*Bernoulli* **26**(4), 2020, 2503–2540
https://doi.org/10.3150/19-BEJ1179

# Nested covariance determinants and restricted trek separation in Gaussian graphical models

MATHIAS DRTON[1], ELINA ROBEVA[2] and LUCA WEIHS[3]

[1] *Department of Mathematics, Technical University of Munich, Munich, Germany.*
*E-mail:* mathias.drton@tum.de
[2] *Department of Mathematics, University of British Columbia, Vancouver, BC, Canada.*
*E-mail:* erobeva@math.ubc.ca
[3] *Department of Statistics University of Washington, Seattle, WA, USA. E-mail:* lucaw@uw.edu

Directed graphical models specify noisy functional relationships among a collection of random variables. In the Gaussian case, each such model corresponds to a semi-algebraic set of positive definite covariance matrices. The set is given via a parametrization, and much work has gone into obtaining an implicit description in terms of polynomial (in)equalities. Implicit descriptions shed light on problems such as parameter identification, model equivalence and constraint-based statistical inference. For models given by directed acyclic graphs, which represent settings where all relevant variables are observed, there is a complete theory: All conditional independence relations can be found via graphical $d$-separation and are sufficient for an implicit description. The situation is far more complicated, however, when some of the variables are hidden (or in other words, unobserved or latent). We consider models associated to mixed graphs that capture the effects of hidden variables through correlated error terms. The notion of trek separation explains when the covariance matrix in such a model has submatrices of low rank and generalizes $d$-separation. However, in many cases, such as the infamous Verma graph, the polynomials defining the graphical model are not determinantal, and hence cannot be explained by $d$-separation or trek-separation. In this paper, we show that these constraints often correspond to the vanishing of nested determinants and can be graphically explained by the (more general) notion of *restricted trek separation*.

*Keywords:* conditional independence; covariance matrix; graphical model; trek separation; Verma constraint

## 1. Introduction

Let $G = (V, \mathcal{E})$ be a directed graph with finite vertex set $V$ and edge set $\mathcal{E} \subseteq V \times V$. The edge set is always assumed to be free of self-loops, so $(i, i) \notin \mathcal{E}$ for all $i \in V$. For each vertex $i$, define a set of parents $\mathrm{pa}(i) = \{j \in V : (j, i) \in \mathcal{E}\}$. The graph $G$ induces a statistical model for the joint distribution of a collection of random variables $X_i$, $i \in V$, indexed by the graph's vertices. The model hypothesizes that each variable is a function of the parent variables and an independent noise term. In this paper, we consider the Gaussian case, in which the functional relationships are linear so that

$$X_i = \lambda_{0i} + \sum_{j \in \mathrm{pa}(i)} \lambda_{ji} X_j + \epsilon_i, \quad i \in V, \tag{1.1}$$

1350-7265 

where the $\epsilon_i$, $i \in V$, are independent and centered Gaussian random variables. The coefficients $\lambda_{0i}$ and $\lambda_{ji}$ are unknown real parameters that are assumed to be such that the system (1.1) admits a unique solution $X = (X_i : i \in V)$. Typically termed a system of structural equations, (1.1) specifies cause-effect relations whose straightforward interpretability is behind the wide-spread use of the models (Spirtes, Glymour and Scheines [22], Pearl [18]).

The random vector $X$ that solves (1.1) follows a Gaussian distribution whose mean vector may be arbitrary through the choice of the parameters $\lambda_{0i}$ but whose covariance matrix is highly structured. The model obtained from (1.1) thus naturally corresponds to the set of covariance matrices, which we denote by $\mathcal{M}(G)$. This set is given parametrically with each covariance being a rational or even polynomial function of the parameters $\lambda_{ji}$ and the variances of the errors $\epsilon_i$, as we detail in Section 2.

While a parametrization is useful to specify a distribution and to optimize the likelihood function, many statistical problems can only be solved with some understanding of an implicit description, that is, a way of telling whether a given covariance matrix lies in the model $\mathcal{M}(G)$. In our setting, an implicit description of the model amounts to a semialgebraic description of the set of covariance matrices that belong to the model through polynomial equations and inequalities, and a combinatorial criterion on the graph which specifies how to obtain them. Specific problems that can be addressed through such an implicit description include model equivalence, parameter identification and constraint-based statistical inference. We refer the reader to the recent work of van Ommen and Mooij [26] and the reviews of Drton [4] and Drton and Maathuis [6].

If the underlying graph $G$ is an acyclic digraph, also termed a directed acyclic graph (DAG), then probabilistic conditional independence yields an implicit description of $\mathcal{M}(G)$ (Lauritzen [16], Studený [23]). For a Gaussian joint distribution, conditional independence corresponds to the vanishing of special subdeterminants of the covariance matrix, namely, subdeterminants that are almost principal in the sense that the row and the column index sets agree in all but one element (Lněnička and Matúš [17], Drton, Sturmfels and Sullivant [8], Chapter 3.1). The conditional independences holding in all distributions in the given model can be found graphically using the concept of $d$-separation. It follows in particular that two DAGs $G$ and $H$ give rise to the same model $\mathcal{M}(G) = \mathcal{M}(H)$ if and only if $G$ and $H$ have the same $d$-separation relations. This combinatorial criterion can be simplified to yield an efficient algorithm: $\mathcal{M}(G) = \mathcal{M}(H)$ if and only if $G$ and $H$ have the same skeleta and the same sets of unshielded colliders (Frydenberg [14], Verma and Pearl [27]).

While DAG models are well-understood, they only pertain to problems where all relevant variables are observed. A long-standing program in the fields of graphical modeling and causal inference seeks to develop combinatorial solutions to problems such as model equivalence in settings with hidden/latent variables. Here, giving a combinatorial explanation for the defining equations of the model of a mixed graph could be used to devise a combinatorial criterion for when two mixed graphs give rise to the same model. Mathematically, if only the variables indexed by a set $A \subset V$ are observed while those indexed by $V \setminus A$ are hidden, then the covariance matrices in the set $\mathcal{M}(G)$ are to be projected on their principal $A \times A$ submatrix. It is well known that conditional independence is no longer sufficient for implicit model description after such a projection.

**Example 1.1.** Let $G$ be the DAG in Figure 1, where vertex 5 indexes a hidden variable. Then no conditional independence involving only the observed $X_1$, $X_2$, $X_3$ and $X_4$ holds for all covari-

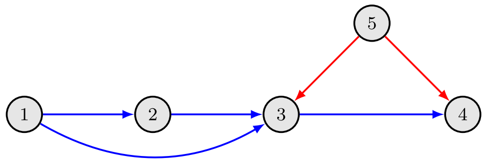


**Figure 1.** A DAG on five vertices. Vertex 5 indexes a hidden variable.

ance matrices in $\mathcal{M}(G)$. Instead, a positive definite $4 \times 4$ matrix $\Sigma = (\sigma_{ij})$ is the projection of a matrix in $\mathcal{M}(G)$ if and only if

$$|\Sigma_{12,34}| := \det(\Sigma_{12,34}) = \sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23} = 0$$

and $\sigma_{j3} = 0$ implies $\sigma_{j4} = 0$ for $j = 1, 2$.

In the example just given the key constraint is a determinant of the covariance matrix that cannot be explained by $d$-separation. A major advance in this decade was the introduction of trek separation, which is a graphical criterion that can be used to decide the vanishing of any subdeterminant of the covariance matrix (Draisma, Sullivant and Talaska [2], Sullivant, Talaska and Draisma [24]). Although more work is required to fully exploit trek separation in model equivalence criteria, the notion has already seen application in parameter identification problems (Weihs et al. [28]).

While greatly generalizing Gaussian conditional independence, determinantal constraints are again not sufficient to describe the sets $\mathcal{M}(G)$ after projection to the covariance matrix of observed variables. The following example is due to Thomas Verma.

**Example 1.2.** Let $G$ be the graph from Figure 2. Then as in the first example no conditional independence that holds for $\mathcal{M}(G)$ involves only the observed variables $X_1$, $X_2$, $X_3$ and $X_4$. Instead, a positive definite $4 \times 4$ matrix $\Sigma = (\sigma_{ij})$ is the projection of a matrix in $\mathcal{M}(G)$ if and only if

$$\begin{aligned} f_{\text{Verma}} = {} & \sigma_{11}\sigma_{13}\sigma_{22}\sigma_{34} - \sigma_{11}\sigma_{13}\sigma_{23}\sigma_{24} - \sigma_{11}\sigma_{14}\sigma_{22}\sigma_{33} + \sigma_{11}\sigma_{14}\sigma_{23}^2 \\ & - \sigma_{12}^2\sigma_{13}\sigma_{34} + \sigma_{12}^2\sigma_{14}\sigma_{33} + \sigma_{12}\sigma_{13}^2\sigma_{24} - \sigma_{12}\sigma_{13}\sigma_{14}\sigma_{23} = 0; \end{aligned} \tag{1.2}$$

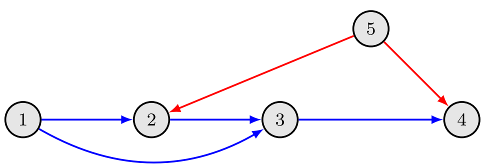


**Figure 2.** The Verma graph. Vertex 5 indexes a hidden variable.

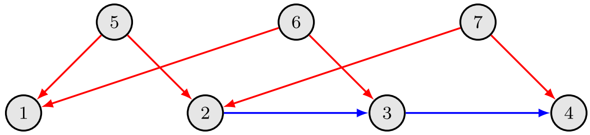


**Figure 3.** Graph based on van Ommen and Mooij [26], Figure 1. Vertices 5, 6 and 7 index hidden variables.

compare Example 3.3.14 in Drton, Sturmfels and Sullivant [8]. The polynomial $f_{\text{Verma}}$ is not a subdeterminant of $\Sigma$ and, therefore, is neither explained by $d$-separation nor by trek-separation.

Another key advance in the area is a graph decomposition result of Tian and Pearl [25]; see also Drton [4], Sections 5–6. This result allows one to derive constraints by applying $d$-separation in certain subgraphs. In particular, the vanishing polynomial $f_{\text{Verma}}$ from (1.2) can be shown to arise from the independence of variables $X_1$ and $X_4$ that holds for the subgraph obtained by removing the edges $1 \to 3$ and $2 \to 3$ from the Verma graph in Figure 2. For further details, we refer the reader to the review of Shpitser et al. [21].

In the next example however, neither Tian's graph decomposition nor trek separation provide any insight.

**Example 1.3.** Let $G$ be the graph from Figure 3. There are four observed variables, and projecting $\mathcal{M}(G)$ gives a set of codimension one. As discussed in van Ommen and Mooij [26], any covariance $\Sigma = (\sigma_{ij}) \in \mathcal{M}(G)$ satisfies the constraint

$$f_{\text{vOM}} = \sigma_{22}\sigma_{34}\sigma_{13} - \sigma_{22}\sigma_{33}\sigma_{14} - \sigma_{23}\sigma_{24}\sigma_{13} + \sigma_{23}^2\sigma_{14} = 0. \tag{1.3}$$

The irreducible polynomial in (1.3) defines the hypersurface that contains the projection of $\mathcal{M}(G)$.

A closer look at Examples 1.2 and 1.3 reveals some common structure. Both constraints are *nested determinants*, by which we mean determinants of a matrix whose entries are determinants themselves. This observation is the point of departure for our paper.

**Example 1.4.** The Verma polynomial from Example 1.2 admits a compact representation through nested determinants, namely,

$$f_{\text{Verma}} = \begin{vmatrix} \frac{|\Sigma_{123,123}|}{|\Sigma_{1,3}|} & \frac{|\Sigma_{123,124}|}{|\Sigma_{1,4}|} \end{vmatrix}. \tag{1.4}$$

Such a representation is generally not unique. For instance,

$$f_{\text{Verma}} = \begin{vmatrix} \frac{|\Sigma_{123,134}|}{|\Sigma_{1,1}|} & \frac{|\Sigma_{123,234}|}{|\Sigma_{1,2}|} \end{vmatrix} = \begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,13}| \\ |\Sigma_{34,12}| & |\Sigma_{34,13}| \end{vmatrix}. \tag{1.5}$$

The polynomial from Example 1.3 is also a nested determinant, namely,

$$f_{\mathrm{vOM}} = \begin{vmatrix} |\Sigma_{23,23}| & |\Sigma_{23,24}| \\ |\Sigma_{1,3}| & |\Sigma_{1,4}| \end{vmatrix}. \tag{1.6}$$

We are not aware of any literature emphasizing these types of representations.

In this paper, we investigate combinatorial conditions on the graph $G$ that entail the vanishing of nested determinants. We give a rigorous definition of the models we study in Section 2, where we also provide background on the current knowledge of their description. In particular, we introduce mixed graph models that play an important role in model selection Drton and Maathuis[6], Section 5.2. Section 3 shows how nested determinants arise under conditions of ancestrality. In Theorem 3.8, we show that such determinants completely describe the model $\mathcal{M}(G)$ for a wide class of mixed graphs that are (nearly) ancestral. Section 4 describes our notion of *restricted trek separation* in the setting of arbitrary acylic mixed graphs. In Section 5, we show how the vanishing of nested determinants can follow from restricted trek separation. The result we present also implies the vanishing of the constraints exhibited for (nearly) ancestral graphs in Section 3. In Section 6, we give examples that involve recursive nesting of determinants. Although these examples are beyond the scope of our results, we can explain them via restricted trek separation. While our focus is on acyclic mixed graphs, our last example shows that a nested determinant may also arise for graphs containing directed cycles. Finally, in Section 7 we discuss future work and open problems.

# 2. Background

## 2.1. Structural equation models

Let $\epsilon = (\epsilon_i : i \in V)$ be the random error vector for the equation system in (1.1). As we are only concerned with the covariance structure, we disregard the offsets $\lambda_{0i}$. The system can then be written as

$$X = \Lambda^T X + \epsilon, \tag{2.1}$$

where the matrix $\Lambda = (\lambda_{ij}) \in \mathbb{R}^{V \times V}$ holds the unknown coefficients. Let $\Omega = (\omega_{ij}) = \mathrm{Var}[\epsilon] \in \mathbb{R}^{V \times V}$ be the covariance matrix of $\epsilon$, which we assume positive definite. Assuming further that $I - \Lambda$ is invertible, the random vector $X = (I - \Lambda)^{-1}\epsilon$ is the unique solution to the linear system in (2.1) and has covariance matrix

$$\mathrm{Var}[X] = (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1}. \tag{2.2}$$

In the Introduction, we focused on the case where the individual error terms $\epsilon_i$ are independent. Their covariance matrix $\Omega$ is then diagonal. In this case, a model postulating that some of the coefficients in $\Lambda$ are zero is conveniently represented by a directed graph, as was our setup in Section 1. Going forward, we also allow for dependence among the $\epsilon_i$ and a possibly nondiagonal matrix $\Omega$. Nonzero off-diagonal terms of $\Omega$ are commonly represented by adding bidirected edges to the considered directed graph.

A *mixed graph* is a triple $G = (V, \mathcal{D}, \mathcal{B})$, where $\mathcal{D} \subset V \times V$ is the set of *directed* edges, and $\mathcal{B}$ is the set of *bidirected* edges which is comprised of unordered pairs of elements of $V$. We denote a directed edge from $i$ to $j$ by $i \to j$, and a bidirected edge by $i \leftrightarrow j$. Let $\mathbb{R}^{\mathcal{D}}$ be the set of $V \times V$ matrices $\Lambda$ with support $\mathcal{D}$, that is,

$$\mathbb{R}^{\mathcal{D}} = \{\Lambda \in \mathbb{R}^{V\times V} : \lambda_{ij} = 0 \text{ if } i \to j \notin \mathcal{D}\}.$$

Let $\mathbb{R}^{\mathcal{D}}_{\mathrm{reg}}$ be the subset of matrices $\Lambda \in \mathbb{R}^{\mathcal{D}}$ for which $I - \Lambda$ is invertible. Let $PD_V$ be the cone of positive definite $V \times V$ matrices, and define $PD(\mathcal{B})$ to be the subcone of matrices supported over $\mathcal{B}$, that is,

$$PD(\mathcal{B}) = \{\Omega = (\omega_{ij}) \in PD_V : \omega_{ij} = 0 \text{ if } i \neq j \text{ and } i \leftrightarrow j \notin \mathcal{B}\}.$$

The mixed graph $G$ is acyclic if its directed part $(V, \mathcal{D})$ does not contain any directed cycles. Such graphs are also referred to as acyclic directed mixed graphs (ADMGs); see, for example, Evans and Richardson [11]. Here, a *directed cycle* is a sequence of vertices $v_1 \to v_2 \to \cdots \to v_k \to v_1$ connected by directed edges, where $k > 1$. If $G$ is acyclic, then its vertex set $V$ can be ordered such that all matrices $\Lambda \in \mathbb{R}^{\mathcal{D}}$ are strictly upper triangular. Thus, the determinant $|I - \Lambda| = 1$ and $\mathbb{R}^{\mathcal{D}} = \mathbb{R}^{\mathcal{D}}_{\mathrm{reg}}$. By Cramer's rule, the covariances in $\mathrm{Var}[X]$ in (2.2) are then polynomial functions of the entries of $\Lambda$ and $\Omega$.

Taking the error $\epsilon$ to be Gaussian, a given mixed graph induces the following statistical model for the joint distribution of $X$.

**Definition 2.1.** The *linear structural equation model* given by a mixed graph $G = (V, \mathcal{D}, \mathcal{B})$ is the family of all multivariate normal distributions on $\mathbb{R}^V$ with covariance matrix in the set

$$\mathcal{M}(G) = \{(I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1} : \Lambda \in \mathbb{R}^{\mathcal{D}}_{\mathrm{reg}}, \Omega \in PD(\mathcal{B})\}.$$

The set $\mathbb{R}^{\mathcal{D}}_{\mathrm{reg}} \times PD(\mathcal{B})$ is semialgebraic. Since $\mathcal{M}(G)$ is the image of this set under a rational map, the Tarski–Seidenberg theorem yields that $\mathcal{M}(G)$ itself is a semialgebraic set, and thus, admits a polynomial description. In this paper, we are interested in studying polynomial equations that are satisfied by the matrices in $\mathcal{M}(G)$. With $\Sigma = (\sigma_{ij})$ interpreted as a symmetric $V \times V$ matrix of indeterminates, define $\mathbb{R}[\Sigma]$ to be the ring of polynomials in the $\sigma_{ij}$. Then the polynomial relations we seek to understand make up the vanishing ideal

$$\mathcal{I}(G) := \{f \in \mathbb{R}[\Sigma] : f(\Sigma) = 0 \text{ for all } \Sigma \in \mathcal{M}(G)\}.$$

Suppose a variable $X_j$, $j \in V$, is hidden. Then the remaining variables $(X_i : i \neq j)$ have their covariance matrix in the set obtained by projecting each matrix in $\mathcal{M}(G)$ onto its $(V \setminus \{j\}) \times (V \setminus \{j\})$ submatrix. Two comments are in order. First, we emphasize that for a fixed $j \in V$, the polynomials $f \in \mathcal{I}(G)$ that do not involve any of the indeterminates indexed by $j$, that is, $f$ is free of $\sigma_{jk}$ for $k \in V$, give precisely the polynomial constraints holding for the model in which random variable $X_j$ is hidden. Second, the paradigm of mixed graphs allows one to directly

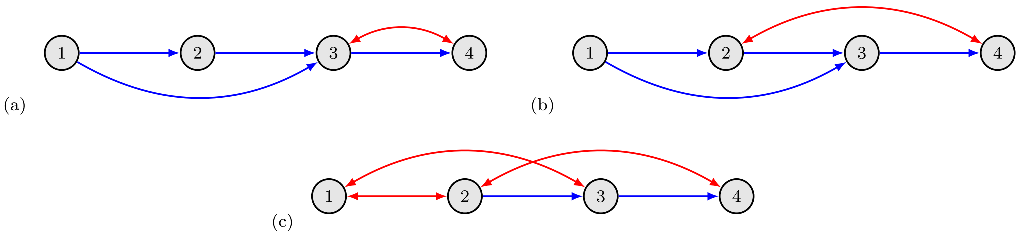


**Figure 4.** (a)–(c) Mixed graphs obtained by latent projection of the DAGs in Figures 1–3, respectively.

capture relations after projection. Indeed, a graphical operation known as "latent projection" creates a new mixed graph $G'$ over the observed variables that represents key relations among covariances of observed variables; see Pearl [18], Section 2.6, Koster [15] or Wermuth [29]. For instance, the three examples from our Introduction would be represented by the three mixed graphs in Figure 4. In these examples, the ideal $\mathcal{I}(G')$ of the given mixed graph coincides with the ideal of polynomial relations among the observed covariances in the hidden variable model given by the original DAG $G$.

## 2.2. Trek rule

Again let $G = (V, \mathcal{D}, \mathcal{B})$ be any mixed graph, possibly cyclic. The starting point for any combinatorial understanding of polynomials in the vanishing ideal $\mathcal{I}(G)$ is the *trek rule*. This rule specifies each entry of the covariance matrix in (2.2) as a sum of monomials associated with certain paths in the graph.

**Definition 2.2.** A *trek* is a path $\tau$ of the form:

(a) $i_\ell \leftarrow \cdots \leftarrow i_1 \leftrightarrow j_1 \rightarrow \cdots \rightarrow j_r$, or
(b) $i_\ell \leftarrow \cdots \leftarrow i_1 = j_1 \rightarrow \cdots \rightarrow j_r$,

for integers $\ell, r \geq 0$ with $\ell + r \geq 1$. Here, a path may visit a vertex more than once. If $\ell = 0$, the trek is simply the directed path $j_1 \rightarrow \cdots \rightarrow j_r$. Similarly, it is $i_\ell \leftarrow \cdots \leftarrow i_1$ if $r = 0$. We call $\tau$ a trek from $i_\ell$ to $j_r$ or also a trek between $i_\ell$ and $j_r$. The sets $\{i_k : k = 1, \dots, \ell\}$ and $\{j_k : k = 1, \dots, r\}$ are the *left side* and the *right side* of $\tau$, respectively.

Let $\Lambda = (\lambda_{ij}) \in \mathbb{R}^{\mathcal{D}}_{\mathrm{reg}}$ and $\Omega = (\omega_{ij}) \in PD(\mathcal{B})$. To any trek $\tau$, specified as in Definition 2.2, associate a *trek monomial*

$$\sigma(\tau) = \omega_{i_1 j_1} \prod_{k=1}^{\ell-1} \lambda_{i_k i_{k+1}} \prod_{k=1}^{r-1} \lambda_{j_k j_{k+1}}. \tag{2.3}$$

The *trek rule* now states that the covariance matrix $\Sigma = (\sigma_{ij}) = (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1}$ has its entries given by

$$\sigma_{ij} = \sum_{\text{treks } \tau \text{ from } i \text{ to } j} \sigma(\tau). \tag{2.4}$$

The rule, which originates in the work of Wright [30], is obtained by observing that $(I - \Lambda)^{-1} = I + \Lambda + \Lambda^2 + \cdots$. The right-hand side of (2.4) is a polynomial when $G$ is acyclic and a (formal) power series otherwise.

## 2.3. Conditional independence and subdeterminants

The notion of $d$-separation allows one to decide by inspection of paths in a mixed graph $G$ whether a conditional independence relation holds for all distributions in the model given by $G$; see, for example, Drton [4], Section 10. In algebraic terms, for a Gaussian joint distribution, variables $X_i$ and $X_j$ are conditionally independent given a subvector $X_S$ with $i, j \notin S$ if and only if the subdeterminant $|\Sigma_{iS,jS}|$ is zero. Here, $iS$ denotes the union of a singleton set $\{i\}$ and the set $S$. Thus, $d$-separation gives a combinatorial characterization of when a subdeterminant of the form $|\Sigma_{iS,jS}|$ belongs to the ideal $\mathcal{I}(G)$. If $G$ is a DAG, then the covariance model $\mathcal{M}(G)$ admits a semialgebraic description by conditional independence. Indeed, $\mathcal{M}(G)$ is the set of positive definite matrices $\Sigma$ for which all conditional independence determinants $|\Sigma_{iS,jS}|$ associated with the graph $G$ vanish. This is also true for mixed graphs that are maximal ancestral (Richardson and Spirtes [19]), but false more generally as the examples in the introduction show.

In seminal work, Sullivant, Talaska and Draisma [24] move beyond conditional independence determinants and give a combinatorial characterization of when an arbitrary subdeterminant $|\Sigma_{A,B}|$ is in $\mathcal{I}(G)$. We briefly review their concept of trek-separation; see also Drton [4], Section 11.

**Definition 2.3.** Two sets $A, B \subseteq V$ are *trek-separated* by the pair $(S_L, S_R)$, where $S_L, S_R \subseteq V$, if every trek between a vertex from $A$ and a vertex from $B$ intersects either $S_L$ on its left side or $S_R$ on its right side.

The vanishing of subdeterminants of the covariance matrix $\Sigma$ in Gaussian DAG models may now be characterized as follows.

**Theorem 2.4 (Theorem 2.17, Sullivant, Talaska and Draisma [24]).** *Let $A, B \subseteq V$. The submatrix $\Sigma_{A,B}$ has rank at most $r$ for all covariance matrices $\Sigma \in \mathcal{M}(G)$ if and only if there exist subsets $S_L, S_R \subseteq V$ such that $|S_L| + |S_R| \le r$ and $(S_L, S_R)$ trek-separates $A$ from $B$. For a generic choice of $\Sigma \in \mathcal{M}(G)$,*

$$\mathit{rank}(\Sigma_{A,B}) = \min\{|S_L| + |S_R| : (S_L, S_R) \text{ trek-separates } A \text{ from } B\}.$$

In the case $|A| = |B| = m$, Theorem 2.4 shows that $|\Sigma_{A,B}| \in \mathcal{I}(G)$ if and only if the sets of vertices $A$ and $B$ are *trek-separated* by a pair $(S_L, S_R)$ with $|S_L| + |S_R| < |A|$.

While trek-separation greatly generalizes $d$-separation and can yield a generating set of $\mathcal{I}(G)$ for some mixed graphs (Fink, Rajchgot and Sullivant [12]), it is in general not sufficient to understand the vanishing ideal $\mathcal{I}(G)$ as we demonstrated in Examples 1.2 and 1.3.

# 3. Ancestral vertices and overdetermined linear systems

We now proceed to a first combinatorial condition (see Proposition 3.4) for the vanishing of very special nested determinants (Definition 3.1). Fix a mixed graph $G = (V, \mathcal{D}, \mathcal{B})$, and let $\Sigma \in PD_V$. For a pair of matrices $\Lambda \in \mathbb{R}^{\mathcal{D}}_{\mathrm{reg}}$ and $\Omega \in PD(\mathcal{B})$, it holds that

$$\Sigma = (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1} \quad \Longleftrightarrow \quad (I - \Lambda)^T\Sigma(I - \Lambda) = \Omega.$$

In turn, $\Sigma \in \mathcal{M}(G)$ if and only if

$$\left[(I - \Lambda)^T\Sigma(I - \Lambda)\right]_{ij} = 0 \quad \forall i, j \in V \text{ with } i \neq j, i \leftrightarrow j \notin \mathcal{B}. \tag{3.1}$$

For some graphs it is known that all entries of $\Lambda$ can be recovered as rational expressions of $\Sigma$, at least for generic choices of positive definite $\Sigma$. For instance, the half-trek criterion (Foygel, Draisma and Drton [13]) and its extensions (Chen [1], Drton and Weihs [9], Weihs et al. [28]) can be used to certify graphically that such *rational identification* of $\Lambda$ from $\Sigma$ is possible and to find rational expressions. If now both the $i$th and the $j$th column of $\Lambda$ are rationally identifiable from $\Sigma$, then the left-hand side of the equation in (3.1) can be expressed as a rational function of $\Sigma$. If $i \neq j$ and $i \leftrightarrow j \notin \mathcal{B}$, then one finds a rational constraint on $\Sigma$ that after clearing denominators yields a polynomial in $\mathcal{I}(G)$. This approach is used, for instance, by van Ommen and Mooij [26].

In this section, we follow a similar approach in which we substitute solutions for some of the entries of $\Lambda$ that appear in (3.1). However, we only linearize the equations and then observe that nested determinantal constraints arise from overdetermined linear equation systems. Specifically, we study the following constraints.

**Definition 3.1.** Let $i$ be a vertex of the mixed graph $G$, and let $J$ be a subset of vertices in $G$. Define a matrix of polynomials of size $(|\operatorname{pa}(i)| + |J|) \times (|\operatorname{pa}(i)| + 1)$ as

$$F_{i,J} = \left(|\Sigma_{\operatorname{pa}(r)\uplus\{r\},\operatorname{pa}(r)\uplus\{c\}}|\right)_{r\in\operatorname{pa}(i)\uplus J, c\in\operatorname{pa}(i)\uplus\{i\}}. \tag{3.2}$$

The *parentally nested determinants* for the pair $(i, J)$ are the minors of order $|\operatorname{pa}(i)| + 1$ of the matrix $F_{i,J}$. When $J = \{j\}$ is a singleton, there is only one parentally nested determinant

$$f_{ij} = \left|\left(|\Sigma_{\operatorname{pa}(r)\uplus\{r\},\operatorname{pa}(r)\uplus\{c\}}|\right)_{r\in\operatorname{pa}(i)\uplus\{j\}, c\in\operatorname{pa}(i)\uplus\{i\}}\right|. \tag{3.3}$$

Here, index sets are treated as multisets with possibly repeated elements, and the determinants are formed according to a prespecified linear order for the vertex set $V$. The symbol $\uplus$ stands for the sum (or disjoint union) of multisets; for example, $\{1, 1, 2\} \uplus \{1, 3\} = \{1, 1, 1, 2, 3\}$.

Suppose $j \in J \cap \operatorname{pa}(i)$. Then $j$ is repeated in the row index set $\operatorname{pa}(i) \uplus J$ for the matrix $F_{i,J}$. In this case $j$ indexes two rows for a minor, which is then zero. In particular, if $j \in \operatorname{pa}(i)$ then $f_{ij} = 0$. We may therefore always restrict the set $J$ to satisfy $J \cap \operatorname{pa}(i) = \varnothing$.

A repeated index may also arise for the column index sets of the matrices whose determinants yield the entries of $F_{i,J}$. Indeed, if $c \in \mathrm{pa}(i) \cup \{i\}$ is also in $\mathrm{pa}(r)$ for $r \in \mathrm{pa}(i) \cup \{j\}$, then the $(r, c)$ entry of $F_{i,J}$ is zero.

**Example 3.2.** It holds that $f_{\mathrm{Verma}} = f_{41}$ in Example 1.2, and $f_{\mathrm{vOM}} = f_{41}$ in Example 1.3.

In the remainder of this section, we identify conditions under which parentally nested determinants vanish.

**Definition 3.3.** A vertex $j$ in the mixed graph $G = (V, \mathcal{D}, \mathcal{B})$ is *ancestral* if (i) $j$ is not on any directed cycle, and (ii) no vertex $k \neq j$ has both $k \leftrightarrow j \in \mathcal{B}$ and a directed path from $k$ to $j$.

Let $\Lambda \in \mathbb{R}^{\mathcal{D}}_{\mathrm{reg}}$ and $\Omega \in PD(\mathcal{B})$, and define $\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1} \in \mathcal{M}(G)$. If $j$ is ancestral, then all treks from a vertex $r \in \mathrm{pa}(j)$ to $j$ end with a directed edge pointing to $j$. The trek rule from (2.4) then implies that

$$\Sigma_{\mathrm{pa}(j),\mathrm{pa}(j)} \Lambda_{\mathrm{pa}(j),j} = \Sigma_{\mathrm{pa}(j),j}. \tag{3.4}$$

For our next result, it is convenient to introduce the set of *siblings* of a vertex $j$, which is $\mathrm{sib}(j) = \{k \in V : k \leftrightarrow i \in \mathcal{B}\}$, the set of neighbors of $j$ in the bidirected part of the graph.

**Proposition 3.4.** *Let $i$ be a vertex of a mixed graph $G = (V, \mathcal{D}, \mathcal{B})$ such that*:

(i) $\mathrm{pa}(i) \cap \mathrm{sib}(i) = \varnothing$,
(ii) *all vertices in* $\mathrm{pa}(i)$ *are ancestral*, *and*
(iii) *the set $J$ of all ancestral vertices in $V \setminus (\mathrm{pa}(i) \cup \mathrm{sib}(i) \cup \{i\})$ is nonempty.*

*Then the parentally nested determinants for the pair $(i, J)$ are in the vanishing ideal $\mathcal{I}(G)$.*

**Proof.** Let $\Lambda = (\lambda_{ab}) \in \mathbb{R}^{\mathcal{D}}_{\mathrm{reg}}$ and $\Omega \in PD(\mathcal{B})$, and define $\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1} \in \mathcal{M}(G)$. Neither $\mathrm{pa}(i)$ nor $J$ contains vertices in $\mathrm{sib}(i)$. Fixing $r \in pa(i) \cup J$, (3.1) implies that

$$\left[(I - \Lambda)^T \Sigma (I - \Lambda)\right]_{ri} = 0. \tag{3.5}$$

This equation becomes

$$\sigma_{ri} - \Lambda^T_{\mathrm{pa}(r),r} \Sigma_{\mathrm{pa}(r),i} - \Sigma_{r,\mathrm{pa}(i)} \Lambda_{\mathrm{pa}(i),i} + \Lambda^T_{\mathrm{pa}(r),r} \Sigma_{\mathrm{pa}(r),\mathrm{pa}(i)} \Lambda_{\mathrm{pa}(i),i} = 0. \tag{3.6}$$

Since all vertices in $\mathrm{pa}(i) \cup J$ are ancestral, we may use (3.4) to get the rational equation

$$\sigma_{ri} - \Sigma_{r,\mathrm{pa}(r)} \Sigma^{-1}_{\mathrm{pa}(r),\mathrm{pa}(r)} \Sigma_{\mathrm{pa}(r),i} - \big(\Sigma_{r,\mathrm{pa}(i)} - \Sigma_{r,\mathrm{pa}(r)} \Sigma^{-1}_{\mathrm{pa}(r),\mathrm{pa}(r)} \Sigma_{\mathrm{pa}(r),\mathrm{pa}(i)}\big) \Lambda_{\mathrm{pa}(i),i} = 0. \tag{3.7}$$

Now observe that for any vertex $c$,

$$\big(\sigma_{rc} - \Sigma_{r,\mathrm{pa}(r)} \Sigma^{-1}_{\mathrm{pa}(r),\mathrm{pa}(r)} \Sigma_{\mathrm{pa}(r),c}\big) |\Sigma_{\mathrm{pa}(r),\mathrm{pa}(r)}| = |\Sigma_{\mathrm{pa}(r)\cup\{r\},\mathrm{pa}(r)\cup\{c\}}|. \tag{3.8}$$

**Figure 5.** (a) A DAG on 4 vertices to illustrate the nested determinants $f_{ij}$. (b) A subgraph with two edges.

Hence, multiplying the equation in (3.7) by $|\Sigma_{\text{pa}(r),\text{pa}(r)}|$ gives

$$|\Sigma_{\text{pa}(r)\cup\{r\},\text{pa}(r)\cup\{i\}}| - \sum_{c\in\text{pa}(i)} |\Sigma_{\text{pa}(r)\cup\{r\},\text{pa}(r)\cup\{c\}}| \cdot \lambda_{ci} = 0. \tag{3.9}$$

With one equation for every $r \in \text{pa}(i) \cup J$, the system is overdetermined and admits a solution only if the matrix $F_{i,J}$ from Definition 3.1 has rank at most $|\text{pa}(i)|$. This in turn implies the vanishing of its minors. Note that in the case that $i$ is not trek reachable from $r$, the last equation is trivial and corresponds to a row of zeros in $F_{i,J}$. □

We now illustrate the vanishing of parentally nested determinants through several examples.

**Example 3.5.** In our first example, consider the graph $G$ from Figure 5(a). This graph is a DAG, and thus all its vertices are ancestral. As there are no bidirected edges, $f_{ij} \in \mathcal{I}(G)$ for all $i \neq j$. As previously noted, for any graph $f_{ij} = 0$ if $j \in \text{pa}(i)$. Here, $f_{21} = f_{32} = f_{42} = f_{43} = 0$. Moreover, $f_{12} = f_{34} = 0$. The nonzero polynomials are

$$f_{13} = |\Sigma_{12,23}|, \qquad f_{31} = \sigma_{11} \cdot |\Sigma_{12,23}|, \qquad f_{23} = -\sigma_{12} \cdot |\Sigma_{12,23}|$$
$$f_{14} = |\Sigma_{123,234}|, \qquad f_{41} = \sigma_{11}\sigma_{22} \cdot |\Sigma_{123,234}|, \qquad f_{24} = -\sigma_{12} \cdot |\Sigma_{123,234}|.$$

The irreducible polynomial $f_{13}$ corresponds to conditional independence of $X_1$ and $X_3$ given $X_2$. The second irreducible polynomial $f_{14}$ encodes conditional independence of $X_1$ and $X_4$ given $(X_2, X_3)$. It turns out that

$$\begin{aligned}\mathcal{M}(G) &= \{\Sigma \in PD_{\{1,2,3,4\}} : f_{13}(\Sigma) = f_{14}(\Sigma) = 0\} \\ &= \{\Sigma \in PD_{\{1,2,3,4\}} : f_{31}(\Sigma) = f_{41}(\Sigma) = 0\} \\ &= \{\Sigma \in PD_{\{1,2,3,4\}} : f_{ij}(\Sigma) = 0 \ \forall i \neq j\}.\end{aligned}$$

In fact, the ideal $\langle f_{ij} : i \neq j\rangle = \langle f_{13}, f_{14}\rangle$ differs from $\mathcal{I}(G)$ only through components that do not vanish at positive definite matrices Roozbehani and Polyanskiy [20], Example 2. Specifically,

$$\langle f_{13}, f_{14}\rangle = \mathcal{I}(G) \cap \langle |\Sigma_{23,23}|, |\Sigma_{13,23}|, |\Sigma_{12,23}|\rangle.$$

Here,

$$\mathcal{I}(G) = \langle |\Sigma_{12,23}|, |\Sigma_{12,24}|, |\Sigma_{12,34}|\rangle$$

is generated by three subdeterminants, two of which are conditional independences.

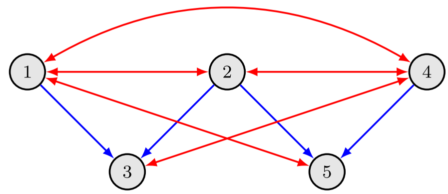


**Figure 6.** An ancestral graph that is not maximal.

For a second example, we pass to a subgraph to emphasize that even for DAGs parentally nested determinants need not factor into conditional independence determinants.

**Example 3.6.** Let $G$ to be the subgraph illustrated in Figure 5(b). Again, this is a DAG and $f_{ij} \in \mathcal{I}(G)$ for all $i \neq j$. We find eight nonzero polynomials. Of these, six correspond to (conditional) independences, namely,

$$f_{13} = f_{31} = \sigma_{13}, \qquad f_{14} = f_{41} = |\Sigma_{14,34}|, \qquad f_{23} = f_{32} = |\Sigma_{12,13}|.$$

However, the remaining two are not conditional independence determinants and also do not factor into such determinants. Instead,

$$f_{24} = f_{42} = -\sigma_{12}\sigma_{14}\sigma_{33} + \sigma_{11}\sigma_{24}\sigma_{33} + \sigma_{12}\sigma_{13}\sigma_{34} - \sigma_{11}\sigma_{23}\sigma_{34}.$$

As in the previous example, $\mathcal{M}(G)$ is comprised exactly of those positive definite matrices at which the $f_{ij}$ vanish. Similarly, the ideal $\mathcal{I}(G) = \langle \sigma_{13}, \sigma_{14}, \sigma_{23}, \sigma_{24} \rangle$ differs from that generated by the $f_{ij}$ from only through components that do not vanish at positive definite matrices:

$$\langle f_{ij} : i \neq j \rangle = \mathcal{I}(G) \cap \langle \sigma_{13}, \sigma_{23}, \sigma_{33} \rangle \cap \langle \sigma_{11}, \sigma_{13}, \sigma_{33} \rangle \cap \langle \sigma_{11}, \sigma_{13}, \sigma_{14} \rangle.$$

Finally, our third example is a mixed graph, whose model cannot be described using conditional independence alone.

**Example 3.7.** The mixed graph in Figure 6 is an ancestral graph, that is, all vertices are ancestral (Richardson and Spirtes [19]). It is not maximal, that is, there are nonadjacent vertices, namely, 3 and 5, that cannot be $d$-separated. There is then no conditional independence constraint associated to the non-adjacency. Precisely two of the $f_{ij}$ are nonzero, namely,

$$f_{35} = \begin{vmatrix} |\Sigma_{1,1}| & |\Sigma_{1,2}| & |\Sigma_{1,3}| \\ |\Sigma_{1,2}| & |\Sigma_{2,2}| & |\Sigma_{2,3}| \\ |\Sigma_{124,245}| & |\Sigma_{224,245}| & |\Sigma_{234,245}| \end{vmatrix} = \begin{vmatrix} |\Sigma_{1,1}| & |\Sigma_{1,2}| & |\Sigma_{1,3}| \\ |\Sigma_{1,2}| & |\Sigma_{2,2}| & |\Sigma_{2,3}| \\ |\Sigma_{124,245}| & 0 & |\Sigma_{234,245}| \end{vmatrix},$$

$$f_{53} = \begin{vmatrix} |\Sigma_{2,2}| & |\Sigma_{2,4}| & |\Sigma_{2,5}| \\ |\Sigma_{4,2}| & |\Sigma_{4,4}| & |\Sigma_{4,5}| \\ |\Sigma_{123,122}| & |\Sigma_{123,124}| & |\Sigma_{123,125}| \end{vmatrix} = \begin{vmatrix} |\Sigma_{2,2}| & |\Sigma_{2,4}| & |\Sigma_{2,5}| \\ |\Sigma_{4,2}| & |\Sigma_{4,4}| & |\Sigma_{4,5}| \\ 0 & |\Sigma_{123,124}| & |\Sigma_{123,125}| \end{vmatrix}.$$

In fact, $f_{35} = f_{53}$, and $\mathcal{I}(G) = \langle f_{35} \rangle = \langle f_{53} \rangle$. We note that there is also the alternative representation of

$$f_{35} = \begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,23}| \\ |\Sigma_{124,245}| & |\Sigma_{234,245}| \end{vmatrix} = \begin{vmatrix} |\Sigma_{24,24}| & |\Sigma_{24,25}| \\ |\Sigma_{123,124}| & |\Sigma_{123,125}| \end{vmatrix} = f_{53}.$$

We now give a model description for a class of graphs that includes all ancestral graphs. It also covers the two graphs from Figure 4(b)(c). We call $G$ *globally identifiable* if for every covariance matrix $\Sigma \in \mathcal{M}(G)$, there are unique parameters $\Lambda \in \mathbb{R}^D$, $\Omega \in PD(B)$ such that $\Sigma = (I - \Lambda)^{-T}\Omega(I - \Lambda)^{-1}$. Recall that a *sink* of a mixed graph $G = (V, \mathcal{D}, \mathcal{B})$ is any vertex that is not a parent of any other vertex. A subgraph of $G$ is a mixed graph $G' = (V', \mathcal{D}', \mathcal{B}')$ with $V' \subseteq V$, $\mathcal{D}' \subseteq \mathcal{D}$, and $\mathcal{B}' \subseteq \mathcal{B}$. It turns out that a graph $G$ is globally identifiable if and only if $G$ is acyclic and none of its subgraphs $G' = (V', \mathcal{D}', \mathcal{B}')$ containing at least 2 vertices has both a connected bidirected part $(V', \mathcal{B}')$ and a unique sink vertex in its directed part $(V', \mathcal{D}')$ (Drton, Foygel and Sullivant [5]).

For any set of polynomials $\mathcal{F} \subset \mathbb{R}[\Sigma]$, we let $\mathcal{V}_{\mathcal{F}} = \{\Sigma : f(\Sigma) = 0 \forall f \in \mathcal{F}\}$ be the algebraic subset it defines in the space of symmetric matrices.

**Theorem 3.8.** *Let $G = (V, \mathcal{D}, \mathcal{B})$ be a globally identifiable mixed graph with vertex set $V = [p] \equiv \{1, \dots, p\}$ enumerated in a topological order. Suppose all vertices in $[p-1]$ are ancestral. Let $\mathcal{F}(G)$ be the set of all parentally nested determinants obtained from the pairs $(i, [i-1] \setminus (\mathrm{pa}(i) \cup \mathrm{sib}(i)))$ for $i \in V$. Then*

$$\mathcal{M}(G) = PD_V \cap \mathcal{V}_{\mathcal{F}(G)}.$$

**Proof.** The inclusion $\mathcal{M}(G) \subseteq PD_V \cap \mathcal{V}_{\mathcal{F}(G)}$ follows from application of Proposition 3.4.

To show the reverse inclusion, we proceed by induction on the number of vertices $p$. The statement is trivial for $p = 1$. In the induction step, let $\Sigma \in PD_V \cap \mathcal{V}_{\mathcal{F}(G)}$. Let $\Sigma_{[p-1],[p-1]}$ be the submatrix obtained by removing the $p$th row and column. Let $G[p-1] = ([p-1], \mathcal{D}[p-1], \mathcal{B}[p-1])$ be the subgraph induced by $[p-1]$. Now, $\Sigma_{[p-1],[p-1]} \in PD_{[p-1]} \cap \mathcal{V}_{\mathcal{F}(G[p-1])}$. The induction hypothesis yields that $\Sigma_{[p-1],[p-1]} \in \mathcal{M}(G[p-1])$. Let $\Sigma_{[p-1],[p-1]} = (I - \Lambda')^{-T}\Omega'(I - \Lambda')^{-1}$ for $\Lambda' \in \mathbb{R}^{\mathcal{D}[p-1]}$ and $\Omega' \in PD(\mathcal{B}[p-1])$.

Consider the matrix $F_{p,[p-1]\setminus(\mathrm{pa}(p)\cup\mathrm{sib}(p))}$ from Definition 3.1 evaluated at the given matrix $\Sigma$. For each $r \in [p-1] \setminus \mathrm{sib}(p)$, divide the corresponding row by $|\Sigma_{\mathrm{pa}(r),\mathrm{pa}(r)}| > 0$. The resulting matrix $\bar{F}$ has entries

$$\sigma_{rc} - \Sigma_{r,\mathrm{pa}(r)}\Sigma^{-1}_{\mathrm{pa}(r),\mathrm{pa}(r)}\Sigma_{\mathrm{pa}(r),c} \tag{3.10}$$

for $r \in [p-1] \setminus \mathrm{sib}(p)$ and $c \in \mathrm{pa}(p) \cup \{p\}$; recall (3.8). Using (3.4), we obtain that

$$\bar{F} = \left[\left(I - \Lambda'\right)^T \Sigma_{[p-1],V}\right]_{[p-1]\setminus\mathrm{sib}(p),\mathrm{pa}(p)\cup\{p\}}. \tag{3.11}$$

Form the submatrix $\bar{F}_{[p-1]\setminus\mathrm{sib}(p),\mathrm{pa}(p)}$, that is, we omit the column indexed by $p$. Then

$$\bar{F}_{[p-1]\setminus\mathrm{sib}(p),\mathrm{pa}(p)} = \left[\Omega'\left(I - \Lambda'\right)^{-1}\right]_{[p-1]\setminus\mathrm{sib}(p),\mathrm{pa}(p)}.$$

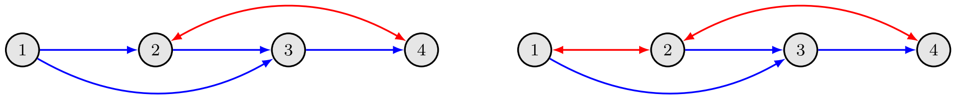


**Figure 7.** Almost ancestral identifiable graphs in the equivalence class of the Verma graph.

Lemma 2 in Drton, Foygel and Sullivant [5] yields that $\bar{F}_{[p-1]\setminus \mathrm{sib}(p),\mathrm{pa}(p)}$ has full column rank.

Since $\Sigma \in \mathcal{V}_{\mathcal{F}(G)}$, the matrix $F_{p,[p-1]\setminus(\mathrm{pa}(p)\cup\mathrm{sib}(p))}$ and thus also $\bar{F}$ do not have full column rank. We conclude that the kernel of $\bar{F}$ contains a vector $x \in \mathbb{R}^{\mathrm{pa}(p)\cup\{p\}}$ for which the last coordinate $x_p \neq 0$. Dividing $x_{\mathrm{pa}(p)}$ by $-x_p$ gives a vector $\lambda_{\mathrm{pa}(p),p} \in \mathbb{R}^{\mathrm{pa}(p)}$ that solves the equation system in (3.9). Define a $p \times p$ matrix $\Lambda \in \mathbb{R}^{\mathcal{D}}$ by using $\lambda_{\mathrm{pa}(p),p}$ to define its last column. Then $\Lambda$ solves (3.5) for $i = p$ and all $r \in [p-1] \setminus \mathrm{sib}(p)$, and thus also (3.1). Therefore, $\Sigma \in \mathcal{M}(G)$. $\square$

Theorem 3.8 allows us to check equivalence of graphs that satisfy the conditions listed in the theorem, that is, mixed graphs that are identifiable and for which the vertices in $[p-1]$ are ancestral. To check equivalence, we create the defining equations for the model given by the first graph (a set of parentally nested determinants), and plug in the parametrization corresponding to the second graph to see if the equations hold there. Then we repeat the procedure with the two graphs switched. An interesting problem for future work is to find a combinatorial criterion that circumvents the algebra and decides the equivalence by only using the graphs.

**Example 3.9.** Consider the Verma graph from Figure 4(b). The equivalence class contains two graphs – the Verma graphs itself and the same graph with the edge $1 \to 2$ changed to a bidirected edge, depicted in Figure 7.

The above facts leverage existence of ancestral vertices. In the next sections, we seek to give a more general condition for the vanishing of nested determinants. The results on vanishing nested determinants from this section can be recovered as a special case; see Proposition 5.8.

## 4. Restricted trek separation

As we reviewed in Section 2.3, the notion of trek separation (Sullivant, Talaska and Draisma [24]) provides a combinatorial characterization of when a subdeterminant of the covariance matrix $\Sigma$ vanishes over a model $\mathcal{M}(G)$. Underlying the trek separation result, we stated in Theorem 2.4 is the observation that determinants correspond to sums of certain products of trek monomials. In this section, we recall this observation and then introduce a notion of *restricted trek separation*, in which separation only needs to occur with respect to treks that avoid certain vertices on their left or right sides. This notion will be used in Section 5 to obtain conditions that imply the vanishing of nested determinants.

Let $A$ and $B$ be two subsets of the vertex set of a mixed graph $G$, with $|A| = |B|$. A *system of treks* from $A$ to $B$ is a set of treks that each are between a vertex in $A$ and a vertex in $B$. Let $\mathcal{T}$ be such a system. Then $\mathcal{T}$ has *no sided intersection* if any two distinct treks in $\mathcal{T}$ have disjoint left sides and disjoint right sides. In particular, each vertex in $A$ and each vertex in $B$ is on precisely one trek, so that $\mathcal{T}$ induces a bijection between $A$ and $B$. Fixing an ordering of the elements of $A$ and $B$, the trek system induces a permutation of $B$ in which the $i$th element of $B$ is mapped to the end point of the trek that starts at the $i$th element of $A$. Write $(-1)^{\mathcal{T}}$ for the sign of this permutation. Now define

$$\mathcal{P}_{A,B} = \sum (-1)^{\mathcal{T}} \prod_{\tau \in \mathcal{T}} \sigma(\tau) \tag{4.1}$$

with the summation being over all systems of treks $\mathcal{T}$ from $A$ to $B$ with no sided intersection; recall the definition of trek monomials from (2.3).

**Theorem 4.1 (Draisma, Sullivant and Talaska, [2]).** *Suppose the underlying graph $G$ is acyclic. Then the determinant of $\Sigma_{A,B}$ equals $\mathcal{P}_{A,B}$.*

This result admits a generalization to the case where the graph $G$ contains directed cycles. Indeed, Draisma, Sullivant and Talaska [2] give a rational expression for the determinant of $\Sigma_{A,B}$ in terms of self-avoiding trek flows, which reduce to trek systems without sided intersection in the acyclic case. As this generalization is more involved, we will not give any details here and focus instead on acyclic graphs only.

We now extend the combinatorial characterization of determinants and the trek separation result from Theorem 2.4 to allow for *restricted* treks.

**Definition 4.2.** Let $A$, $B$, $P$ and $Q$ be subsets of vertices of a mixed graph $G$. A $(P, Q)$-*restricted trek* between $A$ and $B$ is a trek between a vertex in $A$ and a vertex in $B$ that has its left side in $P$ and its right side in $Q$. Let $S_L$ and $S_R$ be two further subsets of vertices. Then $A$ and $B$ are $(P, Q)$-*restricted trek-separated* by $(S_L, S_R)$ if every $(P, Q)$-restricted trek between $A$ and $B$ intersects $S_L$ on its left side or $S_R$ on its right side.

**Example 4.3.** Consider the Verma graph from Figure 4(b). Take $A = \{2, 4\}$, $B = \{2, 3\}$, $P = \{2, 4\}$ and $Q = \{2, 3, 4\}$. Then $A$ and $B$ are $(P, Q)$-restricted trek-separated by $(\{\}, \{2\})$. Indeed, every trek between $A$ and $B$ that only uses $P$ on the left and only uses $Q$ on the right has to go through 2 on the right. Note, however, that this is not true if, for example, $P = Q = V$ or if $3 \in P$.

The main observation of this section is that restricted trek separation is equivalent to a rank constraint on a special matrix. Note also that part (ii) of the theorem is a direct generalization of Theorem 4.1 to the restricted case.

**Theorem 4.4.** *Let $G = (V, \mathcal{D}, \mathcal{B})$ be an acyclic mixed graph, and let $\Lambda \in \mathbb{R}^{\mathcal{D}}$ and $\Omega \in PD(\mathcal{B})$. For $P, Q \subseteq V$, consider the matrix*

$$\Sigma^{(P,Q)} = \left[(I - \Lambda)_{P,P}\right]^{-T} \Omega_{P,Q} \left[(I - \Lambda)_{Q,Q}\right]^{-1},$$

*and its submatrix $\Sigma^{(P,Q)}_{A,B}$ for a choice of row indices $A \subseteq P$ and column indices $B \subseteq Q$.*

(i) *The rank of* $\Sigma^{(P,Q)}_{A,B}$ *is at most*

$$\min\{|S_L| + |S_R| : A \text{ and } B \text{ are } (P,Q)\text{-restricted trek-separated by } (S_L, S_R)\},$$

*and equal to this minimum generically.*

(ii) *If* $|A| = |B|$, *then the determinant of* $\Sigma^{(P,Q)}_{A,B}$ *is equal to*

$$\mathcal{P}_{A,B,(P,Q)} = \sum (-1)^{\mathcal{T}} \prod_{\tau \in \mathcal{T}} \sigma(\tau),$$

*where the summation runs over all systems of treks* $\mathcal{T}$ *that comprise only* $(P,Q)$*-restricted treks from* $A$ *to* $B$ *and have no sided intersection.*

The proof of Theorem 4.4 is located in Appendix A. It is analogous to the proofs of Theorems 2.4 and 4.1 as developed in Sullivant, Talaska and Draisma [24] and Draisma, Sullivant and Talaska [2].

**Example 4.5.** Consider once more the Verma graph from Figure 4(b). Let $A = \{2, 4\}$, $B = \{2, 3\}$, $P = \{2, 4\}$ and $Q = \{2, 3, 4\}$. We saw in Example 4.3 that $A$ and $B$ are $(P, Q)$-restricted trek-separated by $(\{\}, \{2\})$. Now consider the matrix

$$\begin{aligned}\Sigma^{(P,Q)} &= \left[(I - \Lambda)_{24,24}\right]^{-T} \Omega_{24,234} \left[(I - \Lambda)_{234,234}\right]^{-1} \\ &= \begin{pmatrix} \omega_{22} & \omega_{22}\lambda_{23} & \omega_{22}\lambda_{23}\lambda_{34} + \omega_{24} \\ \omega_{24} & \omega_{24}\lambda_{23} & \omega_{24}\lambda_{23}\lambda_{34} + \omega_{44} \end{pmatrix}.\end{aligned}$$

As predicted by Theorem 4.4(i), the submatrix $\Sigma^{(P,Q)}_{A,B} = \Sigma^{(P,Q)}_{24,23}$ has rank 1.

# 5. Nested determinants

In this section, we demonstrate how restricted trek separation may lead to polynomial equations in the vanishing ideal $\mathcal{I}(G)$ of the model $\mathcal{M}(G)$ of an acyclic mixed graph $G$. These equations are in general not determinantal, instead they are given by specific types of nested determinants. In Section 5.1, we introduce a *swapping property* based on which in Theorem 5.3 we show how restricted trek separation gives rise to the vanishing of such nested determinants. In Section 5.2, we show how the swapping property and Theorem 5.3 are sufficient to explain the vanishing of the parentally nested determinants from Proposition 3.4 and Theorem 3.8 in terms of restricted trek separation.

## 5.1. Restricted trek separation and nested determinants

We begin by defining a *swapping property* that allows us to factor certain subdeterminants of $\Sigma$. Recall that $\uplus$ denotes disjoint union (of multisets).

**Definition 5.1.** Let $A_1, \dots, A_k, B_1, \dots, B_k$ be sets of vertices of an acyclic mixed graph $G$ with $|A_i| = |B_i|$ for every $i = 1, \dots, k$. Suppose *every* system of treks without sided intersection between $A_1 \uplus \cdots \uplus A_k$ and $B_1 \uplus \cdots \uplus B_k$ connects $A_i$ to $B_i$ for every $i$. Moreover, suppose that for any two trek systems $\mathcal{T}_1$ and $\mathcal{T}_2$ without sided intersection between $A_1 \uplus \cdots \uplus A_k$ and $B_1 \uplus \cdots \uplus B_k$ if we swap the treks between $A_i$ and $B_i$ from $\mathcal{T}_1$ with those from $\mathcal{T}_2$, then we obtain another two trek systems with no sided intersection between $A_1 \uplus \cdots \uplus A_k$ and $B_1 \uplus \cdots \uplus B_k$. Then we say that $(A_1, B_1), \dots, (A_k, B_k)$ satisfy the *swapping property*.

The next lemma, which is proven in Appendix B.1, shows how the swapping property gives factorizations of subdeterminants of $\Sigma$ into different systems of trek monomial sums.

**Lemma 5.2.** *Assume that $(A_1, B_1), \dots, (A_k, B_k)$ satisfy the swapping property. Then*

$$|\Sigma_{A_1 \uplus \cdots \uplus A_k, B_1 \uplus \cdots \uplus B_k}| = \prod_{i=1}^{k} \mathcal{P}_{A_i, B_i, (C_i, D_i)},$$

*where $C_1, \dots, C_k, D_1, \dots, D_k$ are sets of vertices determined by the trek systems without sided intersection between $A_1 \uplus \cdots \uplus A_k$ and $B_1 \uplus \cdots \uplus B_k$. More specifically, $C_i$ and $D_i$ are some sets of vertices such that $C_i$ contains all left vertices and $D_i$ contains all right vertices of the induced trek system between $A_i$ and $B_i$.*

We now proceed to the main result of this section, Theorem 5.3, which shows that the swapping property for suitable sets of vertices implies that certain nested determinants can be factored into sums of restricted trek monomial systems. Later in the section we will see that though the conditions of this theorem appear to be quite special, they are very natural. In particular, they apply to a multitude of examples, and moreover, the theorem generalizes our results from Proposition 3.4 and Theorem 3.8.

**Theorem 5.3.** *Let $a_1, \dots, a_n, b_1, \dots, b_n \in V$. For $i = 1, \dots, n$, let $A_i, B_i, C_i, D_i \subseteq V$ be four subsets with $|A_i| = |B_i|$ and $|C_i| = |D_i|$. Assume further that for each $i, j$, the sets $(A_i, B_i)$, $(C_j, D_j)$ and $(\{a_i\}, \{b_j\})$ satisfy the swapping property such that*

$$|\Sigma_{A_i \uplus C_j \uplus \{a_i\}, B_i \uplus D_j \uplus \{b_j\}}| = \mathcal{P}_{A_i, B_i, (P_i, Q_i)} \mathcal{P}_{C_j, D_j, (R_j, S_j)} \mathcal{P}_{a_i, b_j, (E_{ij}, F_{ij})}$$

*for sets of vertices $P_i, Q_i, R_j, S_j, E_{ij}, F_{ij}$ and every $i$ and $j$. Assume also that*

$$\left| (\mathcal{P}_{a_i, b_j, (E_{ij}, F_{ij})})_{1 \le i, j \le n} \right| = \mathcal{P}_{\{a_1, \dots, a_n\}, \{b_1, \dots, b_n\}, (E, F)} \tag{5.1}$$

*for some $E, F$. Then*

$$\begin{aligned} &\left| \left( |\Sigma_{A_i \uplus C_j \uplus \{a_i\}, B_i \uplus D_j \uplus \{b_j\}}| \right)_{1 \le i, j \le n} \right| \\ &\quad = \left( \prod_i \mathcal{P}_{A_i, B_i, (P_i, Q_i)} \right) \left( \prod_j \mathcal{P}_{C_j, D_j, (R_j, S_j)} \right) \mathcal{P}_{\{a_1, \dots, a_n\}, \{b_1, \dots, b_n\}, (E, F)}. \end{aligned}$$

**Remark 5.4.** Although the swapping property for $(A_i, B_i), (C_j, D_j), (\{a_i\}, \{b_j\})$ may appear quite restrictive, we will see that it is satisfied in a variety of cases. For example, in Proposition 5.8, we show that it allows one to recover the results from Section 3 that concern graphs with ancestral nodes. The sets used there are $(A_i = \mathrm{pa}(a_i), B_i = \mathrm{pa}(a_i)), (C_j = \emptyset, D_j = \emptyset), (\{a_i\}, \{b_j\})$, where $a_i \in \mathrm{pa}(v) \cup \{w\}$, $b_j \in \mathrm{pa}(v) \cup \{v\}$, $v$ satisfies conditions (i)–(iii) of Proposition 5.8, and $w \in V \setminus (\mathrm{pa}(v) \cup \mathrm{sib}(v) \cup \{v\})$. In addition, the swapping property is satisfied by $(A_i, B_i), (\{a_i\}, \{b_j\})$ whenever, for example, $A_i = \mathrm{pa}(a_i)$ and $B_i$ is a set of cardinality $|A_i|$ consisting of nondescendants of $\mathrm{sib}(a_i) \cup \{a_i\}$.

In general, the role of the swapping property is that it allows us to "get rid of" certain treks when studying the covariance structure. Specifically, it allows us to omit treks between $a_i$ and $b_j$ that use vertices or edges from treks between $A_i$ and $B_i$ or between $C_j$ and $D_j$.

**Proof.** Let $M$ be the matrix with $i, j$th entry equal to $|\Sigma_{A_i \uplus C_j \uplus \{a_i\}, B_i \uplus D_j \uplus \{b_j\}}|$ for $1 \leq i, j \leq n$. Since the entries in the $i$th row of $M$ are divisible by $\mathcal{P}_{A_i, B_i, (P_i, Q_i)}$, we can factor the determinant of $M$ as

$$|M| = \left(\prod_{i=1}^{n} \mathcal{P}_{A_i, B_i, (P_i, Q_i)}\right) \det\big((\mathcal{P}_{C_j, D_j, (R_j, S_j)} \mathcal{P}_{a_i, b_j, (E_{ij}, F_{ij})})_{1 \leq i, j \leq n}\big).$$

Since the $j$th column is divisible by $\mathcal{P}_{C_j, D_j, (R_j, S_j)}$, we can further factor as

$$\begin{aligned} |M| &= \left(\prod_{i} \mathcal{P}_{A_i, B_i, (P_i, Q_i)}\right)\left(\prod_{j} \mathcal{P}_{C_j, D_j, (R_j, S_j)}\right) \det(\mathcal{P}_{a_i, b_j, (E_{ij}, F_{ij})}) \\ &= \left(\prod_{i} \mathcal{P}_{A_i, B_i, (P_i, Q_i)}\right)\left(\prod_{j} \mathcal{P}_{C_j, D_j, (R_j, S_j)}\right) \mathcal{P}_{\{a_1, \dots, a_n\}, \{b_1, \dots, b_n\}, (E, F)} \end{aligned} \tag{5.2}$$

as required. □

Condition (5.1) deserves further discussion. By Theorem 4.4(ii),

$$\big|(\mathcal{P}_{a_i, b_j, (E, F)})_{i,j}\big| = \mathcal{P}_{\{a_1, \dots, a_n\}, \{b_1, \dots, b_n\}, (E, F)}$$

for every $a_1, \dots, a_n, b_1, \dots, b_n \in V$ and $E, F \subseteq V$. However, for this equality, it is not necessary to have the exact sets $E$ and $F$ for all matrix entries on the left-hand side. Instead, slightly different sets $E_{ij}$ and $F_{ij}$ can be used as the following lemma indicates.

**Lemma 5.5.** *Let $a_1, \dots, a_n, b_1, \dots, b_n \in V$ and $E, F \subseteq V$. Then*

$$\big|(\mathcal{P}_{a_i, b_j, (E_{ij}, F_{ij})})_{i,j}\big| = \mathcal{P}_{\{a_1, \dots, a_n\}, \{b_1, \dots, b_n\}, (E, F)} \tag{5.3}$$

*if*

$$E \setminus \{a_k : k \neq i\} \subseteq E_{ij} \subseteq E \cup \big(V \setminus \mathrm{an}(a_i)\big) \quad \text{and} \quad F \setminus \{b_\ell : \ell \neq j\} \subseteq F_{ij} \subseteq F \cup \big(V \setminus \mathrm{an}(b_j)\big).$$

The proof can be found in Appendix B.2. We remark that additional choices of $E_{ij}$ and $F_{ij}$ are certainly possible depending on the graph structure at hand. Nevertheless, this lemma gives us a wide variety of sets $E_{ij}$ and $F_{ij}$ for which equality (5.3) holds.

**Example 5.6.** Recall the Verma graph from Figure 4(b). In the context of Theorem 5.3 let

$$a_1 = 2, \qquad a_2 = 3, \qquad b_1 = 2, \qquad b_2 = 4,$$
$$C_1 = \{1\}, \qquad C_2 = \{1\}, \qquad D_1 = \{1\}, \qquad D_2 = \{3\}, \qquad A_1 = A_2 = B_1 = B_2 = \varnothing.$$

Then $(C_1, D_1), (\{a_i\}, \{b_1\})$ satisfy the swapping property for $i = 1, 2$. The same is true for $(C_2, D_2), (\{a_i\}, \{b_2\})$ for $i = 1, 2$. Moreover,

$$|\Sigma_{12,12}| = \mathcal{P}_{1,1}\mathcal{P}_{2,2,(\{234\},\{234\})}, \qquad |\Sigma_{12,34}| = \mathcal{P}_{1,3}\mathcal{P}_{2,4,(\{234\},\{24\})},$$
$$|\Sigma_{13,12}| = \mathcal{P}_{1,1}\mathcal{P}_{3,2,(\{234\},\{234\})}, \qquad |\Sigma_{13,34}| = \mathcal{P}_{1,3}\mathcal{P}_{3,4,(\{234\},\{24\})}.$$

By Lemma 5.5, we have that

$$\begin{vmatrix} \mathcal{P}_{2,2,(\{234\},\{234\})} & \mathcal{P}_{2,4,(\{234\},\{24\})} \\ \mathcal{P}_{3,2,(\{234\},\{234\})} & \mathcal{P}_{3,4,(\{234\},\{24\})} \end{vmatrix} = \mathcal{P}_{\{2,3\},\{2,4\},(\{234\},\{24\})}.$$

Therefore, the conditions of Theorem 5.3 are satisfied, and

$$f_{\text{Verma}} = \begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,34}| \\ |\Sigma_{13,12}| & |\Sigma_{13,34}| \end{vmatrix} = \mathcal{P}_{1,1}\mathcal{P}_{1,3}\mathcal{P}_{\{2,3\},\{2,4\},(\{234\},\{24\})}.$$

Now $\mathcal{P}_{\{2,3\},\{2,4\},(\{2,3,4\},\{2,4\})}$ is zero because $\{2, 3\}$ and $\{2, 4\}$ are $(\{234\}, \{2, 4\})$-restricted trek separated by $(\{2\}, \varnothing)$. That is, the nested determinant giving $f_{\text{Verma}}$ vanishes because treks between $\{2, 3\}$ and $\{2, 4\}$ that only use $\{2, 3, 4\}$ on the left and $\{2, 4\}$ on the right must all pass through 2 on the left.

The following corollary gives a combinatorial interpretation of the vanishing of a nested determinant like the ones specified in Theorem 5.3.

**Corollary 5.7.** *Suppose the conditions in Theorem 5.3 are satisfied. Define matrix entries* $M_{ij} = |\Sigma_{A_i \uplus C_j \uplus \{a_i\}, B_i \uplus D_j \uplus \{b_j\}}|$. *Then* $|M| = 0$ *if and only if at least one of the following holds*:

(i) *The sets* $\{a_1, \dots, a_n\}$ *and* $\{b_1, \dots, b_n\}$ *are* $(E, F)$*-restricted trek separated by some sets* $(X, Y)$ *with* $|X| + |Y| < n$; *or*
(ii) *For some* $i = 1, \dots, n$, *the sets* $A_i$ *and* $B_i$ *are* $(P_i, Q_i)$*-restricted trek separated by some sets* $(X, Y)$ *with* $|X| + |Y| < |A_i|$; *or*
(iii) *For some* $j = 1, \dots, n$, *the sets* $C_j$ *and* $D_j$ *are* $(R_j, S_j)$*-restricted trek separated by some sets* $(X, Y)$ *with* $|X| + |Y| < |C_j|$.

**Proof.** By Theorem 5.3, $|M|$ factors as in (5.2). Therefore, $|M| = 0$ if and only if one of the factors in this expression vanishes. By Theorem 4.4, each of these factors vanishes if and only if the corresponding restricted trek separation is satisfied. □

## 5.2. Restricted trek separation and ancestral vertices

We return to the parentally nested determinants and specifically the nested determinant $f_{ij}$ defined in (3.3). In Proposition 3.4, we gave conditions that entailed the vanishing of $f_{ij}$. We now see how this result is also implied by restricted trek separation.

**Proposition 5.8.** *Consider the conditions from Proposition 3.4, that is, $i$ is a vertex in $G = (V, \mathcal{D}, \mathcal{B})$ satisfying*:

(i) $\mathrm{pa}(i) \cap \mathrm{sib}(i) = \varnothing$,
(ii) *all vertices in* $\mathrm{pa}(i)$ *are ancestral and*
(iii) *the set $J$ of all ancestral vertices in $V \setminus (\mathrm{pa}(i) \cup \mathrm{sib}(i) \cup \{i\})$ is nonempty.*

*Then for every $j \in J$ the sets $\mathrm{pa}(i) \cup \{j\}$ and $\mathrm{pa}(i) \cup \{i\}$ are $(\mathrm{pa}(i) \cup \{j\}, V)$-restricted trek separated by $(\varnothing, \mathrm{pa}(v))$. This restricted trek separation implies that $f_{ij} \in \mathcal{I}(G)$ for all $j \in J$, that is, all parentally nested determinants for $(i, J)$ lie in $\mathcal{I}(G)$.*

The proof can be found in Appendix B.3.

**Example 5.9.** Again consider the Verma graph from Figure 4(b). We have that

$$f_{\mathrm{Verma}} = \begin{vmatrix} \sigma_{13} & \sigma_{14} \\ |\Sigma_{123,123}| & |\Sigma_{123,124}| \end{vmatrix}.$$

As mentioned in Section 3, $f_{\mathrm{Verma}} = f_{41}$. Indeed, $(\{3,1\}, \{3,4\})$ are $(\{3,1\}, V)$-restricted trek separated by $(\varnothing, \{3\})$.

**Example 5.10.**

Consider the ancestral graph from Figure 8(a), which was studied in more detail in Richardson and Spirtes [19]. Applying Proposition 5.8, we choose $i = 3$ and $j = 4$ to obtain that the corresponding polynomial $f_{34}$ vanishes. Indeed, by Lemma 5.2, we have that

$$\begin{aligned} f_{34} &= \begin{vmatrix} \sigma_{11} & \sigma_{13} \\ |\Sigma_{24,21}| & |\Sigma_{24,23}| \end{vmatrix} \\ &= \begin{vmatrix} \mathcal{P}_{1,1,(\{1,3,4\},\{1,2,3,4\})} & \mathcal{P}_{1,3,(\{1,3,4\},\{1,2,3,4\})} \\ \mathcal{P}_{2,2}\mathcal{P}_{4,1,(\{1,3,4\},\{1,2,3,4\})} & \mathcal{P}_{2,2}\mathcal{P}_{4,3,(\{1,3,4\},\{1,2,3,4\})} \end{vmatrix} = \mathcal{P}_{2,2}^2 \mathcal{P}_{\{1,4\},\{1,3\},(\{1,3,4\},\{1,2,3,4\})}. \end{aligned}$$

As all treks between $\{1,4\}$ and $\{1,3\}$ which avoid 2 on the left must intersect 1 on the right we have that $\mathcal{P}_{\{1,4\},\{1,3\},(\{1,3,4\},\{1,2,3,4\})} = 0$, and thus the above nested determinant is an element of the vanishing ideal for the graph. It can be checked by computational algebra that the above determinant generates the ideal of the model.

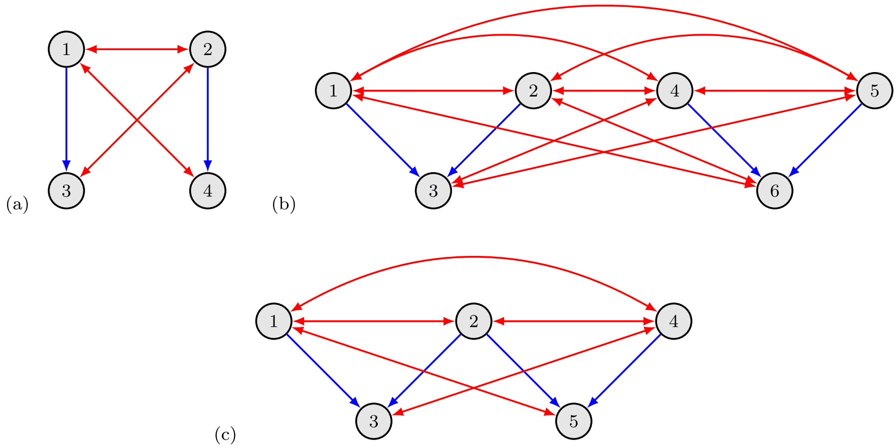


**Figure 8.** Ancestral graph examples.

**Example 5.11.** For the ancestral graph in Figure 8(b), we choose $i = 6$, $j = 3$. Then, using Lemma 5.2, we have that

$$f_{63} = \begin{vmatrix} |\Sigma_{123,124}| & |\Sigma_{123,125}| & |\Sigma_{123,126}| \\ \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \sigma_{54} & \sigma_{55} & \sigma_{56} \end{vmatrix} = \mathcal{P}_{\{1,2\},\{1,2\}} \mathcal{P}_{\{3,4,5\},\{4,5,6\},(\{3,4,5,6\},\{3,4,5,6\})}.$$

As all treks between $\{3, 4, 5\}$ and $\{4, 5, 6\}$ which are restricted to only use $\{3, 4, 5, 6\}$ on their left or right sides must use 4 or 5 on their right side it follows that

$$\mathcal{P}_{\{3,4,5\},\{4,5,6\},(\{3,4,5,6\},\{3,4,5,6\})} = 0,$$

and thus the above determinant is an element of the vanishing ideal.

**Example 5.12.** Consider the graph from Figure 8(c). Lemma 5.2 implies that

$$f_{53} = \begin{vmatrix} |\Sigma_{123,122}| & |\Sigma_{123,124}| & |\Sigma_{123,125}| \\ \sigma_{22} & \sigma_{24} & \sigma_{25} \\ \sigma_{42} & \sigma_{44} & \sigma_{45} \end{vmatrix} = \mathcal{P}_{\{1,2\},\{1,2\}} \mathcal{P}_{\{3,2,4\},\{2,4,5\},(\{2,3,4,5\},\{1,2,3,4,5\})}.$$

As any trek from $\{3, 2, 4\}$ to $\{2, 4, 5\}$ which avoids 1 on the left must use 2 or 4 on the right it follows that $\mathcal{P}_{\{3,2,4\},\{2,4,5\},(\{2,3,4,5\},\{1,2,3,4,5\})} = 0$. Hence $f_{53}$ is in (and generates) the vanishing ideal.

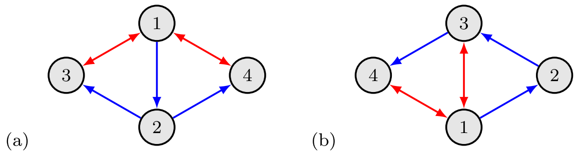


**Figure 9.** Two four-node graphs whose vanishing ideals are known from computational algebra. We are able to write the generators of these vanishing ideals as nested determinant but our combinatorial conditions do not appear to apply.

# 6. Beyond swapping: Recursive nesting, directed cycles and the pentad

In this section, we explore examples that are not covered by Theorem 5.3 and Corollary 5.7 but whose constraints are still nested determinants. In Section 6.1, we consider two mixed graphs for which the constraints could be presented as nested determinants but for which – we argue – a recursive nesting of determinants is more natural and more directly tied to restricted trek separation. In Section 6.2, we discuss an example of a directed graph with a directed cycle for which restricted trek separation also implies a nested determinant constraint. Finally, in Section 6.3, we turn to the pentad from factor analysis (Drton, Sturmfels and Sullivant [7]), and show that it is also defined by a nested determinant.

## 6.1. Restricted trek separation and determinants of recursively nested matrices

It is apparent from Theorem 5.3 and Corollary 5.7 that (singly) nested determinants give a way to express restricted trek systems as factors as follows. Recall the Verma graph and Example 5.6 where we saw that

$$f_{\text{Verma}} = \begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,34}| \\ |\Sigma_{13,12}| & |\Sigma_{13,34}| \end{vmatrix} = \begin{vmatrix} \mathcal{P}_{1,1}\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})} & \mathcal{P}_{1,3}\mathcal{P}_{2,4,(\{2,3,4\},\{2,4\})} \\ \mathcal{P}_{1,1}\mathcal{P}_{3,2,(\{2,3,4\},\{2,3,4\})} & \mathcal{P}_{1,3}\mathcal{P}_{3,4,(\{2,3,4\},\{2,4\})} \end{vmatrix}$$
$$= \mathcal{P}_{1,1}\mathcal{P}_{1,3}\mathcal{P}_{\{2,3\},\{2,4\},(\{2,3,4\},\{2,4\})}.$$

Notice that each of the entries in the $2 \times 2$ matrix whose determinant equals $f_{\text{Verma}}$ is a determinant, which, because of the swapping property, factors out $\mathcal{P}_{1,1}$ or $\mathcal{P}_{1,3}$, and leaves monomials corresponding to restricted trek systems that only use $\{2, 3, 4\}$ on the left and $\{2, 4\}$ on the right.

In other graphs, using single subdeterminants of $\Sigma$ is not enough to factor out restricted trek systems. However, one can instead use recursively nested determinants. We illustrate this in the following two examples.

**Example 6.1.** As is shown by van Ommen and Mooij [26] in their Appendix B, the graph from Figure 9(a) has vanishing ideal generated by

$$f = p_0^2\sigma_{34} + p_0\sigma_{23}p_2 + p_1\sigma_{24}p_0 + p_1\sigma_{22}p_2,$$

where

$$p_0 = |\Sigma_{12,12}|, \qquad p_1 = |\Sigma_{13,21}|, \qquad p_2 = |\Sigma_{21,14}|.$$

One may express $f$ as a nested determinant by noting that

$$-f = \begin{vmatrix} 0 & |\Sigma_{12,12}| & |\Sigma_{12,14}| \\ |\Sigma_{12,12}| & \sigma_{22} & \sigma_{24} \\ |\Sigma_{13,12}| & \sigma_{32} & \sigma_{34} \end{vmatrix} = \begin{vmatrix} |\Sigma_{112,112}| & |\Sigma_{12,12}| & |\Sigma_{12,14}| \\ |\Sigma_{12,12}| & \sigma_{22} & \sigma_{24} \\ |\Sigma_{13,12}| & \sigma_{32} & \sigma_{34} \end{vmatrix}.$$

While the above representation of $f$ suggests applying Theorem 5.3 with

$$\begin{aligned} &a_1 = 2, \qquad a_2 = 2, \qquad a_3 = 3, \qquad b_1 = 2, \qquad b_2 = 2, \qquad b_3 = 4, \\ &A_1 = \{1\}, \qquad B_1 = \{1\}, \qquad C_1 = \{1\}, \qquad D_1 = \{1\}, \\ &A_i = B_i = C_i = D_i = \varnothing \quad (i = 2, 3), \end{aligned}$$

this, unfortunately, does not seem to satisfy the conditions of the theorem. On the other hand, we can express $f$ as the determinant of a matrix whose entries are themselves nested determinants, namely,

$$f = \begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,14}| \\ \begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,13}| \\ \Sigma_{22} & \Sigma_{23} \end{vmatrix} & \begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,13}| \\ \Sigma_{42} & \Sigma_{43} \end{vmatrix} \end{vmatrix}.$$

Moreover, using Lemma 5.2, we get the following factorizations:

$$\begin{aligned} &|\Sigma_{12,12}| = \mathcal{P}_{1,1}\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})}, \qquad |\Sigma_{12,13}| = \mathcal{P}_{1,1}\mathcal{P}_{2,3,(\{2,3,4\},\{2,3,4\})}, \\ &|\Sigma_{12,14}| = \mathcal{P}_{1,1}\mathcal{P}_{2,4,(\{2,3,4\},\{2,3,4\})}. \end{aligned}$$

Thus,

$$\begin{aligned} &\begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,13}| \\ \Sigma_{22} & \Sigma_{23} \end{vmatrix} = \mathcal{P}_{1,1}\omega_{2,2}\lambda_{12}\omega_{1,3} = \mathcal{P}_{1,1}\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})}\mathcal{P}_{3,2,(\{3,4\},\{12,3,4\})}, \\ &\begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,13}| \\ \Sigma_{42} & \Sigma_{43} \end{vmatrix} = \mathcal{P}_{1,1}\omega_{2,2}\lambda_{1,2}\omega_{1,3}\lambda_{2,4} = \mathcal{P}_{1,1}\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})}\mathcal{P}_{3,4,(\{3,4\},\{1,2,3,4\})}. \end{aligned}$$

This implies that

$$\begin{aligned} f &= \mathcal{P}_{1,1}^2\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})} \begin{vmatrix} P_{2,2,(\{2,3,4\},\{2,3,4\})} & \mathcal{P}_{2,4,(\{2,3,4\},\{2,3,4\})} \\ \mathcal{P}_{3,2,(\{34\},\{1,2,3,4\})} & \mathcal{P}_{3,4,(\{34\},\{1,2,3,4\})} \end{vmatrix} \\ &= \mathcal{P}_{1,1}^2\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})}\mathcal{P}_{\{2,3\},\{2,4\},(\{2,3,4\},\{1,2,3,4\})}, \end{aligned}$$

where we have used Lemma 5.5 in the second step. Indeed, the sets $\{2,3\}$ and $\{2,4\}$ are $(\{2,3,4\},\{1,2,3,4\})$-restricted trek-separated by $(\varnothing,\{2\})$.

**Example 6.2.** Now consider the graph in Figure 9(b). Its vanishing ideal is generated by the polynomial

$$f = \begin{vmatrix} |\Sigma_{112,112}| & |\Sigma_{12,13}| & |\Sigma_{12,12}| \\ |\Sigma_{13,12}| & \sigma_{33} & \sigma_{32} \\ |\Sigma_{14,12}| & \sigma_{43} & \sigma_{42} \end{vmatrix}.$$

While the above representation of $f$ suggests applying Theorem 5.3, this, unfortunately, does not seem to satisfy the conditions of the theorem either. On the other hand, we can express $f$ as the determinant of a matrix whose entries are themselves nested determinants:

$$f = \begin{vmatrix} |\Sigma_{12,13}| & |\Sigma_{12,14}| \\ \begin{vmatrix} \Sigma_{2,3} & \Sigma_{3,3} \\ |\Sigma_{12,12}| & |\Sigma_{13,12}| \end{vmatrix} & \begin{vmatrix} \Sigma_{2,4} & \Sigma_{3,4} \\ |\Sigma_{12,12}| & |\Sigma_{13,12}| \end{vmatrix} \end{vmatrix}.$$

Moreover, by Lemma 5.2,

$$|\Sigma_{12,12}| = \mathcal{P}_{1,1}\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})}, \qquad |\Sigma_{12,13}| = \mathcal{P}_{1,1}\mathcal{P}_{2,3,(\{2,3,4\},\{2,3,4\})},$$
$$|\Sigma_{12,14}| = \mathcal{P}_{1,1}\mathcal{P}_{2,4,(\{2,3,4\},\{2,3,4\})}.$$

Consequently,

$$\begin{vmatrix} \Sigma_{2,3} & \Sigma_{3,3} \\ |\Sigma_{12,12}| & |\Sigma_{13,12}| \end{vmatrix} = \mathcal{P}_{1,1}\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})}\mathcal{P}_{3,3,(\{3,4\},\{3,4\})},$$
$$\begin{vmatrix} \Sigma_{2,4} & \Sigma_{3,4} \\ |\Sigma_{12,12}| & |\Sigma_{13,12}| \end{vmatrix} = \mathcal{P}_{1,1}\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})}\mathcal{P}_{3,4,(\{3,4\},\{3,4\})}.$$

Thus, we can write the full determinant as

$$f = \mathcal{P}_{1,1}^2\mathcal{P}_{2,2,(\{2,3,4\},\{2,3,4\})} \begin{vmatrix} \mathcal{P}_{2,3,(\{2,3,4\},\{2,3,4\})} & \mathcal{P}_{2,4,(\{2,3,4\},\{2,3,4\})} \\ \mathcal{P}_{3,3,(\{3,4\},\{3,4\})} & \mathcal{P}_{3,4,(\{3,4\},\{3,4\})} \end{vmatrix}.$$

By Lemma 5.5,

$$f = \mathcal{P}_{1,1}^2\mathcal{P}_{2,2,(2,3,4,2,3,4)}\mathcal{P}_{\{2,3\},\{3,4\},(\{2,3,4\},\{2,3,4\})}.$$

The last term suggests that $\{2,3\}$ and $\{3,4\}$ are $(\{2,3,4\},\{2,3,4\})$-restricted trek separated. Indeed, this is the case, and they are separated by $(\varnothing, \{3\})$.

**Remark 6.3.** Appendix B of van Ommen and Mooij [26] explicitly lists the (minimal) generators of all vanishing ideals of acyclic mixed graphs on 4 nodes. Of these graphs, only those from Figure 9 cannot be immediately recognized as being determinantal constraints on the covariance matrix. From what we have shown above, we now see that all generators of vanishing ideals of acyclic mixed graphs on four nodes can be written as nested determinants with, at most, a single level of nesting (i.e., as determinants of determinants), and moreover can be explained via restricted trek separation.

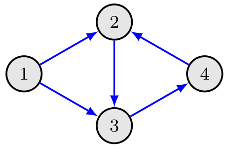

**Figure 10.** A cyclic graph whose model is defined by a nested determinant.

As we also record in Section 7, we believe that restricted trek separation can always be formed as a factor of a vanishing recursively nested determinant. Moreover, we deem it possible that such determinants define all acyclic linear structural equation models. We defer further exploration of these questions to a future study.

## 6.2. Graphs with cycles

Although we believe our results from the previous two sections can be extended to graphs containing cycles, the situation there is a bit more complicated. Even extending Theorem 4.1 to the cyclic case is not a simple task. It was accomplished (along with other results) in a separate article (Draisma, Sullivant and Talaska [2]).

The following model, although it has cycles, is defined by the vanishing of a nested determinant, which can be explained by restricted trek separation. However, we wish to point out that a more sophisticated example might need the definition of further notions, like those that appear in Draisma, Sullivant and Talaska [2].

**Example 6.4.** Consider the graph $G$ in Figure 10, which was treated in Drton [3] where a degree 6 polynomial $f$ generating $\mathcal{I}(G)$ was displayed. This polynomial can be written as the following doubly nested determinant:

$$f = \begin{vmatrix} |\Sigma_{34,12}| & |\Sigma_{34,13}| \\ \begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,34}| \\ |\Sigma_{14,12}| & |\Sigma_{14,34}| \end{vmatrix} & \begin{vmatrix} |\Sigma_{12,13}| & |\Sigma_{12,34}| \\ |\Sigma_{14,13}| & |\Sigma_{14,34}| \end{vmatrix} \end{vmatrix}.$$

We will now show that the vanishing of this determinant corresponds to the fact that $\{4, 2\}$ and $\{2, 3\}$ are $(\{1, 2, 4\}, \{2, 3, 4\})$-restricted-trek separated by $(\varnothing, \{2\})$. For our derivations, we use results from Draisma, Sullivant and Talaska [2], where subdeterminants of $\Sigma$ corresponding to graphs with cycles are given by rational expressions. The entries of the above matrix are:

$$|\Sigma_{34,12}| = \frac{(\lambda_{13} + \lambda_{12}\lambda_{23})\omega_{11}\omega_{44}\lambda_{42}}{(1 - \lambda_{23}\lambda_{34}\lambda_{42})^2} = \frac{\mathcal{P}_{3,1,(\{1,2,3\},\{1,2,3\})}\mathcal{P}_{4,2,(\{1,2,4\},\{2,3,4\})}}{(1 - \lambda_{23}\lambda_{34}\lambda_{42})},$$

$$|\Sigma_{34,13}| = \frac{(\lambda_{13} + \lambda_{12}\lambda_{23})\omega_{11}\omega_{44}\lambda_{42}\lambda_{23}}{(1 - \lambda_{23}\lambda_{34}\lambda_{42})^2} = \frac{\mathcal{P}_{3,1,(\{1,2,3\},\{1,2,3\})}\mathcal{P}_{4,3,(\{1,2,4\},\{2,3,4\})}}{(1 - \lambda_{23}\lambda_{34}\lambda_{42})},$$

$$\begin{vmatrix} |\Sigma_{12,12}| & |\Sigma_{12,34}| \\ |\Sigma_{14,12}| & |\Sigma_{14,34}| \end{vmatrix} = \Sigma_{14,34}\omega_{11}\omega_{22}\frac{1}{1 - \lambda_{23}\lambda_{34}\lambda_{42}} = \Sigma_{14,34}\omega_{11}\mathcal{P}_{2,2,(\{1,2\},\{2,3,4\})},$$

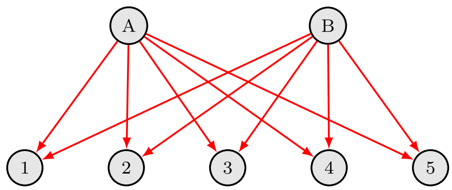


**Figure 11.** The factor analysis model on five nodes with two factors. The vanishing ideal of this model is generated by one degree five polynomial.

$$\begin{vmatrix} |\Sigma_{12,13}| & |\Sigma_{12,34}| \\ |\Sigma_{14,13}| & |\Sigma_{14,34}| \end{vmatrix} = \Sigma_{14,34}\omega_{11}\omega_{22}\lambda_{23}\frac{1}{1-\lambda_{23}\lambda_{34}\lambda_{42}} = \Sigma_{14,34}\omega_{11}\mathcal{P}_{2,3,(\{1,2\},\{2,3,4\})}.$$

It follows that

$$\begin{aligned} f &= \Sigma_{14,34}\omega_{11}\mathcal{P}_{3,1,(\{1,2,3\},\{1,2,3\})}\begin{vmatrix} \mathcal{P}_{4,2,(\{1,2,4\},\{2,3,4\})} & \mathcal{P}_{4,3,(\{1,2,4\},\{2,3,4\})} \\ \mathcal{P}_{2,2,(\{1,2\},\{2,3,4\})} & \mathcal{P}_{2,3,(\{1,2\},\{2,3,4\})} \end{vmatrix} \\ &= \frac{\Sigma_{14,34}\omega_{11}\mathcal{P}_{3,1,(\{1,2,3\},\{1,2,3\})}\mathcal{P}_{\{4,2\},\{2,3\},(\{1,2,4\},\{2,3,4\})}}{(1-\lambda_{23}\lambda_{34}\lambda_{42})}, \end{aligned}$$

where the last equality follows by Lemma 5.5. The last term in the numerator vanishes due to the above mentioned restricted trek separation. We remark that the extra term $(1-\lambda_{23}\lambda_{34}\lambda_{42})$ in the denominator cannot be obtained via the formula given in Theorem 5.3 (even with the usage of geometric series).

## 6.3. Nested determinants with no restricted trek separation

We have not found any examples of acyclic mixed graphs $G$ for which defining equations of the model $\mathcal{M}(G)$ cannot be explained using restricted trek separation. However, there are other, closely related models, for which restricted trek separation does not seem to provide the same combinatorial explanation.

**Example 6.5 (The pentad).** Consider a factor analysis model with five normally distributed observed variables and two latent factors as in Figure 11. Its defining equation is a degree 5 polynomial in the covariance matrix entries:

$$\begin{aligned} f_{\text{pentad}} = {} & \sigma_{12}\sigma_{13}\sigma_{24}\sigma_{35}\sigma_{45} - \sigma_{12}\sigma_{13}\sigma_{25}\sigma_{34}\sigma_{45} - \sigma_{12}\sigma_{14}\sigma_{23}\sigma_{35}\sigma_{45} + \sigma_{12}\sigma_{14}\sigma_{25}\sigma_{34}\sigma_{35} \\ & + \sigma_{12}\sigma_{15}\sigma_{23}\sigma_{34}\sigma_{45} - \sigma_{12}\sigma_{15}\sigma_{24}\sigma_{34}\sigma_{35} + \sigma_{13}\sigma_{14}\sigma_{23}\sigma_{25}\sigma_{45} - \sigma_{13}\sigma_{14}\sigma_{24}\sigma_{25}\sigma_{35} \\ & - \sigma_{13}\sigma_{15}\sigma_{23}\sigma_{24}\sigma_{45} + \sigma_{13}\sigma_{15}\sigma_{24}\sigma_{25}\sigma_{34} - \sigma_{14}\sigma_{15}\sigma_{23}\sigma_{25}\sigma_{34} + \sigma_{14}\sigma_{15}\sigma_{23}\sigma_{24}\sigma_{35}. \end{aligned}$$

This polynomial can be expressed in nested determinantal form as

$$f_{\text{pentad}} = \begin{vmatrix} |\Sigma_{23,45}| & |\Sigma_{25,34}| \\ |\Sigma_{123,145}| & |\Sigma_{125,134}| \end{vmatrix}.$$

Combinatorially, we can see that all trek systems stemming from the second row of the matrix are in one-to-one correspondence with the trek systems from the first row of the matrix, and are obtained by just adding the trek $1 - 1$. However, we have not been able to interpret this nested determinant via restricted trek separation. Note that the mixed graph obtained by latent projection would be a complete graph with a bidirected edge between each $i, j \in \{1, \dots, 5\}$.

# 7. Discussion

We conclude by giving a brief review of the results presented in this paper and then discussing problems for future work.

*Contributions*

This paper demonstrates the importance of nested determinants as constraints on covariance matrices in graphical causal/structural equation models associated to mixed graphs. Nested determinants are determinants of matrices whose entries are determinants themselves. Theorem 3.8 shows that a special class of parentally nested determinants is sufficient for a semialgebraic description of a class of models that is slightly more general than the class of ancestral graph models. Theorem 5.3 provides a framework for explaining the vanishing of more general nested determinants via trek separation under restrictions on the vertices that treks may visit on their left and their right sides.

The examples from Section 6 depict graphs for which the conditions of Theorem 5.3 do not apply. While it is often possible to present the defining equations of such models in terms of (singly) nested determinants, we suggest to instead view the equations as recursively nested determinants. In other words, we consider determinants of smaller matrices whose entries are (recursively) nested determinants. As we exemplified, such recursively nested determinants may admit an explanation by restricted trek separation. We further exhibit an example of a graph with a cycle in which the model is also described by a recursively nested determinant that admits a restricted trek separation interpretation.

*Definition of nested and recursively nested determinants*

Theorem 5.3 is concerned with a particular type of nested determinants where rows and columns of the considered matrix correspond to vertices of the graph/the given random variables. This setup contains as a special case the parentally nested determinants from Section 3. We anticipate that the nested determinants considered in Theorem 5.3 are sufficiently general to describe mixed graph models as long as we allow for a suitable notion of recursive nesting as encountered in the Examples in Section 6.

In a general definition of recursively nested determinants, the subdeterminants of the original covariance matrix would be recursively nested determinants with depth of recursion zero. At depth $k$, we would take determinants of matrices whose entries are recursively nested determinants of depth at most $k - 1$. However, it would be desirable to constrain this construction such that for any recursively nested determinant the rows and columns of the considered matrix can be put in correspondence with two sets of vertices. These sets of vertices may then admit a restricted trek separation.

**Problem 7.1.** Develop a notion of recursively nested determinants for which row and column indices are in correspondence with graph vertices. The depth of recursion should be such that the subdeterminants of the original matrix are the only recursively nested determinants of depth 0. The recursively nested determinants of depth 1 should be of the type encountered in Theorem 5.3.

*Tian decomposition*

In the Introduction, after Example 1.2, we mentioned Tian's graph decomposition, which may yield subgraphs whose covariance matrix can be rationally identified from the covariance matrix for the original graph $G$. Trek separation in the subgraph then gives a rational constraint. Clearing denominators yields a polynomial in $\mathcal{I}(G)$.

**Conjecture 7.2.** Trek separation relations in subgraphs obtained from Tian's graph decomposition correspond to recursively nested determinants.

*Vanishing nested determinants*

The results we have given so far are sufficient conditions for the vanishing of nested determinants.

**Problem 7.3.** Using restricted trek separation, obtain graphical conditions that are necessary for the vanishing of the nested determinants from Theorem 5.3.

If a characterization of the vanishing of nested determinants is established, it can be used to decide model equivalence questions. More generally, it would be desirable to obtain conditions, sufficient and necessary, for the vanishing of recursively nested determinants. We formulate a "hopeful" conjecture for acyclic mixed graphs.

**Conjecture 7.4.** The equality of two models $\mathcal{M}(G)$ and $\mathcal{M}(H)$ can be decided by comparing restricted trek separation relations in $G$ and $H$.

In all examples of graphs $G$ we inspected, the vanishing ideal $\mathcal{I}(G)$ is in fact generated by nested or recursively nested determinants.

**Conjecture 7.5.** The vanishing ideal $\mathcal{I}(G)$ can always be generated by recursively nested determinants.

An important step in the study of nested determinants would be characterizing when two nested determinants are equal (recall, for example, the Verma graph from Example 1.4 and equation (1.5)).

**Problem 7.6.** Give conditions on when two nested determinants are equal. Describe the equivalence class of different representations of an equation as a nested determinant.

*Computing restricted trek separation*

Assuming that restricted trek separation is what characterizes equivalence classes of models, as suggested by Conjecture 7.4, we may need to either use the graphical conditions from Problem 7.3 or to be able to compute restricted trek separation in order to find equivalent graphs.

**Problem 7.7.** Design computationally efficient algorithms for checking/finding restricted trek separations.

*Feedback cycles*

Our focus was on acyclic mixed graphs, for which determinants of the covariance matrix have expansions in terms of systems of treks without sided intersection. However, as the example of Figure 10 shows, (recursively) nested determinants are also relevant for cyclic graphs.

**Problem 7.8.** Generalize Theorem 5.3 to the general possibly cyclic case.

In addition, all problems mentioned above also pertain to graphs with cycles.

# Appendix A: Proofs for Section 4: Restricted trek separation

This section is devoted to proving Theorem 4.4. The proof proceeds through rather minor modifications of the ideas of Sullivant, Talaska and Draisma [24].

## A.1. Proof of Theorem 4.4(i) for directed acyclic graphs

We begin by proving Theorem 4.4 in the case when $G$ is a directed acyclic graph (DAG). We extend it to acyclic mixed graphs in the next section. We first record the following combinatorial interpretation of the entries of $(I - \Lambda_{C,C})^{-1}$ for a subset of vertices $C$.

**Proposition A.1.** *Let $\mathcal{P}(i, j, C)$ be the set of directed paths from $i \in C$ to $j \in C$ that only use vertices from a subset $C \subseteq V$ in the directed graph $G$. For each path $P$, define $\lambda^P = \prod_{i\to j\in P} \lambda_{ij}$. Then*

$$\big[(I - \Lambda_{C,C})^{-1}\big]_{ij} = \sum_{P\in\mathcal{P}(i,j,C)} \lambda^P.$$

**Proof.** The claim follows from Proposition 3.1 in Sullivant, Talaska and Draisma [24] if we consider the induced subgraph of $G$ with vertex set $C$. □

When $G$ is a directed graph, the error covariance matrix $\Omega$ is diagonal. This allows us to show the following lemma. We emphasize that in our discussion a determinant is zero if it is identically zero as a polynomial/function.

**Lemma A.2.** *In a directed graph consider sets of vertices $A \subseteq P$, $B \subseteq Q$ with $|A| = |B|$. Then* $\det \Sigma^{(P,Q)}_{A,B} = 0$ *if and only if for every set $S \subseteq P \cap Q$ with $|S| = |A| = |B|$ either* $\det(((I - \Lambda)_{P,P})^{-1})_{S,A} = 0$ *or* $\det(((I - \Lambda)_{Q,Q})^{-1})_{S,B} = 0$.

**Proof.** Since $\Sigma^{(P,Q)} = ((I - \Lambda)_{P,P})^{-T} \Omega_{P,Q} ((I - \Lambda)_{Q,Q})^{-1}$, we have

$$\Sigma^{(P,Q)}_{A,B} = \bigl(((I - \Lambda)_{P,P})^{-T}\bigr)_{A,P} \Omega_{P,Q} \bigl(((I - \Lambda)_{Q,Q})^{-1}\bigr)_{Q,B}.$$

By the Cauchy–Binet theorem,

$$\det \Sigma^{(P,Q)}_{A,B} = \sum_{S \subseteq P, R \subseteq Q} \det\bigl(((I - \Lambda)_{P,P})^{-T}\bigr)_{A,S} \det(\Omega_{S,R}) \det\bigl(((I - \Lambda)_{Q,Q})^{-1}\bigr)_{R,B},$$

where the sum runs over $S$ and $R$ of cardinality $|A| = |B|$. As $\Omega$ is diagonal, we obtain that

$$\begin{aligned} \det \Sigma^{(P,Q)}_{A,B} &= \sum_{S \subseteq P \cap Q} \det\bigl(((I - \Lambda)_{P,P})^{-T}\bigr)_{A,S} \det(\Omega_{S,S}) \det\bigl(((I - \Lambda)_{Q,Q})^{-1}\bigr)_{S,B} \\ &= \sum_{S \subseteq P \cap Q} \det\bigl(((I - \Lambda)_{P,P})^{-1}\bigr)_{S,A} \det\bigl(((I - \Lambda)_{Q,Q})^{-1}\bigr)_{S,B} \prod_{s \in S} \omega_{s,s}. \end{aligned}$$

Since each monomial $\prod_{s \in S} \omega_{s,s}$ appears only in one term in this expansion, the result follows. □

We now recall the Gessel–Viennot–Lindström lemma.

**Lemma A.3 (Gessel–Viennot–Lindström lemma).** *Suppose $G$ is a DAG with vertex set $\{1, \dots, m\}$. Let $A, B \subseteq \{1, \dots, m\}$ be such that $|A| = |B| = \ell$. Then*

$$\det\bigl((I - \Lambda)^{-1}\bigr)_{A,B} = \sum_{S \in \mathcal{N}(A,B)} (-1)^S \lambda^S,$$

*where $\mathcal{N}(A, B)$ is the set of all nonintersecting systems of $\ell$ directed paths in $G$ from $A$ to $B$, and $(-1)^S$ is the sign of the induced permutation of elements from $A$ to $B$. In particular,* $\det((I - \Lambda)^{-1})_{A,B} = 0$ *if and only if every system of $\ell$ directed paths from $A$ to $B$ has two paths which share a vertex.*

We are going to use this lemma by restricting the original directed acyclic graph $G$ to the induced subgraphs on the subsets $P$ and $Q$. The lemma applies to all these subgraphs because they themselves are directed acyclic graphs.

Let $A \subseteq P$, $B \subseteq Q$ with $|A| = |B| = \ell$. Consider a system $\mathcal{T} = \{\tau_1, \dots, \tau_\ell\}$ of $\ell$ $(P, Q)$-restricted treks from $A \subseteq P$ to $B \subseteq Q$, connecting the $\ell$ distinct vertices in $A$ to the $\ell$ distinct vertices in $B$. Let $\mathrm{top}(\mathcal{T})$ denote the multiset $\{\mathrm{top}(\tau_1,), \dots, \mathrm{top}(\tau_\ell)\}$. Here $\mathrm{top}(\tau)$ is the unique source of the trek $\tau$, that is, the vertex contained in both the left side and the right side of the trek. Note that the trek system $\mathcal{T}$ consists of two systems of directed paths, a path system $S_A$ from $\mathrm{top}(\mathcal{T})$ to $A$ which only uses vertices in $P$, and a path system $S_B$ from $\mathrm{top}(\mathcal{T})$ to $B$ which only uses vertices in $Q$. We say that $\mathcal{T}$ has a *sided intersection* if two paths in $S_A$ share a vertex or if two paths in $S_B$ share a vertex.

**Proposition A.4.** *In a DAG consider sets of vertices $A \subseteq P$ and $B \subseteq Q$ with $|A| = |B|$. Then,*

$$\det\bigl(\Sigma_{A,B}^{(P,Q)}\bigr) = 0$$

*if and only if every system of (simple) $(P, Q)$-restricted treks from $A$ to $B$ has a sided intersection.*

**Proof.** Suppose that $\det(\Sigma_{A,B}^{(P,Q)}) = 0$, and let $\mathcal{T}$ be a $(P, Q)$-restricted trek system from $A$ to $B$. If all elements of the multiset $\mathrm{top}(\mathcal{T})$ are distinct, then Lemma A.2 implies that either $\det(((I - \Lambda)_{P,P})^{-1})_{\mathrm{top}(\mathcal{T}),A} = 0$ or $\det(((I - \Lambda)_{Q,Q})^{-1})_{\mathrm{top}(\mathcal{T}),B} = 0$. If $\mathrm{top}(\mathcal{T})$ has repeated elements, then these determinants are also zero since there are repeated rows. Thus, in both cases, Lemma A.3 implies that there is an intersection in the path system from $\mathrm{top}(\mathcal{T})$ to $A$ or in the path system from $\mathrm{top}(\mathcal{T})$ to $B$. Hence, $\mathcal{T}$ has a sided intersection.

Conversely, suppose that every $(P, Q)$-restricted trek system from $A$ to $B$ has a sided intersection, and let $S \subseteq P \cap Q$. If $R = \mathrm{top}(\mathcal{T})$ for some $(P, Q)$-restricted trek system $\mathcal{T}$ from $A$ to $B$, then either the path system from $\mathrm{top}(\mathcal{T})$ to $A$ or the path system from $\mathrm{top}(\mathcal{T})$ to $B$ has an intersection. If $R$ is not the set of top elements for some $(P, Q)$-restricted trek system $\mathcal{T}$ from $A$ to $B$, then there is no $P$-restricted path system connecting $R$ to $A$ or there is no $Q$-restricted path system from $R$ to $B$. In both cases, Lemma A.3 implies that either $\det(((I - \Lambda)_{P,P})^{-1})_{R,A} = 0$ or $\det(((I - \Lambda)_{Q,Q})^{-1})_{R,B} = 0$. Then, Lemma A.2 implies that $\det(\Sigma_{A,B}^{(P,Q)}) = 0$.

Note that it is sufficient to check the systems of simple treks only. Here, simple indicates that a trek has no repeated vertices. □

We now define a new DAG associated to $G$, denoted $\tilde{G}_{P,Q}$ in order to be able to invoke the Max-Flow–Min-Cut Theorem (see Theorem A.6). Let $P' = \{i' : i \in P\}$ be a set of new vertices, each being the copy of a corresponding vertex in $P$. The vertex set of graph $\tilde{G}_{P,Q}$ is $P' \cup Q$. The edge set of $\tilde{G}_{P,Q}$ includes the edge $i \to j$ for all $i, j \in Q$ such that $i \to j$ is an edge in $G$. Moreover, it includes the edge $j' \to i'$ for all $i, j \in P$ such that $i \to j$ is an edge in $G$, and the edge $i' \to i$ for all $i \in P \cap Q$.

**Proposition A.5.** *The $(P, Q)$-restricted treks in G from $i \in P$ to $j \in Q$ are in bijective correspondence with directed paths from $i'$ to $j$ in $\tilde{G}_{P,Q}$. Simple $(P, Q)$-restricted treks in G from $i$ to $j$ are in bijective correspondence with directed paths from $i'$ to $j$ in $\tilde{G}_{P,Q}$ that use at most one edge from any pair $a \to b$ and $b' \to c'$ where $a, b \in Q$, $b, c \in P$.*

**Proof.** Every trek from $i$ to $j$ is the union of two paths with a common top, the left path in $P$, the right path in $Q$. The part of the trek from the top to $i$ corresponds to the subpath with only vertices in $P'$, and the part of the trek from the top to $j$ corresponds to the subpath with only vertices in $Q$. The unique edge of the form $k' \to k$ corresponds to the top of the trek. Excluding $a \to b$ and $b' \to c'$ implies that a trek never visits the same vertex $b$ twice. □

Menger's theorem, also known as the Max-Flow–Min-Cut theorem, now allows us to turn the sided crossing result on $G$ into a blocking characterization on $\tilde{G}_{P,Q}$.

**Theorem A.6 (Vertex version of Menger's theorem).** *The cardinality of the largest set of vertex disjoint directed paths between two nonadjacent vertices $u$ and $v$ in a DAG is equal to the cardinality of the smallest blocking set, where a blocking set is a set of vertices whose removal from the graph ensures there is no directed path from $u$ to $v$.*

**Proof of Theorem 4.4 for DAGs.** We first focus on the case where $\det \Sigma_{A,B}^{(P,Q)} = 0$ so that the rank is at most $k-1$, where $k = |A| = |B|$. According to Proposition A.4, every system of $k$ $(P, Q)$-restricted treks from $A$ to $B$ must have a sided intersection. That is, the number of vertex disjoint paths from $A'$ to $B$ is at most $k-1$ in the graph $\tilde{G}_{P,Q}$. We add two new vertices to $\tilde{G}_{P,Q}$, one vertex $u$ that points to each vertex in $A'$ and one vertex $v$ that each vertex in $B$ points to $v$. Thus, there are at most $k-1$ vertex disjoint paths from $u$ to $v$. Applying Menger's theorem, there is a blocking set $W$ in $\tilde{G}_{P,Q}$ of cardinality $|W| \leq k-1$. Set $J_A = \{i \in P : i' \in W\}$ and $J_B = \{i \in Q : i \in W\}$. Then, we have that $|J_A| + |J_B| \leq k-1$, and these two sets $(P, Q)$-restricted trek-separate $A$ from $B$.

Conversely, suppose there exist sets $J_A \subseteq P$ and $J_B \subseteq Q$ with $|J_A| + |J_B| \leq k-1$ which $(P, Q)$-restricted trek-separate $A$ from $B$. Then $W = \{i : i \in J_B\} \cup \{i' : i \in J_A\}$ is a blocking set between $u$ and $v$ as above. By Menger's theorem, since $|W| \leq k-1$, there is no vertex disjoint system of $k$ paths from $A'$ to $B$ in $\tilde{G}_{P,Q}$. Thus, every $(P, Q)$-restricted trek system from $A$ to $B$ has a sided intersection so that $\det \Sigma_{A,B}^{(P,Q)} = 0$ by Proposition A.4.

From the special case of determinants, we deduce the general result, because if the smallest blocking set has size $r$, there exists a collection of $r$ disjoint paths between any subset of $A'$ and any subset of $B$, and this is the largest possible number of paths in such a collection. This means that all $(r+1) \times (r+1)$ minors of $\Sigma_{A,B}^{(P,Q)}$ are zero, but at least one $r \times r$ minor is not zero. Hence, $\Sigma_{A,B}^{(P,Q)}$ has rank $r$ for generic choices of the parameters. $\square$

## A.2. Proof of Theorem 4.4(i) for mixed graphs

A standard argument allows us to reduce to the case where there are no bidirected edges in the graph. This can be achieved by subdividing the bidirected edges; that is, for each bidirected edge $i \leftrightarrow j$ in the graph, where $i \leq j$, we replace $i \leftrightarrow j$ with a vertex $v_{i,j}$, directed edges $v_{i,j} \to i$ and $v_{i,j} \to j$. If $i$ or $j$ lie in $P$ or $Q$, then we add $v_{i,j}$ to $P$ or $Q$ respectively. Call the enhanced sets $\overline{P}$ and $\overline{Q}$. The graph $\overline{G}$ obtained from $G$ by subdividing all of its bidirected edges is called the *bidirected subdivision*, or *canonical DAG* associated to $G$.

**Proposition A.7.** *Let $A \subseteq P$, $B \subseteq Q$ be sets of vertices of a mixed graph with $|A| = |B|$.*

(i) *The matrix $\Sigma_{A,B}^{(P,Q)}$ associated to $G$ has the same generic rank as the matrix $\Sigma_{A,B}^{(\overline{P},\overline{Q})}$ associated to $\overline{G}$.*

(ii) *There exist $J_L \subseteq P$, $J_R \subseteq Q$ with $|J_L| + |J_R| = r$ such that $(J_L, J_R)$ $(P, Q)$-restricted trek-separates $A$ from $B$ in $G$ if and only if there exist $\overline{J}_L \subseteq \overline{P}$, $\overline{J}_R \subseteq \overline{Q}$ with $|\overline{J}_L| + |\overline{J}_R| = r$ such that $(\overline{J}_L, \overline{J}_R)$ $(\overline{P}, \overline{Q})$-restricted trek-separates $A$ from $B$ in $\overline{G}$.*

**Proof.** (i) Let $\bar{\Lambda} = (\bar{\lambda}_{k,l})$ and $\bar{\Omega} = (\bar{\omega}_{k,l})$ be parameters for $\bar{G}$. Define parameters for $G = (V, \mathcal{D}, \mathcal{B})$ as follows. For any directed edge $i \to j$ in $G$, set $\lambda_{i,j} = \bar{\lambda}_{i,j}$. For any bidirected edge $i \leftrightarrow j$ in $G$, set

$$\omega_{i,j} = \overline{\omega}_{v_{i,j},v_{i,j}} \overline{\lambda}_{v_{i,j},i} \overline{\lambda}_{v_{i,j},j}. \tag{A.1}$$

Finally, for each vertex $i$ in $G$, set

$$\omega_{i,i} = \overline{\omega}_{i,i} + \sum_{j \leftrightarrow i \in G} \overline{\omega}_{v_{i,j},v_{i,j}} \overline{\lambda}^2_{v_{i,j}i}. \tag{A.2}$$

Clearly, $\Lambda = (\lambda_{i,j}) \in \mathbb{R}^{\mathcal{D}}$. Since all $\overline{\omega}_{i,i} > 0$, the matrix $\Omega = (\omega_{i,j})$ is positive definite and, thus, in $PD(\mathcal{B})$. Let $\Sigma_{A,B}^{(P,Q)}$ be the matrix defined by $(\Lambda, \Omega)$, and let $\Sigma_{A,B}^{(\overline{P},\overline{Q})}$ be the matrix defined by $(\overline{\Lambda}, \overline{\Omega})$. Applying the $(P, Q)$-restricted trek rule to $G$ and $\bar{G}$, respectively, we see that $\Sigma_{A,B}^{(P,Q)} = \Sigma_{A,B}^{(\overline{P},\overline{Q})}$. We conclude that the set of matrices $\Sigma_{A,B}^{(\overline{P},\overline{Q})}$ associated to $\bar{G}$ is contained in the set of matrices $\Sigma_{A,B}^{(P,Q)}$ associated to $G$.

In general the reverse inclusion does not hold (Drton and Yu [10]). Nevertheless, the set of matrices $\Sigma_{A,B}^{(P,Q)}$ for $G$ has the same Zariski closure as the set of $\Sigma_{A,B}^{(\overline{P},\overline{Q})}$ for $\overline{G}$. Let $\mathcal{U} \subset \mathbb{R}^{\mathcal{D}} \times PD(\mathcal{B})$ be a neighborhood of $(0, I)$, that is, we consider matrices $\Lambda$ with entries of small magnitude and $\Omega$ near the identity matrix. To prove equality of the Zariski closures, it suffices to show that every matrix $\Sigma_{A,B}^{(P,Q)}$ given by a choice of $(\Lambda, \Omega) \in \mathcal{U}$ is equal to a matrix $\Sigma_{A,B}^{(\overline{P},\overline{Q})}$ associated to a choice of $\overline{\Lambda}$ and $\overline{\Omega}$ for $\overline{G}$. This in turn will follow from the trek rule if we can find $(\overline{\Lambda}, \overline{\Omega})$ such that (A.1) and (A.2) hold. However, this is possible because near the identity matrix, each off-diagonal entry $\omega_{i,j}$ is small. Specifically, we choose $\overline{\omega}_{v_{i,j},v_{i,j}} = 1$, and set $\overline{\lambda}_{v_{i,j},i} = \sqrt{|\omega_{i,j}|}$ and $\overline{\lambda}_{v_{i,j},j} = \operatorname{sign}(\omega_{i,j})\sqrt{|\omega_{i,j}|}$. When the $\omega_{i,j}$ are small enough, the sum on the right-hand side of (A.2) is smaller than $\omega_{i,i}$, which is near one. Hence, we can find a positive $\overline{\omega}_{i,i}$ satisfying (A.2), which ensures that $\overline{\Omega}$ is a diagonal matrix with positive diagonal entries as required.

(ii) Any pair of sets $S_L$ and $S_R$ that are $(P, Q)$-restricted trek-separating in $G$ are also clearly $(\overline{P}, \overline{Q})$-restricted trek-separating in $\overline{G}$. Conversely, suppose that $(\overline{J}_L, \overline{J}_R)$ is a minimal $(\overline{P}, \overline{Q})$-restricted trek-separating set in $\overline{G}$; that is, if any vertex is deleted from $(\overline{J}_L, \overline{J}_R)$, we no longer have a $(\overline{P}, \overline{Q})$-restricted trek-separating set. We show that such a minimal $(\overline{P}, \overline{Q})$-restricted trek-separating set in $\overline{G}$ corresponds to a $(P, Q)$-restricted trek-separating set in $G$. Define

$$J_L = (\overline{J}_L \cap P) \cup \{i \in P : v_{i,j} \in \overline{J}_L\},$$

$$J_R = (\overline{J}_R \cap Q) \cup \{j \in Q : v_{i,j} \in \overline{J}_R\}.$$

If $\overline{J}_L$ and $\overline{J}_R$ contain none of the vertices $v_{i,j}$, then $J_L$ and $J_R$ clearly $(P, Q)$-restricted trek-separate $A$ and $B$ in $G$. Otherwise, the way that $\{i \in P : v_{i,j} \in \overline{J}_L\}$ and $\{j \in Q : v_{i,j} \in \overline{J}_R\}$ are chosen is important. Given a vertex $v_{i,j} \in \overline{J}_L \cup \overline{J}_R$, let $\mathcal{T}(v_{i,j})$ denote the set of $(P, Q)$-restricted treks $\tau = (\tau_L, \tau_R)$ from $A$ to $B$ such that $\tau_L \cap \overline{J}_L = \{v_{i,j}\}$ or $\tau_R \cap \overline{J}_R = \{v_{i,j}\}$. Since $(\overline{J}_L, \overline{J}_R)$ is minimal, then $\mathcal{T}(v_{i,j})$ must be nonempty. This implies that in every $(P, Q)$-restricted trek $\tau = (\tau_L, \tau_R) \in \mathcal{T}(v_{i,j})$, up to relabeling, $i$ occurs in $\tau_L$ (whose sink lies in $A$) and $j$ occurs in $\tau_R$

(whose sink lies in $B$). For if there were also a trek $\tau = (\tau_L, \tau_R)$ in $\mathcal{T}(v_{i,j})$ which has $j$ in $\tau_L$ or $i$ in $\tau_R$, we could patch two halves of these treks together to find a $(P, Q)$-restricted trek from $A$ to $B$ that does not have a sided intersection with $(\overline{J}_L, \overline{J}_R)$. So, assume $i$ lies in $\tau_L$, and $j$ lies in $\tau_R$ for all $(P, Q)$-restricted treks in $\mathcal{T}(v_{i,j})$. In this case, add $i$ to $J_L$ whenever $v_{i,j} \in \overline{J}_L$, and add $j$ to $J_R$ whenever $v_{i,j} \in \overline{J}_R$. Then, $|J_L| + |J_R| \le |\overline{J}_L| + |\overline{J}_R|$, and $(J_L, J_R)$ $(P, Q)$-restricted trek-separates $A$ from $B$ in $G$. □

To finish the proof of Theorem 4.4(i), note that Proposition A.7 immediately reduces the statement to the case of directed acyclic graphs, which was given in the previous subsection.

### A.3. Proof of Theorem 4.4(ii)

Using first the Cauchy–Binet Theorem and then the Gessel–Viennot–Lindström Lemma A.3, we have that

$$\begin{aligned}
\det\bigl(\Sigma_{A,B}^{(P,Q)}\bigr) &= \det\bigl(\bigl((I-\Lambda)_{P,P}\bigr)^{-T}\bigr)_{A,P}\Omega_{P,Q}\bigl(\bigl((I-\Lambda)_{Q,Q}\bigr)^{-1}\bigr)_{Q,B} \\
&= \sum_{\substack{S\subseteq P, R\subseteq Q,\\ |S|=|R|=|A|}} \det\bigl(\bigl((I-\Lambda)_{P,P}\bigr)^{-T}\bigr)_{A,S} \det(\Omega_{S,R}) \det\bigl(\bigl((I-\Lambda)_{Q,Q}\bigr)^{-1}\bigr)_{R,B} \\
&= \sum_{\substack{S\subseteq P, R\subseteq Q,\\ |S|=|R|=|A|}} \sum_{\substack{\tau_1\in\mathcal{N}(S,A),\\ \tau_2\in\mathcal{N}(R,B)}} (-1)^{\tau_1+\tau_2}\lambda^{\tau_1+\tau_2}\det(\Omega_{S,R}) \\
&= \sum_{\substack{S\subseteq P, R\subseteq Q,\\ |S|=|R|=|A|}} \sum_{\substack{\tau_1\in\mathcal{N}(S,A),\\ \tau_2\in\mathcal{N}(R,B)}} \sum_{\sigma\in\Sigma_{|S|}} (-1)^{\tau_1+\tau_2+\mathrm{sign}(\sigma)}\lambda^{\tau_1+\tau_2}\prod_i \omega_{s_i, r_{\sigma(i)}}.
\end{aligned}$$

The latter sum goes over all trek systems between $A$ and $B$ whose left directed parts have no sided intersection and only use vertices from $P$, whose right directed parts have no sided intersection and only use vertices from $Q$, and use left and right sides are joined via "middle vertices" in $S$ and $R$. Each summand is the product of the trek monomials of the treks in each such system times the sign of the permutation induced by each such trek system. Moreover, note that each trek system with no sided intersection between $A$ and $B$ appears in this sum.

# Appendix B: Proofs for nested determinants

### B.1. Proof of Lemma 5.2

**Proof.** Suppose first that there exist $A_i$ and $A_j$ such that $A_i \cap A_j \neq 0$ for $i \neq j$. The case where two of the $B_i$'s intersect is analogous. Then, $|\Sigma_{A_1\uplus\cdots\uplus A_k, B_1\uplus\cdots\uplus B_k}| = 0$ since this matrix has a repeated row. On the other hand, we can select $C_i = \varnothing$, which makes $\mathcal{P}_{A_i,B_i,(C_i,D_i)} = 0$, so that for any choice of the rest of the $C_j$ and $D_j$, we have that $\prod_{j=1}^{k} \mathcal{P}_{A_j,B_j,(C_j,D_j)} = 0$. Thus, both sides are equal to 0, which establishes the statement.

Now, assume that $A_i \cap A_j = B_i \cap B_j = \varnothing$ for all $i \neq j$. We know by Theorem 4.1 that

$$|\Sigma_{A_1\cup\cdots\cup A_k, B_1\cup\cdots\cup B_k}| = \mathcal{P}_{A_1\cup\cdots\cup A_k, B_1\cup\cdots\cup B_k}.$$

For every $i = 1, \dots, k$ let $C_i^c$, the complement of $C_i$, be the union over all treks in trek systems with no sided intersection between $A_1 \cup \cdots \cup A_k$ and $B_1 \cup \cdots \cup B_k$ of the vertices that take part in the left side of the treks that start at $A_1 \cup \cdots \cup A_{i-1} \cup A_{i+1} \cup \cdots \cup A_k$. Let $D_i^c$ be the union over all treks in trek systems with no sided intersection between $A_1 \cup \cdots \cup A_k$ and $B_1 \cup \cdots \cup B_k$ of the vertices that take part in the right side of the treks that start at $A_1 \cup \cdots \cup A_{i-1} \cup A_{i+1} \cup \cdots \cup A_k$ (and end at $B_1 \cup \cdots \cup B_{i-1} \cup B_{i+1} \cup \cdots \cup B_k$).

By assumption, if we are given two trek systems $\mathcal{T}_1$ and $\mathcal{T}_2$ with no sided intersection between $A_1 \cup \cdots \cup A_k$ and $B_1 \cup \cdots \cup B_k$, then we can swap the treks from $A_i$ to $B_i$ from the first system $\mathcal{T}_1$ with those from the second system $\mathcal{T}_2$ and obtain two other trek systems between $A_1 \cup \cdots \cup A_k$ and $B_1 \cup \cdots \cup B_k$ with no sided intersection. Hence, each summand in $\mathcal{P}_{A_1\cup\cdots\cup A_k, B_1\cup\cdots\cup B_k}$ can be factored uniquely as a product of one element from each of $\mathcal{P}_{A_i,B_i,(C_i,D_i)}$. Conversely, the product of one element from each of $\mathcal{P}_{A_i,B_i,(C_i,D_i)}$ gives an element from $\mathcal{P}_{A_1\cup\cdots\cup A_k, B_1\cup\cdots\cup B_k}$. Thus,

$$|\Sigma_{A_1\cup\cdots\cup A_k, B_1\cup\cdots\cup B_k}| = \prod_{i=1}^{k} \mathcal{P}_{A_i,B_i,(C_i,D_i)},$$

as required. □

## B.2. Proof of Lemma 5.5

**Proof.** Recall that $\mathcal{P}_{\{a_1,\dots,a_n\},\{b_1,\dots,b_n\},(E,F)}$ is the sum of the trek monomials of all trek systems with no sided intersection between $\{a_1, \dots, a_n\}$ and $\{b_1, \dots, b_n\}$ that only use $E$ on the left and $F$ on the right. For each such trek system, the trek starting at $a_i$ only uses $E_{ij}$ on the left, and the trek ending at $b_j$ only uses $F_{ij}$ on the right. On the other hand, the determinant of $(\mathcal{P}_{a_i,b_j,(E_{ij},F_{ij})})_{i,j}$ is the sum of the trek monomials of all trek systems with no sided intersection between $\{a_1, \dots, a_n\}$ and $\{b_1, \dots, b_n\}$ for which the trek starting at $a_i$ only uses $E_{ij}$ on the left, and the trek ending at $b_j$ only uses $F_{ij}$ on the right. Therefore, the two quantities are equal. □

## B.3. Proof of Proposition 5.8

**Proof.** We will show that the determinant of the matrix with entries

$$(|\Sigma_{\mathrm{pa}(u)\cup\{u\},\mathrm{pa}(u)\cup\{x\}})_{u\in\mathrm{pa}(i)\cup\{j\},x\in\mathrm{pa}(i)\cup\{i\}}$$

is divisible by $\mathcal{P}_{\mathrm{pa}(i)\cup\{j\},\mathrm{pa}(i)\cup\{i\},(\mathrm{pa}(i)\cup\{j\},V)}$. Combinatorially, this means that there is a $(\mathrm{pa}(i) \cup \{j\}, V)$-restricted trek separation between the sets $\mathrm{pa}(i) \cup \{j\}$ and $\mathrm{pa}(i) \cup \{j\}$. Indeed, they are $(\mathrm{pa}(i) \cup \{j\}, V)$-restricted trek separated by $(\varnothing, \mathrm{pa}(i))$.

We begin by showing that the sets $(\mathrm{pa}(u), \mathrm{pa}(u))$, $(u, x)$ for $u \in \mathrm{pa}(i) \cup \{j\}$ and $x \in \mathrm{pa}(i) \cup \{i\}$ satisfy the swapping property. Firstly, consider a system of treks with no sided intersection between $\mathrm{pa}(u) \cup \{u\}$ and $\mathrm{pa}(u) \cup \{x\}$. Suppose that in this system it is not the case that $\mathrm{pa}(u)$ is connected to $\mathrm{pa}(u)$ and $u$ is connected to $x$. Then, there must exist a trek between $u$ and an element from $\mathrm{pa}(u)$. Since $u$ is ancestral, the left side of this trek has to end in a directed edge. That means that the left side of this trek contains an element from $\mathrm{pa}(u)$, which is impossible since this creates a sided intersection on the left side of this system. Therefore, we have a contradiction, and any such trek system connects $\mathrm{pa}(u)$ to $\mathrm{pa}(u)$ and $u$ to $x$.

Now, suppose that we have two systems of treks with no sided intersection between $\mathrm{pa}(u) \cup \{u\}$ and $\mathrm{pa}(u) \cup \{i\}$. Call them $\mathcal{T}_1$ and $\mathcal{T}_2$. In both systems, the treks connecting $u$ and $x$ need to start with a bidirected edge at $u$ or with a directed edge away from $u$ in order to avoid intersections on the left. We need to show that we can exchange the part connecting $\mathrm{pa}(u)$ to $\mathrm{pa}(u)$ in $\mathcal{T}_1$ with the corresponding part of $\mathcal{T}_2$, thereby obtaining two new trek systems with no sided intersection. Suppose for contradiction that once we make such an exchange, we get a sided intersection. Then, one of the treks from $u$ to $x$ gets a sided intersection with a trek from $\mathrm{pa}(u)$ to $\mathrm{pa}(u)$. Since the former trek has the form $u(\leftrightarrow) \to \cdots \to x$, the created intersection has to be on its right side. Switch the tails of the two intersecting treks. We get a trek of the form $u \leftrightarrow\to \cdots \to z \in \mathrm{pa}(u)$. But this is a contradiction to $u$ being ancestral.

We have shown that the sets $(\mathrm{pa}(u), \mathrm{pa}(u))$, $(u, x)$ for $u \in \mathrm{pa}(i) \cup \{j\}$ and $x \in \mathrm{pa}(i) \cup \{i\}$ satisfy the swapping property. We now show that

$$|\Sigma_{\mathrm{pa}(u)\cup\{u\},\mathrm{pa}(u)\cup\{x\}}| = \mathcal{P}_{\mathrm{pa}(u),\mathrm{pa}(u)}\mathcal{P}_{u,x,(u,V)}. \tag{B.1}$$

Note that $|\Sigma_{\mathrm{pa}(u)\cup\{u\},\mathrm{pa}(u)\cup\{x\}}| = \mathcal{P}_{\mathrm{pa}(u)\cup\{u\},\mathrm{pa}(u)\cup\{x\}}$. Since the sets $(\mathrm{pa}(u), \mathrm{pa}(u))$, $(u, x)$ for $u \in \mathrm{pa}(i) \cup \{j\}$ and $x \in \mathrm{pa}(i) \cup \{i\}$ satisfy the swapping property, every trek system with no sided intersection between $\mathrm{pa}(u) \cup \{u\}$ and $\mathrm{pa}(u) \cup \{x\}$ splits into a trek system connecting $\mathrm{pa}(u)$ and $\mathrm{pa}(u)$ and a single trek connecting $u$ and $x$. The latter trek only has the vertex $u$ on its left side. In other words, it starts either with a bidirected edge at $u$ or with a directed edge pointing away from $u$.

On the other hand we claim that every trek system connecting $\mathrm{pa}(u)$ to $\mathrm{pa}(u)$ with no sided intersection, and every trek from $u$ to $x$ that only has $u$ on the left can be combined into a trek system connecting $\mathrm{pa}(u) \cup \{u\}$ and $\mathrm{pa}(u) \cup \{x\}$ with no sided intersection. Suppose for contradiction that the combination gives a sided intersection. So, there is a trek from $a \in \mathrm{pa}(u)$ to $b \in \mathrm{pa}(u)$ that has sided intersection with the considered trek from $u$ to $x$. The intersection cannot be on the left since otherwise we would have a loop $u \to \cdots \to a \to u$ which is not allowed. Thus, there is intersection on the right. Swapping the right tails then gives a trek $u(\leftrightarrow) \to \cdots \to b$. But since $u$ is ancestral, we know that every trek between $u$ and its parents has to end with a directed edge at $u$. We have arrived at a contradiction and, thus, the claimed combination into a trek system connecting $\mathrm{pa}(u) \cup \{u\}$ and $\mathrm{pa}(u) \cup \{x\}$ with no sided intersection is possible. This proves (B.1).

Finally, it remains to show that

$$\det(\mathcal{P}_{u,x,(u,V)})_{u\in\mathrm{pa}(i)\cup\{j\},x\in\mathrm{pa}(i)\cup\{i\}} = \mathcal{P}_{\mathrm{pa}(i)\cup\{j\},\mathrm{pa}(i)\cup\{i\},(\mathrm{pa}(i)\cup\{j\},V)}.$$

But this equality follows directly from Lemma 5.5. □

## Acknowledgements

This work started during the 2016 Mathematics Research Communities research conference on Algebraic Statistics. Mathias Drton and Luca Weihs were supported by the US National Science Foundation under Grant No. DMS 1712535. Elina Robeva was supported by a US National Science Foundation Postdoctoral Fellowship No. DMS 1703821.

## References


[1] Chen, B. (2016). Identification and overidentification of linear structural equation models. In *Advances in Neural Information Processing Systems* 29 (D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon and R. Garnett, eds.) 1587–1595. Curran Associates.

[2] Draisma, J., Sullivant, S. and Talaska, K. (2013). Positivity for Gaussian graphical models. *Adv. in Appl. Math.* **50** 661–674. MR3044565 https://doi.org/10.1016/j.aam.2013.03.001

[3] Drton, M. (2009). Likelihood ratio tests and singularities. *Ann. Statist.* **37** 979–1012. MR2502658 https://doi.org/10.1214/07-AOS571

[4] Drton, M. (2018). Algebraic problems in structural equation modeling. In *The* 50*th Anniversary of Gröbner Bases* (T. Hibi, ed.). *Adv. Stud. Pure Math.* **77** 35–86. Tokyo: Math. Soc. Japan. MR3839705 https://doi.org/10.2969/aspm/07710035

[5] Drton, M., Foygel, R. and Sullivant, S. (2011). Global identifiability of linear structural equation models. *Ann. Statist.* **39** 865–886. MR2816341 https://doi.org/10.1214/10-AOS859

[6] Drton, M. and Maathuis, M.H. (2017). Structure learning in graphical modeling. *Annu. Rev. Stat. Appl.* **4** 365–393.

[7] Drton, M., Sturmfels, B. and Sullivant, S. (2007). Algebraic factor analysis: Tetrads, pentads and beyond. *Probab. Theory Related Fields* **138** 463–493. MR2299716 https://doi.org/10.1007/s00440-006-0033-2

[8] Drton, M., Sturmfels, B. and Sullivant, S. (2009). *Lectures on Algebraic Statistics*. *Oberwolfach Seminars* **39**. Basel: Birkhäuser. MR2723140 https://doi.org/10.1007/978-3-7643-8905-5

[9] Drton, M. and Weihs, L. (2016). Generic identifiability of linear structural equation models by ancestor decomposition. *Scand. J. Stat.* **43** 1035–1045. MR3573674 https://doi.org/10.1111/sjos.12227

[10] Drton, M. and Yu, J. (2010). On a parametrization of positive semidefinite matrices with zeros. *SIAM J. Matrix Anal. Appl.* **31** 2665–2680. MR2740626 https://doi.org/10.1137/100783170

[11] Evans, R.J. and Richardson, T.S. (2014). Markovian acyclic directed mixed graphs for discrete data. *Ann. Statist.* **42** 1452–1482. MR3262457 https://doi.org/10.1214/14-AOS1206

[12] Fink, A., Rajchgot, J. and Sullivant, S. (2016). Matrix Schubert varieties and Gaussian conditional independence models. *J. Algebraic Combin.* **44** 1009–1046. MR3566228 https://doi.org/10.1007/s10801-016-0698-2

[13] Foygel, R., Draisma, J. and Drton, M. (2012). Half-trek criterion for generic identifiability of linear structural equation models. *Ann. Statist.* **40** 1682–1713. MR3015040 https://doi.org/10.1214/12-AOS1012

[14] Frydenberg, M. (1990). The chain graph Markov property. *Scand. J. Stat.* **17** 333–353. MR1096723
[15] Koster, J.T.A. (2002). Marginalizing and conditioning in graphical models. *Bernoulli* **8** 817–840. MR1963663
[16] Lauritzen, S.L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. New York: The Clarendon Press. MR1419991
[17] Lněnička, R. and Matúš, F. (2007). On Gaussian conditional independent structures. *Kybernetika* (*Prague*) **43** 327–342. MR2362722
[18] Pearl, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge: Cambridge Univ. Press. MR2548166 https://doi.org/10.1017/CBO9780511803161
[19] Richardson, T. and Spirtes, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166 https://doi.org/10.1214/aos/1031689015
[20] Roozbehani, H. and Polyanskiy, Y. (2014). Algebraic methods of classifying directed graphical models. In 2014 *IEEE International Symposium on Information Theory* 2027–2031.
[21] Shpitser, I., Evans, R., Richardson, T. and Robins, J. (2014). Introduction to nested Markov models. *Behaviormetrika* **41** 3–39.
[22] Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation*, *Prediction*, *and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. MR1815675
[23] Studený, M. (2005). *Probabilistic Conditional Independence Structures*. *Information Science and Statistics*. London: Springer. MR3183760
[24] Sullivant, S., Talaska, K. and Draisma, J. (2010). Trek separation for Gaussian graphical models. *Ann. Statist.* **38** 1665–1685. MR2662356 https://doi.org/10.1214/09-AOS760
[25] Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In *AAAI/IAAI* 567–573.
[26] van Ommen, T. and Mooij, J.M. (2017). Algebraic equivalence of linear structural equation models. In *Proceedings of the* 33*rd Annual Conference on Uncertainty in Artificial Intelligence* (*UAI*-17).
[27] Verma, T.S. and Pearl, J. (1991). Equivalence and synthesis of causal models. In *Uncertainty in Artificial Intelligence* 6 255–268. Elsevier. UCLA Cognitive Systems Laboratory, Technical Report (R-150).
[28] Weihs, L., Robinson, B., Dufresne, E., Kenkel, J., Kubjas, K., II, R.M., Nguyen, N., Robeva, E. and Drton, M. (2018). Determinantal generalizations of instrumental variables. *J. Causal Inference* **6**.
[29] Wermuth, N. (2011). Probability distributions with summary graph structure. *Bernoulli* **17** 845–879. MR2817608 https://doi.org/10.3150/10-BEJ309
[30] Wright, S. (1934). The method of path coefficients. *Ann. Math. Stat.* **5** 161–215.

*Received July 2018 and revised November 2019*