Algebraic Factor Analysis: Tetrads, Pentads and Beyond

Mathias Drton, Bernd Sturmfels and Seth Sullivant

March 10, 2006

Abstract

Factor analysis refers to a statistical model in which observed variables are conditionally independent given fewer hidden variables, known as factors, and all the random variables follow a multivariate normal distribution. The parameter space of a factor analysis model is a subset of the cone of positive definite matrices. This parameter space is studied from the perspective of computational algebraic geometry. Gröbner bases and resultants are applied to compute the ideal of all polynomial functions which vanish on the parameter space. These polynomials, known as model invariants, arise from rank conditions on a symmetric matrix under elimination of the diagonal entries of the matrix. Besides revealing the geometry of the factor analysis model, the model invariants also furnish useful statistics for testing goodness-of-fit.

1 Introduction

Factor analysis is a statistical method which was introduced in the 1920's and is widely used to examine numerical data, especially in the social sciences. In this method, the correlated observed variables are modeled as conditionally independent given some hidden (latent) variables. These hidden variables are the *factors*. Sometimes factor analysis serves as a tool for dimension-reduction; the possibly many observed variables are summarized by fewer factors. However, in many applications the focus is on interpreting the factors as unobservable theorized concepts. In fact, the desire to explain observed correlations between individuals' exam performances by the concept of intelligence was the driving force in the original development of factor analysis (Spearman, 1927).

Currently, statistical inference in factor analysis is often based exclusively on parametric representations and on maximum likelihood estimates computed via the EM algorithm (Rubin and Thayer, 1982). In the early days of factor analysis, however, much attention was directed to *model invariants*, that is, to polynomial equality relations that the model imposes on the entries of the covariance matrix of the observed variables. We refer to Harman (1976, Sect. 1) for some history. It should be noted that factor analysis also leads to inequality constraints (e.g. Bekker and de Leeuw, 1987; Harman, 1976, p. 117), which we do not address here. The best known invariants are the *tetrads*, also called *tetrad differences*, which arise in one-factor models. The name tetrad was chosen because the polynomial arises in one-factor analysis with four observed variables. For example, if $\Psi = (\psi_{ij}) \in \mathbb{R}^{4\times 4}$ is a covariance matrix in the parameter space of a one-factor analysis model, then there are, up to sign change, three tetrads, namely

$$\psi_{12}\psi_{34} - \psi_{14}\psi_{23}, \quad \psi_{14}\psi_{23} - \psi_{13}\psi_{24}, \quad \psi_{12}\psi_{34} - \psi_{13}\psi_{24}, \tag{1}$$

and all three tetrads evaluate to zero. Tetrads have played a major role throughout the history of factor analysis. They also appear in recent research, for example, in work on model identifiability (Grzebyk et al., 2004) and on dichotomized Gaussian models for multivariate binary variables (Cox and Wermuth, 2002). While tetrads are ubiquitous in the literature, there has been very little work attempting to find invariants of models with more than one factor. The work by Kelley (1935) who derived the *pentad*, a fifth degree polynomial vanishing over covariance matrices from two-factor models, constitutes the exception. Since then virtually no progress has been made towards determining the invariants of factor analysis models. Harman (1976, p. 77) summarizes the state of knowledge as follows: *"When the number of factors is greater than two the work of computing determinants of the fourth or higher order becomes so laborious that no explicit conditions corresponding to the tetrads or pentad criterion have been worked out."*

Computational difficulties aside, the apparent ease of data analysis using solely the parametric model representation has inhibited progress on the determination of higher-order invariants. However, parametric approaches are not without their problems. On one hand, the likelihood function of a factor analysis model may have multiple local maxima (Rubin and Thayer, 1982), rendering its maximization difficult. On the other hand, the use of information criteria such as BIC in exploratory factor analysis does not have a theoretically valid basis in the presence of singularities (Geiger et al., 2001). Hence, we believe that a better mathematical understanding of factor analysis models is needed. One step in this direction is the work of Ellis (2004) who applied algebraic topology to study singularities arising in factor analysis. Our interest lies in the algebraic geometry that expresses itself in the model invariants. This has a pragmatic side because the invariants can serve as useful statistics for testing model fit and for constraint-based model selection. Both the desire to find new test statistics as well as the wish for an understanding of the geometry of factor analysis constitute the motivation for this paper.

The paper is outlined as follows. We begin with a review of the factor analysis model (Section 2) and discuss the use of invariants as test statistics (Section 3). In Section 4 we place the problem of determining invariants in the framework of algebraic statistics (Pachter and Sturmfels, 2005; Pistone et al., 2001). In fact, this is one of the first studies in algebraic statistics which deals with continuous rather than discrete random variables.

We shall see in Section 5 that the tetrads form a Gröbner basis of the ideal of invariants of a one-factor model. This implies that all invariants of one-factor models can be written as polynomial combinations of tetrads, which is claimed correctly but without proof in Glymour et al. (1987, p. 84). For models with two or more factors,

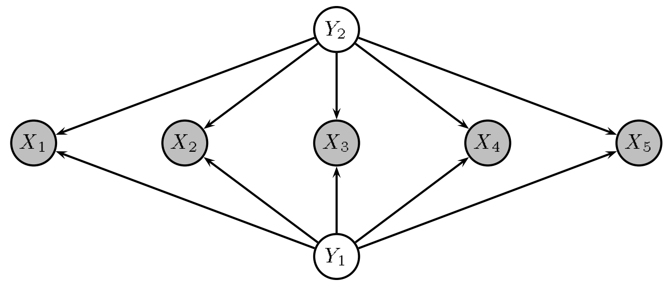

Figure 1: Graphical representation of the factor analysis model $\mathbf{F}_{5,2}$.

we performed extensive Gröbner basis computations, using `Macaulay 2` and `Singular`, for factor analysis models with up to nine observed variables and up to five factors. In Section 6, we show that multilinear resultants provide a useful method of finding individual invariants even when the whole ideal of invariants cannot be determined.

Our computational experiments lead to a series of conjectures and problems presented in Section 7. In particular, we conjecture for two-factor models that 3×3-minors and pentads generate the ideal of invariants. For models with arbitrarily many factors we conjecture that the ideal is generated by polynomials arising from consideration of submatrices whose size depends on the number of factors but not on the number of observed variables. We believe that these conjectures are of independent interest for commutative algebra. In Section 8 we propose some future research directions of statistical interest.

2 Factor analysis

Factor analysis concerns a Gaussian hidden variable model with p observed variables X_i, where $i \in [p] = \{1, \ldots, p\}$, and m hidden variables Y_j, where $j \in [m] = \{1, \ldots, m\}$. It is assumed that (X, Y) follows a joint multivariate normal distribution with positive definite covariance matrix. The *factor analysis model* $\mathbf{F}_{p,m}$ is defined by the requirement that the observed variables X_i, $i \in [p]$, are conditionally independent given the hidden variables Y_j, $j \in [m]$. The factor analysis model $\mathbf{F}_{p,m}$ can be visualized using the graphical model formalism (Lauritzen, 1996), in which the dependence structure between observed and hidden variables is encoded by an acyclic directed graph. This directed graph has the vertex set $\{X_1, \ldots, X_p, Y_1, \ldots, Y_m\}$, and the edges are $Y_j \to X_i$ for all $j \in [m]$ and $i \in [p]$, as shown in Figure 1 for $m = 2$ and $p = 5$.

We start out by deriving the following parametric representation of our model.

Proposition 1. *The factor analysis model* $\mathbf{F}_{p,m}$ *is the family of multivariate normal distributions* $\mathcal{N}_p(\mu, \Psi)$ *on* $\mathbb{R}^p$ *whose mean vector* μ *is an arbitrary vector in* $\mathbb{R}^p$ *and* *whose covariance matrix* Ψ *lies in the (non-convex) cone*

$$\begin{aligned} F_{p,m} &= \{\Sigma + \Lambda\Lambda^t \in \mathbb{R}^{p\times p} \,:\, \Sigma > 0 \text{ diagonal},\ \Lambda \in \mathbb{R}^{p\times m}\} \\ &= \{\,\Sigma + \Gamma \in \mathbb{R}^{p\times p} \,:\, \Sigma > 0 \text{ diagonal},\ \Gamma \geq 0 \text{ symmetric},\ \mathit{rank}(\Gamma) \leq m\}. \end{aligned} \tag{2}$$

Here the notation $A > 0$ means that A is a positive definite matrix (i.e., all eigenvalues are positive), and similarly $A \geq 0$ means that A is a positive semidefinite matrix.

Proof. Consider the joint covariance matrix of the fully observed model underlying $\mathbf{F}_{p,m}$,

$$\text{Cov}\begin{pmatrix} X \\ Y \end{pmatrix} \quad = \quad \begin{pmatrix} \Psi & \Lambda \\ \Lambda^t & \Phi \end{pmatrix}. \tag{3}$$

The entries of this matrix are constrained by the conditional independence statements

$$X_i \perp\!\!\!\perp X_j \,|\, \{Y_1, Y_2, \ldots, Y_m\} \qquad\qquad (1 \leq i < j \leq p), \tag{4}$$

which translate into the vanishing of the corresponding $(m+1)\times(m+1)$-determinants:

$$\det\begin{pmatrix} \psi_{ij} & \Lambda_{i*} \\ \Lambda_{j*}^t & \Phi \end{pmatrix} \quad = \quad \psi_{ij}\cdot\det(\Phi) \,-\, \Lambda_{i*}\cdot\text{adj}(\Phi)\cdot\Lambda_{j*}^t \quad = \quad 0. \tag{5}$$

We refer to Matus (2005) for a general discussion on how to translate conditional independence statements for Gaussian random variables into polynomial algebra.

The determinantal constraint (5) allows us to block-diagonalize the positive definite matrix (3) as follows:

$$\begin{pmatrix} \Sigma & 0 \\ 0 & \Phi \end{pmatrix} \quad = \quad \begin{pmatrix} I_p & -\Lambda\Phi^{-1} \\ 0 & I_m \end{pmatrix}\cdot\begin{pmatrix} \Psi & \Lambda \\ \Lambda^t & \Phi \end{pmatrix}\cdot\begin{pmatrix} I_p & 0 \\ -\Phi^{-1}\Lambda^t & I_m \end{pmatrix}. \tag{6}$$

Upon multiplication by $\det(\Phi) > 0$, the entry of the matrix $\Sigma = \Psi - \Lambda\cdot\Phi^{-1}\cdot\Lambda^t$ in row i and column j is equal to (5), so this positive definite matrix is diagonal if and only if X satisfies the model $\mathbf{F}_{p,m}$. This holds if and only if its covariance matrix Ψ has the form $\Psi = \Sigma + \Lambda\cdot\Phi^{-1}\cdot\Lambda^t$ if and only if Ψ is in the cone $F_{p,m}$. □

In what follows we generally identify the factor analysis model $\mathbf{F}_{p,m}$ with its parameter space $F_{p,m}$. The description given in Proposition 1 shows that $F_{p,m}$ is a parametrically presented subset of the space $\mathbb{R}^{\binom{p+1}{2}}$ of symmetric $p\times p$-matrices. The *dimension* $d = \dim(F_{p,m})$ of the model $F_{p,m}$ is the maximal rank of the Jacobian matrix of that parametrization. The *codimension* of $F_{p,m}$ is $\binom{p+1}{2} - d$.

Theorem 2. *The dimension and the codimension of the factor analysis model are*

$$\dim(F_{p,m}) = \min\left\{p(m+1) - \binom{m}{2}, \binom{p+1}{2}\right\},$$

$$\operatorname{codim}(F_{p,m}) = \max\left\{\binom{p-m}{2} - m, 0\right\}.$$

Thus the codimension of the factor analysis model is positive if and only if

$$p \geq \left\lfloor m + \frac{1}{2}\sqrt{8m+1} + \frac{1}{2} \right\rfloor + 1. \tag{7}$$

Proof. Using orthogonal transformations as in the QR-decomposition, every $\Psi \in F_{p,m}$ can be written as $\Psi = \Sigma + \Lambda\Lambda^t$ with $\Lambda = (\lambda_{ij})$ being lower-triangular in the sense that

$$\Lambda \;\in\; L_{p,m} \;=\; \left\{\Lambda \in \mathbb{R}^{p\times m} \mid \lambda_{ij} = 0 \text{ for all } 1 \leq i < j \leq m\right\};$$

see also Anderson and Rubin (1956, pp. 121 and 124). Thus the factor analysis model $F_{p,m}$ is the image of the following polynomial map:

$$\mathbb{R}^m_{>0} \times L_{p,m} \;\to\; \mathbb{R}^{\binom{p+1}{2}}, \;\; (\Sigma, \Lambda) \;\mapsto\; \Sigma + \Lambda\Lambda^t. \tag{8}$$

The coordinates of the parametrization (8) are

$$\psi_{ij} \;\;=\;\; \begin{cases} \sigma_{ii} + \sum_{r=1}^{\min(i,m)} \lambda_{ir}^2 & \text{if } i = j, \\ \sum_{r=1}^{\min(i,m)} \lambda_{ir}\lambda_{jr} & \text{if } i < j. \end{cases}$$

The dimension of the domain and the image space of (8) are $p(m+1) - \binom{m}{2}$ and $\binom{p+1}{2}$ respectively, so the minimum of these two numbers is an upper bound for $F_{p,m}$. To prove that this upper bound is tight, we will show that the Jacobian matrix J of the parametrization (8) has full rank almost everywhere. The Jacobian matrix has the form

$$J \;\;=\;\; \begin{array}{cc} & \begin{array}{cc} \sigma & \lambda \end{array} \\ \begin{array}{c} \psi_{ii} \\ \psi_{ij} \end{array} & \begin{pmatrix} I_p & B \\ 0 & A \end{pmatrix} \end{array}.$$

The entries in the unit matrix I_p on the upper left are

$$\frac{\partial \psi_{ij}}{\partial \sigma_{tt}} \;\;=\;\; \begin{cases} 1 & \text{if } t = i = j, \\ 0 & \text{else.} \end{cases} \tag{9}$$

The matrix J has full rank if and only if the $\binom{p}{2} \times \left(pm - \binom{m}{2}\right)$-matrix A has full rank. The entries of the latter matrix A are

$$\frac{\partial \psi_{ij}}{\partial \lambda_{st}} \;\;=\;\; \begin{cases} \lambda_{tt} & \text{if } i < j \text{ and } (i,j) = (t,s), \\ \lambda_{jt} & \text{if } i = s = t < j \text{ or } t < i = s < j, \\ \lambda_{it} & \text{if } t < i < j = s, \\ 0 & \text{else.} \end{cases} \tag{10}$$

More precisely, if we set $\psi_{i<} = (\psi_{i,i+1}, \ldots, \psi_{i,p})^t \in \mathbb{R}^{p-i}$ and $\lambda_{>j} = (\lambda_{j+1,j}, \ldots, \lambda_{p,j})^t \in \mathbb{R}^{p-j}$, then the matrix A can be written in the following form

$$
\begin{array}{cc}
 & \begin{matrix} \lambda_{11} & \lambda_{22} & \lambda_{33} & \ldots & \lambda_{mm} & \lambda_{>1} & \lambda_{>2} & \lambda_{>3} & \ldots & \lambda_{>m} \end{matrix} \\
\begin{matrix} \psi_{1<} \\ \psi_{2<} \\ \psi_{3<} \\ \vdots \\ \psi_{m<} \\ \vdots \\ \psi_{p-1,<} \end{matrix} &
\begin{pmatrix}
\lambda_{>1} & & & & & A_{11} & & & & \\
 & \lambda_{>2} & & & & A_{21} & A_{22} & & & \\
 & & \lambda_{>3} & & & A_{31} & A_{32} & A_{33} & & \\
 & & & \ddots & & \vdots & \vdots & \vdots & \ddots & \\
 & & & & \lambda_{>m} & A_{m1} & A_{m2} & A_{m3} & \ddots & A_{mm} \\
 & & & & & \vdots & \vdots & \vdots & \ddots & \vdots \\
 & & & & & A_{p-1,1} & A_{p-1,2} & A_{p-1,3} & \ldots & A_{p-1,m}
\end{pmatrix}
\end{array}
$$

where void entries are zero. The submatrices $A_{ii} = \lambda_{ii}\, I_{p-i}$ are diagonal, and, for $i > j$,

$$
A_{ij} = \begin{array}{cc}
 & \begin{matrix} \lambda_{j+1,j} & \ldots & \lambda_{i-1,j} & \lambda_{i,j} & \lambda_{i+1,j} & \lambda_{i+2,j} & \ldots & \lambda_{p,j} \end{matrix} \\
\begin{matrix} \psi_{i,i+1} \\ \psi_{i,i+2} \\ \vdots \\ \psi_{i,p} \end{matrix} &
\begin{pmatrix}
 & & & \lambda_{i+1,j} & \lambda_{ij} & & & \\
 & & & \lambda_{i+2,j} & & \lambda_{ij} & & \\
 & & & \vdots & & & \ddots & \\
 & & & \lambda_{p,j} & & & & \lambda_{ij}
\end{pmatrix}
\end{array}.
$$

This shows that the matrix A has full rank for almost all matrices $\Lambda \in L_{p,m}$, and we conclude that the stated formula for $\dim(F_{p,m})$ holds. The stated codimension is $\binom{p+1}{2}$ minus $\dim(F_{p,m})$, and the inequality (7) is gotten by solving $\binom{p-m}{2} > m$ for p. $\square$

Example 3. Let us consider the case of two factors $m = 2$. The model $F_{p,2}$ has positive codimension if and only if $p \geq 5$. For $p = 5$, Theorem 2 says that $F_{5,2}$ has codimension 1, so it is a hypersurface in the space of symmetric 5×5-matrix. The hypersurface is defined by the polynomial

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

This is the *pentad constraint* which was first derived by Kelley (1935). If Ψ is the covariance matrix of a distribution in the model $\mathbf{F}_{5,2}$ then $f(\Psi) = 0$, and the pentad f is the unique irreducible polynomial (up to scalar multiplication) with this property. In the next section we discuss the use of such invariants as test statistics, and in the subsequent sections we derive higher invariants using methods of computational algebra.

For $p = 4$, Theorem 2 says that $F_{4,2}$ has codimension 0, so it is full-dimensional in the space of symmetric 4×4-matrices. The theorem does not state that every positive definite matrix Ψ is in the model $F_{4,2}$. All it states is that the decomposition of Proposition 1,

$$\Psi \quad = \quad \begin{pmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{pmatrix} + \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \end{pmatrix} \cdot \begin{pmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & \lambda_{41} \\ \lambda_{12} & \lambda_{22} & \lambda_{32} & \lambda_{42} \end{pmatrix}, \tag{11}$$

imposes *no equality constraints* on the covariance matrix Ψ. But it does impose constraints in the form of inequations $f(\Psi) \neq 0$ and inequalities $f(\Psi) \geq 0$. We will discuss this issue in Section 4, after the algebraic set-up of ideals has been introduced. Note that the statistical problem of *parameter identification* corresponds to the algebraic problem of solving the equations (11) for the unknowns $\sigma_{ii}, \lambda_{ij}$ when the ψ_{ij} are given.

3 Invariants as test statistics

Let $\Psi \in \mathbb{R}^{p\times p}$ be a covariance matrix, that is, a positive definite symmetric $p\times p$-matrix, and let f be a polynomial in the entries ψ_{ij} of Ψ. We write $f(\Psi)$ for the evaluation of f using the numerical values of a particular matrix Ψ. The polynomial f is called an *invariant* of the factor analysis model $\mathbf{F}_{p,m}$ if $f(\Psi) = 0$ for all matrices Ψ in the parameter space $F_{p,m}$. Classical examples of invariants are the tetrad and pentad. If f is an invariant of $\mathbf{F}_{p,m}$ and Ψ is a covariance matrix such that $f(\Psi) \neq 0$ then we can deduce that $\Psi \notin F_{p,m}$. This suggests that model invariants can be used as statistics in tests of model fit. We propose the following approach for putting this on a sound basis.

Assume we observe a sample $X_1, X_2, \ldots, X_N$ of independent random vectors in $\mathbb{R}^p$ that are identically distributed according to the multivariate normal distribution $\mathcal{N}_p(\mu, \Psi)$ with mean vector $\mu \in \mathbb{R}^p$ and positive definite $p \times p$-covariance matrix Ψ. Let $\bar{X} = \frac{1}{N}\sum_{k=1}^N X_k$ be the sample mean vector and consider the sample covariance matrix

$$S \;=\; (s_{ij}) \quad = \quad \frac{1}{N-1}\sum_{k=1}^{N}(X_k - \bar{X})(X_k - \bar{X})^t.$$

Moreover, let f be an invariant of a hypothesized factor analysis model $\mathbf{F}_{p,m}$. The *sample invariant* $f(S)$ provides a consistent estimator of the true invariant evaluation $f(\Psi)$. The variance of $f(S)$, which we denote by $\mathrm{Var}_\Psi[f(S)]$, can be derived by computing appropriate moments of the Wishart distribution according to which the matrix $(N-1)\cdot S$ is distributed; compare Mardia et al. (1979, Sect. 3.4) and Wishart (1928a). The variance $\mathrm{Var}_\Psi[f(S)]$ is a polynomial function of the true covariance matrix Ψ. Replacing Ψ by the sample covariance matrix S in this polynomial yields the estimator $\mathrm{Var}_S[f(S)]$. Using this estimator, we can define the *standardized sample invariant*

$$Z_f \quad = \quad \frac{f(S)}{\sqrt{\mathrm{Var}_S[f(S)]}}. \tag{12}$$

Proposition 4. *Let f be an invariant of the model $\mathbf{F}_{p,m}$, and let Ψ be a covariance matrix such that f vanishes at Ψ but its gradient vector ∇f does not vanish at Ψ. Then, as the sample size N tends to infinity, the standardized sample invariant Z_f converges in distribution to a standard normal distribution:*

$$Z_f \quad \longrightarrow_d \quad \mathcal{N}(0,1),$$

Proof. The vectorization of $\sqrt{N}(S-\Psi)$ converges in distribution to a centered multivariate normal distribution. Applying the delta method (Shorack, 2000, p.279), we obtain that

$$\sqrt{N}\, f(S) \quad = \quad \sqrt{N}\,[f(S) - f(\Psi)] \quad \longrightarrow_d \quad \mathcal{N}(0, v_f)$$

with asymptotic variance

$$v_f = \lim_{N\to\infty} N\, \mathrm{Var}_\Psi[f(S)] > 0.$$

Since $\mathrm{Var}_\Psi[f(S)]/\,\mathrm{Var}_S[f(S)]$ converges in probability to one, it follows from Slutsky's theorem that

$$Z_f = \frac{\sqrt{\mathrm{Var}_\Psi[f(S)]}}{\sqrt{\mathrm{Var}_S[f(S)]}} \cdot \frac{\sqrt{N}\, f(S)}{\sqrt{N\,\mathrm{Var}_\Psi[f(S)]}} \longrightarrow_d \frac{1}{\sqrt{v_f}} \cdot \mathcal{N}(0, v_f) = \mathcal{N}(0,1).$$

$\square$

Remark 5. The sample invariant $f(S)$ is typically a biased estimator of $f(\Psi)$. However, if the expectation of $f(S)$ is of the form $\mathrm{E}_\Psi[f(S)] = h(N) \cdot f(\Psi)$, where $h(N)$ is a function of the sample size only, then one can consider the *bias-corrected sample invariant* $\tilde{f}(S) = f(S)/h(N)$. An analog of Proposition 4 holds when $f(S)$ is replaced by $\tilde{f}(S)$.

Example 6. We derive the standardized and bias-corrected sample invariants for the one-factor model $\mathbf{F}_{p,1}$. Let i, j, k, ℓ be four distinct indices in $[p]$ and consider the tetrad

$$f \quad = \quad \psi_{ik}\psi_{j\ell} - \psi_{i\ell}\psi_{jk}. \tag{13}$$

If Ψ is a covariance matrix in $F_{p,1}$ then the tetrad vanishes, i.e. $f(\Psi) = 0$.

The sample tetrad $f(S) = s_{ik}s_{j\ell} - s_{i\ell}s_{jk}$ is a consistent but biased estimator of $f(\Psi)$. However, the bias can be corrected as described in Remark 5 with the *bias-corrected sample tetrad* being equal to

$$\tilde{f}(S) \quad = \quad \frac{N-1}{N-2}\left(s_{ik}s_{j\ell} - s_{i\ell}s_{jk}\right).$$

For any covariance matrix Ψ, the variance of this unbiased estimator of $f(\Psi)$ is equal to

$$\begin{aligned}\mathrm{Var}_\Psi\left[\tilde{f}(S)\right] = {} & \frac{N+1}{(N-1)(N-2)}\det(\Psi_{\{i,j\}\times\{i,j\}}) \cdot \det(\Psi_{\{k,\ell\}\times\{k,\ell\}}) \\ & - \frac{1}{N-2}\det(\Psi_{\{i,j,k,\ell\}\times\{i,j,k,\ell\}}) + \frac{3}{N-2}\det(\Psi_{\{i,j\}\times\{k,\ell\}})^2.\end{aligned} \tag{14}$$

This expression was first computed by Wishart (1928b). If $\Psi \in F_{p,1}$, then

$$\det(\Psi_{\{i,j\}\times\{k,\ell\}}) \;=\; \psi_{ik}\psi_{j\ell} - \psi_{i\ell}\psi_{jk} \;=\; 0.$$

Thus the last term in (14) vanishes and we can use the estimate

$$\mathrm{Var}_S\left[\tilde{f}(S)\right] = \\ \frac{N+1}{(N-1)(N-2)}\det(S_{\{i,j\}\times\{i,j\}})\cdot\det(S_{\{k,\ell\}\times\{k,\ell\}}) - \frac{1}{N-2}\det(S_{\{i,j,k,\ell\}\times\{i,j,k,\ell\}}).$$

Following the recipe in (12), we introduce the standardized bias-corrected sample tetrad

$$Z_{\tilde{f}} \;=\; \frac{\tilde{f}(S)}{\sqrt{\mathrm{Var}_S\left[\tilde{f}(S)\right]}}.$$

This is an explicit expression which can be evaluated for any sample covariance matrix S arising from data X_i. If at least one of the four entries ψ_{ik}, $\psi_{j\ell}$, $\psi_{i\ell}$, ψ_{jk} is non-zero, then the gradient of the tetrad is non-zero at Ψ. Proposition 4 says that $Z_{\tilde{f}}$ has an asymptotic standard normal distribution $\mathcal{N}(0,1)$ when the sample size N tends to infinity.

Suppose now that f is an arbitrary polynomial invariant of the factor analysis model $\mathbf{F}_{p,m}$, and we wish to test the null hypothesis $H_f : f(\Psi) = 0$. In light of Proposition 4, we can do this by computing the corresponding standardized sample invariant Z_f and by comparing it to the appropriate quantile of the standard normal distribution $\mathcal{N}(0,1)$. More precisely, for chosen significance level $\alpha \in (0,1)$, we can find an interval $[-c_\alpha, c_\alpha]$ which, assuming H_f is true, contains the standardized sample invariant Z_f with (asymptotic) probability $1-\alpha$. If we observe a value z_f of Z_f that falls outside this interval, $z_f \notin [-c_\alpha, c_\alpha]$, then this constitutes evidence against H_f and, in particular, evidence against the hypothesized factor analysis model of which f is an invariant.

If the hypothesized factor analysis model is a hypersurface then consideration of a single invariant f is sufficient. This happens, for instance, in the case $p = 5, m = 2$ discussed in Example 3. Here f is the pentad and we only need to test H_f.

In general, however, the model structure will not be captured in a single polynomial invariant. Then we might want to employ a set I of invariants of the considered model to test model fit. For instance, I could be a set of ideal generators as in Section 4. A simple approach to working with several invariants is to employ Bonferroni's inequality, which suggests the consideration of the interval $[-c_{\alpha/|I|}, c_{\alpha/|I|}]$. This interval simultaneously contains all standardized sample invariants Z_f, $f \in I$, with probability at least $1-\alpha$. Therefore, if one or more observed values z_f fall outside the interval $[-c_{\alpha/|I|}, c_{\alpha/|I|}]$, then we have found statistical evidence against the hypothesized factor analysis model.

Other, perhaps more powerful approaches to simultaneous use of multiple invariants in data analysis are possible. For example, Hipp and Bollen (2003) consider a quadratic form in the tetrads to evaluate fit of a one-factor model. Spirtes et al. (2000) employ tests of vanishing tetrads to define scores for model selection in Gaussian graphical models with hidden variables. Their approach is implemented in the software `TETRAD`.

Tetrads appear to be the only invariants of factor analysis models that have seen routine use in data analysis. However, the approach we have outlined above is feasible also for other invariants such as the pentad and the higher invariants we determine subsequently. The only difficulty involved is the estimation of the variance of the sample invariant. We expect that recent work on moments of the Wishart distribution (Graczyk et al., 2005; Lu and Richards, 2001) can be applied fruitfully to overcome this difficulty.

4 Algebraic setup

We are interested in polynomial relations among the entries of a factor analysis covariance matrix $\Psi \in F_{p,m}$. The mathematical framework for studying such polynomial relations is that of commutative algebra and algebraic geometry (see e.g. Cox et al., 1997). Many algorithms from these fields are implemented in software for symbolic computation, and they provide powerful computational tools for the study of model invariants. The application of these tools in statistics is the focus of algebraic statistics (Pachter and Sturmfels, 2005; Pistone et al., 2001). While algebraic statistics has so far been predominantly occupied with the study of models for discrete random variables, the present study is one of the first in this emerging field which concerns continuous random variables. The set-up to be introduced is fairly general and can be used to study arbitrary Gaussian graphical models, not just the factor analysis model.

We fix the ring of polynomials with real coefficients in the $\binom{p+1}{2}$ indeterminates ψ_{ij}:

$$\mathbb{R}[\psi_{ij},\ i \leq j] \quad = \quad \mathbb{R}[\psi_{11}, \psi_{12}, \ldots, \psi_{1p}, \psi_{22}, \ldots, \psi_{2p}, \psi_{33}, \ldots, \psi_{pp}].$$

For any subset F of the symmetric matrices in $\mathbb{R}^{p \times p}$, let $I(F)$ be the set of all polynomials $f \in \mathbb{R}[\psi_{ij},\ i \leq j]$ such that $f(\Psi) = 0$ for all $\Psi \in F$. Clearly, $I(F)$ is an ideal in $\mathbb{R}[\psi_{ij},\ i \leq j]$; that is, the sum of two polynomials in $I(F)$ is again in $I(F)$, and the product of any polynomial in $\mathbb{R}[\psi_{ij},\ i \leq j]$ with a polynomial in $I(F)$ is in $I(F)$. According to Hilbert's basis theorem, every ideal is generated by a finite list of polynomials. We tacitly assume this finite representation for all the ideals which appear in the following discussion.

The object of our interest is the *ideal of invariants* of the model $\mathbf{F}_{p,m}$, which is the ideal $I_{p,m} = I(F_{p,m})$. Since membership in $F_{p,m}$ depends only on the off-diagonal entries of the matrix Ψ, we can regard $I_{p,m}$ as an ideal in the subring $\mathbb{R}[\psi_{ij},\ i < j]$ of $\mathbb{R}[\psi_{ij},\ i \leq j]$. If I is any ideal in the bigger polynomial ring $\mathbb{R}[\psi_{ij},\ i \leq j]$ then the intersection $I \cap \mathbb{R}[\psi_{ij},\ i < j]$ is an ideal in the smaller polynomial ring $\mathbb{R}[\psi_{ij},\ i < j]$. Passing to this intersection is the process of *elimination* of the variables $\psi_{11}, \ldots, \psi_{pp}$. The following result shows how the ideal $I_{p,m}$ can be computed using elimination.

Theorem 7. *Let $M_{p,m} \subseteq \mathbb{R}[\psi_{ij},\ i \leq j]$ be the ideal that is generated by all $(m+1) \times (m+1)$-minors of a symmetric matrix $\Psi \in \mathbb{R}^{p\times p}$. Then the ideal of invariants equals*

$$I_{p,m} \quad = \quad M_{p,m} \cap \mathbb{R}[\psi_{ij},\ i < j]. \tag{15}$$

Proof. The proof makes use of standard arguments from algebraic geometry, and all varieties $V(\,\cdot\,)$ are understood over the field $\mathbb{C}$ of complex numbers. Recall that the *variety* $V(M_{p,m})$ of the ideal $M_{p,m}$ is the set of common zeroes of the polynomials in $M_{p,m}$. This set coincides with the set of all symmetric $p\times p$-matrices of rank at most m. Let $F'_{p,m}$ denote the set of all $p\times p$-matrices of the form $\Psi = \Sigma + \Gamma$ where Σ is a diagonal matrix and $\Gamma \in V(M_{p,m})$. Thus $F'_{p,m}$ is the superset of our parameter space $F_{p,m}$ gotten by dropping the positive definiteness requirement. Since the cone of positive definite matrices is open (and hence Zariski dense) in the space of all symmetric matrices, we conclude that $F_{p,m}$ and $F'_{p,m}$ have the same Zariski closure in $\mathbb{C}^{p\times p}$. The projection of this Zariski closure onto the space off-diagonal entries is Zariski closed, and it coincides with the variety $V(I_{p,m})$ of the desired ideal $I_{p,m}$. On the other hand, every matrix in this projection is the image of a matrix Γ in $V(M_{p,m})$. We conclude that $V(I_{p,m})$ equals the projection of $V(M_{p,m})$ onto the off-diagonal coordinates.

By the Elimination Theorem (see Cox et al., 1997, §3.2) we have

$$V\big(M_{p,m} \cap \mathbb{R}[\psi_{ij},\ i<j]\big) \quad = \quad V(I_{p,m}). \tag{16}$$

Now, it is known that $M_{p,m}$ is a prime ideal and that the minors form a Gröbner basis for $M_{p,m}$ (see Conca, 1994; Sturmfels and Sullivant, 2005). The primality of $M_{p,m}$ implies that $M_{p,m} \cap \mathbb{R}[\psi_{ij},\ i<j]$ is prime as well. Since, by definition, $I_{p,m}$ is radical, we apply Hilbert's Nullstellensatz to (16) to conclude that $I_{p,m} = M_{p,m} \cap \mathbb{R}[\psi_{ij},\ i<j]$. □

Theorem 7 allows for the derivation of a finite generating set of the ideal $I_{p,m}$ using the method of Gröbner bases. This will be explained in Section 5. In the remainder of this section, we discuss some consequences and geometric aspects of Theorem 7, starting with some polynomials that obviously belong to the ideal $I_{p,m}$. Up to sign, there are

$$\frac{1}{2}\binom{p}{2(m+1)}\binom{2(m+1)}{m+1}$$

off-diagonal $(m+1)\times(m+1)$*-minors* of the matrix Ψ, that is, subdeterminants that do not involve any diagonal entries of Ψ. Such minors of size $m+1$ are trivially in $I_{p,m}$.

Corollary 8. *Let $p \geq 2(m+1)$ and choose two disjoint sets $R, C \subset [p]$ of cardinality $|R| = |C| = m+1$. Then the off-diagonal minor* $\det(\Psi_{R\times C})$ *is in $I_{p,m}$.*

Example 9. Let $m = 1$. Then the 2×2-off-diagonal minors of Ψ belong to $I_{p,1}$. For example, if $R = \{1,2\}$ and $C = \{3,4\}$ then this minor is the tetrad

$$\det(\Psi_{R\times C}) \quad = \quad \det\begin{pmatrix}\psi_{13} & \psi_{14}\\ \psi_{23} & \psi_{24}\end{pmatrix} \quad = \quad \psi_{13}\psi_{24} - \psi_{14}\psi_{23}.$$

If $p < 2(m+1)$ there are no off-diagonal minors. However, if $p \geq 2m+1$, it is still easy to determine some non-zero polynomials in $I_{p,m}$ by considering two minors that contain exactly one common diagonal entry ψ_{ii} and eliminating this diagonal entry.

Corollary 10. *Let $m \geq 2$ and $p \geq 2m+1$, and choose $i \in [p]$. Let (R, C) and $(\bar{R}, \bar{C})$ be two pairs of m-element subsets of $[p]\backslash\{i\}$ that are disjoint: $R \cap C = \emptyset = \bar{R} \cap \bar{C}$. Finally, let Ψ^0 be the symmetric matrix whose off-diagonal entries are the unknowns ψ_{ij} and whose diagonal entries are equal to 0. Then the following polynomial is in $I_{p,m}$:*

$$f_{i,R,C,\bar{R},\bar{C}} \quad = \quad \det(\Psi_{R\times C}) \cdot \det(\Psi^0_{(i,\bar{R})\times(i,\bar{C})}) - \det(\Psi_{\bar{R}\times\bar{C}}) \cdot \det(\Psi^0_{(i,R)\times(i,C)}).$$

The notation $(i, R) \times (i, C)$ indicates that the i-th row and column are arranged first.

Proof. The polynomial $f_{i,R,C,\bar{R},\bar{C}}$ lies in $\mathbb{R}[\psi_{ij},\ i < j]$. The two identities

$$\begin{aligned} \det(\Psi_{(i,R)\times(i,C)}) &= \det(\Psi^0_{(i,R)\times(i,C)}) + \psi_{ii} \cdot \det(\Psi_{R\times C}), \\ \det(\Psi_{(i,\bar{R})\times(i,\bar{C})}) &= \det(\Psi^0_{(i,\bar{R})\times(i,\bar{C})}) + \psi_{ii} \cdot \det(\Psi_{\bar{R}\times\bar{C}}) \end{aligned}$$

imply

$$f_{i,R,C,\bar{R},\bar{C}} \quad = \quad \det(\Psi_{R\times C}) \cdot \det(\Psi_{(i,\bar{R})\times(i,\bar{C})}) - \det(\Psi_{\bar{R}\times\bar{C}}) \cdot \det(\Psi_{(i,R)\times(i,C)}).$$

This is a polynomial linear combination of $(m+1) \times (m+1)$-minors of Ψ, so it lies in $M_{p,m}$. We conclude that $f_{i,R,C,\bar{R},\bar{C}}$ is in the right hand side of (15) and hence in $I_{p,m}$. $\square$

The *linear eliminant* $f_{i,R,C,\bar{R},\bar{C}}$ is a homogeneous polynomial of degree $2m+1$. For $m = 2$ and $m = 3$, the linear eliminants recover the tetrads and the pentads as follows:

Remark 11. If $m = 1$ and $p \geq 4$ then we can choose pairs (R, C) and $(\bar{R}, \bar{C})$ satisfying the assumptions of Corollary 10. The result is a polynomial combination of two tetrads:

$$f_{i,R,C,\bar{R},\bar{C}} \;=\; -\psi_{rc}\psi_{i\bar{r}}\psi_{i\bar{c}} + \psi_{\bar{r}\bar{c}}\psi_{ir}\psi_{ic} \;=\; -\psi_{i\bar{c}} \cdot \det(\Psi_{\{r,\bar{r}\}\times\{c,i\}}) + \psi_{ir} \cdot \det(\Psi_{\{\bar{c},c\}\times\{\bar{r},i\}}).$$

Example 12. Let $m = 2$ and $p = 5$. Then the polynomial $f_{i,R,C,\bar{R},\bar{C}}$ has degree five, and it does not depend on the choices of i, R, C, $\bar{R}$ and $\bar{C}$. Up to sign, it coincides with the pentad f which was displayed in Example 3. Note that the twelve monomials in the pentad f correspond to the twelve labeled cycles on the set of nodes $\{1, 2, 3, 4, 5\}$. The ideal $I_{5,2}$ is the principal ideal generated by the pentad; in symbols, $I_{5,2} = \langle f \rangle$.

The following proposition shows that linear eliminants are non-redundant invariants.

Proposition 13. *Let $m \geq 2$ and $p \geq 2m+1$. If $R \cup C = \bar{R} \cup \bar{C}$, then linear eliminant $f_{i,R,C,\bar{R},\bar{C}}$ is not in the ideal generated by the off-diagonal $(m+1) \times (m+1)$-minors of Ψ.*

Proof. Without loss of generality assume that $i = 1$ and $R \cup C = \{2, 3, \ldots, 2m+1\}$. Since $\mathrm{codim}(F_{2m+1,m}) > 0$, by Theorem 2, we can choose a symmetric matrix $\Psi \in \mathbb{R}^{p\times p}$ such that (i) $f_{1,R,C,\bar{R},\bar{C}}(\Psi) \neq 0$ and (ii) all off-diagonal entries in row $2m+2$ to p are zero. Then all off-diagonal $(m+1)\times(m+1)$-minors of the chosen matrix Ψ are zero. This shows that $f_{1,R,C,\bar{R},\bar{C}}$ cannot be a polynomial combination of the off-diagonal minors. $\square$

We believe that the following converse to Proposition 13 holds. As we will see, Conjecture 14 is part of a general series of finiteness conjectures about the ideals $I_{p,m}$.

Conjecture 14. *Let $m \geq 2$ and $p \geq 2m+2$. If $R\cup C \neq \bar{R}\cup\bar{C}$ then the linear eliminant $f_{i,R,C,\bar{R},\bar{C}}$ is in the ideal generated by the off-diagonal $(m+1)\times(m+1)$-minors of Ψ.*

We call the linear eliminants where $R\cup C = \bar{R}\cup\bar{C}$ the $(2m+1)$*-ads*. So when $m = 2$, we recover the pentads and when $m = 3$ we obtain *septads*.

We close this section with a discussion of the geometric role played by the ideal $I_{p,m}$ in the context of factor analysis. Recall that the variety $V(I_{p,m})$ is the set of all common zeros of the polynomials in the ideal $I_{p,m}$. Consider the following four statements:

(a) A polynomial vanishes on the parameter space $F_{p,m}$ if and only if it lies in $I_{p,m}$.

(b) The factor analysis model $\mathbf{F}_{p,m}$ is represented by the variety $V(I_{p,m})$.

(c) The parameter space $F_{p,m}$ coincides with the variety $V(I_{p,m})$.

(d) The closure of the parameter space $F_{p,m}$ coincides with the variety $V(I_{p,m})$.

Then statement (a) is true because this is how the ideal $I_{p,m}$ was defined. Statement (b) is vague, but it expresses the philosophy of this paper, so we simply declare it to be true. On the other hand, statement (c) is false with respect to every meaningful interpretation of what the statement may mean. In algebraic geometry, $V(I_{p,m})$ denotes the set of zeros of $I_{p,m}$ over the field $\mathbb{C}$ of complex numbers, and this is what was meant in the proof of Theorem 7. Considering the zeros of $I_{p,m}$ among positive-definite matrices, positive semi-definite matrices, or just real symmetric matrices, we get the inclusions

$$V_{\mathrm{pd}}(I_{p,m}) \subset V_{\mathrm{psd}}(I_{p,m}) \subset V_{\mathbb{R}}(I_{p,m}) \subset V(I_{p,m}).$$

So, meaningful interpretations of (c) may be that $F_{p,m}$ equals $V_{\mathrm{pd}}(I_{p,m})$, and that $V(I_{p,m})$ equals the set $F^{\mathbb{C}}_{p,m}$ of complex $p\times p$-matrices $\Psi = \Sigma + \Lambda\Lambda^t$ where Σ is diagonal and $\Lambda \in \mathbb{C}^{p\times m}$. Both of these statements are false as the following example shows.

Example 15. Let $p = 3$ and $m = 1$. Then $F^{\mathbb{C}}_{3,1}$ consists of all 3×3-matrices of the form

$$\begin{pmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{12} & \psi_{22} & \psi_{23} \\ \psi_{13} & \psi_{23} & \psi_{33} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} + \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{pmatrix} \cdot \begin{pmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} \end{pmatrix}. \tag{17}$$

The three off-diagonal identities imply

$$\lambda_{11}^2\psi_{23} - \psi_{12}\psi_{13} \;=\; \lambda_{21}^2\psi_{13} - \psi_{12}\psi_{23} \;=\; \lambda_{31}^2\psi_{12} - \psi_{13}\psi_{23} \;=\; 0. \tag{18}$$

This shows that a matrix in Ψ in $F_{3,1}^{\mathbb{C}}$ cannot have precisely one zero off-diagonal entry. But $V(I_{3,1})$ consists of all symmetric 3×3-matrices since $\mathrm{codim}(F_{3,1}) = 0$ and $I_{3,1} = \{0\}$. Hence $F_{3,1}^{\mathbb{C}}$ is a proper subset of $V(I_{3,1})$, and, likewise, $F_{3,1}$ is a proper subset of $V_{\mathrm{pd}}(I_{3,1})$.

Let us now come to statement (d). This statement is true over the field $\mathbb{C}$ of complex numbers. Every matrix in $V(I_{p,m})$ is the limit of matrices in $F_{p,m}^{\mathbb{C}}$. This follows from a non-trivial algebraic geometry result to the effect that, for the image of any polynomial map over $\mathbb{C}$, the usual closure coincides with the Zariski closure (see Cox et al., 1997, Proposition 7, p. 490). On the other hand, statement (d) is false over the real numbers. Namely, in our example, $V_{\mathrm{pd}}(I_{31})$ is the set of all positive-definite matrices. If Ψ is a positive-definite matrix with $\psi_{12} > 0$, $\psi_{13} > 0$ and $\psi_{23} < 0$, then (18) forces λ_{11} to be the square root of a negative number, so Ψ cannot be in the closure of $F_{3,1}$. A similar (but more complicated) analysis can be performed for the case $p = 4, m = 2$ starting from the equations given in (11).

In summary, in this paper we do not determine all the constraints satisfied by the parameter space $F_{p,m}$ of the factor analysis model. What we do determine is the set $I_{p,m}$ of all polynomial equation constraints. These characterize the closure of $F_{p,m}$ if we allow complex numbers. The polynomials in $I_{p,m}$ are the model invariants, and, as argued in Section 3, they can be used to derive novel test statistics for Gaussian graphical models.

5 Gröbner basis computations

We now focus on computing finite generating sets for the ideals $I_{p,m}$. Following Theorem 7, this can be done by equating to zero all $(m+1) \times (m+1)$-minors of an unknown symmetric $p \times p$-matrix Ψ, and then eliminating the off-diagonal unknowns ψ_{ii} from these equations. In computer algebra, there are two main methods for eliminating unknowns from a system of equations: Gröbner bases and resultants. In this section we present the Gröbner basis approach, while resultants will be featured in the next section. We shall assume familiarity with "Gröbner basics" at the level of Cox et al. (1997).

The complete answer to our problem is currently only known for the one-factor model ($m = 1$). Namely, as we shall see in Theorem 16, the tetrads provide a reduced Gröbner basis for the ideal $I_{p,1}$. For two or more factors ($m \geq 2$), we did numerous computations with the computer algebra systems `Macaulay2` and `Singular`. The results of these computations are presented in this section (see Tables 1 and 2 below). We shall return to these results in Section 7, where we offer some conjectures about the ideals $I_{p,m}$.

Let $m = 1$. For any four indices $i < j < k < \ell$ in $[p]$, we have the tetrads in (1). Since the first tetrad is the difference of the third and the second tetrad, it suffices to pick out the last two tetrads in (1). Let

$$\mathcal{T}_p \quad = \quad \{\underline{\psi_{ij}\psi_{k\ell}} - \psi_{ik}\psi_{j\ell},\ \underline{\psi_{i\ell}\psi_{jk}} - \psi_{ik}\psi_{j\ell} \mid 1 \leq i < j < k < \ell \leq p\}$$

be the set of $2\binom{p}{4}$ tetrads obtained in this way. As described by de Loera et al. (1995), the underlined terms are the leading terms with respect to a certain monomial order $\succ$ on $\mathbb{R}[\psi_{ij},\ i < j]$. (They call this monomial order the *thrackle order*.)

Theorem 16. *If $p \leq 3$ the ideal $I_{p,1}$ is the zero ideal. If $p \geq 4$, the set $\mathcal{T}_p$ is the reduced Gröbner basis of the ideal $I_{p,1}$ with respect to the monomial order $\succ$.*

Proof. The claim follows from Theorem 2.1 in de Loera et al. (1995). □

If we observe $p = 5$ variables, then the set $\mathcal{T}_5$ contains ten tetrads. In Harman (1976, p. 76) it is stated that one can find five of these tetrads such that *"any other conditions must be linearly dependent on the [five tetrads]."* Similarly, Hipp and Bollen (2003, p. 277) state that *"to detect the full set of redundant vanishing tetrads when there are more than four variables requires careful algebraic derivation"* and *"in the case of the five-indicator model there will be five nonredundant vanishing tetrads."* Harman (1976, p. 418) outlines a justification of his claim, which, however, is valid only if all ψ_{ij} are non-zero. In a strict algebraic sense, Harman's claim is incorrect because none of the ten tetrads in $\mathcal{T}_5$ is a polynomial linear combination of the other nine tetrads.

Moving on to the general case $m \geq 2$, we now demonstrate how to compute a minimal generating set of the ideal $I_{p,m}$ by means of two software packages for algebraic geometry.

Example 17 (p=7, m=2 in Macaulay 2). The first software we used is the program `Macaulay 2` due to Grayson and Stillman (1998). To compute a minimal generating set of the ideal $I_{7,2}$ using `Macaulay 2`, we use the following sequence of six commands:

```
R = QQ[p11,p22,p33,p44,p55,p66,p77,p12,p13,p14,p15,p16,p17,p23,p24,p25,p26,p27,
       p34,p35,p36,p37,p45,p46,p47,p56,p57,p67, MonomialOrder=>Eliminate 7];
Psi = matrix{{p11,p12,p13,p14,p15,p16,p17},
             {p12,p22,p23,p24,p25,p26,p27},
             {p13,p23,p33,p34,p35,p36,p37},
             {p14,p24,p34,p44,p45,p46,p47},
             {p15,p25,p35,p45,p55,p56,p57},
             {p16,p26,p36,p46,p56,p66,p67},
             {p17,p27,p37,p47,p57,p67,p77}};
M72 = minors(3,Psi);
I72 = ideal selectInSubring(1,gens gb M72);
mingens I72
codim I72, degree I72
```

The command `I72 = ideal selectInSubring(1,gens gb M72)` performs the actual elimination step of deriving $I_{7,2}$ from $M_{7,2}$. The command `mingens I72` outputs 56 polynomials which minimally generate the ideal $I_{7,2}$. This list includes 35 polynomials of degree three and 21 polynomials of degree five. The latter are 21 pentads like

```
 p36p37p45p47p56-p35p37p46p47p56-p36p37p45p46p57+p35p36p46p47p57
+p34p37p46p56p57-p34p36p47p56p57+p35p37p45p46p67-p35p36p45p47p67
-p34p37p45p56p67+p34p35p47p56p67+p34p36p45p57p67-p34p35p46p57p67.
```

Of the 35 polynomials of degree three, 21 are off-diagonal minors like

```
p26p35p47-p25p36p47-p26p34p57+p24p36p57+p25p34p67-p24p35p67.
```

The remaining 14 polynomials are sums of off-diagonal minors. In fact, we can replace these by 14 off-diagonal minors such that the resulting 35 polynomials are minimal generators of $I_{7,2}$. This validates the entry for $p = 7, m = 2$ in Table 2. Finally, the last command line informs us that the variety $V(I_{7,2})$ has codimension 8 and degree 259.

The notion of "degree" requires an explanation. Next to the codimension, this is the most important invariant of an algebraic variety. Suppose that V is a variety of codimension c in $\mathbb{C}^r$. Then the *degree* of V is the number of points in $V \cap L$ where L is a general affine subspace of dimension c in $\mathbb{C}^r$. The case $c = 1$ is familiar: if V is a hypersurface, defined by the vanishing of one polynomial f, then the degree of V equals the degree of f, and this is the number of intersection points of V with a general line.

Table 1 summarizes what we know about the codimension and the degree of the factor analysis model $V(I_{p,m})$ for $m \leq 5$ and $p \leq 9$. In this section we discuss the $m = 2$ and $m = 3$ columns, and in Section 6 we discuss $(m, p) = (4, 8)$ and $(m, p) = (5, 9)$.

	$m=1$		$m=2$		$m=3$		$m=4$		$m=5$	
p	codim	deg	codim	deg	codim	deg	codim	deg	codim	deg
3	0	1	0	1	0	1	0	1	0	1
4	2	4	0	1	0	1	0	1	0	1
5	5	11	1	5	0	1	0	1	0	1
6	9	26	4	45	0	1	0	1	0	1
7	14	57	8	259	3	91	0	1	0	1
8	20	120	13	1232	7	1368	2	98	0	1
9	27	247	19	5319	12	14232	6	?	1	54

Table 1: Codimensions and degrees for the factor analysis model.

For $m = 2$ our computations suggest that the ideal $I_{p,2}$ is always generated by minors and pentads, but so far we have been unable to prove this for general p. See Section 7 for a discussion of this conjecture. For $m \geq 3$ we found that the computer algebra system `Singular` performs better than `Macaulay 2`. Here is a non-trivial computation.

Example 18 (p=8, m=3 in Singular). To compute the ideal $I_{8,3}$ using `Singular` (Greuel et al., 2001), the following sequence of eight commands can be used:

```
ring R = 0,(p11,p22,p33,p44,p55,p66,p77,p88,
            p12,p23,p34,p45,p56,p67,p78,p18,
            p13,p24,p35,p46,p57,p68,p17,p28,
            p14,p25,p36,p47,p58,p16,p27,p38, p15,p26,p37,p48),dp;
matrix Psi[8][8] = p11,p12,p13,p14,p15,p16,p17,p18,
                   p12,p22,p23,p24,p25,p26,p27,p28,
                   p13,p23,p33,p34,p35,p36,p37,p38,
                   p14,p24,p34,p44,p45,p46,p47,p48,
                   p15,p25,p35,p45,p55,p56,p57,p58,
                   p16,p26,p36,p46,p56,p66,p67,p68,
                   p17,p27,p37,p47,p57,p67,p77,p78,
                   p18,p28,p38,p48,p58,p68,p78,p88;
ideal M83 = minor(Psi,4);
ideal I83 = eliminate(M83, p11*p22*p33*p44*p55*p66*p77*p88);
nvars(basering) - dim(I83); mult(I83); // codimension and degree
ideal I83min = mstd(I83)[2];  // minimal generators
betti(I83min);
I83min;
```

In the first command, which declares the polynomial ring, we list the variables in an order different from the order chosen in the `Macaulay 2` code described in Example 17. Together with the option `dp`, this variable ordering determines a monomial order which we found to be advantageous. Other monomial orders we considered led to slower or even non-terminating computations. With this particular order, a modern workstation requires less than 10 minutes to execute the remaining seven commands.

The function `eliminate` carries out the elimination step of deriving $I_{8,3}$ from $M_{8,3}$. The codimension and degree of the variety $V(I_{8,3})$ are equal to 7 and 1368, respectively, as computed in the line following the elimination. The function `mstd` permits to compute the list of polynomials `I83min`, which minimally generate the ideal $I_{8,3}$. According to the command `betti(I83min)`, this generating set consists of 14 polynomials of degree four, 260 polynomials of degree seven, and 168 polynomials of degree eight. They are:

degree 4: Twelve polynomials are off-diagonal 4×4-minors. The other two polynomials are sums of off-diagonal minors and can be replaced by off-diagonal minors.

degree 7: The 260 polynomials of degree seven come in two flavors.

(a) Of the 260 polynomials, $120 = \binom{8}{7} \cdot 15$ are sums of linear eliminants of the type described in Proposition 13. These linear eliminants are equal to $f_{i,R,C,\bar{R},\bar{C}}$ with $R \cup C = \bar{R} \cup \bar{C}$ in $[p] \setminus \{i\}$ and we call them *septads.* A septad has 168 terms, and an example of a septad appearing in the minimal generator is

```
p23p13p46p68p17p28p47+p67p18p13p24p28p36p47-p13p24p68p17p28p36p47-...
...+p12p17p36p26p37p48^2-p23p17p16p26p37p48^2+p13p16p27p26p37p48^2 .
```

While 102 of the 120 polynomials in our output are septads, the other twelve can be replaced by septads without affecting the minimal generator property.

(b) The remaining $260 - 120 = 140$ polynomials are of ideal-theoretic nature. They are not polynomial linear combinations of the off-diagonal minors and septads. However, the squares of the 140 polynomials are polynomial combinations of off-diagonal minors and septads. Hence these invariants vanish at a covariance matrix Ψ whenever the off-diagonal minors and septads do.

degree 8: These 168 minimal generators are also of ideal-theoretic nature. Their squares are polynomial linear combinations of off-diagonal minors and septads.

Table 2 summarizes our knowledge about the composition of a minimal generating set of the factor analysis ideal $I_{p,m}$ for $m \leq 3$ and $p \leq 9$. This table suggests various conjectures about the ideals $I_{p,m}$ for general p, and we will discuss these in Section 7.

	$m = 1$	$m = 2$		$m = 3$			
p	deg 2	deg 3	deg 5	deg 4	deg 7	deg 7	deg 8
4	2	—	—	—	—	—	—
5	10	0	1	—	—	—	—
6	30	5	6	—	—	—	—
7	70	35	21	0	15	0	20
8	140	140	56	14	120	140	168
9	252	420	126	126	540	1386	756
	tetrad	*minor*	*pentad*	*minor*	*septad*	*ideal-theor.*	*ideal-theor.*

Table 2: The degrees of the minimal generators of the ideals $I_{p,m}$.

6 Multilinear resultants

Resultants are a technique for simultaneously eliminating m unknowns from a system of $m+1$ polynomial equations. While Gröbner bases can be used to perform this task, we must remember that Gröbner bases are a very general method. They often compute too much and are hence too inefficient. The applicability of resultants is much more limited, but wherever they do apply, resultants tend to outperform Gröbner bases. For general introductions to resultants see Gel'fand et al. (1994) and Sturmfels (1997, 2002).

In our algebraic study of factor analysis, we found the Gröbner basis computations of Section 5 to be infeasible for $m \geq 4$. Instead we did some computations using the *multilinear resultant.* In this section we explain this technique, and how it was used to derive the degrees 54 and 98 for $(m,p) = (5,9)$ and $(m,p) = (4,8)$ in Table 1.

Consider a set of $n+1$ multilinear polynomials $f_0, \ldots, f_n$ in n unknowns $x_1, \ldots, x_n$:

$$f_j \quad = \sum_{i_1, i_2, \ldots, i_n \in \{0,1\}} a^j_{i_1 i_2 \cdots i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} \qquad (j = 0, 1, \ldots, n). \tag{19}$$

Here the coefficients $a^j_{i_1 i_2 \cdots i_n}$ are regarded as unknowns. The total number of these coefficients is $2^n \cdot (n+1)$, and they generate a polynomial ring which we denote by

$$\mathbb{R}[\mathbf{a}] \quad := \quad \mathbb{R}\big[\, a^j_{i_1 \cdots i_n} \,:\, i_1, \ldots, i_n \in \{0,1\}, j \in \{0, \ldots, n\}\big].$$

We write $\mathbb{R}[\mathbf{a}, \mathbf{x}]$ for the polynomial ring generated by the coefficients $a^j_{i_1 i_2 \cdots i_n}$ and the unknowns $x_1, \ldots, x_n$, and we consider the ideal $\langle f_0, f_1, \ldots, f_n \rangle$ in $\mathbb{R}[\mathbf{a}, \mathbf{x}]$ which is generated by the multilinear polynomials (19). We have the following result from algebra:

Theorem 19. *The elimination ideal $\langle f_0, f_1, \ldots, f_n \rangle \cap \mathbb{R}[\mathbf{a}]$ is generated by an irreducible polynomial $\mathcal{R}(\mathbf{a})$ which is homogeneous of degree $n!$ in the coefficients of each f_j.*

Proof. This follows from the results in Gel'fand et al. (1994, Sect. 8.2.A), applied to the special case when the toric variety X_A is the product of n projective lines in its Segre embedding. The corresponding polytope Q is the n-dimensional standard cube, which has normalized volume equal to $n!$. The polynomial $\mathcal{R}(\mathbf{a})$ is the Chow form of X_A. $\square$

We call $\mathcal{R}(\mathbf{a})$ the *n-th multilinear resultant.* Here are the first three cases:

Example 20. If $n = 1$ then $f_0 = a^0_0 + a^0_1 x_1$ and $f_1 = a^1_0 + a^1_1 x_1$. Their resultant equals

$$\mathcal{R}(\mathbf{a}) \;=\; a^0_0 a^1_1 - a^0_1 a^1_0 \;=\; a^1_1 \cdot f_0 - a^0_1 \cdot f_1 \;=\; \begin{vmatrix} a^0_0 & a^0_1 \\ a^1_0 & a^1_1 \end{vmatrix}. \tag{20}$$

Example 21. If $n = 2$ then we are considering a system of three bilinear equations

$$\begin{aligned} f_0 &= a^0_{00} + a^0_{10} x_1 + a^0_{01} x_2 + a^0_{11} x_1 x_2, \\ f_1 &= a^1_{00} + a^1_{10} x_1 + a^1_{01} x_2 + a^1_{11} x_1 x_2, \\ f_2 &= a^2_{00} + a^2_{10} x_1 + a^2_{01} x_2 + a^2_{11} x_1 x_2. \end{aligned}$$

Their resultant has the following determinantal representation:

$$\mathcal{R}(\mathbf{a}) = \begin{vmatrix} a^0_{00} & a^0_{10} & a^0_{01} \\ a^1_{00} & a^1_{10} & a^1_{01} \\ a^2_{00} & a^2_{10} & a^2_{01} \end{vmatrix} \cdot \begin{vmatrix} a^0_{10} & a^0_{01} & a^0_{11} \\ a^1_{10} & a^1_{01} & a^1_{11} \\ a^2_{10} & a^2_{01} & a^2_{11} \end{vmatrix} - \begin{vmatrix} a^0_{00} & a^0_{01} & a^0_{11} \\ a^1_{00} & a^1_{01} & a^1_{11} \\ a^2_{00} & a^2_{01} & a^2_{11} \end{vmatrix} \cdot \begin{vmatrix} a^0_{00} & a^0_{10} & a^0_{11} \\ a^1_{00} & a^1_{10} & a^1_{11} \\ a^2_{00} & a^2_{10} & a^2_{11} \end{vmatrix} \tag{21}$$

This polynomial has degree six but it is quadratic in the coefficients of each f_j.

Example 22. If $n = 3$ then the coefficients of f_0, f_1, f_2, f_3 form an 4×8-matrix

$$A \quad = \quad \begin{bmatrix} a^0_{000} & a^0_{001} & a^0_{010} & a^0_{011} & a^0_{100} & a^0_{101} & a^0_{110} & a^0_{111} \\ a^1_{000} & a^1_{001} & a^1_{010} & a^1_{011} & a^1_{100} & a^1_{101} & a^1_{110} & a^1_{111} \\ a^2_{000} & a^2_{001} & a^2_{010} & a^2_{011} & a^2_{100} & a^2_{101} & a^2_{110} & a^2_{111} \\ a^3_{000} & a^3_{001} & a^3_{010} & a^3_{011} & a^3_{100} & a^3_{101} & a^3_{110} & a^3_{111} \end{bmatrix} \tag{22}$$

Let $[ijkl]$ denote the determinant of the 4×4-submatrix of A with columns i, j, k, l. Then the multilinear resultant $\mathcal{R}(\mathbf{a})$ is the determinant of the following 6×6-matrix:

$$\begin{bmatrix} [0124] & [0234] & [0146]-[0245] & [0346]-[0247] & [0456] & [0467] \\[2ex] \begin{matrix}[0125]\\ +[0134]\end{matrix} & \begin{matrix}[1234]\\ +[0235]\end{matrix} & \begin{matrix}[0147]+[0156]\\ -[0345]-[1245]\end{matrix} & \begin{matrix}-[1247]+[0356]\\ -[0257]+[1346]\end{matrix} & \begin{matrix}[1456]\\ +[0457]\end{matrix} & \begin{matrix}[1467]\\ +[0567]\end{matrix} \\[2ex] [0135] & [1235] & [0157]-[1345] & -[1257]+[1356] & [1457] & [1567] \\[2ex] [0126] & [0236] & -[1246]+[0256] & [2346]-[0267] & [2456] & [2467] \\[2ex] \begin{matrix}[0136]\\ +[0127]\end{matrix} & \begin{matrix}[1236]\\ +[0237]\end{matrix} & \begin{matrix}-[1247]-[1346]\\ +[0257]+[0356]\end{matrix} & \begin{matrix}-[0367]-[1267]\\ +[2356]+[2347]\end{matrix} & \begin{matrix}[3456]\\ +[2457]\end{matrix} & \begin{matrix}[2567]\\ +[3467]\end{matrix} \\[2ex] [0137] & [1237] & -[1347]+[0357] & -[1367]+[2357] & [3457] & [3567] \end{bmatrix} \tag{23}$$

The derivation of this 6×6-matrix is explained in Sturmfels (2002, Proposition 4.10). The matrix (23) differs from the 6×6-matrix displayed by Sturmfels (2002) because the latter had some typographical errors. These typos have now been corrected in (23).

These formulas can be applied to algebraic factor analysis as follows. Recall from Theorem 7 that the ideal $I_{p,m}$ is computed by eliminating the diagonal unknowns ψ_{ii} from the $(m+1) \times (m+1)$-minors of an indeterminate covariance matrix $\Psi = (\psi_{ij})$. Multilinear resultants are relevant for this task because the minors are multilinear polynomials in the ψ_{ii}. For instance, the derivation of the pentad constraint can be interpreted as an evaluation of the first multilinear resultant (20). Likewise, the second multilinear resultant (21) can be used to produce non-trivial invariants in $I_{p,m}$ when $p \geq 2$, $m \geq 8$.

The general method for computing invariants of the factor analysis model using resultants is the following. We choose any subset $D = \{d_1, \ldots, d_n\}$ of cardinality n in $[p]$ and any subsets $R_0, \ldots, R_n, C_0, \ldots, C_n$ each of cardinality $m + 1 - n$ in $[p] \backslash D$ such that $R_i \cap C_i = \emptyset$ for all i. This choice of $D, R_\bullet, C_\bullet$ gives rise to an invariant as follows.

Theorem 23. *Let* $x_j = \psi_{d_j d_j}$ *for* $j = 1, \ldots, n$ *and* $f_k = \det(\Psi_{DR_k \times DC_k})$ *for* $k = 0, \ldots, n$. *The evaluation of the n-th multilinear resultant at* $f_0, f_1, \ldots, f_n$ *is an element of* $\mathbb{R}[\psi_{ij},\ i \leq j]$. *This polynomial lies in the ideal* $I_{p,m}$, *and it is either zero or it is homogeneous of degree*

$$(m - \frac{n}{2} + 1) \cdot (n+1)\,! \tag{24}$$

Proof. The proof uses the methods of Gel'fand et al. (1994) and is omitted here. □

It is instructive to examine this theorem for small values of n. If $n = 1$ (as in Example 20) then the degree (24) equals $2m + 1$ and we recover the linear eliminant $f_{i,R,C,\bar{R},\bar{C}}$ of Proposition 10. Here $D = \{i\}$, $R_0 = R$, $R_1 = \bar{R}$, $C_0 = C$ and $C_1 = \bar{C}$. If $n = 2$ (as in Example 21) then the invariant constructed in Theorem 23 is homogeneous of degree $6m$. If $n = 3$ (as in Example 22) then the invariant constructed in Theorem 23 is homogeneous of degree $24m - 12$. In particular, the degree (24) is 108 for $m = 5$.

Example 24 (p=9, m=5 using resultants). This is the case appearing in the lower right corner of Table 1. The variety $V(I_{9,5})$ is a hypersurface of degree 54 in the space of symmetric 9×9-matrices. The irreducible polynomial defining this hypersurface is the greatest common divisor (gcd) of all multilinear resultants constructed for $n = 3$ as in Theorem 23. Each resultant has degree 108, and their gcd was found to have degree 54. In fact, we observed that it suffices to take the gcd of only two such resultants.

For a concrete instance take $D = \{1, 2, 3\}$, $R_0 = \{4, 5, 6\}$, $C_0 = \{7, 8, 9\}$, $R_1 = \{4, 5, 7\}$, $C_1 = \{6, 8, 9\}$, $R_2 = \{4, 6, 8\}$, $C_2 = \{5, 7, 9\}$, $R_3 = \{4, 7, 9\}$, $C_3 = \{5, 6, 8\}$, and let f be the degree 108 polynomial constructed by Theorem 23. It is infeasible to express f as a sum of monomials. (Note that the number of monomials of degree 108 in the 36 unknowns ψ_{ij} exceeds 10^{33}). However, the 6×6-determinant (23) offers an efficient representation of the invariant f. Namely, if the ψ_{ij} are replaced by linear forms in one or two parameters, then this 6×6-determinant can be evaluated rapidly. From this we see that f is non-zero and that it factors as the product of four irreducible polynomials, having degrees 18, 18, 18 and 54. The last factor is the generator of $I_{9,5}$.

Let g denote the degree 54 invariant which generates the ideal $I_{9,5}$. We have represented g as the gcd of several polynomials of degree 108, but this representation specifies g only up to a multiplicative constant. Up to this constant, we can evaluate g numerically at a covariance matrix Ψ by choosing a random matrix Ψ_0, introducing an unknown, say t, computing the gcd of several resultants of the form $f(\Psi + t \cdot \Psi_0) \in \mathbb{R}[t]$, and evaluating the resulting polynomials at $t = 0$. Repeating the computation at a second covariance matrix Ψ' with the same Ψ_0 is an efficient scheme for evaluating the ratio $g(\Psi)/g(\Psi')$.

Example 25 (p=8, m=4 using Macaulay 2 and resultants). The variety $V(I_{8,4})$ has codimension two in the space of symmetric 8×8-matrices. In order to compute its degree, we intersect the projective variety of $I_{8,4}$ with a general two-dimensional plane. We do this computation over a finite field, using the following `Macaulay 2` commands:

```
S = ZZ/101[r,s,t];
R = ZZ/101[p11,p22,p33,p44,p55,p66,p77,p88,r,s,t,MonomialOrder=> Eliminate 8];
f = map(R,S,{r,s,t});
    p12 = f(random(1,S));   p13 = f(random(1,S));   p14 = f(random(1,S));
    p15 = f(random(1,S));   p16 = f(random(1,S));   p17 = f(random(1,S));
    p18 = f(random(1,S));   p23 = f(random(1,S));   p24 = f(random(1,S));
```

```
    p25 = f(random(1,S));   p26 = f(random(1,S));   p27 = f(random(1,S));
    p28 = f(random(1,S));   p34 = f(random(1,S));   p35 = f(random(1,S));
    p36 = f(random(1,S));   p37 = f(random(1,S));   p38 = f(random(1,S));
    p45 = f(random(1,S));   p46 = f(random(1,S));   p47 = f(random(1,S));
    p48 = f(random(1,S));   p56 = f(random(1,S));   p57 = f(random(1,S));
    p58 = f(random(1,S));   p67 = f(random(1,S));   p68 = f(random(1,S));
                                                    p78 = f(random(1,S));
Psi = matrix {{p11,p12,p13,p14,p15,p16,p17,p18},
              {p12,p22,p23,p24,p25,p26,p27,p28},
              {p13,p23,p33,p34,p35,p36,p37,p38},
              {p14,p24,p34,p44,p45,p46,p47,p48},
              {p15,p25,p35,p45,p55,p56,p57,p58},
              {p16,p26,p36,p46,p56,p66,p67,p68},
              {p17,p27,p37,p47,p57,p67,p77,p78},
              {p18,p28,p38,p48,p58,p68,p78,p88}}
G = gens gb minors(5,Psi);   J = ideal selectInSubring(1,G);
codim J, degree J
```

The output of repeated runs verifies that the degree of $V(I_{8,4})$ equals 98. However, this computation gives no information about the minimal generators of $I_{8,4}$, and at present we do not even know the smallest degree of a non-zero polynomial in $I_{8,4}$.

On the other hand, we obtain a large number of non-trivial invariants of degree 24 by applying Theorem 23 with $n = 2$. Namely, for any choice of indices $D, R_\bullet, C_\bullet$ we set

$$f_i(x_1, x_2) \quad = \quad \det \begin{pmatrix} x_1 & \psi_{d_1,d_2} & \psi_{d_1,c_{i0}} & \psi_{d_1,c_{i1}} & \psi_{d_1,c_{i2}} \\ \psi_{d_2,d_1} & x_2 & \psi_{d_2,c_{i0}} & \psi_{d_2,c_{i1}} & \psi_{d_1,c_{i2}} \\ \psi_{r_{i0},d_1} & \psi_{r_{i0},d_2} & \psi_{r_{i0},c_{i0}} & \psi_{r_{i0},c_{i1}} & \psi_{r_{i0},c_{i2}} \\ \psi_{r_{i1},d_1} & \psi_{r_{i1},d_2} & \psi_{r_{i1},c_{i0}} & \psi_{r_{i1},c_{i1}} & \psi_{r_{i1},c_{i2}} \\ \psi_{r_{i2},d_1} & \psi_{r_{i2},d_2} & \psi_{r_{i2},c_{i0}} & \psi_{r_{i2},c_{i1}} & \psi_{r_{i2},c_{i2}} \end{pmatrix} \qquad \text{for } \; i = 0, 1, 2.$$

Here, $R_i = \{r_{i0}, r_{i1}, r_{i2}\}$ and $C_i = \{c_{i0}, c_{i1}, c_{i2}\}$. The invariant of degree 24 is obtained evaluating the formula (21), in which we abbreviate the coefficients of f_i by $a^i_{jk} = a^i_{jk}(\psi)$. Just as in Example 24, it is impossible to write this invariant as a sum of monomials, but it is very easy to evaluate it numerically using the determinantal representation (21).

7 Conjectures about generators of the ideals $I_{p,m}$

In Sections 4-6 we discussed the problem of computing a finite generating set for the ideal $I_{p,m}$ of invariants of the factor analysis model $\mathbf{F}_{p,m}$. Our computational results suggest some natural conjectures and problems about the structure of this generating set, and we believe that these will be of independent interest to commutative algebraists.

The pattern we found for small m, and that we hope is true for larger m, is that as p gets large the generators of $I_{p,m}$ depend on only a certain fixed number of random variables. This type of *finiteness* property has frequent occurrences in algebraic statistics. See Allman and Rhodes (2004) and Santos and Sturmfels (2003) for two instances.

The prototypical conjecture of this type is the following.

Conjecture 26. *The ideal of the two-factor model, $I_{p,2}$, is minimally generated by $5\binom{p}{6}$ off-diagonal 3×3-minors and $\binom{p}{5}$ pentads.*

Conjecture 26 is supported by the numerical evidence compiled in Table 2. Note that all of the minors and pentads described involve at most six random variables. However, unlike in the case of the 1-factor model, there does not seem to be any term order that makes the collection of off-diagonal minors and pentads a Gröbner basis for $I_{p,2}$.

The most natural term order we discovered in our computations has already been introduced in Example 18. In general, this is the lexicographic term order with $\phi_{ij} \succ \phi_{kl}$ if the circular distance between i and j is smaller than the circular distance between k and l. (To compute the circular distance between i and j, place the numbers $1, \ldots, p$ equispaced around a circle and measure the distance by taking the shortest path around the circle between i and j.) If these circular distances are the same, we declare $\phi_{ij} \succ \phi_{kl}$ if $i < k$. We call this term order the *circular lexicographic term order*. The fact that each of the ideals $I_{p,m}$ is an m-th secant ideal, together with the machinery developed in Sturmfels and Sullivant (2005), and our computations led us to the following conjecture.

Conjecture 27. *The circular lexicographic term order is 2-delightful for $I_{p,1}$. More specifically, the reduced Gröbner for $I_{p,2}$ consists of certain explicitly constructed polynomials of odd degree less than p with squarefree initial terms.*

The notion of a delightful term order was developed in Sturmfels and Sullivant (2005) and the technical details are beyond the scope of this short section. However, the basic idea is that, if Conjecture 27 is true, information about the initial ideal and reduced Gröbner basis of the two-factor ideal $I_{p,2}$ can be deduced from the reduced Gröbner basis of the one-factor ideal $I_{p,1}$ using graph theory. We refer the interested reader to Sturmfels and Sullivant (2005) for information about delightful term orders.

Moving to three factors, the situation seems even more complicated. As we have already seen in Table 2, the minors and septads are not enough to generate the entire ideal $I_{p,3}$. The minors and septads do, however, determine the parameter space $F_{p,3}$ set-theoretically in all the examples we were able to compute. This leads us to suspect:

Conjecture 28. *Let $J_{p,3}$ be the ideal generated by all the 4×4 off-diagonal minors and all the septad linear eliminants. Then the radical of $J_{p,3}$ is the prime ideal $I_{p,3}$.*

We do not have any concrete conjectures about the generators or Gröbner bases of $I_{p,3}$. Note that each of the minors and septads involve at most eight random variables. Is it possible that this type of finiteness behavior continues for larger m? For a subset $A \subset [p]$ denote by $I_{A,m}$ the ideal $I_{|A|,m}$ with indices labeled by the elements of A.

Question 29. For each integer $m \geq 1$ does there exist another integer $s(m)$ such that

$$I_{p,m} \quad = \sum_{A \subset [p], |A| = s(m)} I_{A,m} \qquad \text{for all } p > s(m) \ ?$$

Does there exist a different number $t(m)$ where the equality holds up to radical?

Our computations aside, there is some theoretical evidence to suggest that the set-theoretic finiteness result might hold with $t(m) = 2m + 2$. Namely, we can show that the set-theoretic finiteness result does hold in the complement of a certain hypersurface.

Proposition 30. *Let Φ be a symmetric $p \times p$-matrix with $p \geq 2m + 2$ and suppose that no $m \times m$-minor of Φ is zero. Then $\Phi \in F_{p,m}$ if and only if $\Phi_{A,A} \in F_{2m+2,m}$ for all subsets $A \subset [p]$ of cardinality $|A| = 2m + 2$.*

Proof. The "only if" direction is trivial. To prove the "if" direction we first need to refer to two simple results about factor analysis models. First of all, if $\Psi \in F_{p,m}$ and has no $m \times m$ minor equal to zero, then the decomposition $\Psi = \Sigma + \Gamma$ with Σ diagonal and $\operatorname{rank}(\Gamma) = m$ is unique. Furthermore, if Γ is a symmetric rank m matrix and Λ and K are $p \times m$ matrices with $\Gamma = \Lambda\Lambda^t = KK^t$ then there exists an orthogonal matrix Q such that $\Lambda = KQ$. See Anderson and Rubin (1956) for both of these results.

Now we prove the "if" direction by induction on p. The induction base is $p = 2m+2$ in which case the statement is vacuous. Suppose that Φ satisfies $\Phi_{A,A} \in F_{2m+2,m}$ for all subsets $A \subset [p]$ with $|A| = 2m + 2$. Denote by Φ^+ the submatrix $\Phi_{[p-1],[p-1]}$, by Φ_- the submatrix $\Phi_{[p]\setminus\{1\},[p]\setminus\{1\}}$ and by Φ^+_- the submatrix $\Phi_{[p]\setminus\{1,p\},[p]\setminus\{1,p\}}$. In other words, Φ^+ is the upper left $(p-1) \times (p-1)$ submatrix, Φ_- is the lower right $(p-1) \times (p-1)$ submatrix, and Φ^+_- is the $(p-2) \times (p-2)$ submatrix where Φ^+ and Φ_- overlap.

By the induction hypothesis, Φ^+ and Φ_- belong to $F_{p-1,m}$, and so there exist unique Σ^+, Σ_- diagonal and Γ^+, Γ_- of rank m such that $\Phi^+ = \Sigma^+ + \Gamma^+$ and $\Phi_- = \Sigma_- + \Gamma_-$. Furthermore, these provide a unique representation for the overlap $\Phi^+_- = \Sigma^+_- + \Gamma^+_-$, where Σ^+_- and Γ^+_- are the common overlapping portions of Σ^+ and Σ_- and, respectively, Γ^+ and Γ_-. Let $\Gamma^+ = \Lambda\Lambda^t$ and $\Gamma_- = KK^t$ be rank m factorizations of Λ^+ and Λ_-. Because of the uniqueness in the overlap Λ^+_-, and the second result taken from Anderson and Rubin (1956) mentioned above, we can assume that the last $p-2$ rows of Λ coincide with the first $p-2$ rows of K. Now form the $p \times p$ matrices

$$\bar{\Sigma} = \begin{pmatrix} \Sigma^+_{11} & 0 \\ 0 & \Sigma_- \end{pmatrix} \quad \text{and} \quad \bar{\Gamma} = \bar{\Lambda}\bar{\Lambda}^t \quad \text{where} \quad \bar{\Lambda} = \begin{pmatrix} \Lambda_1 \\ K \end{pmatrix}.$$

Set $\bar{\Phi} = \bar{\Sigma} + \bar{\Gamma}$. We claim that $\bar{\Phi} = \Phi$ and hence $\Phi \in F_{p,m}$, which completes the proof. Note that trivially $\bar{\phi}_{ij} = \phi_{ij}$ except possibly for the pair $(i,j) = (1,p)$. So we must show that $\bar{\phi}_{1p} = \phi_{1p}$. Let B and C be disjoint subsets of $[p]\setminus\{1,p\}$ of cardinality

$|B| = |C| = m$. Then the following three $(m+1) \times (m+1)$ minors are equal to zero:

$$\begin{vmatrix} \bar{\Phi}_{1\times C} & \bar{\phi}_{1p} \\ \bar{\Phi}_{B\times C} & \bar{\Phi}_{B\times p} \end{vmatrix} = \begin{vmatrix} \Phi_{1\times C} & \bar{\phi}_{1p} \\ \Phi_{B\times C} & \Phi_{B\times p} \end{vmatrix} = \begin{vmatrix} \Phi_{1\times C} & \phi_{1p} \\ \Phi_{B\times C} & \Phi_{B\times p} \end{vmatrix} = 0.$$

The first and last minors are zero because $\bar{\Phi}_{A\times A}$ and $\Phi_{A\times A}$ belong to $F_{2m+2,p}$ for any A with $|A| = 2m+2$ by assumption. The middle minor is zero because it is entry-wise equal to the first minor. Since by assumption $\det(\Phi_{B\times C}) \neq 0$ we deduce that $\phi_{1p} = \bar{\phi}_{1p}$. □

The proof technique we present does not allow the extra condition in Proposition 30 to be dropped and thus we are a long way from answering Question 29.

8 Computer algebra for Gaussian models: the next steps

The research presented in this paper merely scratches the surface of possible applications of computer algebra techniques for studying the factor analysis model and more general models for Gaussian random variables. In this final section, we highlight two such problems: maximum likelihood (ML) estimation and the study of singularities. Another important direction is the design of test statistics from higher invariants (pentads, septads, etc.) by computing moments of the Wishart distribution (cf. Section 3).

Since much statistical inference in factor analysis is currently based on ML estimation, it seems natural to ask what computational algebra has to say about computing ML estimators for factor analysis. As a starting point, we computed the *maximum likelihood degree* for the one-factor model with four observed variables. The ML degree (Catanese et al., 2005) of a statistical model is the number of nontrivial complex zeros of the critical equations for generic data. We found that the factor analysis model $\mathbf{F}_{4,1}$ has ML degree 57. In what follows, we describe how we discovered the number 57 to be the number of complex zeros of the critical equations for ML estimation in this model.

Consider a matrix $\Psi \in F_{p,m}$ with decomposition $\Psi = \Sigma + \Gamma$, where Σ is diagonal with positive entries and Γ is positive semidefinite with $\mathrm{rank}(\Gamma) \leq m$. Then we have

$$\Psi^{-1} \quad = \quad \Sigma^{-1} - \Psi^{-1}\Gamma\Sigma^{-1}.$$

The matrix $\Psi^{-1}\Gamma\Sigma^{-1}$ is symmetric of rank $\leq m$. Moreover, the fact that $\Psi - \Sigma = \Gamma$ is positive semidefinite implies that $\Sigma^{-1} - \Psi^{-1} = \Psi^{-1}\Gamma\Sigma^{-1}$ is also positive semidefinite. Hence, the inverse of $\Psi \in F_{p,m}$ can be written as

$$\Psi^{-1} \quad = \quad T - KK^t,$$

where T is diagonal with positive entries and $K \in \mathbb{R}^{p\times m}$. In the case of $p = 4$ and $m = 1$, this amounts to writing Ψ^{-1} in terms of $\tau = (\tau_1, \ldots, \tau_4)$ and $\kappa = (\kappa_1, \ldots, \kappa_4)$ as

$$\Psi^{-1}(\tau, \kappa) \quad = \quad \mathrm{diag}(\tau) - \kappa\kappa^T. \tag{25}$$

As in Section 3, let $\bar{X}$ and S be the sample mean vector and the sample covariance matrix computed from a sample of N random vectors. For ML estimation it is more convenient to work with the matrix $\tilde{S} = (N-1)/N \cdot S$ instead of S. The model of multivariate normal distributions $\mathcal{N}(\mu, \Psi)$ has the log-likelihood function

$$\ell(\mu, \Psi) = -\frac{N}{2}\log\det(\Psi) - \frac{N}{2}\operatorname{tr}(\tilde{S}\Psi^{-1}) - \frac{N}{2}(\bar{X}-\mu)^t\Psi^{-1}(\bar{X}-\mu); \qquad (26)$$

compare Mardia et al. (1979, Sect. 4.1.1). This function is maximized in μ by setting $\mu = \bar{X}$, which leads to the vanishing of the quadratic form appearing as third term in (26). Therefore, we can find the maximizer in Ψ by maximizing the expression

$$\log\det(\Psi^{-1}) \;-\; \operatorname{tr}(\tilde{S}\Psi^{-1}). \qquad (27)$$

Here Ψ runs over $F_{4,1}$. By plugging (25) into (27), we can write this expression as a function of the eight unknowns $\tau_1, \tau_2, \tau_3, \tau_4, \kappa_1, \kappa_2, \kappa_3, \kappa_4$. Taking partial derivatives and setting them to zero, we obtain a system of eight equations in eight unknowns. These are the *likelihood equations* of the factor analysis model $\mathbf{F}_{4,1}$ in rational function form:

$$\begin{aligned}
\frac{1}{\det(\Psi^{-1}(\tau,\kappa))}\cdot\frac{\partial\det(\Psi^{-1}(\tau,\kappa))}{\partial\tau_i} &= \operatorname{tr}\left[\tilde{S}\frac{\partial\Psi^{-1}(\tau,\kappa)}{\partial\tau_i}\right], \qquad i = 1,\ldots,4,\\
\frac{1}{\det(\Psi^{-1}(\tau,\kappa))}\cdot\frac{\partial\det(\Psi^{-1}(\tau,\kappa))}{\partial\kappa_i} &= \operatorname{tr}\left[\tilde{S}\frac{\partial\Psi^{-1}(\tau,\kappa)}{\partial\kappa_i}\right], \qquad i = 1,\ldots,4.
\end{aligned}$$

These equations can be made polynomial by multiplying through by $\det(\Psi^{-1}(\tau,\kappa))$. Clearing the denominator introduces many additional solutions to the system, namely all noninvertible matrices of the form $\Psi^{-1}(\tau,\kappa)$. However, these extraneous solutions can be removed using an operation called *saturation*. After saturation, we discover that the solution set of the polynomial likelihood equations consists of 57 isolated (complex) points. Further work needs to be done to determine how many of these can be statistically meaningful local maxima and to extend these results to larger models.

The second problem we wish to illustrate is that of singularities. The singularities of statistical models play an important, though under-appreciated, role. Models with singularities do not form curved exponential families, which invalidates, for example, the theoretical basis of model selection using information criteria like BIC (Geiger et al., 2001). Near a singularity, such criteria require correction terms (Watanabe, 2001). This is especially important in factor analysis because, among other singularities, each parameter space $F_{p,m}$ is singular along $F_{p,m-1}$ making the selection of the number of factors difficult. The fact that $F_{p,m-1}$ is contained in the singular locus of $F_{p,m}$ can be proved by observing that $V(I_{p,m})$ is the m-th secant variety of $V(I_{p,1})$ and by appealing to general results in algebraic geometry about singularities of secant varieties.

As a first step towards a better understanding of singularities, we computed the singular loci of some of the small factor analysis models. The computation of the singular locus is done, in `Macaulay 2` or `Singular`, by augmenting $I_{p,m}$ by the $c \times c$-minors of the Jacobian matrix of any generating set of $I_{p,m}$, where $c = \mathrm{codim}(I_{p,m})$.

Example 31. For $p = 4$ and $m = 1$, the singular locus consists of all matrices Ψ which have at most one non-zero off-diagonal entry. Thus, the singular locus consists of one symmetry class of covariance matrices. For instance, one type of matrix in the singular locus has the form

$$\begin{pmatrix} \psi_{11} & \psi_{12} & 0 & 0 \\ \psi_{12} & \psi_{22} & 0 & 0 \\ 0 & 0 & \psi_{33} & 0 \\ 0 & 0 & 0 & \psi_{44} \end{pmatrix}.$$

Example 31 generalizes to an arbitrary number of observed random variables when the number of factors is fixed at $m = 1$.

Proposition 32. *A matrix $\Psi \in V_{\mathbb{R}}(I_{p,1})$ is a singularity of the one factor model if and only if Ψ has at most one non-zero off diagonal entry.*

Proof. This can be seen by computing derivatives of the tetrads and by using results in toric geometry. We omit the details. $\square$

For more factors, the situation is considerably more complicated.

Example 33. Let $p = 5$ and $m = 2$. In this case $F_{5,2}$ is the pentad hypersurface. The singular locus of this hypersurface has dimension 11, and consists of two symmetry classes of singularities. A representative from the first symmetry class is the set of matrices of the form

$$\begin{pmatrix} \psi_{11} & 0 & 0 & 0 & 0 \\ 0 & \psi_{22} & \psi_{23} & \psi_{24} & \psi_{25} \\ 0 & \psi_{23} & \psi_{33} & \psi_{34} & \psi_{35} \\ 0 & \psi_{24} & \psi_{34} & \psi_{44} & \psi_{45} \\ 0 & \psi_{25} & \psi_{35} & \psi_{45} & \psi_{55} \end{pmatrix}.$$

The second symmetry class is more complicated. A representative set consists of those matrices Ψ that satisfy all tetrads not involving ψ_{12}. Note that this set of singular points contains $F_{5,1}$. In total there are five elements of the first symmetry class and ten elements in the second. To apply algebraic geometry techniques to these singularities for model selection requires a careful analysis of the way the various singular sets intersect.

It is an open problem to determine the singular locus of the factor analysis models $F_{p,m}$ in general. Even the dimension of that singular locus is unknown to us.

Finally we stress that while the algebraic statistical study in this paper was confined to factor analysis models, problems analogous to the ones described here appear in other classes of Gaussian graphical models. The study of such other models, which need not involve hidden variables, opens up a broad range of directions for future research.

Acknowledgements. Much of this work was conducted while Mathias Drton held a postdoctoral position sponsored by the Center of Pure and Applied Mathematics at UC Berkeley. Mathias Drton also acknowledges support from the National Science Foundation (DMS-0505612). Seth Sullivant was supported by an NSF Graduate Research Fellowship. Bernd Sturmfels was also supported by the NSF (DMS-0456960).

References

E. S. Allman and J. A. Rhodes. Phylogenetic ideals and varieties for the general Markov model. preprint, `arXiv:math.AG/0410604`.

T. W. Anderson and H. Rubin. Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. V*, pages 111–150, Berkeley and Los Angeles, 1956. University of California Press.

P. A. Bekker and J. de Leeuw. The rank of reduced dispersion matrices. *Psychometrika*, 52(1): 125–135, 1987.

F. Catanese, S. Hosten, A. Khetan, and B. Sturmfels. The maximum likelihood degree. `arXiv:math.AG/0406533`, to appear in *American Journal of Mathematics*, 2005.

A. Conca. Gröbner bases of ideals of minors of a symmetric matrix. *Journal of Algebra* 166 (2): 406–421, 1994.

D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms.* Springer-Verlag, New York, second edition, 1997.

D. R. Cox and N. Wermuth. On some models for multivariate binary variables parallel in complexity with the multivariate Gaussian distribution. *Biometrika*, 89(2):462–469, 2002.

J. A. de Loera, B. Sturmfels, and R. R. Thomas. Gröbner bases and triangulations of the second hypersimplex. *Combinatorica*, 15:409–424, 1995.

S. Ellis. Instability of factor analysis. *Proc. Amer. Math. Soc.*, 132: 1805-1822, 2004.

D. Geiger, D. Heckerman, H. King, and C. Meek. Stratified exponential families: graphical models and model selection. *Annals of Statistics*, 29(2):505–529, 2001.

I. M. Gel'fand, M. M. Kapranov and A. V. Zelevinsky. *Discriminants, Resultants and Multidimensional Determinants.* Birkhäuser, Boston, 1994.

C. Glymour, R. Scheines, P. Spirtes, and K. Kelly. *Discovering Causal Structure.* Academic Press, London, 1987.

P. Graczyk, G. Letac, and H. Massam. The hyperoctahedral group, symmetric group representations and the moments of the real Wishart distribution. *Journal of Theoretical Probability.* 18 (1):1–42, 2005.

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

G.-M. Greuel, G. Pfister, and H. Schönemann. SINGULAR 2.0. A computer algebra system for polynomial computations. Centre for Computer Algebra, University of Kaiserslautern, 2001. `http://www.singular.uni-kl.de`.

M. Grzebyk, P. Wild, and D. Chouanière. On identification of multi-factor models with correlated residuals. *Biometrika*, 91:141–151, 2004.

H. Harman. *Modern Factor Analysis.* University of Chicago Press, Chicago, third edition, 1976.

J. R. Hipp and K. A. Bollen. Model fit in structural equation models with censored, ordinal, and dichotomous variables: Testing vanishing tetrads. *Sociological Methodology*, 33:267–305, 2003.

T. L. Kelley. *Essential Traits of Mental Life*, volume 26 of *Harvard Studies in Education.* Harvard University Press, Cambridge, MA, 1935.

S. L. Lauritzen. *Graphical Models.* Clarendon Press, Oxford, UK, 1996.

I-L. Lu and D. St. P. Richards. MacMahon's master theorem, representation theory, and moments of Wishart distributions. *Advances in Applied Mathematics.* 27 (2-3): 531–547, 2001.

K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis.* Academic Press, London, 1979.

F. Matus. Conditional independences in Gaussian vectors and rings of polynomials. in G. Kern-Isberner, W. Rödder and F. Kulmann (eds.): *Conditionals, Information, and Inference,* (WCII 2002 Hagen), Lecture Notes in Computer Science, Springer, Volume 3301, pp. 152-161.

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

G. Pistone, E. Riccomagno, and H. P. Wynn. *Algebraic Statistics. Computational Commutative Algebra in Statistics.* Chapman & Hall/CRC, Boca Raton, FL, 2001.

D. B. Rubin and D. T. Thayer. EM algorithms for ML factor analysis. *Psychometrika*, 47(1): 69–76, 1982.

F. Santos and B. Sturmfels. Higher Lawrence configurations. *Journal of Combinatorial Theory: Series A,* **103** (2003) 151-164.

G. R. Shorack. *Probability for Statisticians.* Springer-Verlag, New York, 2000.

C. Spearman. *The Abilities of Man.* Macmillan, New York, 1927.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search.* MIT Press, Cambridge, MA, second edition, 2000.

B. Sturmfels. Introduction to resultants, in: D. Cox, B. Sturmfels (eds.), *Applications of Computational Algebraic Geometry.* Proceedings of Symp. in Applied Math., 53, American Mathematical Society, 1997, pp. 25-39.

B. Sturmfels. *Solving Systems of Polynomial Equations.* American Mathematical Society, CBMS Lecture Series, No 97, Providence, Rhode Island, 2002.

B. Sturmfels and S. Sullivant. Combinatorial secant varieties. `arXiv:math.AC/0506223`, to appear in *Quarterly Journal of Pure and Applied Mathematics*, special issue on the occasion of the sixtieth birthday of Robert MacPherson.

S. Watanabe. Algebraic analysis for non-identifiable learning machines. *Neural Computation* 13: 899–933, 2001.

J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A:32–52, 1928a.

J. Wishart. Sampling errors in the theory of two factors. *The British Journal of Psychology*, 19: 180–187, 1928b.

Authors' addresses:

Mathias Drton, Department of Statistics, University of Chicago, Chicago, IL 60637, USA, `drton@galton.uchicago.edu`

Bernd Sturmfels, Department of Mathematics, University of California, Berkeley, CA 94720, USA, `bernd@math.berkeley.edu`

Seth Sullivant, Society of Fellows and Department of Mathematics, Harvard University, Cambridge, MA 02138, USA `seths@math.harvard.edu`