# GROEBNER BASIS AND STRUCTURAL MODELING

Min Lim

KDU UNIVERSITY COLLEGE

Jerry Brunner

UNIVERSITY OF TORONTO

April 25, 2012

Correspondence should be sent to Jerry Brunner, Department of Statistics, University of Toronto, Toronto, ON M5S 3G3, Canada.

E-Mail: brunner@utstat.toronto.edu
Phone: 905-828-3816
Fax: 905-569-4730
Website: http://www.utstat.toronto.edu/∼brunner

GROEBNER BASIS

**Abstract**

For many structural statistical models, parameter identifiability is established by seeking the roots of systems of multivariate polynomials. The mathematical state of the art is the theory of Groebner basis, which yields algorithmic methods that are widely implemented in symbolic mathematics software. Basic definitions and results are provided, using a notation that is appropriate for statistical modeling. Examples are given, and it is shown how Groebner bases can yield tests of model correctness even when the parameters cannot be uniquely estimated. Strengths and weaknesses of Groebner basis methods are discussed.

Key words: Errors in variables, Model identification, Latent class model, Latent variables, Structural equation modeling, Misclassification, Measurement error, Symbolic computation

## Introduction

A statistical model asserts that the probability distribution $P$ of an observable data set depends upon a parameter $\omega \in \Omega$. The parameter might be a vector that includes an unknown probability distribution, so "non-parametric" cases are included. Any function of $\omega$ that is also a function of $P$ is said to be *identifiable*, and useful estimation and inference for it is possible. For structural statistical models (Koopmans and Reiersøl, 1950) including the classical structural equation models (for example Bentler and Weeks, 1980; Jöreskog, 1978; McArdle and McDonald, 1984; McDonald, 1978), identifiability is not guaranteed, and must be checked on a case by case basis. Suppose the objective is inference about $\boldsymbol{\theta} = g(\omega) \in \Theta \subset \mathbb{R}^t$. For convenience, $\boldsymbol{\theta}$ will be called the "parameter" and $\Theta$ will be called the "parameter space," even though $\boldsymbol{\theta}$ is actually a function of the underlying parameter $\omega$, and $\Theta$ is the image of the underlying parameter space $\Omega$.

In any case, the identifiability of $\boldsymbol{\theta}$ is in question. The standard approach is to show that $\boldsymbol{\theta}$ is a composite function of the distribution $P$, by establishing that it is a function of $\boldsymbol{\Sigma} = m(P) \in \mathcal{M} \subset \mathbb{R}^d$. $\boldsymbol{\Sigma}$ is usually a well-chosen collection of moments or one-to-one functions of the moments, so $\mathcal{M}$ will be called the *moment space*. The notation deliberately suggests that $\boldsymbol{\Sigma}$ is a covariance matrix and this will be the primary application, but $\boldsymbol{\Sigma}$ could include a vector of means as well as the unique elements of a covariance matrix, or for multinomial data it could be a vector of probabilities. Another set of examples is provided by Skrondal and Rabe-Hesketh's (2004) "reduced form" parameters.

When the functions involved are chosen properly, $\boldsymbol{\Sigma}$ depends upon $\boldsymbol{\theta}$, giving rise to a system of *moment structure equations* $\boldsymbol{\Sigma} = \sigma(\boldsymbol{\theta})$. If the function $\sigma$ is one-to-one when restricted to $\Theta$, then $\boldsymbol{\theta}$ is identifiable. If the function $\sigma$ is not onto $\mathcal{M}$, there are points in $\mathcal{M}$ that are not the images of any point in $\Theta$. In this case the model is capable of being falsified by empirical data. Clearly, identifiability is neither a necessary nor a sufficient condition for the possibility of testing model correctness.

For the classical structural equation models and others — for example the polynomial models of Wall and Amemiya (2000, 2003) or measurement models in which both the observed and latent variables are categorical — the moment structure equations are polynomials, or at worst ratios of

polynomials. Writing $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_t)'$ and $\boldsymbol{\Sigma} = (\sigma_1, \ldots, \sigma_d)'$, they take the form

$$\frac{P_j(\boldsymbol{\theta})}{Q_j(\boldsymbol{\theta})} = \sigma_j \iff P_j(\boldsymbol{\theta}) - \sigma_j Q_j(\boldsymbol{\theta}) = 0 \tag{1}$$

for $j = 1, \ldots, d$, where $P_j(\boldsymbol{\theta})$ and $Q_j(\boldsymbol{\theta})$ are polynomials in $\theta_1, \ldots, \theta_t$ with $Q_j(\boldsymbol{\theta}) \neq 0$. Polynomials like the ones set to zero in Expression (1) will be called *moment structure polynomials*.

Thus, the question of model identification can be resolved by seeking the simultaneous roots of a finite set of polynomials. Finding the roots of polynomials is a strictly mathematical topic, with a history that extends to ancient Babylonia around 2000 B.C.E. (Sitwell, 2010, Ch. 6). For systems involving more than two polynomial equations, Bézout's *General theory of algebraic equations* (1779) represented the most notable advance since classical times.

The current mathematical state of the art is represented by the theory of Groebner basis (Buchberger, 1965; translation 2006). For any set of polynomials, a Groebner basis is another set of specially chosen polynomials with the same roots. That is, let $G(\boldsymbol{\theta})$ be a Groebner basis for the polynomials in (1). Then the set of $\boldsymbol{\theta}$ values that satisfy $G(\boldsymbol{\theta}) = \mathbf{0}$ is exactly the solution set of (1). Groebner basis equations can be much easier to solve than the original system, and the existence of multiple solutions can also be easier to diagnose.

From the standpoint of applications, Groebner basis methods are attractive because they are algorithmic, and widely implemented in computer software. Among free open-source offerings, `Singular` (Decker, Greuel, Pfister and Schönemann, 2011) and `Macaulay2` (Grayson and Stillman, 2011) have very good Groebner basis functions. `Sage` (Stein et al., 2010) uses `Singular`'s code. `Sage` has most of `Singular`'s functionality, a more convenient interface, and much broader capabilities in symbolic mathematics. Among commercial alternatives, `Maple` (Maplesoft, 2008) and `Mathematica` (Wolfram, 2003) are comprehensive packages that can calculate Groebner bases. The examples in this paper were carried out with `Sage` 4.3 and `Mathematica` versions 6 through 8.

Groebner basis was named by Bruno Buchberger after his thesis advisor, Wolfgang Gröbner. The anglicized spelling found in this paper is the one used by most software packages. Groebner basis techniques have been applied to structural equation modeling by García-Puente, Spielvogel and Sullivant (2010), and the authors are pleased to acknowledge a helpful conversation with Seth Sullivant (personal communication, 2007).

Groebner basis is a very active area of mathematical research, and has percolated down to the textbook level. Cox, Little and O'Shea's classic *Ideals, varieties and algorithms* (2007) is by far the most accessible text, and is a good next step for readers desiring a rigorous treatment of material presented in this paper. Familiarity with abstract algebra is helpful for reading Cox et al., but not essential. For *Mathematica* users, the free add-on package described in Appendix *C* of Cox et al. is highly recommended.

The plan of this paper is to introduce Groebner basis methods for an audience of psychometricians and applied statisticians, and to evaluate the methods as tools for structural modeling. The primary emphasis is upon model identification, but applications to model testing are also indicated. First, some notation and basic definitions are given, along with a collection of theorems that are useful for structural statistical modeling. In an axiomatic development of Groebner basis, some of these theorems would be more properly described as lemmas and corollaries, and many intermediate results (for example, the beautiful Hilbert Basis Theorem) are skipped because they are used only to prove other theorems.

A non-standard Groebner basis notation is introduced in this paper, with the goal of making the connection to statistical modeling more transparent. Also, explicitness is generally chosen over the compactness favored by algebraists. But the somewhat arcane vocabulary of the field is retained, to help the reader make the transition to more mathematically oriented material. Acquaintance with the standard vocabulary also makes Groebner basis software easier to use.

## Groebner Basis

In this section, it will be assumed that the quantities $\sigma_j$ in the moment structure equations (1) are fixed constants, while the parameters $\theta_1, \ldots, \theta_t$ are variables that determine a set of co-ordinate axes in high-dimensional space. Other arrangements are possible and sometimes quite useful; they will be discussed in due course.

### *Definitions*

*Definition 1.* A *monomial* is a product of the form $\theta_1^{\alpha_1}\theta_2^{\alpha_2}\cdots\theta_t^{\alpha_t}$, where the exponents $\alpha_1, \ldots \alpha_t$ are non-negative integers.

For statistical applications, the $\theta_j$ quantities are real-valued because they represent parameters, but a sizable portion of Groebner basis theory holds only if they are complex variables. That point will be noted when it is reached. For the present, $\theta_1, \ldots, \theta_t$ have no imaginary part.

Polynomials are weighted sums of monomials. To work effectively with polynomials, it is necessary to write their monomials in a consistent order. This is established by choice of a *monomial ordering*. First, one always lists the variables $\theta_1, \ldots, \theta_t$ in a particular order from left to right. The order of the variables can matter a great deal, with different orders sometimes providing different information about the same model. Variables that appear to be missing from a monomial are actually present, but raised to the power zero.

Monomial orderings are defined in terms of the vector of exponents $\boldsymbol{\alpha} = (\alpha_1, \ldots \alpha_t)'$. For solving systems of polynomial equations, the most useful is the *lexicographic* (lex) order.

*Definition 2.* Let

$$m_1 = \theta_1^{\alpha_{1,1}} \theta_2^{\alpha_{1,2}} \cdots \theta_t^{\alpha_{1,t}}, \text{ and}$$
$$m_2 = \theta_1^{\alpha_{2,1}} \theta_2^{\alpha_{2,2}} \cdots \theta_t^{\alpha_{2,t}}.$$

The monomial $m_1$ will be said to be greater than $m_2$ with respect to lexicographic order if and only if in the vector of differences $(\alpha_{1,1} - \alpha_{2,1}, \ldots, \alpha_{1,t} - \alpha_{2,t})'$, the leftmost non-zero entry is positive.

Other monomial orderings are sometimes useful. Define the *total degree* of a monomial as the sum of its exponents. *Graded lexicographic order* (grlex) first sorts monomials by total degree, and then breaks ties if necessary by lexicographic order. Computational efficiencies are often realized by *graded reverse lexicographic order* (grevlex), which first sorts monomials by total degree and then breaks ties if necessary by the length of the vector $\boldsymbol{\alpha}$ in $\mathbb{R}^t$.

*Definition 3.* A *polynomial* in $\theta_1, \ldots, \theta_t$ is a finite linear combination of monomials. Suppose there are $k$ monomials, with exponents $\alpha_{j,1}, \ldots \alpha_{j,t}$ for $j = 1, \ldots, k$. Then the polynomial is written

$$f(\boldsymbol{\theta}) = \sum_{j=1}^{k} a_j \theta_1^{\alpha_{j,1}} \theta_2^{\alpha_{j,2}} \cdots \theta_t^{\alpha_{j,t}}$$

The quantities being added are the *terms* of the polynomial. A term is a monomial multiplied by a *coefficient*. The ordering of the terms in a polynomial corresponds to the order of the monomials with respect to the chosen monomial ordering.

*Definition 4.* The *leading term* of a polynomial $f = f(\boldsymbol{\theta})$ is the one with the largest monomial. It is denoted $LT(f)$. The *leading monomial* is denoted $LM(f)$, and the *leading coefficient* is denoted $LC(f)$. In terms of Definition 3,

$$\begin{aligned} LC(f) &= a_1 \\ LM(f) &= \theta_1^{\alpha 1,1}\theta_2^{\alpha 1,2}\cdots\theta_t^{\alpha_{1,t}} \\ LT(f) &= a_1\theta_1^{\alpha 1,1}\theta_2^{\alpha 1,2}\cdots\theta_t^{\alpha_{1,t}} \end{aligned}$$

The coefficients $a_1, \ldots, a_n$ are members of a *field*. A field is a set of objects, equipped with operations corresponding to addition, subtraction, multiplication and division that satisfy the usual rules. The set of real numbers is a field, as is the set of complex numbers. For moment structure polynomials, the coefficients belong to a field that includes the moments $\sigma_j \in \mathbb{R}$, so the field of real numbers is a good choice for theoretical purposes. But for statistical applications, the field of rational numbers may offer computational advantages.

*Definition 5.* Let $\mathcal{F}$ be a field. Then the set of all polynomials in $\theta_1, \ldots, \theta_t$ with coefficients in $\mathcal{F}$ is denoted $\mathcal{F}[\theta_1, \ldots, \theta_t]$.

So, $\mathbb{R}[\theta_1, \ldots, \theta_t]$ is the set of all polynomials in $\theta_1, \ldots, \theta_t$ with real coefficients. Such a set does not itself form a field, because only constant polynomials can have a multiplicative inverse. They form a *ring*, specifically a commutative ring with unity. Groebner basis software, especially of the non-commercial variety, may require the user to choose a polynomial ring.

*Definition 6.* Let $\mathcal{F}$ be a field and $t$ be a positive integer. The $t$-dimensional *affine space* is defined as

$$\mathcal{F}^t = \{(x_1, \ldots, x_t) : x_j \in \mathcal{F} \text{ for } j = 1, \ldots, t\}$$

The main examples are $\mathbb{R}^t$ and $\mathbb{C}^t$.

*Ideal and Variety*

Given a collection of polynomials in the variables (parameters) $\theta_1, \ldots, \theta_t$, the *roots* are the $\theta$ values for which all the polynomials equal zero. If this set consists of just one point, all the parameters are identifiable. The set of roots is called the *variety*.

*Definition 7.* The (affine) *variety* of a set of polynomials $f_1, \ldots, f_d \in \mathcal{F}[\theta_1, \ldots, \theta_t]$ is the set of points $\boldsymbol{\theta} \in \mathcal{F}^t$ where $f_j(\boldsymbol{\theta}) = 0$ for $j = 1, \ldots, t$.

Given a finite set of polynomials, it is clear that a much larger set of polynomials share the same variety. In fact, if $f_j(\boldsymbol{\theta}) = 0$, then the product of $f_j$ and any other polynomial will also equal zero. This leads to an idea that is analogous to the concept of a vector space spanned by a set of basis vectors, except that instead of being multiplied by constants, the basis polynomials are multiplied by other polynomials.

*Definition 8.* Let $f_1, \ldots, f_d \in \mathcal{F}[\theta_1, \ldots, \theta_t]$. The *ideal generated by* the polynomials $f_1, \ldots, f_d$ is defined by

$$\langle f_1, \ldots, f_d \rangle = \left\{ \sum_{i=1}^{d} h_i(\boldsymbol{\theta}) f_i(\boldsymbol{\theta}) : h_1, \ldots, h_d \in \mathcal{F}[\theta_1, \ldots, \theta_t] \right\}$$

The set $\langle f_1, \ldots, f_d \rangle$ is closed under addition and multiplication, which is the definition of an "ideal in a ring" from abstract algebra. The ideal generated by a collection of polynomials represents all the *polynomial consequences* of setting the polynomials in the generating set to zero. Some of these consequences may be simpler than any member of the generating set, because multiplying polynomials and adding products can result in cancellations. For example, consider the polynomials

$$\begin{aligned} f_1 &= \theta_1^3\theta_2^2 + \theta_1^2\theta_2^3 - 2\,\theta_1^2\theta_2 - 2\,\theta_1\theta_2^2 + \theta_1 + \theta_2 \\ f_2 &= -\theta_1^2\theta_2 - \theta_1\theta_2^2 + 2\,\theta_1 + \theta_2 - 1. \end{aligned} \tag{2}$$

Using the well-chosen "weights" $h_1 = \theta_1 + \theta_2$ and $h_2 = \theta_1^2\theta_2 + \theta_1\theta_2^2 - \theta_2 - 1$, the polynomial combination $g = h_1 f_1 + h_2 f_2 = \theta_1^2 - 2\,\theta_1 + 1 = (\theta_1 - 1)^2$. The variable $\theta_2$ is eliminated from $g$, and

substituting $\theta_1 = 1$ into $f_1$ and $f_2$ shows that the variety of $\{f_1, f_2\}$ consists of just two points: $(\theta_1 = 1, \theta_2 = 1)$ and $(\theta_1 = 1, \theta_2 = -1)$. Thus, the ideal generated by a set of polynomials may contain polynomials that are more helpful than any member of the generating set. In general, the challenge is to find a set of nice, simple polynomials $g_1, \ldots, g_s \in \langle f_1, \ldots, f_d \rangle$ that have exactly the same roots as $f_1, \ldots, f_d$.

The polynomials $f_1, \ldots, f_d$ form a *basis* for the ideal they generate. Just as an ordinary vector space has more than one possible basis, so does an ideal.

*Definition 9.* A set of polynomials $p_1, \ldots, p_s \in \mathcal{F}[\theta_1, \ldots, \theta_t]$ is said to be a *basis* of an ideal $I$ if $I = \langle p_1, \ldots, p_s \rangle$.

What makes it useful to seek another basis of the ideal generated by a set of polynomials is that if two different sets of polynomials are bases of the same ideal, then they have the same roots. That is, the solutions of the equations formed by setting all the polynomials to zero are the same for the two sets.

*Theorem 1.* Let $\mathbf{f} = \{f_1, \ldots, f_d\}$ and $\mathbf{g} = \{g_1, \ldots, g_s\}$ be sets of polynomials in $\mathcal{F}[\theta_1, \ldots, \theta_t]$. If $\langle f_1, \ldots, f_d \rangle = \langle g_1, \ldots, g_s \rangle$, then the varieties of $\mathbf{f}$ and $\mathbf{g}$ are the same.

### *Groebner basis*

For the two polynomials (2), the useful polynomial combination $g = \theta_1^2 - 2\theta_1 + 1 \in \langle f_1, f_2 \rangle$ is one of the polynomials of a *Groebner basis* for $\langle f_1, f_2 \rangle$.

*Definition 10.* Given an ideal in $I \subset \mathcal{F}[\theta_1, \ldots, \theta_t]$, a Groebner basis for $I$ is a finite set of polynomials $G = \{g_1, \ldots, g_s\} \subset I$ such that for each polynomial $f \in I$, $LT(f) = h\, LT(g_i)$, for some $i \in \{1, \ldots, s\}$, where $h$ is a polynomial in $\mathcal{F}[\theta_1, \ldots, \theta_t]$.

Since ordinary long division may be applied to polynomials, one can say that the leading term of each polynomial in the ideal is divisible by the leading term of some polynomial in the Groebner basis. Note that which term of a polynomial is the leading term depends upon the monomial

ordering, so one speaks a Groebner basis with respect to a particular monomial ordering. The next theorem says that every non-zero ideal possesses a Groebner basis.

*Theorem 2.* For any non-zero ideal $I \subset \mathcal{F}[\theta_1, \ldots, \theta_t]$, there is a Groebner basis $\{g_1, \ldots, g_s\}$ with $g_j \in I$ and $\langle g_1, \ldots, g_s \rangle = I$.

*Long division*

The existence of Groebner bases and some of their properties can be proved without ever seeing one, but a systematic way of finding them depends upon long division of polynomials. Long division of a polynomial by one other polynomial works just like ordinary long division, yielding a unique quotient and remainder. There is also a standard algorithm (described by Cox et al. among others, and implemented in many software packages) that divides a single polynomial $f$ by a set $F = \{f_1, \ldots, f_d\}$, allowing $f$ to be written

$$f(\boldsymbol{\theta}) = q_1(\boldsymbol{\theta}) f_1(\boldsymbol{\theta}) + \cdots + q_d(\boldsymbol{\theta}) f_d(\boldsymbol{\theta}) + r(\boldsymbol{\theta}), \tag{3}$$

where the polynomial $q_j$ is the *quotient* corresponding to $f_j$ for $j = 1, \ldots, d$, and the polynomial $r$ is the *remainder*. The quotients and remainder are all elements of $\mathcal{F}[\theta_1, \ldots, \theta_t]$.

Unfortunately, this expression is not unique, and depends upon the ordering of $f_1, \ldots, f_d$. Worse, it is easy to divide $f \in \langle f_1, \ldots, f_d \rangle$ by $f_1, \ldots, f_d$ and still obtain a remainder that is not zero. However, the long division algorithm is more satisfactory when one is dividing by the polynomials of a Groebner basis.

*Theorem 3.* Let $G = \{g_1, \ldots, g_s\}$ be a Groebner basis for the ideal $I \subset \mathcal{F}[\theta_1, \ldots, \theta_t]$, and let the polynomial $f \in \mathcal{F}[\theta_1, \ldots, \theta_t]$. Then the remainder upon division of $f$ by $G$ is unique, and does not depend upon the ordering of $g_1, \ldots, g_s$.

A Groebner basis for the ideal generated by a set of polynomials is obtained by iteratively producing polynomials with simpler leading terms. This is accomplished using *S-Polynomials* (S stands for subtraction), which can eliminate variables when the monomial ordering is lexicographic, producing members of the ideal whose roots are potentially easier to find.

*Definition 11.* Let $m_1 = \theta_1^{\alpha_1} \cdots \theta_t^{\alpha_t}$ and $m_2 = \theta_1^{\beta_1} \cdots \theta_t^{\beta_t}$ be monomials. The *least common multiple* of $m_1$ and $m_2$ is defined by $LCM(m_1, m_2) = \theta_1^{\gamma_1} \cdots \theta_t^{\gamma_t}$, where $\gamma_i = \max(\alpha_i, \beta_i)$ for $i = 1, \ldots, t$.

*Definition 12.* Let the polynomials $p, q \in \mathcal{F}[\theta_1, \ldots, \theta_t]$. The *S-Polynomial* is a combination of $p$ and $q$ defined by

$$S(p, q) = \frac{\ell}{LT(p)} \cdot p - \frac{\ell}{LT(q)} \cdot q,$$

where $\ell = LCM\left(LM(p), LM(q)\right)$.

Again, $LM(f)$ is the leading monomial of $f$.

To see how this works, adopt the lexicographic monomial ordering on the polynomials of (2), with $\theta_1$ coming before $\theta_2$. The least common multiple of $LM(f_1)$ and $LM(f_2)$ is just the leading monomial of $f_1$, and the $S$-polynomial is

$$S(f_1, f_2) = \frac{\theta_1^3\theta_2^2}{\theta_1^3\theta_2^2} \cdot f_1 - \frac{\theta_1^3\theta_2^2}{-\theta_1^2\theta_2} \cdot f_2 = f_1 + \theta_1\theta_2 f_2 = -\theta_1\theta_2^2 - \theta_1\theta_2 + \theta_1 + \theta_2.$$

The $S$-polynomial is a member of $\langle f_1, f_2 \rangle$ that is simpler than either of the generating polynomials, and is a step in the right direction.

Even though $S(f_1, f_2)$ is clearly a combination of the form (3) with quotients $q_1 = 1$, $q_2 = \theta_1\theta_2$ and remainder $r = 0$, the long division algorithm cannot detect it, because it divides $S(f_1, f_2)$ by $f_1$ and $f_2$ one at a time. In either order, the remainder is just $S(f_1, f_2)$. This cannot happen when the input polynomials are part of a Groebner basis.

*Theorem 4. Buchberger's Criterion*: Let $I$ be a non-zero ideal of polynomials in $\mathcal{F}[\theta_1, \ldots, \theta_t]$. Then a set of polynomials $G = \{g_1, \ldots, g_s\}$ is a Groebner basis for $I$ if and only if for all pairs $i \neq j$, the remainder upon division of $S(g_i, g_j)$ by $G$ equals zero.

Zero division upon remainder by a Groebner basis is a characteristic that the $S$-polynomials share with all polynomials in the ideal.

*Theorem 5.* Let $I \subset \mathcal{F}[\theta_1, \ldots, \theta_t]$ be an ideal, let $G = \{g_1, \ldots, g_s\}$ be a Groebner basis for $I$, and let $f \in \mathcal{F}[\theta_1, \ldots, \theta_t]$. Then $f \in I$ if and only if the remainder upon division of $f$ by $G$ equals zero.

Theorem 5 provides a general solution to the *ideal membership problem*. Given a set of polynomials $f_1, \ldots, f_d$ and another polynomial $f$, how can one tell whether $f$ is in the ideal generated by $f_1, \ldots, f_d$? That is, if $f_1(\boldsymbol{\theta}) = \cdots = f_d(\boldsymbol{\theta}) = 0$, does it follow that $f(\boldsymbol{\theta}) = 0$? One divides $f$ by a Groebner basis for $\langle f_1, \ldots, f_d \rangle$; the answer is yes if and only if the remainder is zero.

The missing ingredient is a systematic way of finding Groebner basis for the ideal generated by a set of polynomials. The key is an iterative use of Buchberger's Criterion. One computes $S$-polynomials for all pairs of polynomials in the input set, dividing each $S$-polynomial by the entire input set. If all remainders are zero, the set of polynomials is a Groebner basis by Theorem 4, and the process terminates. Any remainder that is not zero is added to the input set, and the process repeats. The following theorem says that this procedure converges in a finite number of steps, yielding a set of polynomials that form a Groebner basis for $\langle f_1, \ldots, f_d \rangle$.

*Theorem 6. Buchberger's Algorithm* Let $I = \langle f_1, \ldots, f_d \rangle$, where $f_1, \ldots, f_d \in \mathcal{F}[\theta_1, \ldots, \theta_t]$. A Groebner basis $G = \{g_1, \ldots, g_s\}$ can be constructed in a finite number of steps by the algorithm of Figure 1.

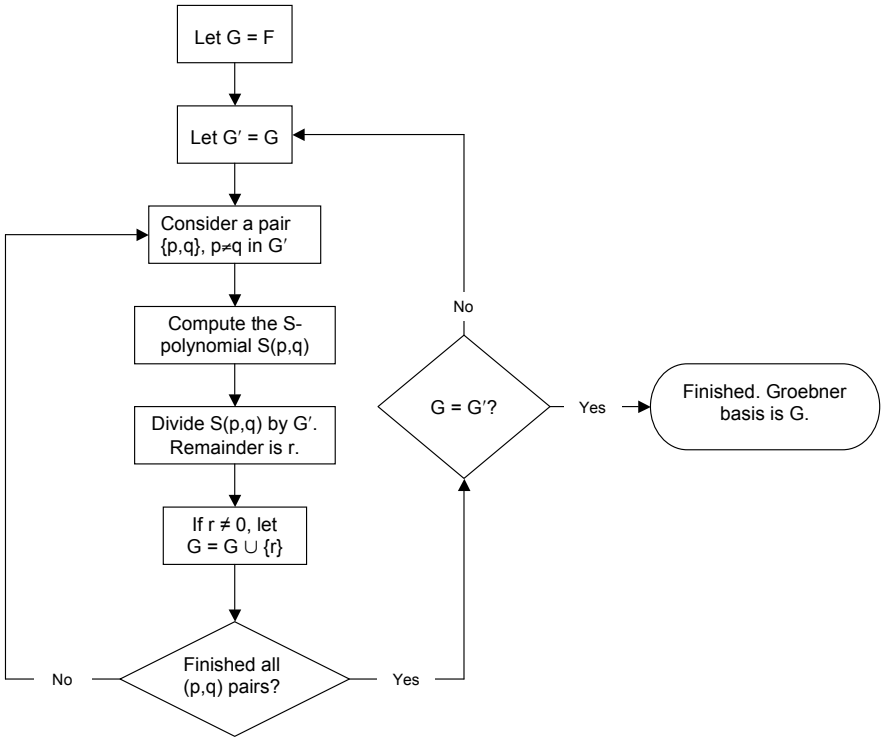
*Reduction*

Figure 1 portrays Buchberger's original algorithm, which was designed more for proving convergence than for computational efficiency. It can be and has been improved in various ways, for example by not re-computing zero remainders; also see Tran (2000) and the references therein. But all existing algorithms for computing Groebner bases work by adding polynomials to the generating set, usually introducing some redundancy. In general Groebner bases are not unique, and it is often possible to eliminate some of the polynomials and still have a Groebner basis for the ideal in question.

FIGURE 1.

Buchberger's Algorithm: Input is a generating set of polynomials $F = \{f_1, \ldots, f_d\}$. Output is a Groebner basis $G = \{g_1, \ldots, g_s\}$

.




*Theorem 7.* Let $G = \{g_1, \ldots, g_s\}$ be a Groebner basis for the ideal $I \in \mathcal{F}[\theta_1, \ldots, \theta_t]$. If $g \in G$ is a polynomial whose leading monomial is a multiple of the leading monomial of some other polynomial in $G$, then $G \cap \{g\}^c$ is also a Groebner basis for $I$.

So by simply examining the leading terms, one can often locate redundant polynomials in a Groebner basis and discard them. Usually, the polynomials that are discarded are earlier in the list; the result is often that some or all of the original polynomials $f_1, \ldots, f_d$ disappear, and are replaced by simpler ones.

Reducing a Groebner basis happens in two steps. First, redundant members of the basis are eliminated, and then the remaining ones are simplified one more time. The first step produces a *minimal Groebner basis*, and the second step produces the *reduced Groebner basis*, which is unique. These names are standard but unfortunate, because one would expect a "minimal" quantity to simpler and more compact than a "reduced" one. But given a monomial order and an ordering of variables, there are infinitely many minimal Groebner bases for a given ideal, each with the same number of polynomials. One of these minimal bases is "reduced," and is usually the most informative.

*Definition 13.* A *minimal Groebner basis* for a non-zero polynomial ideal is a Groebner basis $G$ for $I$ such that for every polynomial $g \in G$ (a) The leading coefficient of $g$ equals one, and (b) The leading term of $g$ is not a multiple of the leading term of any other polynomial in $G$.

*Definition 14.* A *reduced Groebner basis* for a non-zero polynomial ideal $I$ is a Groebner basis $G$ for $I$ such that for every polynomial $g \in G$, (a) The leading coefficient of $g$ equals one, and (b) No monomial of $g$ is a multiple of the leading term of any other polynomial in $G$.

*Theorem 8.* Let $G$ be a minimal Groebner basis for a non-zero polynomial ideal $I$. Replacing each polynomial in $G$ with its remainder upon division by the other polynomials in $G$ yields a reduced Groebner basis for $I$.

*Theorem 9.* A reduced Groebner basis is unique up to a monomial ordering and an ordering of variables.

For the polynomials (2), the slightly improved Buchberger algorithm described by Cox et al. in Section 2 of Chapter 9 yields the following Groebner basis with respect to lexicographic order.

$$
\begin{aligned}
\widetilde{g}_1 &= \theta_1^3\theta_2^2 + \theta_1^2\theta_2^3 - 2\theta_1^2\theta_2 - 2\theta_1\theta_2^2 + \theta_1 + \theta_2 \\
\widetilde{g}_2 &= -\theta_1^2\theta_2 - \theta_1\theta_2^2 + 2\theta_1 + \theta_2 - 1 \\
\widetilde{g}_3 &= -\theta_1\theta_2^2 - \theta_1\theta_2 + \theta_1 + \theta_2 \\
\widetilde{g}_4 &= \theta_1^3 - 2\theta_1^2 + \theta_1\theta_2 - \theta_2 + 1 \\
\widetilde{g}_5 &= \theta_1^2 - 2\theta_1 + 1 \\
\widetilde{g}_6 &= -\theta_1\theta_2 + \theta_2^3 - 2\theta_2^2 + 2 \\
\widetilde{g}_7 &= -\theta_1 + 2\theta_2^3 - 3\theta_2^2 - 2\theta_2 + 4 \\
\widetilde{g}_8 &= -2\theta_2^3 + 2\theta_2^2 + 2\theta_2 - 2
\end{aligned}
\tag{4}
$$

First, notice that $\widetilde{g}_1 = f_1$ and $\widetilde{g}_2 = f_2$; the polynomials $\widetilde{g}_3, \ldots, \widetilde{g}_8$ have been added to the original set to form a Groebner basis. Next, observe that the leading monomials of $\widetilde{g}_1, \ldots, \widetilde{g}_6$ are all multiples of $LM(\widetilde{g}_7)$. So by Theorem 7, they may be discarded and the result is still a Groebner basis for $\langle f_1, f_2 \rangle$.

Dividing by leading coefficients yields a minimal Groebner basis.

$$
\begin{aligned}
-\widetilde{g}_7 &= \theta_1 - 2\theta_2^3 + 3\theta_2^2 + 2\theta_2 - 4 \\
-\frac{1}{2}\,\widetilde{g}_8 &= \theta_2^3 - \theta_2^2 - \theta_2 + 1
\end{aligned}
\tag{5}
$$

To convert this minimal Groebner basis to the reduced Groebner basis, each polynomial is replaced by its remainder upon division by the other one. Dividing $-\widetilde{g}_7$ by $-\frac{1}{2}\,\widetilde{g}_8$ yields remainder $\theta_1 + \theta_2^2 - 2$, while dividing $-\frac{1}{2}\,\widetilde{g}_8$ by $-\widetilde{g}_7$ just returns the remainder $-\frac{1}{2}\,\widetilde{g}_8$. Thus, by Theorem 8, the reduced Groebner basis is

$$
\begin{aligned}
g_1 &= \theta_1 + \theta_2^2 - 2 \\
g_2 &= \theta_2^3 - \theta_2^2 - \theta_2 + 1 = (\theta_2 + 1)\,(\theta_2 - 1)^2\,.
\end{aligned}
\tag{6}
$$

The variable $\theta_1$ is eliminated from $g_2$, and substituting the solutions for $g_2 = 0$ into $g_1$ shows that the original polynomial equations $f_1 = f_2 = 0$ have exactly two real solutions: $\theta_1 = 1, \theta_2 = 1$ and $\theta_1 = 1, \theta_2 = -1$. This was not at all obvious from (2).

In this example, there are two polynomials in the original generating set and two polynomials in the reduced Groebner basis, but that is a coincidence. There is no necessary relationship between the number of generating polynomials and the number of polynomials in the reduced Groebner basis.

*Elimination*

In the reduced Groebner basis (6), the second polynomial is a function of $\theta_2$ only, while the first is a function of both $\theta_1$ and $\theta_2$. This is no accident. With the lexicographic monomial ordering, the reduced Groebner basis is designed to eliminate variables one at a time in a manner similar to the way Gaussian row reduction is used to solve systems of linear equations. The *elimination ideal* formalizes the idea of eliminating variables. Let $f_1, \ldots, f_d$ be polynomials in $\theta_1, \ldots, \theta_t$. The $k$th elimination ideal is the set of polynomial consequences of $f_1 = \cdots = f_d = 0$ that do not involve $\theta_1, \ldots, \theta_k$, where $k < t$. That is, the first $k$ variables are eliminated.

*Definition 15.* Let the polynomial ideal $I = \langle f_1, \ldots, f_d \rangle \subset \mathcal{F}[\theta_1, \ldots, \theta_t]$. The $k$th elimination ideal is defined by $I_k = I \cap \mathcal{F}[\theta_{k+1}, \ldots, \theta_t]$.

Theorem 10 is called the *Elimination Theorem*. It says that with the lexicographic monomial ordering, the Groebner basis successively eliminates $\theta_1, \ldots, \theta_{t-1}$, provided such elimination is possible.

*Theorem 10.* Let $G$ be a Groebner basis for the non-zero polynomial ideal $I = \langle f_1, \ldots, f_d \rangle \subset \mathcal{F}[\theta_1, \ldots, \theta_t]$ with respect to the lexicographic monomial ordering, with the ordering of variables $\theta_1, \ldots, \theta_t$. For every $1 \leq k \leq t-1$, if the $k$th elimination ideal $I_k \neq \emptyset$, then the set $G_k = G \cap \mathcal{F}[\theta_{k+1}, \ldots, \theta_t] \neq \emptyset$, and $G_k$ is a Groebner basis for $I_k$.

Let $G = \{g_1, \ldots, g_s\}$ be the reduced Groebner basis with respect to lexicographic order for $\langle f_1, \ldots, f_d \rangle$, and list $G$ in order of leading monomials, with the polynomial having the "largest"

leading monomial coming first. If elimination of $\theta_1, \ldots, \theta_{t-1}$ from the system of equations is possible, then $g_s$ (the last Groebner basis polynomial) will be a function of $\theta_t$ only. Suppose that in addition, elimination of $\theta_1, \ldots, \theta_{t-2}$ is possible. Then $g_{s-1}$ will be a function of $\theta_{t-1}$ and possibly $\theta_t$, but not $\theta_1, \ldots, \theta_{t-2}$. The pattern continues, with only $g_1$ being potentially a function of all $t$ parameters, again supposing that elimination is possible at each step. The result is a system of polynomials in an upper triangular form similar to the row echelon form resulting from Gaussian elimination in linear systems. As in the linear case, the solutions may be obtained by a series of simple substitutions. By Theorem 1, these are also the solutions of the original system of equations. Incidentally, calculation of the reduced Groebner basis for a set of polynomials that are linear yields exactly the reduced row echelon form.

The ordering of variables $\theta_1, \ldots, \theta_t$ has a profound effect upon the form of a Groebner basis with respect to lexicographic order, because it determines which variables are eliminated. The results of varying parameter order will be illustrated in the examples. Of course by Theorem 1, the ordering of parameters ultimately has no effect upon the variety.

Sometimes, a variable cannot be eliminated, and two or more parameters appear for the first time (reading from the bottom) in the same equation, possibly indicating that the system has infinitely many solutions. This will be made precise in the next theorem.

*Finiteness*

Up to this point, all the definitions and theorems apply to an arbitrary field $\mathcal{F}$, which could be the set of real numbers. That is, the parameters $\theta_1, \ldots, \theta_t$ may be real valued, as may the coefficients in the set of polynomials $\mathcal{F}[\theta_1, \ldots, \theta_t]$. But a rich and substantial portion of Groebner basis theory applies only when the variables and coefficients are complex-valued, with potentially both a real and an imaginary part. Many results in algebra are cleaner and more general as they apply to complex numbers.

This account omits most parts of Groebner basis theory that require the parameters to be complex variables. However, one result is useful in practice even though the parameters in most statistical models are real-valued. Theorem 11 gives a necessary and sufficient condition for a system of polynomial equations to have finitely many *complex* solutions. Since finitely many

complex solutions implies finitely many real solutions, the theorem allows one to rule out infinitely many real solutions for some models.

*Theorem 11.* Let $V \subset \mathbb{C}^t$ be the variety of the nonzero polynomial ideal $I \subset \mathbb{C}[\theta_1, \ldots, \theta_t]$, and let $G = \{g_1, \ldots, g_s\}$ be a Groebner basis for $I$. Then $V$ is a finite set if and only if for each $j$, $j = 1, \ldots, t$ there is some $m_j \geq 0$ and some $g \in G$ such that $LM(g) = \theta^{m_j}$.

That is, if each parameter $\theta_j$ appears to some non-zero power by itself as the leading monomial of at least one Groebner basis polynomial, the system of polynomial equations has only finitely many complex solutions. The case $m_j = 0$ corresponds to a constant, non-zero polynomial. Setting this polynomial to zero implies that the system has no solutions – and zero is a finite number. This does not occur in structural modeling, because the system always has at least one solution.

If any variable fails to appear by itself in a leading monomial, the system has infinitely many complex solutions. There might be infinitely many real solutions, or there might be finitely many, or only one. Further analysis is required.

For the example of the polynomials (2), a glance at either the raw Groebner basis (4), the minimal basis (5) or the reduced basis (6) establishes that the system has finitely many complex solutions and therefore finitely many real solutions.

## Applications to structural modeling

Groebner basis methods clearly have the potential to reveal the identification status of models to which standard rules do not apply. Less expected is their ability to yield statistics that can be used to test model correctness, even for non-identifiable models.

*A single-factor model*

To illustrate the methods on a simple example, consider a confirmatory factor analysis model with one factor and four observed variables. The model may be written

$$\begin{aligned} X_1 &= \lambda_1 F + e_1 \\ X_2 &= \lambda_2 F + e_2 \\ X_3 &= \lambda_3 F + e_3 \\ X_4 &= \lambda_4 F + e_4, \end{aligned} \tag{7}$$

where all expected values equal zero, $Var(F) = \phi$, $Var(e_j) = \psi_j$ for $j = 1, \ldots, 4$, and $F$, $e_1$, $e_2$, $e_3$ and $e_4$ are mutually independent. With or without a normal assumption, in practice the parameters of this model will be identified from the covariance matrix of the manifest variables or not at all. The parameter vector (actually, a *function* of the parameter vector if the distributions of the exogenous variables are unknown) is $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \phi, \psi_1, \psi_2, \psi_3, \psi_4) \in \Theta$, and the covariance matrix is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \lambda_1^2\phi + \psi_1 & \lambda_1\lambda_2\phi & \lambda_1\lambda_3\phi & \lambda_1\lambda_4\phi \\ \lambda_1\lambda_2\phi & \lambda_2^2\phi + \psi_2 & \lambda_2\lambda_3\phi & \lambda_2\lambda_4\phi \\ \lambda_1\lambda_3\phi & \lambda_2\lambda_3\phi & \lambda_3^2\phi + \psi_3 & \lambda_3\lambda_4\phi \\ \lambda_1\lambda_4\phi & \lambda_2\lambda_4\phi & \lambda_3\lambda_4\phi & \lambda_4^2\phi + \psi_4 \end{bmatrix}. \tag{8}$$

The parameters of this model are not identifiable from the covariance matrix, for letting $\lambda_j' = a\lambda_j$ and $\phi' = \phi/a^2$ for any $a \neq 0$ yields the same $\boldsymbol{\Sigma}$ as (8). For this model, identification can be obtained in two standard ways — by letting one of the factor loadings equal one ("setting the scale" of $F$, in the language of Bollen 1989), or by letting $\phi = 1$ and choosing a sign for one of the loadings. With either re-parameterization, the model becomes over-identified, with two over-identifying restrictions:

$$\sigma_{1,2}\sigma_{3,4} = \sigma_{1,3}\sigma_{2,4} = \sigma_{1,4}\sigma_{2,3}, \tag{9}$$

where $\sigma_{i,j}$ refers to element $(i, j)$ of $\boldsymbol{\Sigma}$.

In the following, the model will be left in its original non-identifiable and arguably more plausible form. The polynomials corresponding to the moment (covariance) structure equations

are

$$\begin{aligned}
f_1 &= \lambda_1^2\phi + \psi_1 - \sigma_{1,1} & f_6 &= \lambda_2\lambda_3\phi - \sigma_{2,3} \\
f_2 &= \lambda_1\lambda_2\phi - \sigma_{1,2} & f_7 &= \lambda_2\lambda_4\phi - \sigma_{2,4} \\
f_3 &= \lambda_1\lambda_3\phi - \sigma_{1,3} & f_8 &= \lambda_3^2\phi + \psi_3 - \sigma_{3,3} \\
f_4 &= \lambda_1\lambda_4\phi - \sigma_{1,4} & f_9 &= \lambda_3\lambda_4\phi - \sigma_{3,4} \\
f_5 &= \lambda_2^2\phi + \psi_2 - \sigma_{2,2} & f_{10} &= \lambda_4^2\phi + \psi_4 - \sigma_{4,4}
\end{aligned} \tag{10}$$

Different algorithms will yield unreduced Groebner bases that look quite dissimilar. In this case Tran's (2000) Groebner walk algorithm as implemented in `Mathematica` is a fortunate choice. With the ordering of variables $\psi_4, \psi_3, \psi_2, \psi_1, \phi, \lambda_4, \lambda_3, \lambda_2, \lambda_1$, it yields a Groebner basis with respect to lexicographic order consisting of 24 polynomials.

$$\begin{aligned}
g_1 &= \psi_4 + \phi\lambda_4^2 - \sigma_{4,4} & g_{13} &= \sigma_{2,4}\phi\lambda_1^2 - \sigma_{14,}\sigma_{1,2} \\
g_2 &= \psi_3 + \phi\lambda_3^2 - \sigma_{3,3} & g_{14} &= \sigma_{2,3}\phi\lambda_1^2 - \sigma_{1,3}\sigma_{1,2} \\
g_3 &= \psi_2 + \phi\lambda_2^2 - \sigma_{2,2} & g_{15} &= \sigma_{2,3}\lambda_4 - \sigma_{3,4}\lambda_2 \\
g_4 &= \psi_1 + \phi\lambda_1^2 - \sigma_{1,1} & g_{16} &= \sigma_{1,3}\lambda_4 - \sigma_{3,4}\lambda_1 \\
g_5 &= \phi\lambda_4\lambda_3 - \sigma_{3,4} & g_{17} &= \sigma_{1,2}\lambda_4 - \sigma_{2,4}\lambda_1 \\
g_6 &= \phi\lambda_4\lambda_2 - \sigma_{2,4} & g_{18} &= \sigma_{2,4}\lambda_3 - \sigma_{3,4}\lambda_2 \\
g_7 &= \phi\lambda_4\lambda_1 - \sigma_{1,4} & g_{19} &= \sigma_{1,4}\lambda_3 - \sigma_{3,4}\lambda_1 \\
g_8 &= \phi\lambda_3\lambda_2 - \sigma_{2,3} & g_{20} &= \sigma_{1,2}\lambda_3 - \sigma_{2,3}\lambda_1 \\
g_9 &= \phi\lambda_3\lambda_1 - \sigma_{1,3} & g_{21} &= \sigma_{1,4}\lambda_2 - \sigma_{2,4}\lambda_1 \\
g_{10} &= \sigma_{3,4}\phi\lambda_2^2 - \sigma_{2,4}\sigma_{2,3} & g_{22} &= \sigma_{1,3}\lambda_2 - \sigma_{2,3}\lambda_1 \\
g_{11} &= \phi\lambda_2\lambda_1 - \sigma_{1,2} & g_{23} &= \sigma_{3,4}\sigma_{1,2} - \sigma_{2,3}\sigma_{1,4} \\
g_{12} &= \sigma_{3,4}\phi\lambda_1^2 - \sigma_{1,4} & g_{24} &= \sigma_{2,4}\sigma_{1,3} - \sigma_{2,3}\sigma_{1,4}
\end{aligned} \tag{11}$$

It is helpful to read the Groebner basis from the end, because the upper triangular arrangement imposed by the lexicographic monomial ordering means that the later polynomials contain fewer variables. Here, the last two polynomials involve only constants; they are free of the parameters in $\boldsymbol{\theta}$. Thus a by-product of the algorithm in this case is a pair of relations among constants that must be satisfied if the system of covariance structure equations is to have any solutions at all; this can happen at an intermediate stage in Gaussian elimination for linear systems, too.

Remarkably, setting the Groebner basis polynomials $g_{23}$ and $g_{24}$ to zero gives exactly the over-identifying restrictions (9) that hold when this model is re-parameterized in either of the two standard ways. Using the covariance matrix (8), it is easy to verify that these restrictions hold for the non-identifiable version of the model as well. Thus, even a non-identifiable model can impose testable equality constraints upon the moments, and these constraints may be revealed by a Groebner basis. This suggests a general method for testing the correctness of models whose parameters are not identifiable. Details are given in the Discussion section.

Polynomials $g_{15}$ through $g_{22}$ show the circumstances under which *ratios* of factor loadings are identifiable. For example, setting $g_{21} = g_{22} = 0$, it is possible to solve for $\frac{\lambda_2}{\lambda_1}$ at those points in the parameter space where $\lambda_1$ and at least two of $\lambda_2$, $\lambda_3$ and $\lambda_4$ are not zero.

The set of polynomials $g_{15}$ through $g_{22}$ suggest the device of "setting the scale" of the factor by letting one of the loadings equal unity (Bollen, 1989). For example, with $\lambda_1 = 1$, the equations corresponding to the Groebner basis polynomials are even easier to solve than (10). It is clear from (11) that this is equivalent to a re-parameterization in which the factor loadings are expressed in units of $\lambda_1$, $\phi$ is expressed in units of $\frac{1}{\lambda_1^2}$, and $\psi_1$ through $\psi_4$ are unchanged.

Groebner basis polynomials $g_{12}$ through $g_{14}$ reveal the circumstances under which another function of the parameters is identifiable: $\phi\lambda_1^2$. It is possible to solve for this quantity provided at least one of $\sigma_{2,3}$, $\sigma_{2,4}$ and $\sigma_{3,4}$ is non-zero, and hence that at least two of $\lambda_2$, $\lambda_3$ and $\lambda_4$ are non-zero.

Polynomials $g_{12}$ through $g_{14}$ suggest a second popular restriction commonly used to purchase identification for confirmatory factor analysis models, namely setting $\phi = 1$. Examination of $g_5$ through $g_{14}$ shows that if in addition the sign of one factor loading is known, this makes all the factor loadings identifiable provided that at least three of them are not zero – a well known three variable rule. It is also apparent from the Groebner basis that this is equivalent to a re-parameterization in which the factor loadings are expressed in units of the standard deviation of the underlying factor. All this is fairly obvious for a simple, familiar model like (7). What is noteworthy is how easy it is to see from an unreduced Groebner basis.

The picture is much clearer for $\lambda_1$ than for the other factor loadings even though by symmetry, similar conclusions apply to all four loadings. The reason is that $\lambda_1$ is listed last in the

ordering of variables, so by the Elimination Theorem (Theorem 10), it plays a "starring role" in the Groebner basis. With the lexicographic monomial ordering, conclusions appear most explicitly for the variable listed last, with results for the other variables often appearing in terms of the last variable.

Thus, it is often helpful to try more than one ordering of variables. In the present example, listing $\psi_1, \ldots, \psi_4$ last yields a Groebner basis with fifty (as opposed to 24) polynomials, twelve of which show exactly where in the parameter space these four parameters are identifiable. This is a bit more convenient than using $g_1, \ldots, g_4$ in (11), but at the same time conclusions about $\lambda_1, \ldots, \lambda_4$ become less obvious. It is important to reiterate that all the possible Groebner bases for a problem contain the same information in the sense that by Theorem 1, their polynomials share a common set of roots. But different orderings of variables will cause this information to be expressed differently, and can minimize the need for hand calculation.

So far, the unreduced Groebner bases for this factor analysis example have shown that $\psi_1, \ldots, \psi_4$ are identifiable almost everywhere in the parameter space, but have not yet revealed lack of identifiability for the other parameters. In the Groebner basis (11), the failure of $\lambda_1$ to appear by itself in the leading term of any polynomial is a clue; by Theorem 11, this establishes that the original set of polynomials (10) has infinitely many *complex* roots. But finitely many real roots (or even a single real root) is still a mathematical possibility.

A firm conclusion comes from using Theorem 7 to discard polynomials whose leading terms are multiples of other leading terms. To simplify the discussion, it will be assumed that all the covariances are non-zero, limiting what follows to points where all the factor loadings are non-zero; this applies to all but a set of volume zero in the parameter space. Working from the bottom of the Groebner basis (11), the leading terms of $g_6, g_{10}, g_{11}$ and $g_{21}$ are all multiples of $LT(g_{22})$. Continuing in this fashion leaves $g_1, g_2, g_3, g_4, g_{14}, g_{17}, g_{20}$ and $g_{22}$, as well as $g_{23}$ and $g_{24}$. Setting the last two polynomials to zero gives side conditions which must hold if the model is correct, while the remaining Groebner basis polynomials correspond to eight equations in nine unknowns. By the parameter count rule (see Appendix 5 of Fisher, 1966), these equations have infinitely many real solutions, except possibly on a set of volume zero in $\mathbb{R}^9$, and hence in the parameter space $\Theta$. So, the vector of parameters $(\phi, \lambda_1, \lambda_2, \lambda_3, \lambda_4)'$ is not identifiable, a conclusion

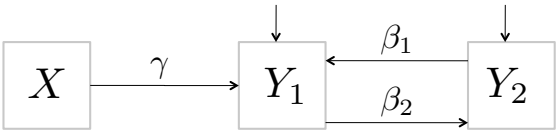
that holds almost everywhere in the parameter space.

*A cyclic model with observed variables*

Figure 2 shows a small observed variable model with a feedback loop.

FIGURE 2.

A non-recursive model




The model equations are

$$\begin{aligned} Y_1 &= \gamma X + \beta_1 Y_2 + \zeta_1 \\ Y_2 &= \beta_2 Y_1 + \zeta_2 \end{aligned} \tag{12}$$

where all expected values equal zero, $Var(X) = \phi$, $Var(\zeta_1) = \psi_1$, $Var(\zeta_2) = \psi_2$, and $X$, $\zeta_1$, and $\zeta_2$ are mutually independent. All the variances are greater than zero. In matrix form, (12) is

$$\begin{aligned} &\mathbf{Y} = \mathbf{\Gamma X} + \mathbf{B Y} + \boldsymbol{\zeta} \\ &\Leftrightarrow (\mathbf{I} - \mathbf{B})\mathbf{Y} = \mathbf{\Gamma X} + \boldsymbol{\zeta}. \end{aligned}$$

Rather than "assuming" $(\mathbf{I} - \mathbf{B})$ non-singular to calculate the covariance matrix of $\mathbf{Y}$, notice how $|\mathbf{I} - \mathbf{B}| = 1 - \beta_1\beta_2 \neq 0$ follows from the model assumptions. Substituting the first equation of (12) into the second yields $Y_2(1 - \beta_1\beta_2) = \gamma\beta_2 X + \beta_2\zeta_1 + \zeta_2$. If $\beta_1\beta_2 = 1$, then $\gamma^2\beta_2^2\phi + \beta_2^2\psi_1 + \psi_2 = 0$, implying $\psi_2 = 0$. The model stipulates non-zero variances, so it implies that the surface $\beta_1\beta_2 = 1$ is not part of the parameter space. Such holes in the parameter space are typical of cyclic models.

The covariance matrix for Model (12) is

$$\begin{bmatrix} \phi & -\frac{\gamma\phi}{(\beta_1\beta_2-1)} & -\frac{\beta_2\gamma\phi}{(\beta_1\beta_2-1)} \\ -\frac{\gamma\phi}{(\beta_1\beta_2-1)} & \frac{\left(\beta_1^2\psi_2+\gamma^2\phi+\psi_1\right)}{(\beta_1\beta_2-1)^2} & \frac{\left(\beta_2\gamma^2\phi+\beta_1\psi_2+\beta_2\psi_1\right)}{(\beta_1\beta_2-1)^2} \\ -\frac{\beta_2\gamma\phi}{(\beta_1\beta_2-1)} & \frac{\left(\beta_2\gamma^2\phi+\beta_1\psi_2+\beta_2\psi_1\right)}{(\beta_1\beta_2-1)^2} & \frac{\left(\beta_2^2\gamma^2\phi+\beta_2^2\psi_1+\psi_2\right)}{(\beta_1\beta_2-1)^2} \end{bmatrix}, \tag{13}$$

yielding a set of six moment structure polynomials.

$$\begin{aligned}
f_1 &= \phi - \sigma_{1,1} & f_4 &= -\beta_1^2\beta_2^2\sigma_{2,2} + \beta_1^2\psi_2 + 2\,\beta_1\beta_2\sigma_{2,2} + \gamma^2\phi + \psi_1 - \sigma_{2,2} \\
f_2 &= \beta_1\beta_2\sigma_{1,2} + \gamma\phi - \sigma_{1,2} & f_5 &= -\beta_1^2\beta_2^2\sigma_{2,3} + \beta_2\gamma^2\phi + 2\,\beta_1\beta_2\sigma_{2,3} + \beta_1\psi_2 + \beta_2\psi_1 - \sigma_{2,3} \\
f_3 &= \beta_1\beta_2\sigma_{1,3} + \beta_2\gamma\phi - \sigma_{1,3} & f_6 &= -\beta_1^2\beta_2^2\sigma_{3,3} + \beta_2^2\gamma^2\phi + 2\,\beta_1\beta_2\sigma_{3,3} + \beta_2^2\psi_1 + \psi_2 - \sigma_{3,3}
\end{aligned} \tag{14}$$

In this example, as in the general case (1), it is necessary to exercise some caution when multiplying through by denominators to obtain polynomials. Groebner basis methods do not "know" that the denominators cannot equal zero, so it is possible to introduce phantom solutions to th moment structure equations. Here, it will be easy to discard solutions that include $\beta_1\beta_2 = 1$. Also note that for acyclic linear structural equation models, the moment structure polynomials have constant denominators and this issue does not arise.

Using the ordering of variables $\psi_1, \psi_2, \phi, \gamma, \beta_1, \beta_2$, a Groebner basis with respect to lexicographic order has eight polynomials, with leading terms

$$\begin{array}{llll}
LT(\widetilde{g}_1) = \psi_1 & LT(\widetilde{g}_2) = \psi_2\beta_1\beta_2 & LT(\widetilde{g}_3) = \phi & LT(\widetilde{g}_4) = \sigma_{1,1}\gamma \\
LT(\widetilde{g}_5) = \sigma_{2,3}\beta_1^3\beta_2^3 & LT(\widetilde{g}_6) = \sigma_{1,3}\sigma_{2,3}\beta_1^3\beta_2^2 & LT(\widetilde{g}_7) = (\sigma_{12}\sigma_{13}\sigma_{33} - \sigma_{13}^2\sigma_{23})\beta_1^3\beta_2 & LT(\widetilde{g}_8) = \sigma_{12}\beta_1\beta_2^2.
\end{array}$$

The parameters $\beta_1$, $\beta_2$ and $\psi_2$ do not each appear alone in a leading monomial. By Theorem 11, it follows that the system has infinitely many complex solutions. However, it is often helpful to factor a Groebner basis, especially when expressions for the moments have non-constant denominators as in (13). This is easy to do with software, and the result is

$$\begin{aligned}
\widetilde{g}_1 &= \psi_1 + \psi_2\beta_1^2 - \sigma_{12}\gamma\beta_1\beta_2 + \sigma_{12}\gamma - \sigma_{22}\beta_1^2\beta_2^2 + 2\,\sigma_{22}\beta_1\beta_2 - \sigma_{22} \\
\widetilde{g}_2 &= (\beta_1\beta_2 - 1)\big(\beta_1\beta_2^2\sigma_{23} - \beta_1\beta_2\sigma_{33} - \beta_2\sigma_{23} - \psi_2 + \sigma_{33}\big) \\
\widetilde{g}_3 &= \phi - \sigma_{11} \\
\widetilde{g}_4 &= \sigma_{11}\gamma + \beta_1\beta_2\sigma_{12} - \sigma_{12} \\
\widetilde{g}_5 &= (\beta_1\beta_2 - 1)^2(\beta_1\beta_2\sigma_{23} - \beta_1\sigma_{33} - \beta_2\sigma_{22} + \sigma_{23}) \\
\widetilde{g}_6 &= (\beta_1\beta_2 - 1)\big(\beta_1^2\beta_2\sigma_{13}\sigma_{23} - \beta_1^2\sigma_{13}\sigma_{33} - \beta_1\beta_2\sigma_{13}\sigma_{22} + \beta_1\sigma_{12}\sigma_{33} - \sigma_{12}\sigma_{23} + \sigma_{13}\sigma_{22}\big) \\
\widetilde{g}_7 &= (\beta_1\beta_2 - 1)(\beta_1\sigma_{13} - \sigma_{12})(\beta_1\sigma_{12}\sigma_{33} - \beta_1\sigma_{13}\sigma_{23} - \sigma_{12}\sigma_{23} + \sigma_{13}\sigma_{22}) \\
\widetilde{g}_8 &= (\beta_1\beta_2 - 1)(\beta_2\sigma_{12} - \sigma_{13})
\end{aligned}$$

Since the surface $\beta_1\beta_2 = 1$ is not part of the parameter space, the polynomials $\widetilde{g}_2$ and $\widetilde{g}_5$ through $\widetilde{g}_8$ may be divided by the appropriate powers of $\beta_1\beta_2 - 1$ to eliminate phantom solutions. The result is a set of eight polynomials that generate an ideal containing $\langle f_1, \ldots, f_6\rangle$, and whose variety is therefore contained in that of $\langle f_1, \ldots, f_6\rangle$. The difference between the two varieties is an infinite set of points that lie outside the parameter space.

A Groebner basis for the new set of polynomials is

$$\begin{aligned}
g_1 &= \psi_1 - \sigma_{1,3}\gamma\beta_1 + \sigma_{1,2}\gamma - \sigma_{3,3}\beta_1^2 + 2\,\sigma_{2,3}\beta_1 - \sigma_{2,2} \\
g_2 &= \psi_2 - \sigma_{2,2}\beta_2^2 + 2\,\sigma_{2,3}\beta_2 - \sigma_{3,3} \\
g_3 &= \phi - \sigma_{1,1} \\
g_4 &= \sigma_{1,1}\gamma + \sigma_{1,3}\beta_1 - \sigma_{1,2} \\
g_5 &= (\sigma_{1,2}\sigma_{3,3} - \sigma_{1,3}\sigma_{2,3})\beta_1 - \sigma_{1,2}\sigma_{2,3} + \sigma_{1,3}\sigma_{2,2} \\
g_6 &= \sigma_{1,2}\beta_2 - \sigma_{1,3}.
\end{aligned} \tag{15}$$

Setting all the polynomials in (15) to zero and working from the bottom, it is easy to see the system of equations has a single solution provided that the leading coefficients are non-zero. This is only an issue for $g_5$ and $g_6$. Solving $g_6 = 0$ for $\beta_2$ requires $\sigma_{1,2} \neq 0$ and hence $\gamma \neq 0$, while solving $g_5 = 0$ for $\beta_1$ requires

$$\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23} = \frac{\gamma\phi\psi_2}{(\beta_1\beta_2 - 1)^2} \neq 0,$$

which again is true if and only if $\gamma \neq 0$. Thus, the parameter vector is identifiable at all points in the parameter space except where $\gamma = 0$. The final conclusion may be a bit surprising; assuming $\gamma \neq 0$, this cyclic model is *just identifiable*, and is indistinguishable from the natural acyclic (recursive) alternative on the basis of empirical data.

### *A categorical measurement model*

Suppose that a characteristic is either present or absent, and that three equivalent judges designate it as present or absent. Judgments are independent, conditionally upon the true presence or absence of the characteristic. Identifiability for variants of this model has been

considered by numerous authors including Anderson (1954), Margolin, Kim and Risko (1989) and Fujisawa and Izumi (2000).

Letting $F$ be a latent binary random variable and $X_1$, $X_2$ and $X_3$ be observable binary random variables, define the *prevalence* of the characteristic as $\theta_1 = Pr(F = 1)$, the *sensitivity* of the judges as $\theta_2 = Pr(X_j = 1|F = 1)$, and the *specificity* of the judges as $\theta_3 = Pr(X_j = 0|F = 0)$, for $j = 1, 2, 3$. The moments in this problem are probabilities in a $2 \times 2 \times 2$ contingency table, with $p_{ijk} = Pr(X_1 = i, X_2 = j, X_3 = k)$, for $i, j, k = 0, 1$. Elementary conditional probability calculations give the moment structure equations

$$\begin{aligned}
p_{000} &= \theta_3^3(1-\theta_1) + \theta_1(1-\theta_2)^3 \\
p_{001} &= \theta_3^2(1-\theta_1)(1-\theta_3) + \theta_1\theta_2(1-\theta_2)^2 \\
p_{010} &= \theta_3^2(1-\theta_1)(1-\theta_3) + \theta_1\theta_2(1-\theta_2)^2 \\
p_{011} &= \theta_3(1-\theta_1)(1-\theta_3)^2 + \theta_1\theta_2^2(1-\theta_2) \\
p_{100} &= \theta_3^2(1-\theta_1)(1-\theta_3) + \theta_1\theta_2(1-\theta_2)^2 \\
p_{101} &= \theta_3(1-\theta_1)(1-\theta_3)^2 + \theta_1\theta_2^2(1-\theta_2) \\
p_{110} &= \theta_3(1-\theta_1)(1-\theta_3)^2 + \theta_1\theta_2^2(1-\theta_2) \\
p_{111} &= (1-\theta_1)(1-\theta_3)^3 + \theta_1\theta_2^3.
\end{aligned} \tag{16}$$

With the ordering of variables $\theta_1, \theta_2, \theta_3$, Tran's (2000) Groebner walk algorithm yields a Groebner basis with 16 polynomials. The last 5 contain only constants, and give the relations among moments implied by the model.

$$\begin{aligned}
\widetilde{g}_{12} &= p_{000} + 3\,p_{100} + 3\,p_{110} + p_{111} - 1 \\
\widetilde{g}_{13} &= p_{001} - p_{100} \\
\widetilde{g}_{14} &= p_{010} - p_{100} \\
\widetilde{g}_{15} &= p_{011} - p_{110} \\
\widetilde{g}_{16} &= p_{101} - p_{110}
\end{aligned} \tag{17}$$

These say that measurements are equivalent, and the probabilities add to one.

The remaining 11 basis polynomials are very messy and uninformative, the longest having 102 terms; they will not be shown. Reducing the Groebner basis results in 15 polynomials rather than 16, and all but the last five are long and difficult to look at. The polynomials do not factor, and re-ordering the parameters makes no appreciable difference. This illustrates an unfortunate reality that must be acknowledged. While a Groebner basis can yield valuable insight into some problems, for others the result is simply unusable, usually because of the volume of output. Here, the number of polynomials in the Groebner basis is modest, but they are long. In the final example of this paper, the most natural way of expressing the problem results in a Groebner basis with so many polynomials that the computation never finishes.

*Skipping the sigmas*

For the categorical measurement model being considered, Groebner basis methods are successful when the problem is set up differently. Returning to the general notation of the Introduction, suppose that $\boldsymbol{\theta}$ is a point in the parameter space $\Theta$, and $\boldsymbol{\Sigma} = \sigma(\boldsymbol{\theta})$. Now let $\mathbf{x}$ be another point in $\Theta$ with $\sigma(\mathbf{x}) = \boldsymbol{\Sigma}$. If this implies $\mathbf{x} = \boldsymbol{\theta}$, then the parameter vector is identifiable. The method of *skipping the sigmas* is to bypass the vector of moments $\boldsymbol{\Sigma}$ altogether, writing

$$\sigma(\boldsymbol{\theta}) = \sigma(\mathbf{x}). \tag{18}$$

One solution will be $\mathbf{x} = \boldsymbol{\theta}$. The question is whether it is the only one.

For the categorical measurement model, the polynomials corresponding to (18) are

$$\begin{aligned}
f_1 &= \theta_3^3(1-\theta_1) + \theta_1(1-\theta_2)^3 - x_3^3(1-x_1) - x_1(1-x_2)^3 \\
f_2 &= \theta_3^2(1-\theta_1)(1-\theta_3) + \theta_1\theta_2(1-\theta_2)^2 - x_3^2(1-x_1)(1-x_3) - x_1x_2(1-x_2)^2 \\
f_3 &= \theta_3^2(1-\theta_1)(1-\theta_3) + \theta_1\theta_2(1-\theta_2)^2 - x_3^2(1-x_1)(1-x_3) - x_1x_2(1-x_2)^2 \\
f_4 &= \theta_3(1-\theta_1)(1-\theta_3)^2 + \theta_1\theta_2^2(1-\theta_2) - x_3(1-x_1)(1-x_3)^2 - x_1x_2^2(1-x_2) \\
f_5 &= \theta_3^2(1-\theta_1)(1-\theta_3) + \theta_1\theta_2(1-\theta_2)^2 - x_3^2(1-x_1)(1-x_3) - x_1x_2(1-x_2)^2 \\
f_6 &= \theta_3(1-\theta_1)(1-\theta_3)^2 + \theta_1\theta_2^2(1-\theta_2) - x_3(1-x_1)(1-x_3)^2 - x_1x_2^2(1-x_2) \\
f_7 &= \theta_3(1-\theta_1)(1-\theta_3)^2 + \theta_1\theta_2^2(1-\theta_2) - x_3(1-x_1)(1-x_3)^2 - x_1x_2^2(1-x_2) \\
f_8 &= (1-\theta_1)(1-\theta_3)^3 + \theta_1\theta_2^3. - (1-x_1)(1-x_3)^3 - x_1x_2^3.
\end{aligned} \tag{19}$$

Ordering of variables is critical, and it is helpful to treat the elements of both $\mathbf{x}$ and $\boldsymbol{\theta}$ as variables in order to alternate between them. This will yield the known solution $\mathbf{x} = \boldsymbol{\theta}$ in a clearer form. Here is the reduced and factored Groebner basis for (19) with respect to lexicographic order, using the ordering $\theta_1, x_1, \theta_2, x_2, \theta_3, x_3$. As usual, it should be read from the bottom.

$$
\begin{aligned}
g_1 &= \theta_1 x_1 x_2^2 + 2\theta_1 x_1 x_2 \theta_3 - 2\theta_1 x_1 x_2 + 2\theta_1 x_1 \theta_3 x_3 - 2\theta_1 x_1 \theta_3 - \theta_1 x_1 x_3^2 + \theta_1 x_1 + \theta_1 \theta_3^2 - 2\theta_1 \theta_3 x_3 \\
&\quad + \theta_1 x_3^2 - x_1^2 x_2^2 - 2x_1^2 x_2 x_3 + 2x_1^2 x_2 - x_1^2 x_3^2 + 2x_1^2 x_3 - x_1^2 - 2x_1 x_2 \theta_3 + 2x_1 x_2 x_3 - 2x_1 \theta_3 x_3 \\
&\quad + 2x_1 \theta_3 + 2x_1 x_3^2 - 2x_1 x_3 - \theta_3^2 + 2\theta_3 x_3 - x_3^2 \\
g_2 &= (x_1 - 1)(-\theta_3 + x_3)^2(\theta_1 x_2 + \theta_1 x_3 - \theta_1 + x_1 x_2 + x_1 x_3 - x_1 - \theta_2 - x_2 - 2x_3 + 2) \\
g_3 &= (x_1 - 1)(-\theta_3 + x_3)^3(-\theta_1 \theta_3 + \theta_1 x_3 + x_1 x_2 + x_1 x_3 - x_1 - \theta_2 + \theta_3 - 2x_3 + 1) \\
g_4 &= \theta_1 \theta_2 + \theta_1 \theta_3 - \theta_1 - x_1 x_2 - x_1 x_3 + x_1 - \theta_3 + x_3 \\
g_5 &= (\theta_1 - 1)(-\theta_3 + x_3)^2(x_2 + \theta_3 - 1)^2 \\
g_6 &= (x_1 - 1)(\theta_3 - x_3)^2(-\theta_2 - x_3 + 1)^2 \\
g_7 &= x_1 \theta_2 x_2 + x_1 \theta_2 x_3 - x_1 \theta_2 - x_1 x_2^2 - x_1 x_2 \theta_3 + x_1 x_2 - x_1 \theta_3 x_3 + x_1 \theta_3 + x_1 x_3^2 - x_1 x_3 \\
&\quad + \theta_2 \theta_3 - \theta_2 x_3 + \theta_3 x_3 - \theta_3 - x_3^2 + x_3 \\
g_8 &= (\theta_3 - x_3)(-x_2 - \theta_3 + 1)(-x_1 x_2 - x_1 x_3 + x_1 + \theta_2 + x_3 - 1) \\
g_9 &= (\theta_3 - x_3)(x_2 + \theta_3 - 1)(-\theta_2 - x_3 + 1)(-\theta_2 + x_2)
\end{aligned}
\tag{20}
$$

The roots of these polynomials, which are the same as the roots of (19), may be obtained by a set of simple substitutions and re-factoring. Fix a point $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in \Theta$, and first consider the case $\theta_3 = 1 - \theta_2$. If $x_2 = \theta_2$ and $x_3 = \theta_3$, all the polynomials become zero, and $x_1$ can be any number between zero and one. So for any point $\boldsymbol{\theta}$ in the parameter space with $\theta_2 + \theta_3 = 1$, there are infinitely many other points that yield the same probability distribution for the observed data. In fact, when $\theta_3 = 1 - \theta_2$ in the moment structure equations (16), they are completely free of $\theta_1$, and the data contain no information about the prevalence of the characteristic being assessed. This is natural, since $\theta_2 + \theta_3 = 1$ corresponds to independence of the latent and observed variables.

In the case where $\theta_2 + \theta_3 \neq 1$, the polynomials in (20) have two roots: $x_1 = \theta_1, x_2 = \theta_2, x_3 = \theta_3$ and $x_1 = 1 - \theta_1, x_2 = 1 - \theta_3, x_3 = 1 - \theta_2$. So for every point in the

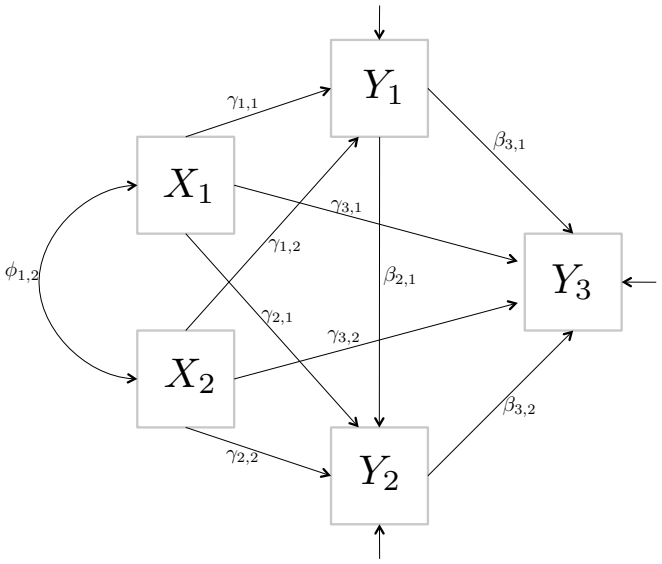
parameter space where $\theta_2 + \theta_3 \neq 1$, there is a point on the other side of the plane $\theta_2 + \theta_3 = 1$ that yields the same set of probabilities for the observed data. Consequently, the likelihood function will have exactly two maxima, one on either side of the plane. In practice, numerical search can usually be limited to the set where $\theta_2 + \theta_3 > 1$, because as Zelen and Haitovsky (1991) observe, $\theta_2 + \theta_3 < 1$ is equivalent to a negative association between manifest and latent variables.

*A recursive path model with observed variables*

Figure 3 shows the path diagram of a saturated acyclic linear structural equation model with two exogenous variables, three endogenous variables and independent errors. The parameters of this model are well known to be just identifiable (for example by Bollen's 1989 Recursive Rule), and it is smaller than most models for real data. But the model of Figure 3 presents a serious computational challenge for Groebner basis methods. Using standard LISREL notation (for

FIGURE 3.

A recursive model



example Jöreskog, 1978), the moment structure polynomials may be written as

$$
\begin{aligned}
f_1 &= \phi_{1,1} - \sigma_{1,1} \\
&\vdots \\
f_{15} &= \beta_{2,1}^2\beta_{3,2}^2\gamma_{1,1}^2\phi_{1,1} + 2\,\beta_{2,1}^2\beta_{3,2}^2\gamma_{1,1}\gamma_{1,2}\phi_{1,2} + \beta_{2,1}^2\beta_{3,2}^2\gamma_{1,2}^2\phi_{2,2} + 2\,\beta_{2,1}\beta_{3,1}\beta_{3,2}\gamma_{1,1}^2\phi_{1,1} \\
&\quad +4\,\beta_{2,1}\beta_{3,1}\beta_{3,2}\gamma_{1,1}\gamma_{1,2}\phi_{1,2} + 2\,\beta_{2,1}\beta_{3,1}\beta_{3,2}\gamma_{1,2}^2\phi_{2,2} + 2\,\beta_{2,1}\beta_{3,2}^2\gamma_{1,1}\gamma_{2,1}\phi_{1,1} + 2\,\beta_{2,1}\beta_{3,2}^2\gamma_{1,1}\gamma_{2,2}\phi_{1,2} \\
&\quad +2\,\beta_{2,1}\beta_{3,2}^2\gamma_{1,2}\gamma_{2,1}\phi_{1,2} + 2\,\beta_{2,1}\beta_{3,2}^2\gamma_{1,2}\gamma_{2,2}\phi_{2,2} + \beta_{2,1}^2\beta_{3,2}^2\psi_1 + 2\,\beta_{2,1}\beta_{3,2}\gamma_{1,1}\gamma_{3,1}\phi_{1,1} \\
&\quad +2\,\beta_{2,1}\beta_{3,2}\gamma_{1,1}\gamma_{3,2}\phi_{1,2} + 2\,\beta_{2,1}\beta_{3,2}\gamma_{1,2}\gamma_{3,1}\phi_{1,2} + 2\,\beta_{2,1}\beta_{3,2}\gamma_{1,2}\gamma_{3,2}\phi_{2,2} + \beta_{3,1}^2\gamma_{1,1}^2\phi_{1,1} \\
&\quad +2\,\beta_{3,1}^2\gamma_{1,1}\gamma_{1,2}\phi_{1,2} + \beta_{3,1}^2\gamma_{1,2}^2\phi_{2,2} + 2\,\beta_{3,1}\beta_{3,2}\gamma_{1,1}\gamma_{2,1}\phi_{1,1} + 2\,\beta_{3,1}\beta_{3,2}\gamma_{1,1}\gamma_{2,2}\phi_{1,2} + 2\,\beta_{3,1}\beta_{3,2}\gamma_{1,2}\gamma_{2,1}\phi_{1,2} \\
&\quad +2\,\beta_{3,1}\beta_{3,2}\gamma_{1,2}\gamma_{2,2}\phi_{2,2} + \beta_{3,2}^2\gamma_{2,1}^2\phi_{1,1} + 2\,\beta_{3,2}^2\gamma_{2,1}\gamma_{2,2}\phi_{1,2} + 2\,\beta_{3,1}\beta_{3,2}\gamma_{1,2}\gamma_{2,2}\phi_{2,2} + \beta_{3,2}^2\gamma_{2,1}^2\phi_{1,1} \\
&\quad +2\,\beta_{3,2}^2\gamma_{2,1}\gamma_{2,2}\phi_{1,2} + \beta_{3,2}^2\gamma_{2,2}^2\phi_{2,2} + 2\,\beta_{2,1}\beta_{3,1}\beta_{3,2}\psi_1 + 2\,\beta_{3,1}\gamma_{1,1}\gamma_{3,1}\phi_{1,1} + 2\,\beta_{3,1}\gamma_{1,1}\gamma_{3,2}\phi_{1,2} \\
&\quad +2\,\beta_{3,1}\gamma_{1,2}\gamma_{3,1}\phi_{1,2} + 2\,\beta_{3,1}\gamma_{1,2}\gamma_{3,2}\phi_{2,2} + 2\,\beta_{3,2}\gamma_{2,1}\gamma_{3,1}\phi_{1,1} + 2\,\beta_{3,2}\gamma_{2,1}\gamma_{3,2}\phi_{1,2} + 2\,\beta_{3,2}\gamma_{2,2}\gamma_{3,1}\phi_{1,2} \\
&\quad +2\,\beta_{3,2}\gamma_{2,2}\gamma_{3,2}\phi_{2,2} + \beta_{3,1}^2\psi_1 + \beta_{3,2}^2\psi_2 + \gamma_{3,1}^2\phi_{1,1} + 2\,\gamma_{3,1}\gamma_{3,2}\phi_{1,2} + \gamma_{3,2}^2\phi_{2,2} + \psi_3 - \sigma_{5,5},
\end{aligned}
$$

where the variances and covariances $\sigma_{i,j}$ covariances are treated as constants.

When Groebner basis calculations are successful, they typically finish in a moment or two. Here, using `Sage` version 4.3 and `Mathematica` versions 6.0, 7.0 and 8.0 on a variety of platforms and operating systems, the calculation was stopped in each case after 24 hours. The algorithms used were sophisticated versions, employing a variety of tricks to reduce the amount of computation. A general idea of what happened can be obtained by tracing the behavior of a slightly enhanced version of the Buchberger algorithm depicted in Figure 1; a convenient choice is the `Buch` command in Gryc and Krauss' `Mathematica` notebook for `Mathematica` 6 (`http://www.cs.amherst.edu/~dac/iva.html`). The only modification of the original Buchberger algorithm is that S-polynomials which have already been computed are not computed again, and when a remainder is zero, the division of that S-polynomial by subsequent sets of polynomials is skipped.

The first time through the main loop of Figure 1, $\binom{15}{2} = 105$ S-polynomials are calculated, and each is divided by the original set of 15 moment structure polynomials. Three remainders are zero, so that 102 remainders are added to the set of polynomials. The next time through the loop,

there are $\binom{117}{2} = 6,786$ S-polynomials, of which 105 may be skipped. So $6,786 - 105 = 6,681$ divisions are performed. Only 200 remainders are zero, and the other 6,481 are added to the input set of polynomials. There are now $15 + 102 + 6,481 = 6,598$ polynomials, the longest with 441 terms. The third time through the loop, there are $\binom{6,598}{2} = 21,763,503$ S-polynomials. Skipping the ones that have already been computed, there are $21,763,503 - 6,786 = 21,756,717$ remaining, each of which must be divided by the set of 6,598 polynomials. More than 24 hours of computation are required, and in practice step 3 never finishes.

This example illustrates how rapidly the number of polynomials in an initial Groebner basis can explode. The process is mathematically guaranteed to terminate eventually and reduction along the way sometimes helps, but given current hardware and software, Groebner basis methods sometimes just fail. While the initial number of polynomials is important, an even more critical factor is the proportion of S-polynomials that have non-zero remainders. This, in turn, is determined by the detailed structure of the polynomials and the way the problem is set up. For example, when the method of "skipping the sigmas" (see Expression 18) is applied to the model of Figure 3, computation is very fast and the Groebner basis consists of just 62 polynomials. Once these are factored to locate and discard roots on the boundary of the parameter space, identification follows immediately.

## Discussion

For many structural statistical models, identifiability is determined by whether a system of multivariate polynomial equations has more than one solution – or equivalently, whether a set of multivariate polynomials has more than one simultaneous root. A Groebner basis for the ideal generated by such a set of polynomials is another set of polynomials with the same set of simultaneous roots, and those roots are often much easier to find starting with the Groebner basis. For many models, a Groebner basis gives a clear picture of how identifiability changes in different regions of the parameter space, and reveals functions of the parameters that are identifiable even when the entire model is not.

Groebner basis theory reduces the process of simplifying multivariate polynomials to a massive clerical task of the sort that is best handled by computer. Many symbolic mathematics

programs have Groebner basis capability, and are very helpful for other modeling tasks such as calculating covariance matrices, asymptotic standard errors and the like. Using the software is no more difficult than using a statistics package, and familiarity with the material in this paper is sufficient to allow informed application of the methods.

Problems that are resistant to elementary mathematics need present no particular difficulty. For example, while the categorical measurement model discussed here is a familiar one that is known to be identifiable with the appropriate restriction, the actual proofs of identifiably are somewhat demanding (Anderson 1954; Teicher 1963). In contrast, a Groebner basis reveals the same information for this model with minimal effort. The experience suggests that Groebner basis methods may be helpful for studying global (rather than merely local) identifiability for less tractable cases such as constrained categorical models with polytomous observed variables.

*Testing fit of non-identifiable models*

The factor analysis example shows how even a non-identifiable model can imply constraints upon the moments, making it capable of being falsified by empirical data even though unique estimation of its parameters is impossible. In this case and in many others, the constraints are convenient by-products of Tran's (2000) Groebner walk algorithm. But it is desirable to have a method that is not tied to a particular algorithm, and to be certain that all the polynomial relations among moments are represented, and that none of them is redundant. These goals may be attained by considering the moments as variables rather than constants, and obtaining a Groebner basis with respect to the lexicographic monomial ordering, making sure that the moments appear after the parameters in the ordering of variables.

For a structural model with parameters $\theta_1, \ldots, \theta_t$ and moments $\sigma_1, \ldots, \sigma_d$, suppose the moment structure equations have the form $f_1 = \cdots = f_d = 0$, where

$$f_1, \ldots, f_d \in \mathbb{Q}[\theta_1, \ldots, \theta_t, \sigma_1, \ldots, \sigma_d].$$

The notation $\mathbb{Q}$ indicates that the coefficients of the polynomials belong to the field of rational numbers; in fact, they are usually integers.

The model-induced polynomial relations among $\sigma_1, \ldots, \sigma_d$ are exactly the consequences of

$f_1 = \cdots = f_d = 0$ that do not involve $\theta_1, \ldots, \theta_t$. That is, they form an elimination ideal. Let $G = \{g_1, \ldots, g_s\}$ be a Groebner basis for $\langle f_1, \ldots, f_d \rangle$ with respect to lexicographic order. By Theorem 10, the set of polynomials in $G$ that are free of $\theta_1, \ldots, \theta_t$ form a Groebner basis for the elimination ideal $\langle f_1, \ldots, f_d \rangle \cap \mathbb{Q}[\sigma_1, \ldots, \sigma_d]$, and represent the equality constraints of the model.

For compactness and interpretability, it is desirable to express the Groebner basis for the elimination ideal in reduced form. Carrying out the entire procedure for the factor analysis example yields exactly (9), so that in this case the Groebner walk algorithm produces the desired constraints in an optimal form even when the moments are treated as constants. The Groebner walk algorithm does not always accomplish this for more complicated models, so it is preferable to obtain the constraints by treating the moments as variables and calculating the reduced Groebner basis for an elimination ideal.

The resulting constraints on the moments are the same as the null hypothesis that is tested in a standard likelihood ratio test for goodness of model fit, provided that the maximum of the likelihood function is in the interior of the parameter space. But Groebner basis methods yield the constraints in an explicit form without the need to estimate model parameters. This means models that are not identifiable may be confronted by data in a convenient way, without the need to seek optimal constraints. When a non-identifiable model is consistent with the data, further inference may be carried out entirely in the moment space. It is not necessary to assume a normal distribution. For linear structural equation models, distribution-free versions of the tests may be obtained using the Central Limit Theorem, assuming only the existence of fourth moments.

*Drawbacks of Groebner basis methods*

Groebner basis is not a panacea. One disadvantage comes from the very generality of the theory. When using it to find the solutions of a set of moment structure equations, it is often difficult to limit the answer to the parameter space. In the cyclic example and the categorical example, it was possible to locate and discard solutions on the boundary of the parameter space by factoring the Groebner basis. But for some models, even restricting solutions to the set of real numbers can be a challenge.

Another disadvantage arises from the sheer volume of material that can be produced by

Groebner basis software. While the polynomials in a Groebner basis are often simpler than those the generating set, sometimes they are huge and ugly. This happened in the first try at the categorical example. And because existing algorithms form an initial Groebner basis by adding polynomials to the generating set, the number of polynomials can quickly become very large. As a mathematical certainty, the algorithms will arrive at a Groebner basis in a finite number of steps — but even with the fastest hardware currently available, there is no guarantee that this will happen during one human lifetime. In the recursive path model example, the number of polynomials exploded geometrically, and the computation never finished. This sort of difficulty is well documented, and a great deal of effort has gone into methods for increasing the efficiency of the computation. A special issue of the *Journal of symbolic computation* (Tran 2007) collects recent developments as well as extensive references to the earlier literature.

For the recursive path model example as well as the categorical example, the Groebner basis was unmanageable when the generating set of polynomials included symbols for the moments as well as the parameters, but the "skipping the sigmas" setup of Expression 18 produced useable results. This often happens, but skipping the sigmas discards information about model-induced constraints on the moments. And for some larger problems, the computation still can take an inordinately long time. For example, the industrialization and political democracy example discussed by Bollen (1989, p. 333) features a three-variable recursive latent model and a non-standard measurement model with several correlations between measurement errors. It has eleven manifest variables and sixteen parameters. Groebner basis calculation for the entire model never finishes even with skipping the sigmas, though it yields easily to a two-step approach in which the latent variable model and the measurement model are analyzed separately. It is also successful when the latent variable and measurement models are combined but the problem is broken into parts by setting aside the diagonal elements of the covariance matrix.

The lesson is clear, if a little sobering. With current computer technology, simply encoding a model with scores of parameters and throwing it at Groebner basis software is unlikely to succeed. For such a model, one must resort to the same devices that would be helpful for doing the problem by hand, such as breaking it into smaller pieces and reducing the number of variables by making substitutions. Groebner basis will be most helpful with smaller parts of the problem that

are unfamiliar or mathematically difficult.

One may confidently expect that as computer hardware continues to become faster and algorithms continue to improve, Groebner basis methods will be applicable to larger and larger problems. Even now, they can be a valuable tool for structural statistical modeling.

**References**

Anderson, T.W. (1954). On estimation of parameters in latent structure analysis. *Psychometrika* 19(1), 1-10.

Bentler, P.M. & Weeks, D.G. (1980) Linear Structural Equations with Latent Variables, *Psychometrika*, 45(3), 289-308.

Bézout, E. (2006). *General theory of algebraic equations* (Eric Feron, Trans.) Princeton, N.J.: University Press. (Original work published 1779)

Bollen, K. A. (1989), *Structural equations with latent variables,* New York: Wiley.

Buchberger, B. (1965) *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem Nulldimensionalen Polynomideal (An Algorithm for Finding the Basis Elements of the Residue Class Ring of a Zero Dimensional Polynomial Ideal)*, Ph.D. Thesis, University of Innsbruck. English translation by M. Abramson (2006) in *Journal of Symbolic Computation* 41, 471-511.

Cox, D., Little, J. and O'Shea, D. (2007). *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 3rd Edition, New York: Springer-Verlang, Inc.

Decker, W., Greuel, G.-M., Pfister, G. & Schönemann, H. (2011). SINGULAR 3-1-1 — A computer algebra system for polynomial computations. http://www.singular.uni-kl.de.

Fisher, F. M. (1966) *The Identification Problem in Econometrics*, New York: McGraw-Hill.

Fujisawa, H. & Izumi, S. (2000). Inference about Misclassification Probabilities from Repeated Binary Responses. *Biometrics*, 56(3), 706-711.

García-Puente, L. D., Spielvogel, S. and Sullivant, S. (2010). Identifying Causal Effects with Computer Algebra, *Proceedings of the 2010 Conference on Uncertainty in Artificial Intelligence*, Catalina Island, California, USA, July 8-11, 2010.

Grayson, D.R. & Stillman, M.E. (2011). MACCAULAY2 — A software system for research in algebraic geometry. http://www.math.uiuc.edu/Macaulay2/.

Jöreskog, K.G. (1978). Structural Analysis of Covariance and Correlation Matrices, *Psychometrika*, 43(4), 443-477.

Koopmans, T.C. & Reiersøl, O. (1950). The Identification of structural characteristics. *Annals of Mathematical Statistics*, 21(2), 165-181.

Margolin, B.H., Kim, B.S., & Risko, K.J. (1989). The Ames Salmonella/microsome mutagenicity assay: Issues of inference and validation. *Journal of the American Statistical Association* 84, 651-661.

Maplesoft, Inc. (2008). *Maple 12 user manual*. Waterloo, Maplesoft.

McArdle, J.J. & McDonald, R. P. (1984). Some Algebraic Properties of the Reticular Action Model, *British Journal of Mathematical and Statistical Psychology*, 37, 234-251.

McDonald, R.P. (1978). A Simple Comprehensive Model for the Analysis of Covariance Structures, *British Journal of Mathematical and Statistical Psychology*, 31, 59-72.

Sitwell, J. (2010) *Mathematics and its history*, 3d Edition. New York: Springer.

Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, Longitudinal, and Structural Equation Models.* New York: Chapman and Hall.

Stein, W.A. et al. (2010). *Sage Mathematics Software (Version 4.2.1)*, The Sage Development Team, http://www.sagemath.org.

Teicher, H. (1963). Identifiability of finite mixtures. *Annals of Mathematical Statistics* 34, 1265-1269.

Tran, Q.N. (2000). A fast algorithm for Gröbner basis conversion and its applications. *Journal of Symbolic Computation*, 30, 451-467.

Tran, Q. N. (2007). Special issue on applications of computer algebra. *Journal of symbolic computation*, 42(5), 495-496.

Wall, M.M. & Amemiya, Y. (2000). Estimation for polynomial structural equations, *Journal of the American Statistical Association*, 95(451), 929-940.

Wall, M.M. & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, 56, 4763.

Wolfram, S. (2003). *The Mathematica Book (5th Edition)*, Champaign, Illinois: Wolfram Media, Inc.

Zelen, M. and Haitovsky, Y. (1991). Testing hypotheses with binary data subject to misclassification errors: Analysis and experimental design. *Biometrika* 78, 857-865.