ESTIMABILITY OF COLOURED GAUSSIAN GRAPHICAL MODELS: THE SMALLEST COUNTEREXAMPLE TO THE FORBES–LAURITZEN BOUND

KOREA SUPERINTELLIGENCE LABS

Abstract. Forbes and Lauritzen study the score matching estimator (SME) of a linear concentration model, a linear subspace L of the symmetric $p \times p$ matrices, and call L *n-estimable* when the SME exists with probability one, equivalently when $\det M(x) > 0$ for some sample $x \in \mathbb{R}^{p\times n}$. They prove that $\dim L > n(2p - n + 1)/2$ forces non-estimability, and conjecture the converse for the models $S(\mathcal{N}, \mathcal{E})$ of coloured graphs — the RCON models of Højsgaard and Lauritzen — where $\dim L = |\mathcal{N}| + |\mathcal{E}|$. In its uncoloured form the conjecture is false: Gross and Sullivant refute it with K_4 plus a pendant vertex at $n = 3$ on five vertices, and their rigidity criterion explains why. None of that theory reaches the coloured case, which the recent survey literature records as open. We settle its bottom layer. At $n = 1$ the bound forces an uncoloured graph to have no edges at all, so the uncoloured conjecture is *true* there; with colours it is false on four vertices, and we exhibit two counterexamples with $|\mathcal{N}| + |\mathcal{E}| = 4 = 1 \cdot (2 \cdot 4 - 1 + 1)/2$, one carrying a single edge and one connected (a coloured K_4). An exhaustive census of all coloured graphs on at most four vertices shows that no counterexample is smaller, in the number of vertices or in $|\mathcal{N}| + |\mathcal{E}|$. We also show that the hereditary strengthening of the conjecture — impose the count on every induced subgraph — fails on four vertices, where the simplest uncoloured witness named by Gross and Sullivant, the double banana, has eight. Every estimability verdict we prove is formally verified in Lean 4, and we say exactly which claims rest on exhaustive computation instead.

1. Introduction

Linear concentration models and the score matching estimator. A *linear concentration model* on p variables is a linear subspace L of the symmetric $p \times p$ real matrices $\mathrm{Sym}_p(\mathbb{R})$; the Gaussian family it defines is $\{N(0, K^{-1}) : K \in L,\ K \succ 0\}$, the concentration matrix being the parameter that varies linearly. Forbes and Lauritzen [3] study the *score matching estimator* (SME) in this family — the estimator obtained from Hyvärinen's score matching criterion [6], which for Gaussian linear concentration models is a system of *linear* estimating equations. Given observations $x = (x^1, \ldots, x^n) \in \mathbb{R}^{p\times n}$, those equations have a unique solution exactly when the quadratic form

$$D_2(K) \;=\; \sum_{i=1}^{n} \left\| K x^i \right\|^2$$

is positive definite on L. Fixing an orthogonal basis $e^1, \ldots, e^d$ of L, the matrix of D_2 in that basis is

$$m_{uv}(x) \;=\; \sum_{i=1}^{n} \langle e^u x^i,\, e^v x^i \rangle, \qquad 1 \le u, v \le d, \tag{1}$$

and [3] make the definition we work with throughout, verbatim:

> We shall say that the linear space L is *n-estimable* if the SME exists with probability one, or equivalently, if there is an $x = (x^1, \ldots, x^n) \in \mathbb{R}^{p\times n}$ such that $\det M(x) > 0$.

The equivalence is immediate from (1): $\det M(\cdot)$ is a polynomial in the entries of x, so it either vanishes identically or is nonzero off a Lebesgue-null set. The SME is not the maximum likelihood estimator. Forbes and Lauritzen prove that existence of the SME implies existence of the MLE ([3, Proposition 3]) and note that the converse fails — "the MLE may well exist even if the SME does not exist" — so a model can fail to be n-estimable and still admit the MLE at some samples; we say "n-estimable" below only in their sense.

The obstruction they isolate is a dimension count. Write $T_k = k(k+1)/2$. A generic n-plane in $\mathbb{R}^p$ is annihilated by a T_{p-n}-dimensional space of symmetric matrices, so if $\dim L$ exceeds $T_p - T_{p-n} = n(2p-n+1)/2$ then every sample admits a nonzero $K \in L$ with $Kx^i = 0$ for all i, and $\det M(x) = 0$ identically.

Proposition 1.1 ([3, Proposition 4]). *If $\dim L > n(2p-n+1)/2$ then L is not n-estimable.*

Coloured graphs. A *coloured graph* $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ on a vertex set V with $|V| = p$ consists of a simple graph $G = (V, E)$ together with a partition $\mathcal{N}$ of V into *vertex colour classes* and a partition $\mathcal{E}$ of E into *edge colour classes*. These are the RCON models of Højsgaard and Lauritzen [5]: the associated linear concentration model is

$$S(\mathcal{N}, \mathcal{E}) = \left\{ K \in \mathrm{Sym}_p(\mathbb{R}) \;:\; \begin{array}{l} k_{ab} = 0 \text{ for } a \neq b,\ ab \notin E; \quad k_{aa} = k_{bb} \text{ for } a, b \text{ in one class of } \mathcal{N}; \\ k_{ab} = k_{cd} \text{ for } ab, cd \text{ in one class of } \mathcal{E} \end{array} \right\},$$

whose 0/1 indicator matrices of the colour classes form an orthogonal basis; hence $\dim S(\mathcal{N}, \mathcal{E}) = |\mathcal{N}| + |\mathcal{E}|$. Taking every class to be a singleton recovers the ordinary Gaussian graphical model of G, with $\dim L = |V| + |E|$.

For these models Forbes and Lauritzen state a belief, which is the subject of this paper ([3], end of their section on Gaussian graphical models with symmetries, verbatim, except that their symbol $\mathcal{V}$ for the vertex-colour partition is written $\mathcal{N}$ here):

> We note that our Proposition 4 implies that the SME does not exist if $|\mathcal{N}| + |\mathcal{E}| > n(2|V| - n + 1)/2$ and believe that the condition $|\mathcal{N}| + |\mathcal{E}| \leq n(2|V| - n + 1)/2$ is sufficient to ensure n-estimability for this particular class of linear concentration models, but have not been able to show this except for the case of $n = p - 1$, cf. Proposition 5.

Write

$$\beta(p, n) \;=\; \frac{n(2p - n + 1)}{2} \;=\; T_p - T_{p-n}, \tag{2}$$

and call the assertion *if $\mathcal{G}$ is a coloured graph on p vertices with $|\mathcal{N}| + |\mathcal{E}| \leq \beta(p, n)$ then $S(\mathcal{N}, \mathcal{E})$ is n-estimable* the *Forbes–Lauritzen conjecture.*

What is known. In its *uncoloured* form the conjecture is refuted in print. Gross and Sullivant [4] define the score matching threshold $\mathrm{smt}(G)$ of a graph, prove $\mathrm{smt}(G) = \mathrm{rank}(G)$ (their Theorem 6.3); combining that with their Theorem 2.5 gives that an uncoloured G is n-estimable if and only if $E(G)$ is independent in the generic $(n-1)$-dimensional rigidity matroid $A(n-1)$. Their Conjecture 6.7 is the all-singleton-classes case of the belief above — they attribute it to a 2014 Prague lecture of Lauritzen — and their Example 6.8 refutes it: K_4 with a pendant vertex has $p = 5$, $\#V + \#E = 12 = \beta(5, 3)$, and is not 3-estimable, because it contains the clique K_4 and $4 > 3$. On eight vertices the *double banana* — two copies of K_5 glued along two vertices with the joining edge deleted — has $\#V + \#E = 26 = \beta(8, 4)$ and is likewise not 4-estimable, this time for a reason that is not a clique ([4], p. 404).

None of this reaches coloured graphs. Gross and Sullivant's theory, and the rigidity-theoretic maximum-likelihood-threshold literature that grew from it [1], is about uncoloured graphs throughout; the coloured (RCON) case is recorded as open in the recent survey [2], which asks for methods of computing and bounding the maximum likelihood thresholds of RCON models in terms of the combinatorics of the graph and its vertex/edge partitions.

Coloured models do appear in the earlier estimability literature as examples rather than as counterexamples. Uhler's Table 2 [10] tabulates coloured four-cycle models on four vertices together with the number of observations needed for the MLE to exist, and Forbes and Lauritzen, after discussing that table, write: "However, none of the examples in [37] or [45] provide counterexamples to this conjecture." Their [37] is Jensen [7] and their [45] is Uhler [10]. The census of Section 7 confirms this on four vertices: no counterexample there at $n = 1$ has the four-cycle as its underlying graph.

Results. The coloured case is genuinely different at the bottom of its range, and that is what we settle. At $n = 1$ the bound (2) reads $|\mathcal{N}| + |\mathcal{E}| \le p$. For an *uncoloured* graph this forces $|\mathcal{N}| = p$, hence $|\mathcal{E}| = 0$: the graph has no edges, L is the space of diagonal matrices, and L is 1-estimable. So the uncoloured conjecture is *true* at $n = 1$, vacuously, and Gross–Sullivant's criterion says nothing there — $A(0)$ is the rank-zero matroid. With colours the conjecture is false already at $n = 1$ and already on four vertices.

- Theorem 3.1: the coloured graph on $V = \{0, 1, 2, 3\}$ with the single edge $\{0, 1\}$ and vertex colour classes $\{0\}, \{1\}, \{2, 3\}$ has $|\mathcal{N}|+|\mathcal{E}| = 4 = \beta(4, 1)$ and is not 1-estimable. Corollary 3.2: the Forbes–Lauritzen conjecture is false.
- Theorem 3.3: so does a *connected* coloured graph of the same size — K_4 with vertex classes $\{0\}$ and $\{1, 2, 3\}$ and edge classes "the star at 0" and "the triangle on $\{1, 2, 3\}$". Its kernel certificate is explicit and needs no linear solve.
- Theorem 6.3: the *hereditary* strengthening of the conjecture, which imposes the count $\dim L_S \le \beta(|S|, n)$ on every induced sub-model, also fails, and the same coloured K_4 is a witness. Gross and Sullivant write that in the uncoloured case "the simplest such graph is the double banana", on eight vertices; with colours four vertices suffice.
- Section 7: an exhaustive census over *all* coloured graphs on at most four vertices — all graphs, all partitions of the vertices, all partitions of the edges, all n — finds no counterexample on three or fewer vertices, and finds that at $p = 4$ the least possible value of $|\mathcal{N}| + |\mathcal{E}|$ among counterexamples is 4, attained only at $n = 1$. With Forbes–Lauritzen's own theorem for $n = p - 1$ this pins the smallest counterexample: $p = 4$, $|\mathcal{N}| + |\mathcal{E}| = 4$, $n = 1$, and necessarily coloured. This is the one claim of the paper that rests on exhaustive computation, and Section 7 says exactly what was run and in which direction the verdicts are certain.

Sections 4 and 5 record the surrounding certificates that make these statements say what they appear to say: models sitting exactly at the bound that *are* estimable (so that the bound is not secretly an equality obstruction), and the known uncoloured counterexamples — Gross–Sullivant's Example 6.8 and the double banana — proved in the same framework, the double banana by a rank-two certificate valid at every sample.

Every verdict about a specific model stated below as a theorem or proposition — every claim of the form "this model is" or "is not n-estimable" — is formally verified in Lean 4 against *Mathlib*, the one exception being the census of Section 7, which is a computation and is labelled as such; Section 8 describes the development and lists what lies outside it. (Proposition 1.1 is quoted from [3] and is not part of the development.) In particular the negative certificates are proved *at every sample separately*: no genericity argument, no polynomial identity theorem, and no Zariski density appears anywhere in the development.

Organisation. Section 2 fixes notation and proves the certificate lemmas. Section 3 contains the two counterexamples at $n = 1$ and the uncoloured contrast, Section 4 the positive certificates, Section 5 the uncoloured counterexamples in the same framework, Section 6 the hereditary strengthening, Section 7 the census and the minimality claim, Section 8 the formal verification and Section 9 what is left open.

2. Preliminaries

Models, samples, and the Gram matrix. Fix $p, d, n \ge 1$. We index vertices by $V = \{0, 1, \ldots, p-1\}$, matching the formal development. A sample is a matrix $X \in \mathbb{R}^{p\times n}$ whose columns are the observations $x^1, \ldots, x^n$. For a coloured graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ and a colour class u let $e^u \in \mathrm{Sym}_p(\mathbb{R})$ be its 0/1 indicator matrix:

$$e^u = \sum_{a\in u} E_{aa} \quad (u \in \mathcal{N}), \qquad e^u = \sum_{ab\in u} \big(E_{ab} + E_{ba}\big) \quad (u \in \mathcal{E}),$$

where E_{ab} is a matrix unit. Distinct classes have disjoint supports, so $e^1, \dots, e^d$ ($d = |\mathcal{N}| + |\mathcal{E}|$) are pairwise orthogonal and form a basis of $L = S(\mathcal{N}, \mathcal{E})$. We write $K(\kappa) = \sum_u \kappa_u e^u$ for the element of L with parameter vector $\kappa \in \mathbb{R}^d$, and $M(X) = \big(m_{uv}(X)\big)_{u,v}$ for the matrix (1), which in this notation reads

$$m_{uv}(X) \;=\; \sum_{a \in V} \sum_{j=1}^{n} \big(e^u X\big)_{aj} \big(e^v X\big)_{aj}.$$

Remark 2.1. The indicator basis is orthogonal but not normalised, and nothing depends on either. If A is the change-of-basis matrix between two bases of L then the two Gram matrices differ by $M' = A^\mathsf{T} M A$, so $\det M' = (\det A)^2 \det M$ and the sign of $\det M(X)$ — hence n-estimability — does not depend on the choice.

Let $D(X) : \mathbb{R}^d \to \mathbb{R}^{p \times n}$, $\kappa \mapsto K(\kappa)X$, be the design map of L written in the basis e^u. Directly from the definitions, $M(X) = D(X)^\mathsf{T} D(X)$.

Lemma 2.2. *$M(X)$ is positive semidefinite; in particular $\det M(X) \geq 0$, and L is n-estimable if and only if $\det M(X) \neq 0$ for some $X \in \mathbb{R}^{p \times n}$. Moreover $\ker M(X) = \ker D(X)$.*

Proof. $M(X) = D(X)^\mathsf{T} D(X)$ is a Gram matrix, so it is positive semidefinite and its determinant is nonnegative; a nonnegative number is positive exactly when it is nonzero. If $M(X)\kappa = 0$ then $\|D(X)\kappa\|^2 = \kappa^\mathsf{T} M(X)\kappa = 0$, so $D(X)\kappa = 0$; the reverse inclusion is trivial. ☐

Lemma 2.2 is what makes a single integer sample a complete positive certificate: the entries of $M(X)$ are integer polynomials in the entries of X with the same coefficients over any commutative ring, because the basis is 0/1, so an integer sample with $\det M(X) \neq 0$ over $\mathbb{Z}$ certifies estimability over $\mathbb{R}$.

The kernel criterion. All the negative results below are instances of one lemma, and it is proved for the sample at hand, not for a generic one.

Lemma 2.3. *Let $X \in \mathbb{R}^{p \times n}$ and $\kappa \in \mathbb{R}^d$ with $\kappa \neq 0$. If $K(\kappa)X = 0$ then $\det M(X) = 0$. Consequently, if every $X \in \mathbb{R}^{p \times n}$ admits such a κ, then L is not n-estimable.*

Proof. From $M(X) = D(X)^\mathsf{T} D(X)$ and $D(X)\kappa = K(\kappa)X = 0$ we get $M(X)\kappa = 0$. The adjugate identity $\operatorname{adj}(M)M = (\det M)I$ then gives $(\det M(X))\,\kappa = \operatorname{adj}(M(X))\big(M(X)\kappa\big) = 0$, and since some coordinate of κ is nonzero, $\det M(X) = 0$. The last sentence is Lemma 2.2. ☐

Lemma 2.4. *Let $X \in \mathbb{R}^{p \times n}$.*

1. *If $u \in \mathbb{R}^p$ satisfies $X^\mathsf{T} u = 0$ and $uu^\mathsf{T} = K(\kappa)$ for some κ, then $K(\kappa)X = 0$.*
2. *If $y, z \in \mathbb{R}^p$ satisfy $X^\mathsf{T} y = X^\mathsf{T} z = 0$ and $s\,yy^\mathsf{T} + t\,zz^\mathsf{T} = K(\kappa)$ for some κ and scalars s, t, then $K(\kappa)X = 0$.*

Proof. $(uu^\mathsf{T})X = u\,(u^\mathsf{T} X) = 0$, and the second statement is the same computation applied twice and added. ☐

We call the first shape a *rank-one certificate* and the second a *rank-two certificate*. The vectors are produced by counting:

Lemma 2.5. *Let $A \in \mathbb{R}^{m \times k}$ with $k > m$. Then $Av = 0$ for some $v \neq 0$.*

Proof. Pad A with $k - m$ zero rows to a square $k \times k$ matrix A'; then $\det A' = 0$, so A', hence A, has a nonzero kernel vector. ☐

Remark 2.6. For an *uncoloured* graph, $uu^\mathsf{T} \in L$ says exactly that $\operatorname{supp}(u)$ is a clique of G: the entries of uu^T outside $\operatorname{supp}(u) \times \operatorname{supp}(u)$ vanish, and the entries inside must be supported on edges. Since Lemma 2.5 produces a nonzero u supported on any prescribed set of more than n vertices with $X^\mathsf{T} u = 0$, every clique on more than n vertices is an obstruction. This is not new: it is the certificate-side form of the Laman-type necessary condition ([4, Theorem 3.2]), which K_{n+1} violates, and Gross and Sullivant use its $n = 3$ case in their Example 6.8, citing their Corollary 6.4. *Colouring changes what the condition asks*: the products $u_i u_j$ need only be constant on colour classes, and merging classes can arrange that — while at the same time lowering $|\mathcal{N}| + |\mathcal{E}|$, which is what lets a graph with edges satisfy the count at all. In Theorem 3.1 the support of u is the edge $\{0, 1\}$, a clique on $2 > n = 1$ vertices, and the colouring's role is purely to make the count fit, by merging two vertices into one class; in Theorem 3.3 the support is all of V and it is the colour constraints on the products $u_i u_j$ that pin the certificate down. Neither route is available uncoloured at $n = 1$, where the count leaves no edges at all.

The formal encoding, and why it is a restriction. In the formal development a model on p vertices with d colour classes is a symmetric *class map* $c : V \times V \to \{1, \ldots, d\} \cup \{\perp\}$, where $c(a, b) = u$ says that the (a, b) entry of $K \in L$ is the free parameter of class u and $c(a, b) = \perp$ says the entry is structurally zero. A class map comes from a coloured graph precisely when

(1) every class is used — so that d really is $|\mathcal{N}| + |\mathcal{E}|$ — and
(2) (*purity*) a class consists of diagonal positions or of off-diagonal positions, never both.

Purity is what separates vertex classes from edge classes; without it one could equate a diagonal and an off-diagonal entry, which is a linear concentration model but not a coloured graph. The distinction is not idle: Forbes and Lauritzen exhibit a linear concentration model with $p = 4$ and $\dim L = 4 = \beta(4, 1)$ that is not 1-estimable — their equation (15), a Jordan subalgebra they attribute to Jensen [7], namely

$$L = \left\{ \begin{pmatrix} a & c & 0 & f \\ c & b & -f & 0 \\ 0 & -f & a & c \\ f & 0 & c & b \end{pmatrix} \ : \ a, b, c, f \in \mathbb{R} \right\}.$$

The basis element carrying f has entries of both signs, so this L is not the model of a coloured graph, its off-diagonal classes not being 0/1 indicator matrices. That is exactly why their belief is restricted to that class. Purity is a real restriction, and the refutations below therefore refute a non-vacuous statement:

Proposition 2.7. *The class map on two vertices that assigns the single colour class to all four positions — both diagonal entries and both off-diagonal ones — violates the purity condition.*

Proof. A finite check on the four positions: the class of $(0, 0)$ equals the class of $(0, 1)$ and is not $\perp$, but $0 = 0$ while $0 \neq 1$. $\square$

3. Two counterexamples at $n = 1$

At $n = 1$ a sample is a single column $x \in \mathbb{R}^p$ and the bound is $\beta(p, 1) = p$.

Theorem 3.1. *Let $\mathcal{G}$ be the coloured graph on $V = \{0, 1, 2, 3\}$ whose only edge is $\{0, 1\}$, with vertex colour classes $\{0\}, \{1\}, \{2, 3\}$ and the single edge colour class $\{\{0, 1\}\}$. Then*

$$|\mathcal{N}| + |\mathcal{E}| = 3 + 1 = 4 = \frac{1 \cdot (2 \cdot 4 - 1 + 1)}{2} = \beta(4, 1),$$

and $S(\mathcal{N}, \mathcal{E})$ is not 1-estimable.

Proof. Let $x = (x_0, x_1, x_2, x_3)^\mathsf{T} \in \mathbb{R}^4$ be any sample. The single linear equation $x_0 v_0 + x_1 v_1 = 0$ in two unknowns has a nonzero solution $v \in \mathbb{R}^2$ by Lemma 2.5. Put $u = (v_0, v_1, 0, 0)^\mathsf{T}$, so that $x^\mathsf{T} u = 0$, and let $K = uu^\mathsf{T}$, i.e.

$$k_{00} = v_0^2, \qquad k_{11} = v_1^2, \qquad k_{01} = k_{10} = v_0 v_1, \qquad k_{ab} = 0 \text{ otherwise.}$$

Then $K \in S(\mathcal{N}, \mathcal{E})$: the two vertices $2, 3$ of the merged class both receive the diagonal value 0, and every position outside $\{0, 1\} \times \{0, 1\}$ is a non-edge and is zero. Its parameter vector is $\kappa = (v_0^2,\, v_1^2,\, 0,\, v_0 v_1)$ in the order $\{0\}, \{1\}, \{2, 3\}, \{\{0, 1\}\}$, and $\kappa \neq 0$ because $v \neq 0$ forces $v_0^2 \neq 0$ or

$v_1^2 \neq 0$. By Lemma 2.4(1), $Kx = 0$; by Lemma 2.3, $\det M(x) = 0$. As x was arbitrary, $S(\mathcal{N}, \mathcal{E})$ is not 1-estimable. $\square$

Corollary 3.2. *The Forbes–Lauritzen conjecture is false: there are a coloured graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ on p vertices and an $n \geq 1$ with $|\mathcal{N}| + |\mathcal{E}| \leq \beta(p, n)$ such that $S(\mathcal{N}, \mathcal{E})$ is not n-estimable.*

Proof. Theorem 3.1 with $p = 4$, $n = 1$. $\square$

The model of Theorem 3.1 is disconnected, and one might read the failure as an artefact of the isolated pair. It is not.

Theorem 3.3. *Let $\mathcal{G}$ be K_4 on $V = \{0, 1, 2, 3\}$ with vertex colour classes $\{0\}$ and $\{1, 2, 3\}$ and edge colour classes $\big\{\{0,1\}, \{0,2\}, \{0,3\}\big\}$ (the star at 0) and $\big\{\{1,2\}, \{1,3\}, \{2,3\}\big\}$ (the triangle on $\{1, 2, 3\}$). Then $\mathcal{G}$ is connected,*

$$|\mathcal{N}| + |\mathcal{E}| = 2 + 2 = 4 = \beta(4, 1),$$

and $S(\mathcal{N}, \mathcal{E})$ is not 1-estimable.

Proof. Let $x \in \mathbb{R}^4$ and put $w = x_0$ and $s = x_1 + x_2 + x_3$.

If $w = s = 0$, take $K = E_{00}$, the indicator of the vertex class $\{0\}$; then $K \neq 0$, and Kx has single nonzero coordinate $x_0 = w = 0$, so $Kx = 0$.

Otherwise put $v = (s, -w, -w, -w)^\mathsf{T}$. Then $x^\mathsf{T} v = sx_0 - w(x_1 + x_2 + x_3) = sw - ws = 0$ *identically* — no linear solve is needed — and $K = vv^\mathsf{T}$ has

$$k_{00} = s^2, \qquad k_{ii} = w^2 \ (i \geq 1), \qquad k_{0i} = -sw \ (i \geq 1), \qquad k_{ij} = w^2 \ (i \neq j,\ i, j \geq 1),$$

which is constant on each of the four colour classes; so $K = K(\kappa)$ with $\kappa = (s^2,\ w^2,\ -sw,\ w^2)$ in the order $\{0\}, \{1, 2, 3\}, \text{star}, \text{triangle}$. Here $\kappa = 0$ would force $s^2 = w^2 = 0$, contrary to the case assumption, so $\kappa \neq 0$; and $Kx = v(v^\mathsf{T} x) = 0$ by Lemma 2.4(1).

In both cases Lemma 2.3 gives $\det M(x) = 0$, for every x. $\square$

Remark 3.4. The vector $v = (x_1 + x_2 + x_3,\ -x_0,\ -x_0,\ -x_0)$ is the "centre versus rim" vector of the colouring. The colouring's automorphism group is the S_3 permuting $\{1, 2, 3\}$ and fixing 0, and its invariant subspace in $\mathbb{R}^4$ is the plane P spanned by $(1, 0, 0, 0)$ and $(0, 1, 1, 1)$; the certificate is simply the line of P orthogonal to x, which is nonzero unless $x \perp P$. It is therefore uniform in x, which is why the proof is a computation rather than a search.

Both counterexamples are coloured, and necessarily so:

Proposition 3.5. *The uncoloured empty graph on four vertices has $|\mathcal{N}| + |\mathcal{E}| = 4 + 0 = 4 = \beta(4, 1)$ and is 1-estimable: for $x = (1, 1, 1, 1)^\mathsf{T}$ one has $M(x) = I_4$, of determinant 1.*

Proof. L is the space of diagonal matrices with basis $e^u = E_{uu}$, and $m_{uv}(x) = \sum_a \sum_j (e^u x)_a (e^v x)_a = \delta_{uv} x_u^2 = \delta_{uv}$. $\square$

Remark 3.6. The same two lines prove more, and we record it as an elementary observation rather than as a theorem, because only the case $p = 4$ of Proposition 3.5 is part of the formal development. At $n = 1$ the bound reads $|\mathcal{N}| + |\mathcal{E}| \leq p$; an uncoloured graph on p vertices has $|\mathcal{N}| = p$, so the bound forces $|\mathcal{E}| = 0$, i.e. no edges, and then $M\big((1, \ldots, 1)^\mathsf{T}\big) = I_p$. *The uncoloured Forbes–Lauritzen conjecture is true at $n = 1$ for every p*, and the whole $n = 1$ layer of the coloured problem is invisible to the uncoloured theory — Gross–Sullivant's criterion at $n = 1$ concerns $A(0)$, the rank-zero rigidity matroid, in which no nonempty edge set is independent.

Proposition 3.5 is also the sharpest control on Theorems 3.1 and 3.3: same p, same n, same $|\mathcal{N}| + |\mathcal{E}| = 4$, opposite verdict. The failure is therefore not an artefact of the arithmetic of the bound at $(p, n) = (4, 1)$.

4. Positive certificates at the bound

A negative result about a bound is only as good as the evidence that the bound is not vacuously tight. The following two models sit at or below the bound and *are* estimable, each certified by one explicit integer sample; by Lemma 2.2 a single such sample is a complete proof.

Proposition 4.1. *The uncoloured path* $0-1-2$ *on three vertices has* $|\mathcal{N}|+|\mathcal{E}| = 3+2 = 5 = \beta(3,2)$*, exactly at the bound, and is* 2*-estimable: for* $x^1 = (1,0,-1)^\mathsf{T}$*,* $x^2 = (0,-1,0)^\mathsf{T}$ *the Gram matrix is* $\operatorname{diag}(1,1,1) \oplus \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$*, of determinant* 3.

Proposition 4.2. *The uncoloured four-cycle* $0-1-2-3-0$ *has* $|\mathcal{N}|+|\mathcal{E}| = 4+4 = 8 \le 9 = \beta(4,3)$ *and is* 3*-estimable: for* $x^1 = (0,-1,0,0)^\mathsf{T}$*,* $x^2 = (0,0,0,1)^\mathsf{T}$*,* $x^3 = (-1,0,1,0)^\mathsf{T}$ *the Gram matrix is* $I_4 \oplus \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \oplus \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$*, of determinant* 9.

Proposition 4.2 is not new: the four-cycle is the model of Forbes and Lauritzen's own Figure 1, and they write of it, verbatim, "For $n = 2$ we have $T_p - T_r = 10 - 3 = 7$, and since $\dim L = 8$, Proposition 4 yields that the SME does not exist. For $n = 3$ and above both the SME and MLE exist with probability one, the SME being a solution to a system of eight linear equations." We reproduce the second sentence with an explicit certificate; the first is Proposition 1.1.

Proof of Propositions 4.1 and 4.2. In both cases the displayed sample is an integer matrix; evaluating (1) over $\mathbb{Z}$ gives the displayed Gram matrix, whose determinant is nonzero. Because the basis is 0/1, the same polynomial formula computes M over $\mathbb{R}$, so $\det M \neq 0$ there too, and Lemma 2.2 upgrades this to $\det M > 0$. In the formal development the determinant is never expanded: a right inverse $MB = cI$ over $\mathbb{Z}$ with $c = 3$ is exhibited in each case, and $\det M \cdot \det B = c^d \neq 0$. □

5. The uncoloured counterexamples, in the same framework

For contrast, and because Section 6 needs them, we record the known uncoloured counterexamples with certificates of the same shape. None of the three facts in this section is new; what is ours is the proof in the formal development, and in the case of Proposition 5.3 the certificate.

Proposition 5.1. *The uncoloured triangle on* $\{0,1,2\}$ *together with an isolated vertex* 3 *has* $|\mathcal{N}| + |\mathcal{E}| = 4+3 = 7 = \beta(4,2)$ *and is not* 2*-estimable.*

Proof. Given $X \in \mathbb{R}^{4\times 2}$, the two equations $\sum_{k\le 2} v_k X_{kj} = 0$, $j = 1,2$, in the three unknowns v_0, v_1, v_2 have a nonzero solution (Lemma 2.5). Put $u = (v_0, v_1, v_2, 0)$; then $\operatorname{supp}(u)$ lies in the triangle, so $uu^\mathsf{T} \in L$, and Lemmas 2.4(1) and 2.3 apply. □

Proposition 5.1 is the $n = 2$ instance of Gross–Sullivant's criterion — $A(1)$ is the graphic matroid and a triangle is a circuit in it — and we claim no novelty for it; it is stated because it is the smallest *uncoloured* counterexample and therefore the natural yardstick for Theorems 3.1 and 3.3.

Proposition 5.2 ([4, Example 6.8])**.** *The uncoloured graph* K_4 *on* $\{0,1,2,3\}$ *with a pendant edge* $\{3,4\}$ *has* $|\mathcal{N}| + |\mathcal{E}| = 5+7 = 12 = \beta(5,3)$ *and is not* 3*-estimable.*

Proof. As for Proposition 5.1, with u supported on the clique $\{0,1,2,3\}$: three equations in four unknowns. Gross and Sullivant reach the same conclusion differently, from their Corollary 6.4, "since the complete graph K_4 is a subgraph of G". □

Proposition 5.3 ([4, p. 404])**.** *The double banana — two copies of* K_5 *glued along the two vertices* $\{0,1\}$ *with the joining edge deleted, so* $p = 8$ *and* $|E| = 18$ *— has* $|\mathcal{N}| + |\mathcal{E}| = 8+18 = 26 = \beta(8,4)$ *and is not* 4*-estimable.*

Gross and Sullivant state this, verbatim: "the double banana graph, which satisfies all these inequalities for $n = 4$ but is not a rigid graph in $A(3)$. So the Gaussian graphical model associated to the double banana graph is not 4-estimable." Their proof is the rigidity criterion. Ours is a certificate, and it is necessarily of a different shape from the ones above, because the largest clique of the double banana has four vertices, which is *not* more than $n = 4$: the counting argument of Remark 2.6 produces no rank-one certificate here. We sketch it, since it is the only certificate in this paper that is not rank one.

Proof sketch of Proposition 5.3. Label the vertices so that the non-edges are $\{0, 1\}$ and the nine pairs $\{2, 3, 4\} \times \{5, 6, 7\}$. Given $X \in \mathbb{R}^{8\times 4}$, Lemma 2.5 produces three nonzero vectors killed by $X^\top$ — four equations in five unknowns each — supported respectively on $\{0, 1, 5, 6, 7\}$, on $\{0, 1, 2, 3, 4\}$ and on $\{2, 3, 4, 5, 6\}$; call them a, b, c. Four branches, exhausting the possibilities:

branch	K	why $K \neq 0$
$a_0a_1 = 0$	$aa^\top$	$a \neq 0$
$b_0b_1 = 0$	$bb^\top$	$b \neq 0$
otherwise, $a_v \neq 0$ for some $v \in \{5, 6, 7\}$	$(b_0b_1)\, aa^\top - (a_0a_1)\, bb^\top$	$k_{vv} = (b_0b_1)a_v^2 \neq 0$
otherwise ($\operatorname{supp}(a) \subseteq \{0, 1\}$)	$ac^\top + ca^\top$	$k_{0j} = a_0c_j$, with $a_0 \neq 0$, $c \neq 0$

In each branch K vanishes at all ten non-edge positions — for the first two because the vector vanishes on one banana's private vertices, for the third because the $\{0, 1\}$ entries cancel by construction and the cross entries vanish termwise, for the fourth because c vanishes on $\{0, 1\}$ and a on $\{2, \dots, 7\}$ — so $K \in L$, and K annihilates X by Lemma 2.4. All four are instances of the rank-two shape, since $ac^\top + ca^\top = \frac{1}{2}(a + c)(a + c)^\top - \frac{1}{2}(a - c)(a - c)^\top$. Lemma 2.3 finishes. Note that the case analysis is on the sample, and each branch produces a kernel element for the sample at hand; no genericity is used. □

6. The hereditary strengthening fails on four vertices

Proposition 1.1 applies to every sub-model. For $S \subseteq V$ put

$$L_S = \{K \in L \ : \ k_{ij} = 0 \text{ unless } i, j \in S\},$$

whose dimension is the number of colour classes entirely contained in $S\times S$. If $\dim L_S > \beta(|S|, n)$ for some S with $|S| > n$ then L is not n-estimable, for the trivial reason that a nonzero kernel element of the sub-model is one of L; call such an obstruction *local*. (For $|S| \le n$ the sub-model has at least as many observations as variables and the count is uninformative.) This suggests strengthening the conjecture by imposing the count everywhere.

Definition 6.1. L is *hereditarily admissible at* n if $\dim L_S \le \beta(|S|, n)$ for every $S \subseteq V$ with $|S| > n$.

Conjecture 6.2 (the hereditary strengthening)**.** If $L = S(\mathcal{N}, \mathcal{E})$ is hereditarily admissible at n then L is n-estimable.

For uncoloured graphs Gross and Sullivant record that this too is false, and name the smallest witness they know ([4], p. 404, verbatim):

> Even with the stronger condition that the inequality $\#V' + \#E' \le n\#V' - \binom{n}{2}$ for every induced subgraph $G' = (V', E')$ of G, there are known counterexamples of graphs satisfying all of these inequalities but not being rigid. The simplest such graph is the double banana …

The double banana has eight vertices. With colours, four suffice.

Theorem 6.3. *Let $\mathcal{G}$ be the coloured K_4 of Theorem 3.3. Then $S(\mathcal{N}, \mathcal{E})$ is hereditarily admissible at $n = 1$ and is not 1-estimable, so Conjecture 6.2 is false on four vertices. By contrast the models of Theorem 3.1 and of Propositions 5.1 and 5.2 are all local: their hereditary counts fail at $S = \{0, 1\}$, at $S = \{0, 1, 2\}$ and at $S = \{0, 1, 2, 3\}$ respectively, i.e. exactly at their cliques.*

Proof. The hereditary counts are finite: for the coloured K_4 one checks the eleven subsets S with $|S| > 1$. The class $\{0\}$ is contained in $S \times S$ iff $0 \in S$; the vertex class $\{1,2,3\}$ and the triangle class iff $\{1,2,3\} \subseteq S$; the star class iff $S = V$. So for $|S| = 2$ we get $\dim L_S \leq 1 \leq \beta(2,1) = 2$; for $|S| = 3$ either $S = \{1,2,3\}$ and $\dim L_S = 2 \leq \beta(3,1) = 3$, or $0 \in S$ and $\dim L_S = 1$; and for $S = V$ all four classes are counted, $4 = \beta(4,1)$. Non-estimability is Theorem 3.3. For the three local models the displayed subsets carry $3 > \beta(2,1) = 2$, $6 > \beta(3,2) = 5$ and $10 > \beta(4,3) = 9$ colour classes respectively. □

Both halves of Theorem 6.3 are finite decidable computations over the colour tables, evaluated by the kernel of the proof assistant rather than by an external program; what is pinned to the development is hereditary admissibility and its negation for the four models, the violating subsets displayed above being read off the same evaluation. The same check applied to the double banana confirms Gross and Sullivant's clause:

Proposition 6.4. *The double banana is hereditarily admissible at* $n = 4$*: every subset* S *of its eight vertices with* $|S| > 4$ *satisfies* $\dim L_S \leq \beta(|S|, 4)$.

Proof. A finite check over all $2^8 = 256$ subsets. □

So the coloured K_4 of Theorem 3.3 plays, on four vertices, the role that the double banana plays on eight; and by Section 7 no coloured graph on fewer than four vertices refutes even the unstrengthened conjecture, so four is optimal for Conjecture 6.2 as well. What the two witnesses have in common is that their obstruction is not a clique; what they do not have in common is the mechanism, and Section 9 asks whether there is one.

Remark 6.5. The implication "a violated hereditary count implies not n-estimable" is Proposition 1.1 applied to a sub-model; it is a dimension count and is not part of the formal development. What is formalised is the count itself, which is decidable, and the estimability verdicts, which are proved. Theorem 6.3 does not use that implication in either direction: it asserts a numerical inequality for one model and its failure for three others.

7. Minimality: an exhaustive census on at most four vertices

This section is the one place where a claim of this paper rests on exhaustive computation, and it is not part of the formal development. We state it as such.

Proposition 7.1 (computational)**.** *Over all coloured graphs on* $p \leq 4$ *labelled vertices — all* $2^{\binom{p}{2}}$ *graphs, all set partitions of the vertex set, all set partitions of the edge set, and all* n *with* $1 \leq n < p$ *— there is no counterexample to the Forbes–Lauritzen conjecture with* $p \leq 3$*, and at* $p = 4$ *the least value of* $|\mathcal{N}| + |\mathcal{E}|$ *among counterexamples is* 4*, attained exactly at* $n = 1$.

What was run. For each admissible tuple $(G, \mathcal{N}, \mathcal{E}, n)$ with $|\mathcal{N}| + |\mathcal{E}| \leq \beta(p, n)$, the verdict was decided by computing the rank of the design map $K \mapsto KX$ of Lemma 2.2 at up to four random integer samples X, over the prime field $\mathbb{F}_q$ with $q = 2^{61} - 1$. One pruning step was used, and it is itself a certificate: if the finest (uncoloured) colouring of G is n-estimable then so is every coarsening of it, because $S(\mathcal{N}, \mathcal{E})$ is then a subspace of an n-estimable space — this is Forbes–Lauritzen's subspace lemma, [3, Lemma 2]: *if* L *is* n*-estimable and* $L_0 \subseteq L$*, then* L_0 *is* n*-estimable.* The counts are

p	models tested	counterexamples	least $\lvert\mathcal{N}\rvert + \lvert\mathcal{E}\rvert$
2	1	0	—
3	58	0	—
4	17,783	20	4 (at $n = 1$)

where "models tested" counts the tuples surviving the dimension filter and the pruning step. The $p = 4$ pass takes 5.4 seconds on one core. Of the twenty counterexamples at $p = 4$, ten have $n = 1$ and $|\mathcal{N}| + |\mathcal{E}| = 4$ and fall into two shapes — six labelled copies of the model of Theorem 3.1 and four of the model of Theorem 3.3 — and ten have $n = 2$ with $|\mathcal{N}| + |\mathcal{E}| \geq 7$. The census was then repeated with an independent enumeration, over graphs up to isomorphism (11 graphs on four vertices) and with the number of edge classes capped at $\beta(p,n) - |\mathcal{N}|$: 40 models at $p = 3$ and 8,029 at $p = 4$, giving 0 and 12 counterexamples, and agreeing with the first pass on both the emptiness at $p \leq 3$ and the least $|\mathcal{N}| + |\mathcal{E}|$ at $p = 4$ (4 at $n = 1$, 7 at $n = 2$). Separately, the local/non-local test of Section 6 was run over all 2^p subsets for each of the twenty counterexamples: exactly ten are non-local, and all ten are colourings of K_4 — four labelled copies at $n = 1$, which form the single isomorphism type of Theorem 3.3, and six at $n = 2$, which form one type, K_4 with a merged vertex pair.

The direction of certainty. An "estimable" verdict is certain: full rank of an integer design matrix modulo q implies full rank over $\mathbb{Q}$ at the same integer sample, which is therefore a genuine witness. That is the direction Proposition 7.1 needs, since "no counterexample with $|\mathcal{N}| + |\mathcal{E}| \leq 3$" is the assertion that a list of models is estimable. A "not estimable" verdict from the census is, in contrast, a Schwartz–Zippel statement: $\det M(X)$ has degree $2d$ in the entries of X, and the dimension filter keeps $d \leq \beta(4,3) = 9$ throughout the census, so a single random point of $\mathbb{F}_q^{p\times n}$ errs with probability at most $2d/q \leq 18/(2^{61}-1) < 10^{-17}$, and up to four independent points were used. The two verdicts this paper actually claims at the minimum, Theorems 3.1 and 3.3, are not left in that state: they are proved for all samples, in the formal development. All four counterexample models named in this paper were in addition re-checked with exact integer determinants, at six random integer samples each.

From $p \leq 4$ to all p. A counterexample needs $n \leq p-2$. Indeed for $n \geq p$ take $X = [\,I_p \;\; 0\,]$: then $KX = 0$ forces $K = 0$, so the design map is injective and $\det M(X) > 0$ by Lemma 2.2; and for $n = p-1$ the conjecture is a theorem of Forbes and Lauritzen — [3, Proposition 5]: *if $L \subseteq \mathrm{Sym}_p(\mathbb{R})$ meets the positive definite cone and $\dim L \leq T_p - T_r$ with $r = p - n = 1$, then L is n-estimable* — whose hypothesis $S(\mathcal{N},\mathcal{E})$ satisfies, since $I_p = \sum_{u\in\mathcal{N}} e^u \in L$. Hence $p \geq n+2 \geq 3$, and $p = 3$ forces $n = 1$, which Proposition 7.1 clears. Combining with Theorems 3.1 and 3.3: *the smallest counterexample to the Forbes–Lauritzen conjecture has $p = 4$ vertices and $|\mathcal{N}| + |\mathcal{E}| = 4$, occurs at $n = 1$, and is necessarily coloured.* The elementary argument for $n \geq p$ just given, and the appeal to the $n = p-1$ theorem, are also outside the formal development.

8. Verification

Every result of this paper that states an estimability verdict, a hereditary count or a property of the class of coloured models — Theorems 3.1, 3.3 and 6.3, Corollary 3.2, and Propositions 2.7, 3.5, 4.1, 4.2, 5.1, 5.2, 5.3 and 6.4 — has been formally verified in Lean 4 against *Mathlib* [8, 9]; repository reference to be added at submission. Proposition 1.1 is quoted from [3] and is not formalised; of the preliminaries, the identity $M(X) = D(X)^{\mathsf{T}} D(X)$, the nonnegativity of $\det M(X)$ and the equivalence in Lemma 2.2, the kernel criterion Lemma 2.3, and both certificate shapes of Lemma 2.4 are formalised — the kernel criterion over an arbitrary commutative integral domain and the certificate shapes over an arbitrary commutative ring, the statements about $\det M(X) \geq 0$ over $\mathbb{R}$; Lemma 2.5 is formalised in the padded square form used in the proofs, and the kernel equality $\ker M = \ker D$ of Lemma 2.2 and the change-of-basis Remark 2.1 are expository and are not. The development is five modules: the framework (the class map, the indicator basis, $M(X)$, the kernel criterion, the rank-one and rank-two certificate shapes, the bound β and the conjecture as a proposition), the counterexamples of Section 3 together with Proposition 5.1 and Proposition 5.2, the double banana, the positive certificates, and the hereditary counts. `#print axioms` was run on every declaration to which a statement of this paper is pinned and returns in each case only axioms from {`propext`, `Classical.choice`, `Quot.sound`}: no `sorry`, and in particular no compiled evaluation — there is no use of `native_decide`, no external solver and no stored certificate anywhere in the development. The only uses of `decide` are on finite tables of colour classes and on small integer matrix identities, and they are checked by the kernel.

Three features of the formalisation are worth stating because they are what the guarantee covers. First, the negative results are proved *sample by sample*: Lemma 2.3 is applied to an arbitrary X

and the kernel element is constructed from that X, via the adjugate identity $\mathrm{adj}(M)M = (\det M)I$ over a commutative ring. There is no genericity argument, no polynomial identity theorem and no Zariski density anywhere, so the statements "not n-estimable" are universally quantified over samples exactly as Forbes–Lauritzen's definition asks. Second, the existence of the kernel vectors — "n equations in more than n unknowns" — is discharged by Lemma 2.5 in the padded form: a square matrix with a zero row has determinant zero, hence a nonzero kernel vector. No dimension theory enters. Third, the positive certificates are checked in exact integer arithmetic and transported to $\mathbb{R}$ by the observation that the indicator basis is 0/1, so M commutes with any ring homomorphism; determinants are never expanded, an explicit integer right inverse $MB = cI$ being exhibited instead.

Outside the guarantee, and marked where they occur, are: the census of Section 7, which quantifies over 17,783 models and is a computation in exact integer and finite-field arithmetic performed by two independent enumerators, not a formal proof; the general-p observation of Remark 3.6, of which only the case $p = 4$ is formalised (Proposition 3.5); the implication of Remark 6.5, which is Forbes–Lauritzen's Proposition 1.1 applied to a sub-model; and the two statements about $n \geq p$ and $n = p - 1$ used at the end of Section 7. Quotations from [3] and [4] are, of course, readings of those papers; the mathematical content we attribute to them — Example 6.8, the double banana verdict, and the hereditary clause — has been independently recomputed here, the first two as formal proofs (Propositions 5.2 and 5.3) and the third as a kernel-checked finite count (Proposition 6.4).

9. Open questions

(1) *A criterion for coloured models.* Gross–Sullivant's theorem makes n-estimability of an uncoloured graph a matroid condition on $E(G)$. Is there a matroid-theoretic criterion for n-estimability of $S(\mathcal{N}, \mathcal{E})$ extending it? Our Theorem 6.3 constrains the answer twice over: it cannot be a condition on the underlying graph alone — the coloured K_4 of Theorem 3.3 and the uncoloured K_4 have the same graph and different verdicts at $n = 1$ — and it cannot be a hereditary counting condition either. A natural first target is a complete characterisation of 1-estimability of coloured graphs.

(2) *The shape of the non-local obstructions.* At $p = 4$ every non-local counterexample found by the census is a colouring of K_4, and the certificate of Theorem 3.3 is the "centre versus rim" vector of the colouring (Remark 3.4), which lives in the invariant subspace of the colour group. The double banana's certificate is not of that shape. Is every non-local coloured obstruction at $n = 1$ a representation-theoretic phenomenon of the colour group?

(3) *The census at $p = 5$.* Whether the uniformity at $p = 4$ — all non-local counterexamples are colourings of K_4 — persists. The corresponding enumeration is dominated by the edge partitions, Bell(10) = 115,975 per vertex partition for the densest graphs, and we did not complete it.

(4) *General n, uncoloured.* For $n \geq 2$ the graph K_{n+1} plus one isolated vertex on $p = n + 2$ vertices satisfies the count and is not n-estimable, by Remark 2.6, uniformly in n; this is Gross–Sullivant's clique obstruction and is not new, but it is not formalised here, since it needs an index type for colour classes that varies with n.

References

[1] D. I. Bernstein, S. Dewar, S. J. Gortler, A. Nixon, M. Sitharam and L. Theran, *Maximum likelihood thresholds via graph rigidity*, Ann. Appl. Probab. **34** (2024), no. 3, 3288–3319; arXiv:2108.02185v3.

[2] D. I. Bernstein, *Rigidity theory in statistical inference*, arXiv:2601.10864v1 (2026).

[3] P. G. M. Forbes and S. Lauritzen, *Linear estimating equations for exponential families with application to Gaussian linear concentration models*, Linear Algebra Appl. **473** (2015), 261–283; arXiv:1311.0662v2. All quotations in this paper, and all proposition, lemma and equation numbers attributed to it, are those of the e-print v2.

[4] E. Gross and S. Sullivant, *The maximum likelihood threshold of a graph*, Bernoulli **24** (2018), no. 1, 386–407; arXiv:1404.6989v3. Section 6, in which the results quoted here appear, is present from v3 of the e-print onwards; the numbering and the page reference are those of the published version, which agree with v3.

[5] S. Højsgaard and S. L. Lauritzen, *Graphical Gaussian models with edge and vertex symmetries*, J. R. Stat. Soc. Ser. B **70** (2008), no. 5, 1005–1027.
[6] A. Hyvärinen, *Estimation of non-normalized statistical models by score matching*, J. Mach. Learn. Res. **6** (2005), 695–709.
[7] S. T. Jensen, *Covariance hypotheses which are linear in both the covariance and the inverse covariance*, Ann. Statist. **16** (1988), no. 1, 302–322.
[8] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science, vol. 12699, Springer, 2021, pp. 625–635.
[9] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381.
[10] C. Uhler, *Geometry of maximum likelihood estimation in Gaussian graphical models*, Ann. Statist. **40** (2012), no. 1, 238–261; arXiv:1012.2643.