Higher Moment Varieties of Non-Gaussian Graphical Models

Carlos Améndola, Mathias Drton, Alexandros Grosdos,
Roser Homs, Elina Robeva

February 2020

Abstract

We study non-Gaussian graphical models from a perspective of algebraic statistics. Our focus is on algebraic relations among second and third moments in graphical models given by linear structural equations. We show that when the graph is a tree these relations form a toric ideal. Furthermore, from the covariance matrix and the third moment tensor, we construct explicit matrices whose 2×2 minors define the model set-theoretically. The matrices are associated to treks and multi-treks in the path, and computational experiments support our conjecture that their minors also generate the vanishing ideal of the model. We use the computer algebra software `Macaulay2` to compute concrete examples and describe the ideals for models with hidden variables.

1 Introduction

In this paper we study algebraic-geometric aspects of directed graphical models, which are a class of statistical models that hypothesize noisy functional (or in other words, cause-effect) relations among a set of random variables [3]. The models are graphical in the sense of using a graph to encode in an intuitive manner which variables appear in each functional relation, as we detail below. For a number of statistical tasks, including model selection, it has proven very useful to have insights about algebraic structure in characteristics of the joint distribution [1]. One prominent example are results on algebraic relations among second moments, i.e., covariances, in models that postulate linear functional relationships among the variables; see e.g. [2, 7].

In contrast to earlier work in algebraic statistics, we conduct here a first systematic study on algebraic relations that also involve higher moments, specifically we focus on second and third moments. Higher moments are of crucial importance for the case of non-Gaussian models, where different models may have identical structure in second but not in all higher moments [5, 6]. For instance, a particular type of algebraic relations among second and third moments was used in [8] to design a model selection method suitable for high-dimensional data.

We begin by reviewing the models of interest. Let $G = (V, E)$ be a directed acyclic graph (DAG), and let X_i, $i \in V$, be a collection of random variables indexed by the vertices in V. A vertex $j \in V$ is a parent of vertex i if there is an edge pointing from j to i, i.e., if $(j, i) \in E$. We denote the set of all *parents* of i by $\mathrm{pa}(i)$. The graph G gives rise to the linear structural equation model that postulates that

$$X_i = \sum_{j \in \mathrm{pa}(i)} \lambda_{j,i} X_j + \varepsilon_i, \quad i \in V.$$

where ε_i are mutually independent random variables that represent stochastic errors and have expectation $\mathbb{E}[\varepsilon_i] = 0$. The coefficients $\lambda_{j,i}$ are unknown parameters, and we fill them into a matrix $\Lambda = (\lambda_{j,i})$ by adding a zero entry when $(j, i) \notin E$. We denote the set of all such matrices as $\mathbb{R}^E$. We note that for simplicity, and without loss of generality, the above equations do not include a constant term, so $\mathbb{E}[X_i] = 0$ for all i.

Definition 1.1. *The linear structural equation model $\mathcal{M}^{(2,3)}(G)$ of second and third order moments corresponding to a directed acyclic graph $G = (V, E)$ with $|V| = n$ is defined as*

$$\begin{aligned}\mathcal{M}^{(2,3)}(G) = \{(S = (I-\Lambda)^{-T}\Omega^{(2)}(I-\Lambda)^{-1},\\ T = \Omega^{(3)} \bullet (I-\Lambda)^{-1} \bullet (I-\Lambda)^{-1} \bullet (I-\Lambda)^{-1}) :\\ \Omega^{(2)} \text{ is } n \times n \text{ positive definite diagonal matrix,}\\ \Omega^{(3)} \text{ is } n \times n \times n \text{ diagonal 3-way tensor, and } \Lambda \in \mathbb{R}^E\}.\end{aligned}$$

Here, $\bullet$ denotes the Tucker product*, see [4, Lemma 1]. Furthermore, denote by $\mathcal{I}^{(2,3)}(G)$ the ideal of polynomials in the entries s_{ij} of S and t_{ijk} of T that vanish on the model $\mathcal{M}^{(2,3)}(G)$.*

2 Parametrization

We always work over DAGs G with the edges topologically ordered so that if $i \to j \in E(G)$, then $i < j$.

Definition 2.1. *A* trek *(or* 2-trek*) τ from i to j is an alternating sequence of nodes and edges of the form*

$$i \leftarrow i_l \leftarrow \cdots \leftarrow i_1 \longleftarrow i_0 \longrightarrow j_1 \to \cdots \to j_r \to j.$$

(a trek takes you up and down a 'mountain'). The top *of the trek is* $\operatorname{top}(\tau) = i_0$*. We can generalize this notion to a n-*trek *between n vertices $i_1, \ldots, i_n$, as an ordered collection of n directed paths $T = (P_1, \ldots, P_n)$ where P_r has sink i_r and they all share the same top vertex as source,* $\operatorname{top}(T)$*.*

Define $T(i_1, \ldots, i_n)$ to be the collection of minimal –in the sense that it does not factor though another n-trek in the set– n-treks between $i_1, \ldots, i_n$. Consider the ring morphism

$$\begin{aligned}\phi_G : \quad \mathbb{C}[s_{ij}, t_{ijk} \mid 1 \le i \le j \le k \le n] &\to \mathbb{C}[a_i, b_i, \lambda_{ij} \mid i \to j \in E(G)]\\ s_{ij} &\mapsto \sum_{T \in T(i,j)} a_{\operatorname{top}(T)} \prod_{k \to l \in T} \lambda_{kl},\\ t_{ijk} &\mapsto \sum_{T \in T(i,j,k)} b_{\operatorname{top}(T)} \prod_{m \to l \in T} \lambda_{ml}.\end{aligned}$$

Proposition 2.2. *[Simple trek rule parametrization] Let $G = (V, E)$ be a DAG and ϕ_G the ring morphism defined above. Then $\mathcal{I}^{(2,3)}(G) = \ker \phi_G$.*

Proof. One can generalize the analogous construction for the variables $s_{i,j}$ that can be found in [7, Proposition 2.3]. □

For trees, the shortest n-trek between $i_1, \ldots, i_n$, if it exists, is unique. Therefore the notion of top is well-defined just by setting the end vertices, that is, $\operatorname{top}(i_1, \ldots, i_n)$ is the largest t such that the corresponding vertex is the top of the shortest n-trek between $i_1, \ldots, i_n$.

Corollary 2.3. *If G is a tree, $I^{(2,3)}(G)$ is a toric ideal.*

Proof. For trees, the set of minimal treks $T(i_1, \ldots, i_n)$ has at most one element: the unique shortest n-trek, if it exists. Therefore, the image of ϕ_G is generated by monomials. □

3 Tree DAGs

In this section we find equations defining the model $\mathcal{M}^{(2,3)}(G)$ whenever G is a tree, i.e., its underlying undirected graph is a tree. In this case, the model has a monomial parametrization, and we show that it is cut out by quadratic binonmials.

Definition 3.1. *Let $i, j \in V$ be two vertices such that a 2-trek between i and j exists. We define the matrix*

$$A_{i,j} := \begin{bmatrix} s_{ik_1} & \cdots & s_{ik_r} & t_{i\ell_1 m_1} & \cdots & t_{i\ell_q m_q} \\ s_{jk_1} & \cdots & s_{jk_r} & t_{j\ell_1 m_1} & \cdots & t_{j\ell_q m_q} \end{bmatrix},$$

where

- *$k_1, \ldots, k_r$ are all vertices such that $\mathrm{top}(i, k_a) = \mathrm{top}(j, k_a)$ and*
- *$(l_1, m_1), \ldots, (l_q, m_q)$ are all pairs of vertices such that $\mathrm{top}(i, l_b, m_b) = \mathrm{top}(j, l_b, m_b)$.*

Theorem 3.2. *Let $G = (V, E)$ be a tree DAG. For all $i, j \in V$ such that there is no trek between i and j, and for all $i, j, k \in V$ such that there is no 3-trek between i, j and k, we have that $s_{ij}, t_{ijk} \in \mathcal{I}^{(2,3)}(G)$. For all $i, j \in V$ such that there is a 2-trek between i and j, the 2×2 subdeterminants of A_{ij} lie in $\mathcal{I}^{(2,3)}(G)$.*

Proof. For all $i, j \in V$ such that there is no trek between them, by the trek rule, $s_{ij} = 0$. Similarly, by the 3-trek-rule, if there is no 3-trek between i, j, k, then $t_{ijk} = 0$.

Consider a matrix $A_{i,j}$ as in Definition 3.1. Fix any two columns of the matrix with indexes k (s columns) and l, m (t columns). Set $a := \mathrm{top}(i, k) = \mathrm{top}(j, k)$ and $b := \mathrm{top}(i, l, m) = \mathrm{top}(j, l, m)$.

The 2-treks between i and j with tops a and b, respectively, factor through the shortest 2-trek with top $\mathrm{top}(i, j)$. Moreover, there is exactly one path between two nodes of a tree, hence both the 2-treks (encoded in $s_{i,k}$ and s_{jk}) and 3-treks (encoded in t_{ilm} and t_{jlm}) factor through $\mathrm{top}(i, j)$ as follows:

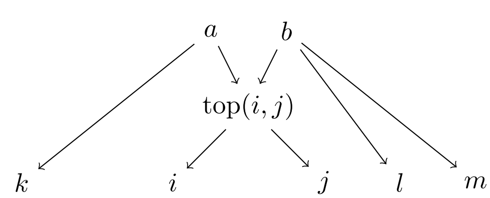

Figure 1: Factorization of 2-treks between i and j with tops a and b through unique shortest 2-trek.

From Figure 1 and Proposition 2.2, it follows that $s_{ik}t_{jlm} - s_{jk}t_{ilm}$ belongs to $\ker \phi_G$.

Cases corresponding to two s columns or two t columns are dealt anagously, since only central part of Figure 1 matters. $\square$

Remark 3.3. Note that $\mathrm{top}(i, j)$ is a *choke point* between $I = \{k, i\}$ and $J = \{j, l, m\}$ as defined in [7].

For the covariance case, all quadratic binomials arise from 2-minors of A_{ij} as in [7, Tetrad representation]. However this is no longer the case for third moments or combinations of covariance and moments:

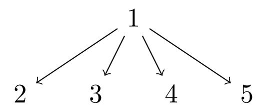

Figure 2: Not all quadratic binomials arise from 2-minors of $A_{i,j}$.

Example 3.4. In the graph below $f = s_{2,3}t_{1,4,5} - s_{4,5}t_{1,2,3} \in I_G$ and because of the lack of repeated indices in s, this cannot come from a 2-minor of A_{ij}. However, this binomial is not necessary in order to generate the vanishing ideal of the model. Indeed,

$$f = (s_{2,3}t_{1,4,5} - s_{3,4}t_{1,2,5}) + (s_{3,4}t_{1,2,5} - s_{4,5}t_{1,2,3})$$

is the sum of 2-minors of $A_{2,4}$ and $A_{3,5}$, respectively.

Proposition 3.5. *All quadratic binomials in $\mathcal{I}^{(2,3)}(G)$ are linear combinations of 2-minors of matrices $A_{i,j}$.*

Proof. Let $f = m_1 + m_2 \in I_G$ be a quadratic binomial. By 2.2, the tops of the underlying 2- or 3-treks of m_1 must be pairwise equal to the tops associated to m_2. Moreover, the indices of the variables in m_2 must be a permutation of the indices of m_1.

If f is a quadratic binomial on variables t, then $f = t_{i,j,k}t_{l,m,n} - t_{\sigma(i),\sigma(j),\sigma(k)}t_{\sigma(l),\sigma(m),\sigma(n)}$, where the $\{i,j,k\} \cup \{\sigma(i),\sigma(j),\sigma(k)\}$ has either one or two elements.

If the intersection consists of a single element, then we can assume $f = t_{i,j,k}t_{l,m,n} - t_{i,l,m}t_{j,k,n}$. If $\mathrm{top}(i,j,k) = \mathrm{top}(j,k,n)$ and $\mathrm{top}(i,l,m) = \mathrm{top}(l,m,n)$, then

$$\begin{array}{c|cc} & (j,k) & (l,m) \\ \hline i & t_{ijk} & t_{ilm} \\ n & t_{jkn} & t_{lmn} \end{array} \tag{1}$$

is a 2-minor of $A_{i,n}$. Otherwise, we need to exchange the elements in the skew-diagonal in order to have equal tops $b_1 := \mathrm{top}(i,j,k) = \mathrm{top}(i,l,m)$ and $b_2 := \mathrm{top}(j,k,n) = \mathrm{top}(l,m,n)$ arranged in columns. The resulting matrix

$$\begin{array}{c|cc} & i & n \\ \hline (j,k) & t_{ijk} & t_{jkn} \\ (l,m) & t_{ilm} & t_{lmn} \end{array} \tag{2}$$

has not the desired shape. Consider the two 4-treks between j,k,l and m, one with top b_1 and the other with top b_2. Because G is a tree, one must factor through the other. Assume $b_1 \leq b_2$, then the determinant of (2) can be written as the sum of the determinants of the corresponding 2-minors in $A_{k,l}$ and $A_{j,m}$.

If the intersection consists of two elements, we can proceed analogously.

If f is a quadratic binomial on both variables s and t, it can be proved that either f arises as a 2-minor of a matrix in 3.1 or it is a sum of two such matrices. $\square$

Remark 3.6. In (2) of the proof above, we see that $t_{i,j,k}t_{l,m,n} - t_{i,l,m}t_{j,k,n}$ is the sum of the determinants of

$$\begin{array}{c|cc} A_{k,l} & (i,j) & (m,n) \\ \hline k & t_{ijk} & t_{kmn} \\ l & t_{ijl} & t_{lmn} \end{array} \quad \text{and} \quad \begin{array}{c|cc} A_{j,m} & (i,l) & (k,n) \\ \hline j & t_{ijl} & t_{jkn} \\ m & t_{ilm} & t_{kmn} \end{array}. \tag{3}$$

The indices in $A_{k,l}$ and $A_{j,m}$ indicate which paths in the 3-trek we exchange between the two terms of the determinant, as represented in Figure 3.

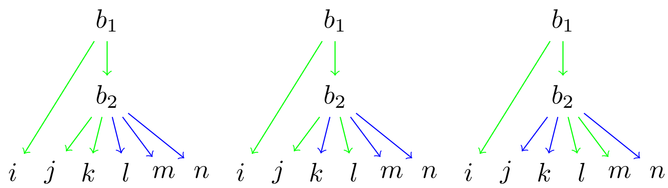

Figure 3: From left to right: pairs of 3-treks representing terms $t_{ijk}t_{lmn}$, $t_{ijl}t_{kmn}$, $t_{ilm}t_{jkn}$. First two pictures correspond to the determinant of the 2-minor of A_{kl} displayed in (3) and last two correspond to A_{jm}.

Let I be the ideal generated by matrices $A_{i,j}$ in Definition 3.1 and linear generators s_{ij} and t_{ijk} when there is no 2-trek between i, j or no 3-trek between i, j, k.

Conjecture 3.7. *All binomials in $\mathcal{I}^{(2,3)}(G)$ are generated by quadratic binomials, that is $\mathcal{I}^{(2,3)}(G) = I$.*

Although the equality of the ideals $\mathcal{I}^{(2,3)}(G)$ and I is yet to be proved, the next result shows that the model $\mathcal{M}^{(2,3)(G)}$ is the variety $V(I)$ intersected with the cone of positive definite matrices $PD(n)$. In fact, it is not necessary to consider all quadratic binomials arising from matrices $A_{i,j}$ in order to prove this set-theoretical equality.

Theorem 3.8. *Let J be the ideal generated by the linear generators of I but only matrices $A_{i,j}$ such that there is a directed path between i and j. Then the model $\mathcal{M}^{(2,3)}(G)$ is the variety $V(J)$ intersected with the cone $PD(n)$.*

Proof. Let S and T be the second and third moments of a distribution such that $(S,T) \in \mathcal{M}^{(2,3)}(G)$. For any edge $i \to j \in E$, consider the matrix $A_{i,j}$. Note that it contains as a column the vector $\left[s_{ii}, s_{ij}\right]^T$. Let $\lambda_{ij} = \frac{s_{ij}}{s_{ii}}$, and let $\Lambda \in \mathbb{R}^E$ contain these as its entries. Consider

$$S' = (I-\Lambda)^T S (I-\Lambda), \quad T' = T \bullet (I-\Lambda) \bullet (I-\Lambda) \bullet (I-\Lambda).$$

We will now show that S' and T' are diagonal. For any $i \neq j$, we have that

$$s'_{ij} = \sum_{k,\ell} (I-\Lambda)_{ki} s_{k\ell} (I-\Lambda)_{\ell j}$$

$$= s_{ij} - \sum_{k \to i \in E} \lambda_{ki} s_{kj} - \sum_{\ell \to j \in E} \lambda_{lj} s_{i\ell} + \sum_{k \to i, \ell \to j \in E} \lambda_{ki} \lambda_{\ell j} s_{k\ell}.$$

We are going to consider 3 cases.

Case 1: The shortest path between i and j has a collider. In this case, $s_{ij} = 0$. Furthermore, for any $k \to i$, the shortest path between k and j also has a collider: it either goes through i; or k is on the shortest path between i and j, but since k cannot be a collider, then, the shortest path between k and j has a collider. Thus, $s_{kj} = 0$, and the first sum is 0. Similarly, the other two sums are also 0. Therefore, $s'_{ij} = 0$.

Case 2: The shortest path between i and j is a trek of the form $i \leftarrow i_0 \cdots j_0 \to j$. Then, in the sums $\sum_{k \to i \in E} \lambda_{ki} s_{kj}$, $\sum_{\ell \to j \in E} \lambda_{lj} s_{i\ell}$, and $\sum_{k \to i, \ell \to j \in E} \lambda_{ki} \lambda_{\ell j} s_{k\ell}$ the only nonzero terms appear when $k = i_0$ and $\ell = j_0$ (otherwise, for example, the shortest path between k and j has a collider, namely, i, so $s_{kj} = 0$). When $k = i_0$ and $\ell = j_0$, $\lambda_{ki} = \frac{s_{ki}}{s_{kk}} = \frac{s_{ij}}{s_{kj}}$ since $\mathrm{top}(i,j) = \mathrm{top}(k,j)$, therefore, $(s_{ik}, s_{kk})^T, (s_{ij}, s_{kj})^T$ are columns in the matrix $A_{i,k}$. Similarly, $(s_{ji}, s_{\ell i})^T, (s_{jk}, s_{\ell k})^T, (s_{\ell j}, s_{\ell \ell})^T$ are columns in the matrix $A_{\ell,j}$ therefore, $\lambda_{\ell j} = \frac{s_{ij}}{s_{i\ell}} = \frac{s_{kj}}{s_{k\ell}}$. Therefore,

$$s'_{ij} = s_{ij} - \frac{s_{ij}}{s_{kj}} s_{kj} - \frac{s_{ij}}{s_{i\ell}} s_{i\ell} + \frac{s_{ij}}{s_{kj}} \frac{s_{kj}}{s_{k\ell}} s_{k\ell} = 0.$$

Note that either $s_{kj} \neq 0$ or both $s_{kj} = s_{ij} = 0$, hence s'_{ij} is either way well-defined. This situation is often repeated along the rest of the proof.

Case 3: The shortest path between i and j is a trek of the form: $i \to \cdots \to j_0 \to j$ or $i \leftarrow i_0 \cdots \leftarrow j$. WLOG, consider the former case. Then, the sums $\sum_{\ell \to j \in E} \lambda_{lj} s_{i\ell}$ and $\sum_{k \to i, \ell \to j \in E} \lambda_{ki}\lambda_{\ell j} s_{k\ell}$ are nonzero only when $\ell = j_0$ (otherwise, $s_{i\ell}$ and $s_{k\ell}$ are 0 because j is a collider on the shortest path between i and ℓ and the one between j and ℓ). Thus, we have that

$$s'_{ij} = s_{ij} - \sum_{k \in \mathrm{pa}(i)} \lambda_{ki} s_{kj} - \lambda_{j_0 j} s_{ij_0} + \sum_{k \in \mathrm{pa}(i)} \lambda_{ki}\lambda_{j_0 j} s_{kj_0}.$$

Note that $(s_{ji}, s_{j_0 i})^T, (s_{jj_0}, s_{j_0 j_0})^T$ are both columns in the matrix A_{j,j_0}, since $\mathrm{top}(i, j_0) = \mathrm{top}(i, j) = i$. Thus,

$$s'_{ij} = s_{ij} - \frac{s_{ij}}{s_{ij_0}} s_{ij_0} + \sum_{k \in \mathrm{pa}(i)} \lambda_{ki}(\lambda_{j_0 j} s_{kj_0} - s_{kj}).$$

Furthermore, since $(s_{kj}, s_{kj_0})^T, (s_{j_0 j}, s_{j_0 j_0})^T$ are columns in $A_{j_0,j}$, then

$$= \sum_{k \in \mathrm{pa}(i)} \lambda_{ki}\left(\frac{s_{kj}}{s_{kj_0}} s_{kj_0} - s_{kj}\right) = 0.$$

Therefore, S' is diagonal.

We now proceed to showing that T' is diagonal. Note that, because, G is a tree DAG, for any three vertices i, j, k for which there is a 3-trek between them, there is a unique shortest 3-trek. We have that

$$t'_{ijk} = \sum_{a,b,c} (I - \Lambda)_{ai}(I - \Lambda)_{bj}(I - \Lambda)_{ck} t_{abc}$$

$$= t_{ijk} - \sum_{a \to i \in E} \lambda_{ai} t_{ajk} - \sum_{b \to j \in E} \lambda_{bj} t_{ibk} - \sum_{c \to k \in E} \lambda_{ck} t_{ijc} + \sum_{a \to i, b \to j \in E} \lambda_{ai}\lambda_{bj} t_{abk}$$

$$+ \sum_{a \to i, c \to k \in E} \lambda_{ai}\lambda_{ck} t_{ajc} + \sum_{b \to j, c \to k \in E} \lambda_{bj}\lambda_{ck} t_{ibc} - \sum_{a \to i, b \to j, c \to k \in E} \lambda_{ai}\lambda_{bj}\lambda_{ck} t_{abc}$$

Case 1: There is no 3-trek between i, j, k, i.e., $t_{ijk} = 0$. Then, all of the 7 sums in the above expression will also be 0. This is because if there is a 3-trek between a, b, c such that $a = i$ or $a \to i$, and the same for b, j and c, k, then, we can complete this 3-trek to a 3-trek between i, j, k. Therefore, if there is no 3-trek between i, j, k, then, $t'_{ijk} = 0$.

Case 2: There is a 3-trek between i, j, k. Thus, i, j, k have a unique common closest ancestor, call it v.

Case 2.0: Suppose first that $v \neq i, j, k$. Therefore, the shortest 3-trek between i, j, k consists of 3 nontrivial directed paths $v \to \cdots \to i_0 \to i, v \to \cdots \to j_0 \to j$, and $v \to \cdots \to k_0 \to k$. Because G is a tree DAG, the seven sums above can be nonzero if and only if $a = i_0, b = j_0, c = k_0$ (otherwise, there cannot be a 3-trek between a, b, and c). Thus,

$$t'_{ijk} = t_{ijk} - \lambda_{i_0 i} t_{i_0 jk} - \lambda_{j_0 j} t_{ij_0 k} - \lambda_{k_0 k} t_{ijk_0} + \lambda_{i_0 i}\lambda_{j_0 j} t_{i_0 j_0 k}$$

$$+ \lambda_{i_0 i}\lambda_{k_0 k} t_{i_0 jk_0} + \lambda_{j_0 j}\lambda_{k_0 k} t_{ij_0 k_0} - \lambda_{i_0 i}\lambda_{j_0 j}\lambda_{k_0 k} t_{i_0 j_0 k_0}.$$

Since $(S, T) \in \mathcal{V}(I)$, we have that, for example, $(s_{i_0 i}, s_{i_0 i_0})^T, (t_{ijk}, t_{i_0 jk})^T$ are both columns in A_{i,i_0}, and, therefore, $\lambda_{i_0 i} = \frac{t_{ijk}}{t_{i_0 jk}}$ (and similarly for the other λ's in the expression above). Thus,

$$t'_{ijk} = t_{ijk} - \frac{t_{ijk}}{t_{i_0 jk}} t_{i_0 jk} - \frac{t_{ijk}}{t_{ij_0 k}} t_{ij_0 k} - \frac{t_{ijk}}{t_{ijk_0}} t_{ijk_0} + \frac{t_{ij_0 k}}{t_{i_0 j_0 k}} \frac{t_{ijk}}{t_{ij_0 k}} t_{i_0 j_0 k}$$

$$+ \frac{t_{ijk_0}}{t_{i_0 jk_0}} \frac{t_{ijk}}{t_{ijk_0}} t_{i_0 jk_0} + \frac{t_{ijk_0}}{t_{ij_0 k_0}} \frac{t_{ijk}}{t_{ijk_0}} t_{ij_0 k_0} - \frac{t_{ijk}}{t_{i_0 jk}} \frac{t_{i_0 jk}}{t_{i_0 j_0 k}} \frac{t_{i_0 j_0 k}}{t_{i_0 j_0 k_0}} t_{i_0 j_0 k_0} = 0.$$

Case 2.1: $v = i \neq j, k$. In this case, the shortest 3-trek between i, j, k consists of the trivial path i and paths $i \to \cdots \to j_0 \to j$, and $i \to \cdots \to k_0 \to k$. Therefore, we have that

$$\begin{aligned} t'_{ijk} &= t_{ijk} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}t_{ajk} - \lambda_{j_0j}t_{ij_0k} - \lambda_{k_0k}t_{ijk_0} + \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\lambda_{j_0j}t_{aj_0k} \\ &+ \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\lambda_{k_0k}t_{ajk_0} + \lambda_{j_0j}\lambda_{k_0k}t_{ij_0k_0} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\lambda_{j_0j}\lambda_{k_0k}t_{aj_0k_0} \\ &= t_{ijk} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}t_{ajk} - \frac{t_{ijk}}{t_{ij_0k}}t_{ij_0k} - \frac{t_{ijk}}{t_{ijk_0}}t_{ijk_0} + \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\frac{t_{ajk}}{t_{aj_0k}}t_{aj_0k} \\ &+ \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\frac{t_{ajk}}{t_{ajk_0}}t_{ajk_0} + \frac{t_{ijk}}{t_{ij_0k}}\frac{t_{ij_0k}}{t_{ij_0k_0}}t_{ij_0k_0} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\frac{t_{ajk}}{t_{aj_0k}}\frac{t_{aj_0k}}{t_{aj_0k_0}}t_{aj_0k_0} = 0. \end{aligned}$$

Case 2.2: $v = i = j \neq k$. In this case, the shortest 3-trek between i, j, k consists of the two trivial paths i and j, and $i \to \cdots \to k_0 \to k$. Therefore, we have that

$$\begin{aligned} t'_{iik} &= t_{iik} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}t_{aik} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}t_{iak} - \lambda_{k_0k}t_{iik_0} + \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}^2 t_{aak} \\ &+ \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\lambda_{k_0k}t_{aik_0} + \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\lambda_{k_0k}t_{iak_0} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}^2\lambda_{k_0k}t_{aak_0} \\ &= t_{iik} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}t_{aik} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}t_{iak} - \frac{t_{iik}}{t_{iik_0}}t_{iik_0} + \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}^2 t_{aak} \\ &+ \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\frac{t_{aik}}{t_{aik_0}}t_{aik_0} + \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}\frac{t_{iak}}{t_{iak_0}}t_{iak_0} - \sum_{a\in\mathrm{pa}(i)} \lambda_{ai}^2\frac{t_{aak}}{t_{aak_0}}t_{aak_0} = 0. \end{aligned}$$

Therefore, in all cases, $t'_{ijk} = 0$ unless $i = j = k$.

We have shown that

$$S' = (I-\Lambda)^T S(I-\Lambda), \quad T' = T \bullet (I-\Lambda) \bullet (I-\Lambda) \bullet (I-\Lambda)$$

are diagonal, and S' is positive definite since S is, which means that

$$S = (I-\Lambda)^{-T} S'(I-\Lambda)^{-1}, \quad T = T' \bullet (I-\Lambda)^{-1} \bullet (I-\Lambda)^{-1} \bullet (I-\Lambda)^{-1}$$

lie in our model. □

Example 3.9. Recall the graph in Example 3.4. Computations in `Macaulay2` show that $\mathcal{I}^{(2,3)}(G) = I = (J : s_{11}^{\infty})$. In particular, $\mathcal{M}^{(2,3)}(G) = V(I) \cap PD(5) = V(J) \cap PD(5)$.

4 An application: trees with latent variables

In this section we apply our computational tools and theoretical findings to find the ideals of factor analysis models with latent (or hidden) variables.

Definition 4.1. *Let $H \cup O$ be a partition of the nodes of the DAG G. The hidden nodes H are said to be* upstream *from the observed nodes O in G if there are no edges $o \leftarrow h$ in G with $o \in O$ and $h \in H$.*

Example 4.2. Recall the graph in Example 3.4. Consider the partition $G = H \cup O$ with $H = \{1\}$ and $O = \{2, 3, 4, 5\}$. Let I_O be the vanishing ideal of the model over the observed variables. `Macaulay2` computations show that $I_O = I \cap \mathbb{C}[s_{ij}, t_{ijk} : i, j, k \in O] = K$, where K is the ideal generated by the $A_{i,j}$ matrices associated to G after removing all columns and rows containing variables with index 1. In particular, it is a binomial ideal minimally generated by 126 quadratic binomials.

The behavior displayed in Example 4.2 is known to be true for vanishing ideals of partially observed Gaussian models arising from trees [7, Section 6]. Assuming Conjecture 3.7 to be true, then $I_O^{(2,3)}(G)$ for any DAG tree G is generated by the minors of the submatrices of $A_{i,j}$ with i, j both in O, with columns indexed by k and l, m where k, l, m are all in O. In the rest of this section we explain how to prove this result.

For an upstream partition we define here a grading on the ring $\mathbb{C}[a, b, \lambda]$, which induces a grading on $\mathbb{C}[s, t]$ such that the corresponding moment ideal is homogeneous. Let $\deg a_h = (1, 1)$ for all $h \in H$ and $\deg a_o = (1, 3)$ for all $o \in O$. Similarly let $\deg b_h = (1, 0)$ for all $h \in H$ and $\deg b_o = (1, 3)$ for all $o \in O$. Finally let $\deg \lambda_{ho} = (0, 1)$ for all $h \in H$ and $o \in O$, and $\deg \lambda_{i,j} = (0, 0)$ otherwise.

Lemma 4.3. *For such an upstream partition, the above grading induces a grading on $\mathbb{C}[s, t]$ with*

$$\deg s_{ij} = (1, 1 + \textit{number of elements in the multiset } \{i, j\} \textit{ in } O)$$

and

$$\deg t_{ijk} = (1, \textit{number of elements in the multiset } \{i, j, k\} \textit{ in } O).$$

Thus the ideal $I^{(2,3)}(G)$ is homogeneous with respect to this grading.

Using this lemma and assuming Conjecture 3.7 to be correct, we can prove the claim because the only homogeneous polynomials of degree $(n, 3n)$ in $I^{(2,3)}$ must entirely consist of variables $s_{i,j}$, $t_{i,j,k}$ with all indices i, j, k in O.

Acknowledgements

This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 883818).

References

[1] Mathias Drton. Algebraic problems in structural equation modeling. In *The 50th anniversary of Gröbner bases*, volume 77 of *Adv. Stud. Pure Math.*, pages 35–86. Math. Soc. Japan, Tokyo, 2018.

[2] Mathias Drton, Elina Robeva, and Luca Weihs. Nested covariance determinants and restricted trek separation in Gaussian graphical models. *Bernoulli*, 26(4):2503–2540, 2020.

[3] Marloes Maathuis, Mathias Drton, Steffen Lauritzen, and Martin Wainwright, editors. *Handbook of graphical models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL, 2019.

[4] Elina Robeva and Jean-Baptiste Seby. Multi-trek separation in linear structural equation models. *arXiv:2001.10426*, 2020.

[5] Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.*, 7:2003–2030, 2006.

[6] Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvärinen, Yoshinobu Kawahara, Takashi Washio, Patrik O. Hoyer, and Kenneth Bollen. DirectLiNGAM: a direct method for learning a linear non-Gaussian structural equation model. *J. Mach. Learn. Res.*, 12:1225–1248, 2011.

[7] Seth Sullivant. Algebraic geometry of Gaussian bayesian networks. *Advances in Applied Mathematics*, 40(4):482–513, 2008.

[8] Y. Samuel Wang and Mathias Drton. High-dimensional causal discovery under non-Gaussianity. *Biometrika*, 107(1):41–59, 2020.