Bernoulli **23**(2), 2017, 1102–1129
DOI: 10.3150/15-BEJ771

Automorphism groups of Gaussian Bayesian networks

JAN DRAISMA[1] and PIOTR ZWIERNIK[2]

[1]*Department of Mathematics and Computer Science, TU Eindhoven, PO Box 513,5600 MB Eindhoven and CWI Amsterdam, The Netherlands. E-mail: j.draisma@tue.nl*
[2]*Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain. E-mail: piotr.zwiernik@upf.edu*

In this paper, we extend earlier work on groups acting on Gaussian graphical models to Gaussian Bayesian networks and more general Gaussian models defined by chain graphs with no induced subgraphs of the form $i \to j - k$. We fully characterise the maximal group of linear transformations which stabilises a given model and we provide basic statistical applications of this result. This includes equivariant estimation, maximal invariants for hypothesis testing and robustness. In our proof, we derive simple necessary and sufficient conditions on vanishing subminors of the concentration matrix in the model. The computation of the group requires finding the essential graph. However, by applying Stúdeny's theory of imsets, we show that computations for DAGs can be performed efficiently without building the essential graph.

Keywords: chain graphs; Gaussian graphical models; group action; equivariant estimator; invariant test; transformation family

1. Introduction

Having an explicit group action on a parametric statistical model gives a better understanding of equivariant estimation or invariant testing for the model under consideration [4,10,18,27]. In [6], we have identified the largest group that acts on an undirected Gaussian graphical model and we have shown how this group can be used to study equivariant estimators of the covariance matrix in this model class. In the present paper, we extend our discussion to Gaussian Bayesian networks by embedding this class in a special family of chain graph models.

A chain graph $\mathcal{H}$ is a graph with both directed edges and undirected edges that contains no semi-directed cycles, that is, no sequences of nodes $i_1, \ldots, i_k, i_{k+1} = i_1$ such that for every $j = 1, \ldots, k$ either $i_j - i_{j+1}$ or $i_j \to i_{j+1}$ and where the latter happens at least once. Chain graph models were introduced as a natural generalization of models on directed acyclic graphs (DAGs) (see [17]), and they become increasingly popular in applications; see, for example, [5, 11,30]. One of the original motivations for introducing chain graph models was that inclusion of undirected edges allowed the modelling of "associative relations" as distinct from causal relations represented by directed edges; see [16].

In this paper, we focus on *chain graphs without flags* (NF-CGs), that is, with no induced subgraphs of the form $i \to j - k$. For more details on these graph-theoretic notions, see Section 2.1. Note that both undirected graphs and directed acyclic graphs are chain graphs without flags, and hence this constrained family of graphs generalizes both classes. Chain graphs with no flags first

appeared in [2] in the context of an alternative definition of chain graph models. They became important theoretical concepts in the study of DAG models; see, for example, [20,21,25].

Gaussian models on chain graphs constitute a flexible family of graphical models, which contains both undirected Gaussian graphical models and Gaussian Bayesian networks defined by directed acyclic graphs. Let $\mathcal{H}$ be a chain graph. Let $\mathbb{R}^{\mathcal{H}}$ denote the space of all $m \times m$ matrices $\Lambda = [\lambda_{ij}]$ such that $\lambda_{ij} = 0$ if $i \nrightarrow j$ in $\mathcal{H}$; let $\mathcal{S}_m^+$ denote the space of all $m \times m$ symmetric positive definite matrices and let $\mathcal{S}_{\mathcal{H}}^+$ be the subspace of $\mathcal{S}_m^+$ consisting of matrices $\Omega = [\omega_{ij}]$ such that $\omega_{ij} = 0$ if $i \neq j$ and $i - \!\!/ j$ in $\mathcal{H}$.

Definition 1.1. *The Gaussian chain graph model $M(\mathcal{H})$ on a chain graph $\mathcal{H}$ consists of all Gaussian distributions on $\mathbb{R}^m$ with mean zero and concentration matrices K of the form*

$$K = (I - \Lambda)\Omega\big(I - \Lambda^T\big) \qquad \text{such that } \Lambda \in \mathbb{R}^{\mathcal{H}}, \Omega \in \mathcal{S}_{\mathcal{H}}^+. \tag{1.1}$$

The set of all matrices of this form will be denoted by $\mathcal{K}(\mathcal{H})$.

Two non-equivalent definitions of chain graph models can be found in the literature, referred to as LWF or AMP chain graph models in [2]. This refers to Lauritzen–Wermuth–Frydenberg [12, 17] and Andersson–Madigan–Perlman [2] (Alternative Markov Properties), respectively. These two definitions differ in how exactly a graph encodes the defining set of conditional independence statements. Our definition above follows the AMP formulation. However, since we work only with chain graphs with no flags, both formulations are equivalent by the following result; see [2], Theorems 1 and 4.

Theorem 1.2. *If $\mathcal{H}$ is a chain graph with no flags, then both the LWF and the AMP definition of chain graph models coincide.*

Let $X = (X_1, \ldots, X_m)$ be a Gaussian vector with the covariance matrix $\Sigma \in M(\mathcal{H})$. By $\mathrm{GL}_m(\mathbb{R})$ denote the *general linear group of* $\mathbb{R}^m$, or equivalently, the set of all invertible real $m \times m$ matrices. A linear transformation $g \in \mathrm{GL}_m(\mathbb{R})$ yields another Gaussian vector $Y = gX$. A basic question of equivariant inference is for which g the covariance matrix $g\Sigma g^T$ of Y still lies in $M(\mathcal{H})$. More formally, the general linear group $\mathrm{GL}_m(\mathbb{R})$ acts on $\mathcal{S}_m^+$ by $g \cdot \Sigma := g\Sigma g^T$. Fix a chain graph $\mathcal{H}$ with no flags. We study the problem of finding

$$G := \big\{g \in \mathrm{GL}_m(\mathbb{R}) | g \cdot M(\mathcal{H}) \subseteq M(\mathcal{H})\big\}. \tag{1.2}$$

In other words, find the stabilizer of $M(\mathcal{H})$ in $\mathrm{GL}_m(\mathbb{R})$.

The fact that G forms a group follows from the following technical remark.

Remark 1.1. The set G is a closed algebraic subgroup of $\mathrm{GL}_m(\mathbb{R})$, and in particular has the structure of a Lie group: First, it is clear from the definition that G is closed under matrix multiplication. To see that it is closed under inversion and closed in the Zariski topology, we argue as follows. Let $\overline{M(\mathcal{H})}$ denote the Zariski closure of $M(\mathcal{H})$ in $\mathbb{R}^{m \times m}$, that is, the set of matrices in $\mathbb{R}^{m \times m}$ whose entries satisfy all polynomial equations that hold identically on $M(\mathcal{H})$. Suppose that $g \in \mathrm{GL}_m(\mathbb{R})$ maps $\overline{M(\mathcal{H})}$ into itself. Then, since acting with g preserves positive definite

matrices and since $M(\mathcal{H})$ consists of all positive definite matrices in $\overline{M(\mathcal{H})}$ (see [8], Proposition 3.3.13) for the case of DAGs; the general chain graph case is similar), g also preserves $M(\mathcal{H})$. Thus, G may be characterised as the stabilizer of the real algebraic variety $\overline{M(\mathcal{H})}$. This shows that G is Zariski closed. To see that it is also closed under inversion, note that $g \cdot \overline{M(\mathcal{H})}$ is a real algebraic variety of the same dimension as $\overline{M(\mathcal{H})}$ and contained in $\overline{M(\mathcal{H})}$, hence equal to $\overline{M(\mathcal{H})}$. But then also $g^{-1} \cdot \overline{M(\mathcal{H})}$ equals $\overline{M(\mathcal{H})}$.

The problem in (1.2) can be alternatively phrased in terms of concentration matrices, which will be more useful in our case. Let $\mathrm{GL}_m(\mathbb{R})$ act on $\mathcal{S}_m^+$ by $g \cdot K := g^{-T} K g^{-1}$. Now find all $g \in \mathrm{GL}_m(\mathbb{R})$ such that $g \cdot \mathcal{K}(\mathcal{H}) \subseteq \mathcal{K}(\mathcal{H})$.

Example 1.3. The DAG $\overset{1}{\bullet} \to \overset{2}{\bullet} \to \overset{3}{\bullet}$ defines a model given by a single conditional independence statement $X_1 \perp\!\!\!\perp X_3 | X_2$, and hence is equal to the model on the undirected graph $\overset{1}{\bullet} - \overset{2}{\bullet} - \overset{3}{\bullet}$. Since the directed part of this graph is empty, by (1.1) the model consists of all covariance matrices such that the corresponding concentration matrices are of the form

$$K = \Omega = \begin{bmatrix} * & * & 0 \\ * & * & * \\ 0 & * & * \end{bmatrix}.$$

By [6], Theorem 1.1, G in this case consists of invertible matrices of the form

$$\begin{bmatrix} * & * & 0 \\ 0 & * & 0 \\ 0 & * & * \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 & * & * \\ 0 & * & 0 \\ * & * & 0 \end{bmatrix}.$$

1.1. The group G

Example 1.3 showed that two different chain graphs may define the same chain graph model. We discuss this phenomenon in more detail in Section 2.2. For any NF-CG $\mathcal{H}$, denote by $\mathcal{H}^*$ the *unique* graph without flags with the largest number of undirected edges which induces the same Gaussian model as $\mathcal{H}$. The fact that such a unique graph exists follows from Proposition 2.10 given later. For example, for the DAG in Example 1.3, such a graph is given by the undirected graph $\overset{1}{\bullet} - \overset{2}{\bullet} - \overset{3}{\bullet}$. By $c^*(i)$ we denote the children of i in $\mathcal{H}^*$, so $c^*(i) = \{j : i \to j \text{ in } \mathcal{H}^*\}$. Similarly, by $n^*(i)$ we denote the set of neighbours of i in $\mathcal{H}^*$, that is, nodes j connected to i by an undirected edge, which we denote by $i - j$. Write

$$N^*(i) := \{i\} \cup n^*(i) \cup c^*(i).$$

Our main results can be summarized as follows. For a fixed chain graph without flags $\mathcal{H}$ with set of nodes given by $[m] := \{1, \ldots, m\}$, consider the set G^0 of invertible matrices given by

$$G^0 := \big\{ g = [g_{ij}] \in \mathrm{GL}_m(\mathbb{R}) : g_{ij} = 0 \text{ if } N^*(i) \not\subseteq N^*(j) \big\}. \tag{1.3}$$

In particular, all invertible diagonal matrices are in G^0.

An automorphism of a chain graph is any permutation of its nodes that maps directed edges to directed edges and undirected edges to undirected edges.

Theorem 1.4. *Let $\mathcal{H}$ be a chain graph without flags. The group G in (1.2) is generated by its connected normal subgroup G^0 and the group* $\mathrm{Aut}(\mathcal{H}^*)$ *of automorphisms of the essential graph $\mathcal{H}^*$.*

The fact that G^0 forms a connected subset of $\mathbb{R}^{m\times m}$ follows directly from the definition in (1.3). The fact that G^0 forms a normal subgroup of G is part of the theorem.

In the undirected case, this theorem reduces to [6], Theorem 1.1. However, the proof in our current, more general setting is much more involved, first because the set $\mathcal{K}(\mathcal{H})$ is not a linear space, and second because the characterization is in terms of the essential graph rather than the graph itself.

For some graphs, there may be two nodes i, j such that $N^*(i) = N^*(j)$. We show in Section 4 that in this case the permutation switching i with j is a valid graph automorphism of $\mathcal{H}^*$. By (1.3), the transposition of i and j also lies in G^0, which shows that the groups G^0 and $\mathrm{Aut}(\mathcal{H}^*)$ may have a non-trivial intersection. In Section 4, we prove a more refined version of Theorem 1.4 that gets rid of this redundancy.

Given a set of edges defining a chain graph without flags $\mathcal{H}$, we would like to find G^0 by listing all pairs (i, j) for $i, j \in [m]$ such that $g_{ij} = 0$ for all $g \in G^0$. Since our theorem depends on computing the essential graph $\mathcal{H}^*$, a natural question arises on complexity of this computation. In Section 5, we show how G^0 can be efficiently computed in the case of DAGs. We propose an algorithm that does not require computing the essential graph $\mathcal{H}^*$ and runs in $O(m^2)$ steps.

1.2. Existence and robustness of equivariant estimators

There are many situations in statistical inference, when the maximum likelihood estimator is not available or hard to compute. Whenever a group G acting on the model and it state space is available, a minimal property that we may want to require is equivariance; see, for example, [10], Section 3.3. This is a common requirement in constructions of robust alternatives for the maximum likelihood estimators. Let $\mathbf{X} = (X^{(1)}, \ldots, X^{(n)})$ denote a random sample of length n from the model $M(\mathcal{H})$, where $X^{(i)}$ denotes the ith independent copy of a Gaussian vector X. An estimator of the covariance matrix of X given the sample $\mathbf{X}$ is any map $T_n : (\mathbb{R}^m)^n \to M(\mathcal{H})$. We say that T_n is an *equivariant estimator* (or *G-equivariant* if we want to make G explicit), if it satisfies

$$T_n(g \cdot \mathbf{X}) = g \cdot T_n(\mathbf{X}) \qquad \text{for every } g \in G, \mathbf{X} \in \left(\mathbb{R}^m\right)^n, \tag{1.4}$$

where, in our case, the action of G on $(\mathbb{R}^m)^n$ is

$$g \cdot (x_1, \ldots, x_n) = (gx_1, \ldots, gx_n).$$

An important example of an equivariant estimator is the maximum likelihood estimator (MLE).

In situations when $n << m$, establishing existence of the MLE may be a hard task. For decomposable undirected graphs, the MLE exists with probability one if and only if n is at least

the size of the maximal clique of the given graph. However, in general, whether the MLE exists, with probability one, for a given sample size and a given graph is a subtle matter; see [14,28] and references therein. By contrast, the question whether for a given sample size an equivariant estimator exists, turns out to have a remarkably elegant answer for any chain graph with no flags. This, in particular, gives us lower bounds on the sample size that guarantees that the MLE for Gaussian DAG models exists with probability one. To state our criterion, for a vertex i of $\mathcal{H}$ we define

$$\downarrow i := \big\{ j : N^*(i) \subseteq N^*(j) \big\}.$$

Example 1.5. Consider the graph $\overset{1}{\bullet} \to \overset{2}{\bullet} \leftarrow \overset{3}{\bullet}$. This graph is essential with $N^*(1) = \{1, 2\}$, $N^*(2) = \{2\}$, $N^*(3) = \{2, 3\}$ and so $\downarrow 1 = \{1\}, \downarrow 3 = \{3\}$ and $\downarrow 2 = \{1, 2, 3\}$.

Theorem 1.6. *Let $\mathcal{H}$ be a chain graph without flags with set of nodes $[m]$. There exists a G-equivariant estimator $T_n : (\mathbb{R}^m)^n \to M(\mathcal{H})$ of the covariance matrix in the model $M(\mathcal{H})$ if and only if*

$$n \geq \max_{i \in [m]} |\downarrow i|.$$

Note that the threshold in Theorem 1.6 can also be read off from G^0: by Theorem 1.4 it is the maximal cardinality of the support of a row of a matrix in G^0.

Example 1.7. To get a better understanding of the condition in Theorem 1.6, consider a graph $\mathcal{H}$ with m nodes and edges $i - m$ for $i = 1, \ldots, m-1$. We have $\mathcal{H} = \mathcal{H}^*$ and $N^*(m) = [m]$, $N^*(i) = \{i, m\}$ for $i = 1, \ldots, m-1$. Hence,

$$\max_{i \in [m]} |\downarrow i| = |\downarrow 1| = 2$$

so the threshold for this graph is low. We note that, if $n \geq 2$, then also the MLE exists with probability 1, which follows from standard results for decomposable graphical models. Suppose now that $\mathcal{H}$ is a DAG with edges $i \to m$ for $i = 1, \ldots, m-1$. This graph is an essential DAG, that is, $\mathcal{H} = \mathcal{H}^*$. We have $N^*(m) = \{m\}$ and $N^*(i) = \{i, m\}$ for $i = 1, \ldots, m-1$. Therefore,

$$\max_{i \in [m]} |\downarrow i| = |\downarrow m| = m$$

and an equivariant estimator exists if and only if $n \geq m$. In that case, the sample covariance matrix is positive definite with probability one, and hence also the MLE exists with probability one. If we add a vertex $m+1$ and an arrow $m \to m+1$, then for the new graph we have

$$\max_{i \in [m+1]} |\downarrow i| = \big|\downarrow (m+1)\big| = 2,$$

and so the threshold drops dramatically.

Our next result concerns G^0-invariants. In general, when a group acts on a set, an invariant is an (arbitrarily-valued) function on that set which is constant on orbits of the group. In our setting, we consider functions on (open, dense subsets) of the space $(\mathbb{R}^m)^n = \mathbb{R}^{m\times n}$ of n-tuples of samples, where we identify two functions when they coincide on an open, dense subset. Then there turns out to be a unique maximal G^0-invariant, through which every other invariant factorises. In other words, every invariant is a function of this so-called *maximal invariant*. For more on this maximal invariant, see [18], Section 6.2, [6], Section 1.3. An important application of the maximal invariant is in the construction of invariant tests; see [10,18]. Suppose, for instance, that we want to test the hypothesis that the distribution of the multivariate Gaussian random vector X lies in $M(\mathcal{H})$ against the alternative that it does not, and suppose that for the sample of size n, $\mathbf{X} = \mathbf{x}$, the test would accept the hypothesis. Then, since $M(\mathcal{H})$ is stable under the action of any $g \in G$, it is natural to require that our test also accepts the hypothesis on observing $g\mathbf{x}$. Thus, the test itself would have to be G-invariant, which means that it needs to be a function of the maximal invariant.

Our result uses the equivalence relation $\sim$ on $[m]$ defined by

$$i \sim j \quad \text{if and only if} \quad N^*(i) = N^*(j). \tag{1.5}$$

We write $\bar{i}$ for the equivalence class of $i \in [m]$ and $[m]/\sim$ for the set of all equivalence classes. Apart from the next result, this relation will also feature in Section 4.

Theorem 1.8. *Let $\mathcal{H}$ be a chain graph without flags. Suppose that $n \geq \max_i |{\downarrow}i|$. Then the map $\tau : \mathbb{R}^{m\times n} \to \prod_{\bar{i}\in[m]/\sim} \mathbb{R}^{n\times n}$ given by*

$$\mathbf{x} \mapsto \left(\mathbf{x}[{\downarrow}i]^T \left(\mathbf{x}[{\downarrow}i]\mathbf{x}[{\downarrow}i]^T\right)^{-1}\mathbf{x}[{\downarrow}i]\right)_{\bar{i}\in[m]/\sim},$$

where $\mathbf{x}[{\downarrow}i] \in \mathbb{R}^{|{\downarrow}i|\times n}$ is the submatrix of $\mathbf{x}$ given by all rows indexed by ${\downarrow}i$, is the maximal G^0-invariant.

Example 1.9. Consider again Example 1.5. The corresponding maximal invariant statistic has the following three components

$$\mathbf{x}[1]^T\left(\mathbf{x}[1]\mathbf{x}[1]^T\right)^{-1}\mathbf{x}[1], \qquad \mathbf{x}^T\left(\mathbf{x}\mathbf{x}^T\right)^{-1}\mathbf{x}, \qquad \mathbf{x}[3]^T\left(\mathbf{x}[3]\mathbf{x}[3]^T\right)^{-1}\mathbf{x}[3],$$

where $\mathbf{x} \in \mathbb{R}^{3\times n}$ is a matrix whose columns are data points, and $\mathbf{x}[i]$ denotes the ith row of this matrix. Here, $\frac{1}{n}\mathbf{x}[i]\mathbf{x}[i]^T$ is just the sample variance of X_i and $\frac{1}{n}\mathbf{x}\mathbf{x}^T$ is the sample covariance matrix.

In [6], we used the structure of the group to provide non-trivial bounds on the finite sample breakdown point for all equivariant estimators of the covariance matrix for undirected Gaussian graphical models. These results extend to chain graphs without flags; see [6], Section 3.2, for the motivation and more details.

Proposition 1.10. *Assume that $n \geq \max_i |{\downarrow}i|$. Then for any G-equivariant estimator $T : \mathbb{R}^{m\times n} \to \mathcal{S}^+_{\mathcal{G}}$, the finite sample breakdown point at a generic sample $\mathbf{x}$ is at most $\lceil (n - \max_i |{\downarrow}i| + 1)/2 \rceil / n$.*

Here, by a generic sample, we mean a point in $\mathbb{R}^{m\times n}$ that lies outside of a lower-dimensional algebraic subspace.

Unlike the proof of Theorem 1.4, the proofs of Theorem 1.6, Theorem 1.8 and Proposition 1.10 are similar to the undirected case because they depend on G only through the induced poset defined by the ordering relation $N^*(i) \subseteq N^*(j)$, which drives the zero pattern of the group G^0. The proofs of these three results will be therefore omitted; see [6] for details.

Organization of the paper

In Section 2.1, we provide some basic graph-theoretical definitions. The theory of Markov equivalence of chain graphs will be discussed in Section 2.2. In Section 3, we provide new results that give necessary and sufficient vanishing conditions for subdeterminants of the concentration matrix $K \in \mathcal{K}(\mathcal{H})$. In Section 4, we analyze the structure of the group G in order to prove Theorem 1.4. In Section 5, we show that in the case of DAG models, structural imsets give us all the required information to identify G without constructing the essential graphs. Section 6 contains some simple examples illustrating Theorem 1.4.

2. Preliminaries

In this section, we discuss basic notions of the theory of chain graphs and chain graph models.

2.1. Basics of chain graphs

Let $\mathcal{H}$ be a *hybrid graph*, that is, a graph with both directed and undirected edges, but neither loops nor multiple edges. This excludes also a situation when two nodes are connected by both an undirected edge and a directed edge. We assume that the set of nodes of $\mathcal{H}$ is labelled with $[m] = \{1, \dots, m\}$. A directed edge (arrow) from i to j is denoted by $i \to j$ and an undirected edge between i and j is denoted by $i - j$. We write $i \cdots j$, and say that i and j are *linked*, whenever we mean that either $i \to j$ or $i \leftarrow j$, or $i - j$.

An *undirected path* between i and j in a hybrid graph $\mathcal{H}$ is any sequence $k_1, \dots, k_n$ of nodes such that $k_1 = i$, $k_n = j$ and $k_i - k_{i+1}$ in $\mathcal{H}$ for every $i = 1, \dots, n-1$. A *semi-directed path* between i and j is any sequence $k_1, \dots, k_n$ of nodes such that $k_1 = i$, $k_n = j$ and either $k_i - k_{i+1}$ or $k_i \to k_{i+1}$ in $\mathcal{H}$ for every $i = 1, \dots, n-1$ and $k_i \to k_{i+1}$ for at least one i. A *directed path* between i and j in a hybrid graph $\mathcal{H}$ is any sequence $k_1, \dots, k_n$ of nodes such that $k_1 = i$, $k_n = j$ and $k_i \to k_{i+1}$ in $\mathcal{H}$ for every $i = 1, \dots, n-1$. A *semi-directed cycle* in a hybrid graph $\mathcal{H}$ is a sequence $k_1, \dots, k_{n+1} = k_1$, $n \geq 3$ of nodes in $\mathcal{H}$ such that $k_1, \dots, k_n$ are distinct, and this sequence forms a semi-directed path. In a similar way, we define a *undirected cycle* and directed cycle.

Definition 2.1. *A chain graph* (*or CG*) *is a hybrid graph without semi-directed cycles.*

A set of nodes T is *connected* in $\mathcal{H}$, if for every $i, j \in T$ there exists an *undirected* path between i and j. Maximal connected subsets in $\mathcal{H}$ with respect to set inclusion are called *components* in $\mathcal{H}$. The class of components of $\mathcal{H}$ is denoted by $\mathcal{T}(\mathcal{H})$. The elements of $\mathcal{T}(\mathcal{H})$ form a partition of the set of nodes of $\mathcal{H}$. For any subset $A \subseteq [m]$ of the set of vertices, we define the *induced graph* on A, denoted by $\mathcal{H}_A$, as the graph with set of nodes A and for any two $i, j \in A$ we have $i \to j$, $j \to i$ or $i - j$ if and only if $i \to j$, $j \to i$ or $i - j$ in $\mathcal{H}$, respectively.

Define the set of *parents* of $A \subseteq [m]$, denoted by $p_{\mathcal{H}}(A)$, as the set of $i \in [m]$ such that $i \to a$ in $\mathcal{H}$ for some $a \in A$. The set of *children* $c_{\mathcal{H}}(A)$ is the set of $i \in [m]$ such that $a \to i$ in $\mathcal{H}$ for some $a \in A$; and the set of neighbours $n_{\mathcal{H}}(A)$ is the set of all $i \in [m]$ such that $i - a$ in $\mathcal{H}$ for some $a \in A$. In addition, we define

$$N_{\mathcal{H}}(i) := \{i\} \cup n_{\mathcal{H}}(i) \cup c_{\mathcal{H}}(i).$$

If C is a connected set in a chain graph $\mathcal{H}$, then there are no arrows between elements in C, for otherwise there would exist a semi-directed cycle. In particular, the induced graph $\mathcal{H}_C$ on C is an undirected graph and $p_{\mathcal{H}}(C)$ is disjoint from C for any $C \in \mathcal{T}(\mathcal{H})$. In addition, for every $A \subseteq [m]$ the induced subgraph $\mathcal{H}_A$ of a chain graph $\mathcal{H}$ is a chain graph itself. A *clique* in an undirected graph is a subset of nodes such that any two nodes are linked. We say that a clique is *maximal* if it is maximal with respect to inclusion.

Definition 2.2. *For any CG $\mathcal{H}$, an immorality is any induced subgraph of $\mathcal{H}$ of the form $i \to j \leftarrow k$. A flag is any induced subgraph of the form $i \to j - k$. A chain graph without flags is abbreviated by NF-CG.*

Undirected graphs and DAGs are chain graphs without flags. In a general NF-CG, whenever we see the structure $i \to j - k$, there is automatically an arrow $i \to k$. This gives rise to the following basic fact.

Lemma 2.3. *If $\mathcal{H}$ is a NF-CG, then $p_{\mathcal{H}}(A) = p_{\mathcal{H}}(T)$ for every $T \in \mathcal{T}(\mathcal{H})$ and non-empty $A \subseteq T$. In particular, for any two $i, j \in [m]$ such that $i - j$ in $\mathcal{H}$ we have $p_{\mathcal{H}}(i) = p_{\mathcal{H}}(j)$.*

Definition 2.4. *Let $\mathcal{H}$ be a chain graph. For any two distinct components $T, T' \in \mathcal{T}(\mathcal{H})$, consider the set of all arrows between T and T'. If this set is non-empty, then we call it a meta-arrow and denote by $T \Rightarrow T'$. That is,*

$$T \Rightarrow T' := \{i \to j : i \in T, j \in T', i \to j \text{ in } \mathcal{H}\}.$$

The notion of meta-arrow is important in the considerations of equivalence classes of chain graphs, which we discuss in the next section.

2.2. Equivalence classes of chain graphs

A chain graph model is given by all concentration matrices of the form (1.1). In Example 1.3, we saw that two different chain graphs may give the same Gaussian models, or equivalently,

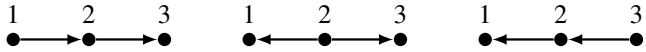

Figure 1. Three equivalent DAGs.

the same set of conditional independence statements. If two NF-CGs $\mathcal{G}$ and $\mathcal{H}$ define the same chain graph model, we say that they are *equivalent*. For example, the three DAGs in Figure 1 are equivalent.

The equivalence class of $\mathcal{H}$ in the set of NF-CGs is denoted by $\langle\mathcal{H}\rangle$:

$$\langle\mathcal{H}\rangle = \{\mathcal{G} : \mathcal{G} \text{ is a NF-CG equivalent to } \mathcal{H}\}.$$

Equivalence of CGs and DAGs was discussed in many papers, for example, [1,12,21,29]. We briefly list the most relevant results.

Definition 2.5. *The skeleton of a chain graph $\mathcal{H}$ is the undirected graph such that $i - j$ whenever $i \cdots j$ in $\mathcal{H}$.*

Theorem 2.6. *Two NF-CGs with the same set of nodes are equivalent if and only if they have the same skeleton and the same immoralities.*

The original statement of this result, given by Frydenberg in [12], is more general and applies to any chain graph in the LWF definition of chain graph models; see also Theorem 1.2.

As was remarked in [21] considering meta-arrows helps to understand equivalence classes of chain graphs. Suppose that we want to obtain one chain graph from another with the same skeleton by changing some of the arrows $i \to j$ to $i - j$ or $i \leftarrow j$. Changing only a subset of arrows in a meta-arrow $T \Rightarrow T'$ is not permitted as it would introduce semi-directed cycles. Hence, the only permitted operations on arrows of $\mathcal{H}$, if we work in the class of CGs, is either changing the directions of all the elements of $T \Rightarrow T'$ or changing all arrows of $T \Rightarrow T'$ into undirected edges. The following basic operation on a chain graph was defined in [21,22].

Definition 2.7. *Let $\mathcal{H}$ be a NF-CG and let $T \Rightarrow T'$ be a meta-arrow in $\mathcal{H}$ where $T, T' \in \mathcal{T}(\mathcal{H})$. Merging of T and T' is an operation of changing all elements of the meta-arrow $T \Rightarrow T'$ into undirected edges. Merging is called legal if*:

(a) $p_{\mathcal{H}}(T') \cap T$ *is a clique of* T;
(b) $p_{\mathcal{H}}(T') \setminus T = p_{\mathcal{H}}(T)$.

Lemma 2.8. *Let $\mathcal{H}$ be a NF-CG and let $\mathcal{H}'$ be a graph obtained from $\mathcal{H}$ by legal merging of two connected components. Then $\mathcal{H}' \in \langle\mathcal{H}\rangle$.*

Proof. See, for example, the proof of Lemma 22 in [24]. □

For two distinct CGs $\mathcal{G}, \mathcal{H}$ with the same skeleton, we write $\mathcal{G} \subseteq \mathcal{H}$ if, whenever $i \to j$ in $\mathcal{G}$, then either $i \to j$ or $i - j$ in $\mathcal{H}$, and whenever $i - j$ in $\mathcal{G}$, then $i - j$ in $\mathcal{H}$. We write $\mathcal{G} \subset \mathcal{H}$ if $\mathcal{G} \subseteq \mathcal{H}$ and $\mathcal{G} \neq \mathcal{H}$.

Theorem 2.9 (Roverato, Studeny [21,22]). *Let $\mathcal{G}$ and $\mathcal{H}$ be two equivalent NF-CGs such that $\mathcal{G} \subset \mathcal{H}$. Then there exists a finite sequence $\mathcal{G} = \mathcal{G}_0 \subset \cdots \subset \mathcal{G}_r = \mathcal{H}$, with $r \geq 1$, of equivalent NF-CGs such that, for all $i = 1, \ldots, r$, $\mathcal{G}_i$ can be obtained from $\mathcal{G}_{i-1}$ by a legal merging of two connected components of $\mathcal{G}_{i-1}$.*

The following proposition shows a beautiful property of models of chain graphs with no flags, namely, there is always a unique NF-CG representing $\langle \mathcal{H} \rangle$ with the largest number of undirected edges.

Proposition 2.10 (Roverato, Studeny [21,22]). *There exists a unique element $\mathcal{H}^*$ in $\langle \mathcal{H} \rangle$ that is maximal in the sense that $\mathcal{H}' \subseteq \mathcal{H}^*$ for every $\mathcal{H}' \in \langle \mathcal{H} \rangle$.*

Definition 2.11. *Let $\mathcal{H}$ be a NF-CG. The graph $\mathcal{H}^*$ of Proposition 2.10 is called the essential graph. The directed arrows in $\mathcal{H}^*$ are called essential. For notational convenience, we write $p^*(A)$, $n^*(A)$ and $c^*(A)$ for $p_{\mathcal{H}^*}(A)$, $n_{\mathcal{H}^*}(A)$ and $c_{\mathcal{H}^*}(A)$, respectively.*

By definition, $\mathcal{H}^*$ has the same skeleton as $\mathcal{H}$, and an edge is essential if and only if it occurs as an arrow with the same orientation in every $\mathcal{H}' \in \langle \mathcal{H} \rangle$; all other edges are undirected. For example, the essential graph for any of the graphs in Figure 1 is the undirected graph $\overset{1}{\bullet} - \overset{2}{\bullet} - \overset{3}{\bullet}$, whereas the essential graph of $\mathcal{H} = \overset{1}{\bullet} \rightarrow \overset{2}{\bullet} \leftarrow \overset{3}{\bullet}$ is $\mathcal{H}$ itself. By Theorem 2.6, every arrow that participates in an immorality in $\mathcal{H}$ is essential, but $\mathcal{H}$ may contain other essential arrows. For example, in the DAG in Figure 2 all arrows are essential but not all of them form immoralities.

The following result has been independently observed in [21,22].

Theorem 2.12. *If $\langle \mathcal{H} \rangle$ contains a DAG $\mathcal{G}$, then the essential graph $\mathcal{H}^*$ is equal to the essential graph of a DAG as defined in* [1].

Remark 2.1. Our terminology is consistent with [21]. However, in [3] the essential graph for a chain graph is defined in a different way and it corresponds to the essential graph $\mathcal{H}^*$ only if $\langle \mathcal{H} \rangle$ contains a DAG.

3. Subdeterminants of concentration matrices

In this section, we offer a detour into the combinatorial commutative algebra of sub-determinants that vanish identically on concentration matrices of chain graph models. Our main result in this

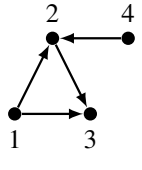

Figure 2. A NF-CG whose arrows are all essential but not all part of immoralities.

section, Proposition 3.5, is the concentration-matrix counterpart to the results in [7,26] for covariance matrices. It provides simple necessary and sufficient conditions for a concentration matrix to lie in $\mathcal{K}(\mathcal{H})$, where $\mathcal{H}$ is a general chain graph on $[m]$. These conditions will be used in Appendix 7 to prove Theorem 4.4, but are also of interest in their own right. We will use the following combinatorial notions.

Definition 3.1. *A* cup *in $\mathcal{H}$ is a quadruple (i, j, k, l) of vertices in $\mathcal{H}$ where*:

1. *either $i = j$ or $i \to j$*; and
2. *either $j = k$ or $j - k$*; and
3. *either $k = l$ or $k \leftarrow l$.*

We say that the cup starts *in i and* ends *in l.*

Note that we do not require $i \to j - k \leftarrow l$ to be an induced subgraph of $\mathcal{H}$ and so $j \neq k$ is possible in a chain graph with no flags.

Definition 3.2. *Let A and B be sets of vertices of $\mathcal{H}$ of the same cardinality d. A* cup system *from A to B is a set U of d cups in $\mathcal{H}$ whose starting points exhaust A and whose end points exhaust B. The cup system U from A to B gives rise to a bijection $A \to B$ that sends $a \in A$ to the end point of the cup in U that starts with a. After fixing labellings $A = \{a_1, \ldots, a_d\}$ and $B = \{b_1, \ldots, b_d\}$, this bijection gives rise to a permutation of $[d]$; define $\operatorname{sgn}(U)$ to be the sign of this permutation. The cup system U from A to B is said to be* self-avoiding *if, for each $k = 1, 2, 3, 4$, the elements $u_k \in [m]$ of $u = (u_1, u_2, u_3, u_4) \in U$ are all distinct.*

For the graph $\overset{1}{\bullet} \to \overset{2}{\bullet} \leftarrow \overset{3}{\bullet}$, there is no self-avoiding cup system from $\{1, 2\}$ to $\{2, 3\}$ but there is such a system between $\{1\}$ and $\{3\}$.

Definition 3.3. *Let λ_{ij} be the parameters corresponding to arrows $i \to j$ in $\mathcal{H}$ and let ω_{ij} be the parameters corresponding to undirected edges $i - j$ and to the diagonal (ω_{ii}). The* weight *of a cup (i, j, k, l) in $\mathcal{H}$ is the product of the (i, j) entry of $(I - \Lambda)$, the (j, k)-entry of Ω and the (k, l)-entry of $(I - \Lambda)^T$, which is the (l, k)-entry of $(I - \Lambda)$. The weight of a cup system U from A to B, denoted $w(U)$, is the product of the weights of the cups in U. This is a monomial of degree k in the ω_{ij} times a monomial of degree at most k in the variables $-\lambda_{ij}$.*

Let $K[A, B]$ denote the $A \times B$-submatrix of $K = (I - \Lambda)\Omega(I - \Lambda)^T$. By expanding the entries, we find that

$$\det K[A, B] = \sum_U \operatorname{sgn}(U) w(U), \tag{3.1}$$

where the sum is over all cup systems U from A to B. In this expression cancellation can occur because of the signs $\operatorname{sgn}(U)$ (not because of the signs in the $-\lambda_{ij}$, which we might as well have taken as new variables). The following proposition captures exactly which terms cancel. For more details on the arguments, we refer to [7,26].

Proposition 3.4. *Relative to the fixed labellings of A and B, the $A \times B$-subdeterminant of K equals*

$$\det K[A, B] = \sum_{U \text{ self-avoiding}} \operatorname{sgn}(U) w(U).$$

Moreover, for any two self-avoiding cup systems U and U' with $w(U) = w(U')$, we have $\operatorname{sgn}(U) = \operatorname{sgn}(U')$.

Proof. To see that the sum in (3.1) can be restricted to self-avoiding cup systems U, we proceed as in the Lindström–Gessel–Viennot lemma ([13], Theorem 1) and give a sign-reversing involution σ on the set of non-self-avoiding cup systems, as follows. Order any cup system U from A to B as $\{u_1, \ldots, u_d\}$ where u_i starts in a_i. If U is not self-avoiding, let $a \in \{2, 3\}$ be minimal such that the entries $u_{ia}, i \in [d]$ are not all distinct, and let (i, i') be a lexicographically minimal pair such that $u_{ia} = u_{i'a}$. Then $\sigma(U)$ is the cup system obtained from U by replacing u_i and $u_{i'}$ by their swaps at position a. For instance, if $a = 2$, then $u'_i = (u_{i1}, u_{i2} = u_{i'2}, u_{i'3}, u_{i'4})$ and $u'_{i'} = (u_{i'1}, u_{i'2} = u_{i2}, u_{i3}, u_{i4})$; and similarly for $a = 3$. Now $\operatorname{sgn}(U') = -\operatorname{sgn}(U)$ and σ is indeed an involution. This proves the expression in the proposition. The second statement is more subtle, but it follows by applying [7], Theorem 3.3, to the DAG obtained from $\mathcal{H}$ by reversing all arrows and replacing all undirected edges $i - j$ by a pair $i \leftarrow k \rightarrow j$ of arrows, where k is a new vertex. Indeed, self-avoiding cup systems in $\mathcal{H}$ correspond to special types of trek systems without sided intersection in that new graph. □

Proposition 3.5. *The subdeterminant* $\det K[A, B]$ *is identically zero on the model corresponding to $\mathcal{H}$ if and only if there does not exist a self-avoiding cup system from A to B in $\mathcal{H}$. These vanishing conditions completely specify the model $M(\mathcal{H})$.*

Proof. The first statement follows directly from Proposition 3.4. To show that the vanishing subdeterminants of K describe $M(\mathcal{H})$ we argue as follows. Because $M(\mathcal{H})$ is a conditional independence model, mapping covariance matrices to concentration matrices via $\Sigma \mapsto \Sigma^{-1}$ gives another conditional independence model. This follows from [9], Lemma 2.1; see also [19]. Moreover, every conditional independence model is defined by vanishing subdeterminants; see, for example, [8], Proposition 3.1.13. □

4. The group G

As we noted in the Introduction, the group G has a structure of a Lie group. The connected component of G that contains the identity matrix is called the identity component. In this section, we describe G by first describing its identity component and then the remaining part of the group.

4.1. The identity component

Denote by E_{ij} the matrix in $\mathbb{R}^{m \times m}$ with all entries zero apart from the (i, j)th element which is 1. By G^0 denote the normal subgroup of G which forms the connected component of the identity

matrix. In the following, we show that G^0 is precisely the set of matrices defined in (1.3). The subgroup T^m of all invertible diagonal matrices is contained in the group G because scaling of vector $\mathbf{X}$ does not affect conditional independencies. By [6], Lemma 2.1, to compute G^0, it suffices to check for which $(i, j) \in [m] \times [m]$ the one-parameter groups $(I + tE_{ij})$, $t \in \mathbb{R}$, lie in G; or equivalently $E_{ij} \in \mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra of G. We provide this result for reader's convenience.

Lemma 4.1 (Lemma 2.1, [6]). *Let $H \subseteq \mathrm{GL}_m(\mathbb{R})$ be a real algebraic matrix group containing the group T^m. Then the Lie algebra of H has a basis consisting of matrices E_{ij} with (i, j) running through some subset I of $[m] \times [m]$. Moreover, the set I defines a pre-order on $[m]$ in the sense that (i, i) lies in I for all $i \in [m]$ and that $(i, j), (j, k) \in I \Rightarrow (i, k) \in I$. Conversely, the E_{ij} with (i, j) running through any set $I \subseteq [m] \times [m]$ defining a pre-order on $[m]$ span the Lie algebra of a unique closed connected subgroup of $\mathrm{GL}_m(\mathbb{R})$ containing T^m, namely, the group of all $g \in \mathrm{GL}_m(\mathbb{R})$ with $g_{ij} = 0$ unless $(i, j) \in I$.*

Before we provide the main result of this section, we also recall ([6], Proposition 2.2).

Proposition 4.2 (Proposition 2.2, [6]). *Let $\mathcal{H}$ be an undirected graph. For $i, j \in [m]$ the matrix E_{ij} lies in $\mathfrak{g}$ if and only if $N_{\mathcal{H}}(i) \subseteq N_{\mathcal{H}}(j)$.*

If $\mathcal{H}$ is a NF-CG such that $\mathcal{H}^*$ is an undirected graph, then Proposition 4.2 can be used to characterise G^0 for $\mathcal{H}$ by passing to the essential graph. However, it is not immediately clear how this result extends to all chain graphs without flags. We first note that one direction of the above result holds in general.

Lemma 4.3. *Let $\mathcal{H}$ be an NF-CG. If $N_{\mathcal{H}}(i) \subseteq N_{\mathcal{H}}(j)$, then $E_{ij} \in \mathfrak{g}$.*

Proof. If $i = j$, then the statement is clear so suppose that $i \neq j$. We have $N_{\mathcal{H}}(i) \subseteq N_{\mathcal{H}}(j)$ only if either $j \to i$ or $i - j$ in $\mathcal{H}$. Suppose first that $j \to i$. We have

$$(I - tE_{ji})(I - \Lambda)\Omega(I - \Lambda)^T(I - tE_{ij}) = (I - \tilde{\Lambda})\Omega(I - \tilde{\Lambda})^T,$$

where $\tilde{\Lambda} = -\Lambda - tE_{ji} + tE_{ji}\Lambda$; $\tilde{\lambda}_{uv} = \lambda_{uv}$ if $u \neq j$; $\tilde{\lambda}_{jv} = \lambda_{jv} - t\lambda_{iv}$ if $v \neq i$; and $\tilde{\lambda}_{ji} = \lambda_{ji} + t$. The fact that $\tilde{\Lambda}$ lies in $\mathbb{R}^{\mathcal{H}}$ follows from $c_{\mathcal{H}}(i) \subseteq c_{\mathcal{H}}(j)$ and hence for every v, if $\lambda_{jv} = 0$, then $\lambda_{iv} = 0$.

If $i - j$ in $\mathcal{H}$, then $i \cup n_{\mathcal{H}}(i) \subseteq j \cup n_{\mathcal{H}}(j)$ and $p_{\mathcal{H}}(i) = p_{\mathcal{H}}(j)$ by Lemma 2.3. By Proposition 4.2 applied to the undirected part of $\mathcal{H}$, we can write $\Omega = (I + tE_{ji})\tilde{\Omega}(I + tE_{ij})$ for some $\tilde{\Omega} \in \mathcal{S}^+_{\mathcal{H}}$. Therefore,

$$\begin{aligned}&(I - tE_{ji})(I - \Lambda)\Omega(I - \Lambda)^T(I - tE_{ij})\\ &\quad = (I - tE_{ji})(I - \Lambda)(I + tE_{ji})\tilde{\Omega}(I + tE_{ij})(I - \Lambda)^T(I - tE_{ij}),\end{aligned}$$

where we now show that there exists $\tilde{\Lambda} \in \mathbb{R}^{\mathcal{H}}$ such that

$$(I - tE_{ji})(I - \Lambda)(I + tE_{ji}) = (I - \tilde{\Lambda}).$$

Indeed,

$$\tilde{\Lambda} = \Lambda + t\Lambda E_{ji} - tE_{ji}\Lambda + t^2 E_{ji}\Lambda E_{ji},$$

where the last term must vanish because $\lambda_{ij} = 0$. Hence, $\tilde{\Lambda}$ is obtained from Λ by adding a multiple of the jth column to the ith column and by adding a multiple of the ith row to the jth row. The fact that $\tilde{\Lambda}$ lies in $\mathbb{R}^{\mathcal{H}}$ follows from the fact that $c_{\mathcal{H}}(i) \subseteq c_{\mathcal{H}}(j)$ and $p_{\mathcal{H}}(i) = p_{\mathcal{H}}(j)$, that is, the ith column has the same support as the jth column and the support of the ith row is contained in the support of the jth row. □

The converse of the lemma does not hold for a general NF-CG $\mathcal{H}$. Consider for instance $\overset{1}{\bullet} \to \overset{2}{\bullet} \to \overset{3}{\bullet}$. By Example 1.3, the element $I + tE_{12}$ lies in G^0 but $\{1, 2\} \not\subseteq \{2, 3\}$. Nevertheless, the converse of the lemma above does hold when $\mathcal{H}$ is essential, that is, when $\mathcal{H} = \mathcal{H}^*$; this is the main result of this section.

Theorem 4.4. *Let $\mathcal{H}$ be an essential NF-CG. Then $E_{ij} \in \mathfrak{g}$ if and only if $N_{\mathcal{H}}(i) \subseteq N_{\mathcal{H}}(j)$.*

As it is somewhat technical, we postpone the proof to the Appendix. As we noted in the beginning of this section, the set of all $E_{ij} \in \mathfrak{g}$ gives already the complete information on the group G^0. Hence, Theorem 4.4 gives the description of G^0 in (1.3).

Example 4.5. Consider a DAG $\mathcal{H} = \overset{1}{\bullet} \to \overset{2}{\bullet} \leftarrow \overset{3}{\bullet}$ discussed in Example 1.5. Since $N_{\mathcal{H}}(2) \subseteq N_{\mathcal{H}}(1)$ and $N_{\mathcal{H}}(2) \subseteq N_{\mathcal{H}}(3)$, both E_{21} and E_{23} lie in $\mathfrak{g}$ but no other off-diagonal elements of matrices in G^0 can be non-zero.

4.2. The rest of G

Note that G^0 given in Theorem 4.4 in general is not the whole group G. For example, both for the model $\overset{1}{\bullet} \to \overset{2}{\bullet} \leftarrow \overset{3}{\bullet}$ and for any of the equivalent DAGs in Figure 1 the permutation matrix

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

lies in G but not in G^0 because the $(1, 3)$-entry of matrices in G^0 is zero by the example above. The following result shows that permutation matrices form the basis for understanding the remaining part of the group G. For the proof, see [6], Proposition 2.5.

Proposition 4.6. *Every element $g \in G$ can be written as $g = \sigma g_0$, where $g_0 \in G^0$ and σ is a permutation matrix contained in G.*

An automorphism of a hybrid graph is any bijection $\sigma : [m] \to [m]$ of its nodes such that for every $i, j \in [m]$ we have $\sigma(i) - \sigma(j)$ if and only if $i - j$ and $\sigma(i) \to \sigma(j)$ if and only if $i \to j$.

Lemma 4.7. *Let $\mathcal{H}$ be a NF-CG and $\mathcal{H}^*$ its essential graph. Let $\sigma \in \mathrm{GL}_m(\mathbb{R})$ be a permutation matrix. Then $\sigma \in G$ if and only if σ is an automorphism of $\mathcal{H}^*$.*

Proof. The model $M(\mathcal{H})$ is uniquely defined by the set of conditional independence statements (see for example [15]). Given a set of such statements that come from a chain graph $\mathcal{H}$ the equivalence class $\langle\mathcal{H}\rangle$ is determined uniquely. The essential graph $\mathcal{H}^*$ is the unique representative of $\langle\mathcal{H}\rangle$ with the largest number of undirected edges. Since any permutation σ applied to $\mathcal{H}^*$ gives a NF-CG with the same number of undirected and directed edges (it simply relabels the nodes), σ lies in the model if and only if σ is an automorphism of $\mathcal{H}^*$. □

By Lemma 4.7, G is generated by G^0 and the automorphism group of $\mathcal{H}^*$, which proves Theorem 1.4.

Recall the equivalence relation $\sim$ on $[m]$ defined by (1.5). As explained in the Introduction, the expression $G = \mathrm{Aut}(\mathcal{H}^*)G^0$ is not minimal in the sense that $\mathrm{Aut}(\mathcal{H}^*)$ and G^0 may intersect. To get rid of that intersection, we define $\widetilde{\mathcal{H}}^*$ to be the graph with vertex set $[m]/\sim$ and $\bar{i} \to \bar{j}$ ($\bar{i} - \bar{j}$) in $\widetilde{\mathcal{H}}^*$ if and only if $i \to j$ ($i - j$) in $\mathcal{H}$. We first show that $\widetilde{\mathcal{H}}^*$ is well-defined.

Lemma 4.8. *Let $\mathcal{H}$ be a NF-CG and $\mathcal{H}^*$ its essential graph. Two elements $i, j \in [m]$ are equivalent if and only if $\{i\} \cup n^*(i) = \{j\} \cup n^*(j)$, $p^*(i) = p^*(j)$ and $c^*(i) = c^*(j)$. In particular, the graph $\widetilde{\mathcal{H}}^*$ is well-defined.*

Proof. If $N^*(i) = N^*(j)$, then i and j are necessarily linked. Since $i \in N^*(j)$ and $j \in N^*(j)$, we conclude that in fact $i - j$ in $\mathcal{H}^*$. By Lemma 2.3, since $i - j$, we also have $p^*(i) = p^*(j)$. Hence, $i \sim j$ if and only if $\{i\} \cup n^*(i) = \{j\} \cup n^*(j)$, $c^*(i) = c^*(j)$ and $p^*(i) = p^*(j)$, which shows that the definition of the arrows and edges in $\mathcal{H}$ is independent of the representative i and j. □

Define $c : [m]/\sim \to \mathbb{N}, \bar{i} \mapsto |\bar{i}|$ and view c as a coloring of the vertices of $\widetilde{\mathcal{H}}^*$ by natural numbers. Let $\mathrm{Aut}(\widetilde{\mathcal{H}}^*, c)$ denote the group of automorphisms of $\widetilde{\mathcal{H}}^*$ preserving the coloring. There is a lifting $\ell : \mathrm{Aut}(\widetilde{\mathcal{H}}^*, c) \to \mathrm{Aut}(\mathcal{H}^*)$ defined as follows: the element $\tau \in \mathrm{Aut}(\widetilde{\mathcal{H}}^*, c)$ is mapped to the unique bijection $\ell(\tau) : [m] \to [m]$ that maps each equivalence class $\bar{i}$ to the equivalence class $\tau(\bar{i})$ by sending the kth smallest element of $\bar{i}$ (in the natural linear order on $[m]$) to the kth smallest element of $\tau(\bar{i})$, for $k = 1, \ldots, |\bar{i}|$.

Example 4.9. Consider a DAG $\mathcal{H}$ and its essential graph $\mathcal{H}^*$ in Figure 3. Since 3 and 4 are equivalent, the induced essential graph $\widetilde{\mathcal{H}}^*$ is equal to $\overset{1}{\bullet} - \overset{2}{\bullet} - \overset{3,4}{\bullet}$. There are no non-trivial automorphisms of this graph preserving cardinality of equivalence classes and $\mathrm{Aut}(\widetilde{\mathcal{H}}^*, c) = \{I\}$. In particular, ℓ is a trivial mapping.

Theorem 4.10. *The group G equals $\ell(\mathrm{Aut}(\widetilde{\mathcal{H}}^*, c))G^0$, and the intersection $\ell(\mathrm{Aut}(\widetilde{\mathcal{H}}^*, c)) \cap G^0$ is trivial, so G is the semidirect product $\ell(\mathrm{Aut}(\widetilde{\mathcal{H}}^*, c)) \ltimes G^0$.*

Proof. It is a standard result from the Lie group theory that the connected component of the identity G^0 is a normal subgroup of G. Hence, to show that $G = G^0 \rtimes \mathrm{Aut}(\widetilde{\mathcal{H}}^*, c)$ we need to show

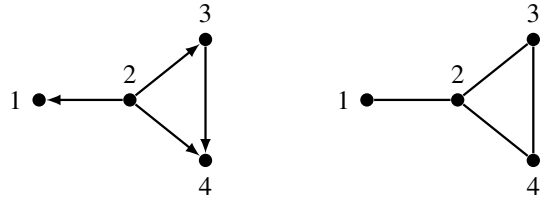

Figure 3. On the left a DAG on four nodes. On the right its essential graph.

that $G = G^0 \cdot \mathrm{Aut}(\widetilde{\mathcal{H}}^*, c)$ and $G^0 \cap \mathrm{Aut}(\widetilde{\mathcal{H}}^*, c) = \{I\}$. The first part follows by Proposition 4.6 and Lemma 4.7. To show that $G^0 \cap \mathrm{Aut}(\widetilde{\mathcal{H}}^*, c) = \{I\}$ note that transpositions of i and j lie in G^0 precisely when i and j are equivalent, and hence, when they do not lie in $\ell(\mathrm{Aut}(\widetilde{\mathcal{H}}^*, c))$. □

Remark 4.1. To the coloured graph $(\widetilde{\mathcal{H}}^*, c)$ we can associate a Gaussian graphical model $M(\mathcal{H}, c)$ with *multivariate nodes*, where node $\bar{i}$ is associated to a Gaussian vector of dimension $c_{\bar{i}}$. This model coincides with $M(\mathcal{H})$. This also shows, conversely, that our framework extends to general Gaussian graphical models of chain graphs with no flags with multivariate nodes.

Computing the essential graph $\mathcal{H}^*$ is not always a simple task. In Section 5 we show how to identify the group G without finding $\mathcal{H}^*$ in the case when $\mathcal{H}$ is a DAG. In Section 6, we illustrate Theorem 4.10 with some basic examples.

5. Efficient computations for DAG models

In this section, we present some efficient techniques for computing the group G^0 in the case when $\mathcal{H}$ is a DAG. The following characterization of essential graphs of DAGs will be useful.

Theorem 5.1 (Roverato, Studený [21,22]). *If $\mathcal{H}$ is a DAG, then each connected component of $\mathcal{H}^*$ is decomposable. Moreover, $\mathcal{H}^*$ coincides with the essential graph of $\mathcal{H}$ as defined in* [1] (*see also Remark* 2.1).

For any DAG $\mathcal{H}$ on the set of nodes $[m]$, the *standard imset* for $\mathcal{H}$ is an integer-valued function $u_{\mathcal{H}} : 2^{[m]} \to \mathbb{Z}$, where $2^{[m]}$ is the set of all subsets of $[m]$, defined by

$$u_{\mathcal{H}} := \delta_{[m]} - \delta_{\varnothing} + \sum_{i \in [m]} (\delta_{p_{\mathcal{H}}(i)} - \delta_{p_{\mathcal{H}}(i) \cup \{i\}}), \tag{5.1}$$

where $\delta_A : 2^{[m]} \to \{0, 1\}$ satisfies $\delta_A(B) = 1$ if $A = B$ and is zero otherwise. For example, it is easy to verify that all DAGs in Figure 1 give rise to the imset represented by Figure 4.

Lemma 5.2 (Corollary 7.1, [23]). *Let $\mathcal{G}, \mathcal{H}$ be two DAGs. Then $\mathcal{H} \in \langle \mathcal{G} \rangle$ if and only if $u_{\mathcal{G}} = u_{\mathcal{H}}$.*

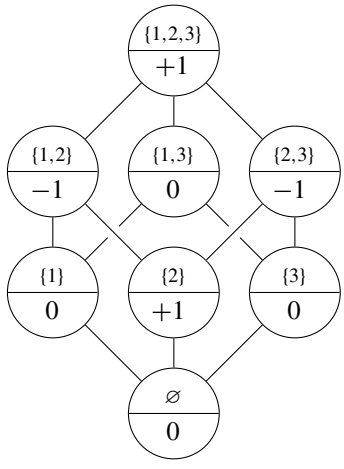

Figure 4. The imset $u_{\mathcal{H}}$, where $\mathcal{H}$ is any of the three equivalent DAGs in Figure 1.

The support of $u_{\mathcal{H}}$ for a DAG $\mathcal{H}$ has been described in [25] directly in terms of the essential graph. To provide this result we introduce some useful notions related to chain graphs.

Definition 5.3. *A set $B \subseteq [m]$ of nodes in a chain graph $\mathcal{H}$ is idle if $i \cdots j$ for all $i, j \in B$; and for every $i \in [m] \setminus B$ and every $j \in B$, $i \to j$ in $\mathcal{H}$.*

By [24], Lemma 18, every chain graph has a unique maximal idle set of nodes (which may be empty), which we denote by $\mathrm{idle}(\mathcal{H})$. The complement of the largest idle set is called the *core* of $\mathcal{H}$ and denoted $\mathrm{core}(\mathcal{H})$. Directly from the definition, it follows that $\mathrm{idle}(\mathcal{H})$ is a union of connected components of $\mathcal{H}$. Therefore, the core is also a union of connected components. The class of core-components, that is, components in $\mathcal{H}$ contained in $\mathrm{core}(\mathcal{H})$ is denoted by $\mathcal{T}_{\mathrm{core}}(\mathcal{H})$.

Lemma 5.4. *If $\langle \mathcal{H} \rangle$ is a NF-CG, then* $\mathrm{idle}(\mathcal{H}^*)$ *forms a clique, that is, all its nodes are connected by an undirected edge.*

Proof. Because there is a directed arrow from any node outside $\mathrm{idle}(\mathcal{H}^*)$ to any node in $\mathrm{idle}(\mathcal{H}^*)$, every component of $\mathcal{H}^*$ lies either inside or outside of $\mathrm{idle}(\mathcal{H}^*)$. Since all nodes in $\mathrm{idle}(\mathcal{H}^*)$ are linked, there is a meta-arrow between any two distinct components of $\mathrm{idle}(\mathcal{H}^*)$ and each component is a clique. Without loss of generality, pick T such that T' is the only child-component of T. First, note that $p^*(T') \cap T = T$ forms a clique. Second, the parent-components of T' are $T \cup p^*(T)$. Indeed, if a component S, such that $S \Rightarrow T'$, lies outside of $\mathrm{idle}(\mathcal{H}^*)$, then $S \subseteq p^*(T)$ by definition. If $S \subseteq \mathrm{idle}(\mathcal{H}^*)$, then $S \subseteq p^*(T)$ because S and T are necessarily linked and T has no other children than T'. Thus, by Definition 2.7, T and T' can be legally merged, which contradicts the fact that $\mathcal{H}^*$ is essential. □

Note that idle($\mathcal{H}^*$) is precisely the set of vertices i such that $\downarrow i = [m]$, where $\downarrow i = \{j : N^*(i) \subseteq N^*(j)\}$.

From now on, $\mathcal{H}$ will always denote a DAG. By Theorem 5.1, each component $T \in \mathcal{T}_{\text{core}}(\mathcal{H}^*)$ induces a decomposable graph $\mathcal{H}^*_T$. We recall that a decomposable graph is an undirected graph with no induced cycles of size ≥ 4. An alternative definition, that will be useful in this section, is that its maximal cliques can be ordered into a sequence $C_1, \ldots, C_p$ satisfying the *running intersection property* (see [15], Proposition 2.17), that is

$$\forall i \geq 2\ \exists k < i, \qquad S_i = C_i \cap \left(\bigcup_{j<i} C_j\right) \subseteq C_k. \tag{5.2}$$

By [23], Lemma 7.2, the collection of sets S_i for $2 \leq i \leq m$ does not depend on the choice of ordering that satisfies (5.2). We call these sets *separators* of the graph. The multiplicity $\nu(S)$ of a separator S is then defined as the number of indices i such that $S_i = S$. This number also does not depend on the choice of an ordering that satisfies (5.2).

By $\mathcal{C}(T)$, denote the collection of maximal cliques of $\mathcal{H}^*_T$, by $\mathcal{S}(T)$ the collection of its separators and by $\nu_T(S)$ the multiplicity of $S \in \mathcal{S}(T)$ in $\mathcal{H}^*_T$. A set $P \subseteq [m]$ is called a *parent set* in $\mathcal{H}^*$ if it is non-empty and there exists a component $T \in \mathcal{T}_{\text{core}}(\mathcal{H}^*)$ with $P = p_{\mathcal{H}^*}(T)$. The multiplicity $\tau(P)$ of P is the number of $T \in \mathcal{T}_{\text{core}}(\mathcal{H}^*)$ with $P = p_{\mathcal{H}^*}(T)$. The collection of all parent sets in $\mathcal{H}^*$ is denoted by $\mathcal{P}_{\text{core}}(\mathcal{H}^*)$. Finally, by $i(\mathcal{H}^*)$ we denote the number of initial components of $\mathcal{H}^*$, that is, the components $T \in \mathcal{T}_{\text{core}}(\mathcal{H}^*)$ such that $p_{\mathcal{H}^*}(T) = \varnothing$.

We refer for the following result to [25], Lemma 5.1.

Lemma 5.5. *Let $\mathcal{H}^*$ be the essential graph of a DAG $\mathcal{H}$. If* $\text{core}(\mathcal{H}^*) = \varnothing$, *then* $u_{\mathcal{H}} = 0$. *If* $\text{core}(\mathcal{H}^*) \neq \varnothing$, *then the standard imset for $\mathcal{H}$ has the form*

$$u_{\mathcal{H}} = \delta_{\text{core}(\mathcal{H}^*)} - \sum_{T \in \mathcal{T}_{\text{core}}(\mathcal{H}^*)} \sum_{C \in \mathcal{C}(T)} \delta_{C \cup p_{\mathcal{H}^*}(T)} + \sum_{T \in \mathcal{T}_{\text{core}}(\mathcal{H}^*)} \sum_{S \in \mathcal{S}(T)} \nu_T(S) \delta_{S \cup p_{\mathcal{H}^*}(T)} +$$
$$+ \sum_{P \in \mathcal{P}_{\text{core}}(\mathcal{H}^*)} \tau(P) \delta_P + \big(i\big(\mathcal{H}^*\big) - 1\big)\delta_{\varnothing}.$$

By Lemma 5.2 in [25], unless $\mathcal{H}^*$ is a complete graph, the terms in the above formula never cancel each other. In particular, the support of $u_{\mathcal{H}}$ is the collection of all sets of the form:

(i) the core of $\mathcal{H}^*$,
(ii) $C \cup p^*(T)$ for $T \in \mathcal{T}_{\text{core}}(\mathcal{H}^*)$ and $C \in \mathcal{C}(T)$,
(iii) $S \cup p^*(T)$ for $T \in \mathcal{T}_{\text{core}}(\mathcal{H}^*)$ and $S \in \mathcal{S}(T)$,
(iv) P for $P \in \mathcal{P}_{\text{core}}(\mathcal{H}^*)$.

The empty set may or may not appear in the support set of $u_{\mathcal{H}}$ but this does not play any role in the following arguments.

Proposition 5.6. *Let $\mathcal{H}$ be a DAG. Then $N^*(i) \subseteq N^*(j)$ if and only if $i \in A$ implies $j \in A$ for every A in the support of $u_{\mathcal{H}}$.*

Proof. Lemma 5.5 gives the support of $u_{\mathcal{H}}$ in terms of $\mathcal{H}^*$, see also items (i)–(iv) above. For the forward direction, first note that if $i \in C$, then $j \in C \cup p^*(T)$, which follows immediately from $i \in N^*(j)$. This implies that, if i lies in the core, then j also lies in the core. Suppose now that $i \in C \cup p^*(T)$ for some $T \in \mathcal{T}_{\text{core}}(\mathcal{H}^*)$ and $C \in \mathcal{C}(T)$. If $i \in C$, then we have just shown that $j \in C \cup p^*(T)$. If $i \in p^*(T)$, then $j \in p^*(T)$ because $c^*(i) \subseteq c^*(j)$. The arguments for the subsets of type (iii) and (iv) above are the same.

For the opposite direction, first note that if $i \in A$ implies $j \in A$ for all A in the support of $u_{\mathcal{H}}$, then taking $A = C \cup p^*(T)$ where T is the connected component of i and $C \in \mathcal{C}(T)$ we find that either $i - j$ or $j \to i$, and hence $i \in N^*(j)$. Let $k \in n^*(i) \cup c^*(i)$. Suppose first that $i - j$. If $k \in n^*(i)$, then $k \in n^*(j)$. To see that take any $C \cup p^*(T)$ such that $i, k \in C$, which implies that $j \in C$. Similarly, if $k \in c^*(i)$, then $k \in c^*(j)$, which follows by considering P a parent set of the component containing k. Consequently, $N^*(i) \subseteq N^*(j)$. The case $j \to i$ is similar. □

Proposition 5.6 gives an efficient procedure of checking when $N^*(i) \subseteq N^*(j)$ without constructing the essential graph $\mathcal{H}^*$, which gives the description of G^0. We present this procedure in the pseudo code below (Algorithm 1).

In addition, note that the size of the support set of $u_{\mathcal{H}^*}$ is $\leq 2m$. The fact that it is $\leq 2m + 2$ is obvious from (5.1). But also any initial vertex i in $\mathcal{H}$ will have $p_{\mathcal{H}}(i) = \varnothing$, and hence $-\delta_{\varnothing}$ and $\delta_{p_{\mathcal{H}}(i)}$ will cancel each other. It follows that the number of operations to build or construct G^0 is quadratic in m. In fact, all loops are linear in $m + |E|$ apart from the penultimate one.

The imset $u_{\mathcal{H}}$ gives, in fact, the complete description of the group G.

Data: a DAG $\mathcal{H} = ([m], E)$
Result: the set of pairs (i, j) such that $N^*(i) \subseteq N^*(j)$
initialization **for** $i \to j$ **in** $\mathcal{H}$ **do**
| add i to $p_{\mathcal{H}}(j)$
end
$u_{\mathcal{H}}(\varnothing) := -1$, $u_{\mathcal{H}}([m]) := 1$, $\mathbf{S} = \varnothing$ **for** $i = 1$ **to** m **do**
| $++u_{\mathcal{H}}(p_{\mathcal{H}}(i)), --u_{\mathcal{H}}(p_{\mathcal{H}}(i) \cup i)$ add $\{p_{\mathcal{H}}(i)\}$ and $\{p_{\mathcal{H}}(i) \cup i\}$ to $\mathbf{S}$
end
forall the *elements* S *of* $\mathbf{S}$ **do**
| if $u_{\mathcal{H}}(S) = 0$, then remove S from $\mathbf{S}$
end
for $i = 1$ **to** m **do**
| $\mathcal{E}_i := \{S \in \mathbf{S} : i \in S\}$
end
for $i \cdots j \in E$ **do**
| $N^*(i) \subseteq N^*(j)$ if and only if $\mathcal{E}_i \subseteq \mathcal{E}_j$
end

Algorithm 1: The computation of G^0 for a DAG $\mathcal{H}$

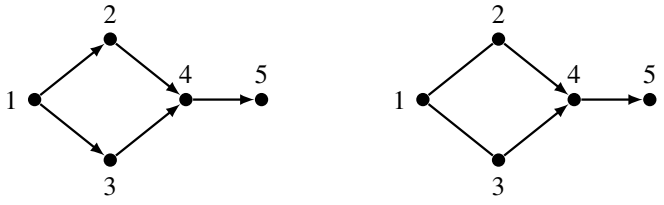

Figure 5. The sprinkle graph on the left and its essential graph on the right.

Lemma 5.7. *Let σ be a permutation. Then $\sigma \in G$ if and only if $u_{\mathcal{H}} = \sigma(u_{\mathcal{H}})$, where*

$$\sigma(u_{\mathcal{H}})(S) = u_{\mathcal{H}}\big(\sigma^{-1}(S)\big).$$

Consequently, by Theorem 1.4 we obtain the complete structure of G.

Proof. This follows from the fact that $u_{\mathcal{H}}$ is in a one-to-one correspondence with a DAG model of $\mathcal{H}$. □

Lemma 5.7 does not provide an efficient algorithm to find the automorphism group of $\mathcal{H}^*$, which in general is a hard problem.

6. Special graphs and small examples

Some DAG models are equivalent to undirected graphical models, in which case we refer to [6], Section 7, for examples. To obtain a new set of examples, we first consider two simple DAGs: the *sprinkle graph* in Figure 5 and the *Verma graph* in Figure 6.

The essential graph of the sprinkle graph is also given in Figure 5. There are no non-trivial equivalence classes and, therefore, $\widetilde{\mathcal{H}}^* = \mathcal{H}^*$. The only non-trivial relation between neighbouring sets is $N^*(5) \subset N^*(4)$, so the matrices in G^0 have only one non-zero off-diagonal element on position $(5, 4)$. The group of automorphisms of $\mathcal{H}^*$ has only one non-trivial element which

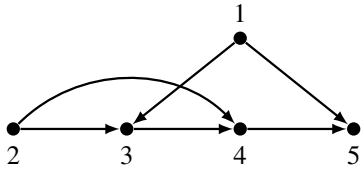

Figure 6. The Verma graph.

permutes 2 and 3. Hence, matrices in G are in either of the two following forms:

$$\begin{bmatrix} * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & * & * \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} * & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & * & * \end{bmatrix}.$$

The essential graph of the Verma graph $\mathcal{H}$ is equal to the Verma graph itself. All equivalence classes are singletons. Moreover, we have $N(3) \subseteq N(2)$, $N(5) \subseteq N(4)$ and $N(5) \subseteq N(1)$, and hence G^0 consists of all invertible matrices of the form

$$\begin{bmatrix} * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ 0 & * & * & 0 & 0 \\ 0 & 0 & 0 & * & 0 \\ * & 0 & 0 & * & * \end{bmatrix}.$$

Since the Verma graph does not admit any non-trivial automorphisms, we have $G = G^0$.

For another important class of examples, consider the DAGs defining factor models as given in Figure 7. This graph is essential and we have $N^*(b_i) \subset N^*(a_i)$ for every i, j and there are no other containment relations. The only non-zero off-diagonal elements of matrices in G^0 are in position (b_i, a_j) for all i, j. For example, if $p = 2$ and $q = 3$, then they are of the form

$$\left[\begin{array}{cc|ccc} * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ \hline * & * & * & 0 & 0 \\ & * & 0 & * & 0 \\ & * & 0 & 0 & * \end{array}\right].$$

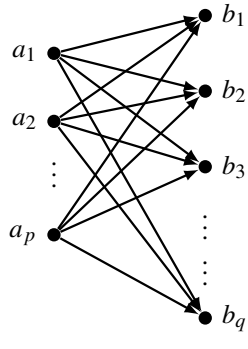

Figure 7. The graph of the factor model.

Any automorphisms of $\mathcal{H}$ is a product of any permutation permuting $\{a_1, \ldots, a_p\}$ and any permutation permuting $\{b_1, \ldots, b_q\}$. Consequently, all matrices in G look like the matrices in G^0 where the two diagonal blocks are replaced by arbitrary monomial matrices.

7. Conclusions

In this paper, we fully characterised the maximal group of linear transformations which stabilises a given Gaussian chain graph model for a chain graph with no flags. Following our earlier work on undirected models (see [6]), we provided basic statistical applications of this result. This includes equivariant estimation, maximal invariants for hypothesis testing and robustness. An example of how it can be used in practice was given in [27] in the case of undirected star tree models. In this particular example, the corresponding group G acts transitively on the model, meaning that each covariance in the model is of the form gg^T for some $g \in G$. This makes the whole analysis relatively simple. Our work is the first, necessary step toward such an analysis for non-transitive cases.

In Section 3, we derive simple necessary and sufficient conditions on vanishing sub-minors of the concentration matrix in the model. This provides an alternative description of Gaussian chain graph models in terms of cup systems that are a bit easier to study than the trek systems.

Our technique for computing the group G requires finding the essential graph of the given chain graph. However, by applying Stúdeny's theory of imsets (see [23]), we show that computations for the special case of DAGs can be performed efficiently without first building the essential graph. This raises a natural question about how this technique extends to other Gaussian conditional independence models. We leave this for our future study.

Appendix: Proof of Theorem 4.4

To prove this theorem, we will use the following two lemmas, in which K is the concentration matrix of the model.

Lemma A.1. *Let A, B be subsets of $[m]$ of the same cardinality satisfying $j \in A$ and $i \notin A$ and either $j \notin B$ or else both $i, j \in B$. If $\det K[A, B]$ is identically zero on the model but $\det K[A - j + i, B]$ is not, then $E_{ij} \notin \mathfrak{g}$.*

Proof. Recall that the one-parameter group $I + tE_{ij}$ acts on K via

$$K \mapsto (I - tE_{ji})K(I - tE_{ij}).$$

In other words, this matrix is obtained from K by adding a multiple of the ith row to the jth row and adding a multiple of the ith column to the jth column. Now consider the effect of this operation on $K[A, B]$. Since either $j \notin B$ or else both $i, j \in B$, adding the ith column to the jth has either no effect on $K[A, B]$ or else is just an elementary column operation on $K[A, B]$. This means that it does not affect the rank of $K[A, B]$. On the other hand, since $\det K[A - j + i, B]$ is non-zero, the rows of $K[A - j, B]$ are linearly independent, and since $\det K[A, B]$ is zero,

the jth row $K[j,B]$ lies in the span of the rows of $K[A-j,B]$. This is not true for the ith row $K[i,B]$, hence the $A\times B$-submatrix of $K+tE_{ji}K+tKE_{ij}$ has full rank for generic K. This means that $I+tE_{ij}$ does not preserve the model, hence it does not lie in the group G. □

Lemma A.2. *Let A,B be subsets of $[m]$ of the same cardinality satisfying $j\in A\cap B$ and $i\notin A\cup B$. If* $\det K[A,B]$ *is identically zero but* $\det K[A-j+i,B]+\det K[A,B-j+i]$ *is not, then $E_{ij}\notin\mathfrak{g}$.*

Proof. Since $K[A,B]=0$, $E_{ij}\in\mathfrak{g}$ only if the determinant of the (A,B)-submatrix of $K_t:=(I-tE_{ji})K(I-tE_{ij})$ is zero. To show that it is not zero, it suffices to show that the linear term of t does not vanish. To study this linear term, we alternatively study the linear term of $(I-sE_{ji})K(I-tE_{ij})$ further specializing to $s=t$. Because E_{ij} has rank 1, the determinant of the (A,B)-submatrix of $(I-sE_{ji})K(I-tE_{ij})$ is a polynomial of order two in s,t. To find its coefficient of the linear term s, we can set $t=0$. Matrix $(I-sE_{ji})K$ is obtained by adding a multiple of the ith row to the jth row. Suppose that the elements of A are $a_1<a_2<\cdots<a_d$ and the elements of B are $b_1<b_2<\cdots<b_d$. Let $1\le k\le d$ be such that $j=a_k$. The determinant if its (A,B)-submatrix can be computed by expanding along the kth row (which corresponds to the jth row of K):

$$\begin{aligned}\det\bigl((I-sE_{ji})K\bigr)[A,B]&=\sum_{l=1}^{d}(-1)^{k+l}(K_{jb_l}-sK_{ib_l})\det K[A-j,B-b_l]\\&=\det K[A,B]-s\det K[A-j+i,B].\end{aligned}$$

Similar computations for the coefficient of t give

$$\det\bigl(K(I-tE_{ij})\bigr)[A,B]=\det K[A,B]-t\det K[A,B-j+i].$$

Hence, the coefficient of t in the determinant of $K_t[A,B]$ is $-\det K[A-j+i,B]-\det K[A,B-j+i]$. If this sum does not identically vanish on the model, then $E_{ij}\notin\mathfrak{g}$. □

Lemma 4.3 gives one direction of the proof of Theorem 4.4; we need only prove that if $i\neq j$ and $N_{\mathcal{H}}(i)\not\subseteq N_{\mathcal{H}}(j)$, then $E_{ij}\notin\mathfrak{g}$. First of all, if there is no cup from j to i, then $K[j,i]$ is identically zero, while $K[i,i]$ is not. Hence, $E_{ij}\notin\mathfrak{g}$ (this is the special case of Lemma A.1 with $A=\{j\}$ and $B=\{i\}$). Thus, in what follows, we may assume that there do exist cups from j to i. We treat the various types of cups from j to i separately; in each case, we assume that cups of the previous types do not exist. Before we get going, we remark that, since there are no flags, for any cup (f,h,k,l) with $f\to h$ also (f,k,k,l) is a cup. The following lemma will be also useful.

Lemma A.3. *Let u be a vertex in a NF-CG $\mathcal{H}$. Let D be the set of children of u together with all their descendants. Then for every vertex $v\notin D\cup\{u\}$ such that there is no link between u and v we have* $\det K[D\cup\{u\},D\cup\{v\}]=0$.

Proof. By Proposition 3.5, it is enough to show that there is no self-avoiding cup system from $D \cup \{u\}$ to $D \cup \{v\}$. It is clear that the second element of every cup starting in $d \in D$ needs to lie in D just because it is either equal to d or it is equal to d' such that $d \to d'$ in $\mathcal{H}$. Also every cup from u needs to have its second entry in D. Indeed, let (u, l_2, l_3, l_4) be such a cup. The node l_2 is either equal to u or it is a child of u, in which case it lies in D. So suppose that $l_2 = u$ and show that this leads to a contradiction. If $l_2 = u$, then l_3 is either u or a neighbour of u. If $l_3 = u$, then l_4 must be a parent of u, which cannot be a vertex of D (because otherwise there is a semi-directed cycle in $\mathcal{H}$) and it cannot be v because there is no arrow $v \to u$ (by assumption). If $l_3 \in n_{\mathcal{H}}(u)$, then l_4 must be a parent of l_3 and by the no flag assumption also a parent of u. This situation is also impossible because l_4 cannot lie in $D \cup \{v\}$. Hence, by the pigeon-hole principle, in any cup system from $D \cup \{u\}$ to $D \cup \{v\}$, two of the elements after one step coincide, and this proves the claim. □

In what follows, we assume that $\mathcal{H}$ is essential.

I. Vertex i lies in $n_{\mathcal{H}}(j) \cup c_{\mathcal{H}}(j)$

In that case, there must exist

$$l \in \big(n_{\mathcal{H}}(i) \cup c_{\mathcal{H}}(i)\big) \setminus \big(n_{\mathcal{H}}(j) \cup c_{\mathcal{H}}(j)\big).$$

Let D denote the set of all children of l together with their descendants. We have $i, j \notin D$ and thus $A := D + j$ and $B := D + l$ have the same cardinalities. By Lemma A.3 with $u = l$, $v = j$, we have $\det K[A, B] = 0$. On the other hand, there does exist a self-avoiding cup system from $A - j + i$ to B that links i directly to l without crossing D and each $d \in D - j$ to itself via (d, d, d, d), and hence $\det K[A - j + i, B] \neq 0$ by Proposition 3.5. Now $E_{ij} \notin \mathfrak{g}$ by Lemma A.1.

II. There is no arrow $i \to j$

In that case, let D be the set of all children of i together with their descendants. Set $A := D + j$ and $B := D + i$. By Lemma A.3, $\det K[A, B] = 0$. But clearly, $\det K[A - j + i, B] = \det K[B, B] \neq 0$.

Mid-proof break

We pause a moment to point out that we have used that $\mathcal{H}$ has no flags, but not yet that it is essential. This will be exploited in the following arguments. Indeed, in the remaining cases, there must be an arrow $i \to j$. This arrow must be essential, hence either the parents of j in the undirected component T of i do not form a clique, or else one of $\{i, j\}$ has a parent outside T that is not a parent of the other. We deal with these cases as follows.

III. There is an arrow $k \to j$ with k in the component of i at distance at least 2

In that case, let D be the set of all children of i together with their descendants. Set $A := D + k$ and $B := D + i$. By Lemma A.3, $\det K[A, B] = 0$. But, as in the first case, $\det K[A - j + i, B] \neq 0$ because there is a self-avoiding cup system from $A - j + i$ to B given by (d, d, d, d) for $d \in D - j + i$ and (j, j, j, k). Again, we conclude that $E_{ij} \notin \mathfrak{g}$.

IV. There is an induced subgraph like in Figure A.1(a)

Let D be the set of all children of k together with their descendants. Set $A = D + k$ and $B = D + l$ and note that both A and B contain j. We again have $\det K[A, B] = 0$ by Lemma A.3. However, both $\det K[A - j + i, B]$ and $\det K[A, B - j + i]$ are non-zero. Even more: the sum of these two determinants is also non-zero because $\det K[A - j + i, B]$ has a monomial that does not appear in $\det K[A, B - j + i]$: consider the cup system from $A - j + i$ to B given by (i, i, l, l), (k, j, j, j) and (d, d, d, d) for all $d \in D - j$. By Proposition 3.4, this system corresponds to a monomial in $\det K[A - j + i, B]$. On the other hand, this monomial cannot appear in $\det K[A, B - j + i]$ because it contains only one element of Λ, namely λ_{kj}, and only one off-diagonal element of Ω, namely ω_{il}. This means that it must correspond to a cup system between A and $B - j + i$ that contains only one undirected edge $i - l$ and one arrow $k \to j$. However, any cup from i to A must contain either an arrow $i \to p$ for some $p \in D$ or an undirected edge $i - k$. By Lemma A.2, we conclude that $E_{ij} \notin \mathfrak{g}$.

V. There is an arrow $k \to j$ with $k \notin T$ and no arrow between k and i

So we have the induced subgraph $i \to j \leftarrow k$. Let D be the set of all children of i together with their descendants. Set $A := D + k$ and $B := D + i$. By Lemma A.3, $\det K[A, B] = 0$. On the other hand, $\det K[A - j + i, B] \neq 0$, because of the self-avoiding cup system from $A - j + i$ to

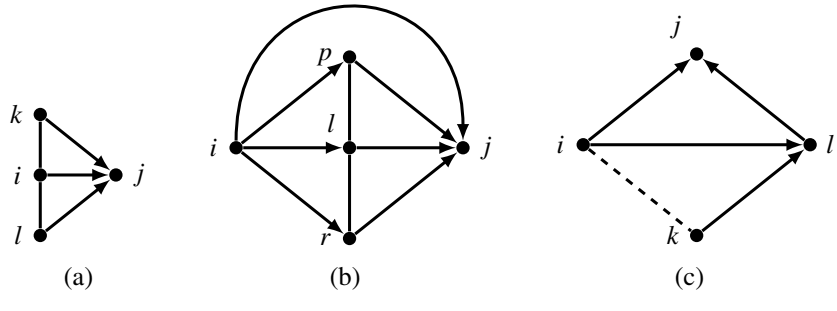

Figure A.1. Some special induced subgraphs considered in the proof.

B consisting of (k, j, j, j) and (i, i, i, i) and (d, d, d, d) for all $d \in D - j$. Again, we may apply Lemma A.1, this time with i, j both in B, to conclude that $E_{ij} \notin \mathfrak{g}$.

VI. There is an arrow $l \to i$ with no arrow from l to j

Pictorially, we have $l \to i \to j$. Let D be the set of children of j together with all their descendants. Set $A = D + j$ and $B = D + l$. By Lemma A.3, we have $K[A, B] = 0$. However, $K[A - j + i, B] \neq 0$, and hence $E_{ij} \notin \mathfrak{g}$ by Lemma A.1.

VII. There is an arrow $i \to l$ and $l \to j$

Without loss of generality, we can assume that l is minimal in the sense that if $i \to l'$ and $l' \to j$, then there is no arrow from l to l'. Since $\mathcal{H}$ is essential, $l \to j$ is an essential arrow. This implies one of the following possibilities:

(i) There exists k in the component of l with distance at least two to l and with $k \to j$.
(ii) There is an induced subgraph like in Figure A.1(b).
(iii) There are arrows $l \to k$, $k \to j$.
(iv) There is an arrow $k \to l$ and no arrow from k to j.

VII(i)

In this case, we have an induced subgraph $k \to j \leftarrow l$. Let D be the set of children of l and all their descendants. Set $A = D + l$ and $B = D + k$. The argument that $\det K[A, B] = 0$ is the same as in the previous cases. By Lemma A.2, $E_{ij} \notin \mathfrak{g}$ because $\det K[A - j + i, B] + \det K[A, B - j + i] \neq 0$. To verify this last statement, note that by Proposition 3.4 $\det K[A - j + i, B]$ contains a monomial corresponding to the cup system (d, d, d, d) for $d \in D - j$, (i, k, k, k) and (l, j, j, j). This monomial contains λ_{ki}, λ_{lj} and no off-diagonal ω's. There is no cup system from A to $B - j + i$ that uses only $k \to j$ and $l \to j$, and hence this monomial does not appear in $\det K[A, B - j + i]$.

VII(ii)

Let D be the set of children of p and all their descendants. Set $A = D + p$ and $B = D + r$. Again $\det K[A, B] = 0$ but $\det K[A - j + i, B] + \det K[A, B - j + i] \neq 0$. For this, we note that $\det K[A - j + i, B]$ contains a monomial corresponding to the cup system (d, d, d, d) for $d \in D - j$, (i, r, r, r) and (p, j, j, j), which does not appear in $\det K[A, B - j + i]$. Now $E_{ij} \notin \mathfrak{g}$ by Lemma A.2.

VII(iii)

Note that in this case no link between i and k is possible (by maximality of l and no semi-directed cycle assumption). But then $E_{ij} \notin \mathfrak{g}$ by Case V.

VII(iv)

Note that in this case by case VI. the arrow $k \to i$ is impossible, and thus we have either $i - k$, $i \to k$ are there is no link between them. The induced subgraph is given in Figure A.1(c), where the dashed edge indicate the three possibilities for the link between i and k. Let D be the set of children of j together with all descendants. Set $A = D + j$, $B = D + k$. Again by Lemma A.3, we have that $\det K[A, B] = 0$. Moreover, $\det K[A - j + i, B] \neq 0$. Now $E_{ij} \notin \mathfrak{g}$ by Lemma A.1. This exhausts all possible cases and hence completes the proof.

Acknowledgements

Both authors acknowledge support of Draisma's Vidi grant from the Netherlands Organisation for Scientific Research (NWO) and hospitality of the Simons Institute for the Theory of Computing in Berkeley during the Fall 2014 program *Algorithms and Complexity in Algebraic Geometry*. PZ was also supported from the European Union Seventh Framework Programme (PIOF-GA-2011-300975). The authors are grateful to Mathias Drton and the referees for helpful remarks.

References

[1] Andersson, S.A., Madigan, D. and Perlman, M.D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* **25** 505–541. MR1439312
[2] Andersson, S.A., Madigan, D. and Perlman, M.D. (2001). Alternative Markov properties for chain graphs. *Scand. J. Stat.* **28** 33–85. MR1844349
[3] Andersson, S.A. and Perlman, M.D. (2006). Characterizing Markov equivalence classes for AMP chain graph models. *Ann. Statist.* **34** 939–972. MR2283399
[4] Barndorff-Nielsen, O., Blæsild, P., Jensen, J.L. and Jørgensen, B. (1982). Exponential transformation models. *Proc. Roy. Soc. London Ser. A* **379** 41–65. MR0643215
[5] Berrington, A., Hu, Y., Smith, P.W.F. and Sturgis, P. (2008). A graphical chain model for reciprocal relationships between women's gender role attitudes and labour force participation. *J. Roy. Statist. Soc. Ser. A* **171** 89–108. MR2412648
[6] Draisma, J., Kuhnt, S. and Zwiernik, P. (2013). Groups acting on Gaussian graphical models. *Ann. Statist.* **41** 1944–1969. MR3127854
[7] Draisma, J., Sullivant, S. and Talaska, K. (2013). Positivity for Gaussian graphical models. *Adv. in Appl. Math.* **50** 661–674. MR3044565
[8] Drton, M., Sturmfels, B. and Sullivant, S. (2009). *Lectures on Algebraic Statistics*. *Oberwolfach Seminars* **39**. Basel: Birkhäuser. MR2723140
[9] Drton, M. and Xiao, H. (2010). Smoothness of Gaussian conditional independence models. *Contemp. Math.* **516** 155–177.
[10] Eaton, M.L. (1989). *Group Invariance Applications in Statistics*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **1**. Hayward, CA: IMS. MR1089423
[11] Ferrándiz, J., Castillo, E.F. and Sanmartín, P. (2005). Temporal aggregation in chain graph models. *J. Statist. Plann. Inference* **133** 69–93. MR2162568
[12] Frydenberg, M. (1990). The chain graph Markov property. *Scand. J. Stat.* **17** 333–353. MR1096723
[13] Gessel, I.M. and Viennot, X.G. (1989). Determinants, paths, and plane partitions. Tech. report, Brandeis Univ.

[14] Gross, E. and Sullivant, S. (2014). The maximum likelihood threshold of a graph. Preprint. Available at arXiv:1404.6989.
[15] Lauritzen, S.L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. New York: The Clarendon Press, Oxford Univ. Press. MR1419991
[16] Lauritzen, S.L. and Richardson, T.S. (2002). Chain graph models and their causal interpretations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 321–361. MR1924296
[17] Lauritzen, S.L. and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31–57. MR0981437
[18] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. New York: Springer. MR2135927
[19] Lněnička, R. and Matúš, F. (2007). On Gaussian conditional independent structures. *Kybernetika* (*Prague*) **43** 327–342. MR2362722
[20] Milan Studenỳ, Hemmecke, R. and Lindner, S. (2010). Characteristic imset: A simple algebraic representative of a Bayesian network structure. In *Proceedings of the* 5*th European Workshop on Probabilistic Graphical Models* (P. Myllymäki, T. Roos and T. Jaakkola, eds.) 257–264. Helsinki Institute for Information Technology HIIT, Helsinki, Finland.
[21] Roverato, A. (2005). A unified approach to the characterization of equivalence classes of DAGs, chain graphs with no flags and chain graphs. *Scand. J. Stat.* **32** 295–312. MR2188675
[22] Studený, M. (2004). Characterization of essential graphs by means of the operation of legal merging of components: New trends in probabilistic graphical models. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems* **12** 43–62. MR2058946
[23] Studený, M. (2005). *Probabilistic Conditional Independence Structures*. *Information Science and Statistics*. London: Springer. MR3183760
[24] Studený, M., Roverato, A. and Štěpánová, Š. (2009). Two operations of merging and splitting components in a chain graph. *Kybernetika* (*Prague*) **45** 208–248. MR2518149
[25] Studený, M. and Vomlel, J. (2009). A reconstruction algorithm for the essential graph. *Internat. J. Approx. Reason.* **50** 385–413. MR2514506
[26] Sullivant, S., Talaska, K. and Draisma, J. (2010). Trek separation for Gaussian graphical models. *Ann. Statist.* **38** 1665–1685. MR2662356
[27] Sun, D. and Sun, X. (2005). Estimation of the multivariate normal precision and covariance matrices in a star-shape model. *Ann. Inst. Statist. Math.* **57** 455–484. MR2206534
[28] Uhler, C. (2012). Geometry of maximum likelihood estimation in Gaussian graphical models. *Ann. Statist.* **40** 238–261. MR3014306
[29] Verma, T.S. and Pearl, J. (1990). Equivalence and Synthesis of Causal Models. In *UAI* '90: *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence* (P.P. Bonissone, M. Henrion, L.N. Kanal and J.F. Lemmer, eds.) 27–29. Cambridge, MA: MIT.
[30] Wan Norsiah Mohamed, Diamond, I. and Smith, P.W.F. (1998). The determinants of infant mortality in Malaysia: A graphical chain modelling approach. *J. Roy. Statist. Soc. Ser. A* **161** 349–366.

Received January 2015 and revised September 2015