 5|4]\!] \\
[\![3 \perp\!\!\!\perp 5|]\!] + [\![4 \perp\!\!\!\perp 5|3]\!] &= [\![4 \perp\!\!\!\perp 5|]\!] + [\![3 \perp\!\!\!\perp 5|4]\!]
\end{aligned}$$

Chapter 5

Toric algebra of conditional probability

5.1 Relations among conditional probabilities

In 1974, Julian Besag [5] studied the "unobvious and highly restrictive consistency conditions" among conditional probabilities. The problem of discovering these conditions is interesting both in general (i.e. for any conditional probability distribution) and after fixing a particular statistical model. In this chapter we give an answer in the discrete case to the question *What conditions must a set of conditional probabilities satisfy in order to be compatible with some joint distribution?*

Let $\Omega = \{1, \ldots, m\}$ be a finite set of singleton events, and let $p = (p_1, \ldots, p_m)$ be a probability distribution on them. Let $\mathscr{E}$ be a set of observable events which will be conditioned on, each a set of at least 2 singleton events. Then for events $I \subset J$ in $\mathscr{E}$, we can assign conditional probabilities for the chance of I given J, denoted $p_{I|J}$. Settling Besag's question then becomes a matter of determining the *relations* that must hold among the quantities $p_{I|J}$. Since there are in general infinitely many such relations, we would like to organize them into an ideal and provide a nice basis for that ideal.

Work on the subject of relations among conditional probabilities has primarily focused on the case of random variables (see Section 5.5), i.e. where the events in $\mathscr{E}$ correspond to observing the states of a subset of n random variables. An ideal of relations among conditional probabilities could be thought of as a generalization of the most basic result on conditional probability, Bayes' rule (Section 5.6). The Hammersley-Clifford theorem provides a partial answer to this question in the case of undirected graphical models with strictly positive probability distributions [5]. These are not the same graphs as those depicted in this chapter. Arnold et. al. [4] develop the theory for both discrete and continuous random variables, particularly in the case of two random variables. Slavkovic and Sullivant [61] consider the case of compatible full conditionals. A little more detail on these will be given after we have developed the necessary terminology in Section 5.5. Matúš [45] considers the case of m discrete events where $\mathscr{E} = \{I \subset [m] : |I| \geq 2\}$ (all possible compound events). He gives a geometric characterization of the space of conditional probabilities through projection to the case $\mathscr{E} = \{I \subset [m] : |I| = 2\}$.

This chapter is organized as follows. In Section 5.2, we introduce some necessary definitions. In Section 5.3, we give compatibility conditions in the general case of m events in a discrete probability space, with any set $\mathscr{E}$ of conditioned-upon events. These conditions come in the form of a universal Gröbner basis, which makes them particularly useful for computations. They may be specialized to the random variable case, the purely conditional case, and other special cases simply by changing $\mathscr{E}$. In earlier chapters, we have seen that permutohedra and generalized permutohedra play a central role in the geometry of conditional independence; the same is true of conditional probability. The geometric results of Matúš [45] map the space of conditional probability distributions (Definition 5.2.1) for all possible conditioned events $\mathscr{E} = \{I \subset [m] : |I| \geq 2\}$ onto the permutohedron $\mathbf{P}_{m-1}$. See Figure 2.3 for a diagram of the 3-dimensional permutohedron. In Section 5.4, we will discuss how to extend this result to general $\mathscr{E}$, in which case we obtain generalized permutohedra as the image. This will be accomplished using a version of the moment map of toric geometry (Theorem 1.2.3). In Section 5.5, we discuss how to specialize our results to the case of n random variables, including as an example how to recover a relation of Besag [5]. Finally, in Section 5.6 we use this specialization to explain the relationship of Bayes' rule to our constructions.

5.2 Conditional probability distributions

Let $\mathscr{E}$ be a collection of subsets I, with $|I| \geq 2$, of $[m] = \{1, \ldots, m\}$. Let $\mathbb{C}[\mathscr{E}]$ denote the *event algebra*, the polynomial ring with indeterminates $p_{i|I}$ for all $I \in \mathscr{E}$ and $i \in I$. Then

$$\|\mathscr{E}\| = \sum_{I \in \mathscr{E}} |I|$$

is the number of variables of $\mathbb{C}[\mathscr{E}]$. We write p_i for $p_{i|[m]}$ when $[m] \in \mathscr{E}$. The unknowns of $\mathbb{C}[\mathscr{E}]$ are meant to represent conditional probabilities, as we now explain. The set $\{1, \ldots, m\}$ indexes the m disjoint events, and a point $(p_1, \ldots, p_m) \in \mathbb{R}^m_{\geq 0}$ with $\sum_j p_j = 1$ represents a probability distribution on these events. When $p_j > 0$ for all j, the *conditional probability* of event i given event I containing it is

$$p_{i|I} = \frac{p_i}{\sum_{j \in I} p_j}. \tag{5.1}$$

To extend this notion to the case $P(I) = \sum_{j\in I} p_j = 0$, and to be able to deal with multiple conditioning sets, we make the following standard definition [7], considered in this form by Matúš [45].

Definition 5.2.1. A *conditional probability distribution* for $\mathscr{E}$ is a point $(p_{i|I} \;:\; i \in I \in \mathscr{E}) \in \mathbb{R}_{\geq 0}^{\|\mathscr{E}\|}$ such that for all $J, K \in \mathscr{E}$ with $J \subset K$,

(i) $\sum_{i\in J} p_{i|J} = 1$

(ii) for all $i \in J$, $p_{i|K} = p_{i|J} \sum_{j\in J} p_{j|K}$.

Observe that (ii) is a relative version of (5.1), as (5.1) follows from (ii) with $K = [m]$, $J = I$, and $\sum_{i\in I} p_i \neq 0$. If on the other hand $\sum_{j\in J} p_{j|K} = 0$, the whole probability simplex $\Delta_J := \{(p_{j|J})_{j\in J} : p_{j|J} \geq 0, \sum_{j\in J} p_{j|J} = 1\}$ satisfies the definition. This freedom is known in probability theory as *versions of conditional probability* [7]. In algebraic geometry, this corresponds to the notion of a blow-up, and the simplex Δ_J to the exceptional divisor. Analogously to the projective version of probability in Section 1.3.1, where we replaced the requirement that probabilities $p_1, \ldots, p_m$ sum to one with viewing them as coordinates of a point in projective space, we now define a multihomogeneous version of Definition 5.2.1. Now, a conditional probability distribution is represented by a point in the product of projective spaces. This product has one $\mathbb{P}^{|I|-1}$ for each event $I \in \mathscr{E}$ which is conditioned upon, and each factor space $\mathbb{P}^{|I|-1}$ is equipped with homogeneous coordinates $(p_{i_1|I} : \cdots : p_{i_{|I|}|I})$.

Definition 5.2.2. A *projective conditional probability distribution* for $\mathscr{E}$ is a point $\mathbf{p} = ((p_{i_1|I} : \cdots : p_{i_{|I|}|I}), I \in \mathscr{E})$ inside $\prod_{I\in\mathscr{E}} \mathbb{P}^{|I|-1}$ such that for all $J, K \in \mathscr{E}$ and $i \in J \subset K$,

$$(\sum_{j\in J} p_{j|J}) p_{i|K} = p_{i|J} (\sum_{j\in J} p_{j|K})$$

Definition 5.2.2 specifies the following ideal in the event algebra $\mathbb{C}[\mathscr{E}]$:

$$J_{\mathscr{E}} = \langle (\sum_{j\in J} p_{j|J}) p_{i|K} - p_{i|J} (\sum_{j\in J} p_{j|K}) \ : \ J, K \in \mathscr{E}, \ i \in J \subset K \rangle.$$

This ideal consists of all polynomial relations that a point $P = (p_{i|I})$ in $\prod_{I\in\mathscr{E}} \mathbb{P}^{|I|-1}$ must satisfy to be a projective conditional probability distribution. If we denote by $\{e_I : I \in \mathscr{E}\}$ a basis of $\mathbb{Z}^{|\mathscr{E}|}$, this ideal $J_{\mathscr{E}}$ is multihomogeneous with respect to the grading $\deg(p_{i|I}) = e_I$ (see e.g. [48] for more on such gradings). In what follows, it will be convenient to abbreviate $p_{J|J} := \sum_{j\in J} p_{j|J}$. Thus $p_{J|J}$ would be equal to 1 for honest distributions, by Definition 5.2.1, but here we regard it as a linear form in $\mathbb{C}[\mathscr{E}]$. Let $\alpha_{\mathscr{E}}$ denote the product $\prod_{i\in I\in\mathscr{E}} p_{i|I}$ of all of the $\|\mathscr{E}\|$ variables in $\mathbb{C}[\mathscr{E}]$, and let $\beta_{\mathscr{E}}$ denote the product $\prod_{I\in\mathscr{E}} p_{I|I}$. Now we define the ideal $I_{\mathscr{E}}$, when $[m] \in \mathscr{E}$, by the saturation

$$I_{\mathscr{E}} := (J_{\mathscr{E}} : (\alpha_{\mathscr{E}}\beta_{\mathscr{E}})^{\infty}).$$

When $[m] \notin \mathscr{E}$, let $\mathscr{E}' = \mathscr{E} \cup [m]$ and set $I_{\mathscr{E}} := I_{\mathscr{E}'} \cap \mathbb{C}[\mathscr{E}]$. In the next section, we describe a matrix $\mathcal{A}_G$ such that $I_{\mathscr{E}}$ arises as the toric ideal $I_{\mathcal{A}_G}$ which was defined in Section 1.2. Our main result will be a universal Gröbner basis for the toric ideal $I_{\mathscr{E}}$.

5.3 A universal Gröbner basis for relations among conditional probabilities

A *Bayes binomial* in $\mathbb{C}[\mathscr{E}]$ is a binomial relation of the form

$$p_{i|K}p_{j|J} - p_{j|K}p_{i|J}$$

for $i, j \in J \subseteq K$, with $J, K \in \mathscr{E}$. Let $I_{\mathrm{Bayes}(\mathscr{E})}$ denote the ideal they generate. Bayes binomials get their name because they come from Bayes' rule; more explanation is given in Section 5.6.

Proposition 5.3.1. *The ideal generated by the Bayes binomials contains $J_{\mathscr{E}}$ and is contained in the saturation of $J_{\mathscr{E}}$ by the probabilities that would sum to one:*

$$J_{\mathscr{E}} \subseteq I_{\text{Bayes}(\mathscr{E})} \subseteq (J_{\mathscr{E}} : (\beta_{\mathscr{E}})^{\infty})$$

and in particular, $I_{\text{Bayes}(\mathscr{E})} \subseteq I_{\mathscr{E}}$.

Proof. The ideal $J_{\mathscr{E}}$ is generated by the degree-2 polynomials $p_{J|J}p_{i|K} - p_{i|J}p_{J|K}$ for $J, K \in \mathscr{E}$ and $i \in J \subseteq K$. For each $i, j \in J$, we have $a = p_{j|J}(p_{J|J}p_{i|K} - p_{i|J}p_{J|K})$ and $b = p_{i|J}(p_{J|J}p_{j|K} - p_{j|J}p_{J|K})$ in $J_{\mathscr{E}}$, so $a - b = p_{J|J}(p_{j|J}p_{i|K} - p_{j|K}p_{i|J})$ is in $J_{\mathscr{E}}$ and $I_{\text{Bayes}(\mathscr{E})} \subseteq (J_{\mathscr{E}} : (\beta_{\mathscr{E}})^{\infty})$. For the first inclusion, if $p_{J|J}p_{i|K} - p_{i|J}p_{J|K}$ is a generator of $J_{\mathscr{E}}$, we may write it as an element $\sum_{j \in J}(p_{i|K}p_{j|J} - p_{j|K}p_{i|J})$ of $I_{\text{Bayes}(\mathscr{E})}$. □

Our universal Gröbner basis of $I_{\mathscr{E}}$ will be given combinatorially by the cycles of a labeled bipartite graph $G(\mathscr{E})$, defined as follows:

Vertices: one vertex u_I for each $I \in \mathscr{E}$ and one vertex v_i for each $i \in \cup_{I \in \mathscr{E}} I$

Edges: a directed edge $u_I \to v_i$ for each $I \in \mathscr{E}$ and $i \in I$

Edge Labels: the edge $u_I \to v_i$ is labeled with the indeterminate $p_{i|I}$.

For example, with $n = 4$, the labeled graph G for $\mathscr{E} = \{\{1,2\},\{1,2,3\},\{1,2,3,4\}\}$ is shown in Figure 5.1. Each oriented cycle C in the undirected version of G defines a binomial f_C as follows: each edge label is on the positive side of the binomial if its edge is directed with the cycle, and on the negative if against. For example, in the graph in Figure 5.1, consider the cycle $(1234, 3, 123, 1, 1234)$. The edges p_3 and $p_{1|123}$ are directed with the cycle and the edges $p_{3|123}$ and p_1 are directed against, so the corresponding binomial is $p_3p_{1|123} - p_{3|123}p_1$. For a higher degree example, with $n = 3$ and

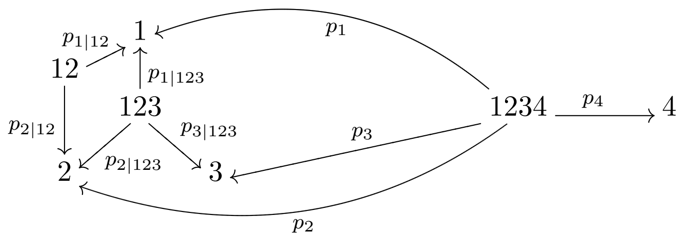


Figure 5.1: Bipartite graph for $\mathscr{E} = \{\{1,2\},\{1,2,3\},\{1,2,3,4\}\}$.

$\mathscr{E} = \{\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$, we get $p_{1|12}p_{3|13}p_{2|23} - p_{2|12}p_{3|23}p_{1|13}$ from the outer cycle, as shown in Figure 5.2. A cycle is *induced* if it has no chord.

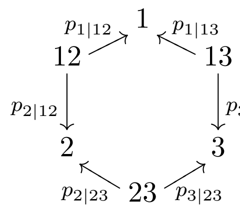


Figure 5.2: Outer cycle of the bipartite graph for $\mathscr{E} = \{\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$.

Theorem 5.3.2. *The binomials defined by the cycles of* $G(\mathscr{E})$ *give a universal Gröbner basis for* $I_{\mathscr{E}}$*. Moreover,* $I_{\mathscr{E}}$ *is generated by the induced cycle binomials, though not necessarily as a Gröbner basis.*

In order to prove Theorem 5.3.2, we first need to recall some facts about unimodular toric ideals, of which $I_{\mathscr{E}}$ is an example. Unimodular matrices and unimodular toric ideals are defined and characterized as follows, following Sturmfels [69]. A *triangulation* of $\mathcal{A}$ is a collection $\mathfrak{F}$ of subsets $\mathcal{B}$ of the columns of $\mathcal{A}$ such that $\{\mathrm{pos}(\mathcal{B}) : \mathcal{B} \in \mathfrak{F}\}$ is the set of cones in a simplicial fan with support $\mathrm{pos}(\mathcal{A})$. A triangulation of $\mathcal{A}$ is *unimodular* if the normalized volume (Section 1.1) is equal to one for all maximal simplices $\mathcal{B}$ in the triangulation. The matrix $\mathcal{A}$ is a *unimodular* matrix if all triangulations of $\mathcal{A}$ are unimodular. We define a unimodular toric ideal in the following definition-proposition.

Proposition 5.3.3. *[69] A toric ideal $I_{\mathcal{A}}$ is called* unimodular *if any of the following equivalent conditions hold.*

(i) Every reduced Gröbner basis of $I_{\mathcal{A}}$ consists of squarefree binomials,

(ii) $\mathcal{A}$ is a unimodular matrix,

(iii) all the initial ideals of $I_{\mathcal{A}}$ are squarefree.

A special class of unimodular matrices are those coming from bipartite graphs [2, 61]. Let $G = (U, V, E)$ be a bipartite graph. In our case, $G(\mathscr{E})$ has

$$U = \{u_I : I \in \mathscr{E}\} \text{ and } V = \{v_i : i \in \cup_{I \in \mathscr{E}} I\}. \tag{5.2}$$

Let $\mathcal{A}$ be the vertex-edge incidence matrix of G: The rows of $\mathcal{A}$ are labeled $u_1, \ldots, u_{|U|}$, $v_1, \ldots, v_{|V|}$, the columns are labeled with the edges, and a_{ij} is 1 if vertex i is in edge j and zero otherwise. For a cycle C in the graph, the cycle binomial f_C is defined (up to sign) as above. Let $\pi_{\mathcal{A}}$ be the map $\mathbb{R}^{\|\mathscr{E}\|} \to \mathbb{R}^{|U|+|V|}$ defined by applying $\mathcal{A}$. We say $u \in \ker(\pi_{\mathcal{A}})$ is a *circuit* if $\text{supp}(u)$ is minimal with respect to inclusion in $\ker(\pi_{\mathcal{A}})$ and the coordinates of u are relatively prime [69]. Equivalently, a circuit is an irreducible binomial $x^{u+} - x^{u-}$ of the toric ideal $I_{\mathcal{A}}$ with minimal support. The *Graver basis* of the ideal $I_{\mathcal{A}}$ consists of all circuits. For $\mathcal{A}$ from a bipartite graph, the circuits of $\mathcal{A}$ are precisely the cycle binomials of the graph [60, 61]. Additionally, a Graver basis is also a universal Gröbner basis in the case of unimodular toric varieties (Proposition 8.11 of [69]). We summarize these results in the following proposition.

Proposition 5.3.4. *The vertex-edge incidence matrix $\mathcal{A}$ of a bipartite graph $G = (U, V, E)$ is unimodular, so $I_{\mathcal{A}}$ is a unimodular toric ideal. The cycle binomials of G are the circuits of*

$\mathcal{A}$, and therefore define the Graver basis of $I_{\mathcal{A}}$. In particular, they give a universal Gröbner basis for $I_{\mathcal{A}}$.

Now we are able to prove our theorem.

Proof of Theorem 5.3.2. Let $\mathcal{A}_{G(\mathscr{E})}$ be the vertex-edge incidence matrix of $G(\mathscr{E})$. By Proposition 5.3.4, its cycle binomials (circuits) give a universal Gröbner basis of $I_{\mathcal{A}_{G(\mathscr{E})}}$. In fact, the induced cycles are enough to generate this ideal [2]. Suppose C is a cycle and e a chord, and split C into two cycles C_1 and C_2, both containing e (but in opposite directions). Associate cycle binomials f_{C_1} and f_{C_2}, respectively. Then the S-polynomial with the e-containing terms leading is f_C. However, this is no longer necessarily a Gröbner basis. For example, let $\mathscr{E} = \{\{12\}, \{23\}, \{123\}\}$ as in Figure 5.3.

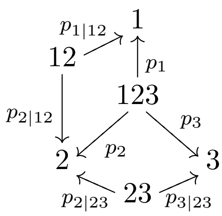


Figure 5.3: Bipartite graph for $\mathscr{E} = \{\{1,2\}, \{2,3\}, \{1,2,3\}\}$.

The outer cycle $C = 1 \to 12 \to 2 \ \to 23 \to 3 \to 123 \to 1$ gives the cycle binomial $f_C = p_{1|12}p_{2|23}p_{3|123} - p_{2|12}p_{3|23}p_{1|123}$. The cycle C has a chord $2 - 123$, and the binomial f_C lies in the ideal of the two binomials

$$p_{1|12}p_{2|123} - p_{2|12}p_{1|123} \text{ and } p_{2|23}p_{3|123} - p_{3|23}p_{2|123}$$

after splitting along the chord. These are both the induced cycles of the graph. However, for a term order prioritizing $p_{2|123}$ (e.g. lexicographic with $p_{2|123} \succ \cdots$), the leading term of

f_C cannot lie in the initial ideal $\langle p_{1|12}p_{2|123}, p_{3|23}p_{2|123}\rangle$ of the ideal generated by the chordal binomials.

Next we show that the graph ideal and conditional probability ideal coincide, $I_{\mathcal{A}_{G(\mathscr{E})}} = I_{\mathscr{E}}$. For the containment $I_{\mathcal{A}_{G(\mathscr{E})}} \supseteq I_{\mathscr{E}}$, first observe that $I_{\text{Bayes}(\mathscr{E})} \subseteq I_{\mathcal{A}_{G(\mathscr{E})}}$. This

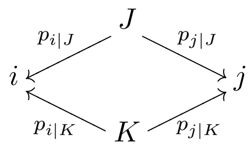


Figure 5.4: Subgraph of $G(\mathscr{E})$ giving a Bayes binomial.

is because if $J, K \in \mathscr{E}$ with $i, j \in J \subseteq K$, we have the subgraph in Figure 5.4, which is a cycle with associated cycle binomial $p_{j|J}p_{i|K} - p_{i|J}p_{j|K}$. Together with Proposition 5.3.1, we now have

$$J_{\mathscr{E}} \subseteq I_{\text{Bayes}(\mathscr{E})} \subseteq I_{\mathcal{A}_{G(\mathscr{E})}}$$

so, since saturation is inclusion-preserving and $I_{\mathcal{A}_{G(\mathscr{E})}}$ is prime,

$$I_{\mathscr{E}} = (J_{\mathscr{E}} : (\alpha_{\mathscr{E}}\beta_{\mathscr{E}})^{\infty}) \subseteq (I_{\mathcal{A}_{G(\mathscr{E})}} : (\alpha_{\mathscr{E}}\beta_{\mathscr{E}})^{\infty}) = I_{\mathcal{A}_{G(\mathscr{E})}}.$$

Now we show the reverse inclusion $I_{\mathcal{A}_{G(\mathscr{E})}} \subseteq I_{\mathscr{E}}$. Again by Proposition 5.3.1, we have

$$I_{\text{Bayes}(\mathscr{E})} \subseteq I_{\mathscr{E}}.$$

Now assume that $[m] \in \mathscr{E}$, so that $p_1, \dots, p_m \in \mathbb{C}[\mathscr{E}]$. We claim that in fact $I_{\mathcal{A}_{G(\mathscr{E})}} \subseteq (I_{\text{Bayes}(\mathscr{E})} : \prod_{i=1}^m p_i)$, from which the result will follow. Let C be an induced cycle of $G(\mathscr{E})$, and f_C its cycle binomial. We must show that this cycle binomial can be obtained from the Bayes binomials, up to multiplication by $\prod_{i=1}^m p_i$. Let C be the cycle

$$i_1 \leftarrow J_1 \rightarrow i_2 \leftarrow J_2 \rightarrow \cdots \rightarrow i_k \leftarrow J_k \rightarrow i_1.$$

With this notation we have $i_1, i_2 \in J_1$, $i_2, i_3 \in J_2$, $\ldots$, $i_1, i_k \in J_k$. Then

$$f_C = p_{i_2|J_1} p_{i_3|J_2} \cdots p_{i_k|J_{k-1}} p_{i_1|J_k} - p_{i_1|J_1} p_{i_2|J_2} \cdots p_{i_k|J_k}.$$

We show the first monomial of $(\prod_{i=1}^{k} p_i) f_C$ is equal to the second mod $I_{\text{Bayes}(\mathscr{E})}$. Pair off as follows:

$$\begin{aligned} & (\underline{p_{i_1}} p_{i_2} p_{i_3} \cdots p_{i_k}) \underline{p_{i_2|J_1}} p_{i_3|J_2} p_{i_4|J_3} \cdots p_{i_k|J_{k-1}} p_{i_1|J_k} && \text{Step 1} \\ = \; & (p_{i_2} \underline{p_{i_2}} p_{i_3} \cdots p_{i_k}) p_{i_1|J_1} \underline{p_{i_3|J_2}} p_{i_4|J_3} \cdots p_{i_k|J_{k-1}} p_{i_1|J_k} && \text{Step 2} \\ = \; & (p_{i_2} p_{i_3} \underline{p_{i_3}} \cdots p_{i_k}) p_{i_1|J_1} p_{i_2|J_2} \underline{p_{i_4|J_3}} \cdots p_{i_k|J_{k-1}} p_{i_1|J_k} && \text{Step 3} \\ & \vdots \end{aligned}$$

where the equalities hold mod $I_{\text{Bayes}(\mathscr{E})}$. Continuing in this fashion, at step $k-1$ we have

$$\begin{aligned} = \; & (p_{i_2} p_{i_3} \cdots p_{i_{k-1}} \underline{p_{i_{k-1}}} p_{i_k}) p_{i_1|J_1} p_{i_2|J_2} \cdots p_{i_{k-2}|J_{k-2}} \underline{p_{i_k|J_{k-1}}} p_{i_1|J_k} && \text{Step } k-1 \\ = \; & (p_{i_2} p_{i_3} \cdots p_{i_{k-1}} p_{i_k} \underline{p_{i_k}}) p_{i_1|J_1} p_{i_2|J_2} \cdots p_{i_{k-2}|J_{k-2}} p_{i_{k-1}|J_{k-1}} \underline{p_{i_1|J_k}} && \text{Step } k \\ = \; & (p_{i_2} p_{i_3} \cdots p_{i_{k-1}} p_{i_k} p_{i_1}) p_{i_1|J_1} p_{i_2|J_2} \cdots p_{i_{k-2}|J_{k-2}} p_{i_{k-1}|J_{k-1}} p_{i_k|J_k} && \text{Step } k+1 \end{aligned}$$

as desired. In terms of $G(\mathscr{E})$, this amounts to breaking up a long cycle into 4-cycles passing through $[m]$, and erasing the overlaps among these cycles. Thus since the induced cycles generate $I_{\mathcal{A}_{G(\mathscr{E})}}$, we have

$$I_{\mathcal{A}_{G(\mathscr{E})}} \subseteq (I_{\text{Bayes}(\mathscr{E})} : \prod_{i=1}^{m} p_i)) \subseteq I_{\mathscr{E}}$$

This proves the result in the special case $[m] \in \mathscr{E}$. In the general case, suppose we have some $\mathscr{E}$ not containing $[m]$, enabling us to obtain relations among 'pure' conditional probabilities (i.e. excluding $p_1, \ldots, p_m$). Let $\mathscr{E}' = \mathscr{E} \cup [m]$ and apply the special case of the Theorem.

Then by [69, Proposition 4.13(c)], since we have a universal Gröbner basis, we just intersect it with the smaller coordinate ring to obtain a universal Gröbner basis of the smaller ring. This corresponds here to removing the set $[m]$ from $\mathscr{E}$ and taking the cycle binomials as our new Gröbner basis. □

5.4 Conditional probability and the moment map

In this section we show how to recover and generalize some results of Matúš [45] using toric geometry. The main result we will expand upon maps the space of conditional probability distributions (Definition 5.2.1) for all possible conditioned events $\mathscr{E} = \{I \subset [m] : |I| \geq 2\}$ onto the permutohedron by first projecting down to events of size 2, $\mathscr{E} = \{I \subset [m] : |I| = 2\}$.

Theorem 5.4.1 (Matúš [45]). *For $\mathscr{E} = \{I \subset [m] : |I| \geq 2\}$ and $\mathbf{p}$ a conditional probability distribution (Definition 5.2.1), the map $W : \mathbb{R}^{\|\mathscr{E}\|} \to \mathbb{R}^m$, given by*

$$W_i(\mathbf{p}) = \sum_{j \in [m] \setminus i} p_{i|ij},$$

restricts to a homeomorphism of the space of conditional probabilities onto the $m-1$ dimensional permutohedron $\mathbf{P}_{m-1}$.

Note that the linear map W is the restriction of $\mathcal{A} = \mathcal{A}_{G(\mathscr{E})}$ to the rows labeled by the vertex set V in G (5.2) and to the columns labeled by two-event conditional probabilities (edges in $G(\mathscr{E})$) $p_{i|ij}$. In fact $\mathcal{A}$, will in general define a map from the space of projective conditional probability distributions onto a generalized permutohedron. This polytope is the Minkowski sum of simplices (MSS) defined in Section 2.4 for the subset $\mathscr{E}$ of $2^{[m]}$ and corresponding set of simplices $\{\Delta_I, I \in \mathscr{E}\}$.

First consider the multiprojective toric variety $\mathcal{Z}_{\mathcal{A}}$ cut out of $\prod_{I\in\mathscr{E}}\mathbb{P}^{|I|-1}$ by the equations of Theorem 5.3.2, i.e. the space of projective conditional probability distributions. In Chapter 1 we defined the affine toric variety $X_{\mathcal{A}}$ associated to an integer matrix $\mathcal{A}$, and the projective toric variety $Y_{\mathcal{A}}$ associated to a $\mathbb{Z}$-graded matrix $\mathcal{A}$ (that is, a matrix $\mathcal{A}$ such that $(1,1,\ldots,1)$ lies in its rowspan). Given a matrix $\mathcal{A}=\mathcal{A}_{G(\mathscr{E})}$, the space of $\mathscr{E}$-projective conditional probability distributions $\mathcal{Z}_{\mathcal{A}}$ is the closure of the image of the map $\mathbf{f}_{\mathcal{A}}:\theta\mapsto\theta^{\mathcal{A}}$, viewed as an element of $\prod_{I\in\mathscr{E}}\mathbb{P}^{|I|-1}$. Equipping this product space with multihomogeneous coordinates $((p_{i_1|I}:\cdots:p_{i_{|I|}|I}), I\in\mathscr{E})$, the variety $\mathcal{Z}_{\mathcal{A}}$ is cut out by the (multihomogeneous) toric ideal $I_{\mathcal{A}}$. Suppose that we have $\cup_{I\in\mathscr{E}}=[m]$. Then because we view the points $((p_{i_1|I}:\cdots:p_{i_{|I|}|I}), I\in\mathscr{E})$ as elements of $\prod_{I\in\mathscr{E}}\mathbb{P}^{|I|-1}$, the dimension of this variety is $m-1$ as expected, though the rank of $\mathcal{A}$ is larger.

We now develop a version of the moment map of toric geometry applicable to the variety of projective conditional probability distributions. Hereafter we index the columns of $\mathcal{A}$ by the conditional probability they represent, i.e. $\mathcal{A}=(a_{\cdot i|I}: i\in I\in\mathscr{E})$. We will require a multigraded notion to play the role of the convex hull conv($\mathcal{A}$) in the moment map. We define

$$\mathrm{mconv}(\mathcal{A})=\{\sum_{I\in\mathscr{E}}\sum_{j\in I}\lambda_{j|I}a_{\cdot j|I}:\lambda_{j|I}\in\mathbb{R}_{\geq 0},\ \sum_{j\in I}\lambda_{j|I}=1\}.$$

Recall from Section 2.4 the following construction. Each subset I of $[m]$ defines a submodular function w_I on $2^{[n]}$ by setting $w_I(J)=1$ if $I\cap J$ is non-empty and $w_I(J)=0$ if $I\cap J$ is empty for $J\in 2^{[n]}$. The corresponding polytope defined by this function by the construction

$$\begin{aligned} Q_w \ := \ & \big\{\, x\in\mathbb{R}^n \,:\, x_1+x_2+\cdots+x_n = w([n]) \\ & \text{and } \textstyle\sum_{i\in I} x_i \leq w(J) \text{ for all } \emptyset\neq J\subseteq[n] \,\big\} \end{aligned}$$

is the simplex $\Delta_I=\mathrm{conv}\{e_k: k\in I\}$. Now consider an arbitrary subset $\mathscr{E}=\{I_1,I_2,\ldots,I_r\}$

of $2^{[m]}$. It defines the submodular function $w_{\mathscr{E}} = w_{I_1} + w_{I_2} + \cdots + w_{I_r}$. The corresponding polytope $Q_{w_{\mathscr{E}}}$ is now the Minkowski sum

$$\Delta_{\mathscr{E}} \quad = \quad \Delta_{I_1} + \Delta_{I_2} + \cdots + \Delta_{I_r}. \tag{5.3}$$

Proposition 5.4.2. *The projection of* $\mathrm{mconv}(\mathcal{A}_{G(\mathscr{E})})$ *to the* V*-coordinates* (5.2) *is* $\Delta_{\mathscr{E}}$.

Proof. The mconv construction is equivalent to translating each simplex that is the convex hull of each set of vectors $\mathcal{A}_I \subset \mathcal{A}$ by setting its U-coordinates (5.2) all to 1, then taking the Minkowski sum. □

Next is a version of Theorem 1.2.3 for varieties $\mathcal{Z}_{\mathcal{A}}$. Note that $|V| = m$ when $\cup_{I\in\mathscr{E}} I = [m]$.

Theorem 5.4.3. *For* $\mathcal{A} = \mathcal{A}_{G(\mathscr{E})}$*, the map* $\nu : \mathcal{Z}_{\mathcal{A}} \to \mathbb{R}^{|V|}$ *defined by*

$$\nu(z) = \sum_{I\in\mathscr{E}} \frac{1}{Z_I(z)} \sum_{i\in I} |z_{i|I}| a_{\cdot i|I},$$

where $Z_I = \sum_{i\in I} |z_{i|I}|$*, maps* $\mathcal{Z}_{\mathcal{A}}$ *onto* $\mathrm{mconv}(\mathcal{A})$*, and is a bijection on* $\mathcal{Z}_{\mathcal{A},\geq 0}$.

Proof. The map ν is the composition of two maps. The first map, $\nu_1 : \mathcal{Z}_{\mathcal{A}} \to \prod_{I\in\mathscr{E}} \Delta_I$, is a product of maps μ_1 corresponding to each submatrix $\mathcal{A}_I$ as in the proof of Theorem 1.2.3, sending $((z_{i_1|I}, \ldots, z_{i_{|I|}|I}), I \in \mathscr{E}) \in \mathcal{Z}_{\mathcal{A}}$ to $\mathbf{p} = (p_{i|I} = \frac{1}{Z_I(z)}|z_{i|I}| : i \in I \in \mathscr{E})$. The second map, ν_2, corresponds to the Minkowski sum, with $\nu_2 : \prod_{I\in\mathscr{E}} \Delta_I \to \mathrm{mconv}(\mathcal{A})$ sending $\mathbf{p}$ to $\mathcal{A}\mathbf{p}$. Whereas in the simplex case (and for a single $\mathcal{A}_I$) in Theorem 1.2.3, μ_1 and μ_2 were identities, here there is additional ambiguity introduced by the Minkowski sum. In particular, let $b \in \Delta_{\mathscr{E}}$ (5.3). Then the preimage of b in $\prod_{I\in\mathscr{E}} \Delta_I$ is

$$P_{\mathcal{A}}(b) = \{\mathbf{p} : \mathcal{A}\mathbf{p} = b\} \cap \prod_{I\in\mathscr{E}} \Delta_I,$$

and in general consists of a polytope. This is illustrated in Figure 5.5, where the polytope $P_{\mathcal{A}}(b)$ is the set of pairs of points in the first and second simplex that add to b. Analogously to the one-factor case (Theorem 1.2.3), we will choose among the points of this fiber by selecting the maximum entropy point (or the point closest in the KL-divergence sense to the point representing a uniform distribution in all simplices).

Setting $D(\mathbf{p}) = D(\mathbf{p}||\mathbf{p}^{uniform})$ so

$$D(\mathbf{p}) = \sum_{i \in I \in \mathscr{E}} p_{i|I} \log p_{i|I} - \sum_{i \in I \in \mathscr{E}} p_{i|I} \log(\frac{1}{|I|}),$$

the Hessian of D is $\frac{1}{p_{i|I}}$ on the diagonal and zero elsewhere. Thus it is positive definite on the interior of $\prod_{I \in \mathscr{E}} \Delta_I$, and on points of the relative interior after restricting to nonzero coordinates. Thus D has a unique minimum $\mathbf{p}^*$ on $\prod_{I \in \mathscr{E}} \Delta_I$. Were there another minimum, the (possibly restricted) Hessian would be positive definite on the open segment connecting it with $\mathbf{p}^*$. We now argue that $\mathbf{p}^* \in \mathcal{Z}_{\mathcal{A}}$.

First suppose $\mathbf{p}^* \in (\prod_{I \in \mathscr{E}} \Delta_I)^\circ$, so that $0 < p_{i|I} < 1$ in all coordinates, and let $u \in \ker \mathcal{A}$. We must show that $p^{u^+} = p^{u^-}$. For small t, $\mathbf{p}^* + tu \in \prod_{I \in \mathscr{E}} \Delta_I$ and

$$D(\mathbf{p}^* + tu) = \sum_{i \in I \in \mathscr{E}} (p_{i|I} + tu_{i|I}) \log(p_{i|I} + tu_{i|I}) - \sum_{i \in I \in \mathscr{E}} (p_{i|I} + tu_{i|I}) \log \frac{1}{|I|}$$

$$\frac{dD}{dt} = \sum_{i \in I \in \mathscr{E}} u_{i|I} \log(p_{i|I} + tu_{i|I}) + \sum_{i \in I \in \mathscr{E}} u_{i|I} + \sum_{i \in I \in \mathscr{E}} u_{i|I} \log \frac{1}{|I|}.$$

Since $\mathcal{A}$ is $\mathscr{E}$-multigraded, the last two terms of $\frac{dD}{dt}$ are zero (i.e. $(1, 1, \ldots, 1) \in \mathbb{R}^{\|\mathscr{E}\|}$ is in the rowspace of $\mathcal{A}$, and $(1, 1, \ldots, 1) \in \mathbb{R}^{|I|}$ is in the rowspace of each $\mathcal{A}_I$). At $t = 0$, the first order condition implies that

$$0 = \frac{dD}{dt} = \sum_{i \in I \in \mathscr{E}} u_{i|I} \log p_{i|I}.$$

Grouping the sum by the sign of $u_{i|I}$ and changing to exponential notation,

$$p^{u^+} = p^{u^-} \tag{5.4}$$

as desired.

Now suppose that $\mathbf{p}^*$ lies on the boundary of $\prod_{I\in\mathscr{E}} \Delta_I$. If the zeros of $\mathbf{p}$ lie outside $\mathrm{supp}(u)$, the argument made above for $\mathbf{p}^*$ in the interior holds after extending D with the limit $p\log(p) \to 0$ as $p \to 0$. If there are zeros on both sides of (5.4), i.e. $p_{i|I} = 0 = p_{j|J}$ for indices $i|I \in \mathrm{supp}(u^+)$ and $j|J \in \mathrm{supp}(u^-)$, then the relation holds with $0 = 0$.

We may assume $p_{i|I} = 0$ for some index $i|I \in \mathrm{supp}(u^+)$ in considering the two remaining cases. The first case has $p_{j|J} = 1$ for some index $j|J \in \mathrm{supp}(u^+)$. Because of the multigrading of $\mathcal{A}$, which requires for any $J \in \mathscr{E}$ and $u \in \ker \mathcal{A}$ that $\sum_{j\in J} u_{j|J} = 0$, it must be that there exists $k|J \in \mathrm{supp}(u^-)$. Then since $\mathbf{p} \in \prod_{I\in\mathscr{E}} \Delta_I$, we have $p_{k|J} = 0$ and the relation (5.4) holds as $0 = 0$.

The second case has $0 \leq p_{j|J} < 1$ for all $j|J \in \mathrm{supp}(u^+)$ and $0 < p_{k|K} \leq 1$ for all $k|K \in \mathrm{supp}(u^-)$. Then for small t, $\mathbf{p}^* + tu \in P_{\mathcal{A}}(b)$. Then we have

$$\frac{dD}{dt} = \sum_{\{i|I\,:\,p_{i|I}=0\}} u_{i|I}(p_{i|I} + tu_{i|I}) + \sum_{\{j|J\,:\,p_{j|J}\neq 0\}} u_{j|J}(p_{j|J} + tu_{j|J}). \tag{5.5}$$

Then the first term on the right hand side of (5.5) approaches negative infinity as $t \to 0$ while the second approaches a constant; this contradicts the optimality of $\mathbf{p}^*$, so this case cannot arise. □

We now give a couple of examples.

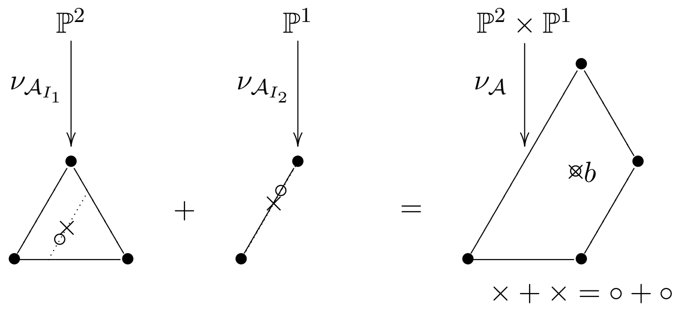


Figure 5.5: Ambiguity arising from Minkowski sum of simplices: two points appearing in the fiber over b in $\prod_{I\in\mathscr{E}} \Delta_I$. For any point on the dotted line, there is a point in the second simplex such that their sum is b. We choose $\times$ among these points by maximizing entropy in the conditional probability distribution. The space of conditional probability distributions is the blowup of $\mathbb{P}^2$ at a labeled point of Figure 1.2, intersected with a triangular prism.

Example 5.4.4. For the case $m = 3$ with $\mathscr{E} = \{12, 13, 23, 123\}$, the matrix $\mathcal{A}$ is

	p_1	p_2	p_3	$p_{1\|12}$	$p_{2\|12}$	$p_{1\|13}$	$p_{3\|13}$	$p_{2\|23}$	$p_{3\|23}$
1	1	0	0	1	0	1	0	0	0
2	0	1	0	0	1	0	0	1	0
3	0	0	1	0	0	0	1	0	1
12	0	0	0	1	1	0	0	0	0
13	0	0	0	0	0	1	1	0	0
23	0	0	0	0	0	0	0	1	1
123	1	1	1	0	0	0	0	0	0

The U-coordinate rows are labeled $1, 2, 3$ and the V-coordinate rows are labeled $12, 13, 23, 123$.

The polytope $\mathrm{mconv}(\mathcal{A})$ is the permutohedron which is the convex hull of the permutations of $(3, 1, 0)$, shown in Figure 5.6. Letting $\mathcal{A}'$ be the last six columns of $\mathcal{A}$ (restriction to $\{I \subseteq [n] : |I| = 2\}$), $\mathrm{mconv}(\mathcal{A}')$ is the regular permutohedron $\mathrm{conv}((2, 1, 0), (2, 0, 1), (1, 0, 2), (1, 2, 0), (0, 2, 1), (0, 1, 2))$, lifted with the last four coordinates all 1. This is illustrated in Figure 5.7.

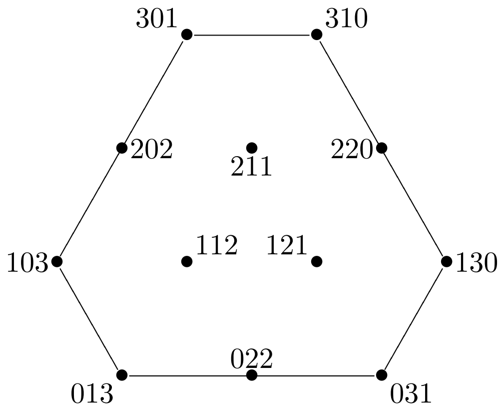


Figure 5.6: Multigraded convex hull of $\mathcal{A}$ for $n = 3$ and $\mathscr{E} = \{I \subseteq [n] : |I| \geq 2\}$. The last four coordinates, not shown, are all 1.

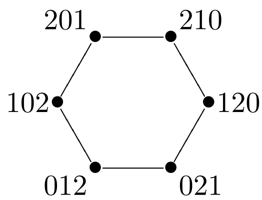


Figure 5.7: Multigraded convex hull of $\mathcal{A}$ for $n = 3$ and $\mathscr{E} = \{I \subseteq [n] : |I| = 2\}$. The last four coordinates, not shown, are all 1

The theorem of Matúš (Theorem 5.4.1) works in this way by projecting first from $\mathscr{E} = \{I : |I| \geq 2\}$ to $\mathscr{E} = \{I : |I| = 2\}$ as in Figure 5.7. Thus the result may be understood as saying that instead of all simplices, we can obtain a regular permutohedron merely as the zonotope given by the Minkowski sum of the 1-simplices.

5.5 The case of discrete random variables

Let $X_1, \ldots, X_n$ be discrete random variables with X_i taking values $x_i^1, \ldots, x_i^{d_i}$. Then the $m = \prod_{i=1}^n d_i$ singleton events in Ω are the elements of the Cartesian product of the sets of states which each random variable may assume. For a subset of random variables $X_{i_1}, \ldots, X_{i_k}$ with $S := \{i_1, \ldots, i_k\} \subseteq [n]$, we write Ω_S for the Cartesian product of the states of this subset of the random variables. We also denote by $x|_S$ the restriction of some global state $x \in \Omega$ to the states of the random variables in S. Then in this important special case of n random variables, the set of events $\mathscr{E}$ has the form:

$$\mathscr{E} = \{x' \in \Omega \ : \ x'|_S = x_S \text{ for some } S \subseteq [n], x_S \in \Omega_S\} \tag{5.6}$$

Let $E(x_S)$ denote the event which is the union of all singleton events with random variables S in state x_S. For example, let $n = 3$, $d_i = 2$ with states denoted 0 and 1, and $S = \{1, 3\}$. Then $E(x_1^0 x_3^1) = \{0010, 0011, 0110, 0111\}$, which corresponds to a 2-face of the 4-cube. With this notation we may abbreviate

$$p_{x_A|x_B} := p_{E(x_A) \cap E(x_B)|E(x_B)}$$

which is convenient for considering, say, the conditional probability of having a disease given a positive test result. The Hammersley-Clifford theorem (Theorem 1.3.3) provides a partial answer to relations among conditional probabilities in the case of undirected graphical models with strictly positive probability distributions [5]. In this setting, n random variables form the vertices of an undirected graph. Then given subsets of variables $A, B \subseteq [n]$ we consider conditioning events $E(x_B)$ such that for some A, we have $B = \text{nbhd}(A)$. Arnold et. al. [4] develop the theory for both discrete and continuous random variables, particularly in the case of two random variables. Slavkovic and Sullivant [61] consider the case of compatible full conditionals using unimodular toric varieties. They study the random variable case with subsets of random variables A and B such that $A \cup B = [n]$ and $A \cap B = \emptyset$.

Besag [5] gives a relation among positive conditional probabilities using this random variable notation:

$$\frac{P(\mathbf{x})}{P(\mathbf{y})} = \prod_{i=1}^{n} \frac{P(x_i|x_1,\ldots,x_{i-1},y_{i+1},\ldots,y_n)}{P(y_i|x_1,\ldots,x_{i-1},y_{i+1},\ldots,y_n)}. \tag{5.7}$$

This is a special case of the relations derived in Theorem 5.3.2, as we now explain.

Denote the event $x_1,\ldots,x_{j-1},y_j,\ldots,y_n$ by j, so the singleton events are $(y_1,\ldots,y_n) = 1,2,\ldots,n+1 = (x_1,\ldots,x_n)$. The set $\mathscr{E}$ consists of the event $\{1,\ldots,n+1\}$ together with the events $\{j,j+1\}$ for $j=1,\ldots,n$. Then the cleared-denominator version of (5.7) is the outer cycle $[n+1] \to 1 \leftarrow 12 \to 2 \leftarrow \cdots \leftarrow n,n+1 \to n+1 \leftarrow [n+1]$ in the graph $G_{\mathscr{E}}$. For example, with three variables we have events $1 = (y_1,y_2,y_3)$, $2 = (x_1,y_2,y_3)$, $3 = (x_1,x_2,y_3)$, and $4 = (x_1,x_2,x_3)$. The relation (5.7) is

$$\frac{p_4}{p_1} = \frac{p_{2|12}p_{3|23}p_{4|34}}{p_{1|12}p_{2|23}p_{3|34}},$$

corresponding to the cycle binomial

$$p_1p_{2|12}p_{3|23}p_{4|34} - p_4p_{1|12}p_{2|23}p_{3|34},$$

which is f_C for the outer cycle C of the graph in Figure 5.8.

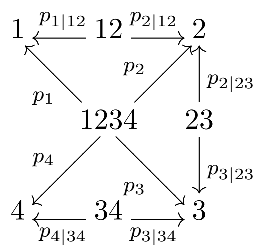


Figure 5.8: Bipartite graph for $\mathscr{E} = \{\{1,2\},\{2,3\},\{3,4\},\{1,2,3,4\}\}$.

5.6 Bayes' rule

Because of the Bayes binomials, on points which are projective conditional probability distributions, we have, with $i, j \subseteq J \subseteq K \subseteq [m]$,

$$p_{i|K} p_{j|J} = p_{j|K} p_{i|J}.$$

This implies, by summing over $j \in J$, that

$$p_{i|K} p_{J|J} = p_{J|K} p_{i|J}. \tag{5.8}$$

Using two copies of (5.8) with different intermediate sets J_1 and J_2, we have

$$(p_{i|J_1} p_{J_1|K}) p_{J_2|J_2} = p_{i|K} p_{J_1|J_1} p_{J_2|J_2} = (p_{i|J_2} p_{J_2|K}) p_{J_1|J_1}$$

which gives a multihomogeneous version of Bayes' rule. Because we consider the point representing a projective conditional probability distribution as an element of $((p_{i_1|I} : \cdots : p_{i_{|I|}|I}), I \in \mathscr{E})$, we may set $p_{J_1|J_1}$ and $p_{J_2|J_2}$ to 1 on an open set containing all probabilistically relevant points, and summing over $i \in I$, this becomes

$$p_{I|J_1} p_{J_1|K} = p_{I|J_2} p_{J_2|K}.$$

Or when $p_{J_1|K} \neq 0$,

$$p_{I|J_1} = \frac{p_{I|J_2} p_{J_2|K}}{p_{J_1|K}}$$

so that in particular, with $A, B \subseteq [m]$, and setting $I = A \cap B$, $J_1 = B$, $J_2 = A$, and $K = [m]$ we have the familiar expression for Bayes' rule

$$p_{A\cap B|B} = \frac{p_{A\cap B|A} p_A}{p_B}.$$

Bibliography

[1] S. Ahnert, K. Willbrand, F. Brown, and T. Fink. Unbiased pattern detection in microarray data series. *Bioinformatics*, 22(1471), 2006.

[2] S. Aoki and A. Takemura. Markov chain monte carlo exact tests for incomplete two-way contingency tables. *Journal of Statistical Computation and Simulation*, 75(10):787–812, 2005.

[3] S. Aoki, A. Takemura, and R. Yoshida. Indispensable monomials of toric ideals and Markov bases. *Journal of Symbolic Computations*, to appear.

[4] B. Arnold, E. Castillo, and J.M. Sarabia. *Conditional Specification of Statistical Models.* Springer, 1999.

[5] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36(2):192–236, 1974.

[6] L. Billera, I. Gelfand, and B. Sturmfels. Duality and minors of secondary polyhedra. *Journal of Combinatorial Theory*, (57):258–268, 1993.

[7] P. Billingsley. *Probability and Measure.* Wiley, 1995.

[8] J. Bokowski and B. Sturmfels. Polytopal and non-polytopal spheres—an algorithmic approach. *Israel Journal of Mathematics*, 57:257–271, 1987.

[9] S. Brady, D. Orlando, J. Lee, J Wang, J. Koch, J. Dinneny, D. Mace, U. Ohler, and P. Benfey. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science*, 318(5851):801–806, 2007.

[10] G. Brightwell and P. Winkler. Counting linear extensions. *Order*, 8(3):225–242, 1991.

[11] F. Brown, T. Fink, and K. Willbrand. On arithmetic and asymptotic properties of up-down numbers. *Discrete Mathematics*, 307(14), 2007.

[12] K. Brown. *Buildings.* Springer, New York, 1989.

[13] W. Bruns and R. Koch. Computing the integral closure of an affine semigroup. effective methods in algebraic and analytic geometry. *Univ. Iagel. Acta Math.*, 39:59–70, 2001.

[14] M. Carr and S. Devadoss. Coxeter complexes and graph associahedra. *Topology and its Applications*, 153:2155–2168, 2006.

[15] C. Chatfield. On analysing time-series data showing cyclic variation. *The Statistician*, 27(1):55–56, 1978.

[16] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko. Lipschitz functions have l_p-stable persistence. Manuscript, Department of Computer Science, Duke University, 2007.

[17] W. Cook and L. Seiford. The geometry of rank-order tests. *The American Statistician*, 37(4):307–311, 1983.

[18] D. Cox. Toric varieties and toric resolutions. In H. Hauser, J. Lipman, F. Oort, and A. Quirós, editors, *Resolutions of Singularities*, pages 259–284. Birkhäuser, Basel-Boston-Berlin, 2000.

[19] A. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society B*, 41:1–31, 1979.

[20] M. Dequéant, S. Ahnert, H. Edelsbrunner, T. Fink, E. Glynn, G. Hattem, A. Kudlicki, Y. Mileyko, J. Morton, A. Mushegian, L. Pachter, M. Rowicka, A. Shiu, B. Sturmfels, and O. Pourquié. Comparing pattern detection methods in microarray time series of the segmentation clock. Submitted to *Bioinformatics*.

[21] M.L. Dequéant, E. Glynn, K. Gaudenz, M. Wahl, J. Chen, A. Mushegian, and O. Pourquié. A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science*, 314(5805):1595–1598, 2006.

[22] P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Annals of Statistics*, 26:363–397, 1998.

[23] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.

[24] G. Ewald. *Combinatorial Convexity and Algebraic Geometry.* Springer, 1996.

[25] S. Fomin and N. Reading. Root systems and generalized associahedra. Lecture notes for the IAS/Park City Graduate Summer School in Geometric Combinatorics, July 2004. math.CO/0505518.

[26] W. Fulton. *Introduction to Toric Varieties.* Princeton University Press, 1993.

[27] L. D. Garcia, M. Stillman, and B. Sturmfels. Algebraic geometry of Bayesian networks. *Journal of Symbolic Computation*, 39:331–355, 2004. Special issue on the occasion of MEGA 2003.

[28] E. Gawrilow and M. Joswig. polymake: a framework for analyzing convex polytopes. In G. Kalai and G.M. Ziegler, editors, *Polytopes — Combinatorics and Computation*, pages 43–74. Birkhäuser, 2000.

[29] D. Geiger, C. Meek, and B. Sturmfels. On the toric algebra of graphical models. *Annals of Statistics*, 34:1463–1492, 2006.

[30] E.F. Glynn, J. Chen, and A. Mushegian. Detecting periodic patterns in unevenly spaced gene expression time series using lomb-scargle periodograms. *Bioinformatics*, 22:310–316, 2006.

[31] D. Grayson and M. Stillman. Macaulay 2, a software system for research in algebraic geometry. Available at `http://www.math.uiuc.edu/Macaulay2/`, 2002.

[32] B. Grünbaum. *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer, second edition, 2003. Prepared and with a preface by V. Kaibel, V. Klee and G.M. Ziegler.

[33] R. Hemmecke, R. Hemmecke, and P. Malkin. 4ti2 version 1.2—computation of Hilbert bases, Graver bases, toric Gröbner bases, and more. Available from `www.4ti2.de`, 2005.

[34] R. Hemmecke, J. Morton, A. Shiu, B. Sturmfels, and O. Wienand. Three counterexamples on semigraphoids. *Combinatorics, Probability, and Computing*, to appear.

[35] R. Hemmecke, A. Takemura, and R. Yoshida. Computing holes in semi-groups. Preprint. math.CO/0607599.

[36] H. Hirai. Sequences of stellar subdivisions. Preprint, 2006.

[37] C. Hohlweg and C. Lange. Realizations of the associahedron and cyclohedron. *Discrete and Computational Geometry*, 37(4):517–543, 2007.

[38] R.R. Klevecz, J. Bolen, G. Forrest, and D.B. Murray. A genomewide oscillation in transcription gates dna replication and cell cycle. *Proceedings of the National Academy of Sciences*, 101(5):1200–1205, 2004.

[39] S.L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press Oxford University Press, New York, 1996. Oxford Science Publications.

[40] Y. Luan and H. Li. Model-based methods for identifying periodically expressed genes based on time. *Bioinformatics*, 20(3):332–339, 2004.

[41] M. Markl. Simplex, associahedron, and cyclohedron. *Contemporary Mathematics*, 227:235–265, 1999.

[42] F. Matúš. Ascending and descending conditional independence relations. *Proceedings of the Eleventh Prague Conference on Inform. Theory, Stat. Dec. Functions and Random Proc.*, pages 189–200, 1992.

[43] F. Matúš. On equivalence of Markov properties over undirected graphs. *Journal of Applied Probability*, 29:745–749, 1992.

[44] F. Matúš. Conditional independences among four random variables. iii. final conclusion. *Combinatorics, Probability, and Computing*, 8:269–276, 1999.

[45] F. Matúš. Conditional probabilities and permutahedron. *Annales de l'Institut H. Poincaré, Probabilités et Statistiques*, 39:687–701, 2003.

[46] F. Matúš. Towards classification of semigraphoids. *Discrete Mathematics*, 277:115–145, 2004.

[47] M. McDonald and M. Rosbash. Microarray analysis and organization of circadian gene expression in drosophila. *Cell*, 107:567–578, 2001.

[48] E. Miller and B. Sturmfels. *Combinatorial Commutative Algebra*, volume 227 of *Graduate Texts in Mathematics*. Springer, New York, 2004.

[49] J. Morton, A. Shiu, L. Pachter, and B. Sturmfels. The cyclohedron test for finding periodic genes in time course expression studies. *Statistical Applications in Genetics and Molecular Biology*, 6:1–21, 2007.

[50] J. Morton, A. Shiu, L. Pachter, B. Sturmfels, and O. Wienand. Geometry of rank tests. *Proceedings of the Third European Workshop on Probabalistic Graphical Models*, pages 207–214, 2006.

[51] J. Morton, A. Shiu, L. Pachter, B. Sturmfels, and O. Wienand. Convex rank tests and semigraphoids. Preprint, 2007. math.CO/0702564.

[52] L. Pachter and B. Sturmfels, editors. *Algebraic Statistics for Computational Biology*. Cambridge University Press, 2005.

[53] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Mateo CA, 1988.

[54] E.J.G. Pitman. Significance tests which may be applied to samples from any populations. *Supplement to the Journal of the Royal Statistical Society*, 4(1):119–130, 1937.

[55] A. Postnikov. Permutohedra, associahedra, and beyond. Preprint, 2005.

[56] A. Postnikov, V. Reiner, and L. Williams. Faces of simple generalized permutohedra. Preprint, 2006. math.CO/0609184.

[57] O. Pourquié. The segmentation clock: converting embryonic time into spatial pattern. *Science*, 301(5631):328–330, 2003.

[58] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005.

[59] N. Sandman. A type-b Tamari poset. *Discrete Applied Math.*, 143:110–122, 2004.

[60] A. Schrijver. *Theory of Integer and Linear Programming*. John Wiley & Sons, New York, 1998.

[61] A. Slavkovic and S. Sullivant. The space of compatible full conditionals is a unimodular toric variety. *Journal of Symbolic Computation*, 41:196–209, 2006.

[62] P. Spellman, G. Sherlock, M. Zhang, V. Iyer, K. Anders, M. Eisen, P. Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Molecular Biology of the Cell*, 9(12):3273–3297, 1998.

[63] R.P. Stanley. *Enumerative combinatorics. Vol. 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

[64] J. Storey, W. Xiao, J. Leek, R. Tompkins, and R. Davis. Significance analysis of time course microarray experiments. *Proceedings of the National Academy of Sciences*, 102:12837–12842, 2005.

[65] M. Studený. Conditional independence relations have no finite completeness characterization. *Kybernetika*, 25:72–79, 1990.

[66] M. Studený. Structural semigraphoids. *International Journal of General Systems*, 22(2):207–217, 1994.

[67] M. Studený. *Probabilistic Conditional Independence Structures.* Springer Series in Information Science and Statistics. Springer, London, 2005.

[68] M. Studený, RR. Bouckaert, and T. Kocka. Extreme supermodular set functions over five variables. Research report n. 1977, Prague, 2000.

[69] B. Sturmfels. *Gröbner Bases and Convex Polytopes.* American Mathematical Society, Providence, 1996.

[70] J. Tits. Le problème des mots dans les groupes de coxeter. *Symposia Math.*, 1:175–185, 1968.

[71] D. Weinreich. The rank ordering of genotypic fitness values predicts genetic constraint on natural selection on landscapes lacking sign epistasis. *Genetics*, 171(3):1397–1405, 2005.

[72] K. Willbrand, F. Radvanyi, J.P. Nadal, J.P. Thiery, and T. Fink. Identifying genes from up-down properties of microarray expression series. *Bioinformatics*, 21(20):3859–3864, 2005.

[73] G.M. Ziegler. *Lectures on polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer, New York, 1995.