The Annals of Statistics
2013, Vol. 41, No. 4, 1944–1969
DOI: 10.1214/13-AOS1130

GROUPS ACTING ON GAUSSIAN GRAPHICAL MODELS

BY JAN DRAISMA[1], SONJA KUHNT AND PIOTR ZWIERNIK[1]

TU Eindhoven, TU Dortmund University and CWI Amsterdam

Gaussian graphical models have become a well-recognized tool for the analysis of conditional independencies within a set of continuous random variables. From an inferential point of view, it is important to realize that they are composite exponential transformation families. We reveal this structure by explicitly describing, for any undirected graph, the (maximal) matrix group acting on the space of concentration matrices in the model. The continuous part of this group is captured by a poset naturally associated to the graph, while automorphisms of the graph account for the discrete part of the group. We compute the dimension of the space of orbits of this group on concentration matrices, in terms of the combinatorics of the graph; and for dimension zero we recover the characterization by Letac and Massam of models that are transformation families. Furthermore, we describe the maximal invariant of this group on the sample space, and we give a sharp lower bound on the sample size needed for the existence of equivariant estimators of the concentration matrix. Finally, we address the issue of robustness of these estimators by computing upper bounds on finite sample breakdown points.

1. Introduction and results. Gaussian graphical models are popular tools for modelling complex associations in the multivariate continuous case. If the graph with vertex set $[m] := \{1, \dots, m\}$ is complete, then the general linear group $\mathrm{GL}_m(\mathbb{R})$, consisting of all invertible $m \times m$-matrices, acts on the space of concentration matrices in the model, as well as on the sample space. The maximum likelihood estimator (MLE) of the concentration matrix is equivariant with respect to this group action, but many other equivariant estimators have been proposed, for example, by Anderson (2003), Donoho (1982), James and Stein (1961), Lopuhaä and Rousseeuw (1991), Stahel (1981). For smaller graphs, only some proper subgroup of $\mathrm{GL}_m(\mathbb{R})$ will act on the set of compatible concentration matrices. In this paper, we describe that subgroup explicitly, and pave the way for its use in designing invariant tests, (robust) equivariant estimators and improved inference procedures.

Having an explicit group acting on a statistical model has numerous advantages. This was first pointed out by Fisher (1934) in the context of the location and scale

Received July 2012; revised May 2013.

[1]Supported by Draisma's Vidi grant from the Netherlands Organisation for Scientific Research (NWO).

MSC2010 subject classifications. Primary 62H99, 62F35; secondary 54H15.

Key words and phrases. Gaussian graphical models, covariance matrix, concentration matrix, robust estimator, breakdown point, equivariant estimator, transformation families.

models, which then led to the notion of a transformation family, that is, a statistical model on which a group acts transitively. Group actions give rise, for example, to the study of model invariants and distributional aspects of the maximum likelihood estimator (MLE) or other equivariant estimators [see Barndorff-Nielsen (1983), Barndorff-Nielsen et al. (1982), Eaton (1989), Fisher (1934), Lehmann and Romano (2005), Reid (1995)]. When a group acts on a model in a nontransitive manner, the model is sometimes called a composite transformation family [see Barndorff-Nielsen et al. (1982)]. In this case, the model can be decomposed into a family of transformation models each corresponding to a fixed value of some parameter.

To set the stage, let $\mathcal{G} = ([m], E)$ be an undirected graph with set of vertices $[m]$ and set of edges $E \subseteq \binom{[m]}{2}$. Denote by $\mathcal{S}_m$ the set of symmetric matrices in $\mathbb{R}^{m\times m}$ and by $\mathcal{S}_m^+ \subseteq \mathcal{S}_m$ the cone of positive definite matrices. Let $\mathcal{S}_\mathcal{G} \subseteq \mathcal{S}_m$ denote the linear space of symmetric matrices whose (i, j) off-diagonal entry is zero if $\{i, j\} \notin E$, and by $\mathcal{S}_\mathcal{G}^+$ the cone of all positive definite matrices in $\mathcal{S}_\mathcal{G}$. As a running example in this Introduction, we take $\mathcal{G}$ to be the path $P_3 : \overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}$. So $\mathcal{S}_{P_3}$ consists of all symmetric matrices of the form

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

Let $X = (X_i)_{i\in[m]}$ be a random vector with multivariate normal distribution $\mathcal{N}(0, \Sigma)$. The *Gaussian graphical model* is the statistical model for X given by

$$M(\mathcal{G}) := \{\mathcal{N}(0, \Sigma) \mid \Sigma^{-1} \in \mathcal{S}_\mathcal{G}^+\};$$

so $\mathcal{S}_\mathcal{G}^+$ is the space of concentration matrices compatible with the model [see Lauritzen (1996)].

The group $\mathrm{GL}_m(\mathbb{R})$ acts on $\mathbb{R}^m$ by matrix-vector multiplication, and this induces an action on $\mathcal{S}_m$ and $\mathcal{S}_m^+$ given by $g \cdot K := g^{-T} K g^{-1}$—indeed, note that this is the concentration matrix of gX if K is the concentration matrix of X.

A leading role in this paper is played by the group

$$G := \{g \in \mathrm{GL}_m(\mathbb{R}) \mid g \cdot \mathcal{S}_\mathcal{G}^+ \subseteq \mathcal{S}_\mathcal{G}^+\}.$$

This is a closed subgroup of the Lie group $\mathrm{GL}_m(\mathbb{R})$ (see Section 2). For example, if $\mathcal{G}$ is the complete graph, then G is all of $\mathrm{GL}_m(\mathbb{R})$. For *any* graph $\mathcal{G}$, the group G contains the invertible diagonal matrices, which correspond to scaling the components of X. Furthermore, G contains elements coming from graph automorphisms of $\mathcal{G}$. Specifically, if $\pi : [m] \to [m]$ is such an automorphism, then the permutation matrix g with ones on the positions $(i, \pi(i))$, $i \in [m]$ lies in G, since its action on $\mathcal{S}_m$ stabilizes the zero pattern prescribed by $\mathcal{G}$. For our running example P_3, the permutation matrix

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

lies in G.

1.1. *The group G.* Our first result is an explicit description of G in terms of $\mathcal{G}$, and requires the pre-order on $[m]$ defined by

$$i \preccurlyeq j \quad \text{if and only if} \quad N(j) \cup \{j\} \subseteq N(i) \cup \{i\},$$

where $N(i) = \{j \in [m] : \{i, j\} \in E\}$ denotes the set of neighbors of i in $\mathcal{G}$. So in our running example P_3 we have $1 \preccurlyeq 2, 3$. Consider the closed subset G^0 of $\mathrm{GL}_m(\mathbb{R})$ defined by

$$G^0 = \{g \in \mathrm{GL}_m(\mathbb{R}) \mid g_{ij} = 0 \text{ for all } j \not\preccurlyeq i\}.$$

We show in Section 2 that this set is a subgroup of $\mathrm{GL}_m(\mathbb{R})$. For $\mathcal{G} = P_3$, it consists of all invertible matrices of the form

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

THEOREM 1.1. *For any undirected graph $\mathcal{G} = ([m], E)$, the group G is generated by the group G^0 and the permutation matrices corresponding to the automorphism group of the graph $\mathcal{G}$.*

For P_3, this theorem says that G is the group of all matrices of the form above, together with all matrices of the form

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

The two subgroups of G in Theorem 1.1 can have a nontrivial intersection. For instance, when $\mathcal{G}$ is the complete graph, the automorphism group of $\mathcal{G}$ is contained in G^0. In Section 2, we state and prove a more refined statement that gets rid of that intersection.

1.2. *Existence and robustness of equivariant estimators.* Now that we know explicitly which matrix group G acts on our graphical model $M(\mathcal{G})$, we can use this group to develop classical notions of multivariate statistics in the general context of graphical models. One of these notions is that of an equivariant estimator [see, e.g., Eaton (1989), Schervish (1995)]. Let $\mathbf{X}$ denote the $m \times n$ matrix, whose columns correspond to n independent copies of the vector X. Then an equivariant estimator for the concentration matrix is a map $T : (\mathbb{R}^m)^n = \mathbb{R}^{m \times n} \to \mathcal{S}^+_{\mathcal{G}}$, that is, a map from the space of n-samples $\mathbf{X}$ to the parameter space of the model, that satisfies $T(g\mathbf{x}) = gT(\mathbf{x})$ for all realisations $\mathbf{x}$ of $\mathbf{X}$. The standard example is the maximum likelihood estimator (MLE). Indeed, the likelihood of concentration matrix K given an n-sample $\mathbf{x}$ equals the likelihood of $g \cdot K$ given $g\mathbf{x}$, for any $g \in G$, and this implies that the MLE is G-equivariant. Other equivariant estimators of the concentration matrix for some special graphical models have been proposed in Sun and Sun (2005).

For decomposable 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 n and a given graph $\mathcal{G}$ is a subtle matter; see the recent paper by Uhler (2012) and the references therein. By contrast, the question whether for a given sample size *any* equivariant estimator exists, turns out to have a remarkably elegant answer for any graph $\mathcal{G}$. To state it, define the *down set* $\downarrow i$ of an element $i \in [m]$ to be the set of all $j \in [m]$ with $j \preccurlyeq i$.

THEOREM 1.2. *Let $\mathcal{G} = ([m], E)$ be an undirected graph. There exists a G-equivariant estimator $T : \mathbb{R}^{m \times n} \to \mathcal{S}_{\mathcal{G}}^{+}$ if and only if $n \geq \max_{i \in [m]} |\downarrow i|$.*

To be precise, when n is at least the bound in the theorem, a G-equivariant T exists that is defined outside some measure-zero set (in fact, an algebraic subvariety of positive codimension), while if n is smaller than that bound, then not even any partially defined equivariant map T exists. For our running example P_3, we have $\downarrow 1 = \{1\}$ and $\downarrow 2 = \{1, 2\}$ and $\downarrow 3 = \{1, 3\}$, so Theorem 1.2 says that an equivariant estimator exists with probability one if and only if the sample size is at least 2, which in this case coincides with the condition for existence of the MLE.

Theorem 1.2 will be proved in Section 3, where we also establish upper bounds on the robustness of equivariant estimators, based on general theory from Davies and Gather (2005).

1.3. *The maximal invariant.* Another classical notion related to a group action on a statistical model is that of *invariants* on the sample space. In our case, these are maps τ defined on $\mathbb{R}^{m \times n}$, possibly outside some measure-zero set, that are constant on G-orbits, that is, that satisfy $\tau(g\mathbf{X}) = \tau(\mathbf{X})$ for all $g \in G$. An invariant τ is called *maximal* if it distinguishes all G-orbits. In formulas this means that for n-samples $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times n}$, outside some set of measure zero, the equality $\tau(\mathbf{x}) = \tau(\mathbf{y})$ implies that there exists a $g \in G$ such that $g\mathbf{x} = \mathbf{y}$. Any invariant map is then a function of τ.

The relevance of maximal invariants in statistics lies in the fact that they facilitate inference for the maximum likelihood estimator in the case of transformation families [see Barndorff-Nielsen (1983), Reid (1995)]. In this case the maximal invariant is an ancillary statistics that one may chose to condition on. These ideas can be used also in the case of composite transformation families, where the inference for the index parameter κ is based on the marginal distribution of the maximal invariant statistics [Barndorff-Nielsen (1983), Section 5].

Another important application of the maximal invariant is in the construction of *invariant tests* [see Eaton (1989), Lehmann and Romano (2005)]. Suppose, for instance, that we want to test the hypothesis that the distribution of the multivariate Gaussian random vector X lies in $M(\mathcal{G})$ against the alternative that it does not, and suppose that for the n-sample $\mathbf{X} = \mathbf{x}$ the test would accept the hypothesis. Then, since $M(\mathcal{G})$ 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.

Our result on maximal invariants uses the equivalence relation $\sim$ on $[m]$ defined by $i \sim j$ if and only if both $i \preccurlyeq j$ and $j \preccurlyeq i$, that is, if and only if $N(i) \cup \{i\} = N(j) \cup \{j\}$. We write $\bar{i}$ for the equivalence class of $i \in [m]$ and $[m]/\sim$ for the set of all equivalence classes.

THEOREM 1.3. *Let $\mathcal{G} = ([m], E)$ be an undirected graph. 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 a maximal G^0-invariant.

The lower bound on n in the theorem ensures that the $|\downarrow i| \times |\downarrow i|$-matrices $\mathbf{x}[\downarrow i]\mathbf{x}[\downarrow i]^T$ are invertible for generic $\mathbf{x}$, and in particular for $\mathbf{x}$ outside a set of measure zero. For the complete graph, Theorem 1.3 reduces to the known statement that $\mathbf{x} \mapsto \mathbf{x}^T(\mathbf{x}\mathbf{x}^T)^{-1}\mathbf{x}$ is a maximal invariant, see Example 6.2.3 in Lehmann and Romano (2005), while for our running example P_3 it says that the rank-one matrix $\mathbf{x}[1]^T(\mathbf{x}[1]\mathbf{x}[1]^T)^{-1}\mathbf{x}[1]$ (recording only the direction of the first row of $\mathbf{x}$) and the rank-two matrices $\mathbf{x}[1,2]^T(\mathbf{x}[1,2]\mathbf{x}[1,2]^T)^{-1}\mathbf{x}[1,2]$ and $\mathbf{x}[1,3]^T(\mathbf{x}[1,3]\mathbf{x}[1,3]^T)^{-1}\mathbf{x}[1,3]$ together form a maximal invariant for G^0.

We stress that Theorem 1.3 gives a maximal invariant under the subgroup G^0, rather than under all of G. The proof of this theorem can be found in Section 4.

1.4. *Orbits of G on $\mathcal{S}_{\mathcal{G}}^+$.* Our final results concern the space $\mathcal{S}_{\mathcal{G}}^+/G$ of G-orbits in $\mathcal{S}_{\mathcal{G}}^+$. When $M(\mathcal{G})$ is a transformation family, this space consists of a single point and hence has dimension zero. Conversely, it turns out that when the dimension of $\mathcal{S}_{\mathcal{G}}^+/G$ is zero, $M(\mathcal{G})$ is a transformation family. By work of Letac and Massam (2007), it is known exactly for which graphs this happens. Our result on $\mathcal{S}_{\mathcal{G}}^+/G$ is a combinatorial expression for its dimension. Rather than capturing that expression in a formula, which we will do in Section 6, we now describe it in terms of a combinatorial procedure.

Let $\mathcal{G} = ([m], E)$ be an undirected graph. Color an edge $\{i, j\} \in E$ red if $i \sim j$, green if $i \preccurlyeq j$ or $j \preccurlyeq i$ but not both, and blue otherwise. Next delete all green edges from $\mathcal{G}$, while retaining their vertices. Then delete the blue edges sequentially, in each step not only deleting a blue edge but also its two vertices together with all further blue and red edges incident to those two vertices. Continue this process

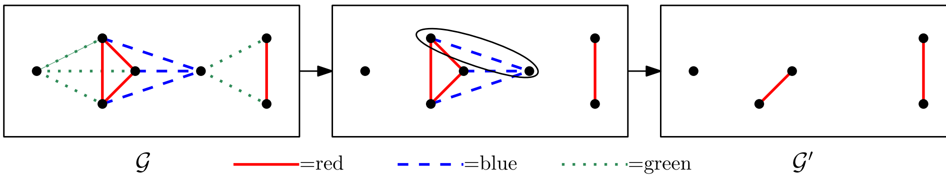


FIG. 1. *An example where the orbit space $\mathcal{S}_{\mathcal{G}}^{+}/G$ has dimension* 1.

until no blue edges are left. Call the resulting graph $\mathcal{G}'$; it consists of red edges only. See Figure 1 for an example. One can show that, up to isomorphism, $\mathcal{G}'$ is independent of the order in which the blue edges with incident vertices were removed—though in general it is larger than the graph obtained by deleting all blue edges, their vertices, and their incident edges at once.

THEOREM 1.4. *The dimension of $\mathcal{S}_{\mathcal{G}}^{+}/G$ equals the number of blue edges in the original graph $\mathcal{G}$ minus the number of red edges in $\mathcal{G}$ plus the number of remaining red edges in $\mathcal{G}'$.*

In other words, that dimension equals the number of blue edges in $\mathcal{G}$ minus the number of red edges deleted in the process going from $\mathcal{G}$ to $\mathcal{G}'$. This number is nonnegative: indeed, if in some step a blue edge $\{i, j\}$ is being deleted together with its vertices, then for each red edge $\{i, k\}$ being deleted along with i there is also a blue edge $\{k, j\}$ being deleted, and for each red edge $\{j, l\}$ being deleted along with j also a blue edge $\{i, l\}$ is deleted. This shows, in particular, that $\dim \mathcal{S}_{\mathcal{G}}^{+}/G$ is zero if and only if $\mathcal{G}$ has no blue edges, that is, if all edges run between vertices that are comparable in the pre-order. This is equivalent to the condition found in Letac and Massam (2007) for $M(\mathcal{G})$ to be a transformation family; see Theorem 5.1 below.

For our running example P_3, the model is a transformation family and similarly for complete graphs. For an example where $\mathcal{S}_{\mathcal{G}}^{+}/G$ has dimension 1, see Figure 1. The proof of Theorem 1.4 can be found in Section 6 and in supplementary materials [Draisma, Kuhnt and Zwiernik (2013)].

Organization of the paper. The remainder of the paper closely follows the structure of this introduction. First, in Section 2 we use structure theory of real algebraic groups to determine G. In Section 3, we derive necessary and sufficient conditions for the existence, with probability one, of equivariant estimators of the concentration (or covariance) matrix, and we give an upper bound on the robustness of those estimators, measured by the finite sample breakdown point for generic samples. In Section 4, we derive the maximal invariant of Theorem 1.3. In Section 5, we discuss in some detail the case where G acts transitively on $\mathcal{S}_{\mathcal{G}}^{+}$ providing general formula for an equivariant estimator, after which Section 6 is devoted to our combinatorial formula for the orbit space dimension in the general case. We conclude the paper with a short discussion. In the supplementary materials [Draisma, Kuhnt and Zwiernik (2013)], we provide the proof of Theorem 1.4. We also discuss further results on the combinatorial structure of the problem that link our work to Andersson and Perlman (1993).

2. The group G. Throughout this paper, we fix an undirected graph $\mathcal{G} = ([m], E)$ and define the group G as in the Introduction:

$$G := \{g \in \mathrm{GL}_m(\mathbb{R}) \mid g \cdot \mathcal{S}_{\mathcal{G}}^+ \subseteq \mathcal{S}_{\mathcal{G}}^+\}.$$

Note that G is, indeed, a subgroup of $\mathrm{GL}_m(\mathbb{R})$: first, if $g, h \in G$, then $(gh) \cdot \mathcal{S}_{\mathcal{G}}^+ \subseteq g \cdot \mathcal{S}_{\mathcal{G}}^+ \subseteq \mathcal{S}_{\mathcal{G}}^+$; and second, if $g \in G$, then since $\mathcal{S}_{\mathcal{G}}^+$ linearly spans $\mathcal{S}_{\mathcal{G}}$, the (linear) action of g must map the linear space $\mathcal{S}_{\mathcal{G}}$ into itself. Since g is invertible, we then have $g \cdot \mathcal{S}_{\mathcal{G}} = \mathcal{S}_{\mathcal{G}}$ (which implies that $g \cdot \mathcal{S}_{\mathcal{G}}^+ = \mathcal{S}_{\mathcal{G}}^+$ holds instead of the apparently weaker defining inclusion). But then also $g^{-1} \cdot \mathcal{S}_{\mathcal{G}} = \mathcal{S}_{\mathcal{G}}$. Finally, the action by g^{-1} preserves positive definiteness, so that $g^{-1} \cdot \mathcal{S}_{\mathcal{G}}^+ = \mathcal{S}_{\mathcal{G}}^+$, as claimed.

The general linear group $\mathrm{GL}_m(\mathbb{R})$ has two natural topologies: the Euclidean topology, and the weaker Zariski topology in which closed sets are defined by polynomial equations in the matrix entries. The subgroup G is closed in both topologies. Indeed, by the above, its elements g are characterized by the condition that $g^T K g \in \mathcal{S}_{\mathcal{G}}$ for all $K \in \mathcal{S}_{\mathcal{G}}$, and this translates into quadratic equations in the entries of g. As a Zariski-closed subgroup of $\mathrm{GL}_m(\mathbb{R})$, the group G is a real algebraic matrix group, and in particular a real Lie group. For basic structure theory of algebraic groups, we refer to Borel (1991). In algebraic groups, the Zariski-connected component containing the identity is always a normal subgroup, the quotient by which is finite. We first determine the identity component and then the quotient.

2.1. *The identity component.* Observe that the group

$$\mathbf{T}^m := (\mathrm{GL}_1(\mathbb{R}))^m \subseteq \mathrm{GL}_m(\mathbb{R})$$

of all invertible diagonal matrices is contained entirely in G—indeed, it just rescales the components of the random vector X and therefore preserves the original conditional independence statements defining $M(\mathcal{G})$. The group $\mathbf{T}^m$ has 2^m components in the Euclidean topology, corresponding to the possible sign patterns of the diagonal entries, but it is connected in the Zariski topology. For this reason, the Zariski topology is slightly more convenient to work with, and in what follows our topological terminology refers to it.

We will use that the connected component of G containing the identity (the *identity component*, for short) is determined uniquely by its Lie algebra $\mathfrak{g}$. The following lemma helps us determine that Lie algebra; we use the standard notation E_{ij} for the matrix that has zeroes everywhere except for a one at position (i, j).

LEMMA 2.1. *Let $H \subseteq \mathrm{GL}_m(\mathbb{R})$ be a real algebraic matrix group containing the group $\mathbf{T}^m$. 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 $\mathbf{T}^m$, namely, the group of all $g \in \mathrm{GL}_m(\mathbb{R})$ with $g_{ij} = 0$ unless $(i, j) \in I$.*

This lemma is well known, so we only sketch the key arguments. The commutative group $\mathbf{T}^m$ acts by conjugation on the Lie algebra of H, which therefore must be a direct sum of simultaneous eigenspaces of the elements of $\mathbf{T}^m$ in their conjugation action on the space of $m \times m$-matrices. These simultaneous eigenspaces are the one-dimensional subspaces spanned by the E_{ij}, so the Lie algebra of H is spanned by some of these matrices. For this argument, see [Borel (1991), Section 8.17]. The inclusion $\mathbf{T}^m \subseteq H$ implies that the E_{ii} are all in the Lie algebra, and for E_{ij}, E_{jk} in the Lie algebra with $i \neq k$, also the commutator $[E_{ij}, E_{jk}] = E_{ik}$ lies in the Lie algebra. The earliest relation to pre-orders that we could find is the paper Malyšev (1977).

Next, we determine which E_{ij} lie in $\mathfrak{g}$.

PROPOSITION 2.2. *For $i, j \in [m]$, the matrix E_{ij} lies in $\mathfrak{g}$ if and only if $j \preccurlyeq i$. As a consequence, the identity component of G is the group $G^0 = \{g \in \mathrm{GL}_m(\mathbb{R}) \mid g_{ij} = 0 \text{ if } j \not\preccurlyeq i\}$ from the Introduction.*

PROOF. The element E_{ij} with $i \neq j$ lies in $\mathfrak{g}$ if and only if the one-parameter group $(I + tE_{ij}),\ t \in \mathbb{R}$ lies in G, that is, maps $\mathcal{S}_\mathcal{G}$ into itself. Pick $K \in \mathcal{S}_\mathcal{G}$ with nonzero entries on the diagonal and at all positions corresponding to edges of $\mathcal{G}$. We have $(I + tE_{ij}) \cdot K = (I - tE_{ji})K(I - tE_{ij})$—this takes into account the inverses and the transpose in the definition of the action. This action has the effect of subtracting t times the ith row of K from the jth row and subtracting t times the ith column from the jth column. For suitable t this will create zeroes at positions corresponding to nonedges of $\mathcal{G}$ *unless* the positions of the nonzeroes in the ith row are among the positions of the nonzeroes in the jth row. This shows that $N(i) \cup \{i\} \subseteq N(j) \cup \{j\}$ is necessary for E_{ij} to lie in $\mathfrak{g}$; and repeating the argument for general K shows that it is also sufficient. The second statement now follows from Lemma 2.1. $\square$

Recall the running example $P_3 : \overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}$ from the Introduction. By Proposition 2.2, the Lie algebra $\mathfrak{g}$ is spanned by E_{11}, E_{22}, E_{33} together with E_{21} and E_{31}. The element E_{21} lies in $\mathfrak{g}^0$ because $N(2) \cup \{2\} = \{1, 2\} \subseteq N(1) \cup \{1\} = \{1, 2, 3\}$.

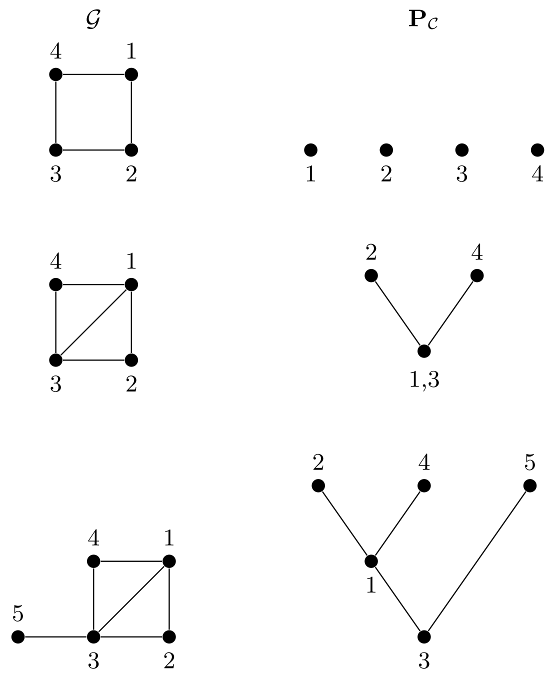


FIG. 2. *Three graphs and Hasse diagrams of the corresponding posets* $\mathbf{P}_{\mathcal{C}}$.

The inverse containment does not hold, so E_{12} does not lie in $\mathfrak{g}^0$. The group G^0 consists of invertible matrices of the form

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

where the asterisk denotes an element which can be nonzero.

It is useful in the remainder of the paper to have a thorough understanding of the pre-order $\preccurlyeq$. It can also be described in terms of the collection $\mathcal{C}$ of *maximal cliques* in the graph $\mathcal{G}$, as follows: $j \preccurlyeq i$ if and only if every $C \in \mathcal{C}$ containing j also contains i. Recall that $\preccurlyeq$ determines an equivalence relation $\sim$ on $[m]$. It also determines a partial order on $[m]/\sim$, still denoted $\preccurlyeq$, defined by $\bar{i} \preccurlyeq \bar{j}$ if $i \preccurlyeq j$. We denote the poset $([m]/\sim, \preccurlyeq)$ by $\mathbf{P}_{\mathcal{C}}$; it was first introduced in Letac and Massam (2007) but appeared also in other related contexts in Andersson and Klein (2010), Drton and Richardson (2008). In Figure 2, we show three graphs and the *Hasse diagrams* of the corresponding posets $\mathbf{P}_{\mathcal{C}}$. We note in passing that not all posets arise as $\mathbf{P}_{\mathcal{C}}$ for some $\mathcal{G}$. Two counterexamples are given in Figure 3. A more detailed study of the structure of $\mathbf{P}_{\mathcal{C}}$ is provided in the supplementary materials [Draisma, Kuhnt and Zwiernik (2013)].

FIG. 3. *Two posets that do not arise as $\mathbf{P}_{\mathcal{C}}$ for any $\mathcal{G}$ with collection $\mathcal{C}$ of maximal cliques.*

REMARK 2.3. Imagine relabeling the vertices of $\mathcal{G}$ by $[m]$ in such a way that the equivalence classes of $\sim$ are consecutive intervals and such that an inequality $\bar{j} \prec \bar{i}$ between equivalence classes implies that the interval corresponding to $\bar{j}$ contains smaller integers than the interval corresponding to $\bar{i}$. Then the matrices in G^0 are block lower triangular with square blocks along the diagonal corresponding to the equivalence classes. From this it is easy to see that G^0 is connected in the Zariski topology, but not in the ordinary Euclidean topology. Its number of Euclidean components is $2^{|[m]/\sim|}$, corresponding to sign patterns of the determinants of the diagonal blocks.

REMARK 2.4. The analogue of G^0 has been studied for other Gaussian models. For lattice conditional independence models this group was named the group of generalized block-triangular matrices with lattice structure [see Andersson and Perlman (1993), Section 2.4]. The link between lattice conditional independence models and certain Gaussian graphical models is discussed in Andersson et al. (1995) and in the supplementary materials [Draisma, Kuhnt and Zwiernik (2013)].

2.2. *The component group.* Now that we have determined the identity component G^0 of G, we set out to describe the quotient G/G^0, known as the component group. In the Introduction we observed that for our running example P_3 the permutation matrix

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

lies in G but not in G^0. The key to generalizing this observation is the following.

PROPOSITION 2.5. *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.*

PROOF. The subgroup $H := g^{-1}\mathbf{T}^m g$ is a maximal (real, split) torus in the real algebraic group G^0. By a standard result in the theory of algebraic groups [see, e.g., Borel (1991), Theorem 15.14], maximal tori are conjugate under G^0. Hence, there exists a $g_0 \in G$ such that $\mathbf{T}^m = g_0^{-1} H g_0$. Then $\mathbf{T}^m = (gg_0)^{-1}\mathbf{T}^m(gg_0)$, that is, gg_0 normalizes $\mathbf{T}^m$. But the normalizer of $\mathbf{T}^m$ in $\mathrm{GL}_m(\mathbb{R})$ consists of monomial matrices, that is, gg_0 equals σt with $t \in \mathbf{T}^m$ and σ some permutation matrix.

Hence, $g = \sigma(tg_0^{-1})$. Here the second factor is an element of G^0, so that σ is a permutation matrix contained in G. $\square$

We can now prove Theorem 1.1.

PROOF OF THEOREM 1.1. By Proposition 2.5 every element of G can be written as σg_0 with g_0 an element of G^0 and σ a permutation matrix belonging to G, that is, preserving the zero pattern of matrices in $\mathcal{S}_{\mathcal{G}}$. The only such permutation matrices are those coming from automorphisms of $\mathcal{G}$. This proves that $G = \mathrm{Aut}(\mathcal{G})G^0$, where we identify the automorphism group $\mathrm{Aut}(\mathcal{G})$ with the group of corresponding permutation matrices. This proves the theorem. $\square$

As explained in the Introduction, the expression $G = \mathrm{Aut}(\mathcal{G})G^0$ is not minimal in the sense that $\mathrm{Aut}(\mathcal{G})$ and G^0 may intersect. To get rid of that intersection, we define $\widetilde{\mathcal{G}}$ to be the graph with vertex set $[m]/\sim$ and an edge between $\bar{i}$ and $\bar{j}$ if there is an edge between i and j in $\mathcal{G}$. Define $c : [m]/\sim \to \mathbb{N}$, $\bar{i} \mapsto |\bar{i}|$ and view c as a coloring of the vertices of $\widetilde{\mathcal{G}}$ by natural numbers. Let $\mathrm{Aut}(\widetilde{\mathcal{G}}, c)$ denote the group of automorphisms of $\widetilde{\mathcal{G}}$ preserving the coloring. There is a lifting $\ell : \mathrm{Aut}(\widetilde{\mathcal{G}}, c) \to \mathrm{Aut}(G)$ defined as follows: the element $\tau \in \mathrm{Aut}(\widetilde{\mathcal{G}}, c)$ 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 $k = 1, \dots, |\bar{i}|$.

THEOREM 2.6. *The group G equals $\ell(\mathrm{Aut}(\widetilde{\mathcal{G}}, c))G^0$, and the intersection $\ell(\mathrm{Aut}(\widetilde{\mathcal{G}}, c)) \cap G^0$ is trivial, so G is the semidirect product $\ell(\mathrm{Aut}(\widetilde{\mathcal{G}}, c)) \ltimes G^0$.*

PROOF. By the proof of Theorem 1.1, any $g \in G$ can be written as σg_0 with $\sigma \in \mathrm{Aut}(\mathcal{G})$ and $g_0 \in G^0$. Since $\sim$ is defined entirely in terms of the graph $\mathcal{G}$, the graph automorphism σ satisfies $i \sim j \Leftrightarrow \sigma(i) \sim \sigma(j)$. This implies that σ determines an automorphism $\tau \in \mathrm{Aut}(\widetilde{\mathcal{G}}, c)$ mapping $\bar{i}$ to $\overline{\sigma(i)}$. Now σ equals $\ell(\tau)\sigma'$ where $\sigma' \in \mathrm{Aut}(\mathcal{G})$ maps each equivalence class $\bar{i}$ into itself. But then σ' lies in G^0 and hence g equals $\ell(\tau)$ times an element $\sigma' g_0$ of G^0. This proves the first statement. As for the second statement, observe that a permutation matrix can have the zero pattern prescribed by G^0 *only* if the permutation maps each equivalence class into itself. The only element of $\ell(\mathrm{Aut}(\widetilde{\mathcal{G}}, c))$ with this property is the identity matrix. $\square$

EXAMPLE 2.7. As an example, we consider a special small graph—the *bull graph*—which is a graph on five vertices depicted in Figure 4. The continuous part of G is given by the poset $\mathbf{P}_{\mathcal{C}}$ depicted on the right. There is only one nontrivial

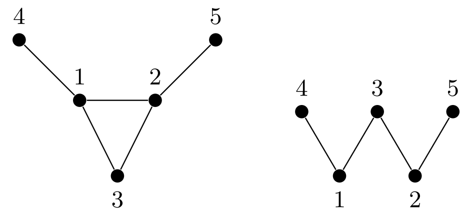


FIG. 4. *The bull graph and the corresponding* $\mathbf{P}_{\mathcal{C}}$ *on the right.*

automorphism of $\mathcal{G}$. It permutes 4 with 5 and 1 with 2. Hence, the group $G \subseteq \mathrm{GL}_5(\mathbb{R})$ consists of matrices of the following two types:

$$\begin{bmatrix} * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & 0 & 0 & * & 0 \\ 0 & * & 0 & 0 & * \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & 0 & 0 & * & 0 \\ 0 & * & 0 & 0 & * \end{bmatrix}$$

To see Theorems 1.1 and 2.6 in some further examples, see Section 7.

REMARK 2.8. To the coloured graph $(\tilde{\mathcal{G}}, c)$ we can associate a Gaussian graphical model $M(\mathcal{G}, c)$ with *multivariate nodes*, where node $\bar{i}$ is associated to a Gaussian vector of dimension $c_{\bar{i}}$. This model coincides with $M(\mathcal{G})$. This also shows, conversely, that our framework extends to general Gaussian graphical models with multivariate nodes.

3. Existence and robustness of equivariant estimators. Suppose that in the inference of the unknown concentration matrix $K \in \mathcal{S}_{\mathcal{G}}^{+}$ the observed n-sample $\mathbf{x} \in \mathbb{R}^{m \times n}$ leads to the estimate $T(\mathbf{x})$. Then it is reasonable to require that the sample $g\mathbf{x}$ leads to the estimate $gT(\mathbf{x})$. Such a map $T : \mathbb{R}^{m \times n} \to \mathcal{S}_{\mathcal{G}}^{+}$, possibly defined only outside some (typically G-stable) measure-zero set and satisfying $T(g\mathbf{X}) = gT(\mathbf{X})$ for all $g \in G$ there, is called a (G-)*equivariant estimator*. In this section we determine a sharp lower bound on n for an equivariant estimator to exist, and then, building on theory from Davies and Gather (2005), we determine a bound on the robustness of such estimators.

The MLE, when it exists, is automatically G-equivariant, since the likelihood function is G-invariant. A *necessary* condition for the MLE to exist with probability 1 is that the sample size n be at least the largest clique size $q = \max_{C \in \mathcal{C}} |C|$. A *sufficient* condition is that n be at least the maximal clique size q^* in a *decomposable cover* of $\mathcal{G}$, that is, a graph $\mathcal{G}^* = ([m], E^*)$ with $E^* \supseteq E$ that does not have induced k-cycles for $k \geq 4$. The exact minimal value of n for which MLE exists is not known explicitly in general, but interesting classes of graphs were analyzed in Barrett, Johnson and Loewy (1996), Buhl (1993) and Uhler (2012).

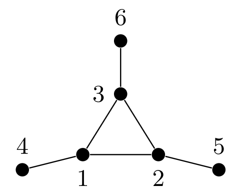


FIG. 5. *For this graph a G-equivariant map exists as soon as $n \geq 2$. However, the MLE exists only when $n \geq 3$.*

Our Theorem 1.2 states that equivariant estimators of the concentration matrix (or, equivalently by taking inverses, of the covariance matrix) exist if and only if $n \geq \max_{i\in[m]} |{\downarrow}i|$. Note that this is weaker than the necessary condition $n \geq q$ for the existence of MLE. Indeed, any down set ${\downarrow}i$ is in fact a clique, because $j, k \in{\downarrow}i$ implies that $j \in N(i) \cup \{i\} \subseteq N(k) \cup \{k\}$, that is, j and k are either equal or connected by an edge. The inequality $\max_{i\in[m]} |{\downarrow}i| \leq \max_{C\in\mathcal{C}} |C|$ can be strict. For example, in the graph of Figure 5 the biggest maximal clique has cardinality 3, while $\max_{i\in[m]} |{\downarrow}i| = 2$. In consequence, our result does not shed new light on the existence of MLE; however, it does provide necessary and sufficient conditions in the search for other equivariant estimators.

3.1. *Existence of equivariant estimators*. We now prepare the proof of Theorem 1.2. In our arguing, we borrow some terminology from algebraic geometry: we say that some property holds for *generic* n-tuples $\mathbf{x} \in \mathbb{R}^{m\times n}$ if it holds for $\mathbf{x}$ outside the zero set of some nonzero polynomial. Note that if a property holds for generic n-tuples, then it holds with probability one for the random sample $\mathbf{X}$ drawn from any nondegenerate probability distribution with continuous density function on $\mathbb{R}^{m\times n}$.

THEOREM 3.1. *The minimal number n for which the stabilizer in G^0 of a generic n-sample $\mathbf{x} \in \mathbb{R}^{m\times n}$ consists entirely of determinant-(± 1) matrices equals $n = \max_{i\in[m]} |{\downarrow}i|$. For that value of n the stabilizer of a generic n-sample is*, *in fact*, *the trivial group $\{I\}$.*

PROOF. The condition that $g \in G^0$ fixes one vector $x = (x_1, \dots, x_m)^T \in \mathbb{R}^m$ translates into m linear conditions on the entries of g, namely:

$$\sum_{j\preccurlyeq i} g_{ij}x_j = x_i \qquad \text{for } i = 1, \dots, m.$$

The ith condition concerns only the entries in the ith row of g. We therefore concentrate on that single row of g, and regard the entries g_{ij}, $j \preccurlyeq i$ as variables to be solved from the linear equations above as x ranges through the given n-tuple $\mathbf{x}$ of vectors. Since the given n-tuple is generic, those equations are linearly independent as long as n is at most the cardinality of ${\downarrow}i$. Hence, they determine the ith row uniquely as soon as n is at least that number. Hence, as soon as n is at least the maximal cardinality of the sets $\downarrow i$ over all i the stabilizer of a generic sample $\mathbf{x}$ will be trivial.

What remains to be checked, is that for smaller n the stabilizer of a generic $\mathbf{x} \in \mathbb{R}^{m \times n}$ does *not* consist entirely of determinant-(± 1) matrices. This is most easily seen by considering the Lie algebra of that stabilizer, which is the set of matrices A in the Lie algebra of G^0 satisfying the linear conditions $A\mathbf{x} = 0$. Let i be a row index for which $\downarrow i$ has more than n elements. Then the linear conditions on A do not fix the ith row of A uniquely. Moreover, by genericity, they do not fix the diagonal entry A_{ii} uniquely, either. As a consequence, they do not determine the trace of A uniquely. This shows that the Lie algebra of the stabilizer is not contained in the Lie algebra of trace-zero matrices. But then the stabilizer is not contained in the Lie group of determinant-one matrices (whose Lie algebra consists of the trace-zero matrices). □

PROOF OF THEOREM 1.2, NECESSITY OF $n \geq \max_i |\downarrow i|$. Assume that there exists a G-equivariant estimator $T : \mathbb{R}^{m \times n} \to \mathcal{S}_{\mathcal{G}}^{+}$, possibly defined outside some measure-zero set. In particular, the G^0-equivariance of T implies that the G^0-stabilizer of a generic sample $\mathbf{x}$ is contained in the G^0-stabilizer of $T(\mathbf{x})$:

$$G^0_{\mathbf{x}} \leq G^0_{T(\mathbf{x})}.$$

Now since $T(\mathbf{x}) \in \mathcal{S}_{\mathcal{G}}^{+}$, the stabilizer on the right-hand side is a generalized orthogonal group, and hence in particular compact in the Eculidean topology. Hence, the stabilizer on the left-hand side must be compact, as well. However, by (the proof of) Theorem 3.1, that stabilizer is the intersection of $\mathrm{GL}_m(\mathbb{R})$ with an affine subspace of $\mathbb{R}^{m \times m}$. Such a set is not compact in the Euclidean topology unless it consists of a single matrix, and this happens only when $n \geq \max_{i \in [m]} |\downarrow i|$. □

To prove that $n \geq \max_i |\downarrow i|$ is also *sufficient* for the existence of a G-equivariant estimator, we introduce the following construction. Fix a natural number $n \geq \max_i |\downarrow i|$ and construct a function $f : [m] \to [n]$ by induction, as follows: if f has been defined on all elements of $\downarrow i \setminus \bar{i}$, then define f on elements of $\bar{i}$ to be the increasing bijection from $\bar{i}$ (with the natural linear order coming from $[m]$) to the $|\bar{i}|$ smallest elements of the set $[n] \setminus f(\downarrow i \setminus \bar{i})$. This automatically guarantees that f is injective on any down set $\downarrow i$ and that $f \circ g = f$ for all $g \in \ell(\mathrm{Aut}(\widetilde{\mathcal{G}}, c))$. Now let $L \subseteq \mathbb{R}^{m \times n}$ be affine space of all matrices $\mathbf{x}$ with the property that first, the matrix $\mathbf{x}[\bar{i}, f(\bar{i})]$ obtained by taking the rows labeled by $\bar{i}$ and the columns labeled by $f(\bar{i})$ is an identity matrix for each $\bar{i} \in [m]/\sim$, and second, the matrices $\mathbf{x}[\bar{j}, f(\bar{i})]$ are zero for all $\bar{j} \prec \bar{i}$.

In our running example P_3, if the sample size n is at least 2, then f maps 1 to 1 and 2, 3 both to 2. The affine space L then consists of all matrices of the form

$$\begin{bmatrix} 1 & * & * & \cdots & * \\ 0 & 1 & * & \cdots & * \\ 0 & 1 & * & \cdots & * \end{bmatrix}.$$

LEMMA 3.2. *For generic $\mathbf{x} \in \mathbb{R}^{m\times n}$, there exists a unique $g \in G^0$ such that $g\mathbf{x} \in L$.*

The geometric content of this lemma is that L is a *slice* transverse to (most of) the orbits of G^0 on $\mathbb{R}^{m\times n}$. In our running example P_3, one goes from a generic sample to a sample in L by first multiplying the first row by x_{11}^{-1} so as to create a one at position $(1,1)$; then subtracting a multiple of the (new) first row from the second to create a zero at position $(2,1)$ and multiplying the second row by a constant to create a one at position $(2,2)$; and then similarly (and independently) for the third row. All of these operations are realized by elements of G^0. The following proof in the general case is a straightforward generalization of this.

PROOF OF LEMMA 3.2. For the existence of such a g, proceed by induction. Assume that the submatrix $\mathbf{x}[\downarrow i \setminus \bar{i}, f(\downarrow i \setminus \bar{i})]$ already has the required shape, and decompose $\mathbf{x}[\downarrow i, f(\downarrow i)]$ into blocks as follows:

$$\begin{bmatrix} \mathbf{x}[\downarrow i \setminus \bar{i}, f(\downarrow i \setminus \bar{i})] & \mathbf{x}[\downarrow i \setminus \bar{i}, f(\bar{i})] \\ \mathbf{x}[\bar{i}, f(\downarrow i \setminus \bar{i})] & \mathbf{x}[\bar{i}, f(\bar{i})] \end{bmatrix}.$$

Then take the block matrix $g \in G^0$ which is the identity outside the blocks labeled by $\downarrow i \times \downarrow i$, and which in those blocks looks like

$$\begin{bmatrix} I & 0 \\ g[\bar{i}, \downarrow i \setminus \bar{i}] & g[\bar{i}, \bar{i}] \end{bmatrix}.$$

Now straightforward linear algebra shows that, under the condition that both $\mathbf{x}[\downarrow i, f(\downarrow i)]$ and $\mathbf{x}[\downarrow i \setminus \bar{i}, f(\downarrow i \setminus \bar{i})]$ are full rank, there are unique choices for the as yet unspecified components of g such that $(g\mathbf{x})[\bar{i}, f(\downarrow i \setminus \bar{i})] = 0$ and $(g\mathbf{x})[\bar{i}, f(\bar{i})] = I$. This shows the existence of g such that $g\mathbf{x} \in L$. Uniqueness can be proved by a similar induction. $\square$

PROOF OF THEOREM 1.2, SUFFICIENCY OF $n \geq \max_i |\downarrow i|$. Now we show that $n \geq \max_{i\in[m]} |\downarrow i|$ is also a sufficient condition for the existence of an equivariant map $T : \mathbb{R}^{m\times n} \to \mathcal{S}^+_{\mathcal{G}}$, defined for generic samples $\mathbf{x}$. Indeed, by construction, the space L is stable under $\ell(\mathrm{Aut}(\widetilde{\mathcal{G}}, c))$. Fix any $\ell(\mathrm{Aut}(\widetilde{\mathcal{G}}, c))$-equivariant map $T : L \to \mathcal{S}^+_{\mathcal{G}}$. Such maps exist and can be found as follows: take $T' : L \to \mathcal{S}^+_{\mathcal{G}}$ *any* map, and then define

$$T(\mathbf{X}) := \frac{1}{|\mathrm{Aut}(\widetilde{\mathcal{G}}, c)|} \sum_{g \in \ell(\mathrm{Aut}(\widetilde{\mathcal{G}}, c))} g \cdot T'(g^{-1}\mathbf{X}),$$

an average over the finite group $\mathrm{Aut}(\widetilde{\mathcal{G}}, c)$.

We claim that T extends to a unique G-equivariant map $\mathbb{R}^{m\times n} \to \mathcal{S}^+_{\mathcal{G}}$ defined almost everywhere. Indeed, this extension is defined as follows: given a generic sample $\mathbf{x}$, find the unique $g_0 \in G^0$ such that $g_0\mathbf{x} \in L$, and set $T(\mathbf{x}) := g_0^{-1} \cdot T(g_0\mathbf{x})$.

Checking that the map T thus defined (almost) everywhere is both $\mathrm{Aut}(\widetilde{\mathcal{G}}, c)$-equivariant and G^0-equivariant is straightforward. This proves the existence part of Theorem 1.2. $\square$

REMARK 3.3. We stress that, apart from giving necessary and sufficient conditions for the existence of a G-equivariant estimator, the proof of Theorem 1.2 actually yields the general structure of *any* such estimator. Of course, the usefulness (bias, robustness, etc.) of an equivariant estimator thus constructed depends on the (free) choice of T', that is, on the restriction of T to the slice L. We do not know at present good conditions on T' that ensure usefulness of T.

REMARK 3.4. Note that the maps $T : \mathbb{R}^{m\times n} \to \mathcal{S}_{\mathcal{G}}^+$ constructed in the proof of Theorem 1.2 are merely G-equivariant, and not necessarily invariant under permutation of the sample points. It is easy to see, though, that the lower bound $n \geq \max_i |{\downarrow} i|$ also implies the existence of G-equivariant estimators that *are* invariant under permutations of the sample points. Indeed, simply replace T by its group average $\mathbf{X} \mapsto \frac{1}{n!}\sum_{\sigma\in S_n} T(\mathbf{X}^\sigma)$.

3.2. *Robustness*. An important notion for the robustness of parameter estimators is that of breakdown points [Donoho and Huber (1983), Hampel (1971)]. In a simple univariate situation, if the estimator is given by the sample mean, then a (large) change made to one of the observations leads to an arbitrarily large change in the value of the estimator. On the other hand, if the estimator is the sample median, then changing one observation in a sample of size larger than two cannot lead to arbitrarily large changes in the estimator. This feature makes the median more robust to outliers in the sample. The (*finite sample*) *breakdown point* of an estimator T at an n-sample $\mathbf{X} = \mathbf{x}$ is the minimal number of components of $\mathbf{x}$ that need to be altered to force arbitrarily large changes in the value of the estimator; this quantity is usually normalized by the sample size n. For example, the sample mean above has breakdown point $1/n$ while the sample median has breakdown point roughly $1/2$ (in fact, both independently of $\mathbf{x}$). So when it comes to robustness, the estimator with the highest breakdown point is preferred.

In the multivariate Gaussian setting, when estimating the concentration matrix (or the covariance matrix), the change in the estimator value is often measured by means of the *pseudo-metric* D on $\mathcal{S}_m^+$ [see, e.g., Davies and Gather (2005)]

$$D(K_1, K_2) = \bigl|\log\det\bigl(K_1 K_2^{-1}\bigr)\bigr|.$$

For graphical models, robustness issues have been rarely looked at so far, although it has been known for some time that the classical estimators and model selection procedures are vulnerable to contaminated data [Gottard and Pacillo (2006), Kuhnt and Becker (2003)]. First, approaches toward robust covariance estimators for undirected Gaussian graphical models can be found in Becker (2005), Gottard and Pacillo (2010). These papers suggest to replace the sample covariance matrix by the reweighted minimum covariance determinant (MCD) estimator. The paper Miyamura and Kano (2006) proposes an M-type estimator instead. Both in Finegold and Drton (2011) and in Vogel and Fried (2011) the assumption of normality is discarded, and replaced by the t-distribution or the general elliptical distribution, respectively, to model heavy tails.

Our modest contribution to robustness issues is an upper bound on the finite sample breakdown point for G-equivariant estimators of the concentration matrix for the graphical model $M(\mathcal{G})$. To this end, we specialize one of the key ideas from Davies and Gather (2005, 2007) to our setting. Suppose we have an n-sample $\mathbf{x} \in \mathbb{R}^{m\times n}$ and an equivariant estimator $T : \mathbb{R}^{m\times n} \to \mathcal{S}_{\mathcal{G}}^{+}$ of the concentration matrix. Assume that there exists an element $g \in G$ with $|\det g| \neq 1$ that fixes (at least) k of the n sample points $\mathbf{x}_1, \dots, \mathbf{x}_k$ of the sample $\mathbf{x}$. Define $d = \lceil \frac{n-k}{2} \rceil$ and let

$$\mathbf{y} = (\mathbf{x}_1, \dots, \mathbf{x}_k, \dots, \mathbf{x}_{n-d}, g^l \mathbf{x}_{n-d+1}, \dots, g^l \mathbf{x}_n).$$

Since $k + d \geq n - d$, for each natural number l both $\mathbf{y}$ and $g^{-l}\mathbf{y}$ contain at least $n - d$ points of the original sample $\mathbf{x}$. By the triangle inequality, we have $D(T(\mathbf{y}), T(g^{-l}\mathbf{y})) \leq D(T(\mathbf{x}), T(\mathbf{y})) + D(T(\mathbf{x}), T(g^{-l}\mathbf{y}))$ and on the other hand

$$\begin{aligned} D\big(T(\mathbf{y}), T\big(g^{-l}\mathbf{y}\big)\big) &= D\big(T(\mathbf{y}), \big(g^T\big)^l T(\mathbf{y}) g^l\big) \\ &= \big|\log\det\big(\big(g^T\big)^l T(\mathbf{y}) g^l T(\mathbf{y})^{-1}\big)\big| = l\big|\log\big(\det g^2\big)\big|, \end{aligned}$$

which is unbounded as $l \to \infty$. Hence, changing not more than $d = \lceil \frac{n-k}{2} \rceil$ of the sample points in $\mathbf{x}$ can already lead to arbitrarily large changes in the estimated concentration matrix, so that the finite sample breakdown point of T at $\mathbf{x}$ is at most d/n. We now state and prove our upper bound on the robustness of equivariant estimators at generic samples.

PROPOSITION 3.5. *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 $\lceil (n - \max_i |{\downarrow} i| + 1)/2 \rceil / n$.*

PROOF. By Theorem 3.1, there exist matrices $g \in G^0$ with determinant $\neq \pm 1$ that fix the first $k = \max_i |{\downarrow} i| - 1$ sample points. Now the proposition follows from the discussion preceding it. $\square$

REMARK 3.6. Writing $q := \max_i |{\downarrow} i|$, note that $q \leq m$ with equality if and only if $\mathcal{G}$ is the complete graph, and that $q \geq 1$, with equality if and only if for each edge $\{i, j\} \in E$ the vertex i has neighbors that are not connected to j (and vice versa). Examples of such graphs are m-cycles with $m \geq 4$. Trees with $m \geq 3$ vertices are examples of graphs with $q = 2$.

Note that for graphs with small q the upper bound in Proposition 3.5 is close to $1/2$, even for relatively small sample sizes n. On the other hand, the MLE, as pointed out for example in Maronna, Martin and Yohai (2006), is typically the least robust with respect to potential outliers in the sample space. Although we do not know whether the upper bound in the proposition is attained for any sensible estimator, our results do suggests the quest for more robust estimators, especially for graphs with small q.

4. The maximal invariant. In this section, we discuss a G^0-invariant map τ on the space $\mathbb{R}^{m\times n}$ of n-samples, defined almost everywhere, and prove that it is maximal in the sense that for two samples $\mathbf{x}, \mathbf{y}$ in the domain of definition of τ the equality $\tau(\mathbf{x}) = \tau(\mathbf{y})$ implies that $\mathbf{x}, \mathbf{y}$ are in the same G^0-orbit.

Recall from the Introduction that τ is defined as

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

where we assume from now on that n is at least $|\downarrow i|$, and where τ is defined on n-samples where $\mathbf{x}[\downarrow i]$ has full rank for all i. Before we proceed, we recall the following known lemma.

LEMMA 4.1. *Let $k \le n$ be natural numbers, and consider the action of* $\mathrm{GL}_k(\mathbb{R})$ *on* $\mathbb{R}^{k\times n}$. *Let U be the open subset of the latter space consisting of matrices of full rank k. Then the map $\varphi : U \mapsto \mathbb{R}^{n\times n}$ mapping $\mathbf{x}$ to $\mathbf{x}^T(\mathbf{x}\mathbf{x}^T)^{-1}\mathbf{x}$ is a maximal invariant for the action of* $\mathrm{GL}_k(\mathbb{R})$ *on U.*

PROOF. First, to see that φ is $\mathrm{GL}_k(\mathbb{R})$-invariant, compute

$$\varphi(g\mathbf{x}) = \mathbf{x}^T g^T \left(g\mathbf{x}\mathbf{x}^T g^T\right)^{-1} g\mathbf{x} = \varphi(x).$$

Second, to see that φ is maximal, note that the row space of $\mathbf{x} \in U$ is also the row space of $\varphi(\mathbf{x})$. Hence, if $\varphi(\mathbf{y}) = \varphi(\mathbf{x})$ for a second $\mathbf{y} \in U$, then $\mathbf{y}$ has the same row space as $\mathbf{x}$. But this means that there exists a $g \in \mathrm{GL}_k(\mathbb{R})$ with $g\mathbf{x} = \mathbf{y}$. $\square$

The proof of the lemma shows that $\varphi(\mathbf{x})$ determines the row space of $\mathbf{x}$ (and is determined by that!). Now we can prove Theorem 1.3, which states that τ is a maximal G^0-invariant. This generalizes Example 6.2.3 in Lehmann and Romano (2005), which deals with the case of complete graphs.

PROOF OF THEOREM 1.3. The G^0-invariance of each of the components of τ follows from the observation that $(g\mathbf{x})[\downarrow i] = g[\downarrow i]\mathbf{x}[\downarrow i]$, together with the computation in the proof of the preceding lemma.

For maximality, assume that $\tau(\mathbf{x}) = \tau(\mathbf{y})$. This means that the row space of $\mathbf{x}[\downarrow i]$ equals that of $\mathbf{y}[\downarrow i]$, for all i. If, by induction, we have replaced $\mathbf{x}$ by an element in its orbit and achieved that $\mathbf{x}[\downarrow i \setminus \bar{i}] = \mathbf{y}[\downarrow i \setminus \bar{i}]$, then it follows that $\mathbf{y}[\bar{i}] = h_1\mathbf{x}[\bar{i}] + h_2\mathbf{x}[\downarrow i \setminus \bar{i}]$ for a suitable invertible $\bar{i} \times \bar{i}$-matrix h_1 and a suitable full-rank $\bar{i} \times (\downarrow i \setminus \bar{i})$-matrix h_2. These matrices h_1, h_2 can be assembled into a block matrix $g_0 \in G^0$ (as in the proof of Lemma 3.2) such that $(g_0\mathbf{x})$ coincides with $\mathbf{x}$ outside the $\bar{i}$-labeled rows and with $\mathbf{y}$ in the $\bar{i}$-labeled rows. Doing this for all equivalence classes $\bar{i}$ from the bottom to the top of $\mathbf{P}_{\mathcal{C}}$, we move $\mathbf{x}$ to $\mathbf{y}$ by an element of G^0. $\square$

Since every invariant test depends on $\mathbf{x}$ only through the value of the maximal invariant [Lehmann and Romano (2005), Section 6.2], Theorem 1.3 paves the way for G^0-invariant tests, for example, for testing the hypothesis that the distribution of X lies in $M(\mathcal{G})$ against the null-hypothesis that it does not. A more general question involves testing two alternative (typically nested) graphical models corresponding to graphs $\mathcal{G}_1, \mathcal{G}_2$ on $[m]$. For this, it is natural to develop tests that are invariant with respect to matrices stabilizing *both* models. The identity component of the group of such matrices consists of all g with $g_{ij} = 0$ unless $j \preccurlyeq i$ in *both* pre-orders coming from $\mathcal{G}_1, \mathcal{G}_2$. The same construction as above, now applied to the intersection of the pre-orders, gives the maximal invariant for this group. A simple example of a G^0-invariant test is the deviance test [see, e.g., Lauritzen (1996), Section 5.2.2].

5. Equivariance in the transitive case. When G acts transitively on $\mathcal{S}^+_{\mathcal{G}}$ then $M(\mathcal{G})$ forms an exponential transformation family [see Barndorff-Nielsen et al. (1982)], which gives very efficient tools for dealing with the ancillary statistics in the hypothesis testing and inference. In particular the p^*-formula of Barndorff-Nielsen (1983), which gives an approximation for the density of the maximum likelihood estimator given the ancillary statistics is exact and the ancillary statistics is given by the maximal invariant $\tau(\mathbf{X})$.

The following result tells us when the graphical Gaussian model $M(\mathcal{G})$ is an exponential transformation family under the group G (cf. Theorem 1.4).

THEOREM 5.1 [Theorem 2.2, Letac and Massam (2007)]. *Let $\mathcal{G} = ([m], E)$ be an undirected graph. Then G acts transitively on $M(\mathcal{G})$ if and only if one of the following equivalent conditions holds*:

- *for any two neighbors $i, j \in [m]$ we have either $i \preccurlyeq j$ or $j \preccurlyeq i$*;
- *$\mathcal{G}$ is decomposable and does not contain a* 4-*chain* $\bullet - \bullet - \bullet - \bullet$ *as an induced subgraph*;
- *the Hasse diagram of $\mathbf{P}_{\mathcal{C}}$ is a tree with a unique minimum.*

As we show in the supplementary materials [Draisma, Kuhnt and Zwiernik (2013)], the transitive case is precisely the case when $M(\mathcal{G})$ corresponds to a lattice conditional independence model. We also prove there the following lemma.

LEMMA 5.2. *If $\mathcal{G}$ satisfies the conditions of Theorem 5.1, then* $\max_i |\downarrow i|$ *is equal to the size of the biggest maximal clique of $\mathcal{G}$. In particular a G-equivariant estimator exists with probability one if and only is the MLE estimator exists with probability one.*

In the transitive case construction of a G-equivariant estimator is particularly straightforward [Eaton (1989), Chapter 6, Example 6.2]. This generalizes the case of a star-shape graph analyzed in Sun and Sun (2005). Let $\widehat{\Sigma}$ be the MLE of the covariance matrix and define $S(\mathbf{X}) = n\widehat{\Sigma}$. Because $S(\mathbf{X})$ is a sufficient statistic, without loss we can assume that every estimator based on the full sample satisfies $T(\mathbf{X}) = T(S(\mathbf{X}))$. Since $S(\mathbf{X})^{-1} \in \mathcal{S}_{\mathcal{G}}^{+}$, there exists $h : \mathbb{R}^{m\times n} \to G^0$ such that $S^{-1}(\mathbf{X}) = h(\mathbf{X})^T h(\mathbf{X})$. The construction of h follows by the fact that in the transitive case there exists a well defined map $\phi : \mathcal{S}_{\mathcal{G}}^{+} \to G/G_I$, where G_I is the stabiliser of the identity matrix. This map is the inverse of the canonical map from G/G_I to $\mathcal{S}_{\mathcal{G}}^{+}$. Then h is just a composition of $S^{-1}(\mathbf{X}) : \mathbb{R}^{m\times n} \to \mathcal{S}_{\mathcal{G}}^{+}$ followed by ϕ. By G-equivariance,

$$T(\mathbf{X}) = T\bigl(S(\mathbf{X})\bigr) = T\bigl(h(\mathbf{X})^T h(\mathbf{X})\bigr) = h(\mathbf{X})^T T(I) h(\mathbf{X}),$$

where $T(I) \in \mathcal{S}_{\mathcal{G}}^{+}$. We have just shown the following result.

PROPOSITION 5.3. *Let $\mathcal{G}$ be a decomposable graph without induced 4-chains. Define $S(\mathbf{X}) = n\widehat{\Sigma}$ as above. Then every G-equivariant estimator of the concentration matrix is of the form*

$$T(\mathbf{X}) = \bigl(h_0 h(\mathbf{X})\bigr)^T h_0 h(\mathbf{X}),$$

where $h_0 \in G^0$ is a constant matrix and $h : \mathbb{R}^{m\times n} \to G^0$ is such that $S(\mathbf{X}) = h(\mathbf{X})^T h(\mathbf{X})$.

Since the function h is uniquely identified the only way to obtain different equivariant estimators is by varying the constant matrix h_0. This can be done with different optimality criteria in mind. An interesting problem is to find h_0 such that T^{-1} is an unbiased estimator of the concentration matrix. Another motivation is that the MLE for lattice conditional independence models (and hence for $M(\mathcal{G})$ in the transitive case by the theorem in the supplementary materials [Draisma, Kuhnt and Zwiernik (2013)]) is not admissible [see Konno (2001)]. A relevant question is to analyse equivariant estimators minimizing risk related to certain loss functions. This analysis has been already done for star-shaped models by Sun and Sun (2005).

6. Orbits of G on $\mathcal{S}_{\mathcal{G}}^{+}$. Given an undirected graph $\mathcal{G} = ([m], E)$, we have determined the group $G \subseteq \mathrm{GL}_m(\mathbb{R})$ of all invertible linear maps $\mathbb{R}^m \to \mathbb{R}^m$ stabilizing the cone $\mathcal{S}_{\mathcal{G}}^{+}$. Theorem 5.1 characterizes when $M(\mathcal{G})$ is a transformation family, that is, when G has a single orbit on $\mathcal{S}_{\mathcal{G}}^{+}$. For general $\mathcal{G}$, the orbit space $\mathcal{S}_{\mathcal{G}}^{+}/G$—like many quotients of manifolds by group actions—can conceivably be very complicated. In this section, we compute its first natural invariant, namely, its dimension. In the zero-dimensional case, we recover the class from Theorem 5.1.

Basic Lie group theory tells us that $\dim \mathcal{S}_{\mathcal{G}}^{+}/G$ equals

$$\dim \mathcal{S}_{\mathcal{G}}^{+} - \dim G + \dim G_K,$$

where G_K is the stabilizer of a generic concentration matrix in $\mathcal{S}_{\mathcal{G}}^{+}$. In this expression, the first term equals $m + |E|$ and the second term equals $\dim G^0 = \sum_{\bar{i}\in[m]/\sim} |\bar{i}| \cdot |\downarrow i|$, so it suffices to determine the generic stabilizer dimension. Note that for the dimension it does not matter whether we consider the stabilizer in G or in G^0. The following theorem makes use of the colored quotient graph $(\widetilde{\mathcal{G}}, c)$ from Section 2.

PROPOSITION 6.1. *The dimension of the stabilizer G_K^0 in G^0 of a generic matrix in $\mathcal{S}_{\mathcal{G}}^{+}$ equals $\sum_{\bar{i}\in[m]/\sim} \binom{n_{\bar{i}}}{2}$, where $n_{\bar{i}}$ is defined by*

$$n_{\bar{i}} := \max\left\{0, |\bar{i}| - \left(\sum_{\bar{j}\in N(\bar{i}), \bar{i}\not\preccurlyeq\bar{j}\not\preccurlyeq\bar{i}} |\bar{j}|\right)\right\},$$

where the sum ranges over all neighbors $\bar{j}$ of $\bar{i}$ in $\widetilde{\mathcal{G}}$ that are not comparable to $\bar{i}$ in the partial order $\preccurlyeq$.

In words: starting from $\widetilde{\mathcal{G}}$, one deletes all edges between vertices that are comparable in the partial order $\preccurlyeq$, and one subtracts from $|\bar{i}|$ the sum of the $|\bar{j}|$ for all neighboring $\bar{j}$ in the new graph. If the result is positive, then this is $n_{\bar{i}}$; otherwise, $n_{\bar{i}}$ is zero.

The expression above suggests that the identity component of G_K^0 is a product of special orthogonal groups of spaces of dimensions n_i, which is indeed what the proof of this proposition, given in the supplementary materials [Draisma, Kuhnt and Zwiernik (2013)], will show. We now use the proposition to explain the combinatorial procedure in the Introduction.

PROOF OF THEOREM 1.4. By Proposition 6.1, we need to compute

$$(m + |E|) - \sum_{\bar{i}\in[m]/\sim} |\bar{i}||\downarrow i| + \sum_{\bar{i}\in[m]/\sim} \binom{n_{\bar{i}}}{2}.$$

The term m cancels against the diagonal entries in G^0 in the second term. Recall that in Theorem 1.4 we colored an edge $\{i, j\}$ in $\mathcal{G}$ blue, green or red according to whether zero, one, or two of the statements $i \preccurlyeq j$ and $j \preccurlyeq i$ hold. The term $|E|$ counts blue plus green plus red. What remains of the second term after cancelling the diagonal entries against m counts green edges once and red edges twice. Thus, the first two terms count blue edges minus red edges. Finally, the last term counts the number of red edges that survive when blue edges are deleted one by one. □

We conclude with few examples of the use of Proposition 6.1.

EXAMPLE 6.2. Let $\widetilde{\mathcal{G}}$ be the bull graph in Figure 4, with each vertex representing an equivalence class in $[m]/\sim$ with cardinality c_i for $i = 1, \dots, 5$. In this case the only pair of connected but not comparable vertices is $(1, 2)$. With the convention that $\binom{m}{2} = 0$ if $m \le 0$, Proposition 6.1 shows that the dimension of the stabilizer of a generic matrix in $\mathcal{S}_{\mathcal{G}}^{+}$ is

$$\binom{c_1 - c_2}{2} + \binom{c_2 - c_1}{2} + \binom{c_3}{2} + \binom{c_4}{2} + \binom{c_5}{2}.$$

EXAMPLE 6.3. Let $\widetilde{\mathcal{G}}$ be a tree, where each vertex v represents an equivalence class with cardinality c_v. In this case, the dimension of the stabilizer of a generic matrix in $\mathcal{S}_{\mathcal{G}}^{+}$ is

$$\sum_{(u,v)\in \text{inner}} \left(\binom{c_u - c_v}{2} + \binom{c_v - c_u}{2} \right) + \sum_{i \in \text{leaves}} \binom{c_i}{2},$$

where the first sum is over all the inner vertices of $\widetilde{\mathcal{G}}$ and the second sum is over all the leaves (vertices of valency 1) of $\widetilde{\mathcal{G}}$. In particular, if for some c we have that $c_i = c$ for all $i \in C$ then this formula degenerates to $l\binom{c}{2}$, where l is the number of leaves.

7. Small examples. Let S_m denote the symmetric group on $[m]$, D_m the dihedral group of graph isomorphisms of an m-cycle. Also recall that $\mathbf{T}^k \simeq (\mathrm{GL}_1(\mathbb{R}))^k$ denotes the group of all diagonal invertible $k \times k$ matrices. In Table 1, we provide the full description of G for all undirected graphs on $m = 2, 3, 4$ vertices.

8. What's next. In this paper, we presented the complete description of the maximal subgroup of $GL_m(\mathbb{R})$ that stabilizes the Gaussian graphical model $M(\mathcal{G})$ for any given graph $\mathcal{G}$. The main motivation for this study was to put Gaussian graphical models into the framework of (composite) transformation families. Group invariance is a classical topic in multivariate statistics and there are many ways that statistical inference can be improved when the group action is better understood. While we have constructed the maximal invariant under this group on sample space, we have not yet used this invariant to develop explicit tests, for example, for model selection; and while we have given theoretical bounds on when equivariant estimators for the concentration matrix exist, and how robust they can be, we have not yet constructed such new estimators. We regard our work as a step toward achieving these goals for general graphs, laying down the theoretical framework. On the other hand, in the case where G acts transitively on the model, we already have a much better understanding. For instance, it seems feasible to extend the work of Sun and Sun (2005) from star-shaped models to general models in the transitive case. Once these transitive models are completely understood, it seems natural to move on to those where the orbit space of G on the model is one-dimensional. Here we expect beautiful mathematics and statistics to go hand in

TABLE 1
Small undirected graphs $\mathcal{G}$, corresponding groups G^0 and $\mathrm{Aut}(\widetilde{\mathcal{G}}, c)$ up to isomorphism

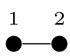
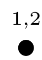
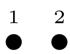
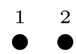
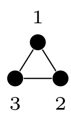
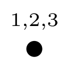
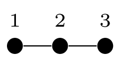
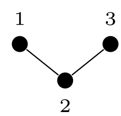
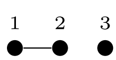
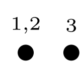
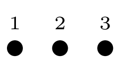
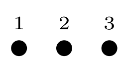
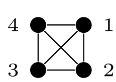
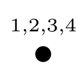
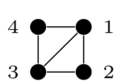
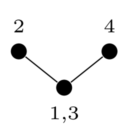
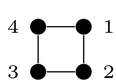
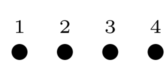
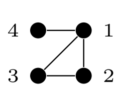
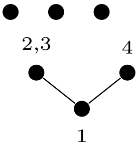
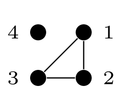
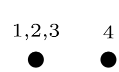
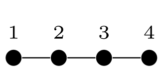
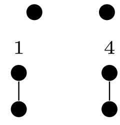
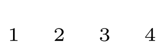
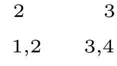
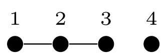
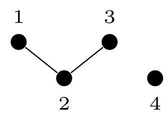
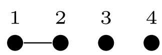
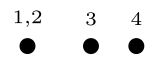
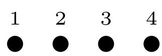
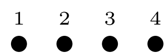
$\mathcal{G}$	$\mathbf{P}_{\mathcal{C}}$	G^0	$\mathrm{Aut}(\widetilde{\mathcal{G}}, c)$
		$\mathrm{GL}_2(\mathbb{R})$	$\{\mathrm{id}\}$
		$\mathbf{T}^2$	S_2
		$\mathrm{GL}_3(\mathbb{R})$	$\{\mathrm{id}\}$
		$\begin{bmatrix} * & * & 0 \\ 0 & * & 0 \\ 0 & * & * \end{bmatrix}$	S_2
		$\mathrm{GL}_2(\mathbb{R}) \times \mathbf{T}^1$	$\{\mathrm{id}\}$
		$\mathbf{T}^3$	S_3
		$\mathrm{GL}_4(\mathbb{R})$	$\{\mathrm{id}\}$
		$\begin{bmatrix} * & 0 & * & 0 \\ * & * & * & 0 \\ * & 0 & * & 0 \\ * & 0 & * & * \end{bmatrix}$	S_2
		$\mathbf{T}^4$	D_4
		$\begin{bmatrix} * & 0 & 0 & 0 \\ * & * & * & 0 \\ * & 0 & 0 & * \end{bmatrix}$	$\{\mathrm{id}\}$
		$\mathrm{GL}_3(\mathbb{R}) \times \mathbf{T}^1$	$\{\mathrm{id}\}$
		$\begin{bmatrix} * & 0 & 0 & 0 \\ * & * & 0 & 0 \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix}$	S_2
		$\mathrm{GL}_2(\mathbb{R}) \times \mathrm{GL}_2(\mathbb{R})$	S_2

TABLE 1
(*Continued*)

$\mathcal{G}$	$\mathbf{P}_{\mathcal{C}}$	G^{0}	$\mathrm{Aut}(\tilde{\mathcal{G}}, c)$
		$\begin{bmatrix} * & * & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & * & * & 0 \\ 0 & 0 & 0 & * \end{bmatrix}$	S_2
		$\mathrm{GL}_2(\mathbb{R}) \times \mathbf{T}^2$	S_2
		$\mathbf{T}^4$	S_4

hand: combinatorics for characterizing which graphs lead to such models, geometry for a better understanding of the one-dimensional orbit space, and statistical inference tailored to the geometry of that space.

Acknowledgments. S. Kuhnt thanks the Deutsche Forschungsgemeinschaft (SFB 823, project B1) for funding.

SUPPLEMENTARY MATERIAL

Proofs and more on the structure of $\mathbf{P}_{\mathcal{C}}$ (DOI: 10.1214/13-AOS1130SUPP; .pdf). We provide the proof of Proposition 6.1 and more results on the structure of the poset $\mathbf{P}_{\mathcal{C}}$ that link our work to Andersson and Perlman (1993).

REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. MR1990662

ANDERSSON, S. A. and KLEIN, T. (2010). On Riesz and Wishart distributions associated with decomposable undirected graphs. *J. Multivariate Anal.* **101** 789–810. MR2584900

ANDERSSON, S. A. and PERLMAN, M. D. (1993). Lattice models for conditional independence in a multivariate normal distribution. *Ann. Statist.* **21** 1318–1358. MR1241268

ANDERSSON, S. A., MADIGAN, D., PERLMAN, M. D. and TRIGGS, C. M. (1995). On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. *J. Statist. Plann. Inference* **48** 25–46. MR1366371

BARNDORFF-NIELSEN, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365. MR0712023

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

BARRETT, W. W., JOHNSON, C. R. and LOEWY, R. (1996). The real positive definite completion problem: Cycle completability. *Mem. Amer. Math. Soc.* **122** viii+69. MR1342017

BECKER, C. (2005). Iterative proportional scaling based on a robust start estimator. In *Classification—The Ubiquitous Challenge* (C. Weihs and W. Gaul, eds.) 248–255. Springer, Berlin.

BOREL, A. (1991). *Linear Algebraic Groups*, 2nd ed. *Graduate Texts in Mathematics* **126**. Springer, New York. MR1102012

BUHL, S. L. (1993). On the existence of maximum likelihood estimators for graphical Gaussian models. *Scand. J. Stat.* **20** 263–270. MR1241392

DAVIES, P. L. and GATHER, U. (2005). Breakdown and groups. *Ann. Statist.* **33** 977–1035. MR2195626

DAVIES, P. L. and GATHER, U. (2007). The breakdown point—Examples and counterexamples. *REVSTAT* **5** 1–17. MR2365930

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. thesis, Harvard Univ.

DONOHO, D. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* 157–184. Wadsworth, Belmont, CA. MR0689745

DRAISMA, J., KUHNT, S. and ZWIERNIK, P. (2013). Supplement to "Groups acting on Gaussian graphical models." DOI:10.1214/13-AOS1130SUPP.

DRTON, M. and RICHARDSON, T. S. (2008). Graphical methods for efficient likelihood inference in Gaussian covariance models. *J. Mach. Learn. Res.* **9** 893–914. MR2417257

EATON, M. L. (1989). *Group Invariance Applications in Statistics*. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 1. IMS, Hayward, CA. MR1089423

FINEGOLD, M. and DRTON, M. (2011). Robust graphical modeling of gene networks using classical and alternative t-distributions. *Ann. Appl. Stat.* **5** 1057–1080. MR2840186

FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **144** 285–307.

GOTTARD, A. and PACILLO, S. (2006). On the impact of contaminations in graphical Gaussian models. *Stat. Methods Appl.* **15** 343–354. MR2345645

GOTTARD, A. and PACILLO, S. (2010). Robust concentration graph model selection. *Comput. Statist. Data Anal.* **54** 3070–3079. MR2727735

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896. MR0301858

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

KONNO, Y. (2001). Inadmissibility of the maximum likelihood estimator of normal covariance matrices with the lattice conditional independence. *J. Multivariate Anal.* **79** 33–51. MR1867253

KUHNT, S. and BECKER, C. (2003). Sensitivity of graphical modeling against contamination. In *Between Data Science and Applied Data Analysis* (M. Schader, W. Gaul and M. Vichi, eds.) 279–287. Springer, Berlin.

LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Oxford Univ. Press, New York. MR1419991

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York. MR2135927

LETAC, G. and MASSAM, H. (2007). Wishart distributions for decomposable graphs. *Ann. Statist.* **35** 1278–1323. MR2341706

LOPUHAÄ, H. P. and ROUSSEEUW, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19** 229–248. MR1091847

MALYŠEV, F. M. (1977). Closed subsets of roots and the cohomology of regular subalgebras. *Mat. Sb.* **104(146)** 140–150, 176. MR0472944

MARONNA, R. A., MARTIN, R. D. and YOHAI, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley, Chichester. MR2238141

MIYAMURA, M. and KANO, Y. (2006). Robust Gaussian graphical modeling. *J. Multivariate Anal.* **97** 1525–1550. MR2275418

REID, N. (1995). The roles of conditioning in inference. *Statist. Sci.* **10** 138–157, 173–189, 193–196. MR1368097

SCHERVISH, M. J. (1995). *Theory of Statistics*. Springer, New York. MR1354146

STAHEL, W. (1981). Robust estimation: Infinitesimal optimality and covariance matrix estimators. Ph.D. thesis, ETH, Zürich.

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

UHLER, C. (2012). Geometry of maximum likelihood estimation in Gaussian graphical models. *Ann. Statist.* **40** 238–261. MR3014306

VOGEL, D. and FRIED, R. (2011). Elliptical graphical modelling. *Biometrika* **98** 935–951. MR2860334

J.DRAISMA
DEPARTMENT OF MATHEMATICS
AND COMPUTER SCIENCE
TU EINDHOVEN
PO BOX 513
5600 MB EINDHOVEN
THE NETHERLANDS
E-MAIL: j.draisma@tue.nl

S. KUHNT
FACULTY OF STATISTICS
TU DORTMUND UNIVERSITY
44221 DORTMUND
GERMANY
E-MAIL: kuhnt@statistik.tu-dortmund.de

P. ZWIERNIK
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA BERKELEY
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: pzwiernik@berkeley.edu