

Symmetries in directed Gaussian graphical models

Visu Makam¹, Philipp Reichenbach², and Anna Seigal³

¹*School of Mathematics and Statistics, University of Melbourne*
e-mail: visu@umich.edu

²*Institut für Mathematik, Technische Universität Berlin*
e-mail: reichenbach@tu-berlin.de

³*School of Engineering and Applied Sciences, Harvard University*
e-mail: aseigal@seas.harvard.edu

Abstract: We define Gaussian graphical models on directed acyclic graphs with coloured vertices and edges, calling them RDAG (restricted directed acyclic graph) models. If two vertices or edges have the same colour, their parameters in the model must be the same. We present an algorithm to find the maximum likelihood estimate (MLE) in an RDAG model, and characterise when the MLE exists, via linear independence conditions. We relate properties of a graph, and its colouring, to the number of samples needed for the MLE to exist and to be unique. We also characterise when an RDAG model is equal to an associated undirected graphical model and study connections to groups and invariant theory. We provide examples and simulations to study the benefits of RDAGs over uncoloured DAGs.

MSC2020 subject classifications: Primary 62F10, 62H22, 62R01; secondary 05C15, 14L24, 15A23, 20G45.

Keywords and phrases: Maximum likelihood estimation, graphical models, graph colouring, Gaussian models, maximum likelihood thresholds, group actions.

Received May 2022.

Contents

Introduction	3970
1 Preliminaries	3972
2 Introducing RDAG models	3974
3 Comparison of RDAG and RCON models	3978
4 MLE: existence, uniqueness, and an algorithm	3983
5 Maximum likelihood thresholds	3989
6 Simulations	3994
7 Conclusion	3995
A Connections to stability	3996
B Connections to Gaussian group models	4003
Acknowledgments	4007
Funding	4007
References	4007

Introduction

The concept of a graph is widely used across the sciences [15]. Graphs are a framework to relate entities: the vertices are the entities of interest, and the edges encode connections between them. A graph is given a statistical meaning in the study of graphical models [23, 25]. Each vertex represents a random variable, and the edges between variables reflect their statistical dependence [42]. In this paper, we study *directed* Gaussian graphical models, also called Gaussian Bayesian networks, or linear structural equation models with independent errors [39]. Such models have been applied to cell signalling [34], gene interactions [17], causal inference [30], and many other contexts.

We define graphical models on directed acyclic graphs (DAGs) with a colouring of their vertices and edges. Vertices or edges with the same colour must have the same parameter values. Thus, the graph colouring imposes symmetries in the model. We call such models *RDAG models*, where the ‘R’ stands for restricted, cf. [22].

Our first motivation for RDAG models is that vertex and edge symmetries appear in various applications, such as in the study of longitudinal data [1, 43], or clustered variables [19, 22]. The coloured directed graph gives an intuitive pictorial description of the symmetry conditions in the model.

Our second motivation is to decrease the maximum likelihood threshold, the minimum sample size required for the maximum likelihood estimate (MLE) to exist almost surely, see [7]. In applications, it is desirable for the MLE to exist when there is only a small number of samples; i.e., for the maximum likelihood threshold to be small. Innovative ideas have been used to find maximum likelihood thresholds in graphical models [5, 12, 20, 41] and for estimating the MLE from too few samples [16, 44]. Removing edges from a graph can lower the threshold [41, 23], but there is a trade-off: removing edges imposes more conditional independence among the variables. This is why, instead, we aim to decrease the maximum likelihood threshold by introducing symmetries.

We will use the following as our running example throughout the paper.

Example 0.1. Consider the coloured graph $\textcircled{1} \leftarrow \boxed{3} \rightarrow \textcircled{2}$, with blue (circular) vertices $\{1, 2\}$, black (square) vertex 3 and two red edges. The RDAG model is

$$y_1 = \lambda y_3 + \epsilon_1, \quad y_2 = \lambda y_3 + \epsilon_2, \quad y_3 = \epsilon_3,$$

where $\epsilon_1, \epsilon_2 \sim N(0, \omega)$ and $\epsilon_3 \sim N(0, \omega')$, i.e. ω is the variance of blue vertices 1 and 2 and ω' is the variance of black vertex 3. The third parameter λ is the regression coefficient given by a red edge. We will see that the MLE exists uniquely (almost surely) given one sample. For comparison, if we remove the colours the resulting model needs two samples for the MLE to exist. We use this example to model the dependence of two daughters’ heights on the height of their mother, and we compute the MLE given some sample data, in Section 4.3.

As far as we are aware, RDAG models have not been defined before in the literature; we comment on some related models. The assumption of equal variances from [31] is the special case of an RDAG model, where all vertex colours

are the same. Special colourings encode exchangeability between variables, or invariance under a group of permutations. A graphical model is combined with group symmetries in the directed setting in [26] and in the undirected setting in [4, 36]. RDAG models also relate to the fused graphical lasso [6], which penalises differences between parameters on different edges, whereas in an RDAG model the parameters on edges of the same colour must be equal.

In this paper, we give a closed-form formula for the MLE in an RDAG model, as a collection of least squares estimators, see Algorithm 1. We characterise the existence and uniqueness of the MLE via linear algebraic properties of the sample data, see Theorem 4.4. We give upper and lower bounds on the threshold number of samples required for existence and uniqueness of the MLE in Theorem 5.3. Our results show that RDAG thresholds are less or equal to the DAG thresholds, and that high symmetry decreases the thresholds. Finally, we compare RDAG MLEs to uncoloured DAG MLEs via simulations in Section 6. Our results hold with an assumption on the graph colouring, which we call *compatibility* (Definition 2.5). It is an open problem to extend our results to the non-compatible setting, as well as to directed graphs with cycles. It is also an open problem to find the exact maximum likelihood thresholds, see Problem 5.4.

The undirected analogue to RDAG models are the RCON models from [22]. Although a motivation for the graph colouring in RCON models is to lower the maximum likelihood threshold, there are relatively few graphs for which the threshold is known: colourings of the four cycle are studied in [41, §6], [38, §5], while an example with five vertices is [41, Example 3.2]. In certain cases, RDAG models are equivalent to RCON models. We determine precisely the conditions under which this occurs in Theorem 3.4. As a consequence, we obtain an entire class of RCON models where conditions for MLE existence and uniqueness can be found by appealing to our results on RDAGs.

This paper has two appendices, where we explain some connections to invariant theory. A Gaussian group model [2] is parametrised by a group. In [2], the authors draw a dictionary between maximum likelihood estimation and stability notions in invariant theory. This dictionary allows for the transfer of tools from the algebraic subjects of representation theory and invariant theory to statistics: maximum likelihood thresholds were computed for matrix normal models in [9] and for tensor normal models in [10].

We extend the dictionary between maximum likelihood estimation and stability notions to RDAGs in Theorem A.2. This requires us to extend the definitions of stability beyond the setting of a group action, see Definition A.1. While not evident in our final presentation, this perspective gave us the understanding needed to obtain many of the results in this paper and we would like to stress its importance for future work. We have far more tools at our disposal when a model is backed by a group action, i.e., when it is a Gaussian group model. We identify RDAGs that are Gaussian group models in Proposition B.2 and exhibit additional tools that one can use in such cases. The two appendices offer two alternative descriptions of the set of MLEs, see Propositions A.3 and B.6.

1. Preliminaries

1.1. Multivariate Gaussian models

We consider m -dimensional Gaussian distributions with mean zero. Such a distribution is determined by its concentration (inverse covariance) matrix Ψ , a real $m \times m$ positive definite matrix. The density function is

$$f_{\Psi}(y) = \frac{1}{\sqrt{\det(2\pi\Psi^{-1})}} \exp\left(-\frac{1}{2}y^{\top}\Psi y\right), \quad y \in \mathbb{R}^m.$$

We refer to a multivariate Gaussian model by the set of concentration matrices in the model. So, a model is a subset of PD_m , the cone of $m \times m$ positive definite matrices.

We study statistical models in PD_m via a set of invertible matrices. We define

$$\mathcal{M}_A := \{a^{\top}a \mid a \in A\}, \quad (1)$$

where A is a subset of GL_m , the real invertible $m \times m$ matrices. Many sets A can correspond to the same model \mathcal{M}_A . For instance, the full cone PD_m is \mathcal{M}_A whenever A contains all invertible upper triangular matrices. When the set A is a group, the model \mathcal{M}_A is called a *Gaussian group model* [2].

1.2. Maximum likelihood estimation

A maximum likelihood estimate (MLE) is a point in the model that maximizes the likelihood of observing some data. For n samples from a Gaussian model $\mathcal{M} \subseteq \text{PD}_m$, the data samples are the columns of a matrix $Y \in \mathbb{R}^{m \times n}$. Assuming independent samples, the likelihood function is

$$L_Y(\Psi) = \prod_{i=1}^n f_{\Psi}(Y_i),$$

where Y_i is the i th column of Y . We work with the log-likelihood function $\log L_Y$, which has the same maximisers as L_Y . The log-likelihood function can be written, up to additive and multiplicative constants, as

$$\ell_Y(\Psi) = \log \det(\Psi) - \text{tr}(\Psi S_Y), \quad (2)$$

where $S_Y = \frac{1}{n} \sum_{i=1}^n Y_i Y_i^{\top}$ is the sample covariance matrix. Four possibilities arise when maximising the log-likelihood:

- (a) ℓ_Y unbounded from above
- (b) ℓ_Y bounded from above
- (c) the MLE exists (i.e. ℓ_Y is bounded from above and attains its supremum)
- (d) the MLE exists and is unique.

The minimal number of samples needed for the MLE to exist almost surely is the MLE existence threshold; the number of samples for the MLE to exist uniquely almost surely is the uniqueness threshold.

Example 1.1. Let $\mathcal{M} = \text{PD}_m$. The unique maximiser of the likelihood is $\hat{\Psi} = S_Y^{-1}$, if S_Y is invertible. If S_Y is not invertible, the likelihood function is unbounded from above, see [39, Proposition 5.3.7]. The existence and uniqueness thresholds are therefore both m , since with m samples the matrix S_Y will almost surely be invertible.

For a model \mathcal{M}_A as in (1), we can rewrite the log-likelihood (2) at sample matrix $Y \in \mathbb{R}^{m \times n}$ as a function of a matrix $a \in A$:

$$\ell_Y(a^\top a) = \log \det(a^\top a) - \frac{1}{n} \|a \cdot Y\|^2, \quad (3)$$

where $\|M\|$ denotes the Frobenius norm of matrix M .

1.3. Directed Gaussian graphical models (DAG models)

A directed graph is $\mathcal{G} = (I, E)$, where I is a finite set of vertices, and E a set of directed edges. We write $j \rightarrow i$ for an edge from j to i ; the absence of such an edge is denoted $j \not\rightarrow i$. A directed acyclic graph (DAG) is a directed graph with no directed cycles; i.e., no cycles of the form $i \rightarrow j \rightarrow \dots \rightarrow k \rightarrow i$. The *parents* and *children* of i in \mathcal{G} are, respectively, the vertex sets

$$\text{pa}(i) = \{j \in I \mid (j \rightarrow i) \in E\} \quad \text{ch}(i) = \{k \in I \mid (i \rightarrow k) \in E\}.$$

We often take the vertex set I to be $[m] = \{1, 2, \dots, m\}$.

We call a directed Gaussian graphical model on \mathcal{G} a *DAG model*. A DAG model is defined by the linear structural equation

$$y = \Lambda y + \varepsilon, \quad \text{i.e.} \quad y_i = \sum_{j \in \text{pa}(i)} \lambda_{ij} y_j + \varepsilon_i, \quad (4)$$

where $y \in \mathbb{R}^m$, $\lambda_{ij} = 0$ for $j \not\rightarrow i$ in \mathcal{G} , and $\varepsilon \sim N(0, \Omega)$ with Ω diagonal. The coefficient λ_{ij} is a regression coefficient, the effect of parent j on child i . The model encodes conditional independence: a node is independent of its non-descendants, after conditioning on its parents [42].

Remark 1.2 (The matrix Λ is strictly upper triangular). Throughout the paper, we choose an ordering on the vertices of \mathcal{G} so that Λ is upper triangular. That is, if edge $j \rightarrow i$ is in E then $j > i$. Such an ordering is possible because \mathcal{G} is acyclic. Thinking of a vertex label as its age, the ordering makes parents older than children.

Solving (4) for y gives $y = (\text{id} - \Lambda)^{-1} \varepsilon$, where id denotes the $m \times m$ identity matrix, and the acyclicity of \mathcal{G} ensures that $(\text{id} - \Lambda)$ is invertible. Hence y is multivariate Gaussian with mean zero and concentration matrix

$$\Psi = (\text{id} - \Lambda)^\top \Omega^{-1} (\text{id} - \Lambda). \quad (5)$$

We define a set of matrices associated to the DAG \mathcal{G}

$$A(\mathcal{G}) = \{a \in \text{GL}_m \mid a_{ij} = 0 \text{ for } i \neq j \text{ with } j \not\rightarrow i \text{ in } \mathcal{G}\}. \quad (6)$$

Recall from (1) the notation $\mathcal{M}_{A(\mathcal{G})} = \{a^\top a : a \in A(\mathcal{G})\}$. The set of concentration matrices of the form (5) is equal to the set $\mathcal{M}_{A(\mathcal{G})}$. We prove this in Lemma 2.9.

1.4. Undirected Gaussian graphical models

Multivariate Gaussian models can also be obtained from undirected graphs. An undirected graph $\mathcal{G} = (I, E)$ is a set of vertices I and *undirected* edges E . The model is the set of distributions with mean zero and concentration matrix Ψ with $\Psi_{ij} = 0$ whenever edge $i - j$ is not in E . That is, the variables at nodes i and j are independent after conditioning on all others, see [39, Proposition 13.1.5].

1.5. Restricted concentration (RCON) models

In [22], the authors introduce restricted concentration (RCON) models, which impose symmetries on the concentration matrix Ψ according to a graph colouring. A coloured undirected graph is a tuple (\mathcal{G}, c) , where $\mathcal{G} = (I, E)$ is an undirected graph and the map

$$c : I \cup E \rightarrow \mathcal{C}$$

assigns a colour to each vertex and to each edge. The vertex $i \in I$ has colour $c(i) \in \mathcal{C}$, and edge $i - j$ has colour $c(ij) \in \mathcal{C}$.

Definition 1.3 (see [22, §3]). The *RCON model* on the coloured undirected graph (\mathcal{G}, c) consists of concentration matrices with

1. $\Psi_{ij} = 0$ whenever $i - j$ is not in E
2. $\Psi_{ii} = \Psi_{jj}$ whenever $c(i) = c(j)$,
3. $\Psi_{ij} = \Psi_{kl}$ whenever $c(ij) = c(kl)$.

2. Introducing RDAG models

A colouring of a DAG assigns colours to the vertices and edges. A coloured DAG is a tuple (\mathcal{G}, c) , where $\mathcal{G} = (I, E)$ is a DAG on vertices I and edges E , and

$$c : I \cup E \rightarrow \mathcal{C}$$

is a colouring of the vertices and edges. Vertex $i \in I$ has colour $c(i) \in \mathcal{C}$, and edge $j \rightarrow i$ has colour $c(ij) \in \mathcal{C}$. We sometimes denote the vertex colour $c(i)$ by $c(ii)$, with no ambiguity because a DAG cannot have loops.

Definition 2.1. The *restricted DAG (RDAG) model* on the coloured DAG (\mathcal{G}, c) is the set of concentration matrices $\Psi = (\text{id} - \Lambda)^\top \Omega^{-1} (\text{id} - \Lambda)$, where $\Lambda \in \mathbb{R}^{m \times m}$ satisfies

1. $\lambda_{ij} = 0$ unless $j \rightarrow i$ in \mathcal{G}
2. $\lambda_{ij} = \lambda_{kl}$ whenever edges $j \rightarrow i$ and $l \rightarrow k$ have the same colour

and the diagonal matrix $\Omega \in \mathbb{R}^{m \times m}$ has positive entries and satisfies

- (3) $\omega_{ii} = \omega_{jj}$ if vertices i and j have the same colour.

The model is given by the linear structural equation $y = \Lambda y + \varepsilon$, where $\varepsilon \sim N(0, \Omega)$.

Example 2.2. Consider the coloured graph $\textcircled{1} \leftarrow \textcircled{3} \rightarrow \textcircled{2}$ from Example 0.1. The RDAG model is parametrised by matrices

$$\Lambda \in \left\{ \begin{pmatrix} 0 & 0 & \lambda \\ 0 & 0 & \lambda \\ 0 & 0 & 0 \end{pmatrix} : \lambda \in \mathbb{R} \right\} \quad \text{and} \quad \Omega \in \left\{ \begin{pmatrix} \omega & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega' \end{pmatrix} : \omega, \omega' > 0 \right\}.$$

We will parametrise the RDAG model on (\mathcal{G}, c) via the set

$$A(\mathcal{G}, c) := \left\{ a \in \text{GL}_m \mid \begin{array}{l} a_{ij} = 0 \text{ for } i \neq j \text{ with } j \not\rightarrow i \text{ in } \mathcal{G} \\ a_{ij} = a_{kl} \text{ whenever } c(ij) = c(kl) \end{array} \right\}. \tag{7}$$

Note that $A(\mathcal{G}, c)$ is contained in the set $A(\mathcal{G})$ from (6): the zero patterns of $A(\mathcal{G})$ and $A(\mathcal{G}, c)$ are the same, and $A(\mathcal{G}, c)$ has further equalities imposed by the colouring c . Before we characterise which RDAG models can be parametrised by $A(\mathcal{G}, c)$, we pause to motivate the use of this alternative parametrisation.

Remark 2.3 (Motivation for the parametrisation via $A(\mathcal{G}, c)$). The RDAG model on (\mathcal{G}, c) will be the set \mathcal{M}_A as in (1), where $A := A(\mathcal{G}, c)$. This point of view is motivated by connections to invariant theory for transitive DAG models in [2, Section 5]. The alternative parametrisation has useful consequences. First, it leads to a condition on the graph colouring, called compatibility, which is indispensable in our results of Sections 4 and 5. Second, it is helpful when comparing directed and undirected models in Section 3. Finally, it enables us to generalise the connections to invariant theory from [2] to the setting of RDAGs, see Appendices A and B.

Example 2.4. Returning to the example $\textcircled{1} \leftarrow \textcircled{3} \rightarrow \textcircled{2}$, we have

$$A(\mathcal{G}, c) = \left\{ \begin{pmatrix} d_1 & 0 & r \\ 0 & d_1 & r \\ 0 & 0 & d_2 \end{pmatrix} : d_1, d_2 \neq 0, r \in \mathbb{R} \right\}.$$

We now introduce a natural assumption on a colouring.

Definition 2.5. A colouring c of a directed graph is *compatible*, if:

- (i) Vertex and edge colours are disjoint; and
- (ii) If edges $j \rightarrow i$ and $l \rightarrow k$ have the same colour, then the child vertices i and k also have the same colour, i.e. $c(ij) = c(kl) \implies c(i) = c(k)$.

Note that compatibility does *not* impose equality of parent colours $c(j)$ and $c(l)$.

Remark 2.6 (Motivation for compatibility). In an RDAG model, we do not impose equalities between Ω and Λ . The entry ω_{ii} is a variance, while λ_{kl} is a regression coefficient, so setting them to be equal would be difficult to interpret. Hence the vertex and edge colours can always be thought of as disjoint, as in compatibility condition (i). Compatibility condition (ii) has the statistical interpretation that the same regression coefficient appearing in an expression for two variables implies that their error variances agree. This extra assumption is indispensable in many of our results and proofs. It is a directed analogue to the condition appearing in [22, Proposition 1].

The first use of compatibility condition (ii) is in relating an RDAG model on (\mathcal{G}, c) to the set $A(\mathcal{G}, c)$. As in (1), we consider the model

$$\mathcal{M}_{A(\mathcal{G}, c)} = \{a^\top a \mid a \in A(\mathcal{G}, c)\}.$$

Proposition 2.7. *Fix a coloured DAG (\mathcal{G}, c) . The RDAG model on (\mathcal{G}, c) is equal to $\mathcal{M}_{A(\mathcal{G}, c)}$ if and only if the colouring c is compatible.*

Before proving the proposition, we recall two matrix decompositions. The LDL decomposition writes a positive definite matrix as $\Psi = LDL^\top$, where D is diagonal with positive entries, and L is lower triangular and unipotent (i.e. has ones on the diagonal). The LDL decomposition is closely related to the factorisation $\Psi = (\text{id} - \Lambda)^\top \Omega^{-1} (\text{id} - \Lambda)$ from (5): the LDL decomposition is $D = \Omega^{-1}$ and $L = (\text{id} - \Lambda)^\top$. Hence an RDAG model imposes zeros and symmetries in the LDL decomposition.

The second matrix decomposition is the Cholesky decomposition. It writes a positive definite matrix as the product $\Psi = a^\top a$, where a is upper triangular with positive diagonal entries. The model $\mathcal{M}_{A(\mathcal{G}, c)}$ imposes zeros and symmetries in the Cholesky decomposition, as follows.

Lemma 2.8. *Fix a coloured DAG (\mathcal{G}, c) with compatible colouring c . Then $\mathcal{M}_{A(\mathcal{G}, c)}$ is the set of matrices with Cholesky decomposition $a^\top a$ for some $a \in A(\mathcal{G}, c)$.*

Proof. The set $\mathcal{M}_{A(\mathcal{G}, c)}$ consists of all matrices Ψ of the form $a^\top a$ for some $a \in A(\mathcal{G}, c)$, see (1). The matrix a is upper triangular by the structure of \mathcal{G} . To get the Cholesky decomposition, it remains to modify a to have positive diagonal entries. We replace a by ka , where k is the diagonal matrix with $k_{ii} = 1$ if $a_{ii} > 0$ and $k_{ii} = -1$ if $a_{ii} < 0$. Then ka flips the sign of all rows of a with negative diagonal entry, hence it has all diagonal entries strictly positive. The compatibility of the colouring ensures that $a_{ij} = a_{kl}$ can only hold in $A(\mathcal{G}, c)$ if $a_{ii} = a_{kk}$. Hence multiplying by k doesn't break any edge compatibility conditions, and $ka \in A(\mathcal{G}, c)$. \square

The LDL and Cholesky decompositions are both unique, since Ψ is (strictly) positive definite. They are related by:

$$\begin{aligned} \text{Cholesky from LDL: } & a = D^{1/2}L^\top, \\ \text{LDL from Cholesky: } & D = \text{diag}(a_{11}^2, \dots, a_{mm}^2), \quad L^\top = D^{-1/2}a. \end{aligned}$$

The following lemma is proved by comparing zero patterns in the two decompositions.

Lemma 2.9. *The DAG model on \mathcal{G} is the model $\mathcal{M}_{A(\mathcal{G})}$.*

Proof. The LDL decomposition of $\Psi = (\text{id} - \Lambda)^\top \Omega^{-1} (\text{id} - \Lambda)$ is given by $D = \Omega^{-1}$ and $L = (\text{id} - \Lambda)^\top$. The Cholesky decomposition has

$$a = D^{1/2} L^\top = \Omega^{-1/2} (\text{id} - \Lambda), \quad \text{i.e.} \quad a_{ij} = \begin{cases} \omega_{ii}^{-1/2} & \text{if } i = j \\ -\omega_{ii}^{-1/2} \lambda_{ij} & \text{if } i \neq j. \end{cases} \quad (8)$$

We have containment of the DAG model inside $\mathcal{M}_{A(\mathcal{G})}$, because if $j \not\rightarrow i$ in \mathcal{G} , then $\lambda_{ij} = 0$ and therefore $a_{ij} = 0$. Conversely, given $\Psi \in \mathcal{M}_{A(\mathcal{G})}$, its Cholesky decomposition is $a^\top a$ for some $a \in A(\mathcal{G})$ by Lemma 2.8, since the colouring that assigns all vertices and edges different colours satisfies $A(\mathcal{G}) = A(\mathcal{G}, c)$ and is compatible. Hence $a_{ij} = 0$ for $i \neq j$ unless $j \rightarrow i$ in \mathcal{G} . We therefore have $(\text{id} - \Lambda)_{ij} = L_{ji} = \frac{1}{a_{ii}} a_{ij} = 0$, where $a_{ii} \neq 0$, and hence $\lambda_{ij} = 0$. \square

We prove Proposition 2.7, by comparing symmetries in the two decompositions.

Proof of Proposition 2.7. Given some Ψ in the RDAG model, we show that its Cholesky decomposition is $\Psi = a^\top a$ with $a \in A(\mathcal{G}, c)$. By Lemma 2.9, $a \in A(\mathcal{G})$ and it remains to show that the colour conditions hold, i.e. that $a_{ij} = a_{kl}$ whenever $c(ij) = c(kl)$. If $c(ii) = c(kk)$, then $\omega_{ii} = \omega_{kk}$ since Ω respects the colouring. This shows that $a_{ii} = a_{kk}$, using (8). If $c(ij) = c(kl)$ for edges $j \rightarrow i$ and $l \rightarrow k$, then $\lambda_{ij} = \lambda_{kl}$ since Λ respects the colouring and, moreover, $\omega_{ii} = \omega_{kk}$ by compatibility. This implies $a_{ij} = a_{kl}$, by (8).

Conversely, given $\Psi \in \mathcal{M}_{A(\mathcal{G}, c)}$, we show that Ψ is in the RDAG model. The Cholesky decomposition is $\Psi = a^\top a$ for $a \in A(\mathcal{G}, c)$, by Lemma 2.8. The entries of Ω and Λ are $\omega_{ii} = a_{ii}^{-2}$ and $\lambda_{ij} = -a_{ii}^{-1} a_{ij}$, by (8), which satisfy the RDAG model conditions. Hence a compatible colouring implies the equivalence of the RDAG model on (\mathcal{G}, c) and $\mathcal{M}_{A(\mathcal{G}, c)}$.

If the colouring is not compatible, we exhibit some Ψ in the RDAG model that is not in $\mathcal{M}_{A(\mathcal{G}, c)}$. Let $\Psi = a^\top a$ be the Cholesky decomposition. If there is some $a' \in A(\mathcal{G}, c)$ with $\Psi = (a')^\top a'$ then, similar to the proof of Lemma 2.8, there is a diagonal matrix o with entries ± 1 with $oa' = a$. First, if Definition 2.5(i) does not hold, there is a vertex k and an edge $j \rightarrow i$ with $c(k) = c(ij)$. The RDAG model imposes no relation between ω_{kk} and λ_{ij} , so let Ψ be given by some Ω and Λ with $\omega_{kk} = 1$ and $\lambda_{ij} = 0$. Then $o_{kk} a'_{kk} = a_{kk} = 1$ and $o_{ii} a'_{ij} = a_{ij} = 0$, by (8). Hence, $a'_{kk} \neq 0 = a'_{ij}$ and therefore $\Psi \notin \mathcal{M}_{A(\mathcal{G}, c)}$. Second, if Definition 2.5(ii) does not hold, then there exist edges $j \rightarrow i$ and $l \rightarrow k$ with $c(ij) = c(kl)$ but $c(i) \neq c(k)$. We choose Ψ given by some Ω and Λ with $\omega_{ii} = 1$, $\omega_{kk} = \frac{1}{4}$ and $\lambda_{ij} = \lambda_{kl} = -1$. Then $o_{ii} a'_{ij} = a_{ij} = 1$ and $o_{kk} a'_{kl} = a_{kl} = 2$, by (8). Thus, $|a'_{ij}| = 1 \neq 2 = |a'_{kl}|$ and, again, we cannot have $\Psi \in \mathcal{M}_{A(\mathcal{G}, c)}$. \square

Example 2.10. We return to the graph $\textcircled{1} \leftarrow \textcircled{3} \rightarrow \textcircled{2}$ from Examples 2.2 and 2.4. The colouring is compatible, because the set {blue, black} of vertex colours

is disjoint from the edge colour set {red}, and the children of both red edges have the same colour. Hence Proposition 2.7 shows that the RDAG model is equal to

$$\mathcal{M}_{A(\mathcal{G},c)} = \left\{ \left(\begin{array}{ccc} d_1^2 & 0 & rd_1 \\ 0 & d_1^2 & rd_1 \\ rd_1 & rd_1 & 2r^2 + d_2^2 \end{array} \right) \mid d_1, d_2 \neq 0 \right\}. \tag{9}$$

Remark 2.11. RDAG models can also be defined over the complex numbers. Here, the parameters Λ can be complex, and we obtain a subset of PD_m by taking conjugate transposes, $\Psi = (\text{id} - \Lambda)^\dagger \Omega^{-1} (\text{id} - \Lambda)$. For the \mathcal{M}_A characterisation, we replace $a^\text{T}a$ by $a^\dagger a$. Many of our results and proofs can be modified to hold in the complex setting. We return to complex RDAGs in Section B.2.

3. Comparison of RDAG and RCON models

Given a directed graph, we can forget the direction of each edge to give an undirected graph. We characterise when the RDAG model on the coloured directed graph is equal to the undirected model on the corresponding coloured undirected graph in Theorem 3.4. We begin by comparing RDAG and RCON models in two examples.

Example 3.1 (RDAG = RCON). We revisit our example $\textcircled{1} \leftarrow \boxed{3} \rightarrow \textcircled{2}$. The corresponding RCON model has coloured undirected graph $\textcircled{1} - \boxed{3} - \textcircled{2}$, with blue (circular) vertices $\{1, 2\}$, black (square) vertex 3, and red edges. By Definition 1.3, the RCON model is the set of positive definite matrices of the form

$$\Psi = \begin{pmatrix} \delta_1 & 0 & \varrho \\ 0 & \delta_1 & \varrho \\ \varrho & \varrho & \delta_2 \end{pmatrix}.$$

Since the colouring is compatible, the RDAG model is equal to $\mathcal{M}_{A(\mathcal{G},c)}$ from (9). Any matrix in $\mathcal{M}_{A(\mathcal{G},c)}$ satisfies the equalities for the RCON model, so we have containment of the RDAG model in the RCON model. Conversely, given Ψ in the RCON model,

$$\det(\Psi) = \delta_1^2(\delta_2 - 2\varrho^2\delta_1^{-1}) > 0 \quad \text{hence} \quad \delta_2 - 2\varrho^2\delta_1^{-1} > 0,$$

since Ψ is positive-definite. Positive definiteness of Ψ also implies $\delta_1, \delta_2 > 0$. We obtain real numbers $d_1 := \sqrt{\delta_1}$, $r := \varrho/d_1$ and $d_2 := \sqrt{\delta_2 - 2\varrho^2\delta_1^{-1}}$, which shows that Ψ is of the form in (9).

Example 3.2 (RDAG \neq RCON). Consider the RDAG model on $\textcircled{1} \leftarrow \textcircled{2}$, the graph with two blue (circular) vertices $\{1, 2\}$ and a red edge. The colouring is compatible, so by Proposition 2.7 the RDAG model is $\mathcal{M}_{A(\mathcal{G},c)}$, where

$$A(\mathcal{G}, c) = \left\{ \begin{pmatrix} d & r \\ 0 & d \end{pmatrix} \mid d \neq 0 \right\}.$$

The corresponding RCON model is given by the coloured undirected graph $\textcircled{1} - \textcircled{2}$. The RCON model consists of all $\Psi \in \text{PD}_2$ with $\Psi_{11} = \Psi_{22}$ and $\Psi_{12} = \Psi_{21}$, by Definition 1.3. Neither model is contained in the other: the RCON model contains

$$\begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & \sqrt{3} \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & \sqrt{3} \end{pmatrix}$$

but the diagonal entries 2 and $\sqrt{3}$ in the Cholesky decomposition do not satisfy the conditions for $A(\mathcal{G}, c)$. Conversely, the matrix

$$\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$$

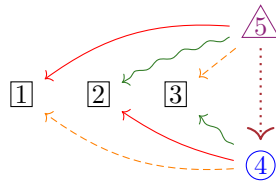
is in the RDAG model, but not the RCON model, since $\Psi_{11} \neq \Psi_{22}$.

To characterize when an RDAG model is equal to its corresponding RCON model, we give two constructions of coloured graphs obtained from some (\mathcal{G}, c) , one that is built from a vertex and the other from an edge. As before, $\mathcal{G} = (I, E)$.

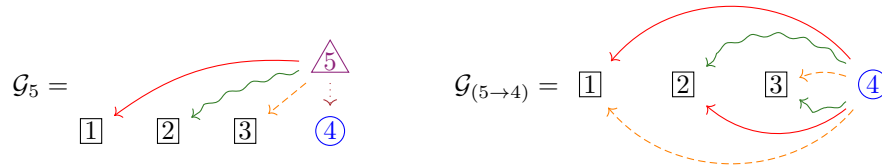
Fix a vertex $i \in I$. Recall that the children of i are the vertices k with $(i \rightarrow k) \in E$. Consider the subgraph on vertex set $\{i\} \cup \text{ch}(i)$ with edges $i \rightarrow k$ for each $k \in \text{ch}(i)$, and colours inherited from (\mathcal{G}, c) . We denote the subgraph by \mathcal{G}_i .

Now fix an edge $(j \rightarrow i) \in E$. Consider vertices $\{i\} \cup (\text{ch}(i) \cap \text{ch}(j))$ with vertex colours inherited from (\mathcal{G}, c) . For each $k \in \text{ch}(i) \cap \text{ch}(j)$, we introduce two edges $i \rightarrow k$, one with colour $c(ki)$ and the other with colour $c(kj)$. We denote this graph by $\mathcal{G}_{(j \rightarrow i)}$.

Example 3.3. Consider the graph¹



The vertex construction at vertex 5 and edge construction at edge $5 \rightarrow 4$ are:



Given a DAG \mathcal{G} , its corresponding undirected graph is denoted \mathcal{G}^u . Similarly, given a coloured DAG (\mathcal{G}, c) , its corresponding undirected coloured graph (where

¹With three vertex colours (blue/circular, black/square, and purple/triangular) and four edge colours (red/solid, green/squiggly, orange/dashed, and brown/dotted).

the colour of a directed edge becomes the colour of the undirected edge) is (\mathcal{G}^u, c) . Two coloured graphs (\mathcal{G}, c) and (\mathcal{G}', c') are *isomorphic* if the coloured graphs are the same up to relabelling vertices. We denote an isomorphism by $\mathcal{G} \simeq \mathcal{G}'$ when the colouring is clear. An *unshielded collider* of a DAG \mathcal{G} is an induced subgraph $j \rightarrow i \leftarrow k$ that has no edge between j and k .

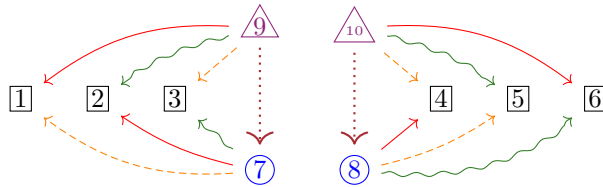
Theorem 3.4. *Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible. The RDAG model on (\mathcal{G}, c) is equal to the RCON model on (\mathcal{G}^u, c) if and only if:*

- (a) \mathcal{G} has no unshielded colliders;
- (b) $\mathcal{G}_i \simeq \mathcal{G}_j$ for every pair of vertices i, j of the same colour; and
- (c) $\mathcal{G}_{(j \rightarrow i)} \simeq \mathcal{G}_{(l \rightarrow k)}$ for every pair of edges $j \rightarrow i$ and $l \rightarrow k$ of the same colour.

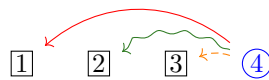
We prove Theorem 3.4 at the end of this section. It is proved via Propositions 3.8 and 3.9. First, we illustrate Theorem 3.4 in two examples.

Example 3.5. Our running example $\textcircled{1} \leftarrow \textcircled{3} \rightarrow \textcircled{2}$ satisfies the conditions of Theorem 3.4: it has no unshielded colliders, the graphs \mathcal{G}_1 and \mathcal{G}_2 both consist of a single blue vertex, and since $\text{ch}(3) \cap \text{ch}(1) = \emptyset$ and $\text{ch}(3) \cap \text{ch}(2) = \emptyset$ hold, the graphs $\mathcal{G}_{(3 \rightarrow 1)}$ and $\mathcal{G}_{(3 \rightarrow 2)}$ consist only of a single blue vertex. The RDAG and RCON models are therefore equivalent, as we saw in Example 3.1.

Example 3.6. The following graph also satisfies the conditions of Theorem 3.4:



- (a) It has no unshielded colliders.
- (b) For the black (square) vertices, the graphs \mathcal{G}_i consist of one black vertex. For the blue (circular) vertices, the \mathcal{G}_i are isomorphic to



The purple (triangular) vertices have \mathcal{G}_i isomorphic to \mathcal{G}_5 from Example 3.3.

- (c) All edges $j \rightarrow i$ have $\text{ch}(j) \cap \text{ch}(i) = \emptyset$, except for the two brown edges. For these, $\mathcal{G}_{(10 \rightarrow 8)}$ and $\mathcal{G}_{(9 \rightarrow 7)}$ are both isomorphic to $\mathcal{G}_{(5 \rightarrow 4)}$ from Example 3.3.

Hence the RDAG model on this coloured graph is equal to the RCON model on the underlying undirected graph. Note that the two connected components of (\mathcal{G}, c) are not isomorphic. We will see why this is not required for the proof of Theorem 3.4, i.e. why we can collapse vertices i and j in the definition of $\mathcal{G}_{(j \rightarrow i)}$.

One ingredient to our proof of Theorem 3.4 is the condition for a DAG model to be equal to its corresponding undirected graphical model.

Theorem 3.7 (Gaussian special case of [3, Theorem 3.1], [18, Theorem 5.6]). *The DAG model on \mathcal{G} is equal to the undirected Gaussian graphical model on \mathcal{G}^u if and only if \mathcal{G} has no unshielded colliders.*

We now prove Theorem 3.4 via two propositions.

Proposition 3.8. *Let (\mathcal{G}, c) be a coloured DAG with compatible colouring c . The RDAG model on (\mathcal{G}, c) is contained in the RCON model on (\mathcal{G}^u, c) if and only if*

- (a) \mathcal{G} has no unshielded colliders;
- (b) $\mathcal{G}_i \simeq \mathcal{G}_j$ for every pair of vertices i, j of the same colour; and
- (c) $\mathcal{G}_{(j \rightarrow i)} \simeq \mathcal{G}_{(l \rightarrow k)}$ for every pair of edges $j \rightarrow i$ and $l \rightarrow k$ of the same colour.

Proof. Since the colouring is compatible, the RDAG model is $\mathcal{M}_{A(\mathcal{G}, c)}$, by Proposition 2.7. Let $a \in A(\mathcal{G}, c)$ be a general matrix. We think of it as having indeterminate entries, one for each vertex colour and one for each edge colour.

If the RDAG model is contained in the RCON model, we must have $(a^\top a)_{ij} = 0$ whenever $a_{ij} = a_{ji} = 0$. This holds iff there are no unshielded colliders in \mathcal{G} , by Theorem 3.7. Moreover, certain equalities must hold on $a^\top a$. We have vertex colour condition $(a^\top a)_{ii} = (a^\top a)_{jj}$ whenever $c(i) = c(j)$ and edge colour condition $(a^\top a)_{ij} = (a^\top a)_{kl}$ whenever $c(ij) = c(kl)$. These give the polynomial identities

$$a_{ii}^2 + \sum_{k \in \text{ch}(i)} a_{ki}^2 = a_{jj}^2 + \sum_{l \in \text{ch}(j)} a_{lj}^2 \quad \text{whenever } c(i) = c(j) \quad (10)$$

$$a_{ii}a_{ij} + \sum_{p \neq i, j}^m a_{pi}a_{pj} = a_{kk}a_{kl} + \sum_{q \neq k, l}^m a_{qk}a_{ql} \quad \text{whenever } c(ij) = c(kl). \quad (11)$$

We show that (10) is equivalent to (b) and that (11) is equivalent to (c).

Given two vertices i, j with $c(i) = c(j)$, we have $a_{ii}^2 = a_{jj}^2$. The sums in (10) are equal if and only if $|\text{ch}(i)| = |\text{ch}(j)|$ and the edge colours in \mathcal{G}_i and \mathcal{G}_j agree (counted with multiplicity). By compatibility, the vertex colours also agree, hence (10) is equivalent to $\mathcal{G}_i \simeq \mathcal{G}_j$.

Next, we consider (11). No terms $a_{ji}a_{jj}$ and $a_{lk}a_{ll}$ appear, since there is no edge $i \rightarrow j$ or $k \rightarrow l$. The compatibility of the colouring gives $a_{ii} = a_{kk}$. Hence $a_{ii}a_{ij} = a_{kk}a_{kl}$. A summand in (11) vanishes unless $p \in \text{ch}(i) \cap \text{ch}(j)$ or $q \in \text{ch}(k) \cap \text{ch}(l)$. The sums are equal if and only if $|\text{ch}(i) \cap \text{ch}(j)| = |\text{ch}(k) \cap \text{ch}(l)|$ and the graphs $\mathcal{G}_{(j \rightarrow i)}$ and $\mathcal{G}_{(l \rightarrow k)}$ are isomorphic on their edge colours. By compatibility of the colouring, the vertex colours of the children also agree and $c(i) = c(k)$, hence (11) is equivalent to $\mathcal{G}_{(j \rightarrow i)} \simeq \mathcal{G}_{(l \rightarrow k)}$. \square

Proposition 3.9. *Let (\mathcal{G}, c) be a coloured DAG with compatible colouring c such that*

- (a) \mathcal{G} has no unshielded colliders;
- (b) $\mathcal{G}_i \simeq \mathcal{G}_j$ for every pair of vertices i, j of the same colour; and
- (c) $\mathcal{G}_{(j \rightarrow i)} \simeq \mathcal{G}_{(l \rightarrow k)}$ for every pair of edges $j \rightarrow i$ and $l \rightarrow k$ of the same colour.

Then the RCON model on (\mathcal{G}^u, c) is contained in the RDAG model on (\mathcal{G}, c) .

Proof. Given a matrix Ψ in the RCON model on (\mathcal{G}^u, c) , we show that it is contained in the RDAG model on (\mathcal{G}, c) by showing that the Cholesky decomposition $\Psi = a^\top a$ satisfies the conditions of the set $A(\mathcal{G}, c)$ in (7). The matrix a is upper triangular. Completing it one column at a time shows that its entries are

$$a_{l,l} = \left(\Psi_{l,l} - \sum_{p \in \text{ch}(l)} a_{p,l}^2 \right)^{1/2} \tag{12}$$

$$a_{i,j} = \left(\Psi_{i,j} - \sum_{p \in \text{ch}(i) \cap \text{ch}(j)} a_{p,i} a_{p,j} \right) a_{i,i}^{-1}. \tag{13}$$

Note that the expression under the square root in (12) is a positive real number, see [40, Lecture 23]. We need to show that $a_{ij} = 0$ for $i \neq j$ with $j \not\rightarrow i$ (the support conditions) and that $a_{ij} = a_{kl}$ whenever $c(ij) = c(kl)$ (the colour conditions).

First we show that a satisfies the support conditions of $A(\mathcal{G}, c)$. If there is no edge from j to i in \mathcal{G} , then \mathcal{G}^u has no edge between i and j , so $\Psi_{i,j} = 0$. Moreover, all products $a_{p,i} a_{p,j}$ vanish, otherwise $i \rightarrow p \leftarrow j$ would be an unshielded collider. Hence $a_{ij} = 0$ if $j \not\rightarrow i$ in \mathcal{G} .

Next, we show that a satisfies the colour conditions of $A(\mathcal{G}, c)$. We prove this inductively over the top left $k \times k$ blocks of a . If $k = 1$ there are no symmetries to check. We assume that the top left $k \times k$ submatrix of a satisfies the symmetries. For the induction step, we compare $a_{1,k+1}, a_{2,k+1}, \dots, a_{k+1,k+1}$ with each other and with $a_{i,j}$, $i, j \in [k]$.

If there is an edge $(k + 1) \rightarrow 1$ with same colour as $j \rightarrow i$ for $i, j \in [k]$, we need to show that $a_{1,k+1} = a_{i,j}$. First, $a_{11} = a_{ii}$ by compatibility and the induction hypothesis, and $\Psi_{i,j} = \Psi_{1,k+1}$ since Ψ is in the RCON model. Moreover, all a_{pq} for $p, q \in [k]$ respect the symmetries. Since $\mathcal{G}_{(j \rightarrow i)} \simeq \mathcal{G}_{(k+1 \rightarrow 1)}$, the expressions (13) for $a_{i,j}$ and $a_{1,k+1}$ are equal. Proceeding inductively, we show that all entries $a_{2,k+1}, \dots, a_{k,k+1}$ respect the symmetries of c .

Finally, if vertex $k + 1$ has same colour as vertex $l \in [k]$, we show $a_{k+1,k+1} = a_{l,l}$. We have $\mathcal{G}_l \simeq \mathcal{G}_{k+1}$ by assumption (b) and $\Psi_{l,l} = \Psi_{k+1,k+1}$, since Ψ is in the RCON model. Furthermore, we have shown that all $a_{p,q}$, where $p \in [k]$ and $q \in [k + 1]$, satisfy colouring c . Hence (12) implies $a_{l,l} = a_{k+1,k+1}$. \square

Proof of Theorem 3.4. If any of conditions (a),(b), and (c) do not hold, this rules out containment of the RDAG model in the RCON model, by Proposition 3.8, and hence rules out the two models being equal. If conditions (a),(b),(c) hold, we have containment of the RDAG model inside the RCON model (by Proposition 3.8) and the reverse containment (by Proposition 3.9). \square

4. MLE: existence, uniqueness, and an algorithm

In this section we characterise the existence and uniqueness of the MLE in an RDAG model via linear dependence conditions on certain matrices. Our description specialises to give the characterisation of existence and uniqueness of the MLE in a usual DAG model in terms of linear dependence among the rows of the sample matrix.

Let α_s be the number of vertices of colour s . For an edge $j \rightarrow i$, vertex j is called the *source* of the edge and i is called the *target*. The *parent relationship colours* are the colours of all edges with a target of colour s :

$$\text{prc}(s) = \{c(ij) \mid \text{there exists } j \rightarrow i \text{ in } \mathcal{G} \text{ with } c(i) = s\}, \quad \beta_s := |\text{prc}(s)|.$$

A sample matrix is $Y \in \mathbb{R}^{m \times n}$; its m rows index the vertices in \mathcal{G} , and its columns are the n samples. For each vertex colour in \mathcal{G} we define an augmented sample matrix.

Definition 4.1. The *augmented sample matrix* of sample matrix Y and vertex colour s , denoted $M_{Y,s}$, has size $(\beta_s + 1) \times \alpha_s n$. We construct it row by row. Each row consists of α_s blocks, each a vector of length n . For notational simplicity, we assume for now that the vertices of colour s are the set $\{1, 2, \dots, \alpha_s\} \subset I$. Then the top row of $M_{Y,s}$ is

$$(Y^{(1)} \quad Y^{(2)} \quad \dots \quad Y^{(\alpha_s)}) \in \mathbb{R}^{\alpha_s n},$$

where $Y^{(i)} \in \mathbb{R}^n$ is the i th row of sample matrix $Y \in \mathbb{R}^{m \times n}$. The other rows of $M_{Y,s}$ are indexed by the parent relationship colours $\text{prc}(s)$. The row indexed by $t \in \text{prc}(s)$ is

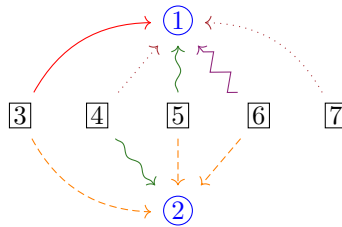
$$\left(\begin{array}{ccc} \sum_{\substack{1 \leftarrow j \\ c(1j)=t}} Y^{(j)} & \sum_{\substack{2 \leftarrow j \\ c(2j)=t}} Y^{(j)} & \dots & \sum_{\substack{\alpha_s \leftarrow j \\ c(\alpha_s j)=t}} Y^{(j)} \end{array} \right).$$

Note that the sum at the k th block is zero if there are no $j \rightarrow k$ of colour t . Let $M_{Y,s}^{(i)}$ denote the i th row of $M_{Y,s}$, where we index from 0 to β_s .

Example 4.2. Our running example $\textcircled{1} \leftarrow \textcircled{3} \rightarrow \textcircled{2}$ has two augmented sample matrices, one for each vertex colour:

$$M_{Y,\textcircled{1}} = \begin{pmatrix} Y^{(1)} & Y^{(2)} \\ Y^{(3)} & Y^{(3)} \end{pmatrix}_{\textcircled{1}} \in \mathbb{R}^{2 \times 2n} \quad \text{and} \quad M_{Y,\textcircled{2}} = (Y^{(3)}) \in \mathbb{R}^{1 \times n}. \quad (14)$$

Example 4.3. The RDAG model on the coloured DAG



$$\text{has } M_{Y,\circ} = \begin{pmatrix} Y^{(1)} & Y^{(2)} & \circ \\ Y^{(3)} & 0 & \rightarrow \\ 0 & Y^{(3)} + Y^{(5)} + Y^{(6)} & \dashrightarrow \\ Y^{(5)} & Y^{(4)} & \curvearrowright \\ Y^{(6)} & 0 & \curvearrowleft \\ Y^{(4)} + Y^{(7)} & 0 & \cdots \rightarrow \end{pmatrix}$$

Theorem 4.4. Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible, and fix sample matrix $Y \in \mathbb{R}^{m \times n}$. The following possibilities characterise maximum likelihood estimation given Y :

- (a) ℓ_Y unbounded from above $\Leftrightarrow \exists s \in c(I): M_{Y,s}^{(0)} \in \text{span}\{M_{Y,s}^{(i)} : i \in [\beta_s]\}$
- (b) MLE exists $\Leftrightarrow \forall s \in c(I): M_{Y,s}^{(0)} \notin \text{span}\{M_{Y,s}^{(i)} : i \in [\beta_s]\}$
- (c) MLE exists uniquely $\Leftrightarrow \forall s \in c(I): M_{Y,s}$ has full row rank.

We prove Theorem 4.4 in Section 4.2.

Example 4.5. For our running example $\textcircled{1} \leftarrow \textcircled{3} \rightarrow \textcircled{2}$, Theorem 4.4 says that the MLE exists uniquely provided $Y^{(3)} \neq 0$ and $(Y^{(1)} \ Y^{(2)})$ is not parallel to $(Y^{(3)} \ Y^{(3)})$. This holds almost surely as soon as we have one sample, as we mentioned in Example 0.1.

Example 4.6. Returning to Example 4.3, the MLE given Y exists provided $M_{Y,\square} = (Y^{(3)} \ \dots \ Y^{(7)}) \neq 0$, and $(Y^{(1)} \ Y^{(2)})$ is not in the span of the other rows of $M_{Y,\circ}$. The MLE is unique if and only if $M_{Y,\circ}$ is full row rank, since this also implies $M_{Y,\square} \neq 0$.

The proof of Theorem 4.4 gives Algorithm 1 for finding the MLE in an RDAG model with compatible colouring. The MLE is returned as entries of the matrices Λ and Ω . We give the MLE in a closed-form formula, as a collection of least squares estimators.

Remark 4.7. In Algorithm 1, the entries of Ω are given as $\{\omega_{ss} : s \in c(I)\}$. The entries of Λ are returned as $\{\lambda_{s,t} : s \in c(I), t \in \text{prc}(s)\}$, which equals the set of edge colours by compatibility, since edge colour t only appears in $\text{prc}(s)$ for one s .

The proof of Theorem 4.4 directly gives a description of the set of MLEs.

Corollary 4.8. Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible, with sample matrix $Y \in \mathbb{R}^{m \times n}$. If (Λ, Ω) and (Λ', Ω') are two MLEs, then $\Omega = \Omega'$ and

$$\sum_{t \in \text{prc}(s)} (\lambda_{s,t} - \lambda'_{s,t}) M_{Y,s}^{(t)} = 0, \quad \text{for all } s \in c(I).$$

4.1. Usual DAG models

We recall the characterisation of MLE existence and uniqueness for usual Gaussian graphical models, as linear dependence conditions on the sample matrix.

Algorithm 1: Computing the MLE for an RDAG model

input : A coloured DAG (\mathcal{G}, c) and sample matrix Y
output: An MLE given Y in the RDAG model on (\mathcal{G}, c) , if one exists

for $s \in c(I)$ **do**
 $\alpha_s := |c^{-1}(s)|$;
 $\beta_s := |\text{prc}(s)|$;
 construct matrix $M_{Y,s} \in \mathbb{R}^{(\beta_s+1) \times \alpha_s n}$;
 $P_{Y,s} :=$ orthogonal projection of $M_{Y,s}^{(0)}$ onto $\text{span}\{M_{Y,s}^{(t)} : t \in [\beta_s]\}$;
 if $P_{Y,s} = M_{Y,s}^{(0)}$ **then**
 return *MLE does not exist*;
 else
 coefficients $\lambda_{s,t}$ are such that $P_{Y,s} = \sum_{t \in \text{prc}(s)} \lambda_{s,t} M_{Y,s}^{(t)}$;
 $\omega_{ss} := (\alpha_s n)^{-1} \|P_{Y,s} - M_{Y,s}^{(0)}\|^2$;
 end
end
return *MLE for Λ and Ω*

For a DAG \mathcal{G} on m nodes, and n data samples, the sample matrix is $Y \in \mathbb{R}^{m \times n}$. For a node i in \mathcal{G} we denote by $Y^{(i)}$ the i th row of Y , by $Y^{(\text{pa}(i))}$ the sub-matrix of Y with rows indexed by the parents of i in \mathcal{G} , and by $Y^{(\text{pa}(i) \cup i)}$ the sub-matrix of Y with rows indexed by node i and its parents.

Theorem 4.9. *Consider the DAG model on \mathcal{G} , with m nodes, and fix sample matrix $Y \in \mathbb{R}^{m \times n}$. The following possibilities characterise maximum likelihood estimation given Y :*

- (a) ℓ_Y unbounded from above $\Leftrightarrow \exists i \in [m] : Y^{(i)} \in \text{span}\{Y^{(j)} : j \in \text{pa}(i)\}$
- (b) MLE exists $\Leftrightarrow \forall i \in [m] : Y^{(i)} \notin \text{span}\{Y^{(j)} : j \in \text{pa}(i)\}$
- (c) MLE exists uniquely $\Leftrightarrow \forall i \in [m] : Y^{(\text{pa}(i) \cup i)}$ has full row rank.

The above result is known to experts working in the area. Closely related results are in [23, Section 5.4.1] and [12], but we are not aware of a statement in the literature that matches the above. The result follows from viewing maximum likelihood estimation in a DAG as a sequence of regression problems. The acyclicity ensures that the sub-problems are uncoupled. We give a proof for completeness and to see its generalisation in Theorem 4.4.

Proof. We denote the entries of the MLEs Λ and Ω by $\hat{\lambda}_{ij}$ and $\hat{\omega}_{kk}$. The negative of the log-likelihood ℓ_Y , in terms of the parameters ω_{ii} and λ_{ij} , is

$$\sum_{i=1}^m \left(\log \omega_{ii} + \frac{1}{n\omega_{ii}} \|Y^{(i)} - \sum_{j \in \text{pa}(i)} \lambda_{ij} Y^{(j)}\|^2 \right).$$

We minimise the above expression. Each parameter only appears in one summand. In the i th summand, the $\hat{\lambda}_{ij}$ always exist: they are coefficients of each $Y^{(j)}$ in the orthogonal projection of $Y^{(i)}$ onto the span of $\{Y^{(j)} : j \in \text{pa}(i)\}$.

The $\hat{\lambda}_{ij}$ are unique if and only if $Y^{(\text{pa}(i))}$ has full row rank. Let ζ_i be the residual $\|Y^{(i)} - \sum_{j \in \text{pa}(i)} \hat{\lambda}_{ij} Y^{(j)}\|^2$. Let $\hat{\omega}_{ii}$ minimise $\log(\omega_{ii}) + \frac{\zeta_i}{n\omega_{ii}}$. If $\zeta_i = 0$, equivalently if $Y^{(i)} \in \text{span}\{Y^{(j)} : j \in \text{pa}(i)\}$, then in the limit $\omega_{ii} \rightarrow 0$, the log-likelihood tends to infinity and $\hat{\omega}_{ii}$ does not exist. Otherwise, the minimum is attained uniquely at $\hat{\omega}_{ii} = \zeta_i/n$. Combining these cases gives the theorem. \square

4.2. Proof of Theorem 4.4

The proof of Theorem 4.4 will be similar to the proof for uncoloured models in Theorem 4.9. We start by proving the following lemma.

Lemma 4.10. Fix $\alpha > 0$ and, for $\gamma \geq 0$, consider the family of functions

$$f_\gamma: \mathbb{R}_{>0} \rightarrow \mathbb{R}, \quad x \mapsto \alpha \log(x) + \frac{\gamma}{x}.$$

- (i) If $\gamma = 0$, then f_γ is neither bounded from below nor bounded from above.
- (ii) If $\gamma > 0$, then f_γ attains a global minimum at $x_0 = \frac{\gamma}{\alpha}$ with function value $f_\gamma(\frac{\gamma}{\alpha}) = \alpha(\log(\gamma) - \log(\alpha) + 1)$.
- (iii) Given $\gamma_1 \geq \gamma_2 > 0$, we have $f_{\gamma_1}(\frac{\gamma_1}{\alpha}) \geq f_{\gamma_2}(\frac{\gamma_2}{\alpha})$ at the global minima.

Proof. Part (i) follows from the properties of log. To prove part (ii), one computes $f'_\gamma(x) = \frac{\alpha}{x} - \frac{\gamma}{x^2}$ for $x > 0$. For $x > 0$ we have

$$f'_\gamma(x) = 0 \iff \frac{\alpha}{x} = \frac{\gamma}{x^2} \iff \alpha x = \gamma \iff x = \frac{\gamma}{\alpha}.$$

Thus $x_0 := \frac{\gamma}{\alpha}$ is the only possible local extremum of f_γ . For $x > 0$,

$$f'_\gamma(x) > 0 \iff \frac{\alpha}{x} > \frac{\gamma}{x^2} \iff \alpha x > \gamma \iff x > \frac{\gamma}{\alpha}.$$

and similarly one has $f'_\gamma(x) < 0$ if and only if $x < \frac{\gamma}{\alpha} = x_0$. Therefore, x_0 is a global minimum of f_γ . One directly verifies the function value for $f_\gamma(x_0)$, and so part (iii) follows from the monotonicity of the logarithm. \square

Proof of Theorem 4.4. Since colouring c is compatible, the RDAG model equals $\mathcal{M}_{A(\mathcal{G},c)}$, by Proposition 2.7. That is, for $\Psi = (\text{id} - \Lambda)^\top \Omega^{-1} (\text{id} - \Lambda)$ in the RDAG model, the matrix $a = \Omega^{-1/2} (\text{id} - \Lambda)$ is in $\mathcal{M}_{A(\mathcal{G},c)}$ and satisfies $\Psi = a^\top a$. As usual, let $\alpha_s := |c^{-1}(s)|$ and $\beta_s := |\text{prc}(s)|$. The entry of Ω at vertex colour s is denoted ω_{ss} and the edge colour entries of Λ that point towards colour s are labelled by $\lambda_{s,t}$, $t \in [\beta_s]$. Using the construction of the matrices $M_{Y,s} \in \mathbb{R}^{(\beta_s+1) \times \alpha_s n}$ and that $\det(\text{id} - \Lambda) = 1$, we compute

$$\begin{aligned} -\ell_Y(\Psi) &= -\log \det(\Psi) + \text{tr}(\Psi S_Y) = \log \det(\Omega) + \frac{1}{n} \|a \cdot Y\|^2 \\ &= \log \left(\prod_{s \in c(I)} \omega_{ss}^{\alpha_s} \right) + \frac{1}{n} \sum_{s \in c(I)} \|\omega_{ss}^{-1/2} (M_{Y,s}^{(0)} - \sum_{t \in [\beta_s]} \lambda_{s,t} M_{Y,s}^{(t)})\|^2 \quad (15) \\ &= \sum_{s \in c(I)} \alpha_s \log(\omega_{ss}) + \frac{1}{n\omega_{ss}} \|M_{Y,s}^{(0)} - \sum_{t \in [\beta_s]} \lambda_{s,t} M_{Y,s}^{(t)}\|^2. \end{aligned}$$

An MLE is a minimiser of the above expression. Each parameter occurs in exactly one of the summands over $s \in c(I)$, because the $\text{pre}(s)$ partition the edge colours by compatibility. We therefore minimise each summand separately. By Lemma 4.10(iii) we can first determine $\hat{\lambda}_{s,t}$, $t \in [\beta_s]$ that minimise

$$\|M_{Y,s}^{(0)} - \sum_{t \in [\beta_s]} \lambda_{s,t} M_{Y,s}^{(t)}\|^2. \quad (16)$$

Such $\hat{\lambda}_{s,t}$ always exist: they are coefficients in the orthogonal projection $P_{Y,s}$ of $M_{Y,s}^{(0)}$ onto $\text{span}\{M_{Y,s}^{(t)} : t \in [\beta_s]\}$; i.e.,

$$P_{Y,s} = \sum_{t \in [\beta_s]} \hat{\lambda}_{s,t} M_{Y,s}^{(t)}.$$

Furthermore, $\hat{\lambda}_{s,t}$, $t \in [\beta_s]$ are unique if and only if the vectors $M_{Y,s}^{(t)}$, $t \in [\beta_s]$ are linearly independent. Denote the minimum value of (16) by ζ_s . We will apply Lemma 4.10 several times with $\gamma_s := \zeta_s/n$.

If $M_{Y,s}^{(0)} \in \text{span}\{M_{Y,s}^{(t)} : t \in [\beta_s]\}$ for some s , then $\zeta_s = 0$ and the summand $\alpha_s \log(\omega_{ss}) + \zeta_s/(n\omega_{ss})$ is not bounded from below for $\omega_{ss} > 0$, by Lemma 4.10(i). Hence ℓ_Y is not bounded from above, e.g. by setting $\omega_{s',s'} = 1$ and $\lambda_{s',t'} = 0$ for all $s' \in c(I) \setminus \{s\}$ and all $t' \in [\beta_{s'}]$. This proves “ \Leftarrow ” of (a).

If $M_{Y,s}^{(0)} \notin \text{span}\{M_{Y,s}^{(t)} : t \in [\beta_s]\}$, equivalently $\zeta_s > 0$, then the summand $\alpha_s \log(\omega_{ss}) + \zeta_s/(n\omega_{ss})$ has unique minimiser $\hat{\omega}_{ss} = \zeta_s/(n\alpha_s)$, by Lemma 4.10(ii). Hence, an MLE exists if $\zeta_s > 0$ for all $s \in c(I)$, which proves “ \Leftarrow ” in (b). As the right-hand sides of (a) and (b) are opposites, we have proved (a) and (b).

Since the $\hat{\omega}_{ss}$ are uniquely determined (if they exist), an MLE (if it exists) is unique if and only if for all $s \in c(I)$ the vectors $M_{Y,s}^{(t)}$, $t \in [\beta_s]$ are linearly independent. In combination with the condition for MLE existence from part (b) we deduce (c). \square

Remark 4.11 (Comparison with usual DAG models). The framework of RDAG models includes usual DAG models as a special case; namely, when each colour is used only once. In this case, Theorem 4.4 specialises to Theorem 4.9. We see in the next section that imposing colours in a DAG reduces the threshold number of samples required for existence and uniqueness of the MLE.

4.3. Illustrative examples

In this section we apply RDAG models to some small illustrative examples. We first apply our running example to model the effect of a mother’s height on her two daughters’ heights.

Example 4.12. The RDAG model on coloured graph $\textcircled{1} \leftarrow \textcircled{3} \rightarrow \textcircled{2}$ is given by

$$y_1 = \lambda y_3 + \epsilon_1, \quad y_2 = \lambda y_3 + \epsilon_2, \quad y_3 = \epsilon_3, \quad \epsilon_1, \epsilon_2 \sim N(0, \omega), \quad \epsilon_3 \sim N(0, \omega').$$

Let variable 3 be the height (in cm) of a woman and let variables 1 and 2 be, respectively, the heights of her younger and older daughter. Vertices 1 and 2 both being blue indicates that, conditional on the mother's height, the variance of the daughter's heights is the same. The edges both being red encodes that the dependence of a daughter's height on the mother's height is the same for both daughters.

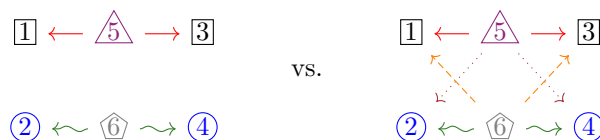
We saw in Example 4.5 that the MLE exists almost surely given one sample. We use Algorithm 1 to find the MLE, given one sample where the the younger daughter's height is 159.75cm, the older daughter's height is 161.56, the mother's height is 155.32, and the population mean height is 163.83cm. Mean-centring the data gives

$$(Y^{(1)} \ Y^{(2)} \ Y^{(3)}) = (-4.08 \ -2.27 \ -8.51).$$

The only black vertex is 3, and it has no parents, hence $\omega' = \|Y^{(3)}\|^2 = 72.42$. The orthogonal projection of $(Y^{(1)} \ Y^{(2)})$ onto $\text{span}\{(Y^{(3)} \ Y^{(3)})\}$ has coefficient $\lambda = 0.37$ and residual $\omega = [(-3.175 + 4.08)^2 + (-3.175 + 2.27)^2]/2 = 0.82$. As we would expect, the regression coefficient λ is positive and the variance of the daughters' heights conditional on the mother's height is lower than the variance of the mother's height.

We now consider multiple measurements taken in each generation.

Example 4.13. We consider measurements of the snout length and head length of dogs. These are the first two of the seven morphometric parameters in the study of clinical measurements of dog breeds in [28]. We compare two RDAG models:



The black/square vertices 1 and 3 are the snout lengths of the two offspring. Blue/circular vertices 2 and 4 are their head lengths. The purple/triangular vertex 5 is the snout length of the parent and grey/pentagonal vertex 6 is the head length of the parent. The edges encode the dependence of the offsprings' traits on those of the parents.

Maximum likelihood estimation in the left hand model is two copies of Example 4.12, one on the three odd variables, and one on the three even variables. Thus, given one sample, a unique MLE exists almost surely. For the right hand model, Theorem 4.4 says an MLE exists provided $Y^{(5)} \neq 0$, $Y^{(6)} \neq 0$ and neither $(Y^{(1)} \ Y^{(3)})$ nor $(Y^{(2)} \ Y^{(4)})$ are in $\text{span}\{(Y^{(5)} \ Y^{(5)}), (Y^{(6)} \ Y^{(6)})\}$. Hence an MLE exists almost surely with one sample. Moreover, the augmented sample matrices $M_{Y,\circ}$ and $M_{Y,\square}$ have full row rank almost surely provided $n \geq 2$, hence the MLE exists uniquely with two samples, by Theorem 4.4.

The vertex colours in an RDAG model could correspond to colours in an experiment, as follows. Fluorescent reporters can be used to take measurements

in a cell. In [37], the authors quantify data in a single living bacteria using fluorescent reporters in red, cyan, yellow, and green, see [37, Figure 2a]. Given such measurements taken for a parent cell and its daughter cells, we could consider analogues of Example 4.13 in which, for example, the red fluorescence in a daughter cell only depends on the red fluorescence in the parent, or larger models in which there can also be dependence on the fluorescence of other colours.

5. Maximum likelihood thresholds

In the previous section we gave a characterisation of the existence and uniqueness of the MLE based on linear independence conditions. Here we turn this into results that depend only on the coloured graph, and that hold for a generic sample matrix. We give upper and lower bounds on the thresholds for almost sure existence and uniqueness of the MLE. The bounds hold whenever the sample matrix doesn't have extra linear dependencies among its rows.

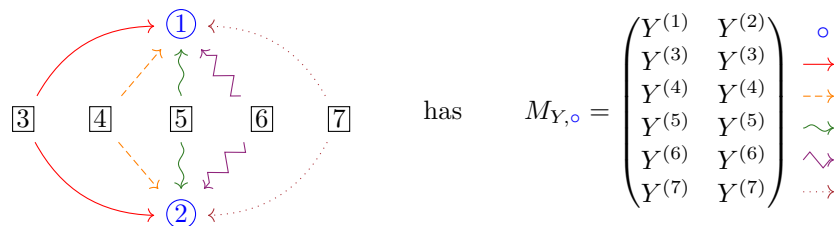
For a fixed number of samples, the MLE in an RDAG model may exist but not be unique almost surely, which cannot happen in an uncoloured model. In fact, Example 5.5 gives a family of models for which the gap between the existence and uniqueness thresholds becomes arbitrarily large.

As well as assuming compatibility of the colouring, we often assume in this section that there are no edges between vertices of the same colour. Some of our bounds involve the notion of generic rank. For this, a set $Z \subseteq \mathbb{R}^N$ is Zariski-closed if it is the common zero locus of some polynomials $p_1, \dots, p_k \in \mathbb{R}[x_1, \dots, x_N]$; see [27, Section 2.1] or [39, Section 3.2]. A property holds for generic $Y \in \mathbb{R}^{m \times n}$ if it holds for all $Y \in \mathbb{R}^{m \times n} \setminus Z$ where $Z \subsetneq \mathbb{R}^{m \times n}$ is Zariski-closed. A generic property holds almost surely with respect to absolutely continuous measures, since a Zariski-closed $Z \subsetneq \mathbb{R}^{m \times n}$ has Lebesgue measure zero.

Definition 5.1. Let M_Y be a matrix whose entries are linear combinations of the entries of a matrix Y . The *generic rank* of M_Y is its rank for generic Y .

For $Y \in \mathbb{R}^{m \times n}$, we often study the generic rank of M_Y by considering it as a symbolic matrix whose entries are linear forms in the mn indeterminates Y_{ij} .

Example 5.2. The graph²



²With two vertex colours (blue/circular and black/square) and five edge colours (red/solid, orange/dashed, green/squiggly, purple/zigzag, and brown/dotted).

When $n = 1$, the matrix $M_{Y,\circ}$ has generic rank two. Removing its top row gives a 5×2 matrix of generic rank one.

Define the augmented sample matrix $M_{Y,s} \in \mathbb{R}^{(\beta_s+1) \times \alpha_s n}$ as in Definition 4.1. We let $M'_{Y,s} \in \mathbb{R}^{\beta_s \times \alpha_s n}$ be obtained from $M_{Y,s}$ by removing its top row. As before, α_s is the number of vertices of colour s and β_s is the number of edge colours of edges to a vertex of colour s . Let r_s be the generic rank of $M'_{Y,s}$ when $n = 1$. Let mlt_e (resp. mlt_u) denote the minimal number of samples needed for almost sure existence (resp. uniqueness) of the MLE.

Theorem 5.3. *Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible, and (\mathcal{G}, c) has no edges between vertices of the same colour. For vertex colour s , set $l(s) := (r_s - 1)(\alpha_s - 1)^{-1}$ if $\alpha_s \geq 2$ and $l(s) := \beta_s$ if $\alpha_s = 1$. We have the following bounds on the thresholds mlt_e and mlt_u*

$$\max_s \lfloor l(s) \rfloor + 1 \leq \text{mlt}_e \leq \max_s \left\lfloor \frac{\beta_s}{\alpha_s} \right\rfloor + 1, \tag{17}$$

$$\max_s \left\lfloor \frac{\beta_s}{\alpha_s} \right\rfloor + 1 \leq \text{mlt}_u \leq \max_s (\beta_s + 2 - r_s). \tag{18}$$

We prove Theorem 5.3 in Section 5.1 below. It remains an open problem to turn the bounds in Theorem 5.3 into formulae for mlt_e and mlt_u .

Problem 5.4. *Determine the maximum likelihood thresholds mlt_e and mlt_u of an RDAG model, as formulae involving properties of the graph \mathcal{G} and its colouring c .*

Note that the upper bounds for existence and uniqueness are both at most $\max_s \beta_s + 1$, which is the threshold for uniqueness in the (uncoloured) DAG case, see [12]. Hence the RDAG thresholds are always at least as small as the DAG threshold.

Example 5.5. We find the existence and uniqueness thresholds for the RDAG on the graph in Example 5.2. The black (square) vertices have no parents, so the matrix $M_{Y,\square}$ is full rank as soon as $n \geq 1$. The thresholds are therefore determined by the matrix $M_{Y,\circ}$. The generic rank of $M'_{Y,\circ}$ is one when $n = 1$, i.e. $r_\circ = 1$. Theorem 5.3 gives the bounds

$$\frac{r_\circ - 1}{\alpha_\circ - 1} + 1 = 1 \leq \text{mlt}_e \quad \text{and} \quad \text{mlt}_u \leq \beta_\circ + 2 - r_\circ = 5 + 2 - 1 = 6.$$

We will later see in Proposition 5.10 that $\text{mlt}_u \leq \beta_\circ + 1 - r_\circ = 5$, since $r_\circ \neq \beta_\circ + 1 - (\beta_\circ/\alpha_\circ)$. In fact, the bounds $1 \leq \text{mlt}_e$ and $\text{mlt}_u \leq 5$ are attained, as follows. When $n = 1$, the row $M_{Y,\circ}^{(0)} = (Y^{(1)}, Y^{(2)})$ is almost surely not contained in the span of the other rows of $M_{Y,\circ}$, hence $\text{mlt}_e = 1$. On the other hand, we need at least $n = 5$ samples for generic linear independence of the rows $(Y^{(3)}, Y^{(3)}), \dots, (Y^{(7)}, Y^{(7)})$.

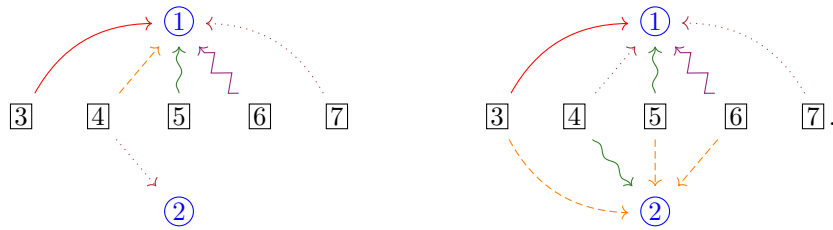
This example extends to an arbitrary number of vertices, i.e. the graph with $k + 2$ vertices, 2 blue/circular and k black/square, and $2k$ edges of k colours

(arranged as in the $k = 5$ case above). Repeating the above argument gives $\text{mlt}_e = 1$ and $\text{mlt}_u = k$.

Remark 5.6. Theorem 5.3 applies to all RDAG models with compatible colouring that are equal to some RCON model, because such models never have edges between vertices of the same colour, as follows. Take some vertex i of minimal index with $i \leftarrow j$ in \mathcal{G} having $c(i) = c(j)$. Then no children of i have colour $c(i)$, therefore $\mathcal{G}_i \neq \mathcal{G}_j$, a contradiction to Theorem 3.4(b).

We modify the edges from Examples 5.2 and 5.5 to see how the thresholds change.

Example 5.7. Consider the following DAGs, both with compatible colouring



Given sample matrix $Y \in \mathbb{R}^{7 \times n}$, we respectively obtain

$$M_{Y,\circ} = \begin{pmatrix} Y^{(1)} & Y^{(2)} & \circ \\ Y^{(3)} & 0 & \rightarrow \\ Y^{(4)} & 0 & \dashrightarrow \\ Y^{(5)} & 0 & \sim \\ Y^{(6)} & 0 & \sphericalangle \\ Y^{(7)} & Y^{(4)} & \cdots \end{pmatrix} \quad M_{Y,\circ} = \begin{pmatrix} Y^{(1)} & Y^{(2)} & \circ \\ Y^{(3)} & 0 & \rightarrow \\ 0 & Y^{(3)} + Y^{(5)} + Y^{(6)} & \dashrightarrow \\ Y^{(5)} & Y^{(4)} & \sim \\ Y^{(6)} & 0 & \sphericalangle \\ Y^{(4)} + Y^{(7)} & 0 & \cdots \end{pmatrix}$$

In both cases we have $\alpha_\circ = 2$, $\beta_\circ = 5$, and $r_\circ = 2$. Thus, Theorem 5.3 gives the bounds

$$2 = \left\lfloor \frac{r_\circ - 1}{\alpha_\circ - 1} \right\rfloor + 1 \leq \text{mlt}_e \leq \left\lfloor \frac{\beta_\circ}{\alpha_\circ} \right\rfloor + 1 = 3$$

$$3 = \left\lfloor \frac{\beta_\circ}{\alpha_\circ} \right\rfloor + 1 \leq \text{mlt}_u \leq \beta_\circ + 2 - r_\circ = 5.$$

In fact, Proposition 5.10 gives an upper bound of 4 on mlt_u , since $r_\circ \neq \beta_\circ + 1 - (\beta_\circ/\alpha_\circ)$.

First, we study the left-hand RDAG. When $n = 2$ the row $Y^{(2)} \in \mathbb{R}^{1 \times 2}$ is generically not in the span of $Y^{(4)}$, hence $M_{Y,\circ}^{(0)} = (Y^{(1)}, Y^{(2)})$ is not in the linear span of the other five rows of $M_{Y,\circ}$, and we deduce $\text{mlt}_e = 2$. For generic uniqueness of an MLE we need $M_{Y,\circ} \in \mathbb{R}^{6 \times 2n}$ to have full row rank six. For $n \geq 2$, the submatrix

$$\begin{pmatrix} Y^{(2)} \\ Y^{(4)} \end{pmatrix} \in \mathbb{R}^{2 \times n}$$

generically has rank two. Therefore, $M_{Y,\circ}$ has rank at most five if $n = 3$, but $n = 4$ suffices for $M_{Y,\circ}$ generically having full row rank. We conclude $\text{mlt}_u = 4$.

Next, we study the right-hand RDAG. For $n = 2$, $M_{Y,\circ}^{(0)} = (Y^{(1)}, Y^{(2)})$ is generically contained in the linear span of the other rows of $M_{Y,\circ}$. From the bounds we conclude that $\text{mlt}_e = 3$. For uniqueness, when $n = 3$ the submatrices

$$\begin{pmatrix} Y^{(3)} \\ Y^{(6)} \\ Y^{(4)} + Y^{(7)} \end{pmatrix}, \quad \begin{pmatrix} Y^{(2)} \\ Y^{(3)} + Y^{(5)} + Y^{(6)} \\ Y^{(4)} \end{pmatrix} \in \mathbb{R}^{3 \times 3}$$

of $M_{Y,\circ}$ generically have rank three, and the zero pattern then ensures that $M_{Y,\circ}$ has full row rank six. Combining with the lower bound $3 \leq \text{mlt}_u$ gives $\text{mlt}_u = 3$.

5.1. Proof of Theorem 5.3

For fixed vertex colour s , we define $\text{mlt}_e(s)$ to be the smallest n such that the top row of $M_{Y,s}$ is almost surely not in the span of the other β_s rows, and define $\text{mlt}_u(s)$ to be the smallest n such that the matrix $M_{Y,s}$ is almost surely of full row rank $\beta_s + 1$. To prove Theorem 5.3, we use the following lemma.

Lemma 5.8. *Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible, and fix a vertex colour s . For $n \geq \beta_s$ and generic $Y \in \mathbb{R}^{m \times n}$ the row vectors $M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}$ are linearly independent.*

Proof. We think of the mn entries of Y as indeterminates. Let $M'_{Y,s} \in \mathbb{R}^{\beta_s \times \alpha_s n}$ be the matrix with rows $M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}$. We construct an invertible $\beta_s \times \beta_s$ submatrix of $M'_{Y,s}$.

Without loss of generality, let $1, 2, \dots, \alpha_s$ be the vertices of colour s . The matrix $M'_{Y,s}$ has α_s many $\beta_s \times n$ blocks. For each parent relationship colour $p_t, t \in [\beta_s]$ there is some vertex $i = i(t) \in [\alpha_s]$ such that there is an edge of colour p_t pointing towards vertex $i = i(t)$. That is, the i^{th} block of $M'_{Y,s}$ has non-zero entries in the t^{th} row. Let C_t be the t^{th} column of that block, which exists as $n \geq \beta_s$. By construction, the t^{th} entry of C_t is non-zero. We show that the matrix $C = (C_1 \ C_2 \ \dots \ C_{\beta_s})$, is invertible.

An entry of C is either a sum of variables or it is zero. By construction, column C_t only contains (sums of) elements of the t^{th} column of Y . The same variable $Y_{j,t}$ cannot occur in two different entries of C_t , because there is at most one edge from j to vertex $i(t)$. Altogether, the entries of C are (possible empty) sums of variables and each variable occurs in at most one entry of C . The determinant is an alternating sum over products of permutations, and it is enough to show that one product is non-zero. By construction, $C_{11}C_{22} \cdots C_{\beta_s\beta_s} \neq 0$. Thus, $M'_{Y,s}$ has rank β_s for $n \geq \beta_s$. \square

Proposition 5.9. *Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible, with no edges between any vertices of colour s . If $\alpha_s = 1$, then*

$\text{mlt}_e(s) = \beta_s + 1$, while if $\alpha_s \geq 2$ we have

$$\left\lfloor \frac{r_s - 1}{\alpha_s - 1} \right\rfloor + 1 \leq \text{mlt}_e(s) \leq \left\lfloor \frac{\beta_s}{\alpha_s} \right\rfloor + 1.$$

Proof. If $\alpha_s = 1$, the equality $\text{mlt}_e(s) = \beta_s + 1$ is known from the uncoloured case, see [12, Theorem 1]. It remains to consider $\alpha_s \geq 2$. For the upper bound, we show that if $n > \frac{\beta_s}{\alpha_s}$, then the top row of $M_{Y,s}$ is generically not in the span of the other rows. Since there are no edges between two vertices of colour s , the $n\alpha_s$ entries of the top row $M_{Y,s}^{(0)}$ are all independent, from each other and from the entries of the other rows. If $\beta_s < \alpha_s n$, the other β_s rows do not span $\mathbb{R}^{n\alpha_s}$, so a generic choice of top row will not lie in their span.

For the lower bound, consider the $\beta_s \times \alpha_s$ matrix $M'_{Y,s}$ for $n = 1$. Its generic rank is denoted r_s . We consider the $1 \times n$ matrix blow up, where the scalar variables $Y^{(i)}$ are replaced by generic row vectors of length n , to give a $\beta_s \times \alpha_s n$ matrix. We consider the rank of this matrix as n increases. The maximum rank is β_s , which occurs for $n \geq \beta_s$ by Lemma 5.8. Moreover, the rank is a (weakly) concave function in n [8, Corollary 2.8]. Since the rank is positive integer valued, it cannot be the same at two distinct n unless it is at its maximum. Hence the rank for fixed n is at least $\min(r_s + n - 1, \beta_s)$. Therefore, the top row is in the span of the others whenever

$$\min(r_s + n - 1, \beta_s) \geq n\alpha_s, \quad \text{in particular, whenever } \alpha_s n \leq r_s + n - 1 \leq \beta_s,$$

i.e. for $n \leq \min\left(\left\lfloor \frac{r_s - 1}{\alpha_s - 1} \right\rfloor, \beta_s + 1 - r_s\right)$. Hence

$$\text{mlt}_e(s) \geq \min\left(\left\lfloor \frac{r_s - 1}{\alpha_s - 1} \right\rfloor + 1, \beta_s + 2 - r_s\right).$$

We exclude the possibility that the smaller of the two arguments in the minimum is $\beta_s + 2 - r_s$ by appealing to Proposition 5.10. \square

Proposition 5.10. *Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible, with no edges between any vertices of colour s . Then*

$$\left\lfloor \frac{\beta_s}{\alpha_s} \right\rfloor + 1 \leq \text{mlt}_u(s) \leq \beta_s + 2 - r_s.$$

Moreover, if $r_s \neq \beta_s + 1 - (\beta_s/\alpha_s)$ then $\text{mlt}_u(s) \leq \beta_s + 1 - r_s$.

Proof. For the lower bound, we observe that if $\alpha_s n \leq \beta_s$, the $\beta_s + 1$ rows of $M_{Y,s}$ will be linearly dependent. Hence, we need $n > \frac{\beta_s}{\alpha_s}$.

For the upper bound, let $M'_{Y,s}$ and r_s be as above. Recall from the proof of Proposition 5.9 that, for n samples, $\text{rk}(M'_{Y,s}) \geq \min(r_s + n - 1, \beta_s)$ generically. Thus, for $n = \beta_s + 1 - r_s$ the matrix $M'_{Y,s}$ generically has full row rank β_s . It remains to consider the top row of $M_{Y,s}$. We must have $\beta_s \leq \alpha_s n$, otherwise the $\beta_s \times \alpha_s n$ matrix could not have full row rank. We look separately at the possible cases: $\beta_s < n\alpha_s$ and $\beta_s = n\alpha_s$. If $\beta_s < n\alpha_s$, the row vector $M_{Y,s}^{(0)} \in \mathbb{R}^{n\alpha_s}$

is generically not in the span of the β_s rows of $M'_{Y,s}$, because there are no edges between vertices of colour s . Thus, $M_{Y,s}$ generically has full row rank $\beta_s + 1$, and $\text{mlt}_u(s) \leq n = \beta_s + 1 - r_s$. If $\beta_s = n\alpha_s$, equivalently if $r_s = \beta_s + 1 - (\beta_s/\alpha_s)$, an additional sample ensures $\text{rk}(M_{Y,s}) = \beta_s + 1$ generically. \square

Proof of Theorem 5.3. We have $\text{mlt}_e = \max_s \text{mlt}_e(s)$ and $\text{mlt}_u = \max_s \text{mlt}_u(s)$ by Theorem 4.4 parts (b) and (c). Taking the maximum of the lower and upper bounds in Propositions 5.9 and 5.10, over all vertex colours, gives the stated bounds. \square

5.2. A randomised algorithm

Proposition 5.11. *For an RDAG model on (\mathcal{G}, c) , where colouring c is compatible, there is a randomised algorithm for computing the thresholds mlt_e and mlt_u .*

Proof. It suffices to give a randomised algorithm to compute $\text{mlt}_e(s)$ and $\text{mlt}_u(s)$ for fixed vertex colour s . The rank of a symbolic matrix can be computed (efficiently) by a randomised algorithm, see e.g. [24, 35]. Hence, thinking of the entries of $Y \in \mathbb{R}^{m \times n}$ as indeterminates, we can compute for any $n \geq 1$ the rank of the symbolic $(\beta_s + 1) \times \alpha_s n$ matrix $M_{Y,s}$ as well as the rank of the symbolic $\beta_s \times \alpha_s n$ matrix $M'_{Y,s}$. We obtain $\text{mlt}_e(s)$ as the smallest n such that $\text{rk}(M_{Y,s}) > \text{rk}(M'_{Y,s})$ and $\text{mlt}_u(s)$ as the smallest n such that $\text{rk}(M_{Y,s}) = \beta_s + 1$. The algorithm terminates by the upper bound of $\beta_s + 1$ for both $\text{mlt}_e(s)$ and $\text{mlt}_u(s)$. \square

6. Simulations

In the previous section, we gave upper and lower bounds for the maximum likelihood thresholds for RDAG models, see Theorem 5.3. Our bounds quantify how the graph colouring serves to decrease the number of samples needed for existence and uniqueness of the MLE. In this section, we assume that the number of samples is above the maximum likelihood threshold. We explore via simulations the distance of an MLE to the true model parameters. We compare the RDAG model estimate from Algorithm 1 to the usual (uncoloured) DAG model MLE.

The details of our simulations are as follows. We used the NetworkX Python package [21] to build an RDAG model via the following steps. We first build a DAG by generating an undirected graph according to an Erdős–Rényi model that includes each edge with fixed probability, and then directing the edges so that $j \rightarrow i$ implies $j > i$. We assign edge colours randomly, after fixing the total number of possible edge colours. We choose the unique vertex colouring with the largest number of vertex colours that satisfies the compatibility assumption from Definition 2.5. We sample edge weights λ_{st} from a uniform distribution on $[-1, -0.25] \cup [0.25, 1]$ and we sample noise variances ω_{ss} uniformly from $[0, 1]$. Our code is available at <https://github.com/seigal/rdag>.

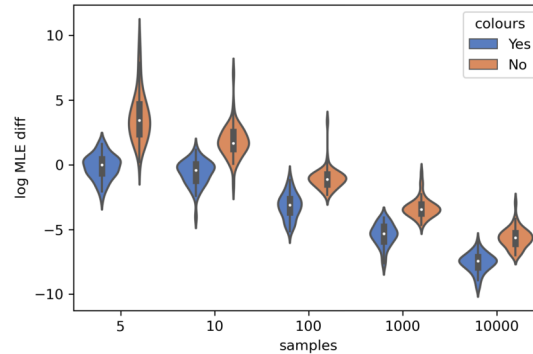


FIG 1. We generated RDAGs on 10 vertices, with each edge present with probability 0.5 and 5 edge colours. We sampled from the distribution $n \in \{5, 10, 100, 1000, 10000\}$ times. For each n we generated 50 random graphs and computed the RDAG MLE and the DAG MLE, comparing them to the true parameter values on a log scale. The results are displayed in a violin plot, with blue for the RDAG MLE and orange for the DAG MLE.

The RDAG MLE is generally closer to the true model parameters than the DAG MLE, see Figure 1. As we would expect, both estimates get closer to the true parameters as the number of samples from the distribution increases. At a high number of samples, the difference between the RDAG MLE and the DAG MLE is smaller than at a low number of samples.

Next we examined how the RDAG MLE was affected by the number of edge colours, see Figure 2. The RDAG MLE is closest to the true parameters when the number of edge colours is small; i.e., when there are fewer parameters to learn. As the number of edge colours increases, the difference between the RDAG MLE and the DAG MLE decreases. Note that the DAG model is the setting where each vertex and edge has its own colour.

Finally, we looked at how the RDAG MLE and DAG MLE are affected by the edge density of the graph, see Figure 3. The RDAG MLEs get closer to the true parameter values as the edge density increases: more edges have the same weight, so more samples contribute to estimating each edge weight. By comparison, the DAG MLEs get further from the true parameters as the edge density increases, because there are more parameters to learn.

7. Conclusion

In this paper, we introduce RDAG models, characterize their maximum likelihood estimation via linear independence conditions and obtain their MLEs via least squares estimators. We bound the maximum likelihood thresholds for RDAG models and relate RDAG models to their undirected analogues. We implement an algorithm to compute the MLE and use it to explore the distance between the MLE and true parameters in simulations.

We hope that our models and results will be useful for settings in which parameters may be shared between different parts of a graphical model, to

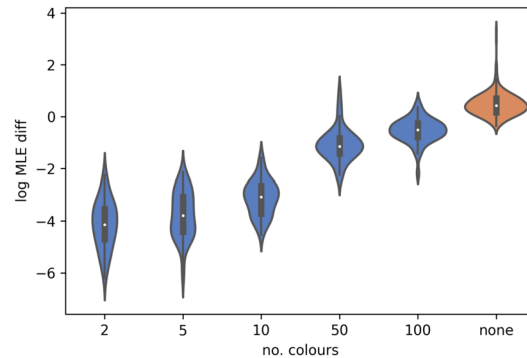


FIG 2. We generated RDAGs on 20 vertices, each edge present with probability 0.5 and number of edge colours in $\{2, 5, 10, 50, 100\}$. We sampled from the distribution 100 times and compared the MLE to the true parameter values on a log scale. The DAG MLE is shown in orange for comparison.

reduce maximum likelihood thresholds and/or to capture known symmetries of a set-up. In particular, our results may apply to situations in which only few samples are available. Furthermore, our results may be used for maximum likelihood estimation in RCON models whenever the conditions of Theorem 3.4 hold.

Future work One limitation of our approach is that we assume the coloured graph is known as input. That is, we study maximum likelihood estimation and do not consider the topic of model selection for RDAGs. This would be an interesting direction for future study, building on related work for uncoloured graphs, such as [33].

Moreover, we require assumptions on the input graph: the absence of bidirected edges (i.e., the independence of errors), acyclicity, and the compatibility of the colouring. There are likely to be situations where these assumptions do not give a well-fitting graph. Therefore, it would be beneficial to be able to estimate parameters in general linear structural equation models (see [14]) with symmetries given by colourings of vertices and (bi)directed edges. It seems natural to try to generalize existing results regarding identifiability of the model parameters [13], as well as results on MLE computation, maximum likelihood thresholds etc. One would hope that symmetries might again give rise to properties such as smaller maximum likelihood thresholds.

Another direction we leave open for future work is to find *exact* maximum likelihood thresholds for RDAG models (Problem 5.4).

Appendix A: Connections to stability

We give a characterization of maximum likelihood estimation for RDAG models, via stability under a group action. We generalise the definitions of stability to

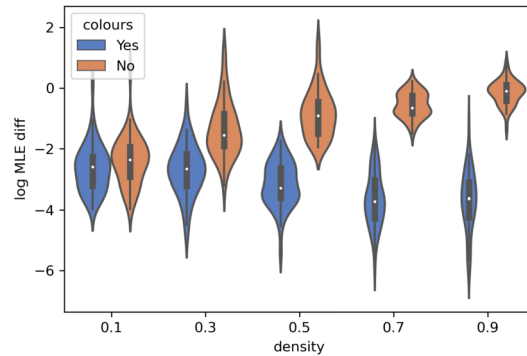


FIG 3. We generated RDAGs on 10 vertices, each edge present with probability in $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ and 5 edge colours. For each edge probability we generated 50 random graphs, sampled from each one 100 times, and compared the RDAG and DAG MLEs. As above, blue is the RDAG MLE and orange is the DAG MLE.

a set rather than a group. This offers an alternative to Corollary 4.8 for the set of MLEs.

Fix a set of invertible $m \times m$ matrices A , with entries in \mathbb{R} . Consider some $Y \in \mathbb{R}^{m \times n}$. By analogy to an orbit and stabiliser under a group action, we define the orbit and stabiliser under a set A to be, respectively,

$$A \cdot Y := \{aY \mid a \in A\}, \quad A_Y := \{a \in A \mid aY = Y\}.$$

This allows us to define stability notions analogous to the group situation.

Definition A.1. We say that the matrix $Y \in \mathbb{R}^{m \times n}$, under the set A , is

- (i) unstable if there exist $a_n \in A$ with $a_n Y \rightarrow 0$ as $n \rightarrow \infty$, i.e. 0 is contained in the Euclidean closure $\overline{A \cdot Y}$
- (ii) semistable if Y is not unstable, i.e. $0 \notin \overline{A \cdot Y}$
- (iii) polystable if $Y \neq 0$ and the set $A \cdot Y$ is Euclidean closed
- (iv) stable if Y is polystable and A_Y is finite.

The above notions of stability are usually studied for A a reductive group [11, 29]. They were studied for reductive and non-reductive groups in [2]. We are not aware of these definitions being used before without any group structure on A .

We relate maximum likelihood estimation for an RDAG model to these stability notions. For coloured graph (\mathcal{G}, c) , we recall the definition of $A(\mathcal{G}, c)$ from (7). Given a set $A \subseteq \text{GL}_m$, we define $A_{\text{SL}} = \{a \in A \mid \det(a) = 1\}$ and $A_{\text{SL}}^{\pm} = \{a \in A \mid \det(a) = \pm 1\}$.

Theorem A.2. Consider an RDAG model on (\mathcal{G}, c) with compatible colouring c and sample matrix $Y \in \mathbb{R}^{m \times n}$. Then stability under $A(\mathcal{G}, c)_{\text{SL}}$ relates to

maximum likelihood estimation:

- (a) Y unstable $\Leftrightarrow \ell_Y$ unbounded from above
- (b) Y semistable $\Leftrightarrow \ell_Y$ bounded from above
- (c) Y polystable \Leftrightarrow MLE exists
- (d) Y stable \Leftrightarrow MLE exists uniquely.

Proposition A.3. Fix the RDAG model on (\mathcal{G}, c) and set $A := A(\mathcal{G}, c)_{\text{SL}}$. If $\lambda a^\top a$ is an MLE given Y , where $a \in A$ and $\lambda > 0$ is a scalar, then the set of MLEs given Y are in bijection with A_Y under mapping $b \in A_Y$ to $\lambda(a + b - \text{id})^\top(a + b - \text{id})$.

Theorem A.2 applies to any DAG model, see Remark 4.11. Therefore, Theorem A.2 generalises [2, Theorem 5.3] in multiple ways. First, it extends from transitive DAGs³ to all DAGs. Second, it extends from uncoloured DAG models to RDAG models. Third, it adds part (d) about stable samples. Moreover, Proposition A.3 gives a bijection between the MLEs and the stabilising set.

To prove Theorem A.2 and Proposition A.3, we first we generalise [2, Proposition 3.4 and Theorem 3.6] to no longer require that A is a group. We say that a set A is closed under non-zero scalar multiples if $a \in A$ implies $ta \in A$ for all $t \in \mathbb{R}^\times$.

Proposition A.4. Let A be a set of real invertible $m \times m$ matrices, closed under non-zero scalar multiples. There is a correspondence between stability under A_{SL}^\pm and maximum likelihood estimation in the model \mathcal{M}_A given sample matrix $Y \in \mathbb{R}^{m \times n}$:

- (a) Y unstable \Leftrightarrow likelihood ℓ_Y unbounded from above
- (b) Y semistable \Leftrightarrow likelihood ℓ_Y bounded from above
- (c) Y polystable \Rightarrow MLE exists.

The MLEs, if they exist, are the matrices $\lambda a^\top a$, where $\|a \cdot Y\| > 0$ is minimal in $A_{\text{SL}}^\pm \cdot Y$ and $\lambda \in \mathbb{R}_{>0}$ is the unique global minimum of

$$\mathbb{R}_{>0} \rightarrow \mathbb{R}, \quad x \mapsto \frac{x}{n} \|a \cdot Y\|^2 - m \log(x).$$

If for all $a \in A$ there exists an orthogonal matrix $o = o(a)$ such that $o^\top a \in A$ and $\det(o^\top a) > 0$, then A_{SL}^\pm can be replaced by A_{SL} .

Proof. This is proved using the same argument as [2, Proposition 3.4 and Theorem 3.6]. Maximising ℓ_Y over \mathcal{M}_A is equivalent to minimising

$$f: A \rightarrow \mathbb{R}, \quad a \mapsto \frac{1}{n} \|a \cdot Y\|^2 - \log \det(a^\top a).$$

We write $a \in A$ as τb , where $\tau \in \mathbb{R}_{>0}$ and $b \in A_{\text{SL}}^\pm$. Setting $x := \tau^2$ we compute

$$f(a) = \frac{\tau^2}{n} \|b \cdot Y\|^2 - \log \det(\tau^2 b^\top b) = \frac{x}{n} \|b \cdot Y\|^2 - m \log(x).$$

³A DAG is transitive if whenever $k \rightarrow j$ and $j \rightarrow i$ are in \mathcal{G} then so is $k \rightarrow i$. It is exactly the condition to make $A(\mathcal{G})$ a group, see [2, Proposition 5.1].

The infimum of the function $x \mapsto xC - m \log(x)$ increases as $C \geq 0$ increases, hence

$$\sup_{a \in A} \ell_Y(a^\top a) = - \inf_{a \in A} f(a) = - \inf_{x \in \mathbb{R}_{>0}} \left(\frac{x}{n} \left(\inf_{b \in A_{\text{SL}}^\pm} \|b \cdot Y\|^2 \right) - m \log(x) \right). \tag{19}$$

We have $\inf_{a \in A} f(a) = -\infty$ if and only if $\inf_{b \in A_{\text{SL}}^\pm} \|b \cdot Y\|^2 = 0$, i.e. if and only if Y is unstable. This shows parts (a) and (b).

To prove (c) assume that Y is polystable under A_{SL}^\pm . Then $C := \inf_{b \in A_{\text{SL}}^\pm} \|b \cdot Y\|^2$ is strictly positive, as Y is semistable. Since $A_{\text{SL}}^\pm \cdot Y$ is closed in $\mathbb{R}^{m \times n}$, we see that C is attained in the compact set $(A_{\text{SL}}^\pm \cdot Y) \cap \{Z \in \mathbb{R}^{m \times n} \mid \|Z\|^2 \leq C+1\}$. Thus, an MLE given Y exists. If an MLE exists, then the inner and outer infima in (19) are attained, and any MLE has the form in the statement.

Finally, assume that for each $a \in A$ there exists an orthogonal matrix $o = o(a)$ with $o^\top a \in A$ and $\det(o^\top a) > 0$. Then we can write $a = \tau ob$, where $\tau := \sqrt[m]{\det(o^\top a)}$ and $b := \tau^{-1} o^\top a$. By construction, $\det(b) = 1$ and as A is closed under non-zero scalar multiples we have $b \in A_{\text{SL}}$. Setting $x := \tau^2$, we derive the same computation for $f(a)$ (now with $b \in A_{\text{SL}}$), since o is orthogonal. The rest of the proof then works with A_{SL} instead of A_{SL}^\pm . \square

Remark A.5. Given an RDAG model $\mathcal{M}_{A(\mathcal{G},c)}$ with compatible colouring, we can always apply Proposition A.4 using stability under $A(\mathcal{G},c)_{\text{SL}}$ instead of the bigger set $A(\mathcal{G},c)_{\text{SL}}^\pm$, as follows. If α_s – the number of vertices of colour s – is even for all $s \in c(I)$, then $A(\mathcal{G},c)$ only contains matrices of positive determinant, so $A(\mathcal{G},c)_{\text{SL}}^\pm = A(\mathcal{G},c)_{\text{SL}}$. If α_s is odd for some vertex colour s , then $A(\mathcal{G},c)$ contains a diagonal orthogonal matrix \hat{o} with determinant -1 . By compatibility, $A(\mathcal{G},c)$ is invariant under left-multiplication with diagonal matrices from $A(\mathcal{G},c)$. Hence, for $a \in A$ we can choose $o(a) = \hat{o}$ if $\det(a) < 0$ and $o(a) = \text{id}$ otherwise, to satisfy the required assumptions in Proposition A.4.

Next, we return to the linear independence condition in Theorem 4.4(b).

Lemma A.6. *Consider the RDAG model on (\mathcal{G},c) where colouring c is compatible, and set $A := A(\mathcal{G},c)_{\text{SL}}$. Assume for a non-zero $Y \in \mathbb{R}^{m \times n}$ that $M_{Y,s}^{(0)} \notin \text{span}\{M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}\}$ for all $s \in c(I)$. Then Y is polystable under A and $A \cdot Y$ is Zariski closed.*

Proof. The hypotheses in the statement imply that the log-likelihood ℓ_Y is bounded from above, by Theorem 4.4(b). Since $A(\mathcal{G},c)$ is closed under non-zero scalar multiples, Y is semistable under $A = A(\mathcal{G},c)_{\text{SL}}$ by Proposition A.4(b) and Remark A.5.

We now study the orbit $A \cdot Y$. Let T be the set of diagonal matrices in A and U the unipotent matrices in A . Then $A = T \cdot U$ by compatibility and, in fact, any $a \in A$ admits a unique decomposition $a = tu$ with $t \in T$, $u \in U$. For $s \in c(I)$, recall the construction of $M_{Y,s} \in \mathbb{R}^{(\beta_s+1) \times \alpha_s n}$ from Definition 4.1. Setting $V_s := \mathbb{R}^{1 \times \alpha_s n}$ we can identify $\mathbb{R}^{m \times n} \cong \bigoplus_s V_s$ such that the rows of vertex colour s belong to V_s . By definition of $M_{Y,s}$, and since the $\text{pre}(s)$ partition

$c(E)$, the set $U \cdot Y$ is $H := \prod_s H_s$ with

$$H_s = \left\{ M_{Y,s}^{(0)} + a_{s,1}M_{Y,s}^{(1)} + \dots + a_{s,\beta_s}M_{Y,s}^{(\beta_s)} \mid a_{s,t} \in \mathbb{R} \right\}.$$

The affine space H_s is $M_{Y,s}^{(0)} + X_s$, where $X_s := \text{span}\{M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}\}$. Since $M_{Y,s}^{(0)} \notin X_s$ for all $s \in c(I)$ by assumption, we have $0 \notin H_s$ and hence H_s has at least codimension one in V_s . We define the linear subspace $K_s := (\mathbb{R}M_{Y,s}^{(0)}) \oplus X_s$ of V_s . Since T acts on each V_s with the non-zero scalar of colour s , we have

$$A \cdot Y = T \cdot (U \cdot Y) = T \cdot H = T \cdot \prod_s H_s \subseteq \bigoplus_s K_s \subseteq \bigoplus_s V_s.$$

It suffices to show that $A \cdot Y$ is Zariski-closed in $\bigoplus_s K_s$. Each H_s is an affine subspace of K_s with codimension one. Therefore, there exists a linear form $p_s \in K_s^*$ such that $H_s = \mathbb{V}_{K_s}(p_s - 1)$, where $\mathbb{V}(\cdot)$ denotes the vanishing locus.

We finish the proof by showing that $A \cdot Y = \mathbb{V}(\prod_s p_s^{\alpha_s} - 1)$ in $\bigoplus_s K_s$. First, given $W = (W_s)_s \in A \cdot Y = T \cdot H$ we can write $W = a \cdot Z$ with $a \in T$ and $Z = (Z_s)_s \in H$. Then

$$\left(\prod_s p_s^{\alpha_s} \right)(W) = \prod_s p_s(W_s)^{\alpha_s} = \prod_s (a_s p_s(Z_s))^{\alpha_s} = \prod_s (a_s)^{\alpha_s} = 1$$

by the choice of $p_s \in K_s^*$ and since $\det(a) = \prod_s a_s^{\alpha_s} = 1$. On the other hand, suppose $W = (W_s)_s \in \mathbb{V}(\prod_s p_s^{\alpha_s} - 1) \subseteq \bigoplus_s K_s$. Set $a_s := p_s(W_s)$, then we have $\prod_s a_s^{\alpha_s} = 1$, so the a_s define some $a \in T$. Moreover, $W' := (a_s^{-1}W_s)_s \in H$ by definition of the a_s and hence $W = a \cdot W'$ is contained in $T \cdot H = A \cdot Y$. \square

Proposition A.7. *Consider an RDAG model on (\mathcal{G}, c) with compatible colouring c and $A := A(\mathcal{G}, c)_{\text{SL}}$. Let $Y \in \mathbb{R}^{m \times n}$ be a sample matrix. Stability under A relates to linear independence conditions on the matrices $M_{Y,s}$:*

- (a) Y unstable $\Leftrightarrow M_{Y,s}^{(0)} \in \text{span}\{M_{Y,s}^{(i)} : i \in [\beta_s]\}$ for some $s \in c(I)$
- (b) Y semistable $\Leftrightarrow Y$ polystable
- (c) Y polystable $\Leftrightarrow M_{Y,s}^{(0)} \notin \text{span}\{M_{Y,s}^{(i)} : i \in [\beta_s]\}$ for all $s \in c(I)$
- (d) Y stable $\Leftrightarrow M_{Y,s}$ has full row rank for all $s \in c(I)$

Proof. Proposition A.4 in combination with Theorem 4.4 yields part (a) and the forwards direction of (c), while Lemma A.6 gives the backwards direction of (c). We obtain part (b) as a direct consequence of (a) and (c).

For part (d), it suffices to see that a polystable Y has a trivial stabiliser A_Y if and only if for all $s \in c(I)$ the rows $M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}$ are linearly independent. So let Y be polystable. By construction of $M_{Y,s}$, a matrix $a \in A$ satisfies $aY = Y$ if and only if

$$a_s M_{Y,s}^{(0)} + \sum_{t \in [\beta_s]} a_{s,t} M_{Y,s}^{(t)} = M_{Y,s}^{(0)}, \tag{20}$$

for all $s \in c(I)$, where $a_s \in \mathbb{R}^\times$ is the entry of a for vertex colour s , and $a_{s,t} \in \mathbb{R}$ is the entry of a for the parent relationship colour encoded by (s, t) , where $t \in [\beta_s]$. We note that (20) implies $a_s = 1$ and $\sum_{t \in [\beta_s]} a_{s,t} M_{Y,s}^{(t)} = 0$, by polystability of Y and part (c).

If $M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}$ are linearly independent, then (20) has exactly one solution, namely $a_s = 1$ and $a_{s,t} = 0$ for all $t \in [\beta_s]$. Thus, if $M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}$ are linearly independent for all $s \in c(I)$, then $A_Y = \{\text{id}\}$. On the other hand, if for some $s \in c(I)$ the rows $M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}$ are linearly dependent, then $\sum_{t \in [\beta_s]} a_{s,t} M_{Y,s}^{(t)} = 0$ has infinitely many solutions. Distinct solutions give distinct unipotent matrices $a \in A$ by using $a_{s,t}$ as the entry for edge colour $t \in \text{prc}(s)$, and setting all other off-diagonal entries of a to zero. Such a unipotent matrix $a \in A$ satisfies $aY = Y$, since the sets $\text{prc}(s)$ are disjoint, so the $a_{s,t}$ do not affect any rows of Y with a different vertex colour. In conclusion, A_Y is infinite if $M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}$ are linearly dependent for some $s \in c(I)$. \square

Proof of Theorem A.2. Combine Proposition A.7 with Theorem 4.4. \square

We now turn to Proposition A.3. As above, $A := A(\mathcal{G}, c)_{\text{SL}}$ for an RDAG with compatible colouring. Denote the set of diagonal (respectively unipotent) matrices in A by T (respectively U). By compatibility of the colouring, $A = T \cdot U$ and in fact any $a \in A$ admits a unique factorisation $a = tu$ with $t \in T$, $u \in U$.

Lemma A.8. *Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible. For $A := A(\mathcal{G}, c)_{\text{SL}}$ write $A = T \cdot U$ as above. If $Y \in \mathbb{R}^{m \times n}$ is polystable under A , then the following hold:*

- (a) $U \cdot Y$ contains a unique element \tilde{Y} of minimal norm.
- (b) For $t \in T$ and $u \in U$, $\|tu \cdot Y\| \geq \|t \cdot \tilde{Y}\|$ with equality if and only if $u \cdot Y = \tilde{Y}$.
- (c) Let $a, \tilde{a} \in A$ be such that $a \cdot Y$ and $\tilde{a} \cdot Y$ are of minimal norm in $A \cdot Y$. Then there is an orthogonal $t \in T$ such that $ta \cdot Y = \tilde{a} \cdot Y$.

Proof. For part (a), we recall that the $\text{prc}(s)$, $s \in c(I)$, partition the edge colours $c(E)$. Therefore, when minimising

$$\|uY\|^2 = \sum_{s \in c(I)} \left\| M_{uY,s}^{(0)} \right\|^2 = \sum_{s \in c(I)} \left\| M_{Y,s}^{(0)} + \sum_{t \in [\beta_s]} u_{s,t} M_{Y,s}^{(t)} \right\|^2$$

over $u \in U$ we can minimise each summand separately. For each $s \in c(I)$, the affine space $M_{Y,s}^{(0)} + \text{span}\{M_{Y,s}^{(t)} : t \in [\beta_s]\}$ has a unique element of minimal norm, call it M_s . Hence, $U \cdot Y$ has a unique element of minimal norm, \tilde{Y} , determined by $M_{\tilde{Y},s}^{(0)} = M_s$ for all $s \in c(I)$. (Note that there may be several $u \in U$ with $uY = \tilde{Y}$.)

To prove part (b), we use (the proof of) part (a) to obtain

$$\|M_{tuY,s}^{(0)}\|^2 = \|t_s M_{uY,s}^{(0)}\|^2 = |t_s|^2 \|M_{uY,s}^{(0)}\|^2 \geq |t_s|^2 \|M_{\tilde{Y},s}^{(0)}\|^2 = \|M_{t\tilde{Y},s}^{(0)}\|^2 \quad (21)$$

for all $s \in c(I)$, hence $\|tu \cdot Y\| \geq \|t\tilde{Y}\|$. The latter inequality is strict if and only if there is strict inequality in (21) for at least one s . By uniqueness of \tilde{Y} , this is the case if and only if $uY \neq \tilde{Y}$.

For (c), write $a = tu$ with $t \in T$ and $u \in U$. Since aY is of minimal norm in $A \cdot Y$, we deduce $uY = \tilde{Y}$ using (b). Thus, $aY \in T \cdot \tilde{Y}$ and similarly $\tilde{a}Y \in T \cdot \tilde{Y}$. As $T \cdot \tilde{Y} \subseteq A \cdot Y$ the matrices aY and $\tilde{a}Y$ are also of minimal norm in $T \cdot \tilde{Y}$. We note that T is a group isomorphic to the reductive group $\{(t_s)_{s \in c(I)} \mid t_s \in \mathbb{R}^\times, \prod_s t_s^{\alpha_s} = 1\}$. Hence the Kempf-Ness theorem, see [2, Theorem 2.2], for the action of T implies that there is some orthogonal $t \in T$ that relates the minimal norm elements aY and $\tilde{a}Y$ in $T \cdot \tilde{Y}$. \square

We conclude this appendix with a proof of Proposition A.3.

Proof of Proposition A.3. Recall that $A = A(\mathcal{G}, c)_{\text{SL}} = T \cdot U$, where T is the diagonal matrices in A , and U the unipotent matrices in A . If $aY = Y$, then for all $s \in c(I)$,

$$M_{Y,s}^0 = a_s M_{Y,s}^{(0)} + \sum_{t \in [\beta_s]} a_{s,t} M_{Y,s}^{(t)}.$$

We have $M_{Y,s}^{(0)} \notin \text{span}\{M_{Y,s}^{(t)} : t \in [\beta_s]\}$, since Y is polystable. Hence $a_s = 1$ for all s , i.e. $a \in U$ and therefore $A_Y = U_Y$. We set $N_Y := U_Y - \text{id}$, which consists of strictly upper triangular matrices. It suffices to show that for fixed MLE $\lambda a^\top a$ the map

$$\begin{aligned} \varphi : N_Y &\rightarrow \{\text{MLEs given } Y\} \\ b &\mapsto \lambda(a+b)^\top(a+b) \end{aligned}$$

is well-defined and bijective. Note that $bY = 0$ for any $b \in N_Y$. Therefore, $(a+b)Y = aY$ is of minimal norm in $A \cdot Y$ and thus $\varphi(b)$ is an MLE by Proposition A.4.

For surjectivity, let $\lambda \tilde{a}^\top \tilde{a}$ be another MLE given Y . Then aY and $\tilde{a}Y$ are of minimal norm in $A \cdot Y$, hence there is an orthogonal $t \in T$ with $aY = t\tilde{a}Y$ by Lemma A.8(c). We set $b := t\tilde{a} - a$ so that $b \cdot Y = 0$ and $(\text{id} + b)Y = Y$. By compatibility of the colouring we have $t\tilde{a} \in A$ and thus all entries of $b = t\tilde{a} - a$ obey the colouring c . However, $bY = 0$ implies $b_s = 0$ for all s by polystability of Y , hence $b \in N_Y$. We compute $\varphi(b) = \lambda(t\tilde{a})^\top(t\tilde{a}) = \lambda \tilde{a}^\top \tilde{a}$ using orthogonality of t .

For injectivity, let $b, b' \in N_Y$ be such that $\varphi(b) = \varphi(b')$. Let $t \in T$ be defined by $t_s = 1$ if $a_s > 0$ and $t_s = -1$ if $a_s < 0$. Then t is orthogonal and thus

$$(ta + tb)^\top(ta + tb) = (a + b)^\top(a + b).$$

Similarly, $(ta + tb')^\top(ta + tb') = (a + b')^\top(a + b')$. Then $\varphi(b) = \varphi(b')$ implies

$$(ta + tb)^\top(ta + tb) = (ta + tb')^\top(ta + tb'). \tag{22}$$

Moreover, tb, tb' are strictly upper triangular and $ta \in A$ has positive diagonal entries, by construction of t . Hence, applying uniqueness of the Cholesky decomposition to (22) gives $ta + tb = ta + tb'$, and we deduce $b = b'$. \square

Appendix B: Connections to Gaussian group models

A model is a Gaussian group model if it is equal to \mathcal{M}_A , see (1), where the set A a group. In this case, the second term in the log-likelihood (3) is the minimisation of the norm over a group orbit. This perspective was used in [2] to relate existence of the MLE to notions of stability under a group action. In this appendix, we characterise when the set of matrices $A(\mathcal{G}, c)$ from (7) is a group. We use Popov's criterion to study stability, and give our third and final description of the set of MLEs in an RDAG model.

B.1. The butterfly criterion

Suppose (\mathcal{G}, c) is a coloured DAG with compatible c . We give necessary and sufficient conditions for $A(\mathcal{G}, c)$ to be a group. We call a subset $A \subseteq \text{GL}_m$ linear if $A = L \cap \text{GL}_m$, where L is a linear subspace of $m \times m$ matrices.

Lemma B.1. *Let $A \subseteq \text{GL}_m$ be a linear subspace of matrices containing the identity matrix. Then A is a group if and only if it is closed under multiplication.*

Proof. A group is closed under multiplication. Conversely, if A is closed under multiplication, to be a group it must also be closed under inverses. For a matrix $a \in A$, its characteristic polynomial is $f_a(t) = t^m + c_1 t^{m-1} + \dots + c_m$. We know $c_m \neq 0$ because c_m is (up to sign) the determinant of a . This means $a \cdot \frac{1}{-c_m}(a^{m-1} + c_1 a^{m-2} + \dots + c_{m-1} \text{id}) = \text{id}$. Suppose L is the linear subspace of matrices such that $A = L \cap \text{GL}_m$. Then $a^{-1} = \frac{1}{-c_m}(a^{m-1} + c_1 a^{m-2} + \dots + c_{m-1} \text{id}) \in L \cap \text{GL}_m = A$, since $\text{id} \in A$. \square

For a pair of vertices $i, j \in I$, define $b(ij) = \{k \mid i \leftarrow k, k \leftarrow j \in \mathcal{G}\}$. Let $\mathcal{G}_{b(ij)}$ denote the coloured subgraph on $\{i\} \cup \{j\} \cup b(ij)$, with edges $i \leftarrow k, k \leftarrow j$ for each $k \in b(ij)$, and colours inherited from c . We call $\mathcal{G}_{b(ij)}$ a *butterfly graph*.

Proposition B.2 (The butterfly criterion). *Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible. The set $A(\mathcal{G}, c)$ is a group if and only if*

- (a) \mathcal{G} is transitive; and
- (b) if $c(ij) = c(kl)$ for edges $j \rightarrow i, l \rightarrow k$ in \mathcal{G} , then $\mathcal{G}_{b(ij)} \simeq \mathcal{G}_{b(kl)}$.

Proof. Observe that $A(\mathcal{G}, c)$ is a group if and only if it is closed under multiplication, since it is a linear subspace of GL_m that contains the identity matrix. Hence, by Lemma B.1, we need to characterize when $A(\mathcal{G}, c)$ is closed under multiplication. We have $gh \in A(\mathcal{G}, c)$ for $g, h \in A(\mathcal{G}, c)$ if and only if

1. $(gh)_{ii} = (gh)_{jj}$ whenever $c(i) = c(j)$;
2. $(gh)_{ij} = (gh)_{kl}$ whenever $j \rightarrow i, l \rightarrow k$ in \mathcal{G} have $c(ij) = c(kl)$; and
3. $(gh)_{ij} = 0$ whenever $j \not\rightarrow i$ in \mathcal{G} .

For (1), observe that $(gh)_{ii} = g_{ii}h_{ii}$. Thus, if $c(i) = c(j)$ then $(gh)_{ii} = (gh)_{jj}$. For (2), take $j \rightarrow i, l \rightarrow k$ in \mathcal{G} with $c(ij) = c(kl)$. Then

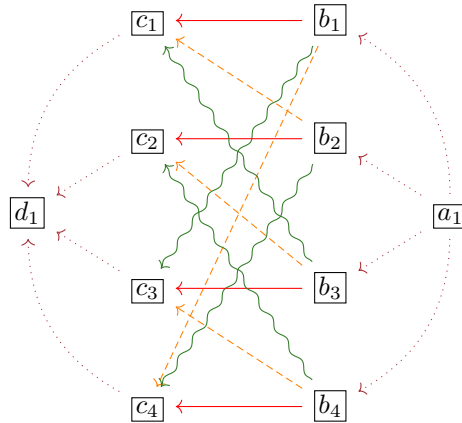
$$(gh)_{ij} = g_{ii}h_{ij} + g_{ij}h_{jj} + \sum_{p \in b(ij)} g_{ip}h_{pj}$$

$$\text{and } (gh)_{kl} = g_{kk}h_{kl} + g_{kl}h_{ll} + \sum_{q \in b(kl)} g_{kq}h_{ql},$$

hence $(gh)_{ij} = (gh)_{kl}$ if $\mathcal{G}_{b(ij)} \simeq \mathcal{G}_{b(kl)}$. Conversely, assume $(gh)_{ij} = (gh)_{kl}$ as a polynomial identity in the unknown entries of matrices g and h . By compatibility, $c(i) = c(k)$, so $g_{ii}h_{ij} = g_{kk}h_{kl}$. Vertex and edge colours are disjoint and the sums over $b(ij)$ and $b(kl)$ only involve edge colours. Thus, $(gh)_{ij} = (gh)_{kl}$ implies $g_{ij}h_{jj} = g_{kl}h_{ll}$, so $h_{jj} = h_{ll}$, and the sum over $b(ij)$ must equal the sum over $b(kl)$. This means $c(j) = c(l)$, and the two collections $(c(ip), c(pj)), p \in b(ij)$ and $(c(kq), c(ql)), q \in b(kl)$ of ordered pairs counted with multiplicity agree. Compatibility ensures the correct colours on the vertices in $b(ij)$ and $b(kl)$ as well, hence $\mathcal{G}_{b(ij)} \simeq \mathcal{G}_{b(kl)}$.

For (3), observe that if $j \not\rightarrow i$ in \mathcal{G} then $g_{ij} = 0 = h_{ij}$ and therefore $(gh)_{ij} = \sum_{k \in b(ij)} g_{ik}h_{kj}$. The latter is zero for all $g, h \in A(\mathcal{G}, c)$ if and only if $b(ij) = \emptyset$. Thus, condition (3) is equivalent to the following: if $j \not\rightarrow i$ in \mathcal{G} , then there does not exist $k \in I$ with $j \rightarrow k$ and $k \rightarrow i$ in \mathcal{G} , i.e. \mathcal{G} must be transitive, by contraposition. We have shown that (1), (2) and (3) are satisfied if and only if conditions (a) and (b) hold. \square

Example B.3. Surprisingly, two graphs can have all the same butterfly graphs without being isomorphic. We present an example. Consider the coloured graph with 10 black (square) vertices, and edges that are red (solid), green (squiggly), orange (dashed) or brown (dotted).



We add some further edges: four purple edges $a_1 \rightarrow c_i$, four blue edges $b_i \rightarrow d_1$, and a yellow edge $a_1 \rightarrow d_1$. Now consider the graph obtained by exchanging the green (squiggly) and orange (dashed) edges.

The butterfly graphs for the two graphs are the same, as follows. On the yellow edge, the butterfly graphs both have four paths consisting of a brown edge followed by a blue edge, and four that are a purple edge followed by a brown edge. Similarly, we can check the butterfly graphs at the other edge colours.

However, the two coloured graphs are not isomorphic. Indeed, the only way to get an isomorphism is to permute the b-layer and the c-layer. The red (solid) edges give the identity permutation, the orange (dashed) edges give the cycle $\sigma = (1\ 4\ 3\ 2)$, and the green (squiggly) edges give σ^2 . Hence an isomorphism would need to consist of permutations μ_1 and μ_2 of $\{1, 2, 3, 4\}$ with $\mu_1 \text{id} \mu_2 = \text{id}$, $\mu_1 \sigma \mu_2 = \sigma^2$, $\mu_1 \sigma^2 \mu_2 = \sigma$. The first condition implies $\mu_2 = \mu_1^{-1}$, hence σ and σ^2 need to be simultaneously conjugate to σ^2 and σ respectively. This implies $(\sigma^2)^2 = \sigma$, a contradiction because $\sigma^4 = \text{id}$.

B.2. Popov's criterion

If $A(\mathcal{G}, c)$ is a group, we can prove the important Lemma A.6 differently, via a criterion of Popov [32, Theorem 4]. The criterion characterises when an orbit under a connected solvable group is closed, provided the underlying field is algebraically closed. Due to the latter assumption, we work in this subsection with RDAG models defined over the complex numbers, and note that many of our results and proofs carry over to the complex case, see Remark 2.11. We start by describing Popov's criterion for the group $G := A(\mathcal{G}, c)_{\text{SL}}$.

Since $G \subseteq \text{GL}_m(\mathbb{C})$ is a group of invertible upper triangular matrices, it is solvable. We decompose $G = T \cdot U \subseteq \text{GL}_m$ as a semi-direct product, where T is the subgroup of diagonal matrices in G , and U is the subgroup of unipotent matrices in G . The group G acts on $\mathbb{C}^{m \times n}$ by left-multiplication. Let $f_{k,l} \in \mathbb{C}[\mathbb{C}^{m \times n}]$, $k \in [m]$, $l \in [n]$, be the coordinate functions on $\mathbb{C}^{m \times n}$. Let x_s be the coordinate function corresponding to vertex colour $s \in c(I)$, and let $x_{s,t}$ be the coordinate function for the edge colour $t \in \text{pre}(s)$. Given a tuple of samples $Y \in \mathbb{C}^{m \times n}$ we consider the orbit map

$$\mu_{G \cdot Y}: \begin{array}{l} G \rightarrow \mathbb{C}^{m \times n} \\ g \mapsto g \cdot Y \end{array} \quad \text{or, on coordinate rings,} \quad \mu_{G \cdot Y}^*(f_{k,l}) = \sum_{j=1}^m x_{c(kj)} Y_{j,l}.$$

We define

$$R_Y := \mu_{G \cdot Y}^*(\mathbb{C}[\mathbb{C}^{m \times n}]) = \mathbb{C} \left[\sum_{j=1}^m Y_{j,l} x_{c(kj)} \mid k \in [m], l \in [n] \right] \subseteq \mathbb{C}[G].$$

Since R_Y is a \mathbb{C} -algebra, we obtain the semigroup

$$\mathfrak{X}_{G \cdot Y} := \left\{ (d_s)_{s \in c(I)} \in \mathfrak{X}(T) \mid \prod_{s \in c(I)} x_s^{d_s} \in R_Y \right\},$$

where $\mathfrak{X}(T) \cong \mathbb{Z}^{|c(I)|} / (\mathbb{Z} \cdot (\alpha_s)_{s \in c(I)})$ is the character group of T .

Theorem B.4 (Popov's Criterion, [32, Theorem 4]). *Let G and Y be as above. The orbit $G \cdot Y$ is Zariski closed if and only if $\mathfrak{X}_{G \cdot Y}$ is a group.*

Remark B.5. The group $G = A(\mathcal{G}, c)_{\text{SL}}$ may not be connected as required in [32, Theorem 4]. However, the orbit $G \cdot Y$ is Zariski-closed if and only if $G^\circ \cdot Y$ is

Zariski-closed, where G° is the identity component of G . Thus, after restricting to $G^\circ = T^\circ U$ we may assume that G is connected. Restricting to T° amounts to restricting to the torsion-free part of $\mathfrak{X}(T)$: if α is the greatest common divisor of all $\alpha_s, s \in c(I)$, then $T^\circ \cong \{(g_s)_{s \in c(I)} \mid \prod_s g_s^{\alpha_s/\alpha} = 1\}$ and $\mathfrak{X}(T^\circ) = \mathbb{Z}^{|c(I)|} / (\mathbb{Z} \cdot (\alpha_s/\alpha)_{s \in c(I)})$.

Second Proof of Lemma A.6. The matrix Y is semistable by Proposition A.4(b) and Theorem 4.4(b). Fix $s \in c(I)$ and let $M_{Y,s}^\dagger$ be the Hermitian transpose of $M_{Y,s}$. Since $M_{Y,s}^{(0)} \notin \text{span}\{M_{Y,s}^{(1)}, \dots, M_{Y,s}^{(\beta_s)}\}$ we have

$$\ker(M_{Y,s}^\dagger) \subseteq \text{span}\{e_1, \dots, e_{\beta_s}\} \subseteq \mathbb{C}^{\beta_s+1}.$$

Therefore, e_0 is in the orthogonal complement of $\ker(M_{Y,s}^\dagger)$, i.e. in the image of $M_{Y,s}$, so there is some $z \in \mathbb{C}^{\alpha_s n}$ with $M_{Y,s} z = e_0$. By construction of the matrix $M_{Y,s}$, the equation

$$\begin{pmatrix} x_s & x_{s,1} & x_{s,2} & \cdots & x_{s,\beta_s} \end{pmatrix} (M_{Y,s} z) = \begin{pmatrix} x_s & x_{s,1} & x_{s,2} & \cdots & x_{s,\beta_s} \end{pmatrix} e_0 = x_s$$

shows that x_s is a \mathbb{C} -linear combination of the $\sum_{j=1}^m x_{c(kj)} Y_{j,l}$, where $k \in c^{-1}(s)$ and $l \in [n]$; the coefficients are given by $z \in \mathbb{C}^{\alpha_s n}$. In particular, $x_s \in R_Y$. Since the coordinate functions $x_s, s \in c(I)$ generate the character group $\mathfrak{X}(T)$ (thinking of characters as algebraic group morphisms $G \rightarrow \mathbb{C}^\times$), we conclude $\mathfrak{X}_{G \cdot Y} = \mathfrak{X}(T)$. Hence, $\mathfrak{X}_{G \cdot Y}$ is a group and $G \cdot Y$ is Zariski closed by Popov's criterion, Theorem B.4. \square

B.3. Bijection between the stabiliser and the set of MLEs

So far we have given two descriptions of the set of MLEs given Y in an RDAG model. Corollary 4.8 gives a linear space of possible Λ , while Proposition A.3 gives an additive bijection between the MLEs and the stabiliser. Here we give an alternative (multiplicative) bijection between the set of MLEs and the stabiliser, when $A(\mathcal{G}, c)$ is a group. This is similar to [2, Proposition 3.9], which gives a multiplicative surjection between the MLEs and the stabiliser, for a Gaussian group model on a reductive group.

Proposition B.6. *Consider the RDAG model on (\mathcal{G}, c) where colouring c is compatible and $A(\mathcal{G}, c)$ is a group. Set $A := A(\mathcal{G}, c)_{\text{SL}}$ and let $Y \in \mathbb{R}^{m \times n}$ be polystable under A . Given an MLE $\lambda a^\top a$, where $a \in A$ and λ is as in Proposition A.4, we have a bijection*

$$\begin{aligned} \varphi: A_Y &\rightarrow \{\text{MLEs given } Y\} \\ g &\mapsto \lambda g^\top a^\top a g. \end{aligned}$$

Proof. For $g \in A_Y$ we have $ag \cdot Y = aY$, which is of minimal norm in $A \cdot Y$ as $\lambda a^\top a$ is an MLE. Hence, $\varphi(g) = \lambda(ag)^\top(ag)$ is another MLE given Y , by Proposition A.4, and we see that φ is well-defined. For surjectivity, let $\lambda \tilde{a}^\top \tilde{a}$ be

another MLE given Y . Then aY and $\tilde{a}Y$ are of minimal norm in $A \cdot Y$, hence there is an orthogonal $t \in T$ with $taY = \tilde{a}Y$ by Lemma A.8(c). We obtain $g := a^{-1}t^{-1}\tilde{a} \in A_Y$, where we crucially used that $A(\mathcal{G}, c)$ (and hence A) is a group. The orthogonality of t gives $\varphi(g) = \lambda\tilde{a}^\top\tilde{a}$.

To prove injectivity, let $g, g' \in A_Y$ be such that $\varphi(g) = \varphi(g')$. The latter implies $g^\top a^\top a g = g'^\top a^\top a g'$, which shows that $h := ag'g^{-1}a^{-1}$ is orthogonal. Therefore, h is diagonal, because any orthogonal upper triangular matrix is diagonal. Moreover, using $g, g' \in A_Y$ we have $haY = aY$, i.e. $h \in A_{aY}$. Note Y and aY have the same orbit (closure), where we again use that A is a group. Thus, aY is polystable as Y is polystable. Combining this with $h(aY) = aY$ and h diagonal implies $h = \text{id}$. Finally, $\text{id} = h = ag'g^{-1}a^{-1}$ shows $g = g'$. \square

Acknowledgments

We are grateful to Dominic Bunnett, Mathias Drton, Robin Evans, Caroline Uhler, and Piotr Zwiernik for helpful discussions. We thank an anonymous referee for helpful comments that improved the paper.

Funding

VM was supported by the University of Melbourne and NSF Grant CCF 1900460. PR was funded by the European Research Council (ERC) under the European's Horizon 2020 research and innovation programme (grant agreement no. 787840).

References

- [1] ABBRUZZO, A., FASONE, V. and SCUDERI, R. (2016). Operational and financial performance of Italian airport companies: A dynamic graphical model. *Transport Policy* **52** 231–237.
- [2] AMÉNDOLA, C., KOHN, K., REICHENBACH, P. and SEIGAL, A. (2021). Invariant theory and scaling algorithms for maximum likelihood estimation. *SIAM Journal on Applied Algebra and Geometry* **5** 304–337. [MR4274830](#)
- [3] ANDERSSON, S., MADIGAN, D. and PERLMAN, M. (1997). On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. *Scandinavian Journal of Statistics* **24** 81–102. [MR1436624](#)
- [4] ANDERSSON, S. and MADSEN, J. (1998). Symmetry and lattice conditional independence in a multivariate normal distribution. *The Annals of Statistics* **26** 525–572. [MR1626059](#)
- [5] BUHL, S. L. (1993). On the existence of maximum likelihood estimators for graphical Gaussian models. *Scandinavian Journal of Statistics* 263–270. [MR1241392](#)
- [6] DANAHER, P., WANG, P. and WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* **76** 373. [MR3164871](#)

- [7] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* 157–175. [MR3931974](#)
- [8] DERKSEN, H. and MAKAM, V. (2017). Polynomial degree bounds for matrix semi-invariants. *Advances in Mathematics* **310** 44–63. [MR3620684](#)
- [9] DERKSEN, H. and MAKAM, V. (2021). Maximum likelihood estimation for matrix normal models via quiver representations. *SIAM Journal on Applied Algebra and Geometry* **5** 338–365. [MR4274831](#)
- [10] DERKSEN, H., MAKAM, V. and WALTER, M. (2020). Maximum likelihood estimation for tensor normal models via castling transforms. *arXiv preprint arXiv:2011.03849*. [MR4447881](#)
- [11] DOLGACHEV, I. (2003). *Lectures on invariant theory*. London Mathematical Society Lecture Note Series **296**. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CB09780511615436>. [MR2004511](#)
- [12] DRTON, M., FOX, C., KÄUFL, A. and POULIOT, G. (2019). The maximum likelihood threshold of a path diagram. *The Annals of Statistics* **47** 1536–1553. [MR3911121](#)
- [13] DRTON, M., FOYCEL, R. and SULLIVANT, S. (2011). Global identifiability of linear structural equation models. *Ann. Statist.* **39** 865–886. [MR2816341](#)
- [14] DRTON, M. et al. (2018). Algebraic problems in structural equation modeling. In *The 50th anniversary of Gröbner bases*, **77** 35–87. Mathematical Society of Japan Tokyo. [MR3839705](#)
- [15] FOULDS, L. R. (2012). *Graph theory applications*. Springer Science & Business Media. [MR1312607](#)
- [16] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.
- [17] FRIEDMAN, N., LINIAL, M., NACHMAN, I. and PE’ER, D. (2000). Using Bayesian networks to analyze expression data. *Journal of Computational Biology* **7** 601–620.
- [18] FRYDENBERG, M. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics* 333–353. [MR1096723](#)
- [19] GAO, X. and MASSAM, H. (2015). Estimation of symmetry-constrained Gaussian graphical models: application to clustered dense networks. *Journal of Computational and Graphical Statistics* **24** 909–929. [MR3432922](#)
- [20] GROSS, E. and SULLIVANT, S. (2018). The maximum likelihood threshold of a graph. *Bernoulli* **24** 386–407. [MR3706762](#)
- [21] HAGBERG, A., SWART, P. and SCHULT, D. (2008). Exploring network structure, dynamics, and function using NetworkX Technical Report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
- [22] HØJSGAARD, S. and LAURITZEN, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 1005–1027. [MR2530327](#)
- [23] LAURITZEN, S. L. (1996). *Graphical models* **17**. Clarendon Press. [MR1419991](#)
- [24] LOVÁSZ, L. (1979). On determinants, matchings, and random algorithms. In *Fundamentals of computation theory (Proc. Conf. Algebraic, Arith. and Categorical Methods in Comput. Theory)*. *Math. Res.* **2** 565–574. Akademie-

- Verlag, Berlin. [MR0563723](#)
- [25] MAATHUIS, M., DRTON, M., LAURITZEN, S. and WAINWRIGHT, M. (2018). *Handbook of graphical models*. CRC Press. [MR3889064](#)
- [26] MADSEN, J. (2000). Invariant normal models with recursive graphical Markov structure. *Annals of Statistics* 1150–1178. [MR1810923](#)
- [27] MICHĄŁEK, M. and STURMFELS, B. (2021). *Invitation to nonlinear algebra* **211**. American Mathematical Soc. [MR4423369](#)
- [28] MOMOZAWA, Y., MERVEILLE, A.-C., BATAILLE, G., WIBERG, M., KOCH, J., WILLESEN, J. L., PROSCHOWSKY, H. F., GOUNI, V., CHETBOUL, V., TIRET, L. et al. (2020). Genome wide association study of 40 clinical measurements in eight dog breeds. *Scientific Reports* **10** 1–11.
- [29] MUMFORD, D., FOGARTY, J. and KIRWAN, F. (1994). *Geometric invariant theory*, third ed. *Ergebnisse der Mathematik und ihrer Grenzgebiete (2) [Results in Mathematics and Related Areas (2)]* **34**. Springer-Verlag, Berlin. <https://doi.org/10.1007/978-3-642-57916-5>. [MR1304906](#)
- [30] PEARL, J. (2009). *Causality*. Cambridge University Press. [MR2548166](#)
- [31] PETERS, J. and BÜHLMANN, P. (2014). Identifiability of Gaussian structural equation models with equal error variances. *Biometrika* **101** 219–228. [MR3180667](#)
- [32] POPOV, V. L. (1989). Closed orbits of Borel subgroups. *Mathematics of the USSR-Sbornik* **63** 375. [MR0937648](#)
- [33] RASKUTTI, G. and UHLER, C. (2018). Learning directed acyclic graph models based on sparsest permutations. *Stat* **7** e183. [MR3796726](#)
- [34] SACHS, K., PEREZ, O., PE’ER, D., LAUFFENBURGER, D. A. and NOLAN, G. P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308** 523–529.
- [35] SCHWARTZ, J. T. (1980). Fast probabilistic algorithms for verification of polynomial identities. *J. Assoc. Comput. Mach.* **27** 701–717. <https://doi.org/10.1145/322217.322225>. [MR0594695](#)
- [36] SHAH, P. and CHANDRASEKARAN, V. (2012). Group symmetry and covariance regularization. In *2012 46th Annual Conference on Information Sciences and Systems (CISS)* 1–6. IEEE.
- [37] SHAO, B., RAMMOHAN, J., ANDERSON, D. A., ALPEROVICH, N., ROSS, D. and VOIGT, C. A. (2021). Single-cell measurement of plasmid copy number and promoter activity. *Nature Communications* **12** 1–9.
- [38] STURMFELS, B. and UHLER, C. (2010). Multivariate Gaussians, semidefinite matrix completion, and convex algebraic geometry. *Annals of the Institute of Statistical Mathematics* **62** 603–638. [MR2652308](#)
- [39] SULLIVANT, S. (2018). *Algebraic statistics. Graduate Studies in Mathematics* **194**. AMS. <https://doi.org/10.1365/s13291-019-00204-x>. [MR3838364](#)
- [40] TREFETHEN, L. N. and BAU, D. (1997). *Numerical linear algebra* **50**. SIAM. [MR1444820](#)
- [41] UHLER, C. (2012). Geometry of maximum likelihood estimation in Gaussian graphical models. *Annals of Statistics* **40** 238–261. [MR3014306](#)
- [42] VERMA, T. and PEARL, J. (1990). Causal networks: Semantics and expres-

- siveness. In *Machine intelligence and pattern recognition*, **9** 69–76. Elsevier. [MR1166827](#)
- [43] VINCIOTTI, V., AUGUGLIARO, L., ABBRUZZO, A. and WIT, E. C. (2016). Model selection for factorial Gaussian graphical models with an application to dynamic regulatory networks. *Statistical Applications in Genetics and Molecular Biology* **15** 193–212. [MR3510366](#)
- [44] WILLE, A., ZIMMERMANN, P., VRANOVÁ, E. et al. (2004). Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biology* **5** 1–13.