ALGEBRAIC COMBINATORICS

Jan Draisma

Partial correlation hypersurfaces in Gaussian graphical models

Volume 2, issue 3 (2019), p. 439-446.

<http://alco.centre-mersenne.org/item/ALCO_2019__2_3_439_0>

Algebraic Combinatorics
Volume 2, issue 3 (2019), p. 439–446
https://doi.org/10.5802/alco.44

Partial correlation hypersurfaces in Gaussian graphical models

Jan Draisma

Abstract We derive a combinatorial sufficient condition for a partial correlation hypersurface in the parameter space of a directed Gaussian graphical model to be nonsingular, and speculate on whether this condition can be used in algorithms for learning the graph. Since the condition is fulfilled in the case of a complete DAG on any number of vertices, the result implies an affirmative answer to a question raised by Lin–Uhler–Sturmfels–Bühlmann.

1. Introduction

DAGs. Let G be a directed, acyclic graph (DAG) with vertex set V and edge set $D \subseteq \{(i,j) \in V^2 \mid i \neq j\}$. We write $i \to j$ if $(i,j) \in D$ and $i \not\to j$ otherwise. A path in G from i to j of length k is a sequence $(i = i_0, i_1, \ldots, i_k = j)$ with $i_l \to i_{l+1}$ for all $l = 0, \ldots, k-1$; we allow $k = 0$. If there exists a path from i to j of length at least 1 we say that j is below i.

Directed Gaussian graphical models. We follow [2, p. 87]. Associated to G is the directed graphical model for jointly Gaussian random variables $X_i,\ i \in V$ related by

$$X_j = \sum_{i:i\to j} a_{ij} X_i + \epsilon_j$$

where the vector $\epsilon \sim \mathcal{N}(0, I)$ and where the $a_{ij} \in \mathbb{R}$ are the parameters of the model. The vector $X = (X_j)_{j\in V}$ satisfies

$$(I - A)^T X = \epsilon$$

where A is the matrix with (i,j)-entry a_{ij} if $i \to j$ and 0 otherwise. Therefore $X \sim \mathcal{N}(0, \Sigma)$ where

$$\Sigma = \Sigma(A) = (I - A)^{-T}(I - A)^{-1}.$$

Note that, since A is nilpotent, this is a matrix whose entries are polynomials in the parameters $a_{ij}, i \to j$. For subsets $I, J \subseteq V$ we write $\Sigma[I,J]$ for $I \times J$-submatrix $(\sigma_{ij})_{i\in I, j\in J}$ of Σ, and we use notation such as $I + i_0 - s := I \cup \{i_0\} \setminus \{s\}$.

Manuscript received 13th June 2018, revised 22nd September 2018, accepted 24th September 2018.

Keywords. partial correlation, Gaussian graphical models, trek separation.

Acknowledgements. The author is partially supported by a Vici grant from the Netherlands Organisation for Scientific Research (NWO).

Partial correlation hypersurfaces. Let $i_0, j_0 \in V$ be distinct and $S \subseteq V \setminus \{i_0, j_0\}$. In [4] the *partial correlation hypersurface* $H_f \subseteq \mathbb{R}^D$ is defined as the zero locus of the polynomial

$$f := \det(\Sigma[S + i_0, S + j_0]);$$

the expression $\operatorname{corr}(i_0, j_0|S) := f/\sqrt{\det(\Sigma[S + i_0, S + i_0]) \det(\Sigma[S + j_0, S + j_0])}$ is the partial correlation of i_0 and j_0 given S.

So the vanishing of f is equivalent to the statement that i_0, j_0 are conditionally independent given S. We assume that f is not identically zero on $\mathbb{R}^D$. This is equivalent to the statement that S does not *d-separate* i_0 and j_0 in G [5, Section 2.3.4]; the trek system expansion of Section 2 yields an equivalent combinatorial characterisation.

The key motivation in [4] for studying H_f is that the behaviour for $\lambda \to 0$ of the volume (relative to some probability measure) of

$$\operatorname{Tube}(\lambda) := \{a \in \mathbb{R}^D : |\operatorname{corr}(i_0, j_0, S)| \leqslant \lambda\}$$

is related to the singularities of H_f. This volume scales linearly with λ if H_f is nonsingular but can be superlinear otherwise—whence the study of the real log-canonical threshold of H_f in [4]. The parameter values a in $\operatorname{Tube}(\lambda)$ correspond to probability distributions that are not λ-*strongly-faithful* to G—distributions where the PC algorithm for learning G might fail. So it is useful to know criteria for nonsingularity of H_f.

Main result. We will establish the following criterion for nonsingularity of H_f; the same applies when the ϵ_i have unequal variances (Proposition 2.5).

Theorem 1.1. *Assume that $i_0 \to j_0$ and that for all $s \in S$ below j_0 we have $i_0 \to s$. Then H_f is nonsingular.*

Corollary 1.2. *If G is the DAG on $\{1, \ldots, n\}$ with $i \to j$ if and only if $i < j$, then H_f is nonsingular, independently of the choice of i_0, j_0, S.*

For $n \leqslant 6$ this is [4, Theorem 4.1], which was established there by extensive computer calculations showing that some power of $\det \Sigma[S + i_0 + j_0, S + i_0 + j_0]$ lies in the ideal generated by f and its partial derivatives. Since $\Sigma[S + i_0 + j_0, S + i_0 + j_0]$ is positive definite and hence has a nonzero determinant for all (real) values of the parameters, this shows that the (real) common vanishing locus of f and its derivatives is empty.

We will follow a similar approach, except that we consider the principal submatrix $\det \Sigma[S + i_0, S + i_0]$, no power is needed, and indeed not f but only some of its partial derivatives are needed.

Organisation. In Section 2 we review the expansion of subdeterminants of Σ in terms of trek systems without sided intersection [6]. In Section 3 we use this to prove the theorem, and we conclude with a brief discussion in Section 4.

2. Background

The trek rule. We recall results from [6]. Suppose we allow the variances of the ϵ_i to be distinct, rather than all equal to 1 as above. In that case, the covariance matrix Σ becomes

$$\Sigma = (I - A)^{-T} \Omega (I - A)^{-1}$$

where Ω is the diagonal matrix with the covariances of the ϵ_i on the diagonal. Using the geometric series for $(I - A)^{-1}$ we find that

$$\sigma_{ij} = \sum_{t: i \to j} w(t)$$

where the sum is over all *treks* from i to j as in the following definition.

Definition 2.1. *A* trek t *in* G *is a pair* (P_U, P_D) *of paths in* G *that start at the same vertex* m*, the* top *of the trek. The paths* P_U, P_D *are called the* up part *and the* down part *of* t*, respectively. If* i_0 *is the last vertex of* P_U *and* j_0 *is the last vertex of* P_D*, then we call* t *is a trek* from i_0 to j_0*,* i_0 *the* starting *vertex of* t*, and* j_0 *the* end *vertex of* t*. The* weight *of* t *equals*

$$w(t) := \left(\prod_{(i,j) \text{ in } P_U} a_{ij}\right) \cdot \omega_m \cdot \left(\prod_{(i,j) \text{ in } P_D} a_{ij}\right).$$

We allow one or both of P_U, P_D *to have length* 0*, in which case the corresponding factor(s) above is (are)* 1*.*

The terminology derives from an informal interpretation of a trek as traversing P_U upwards from i_0 (i.e. against the direction of its edges in G) and then traversing P_D downwards to j_0. In slightly different terms, the *trek rule* above goes back at least to [7].

Trek system expansion. Equip V with an arbitrary linear order. Then for $I, J \subseteq V$ of equal cardinality and $\pi : I \to J$ we define $\operatorname{sgn}(\pi)$ as (-1) to the power the number of *crossings*: pairs $(i_1, i_2) \in I^2$ with $i_1 < i_2$ but $\pi(i_1) > \pi(i_2)$.

Definition 2.2. *Let* $I, J \subseteq V$ *with* $|I| = |J| = k$*. A* trek system T *from* I *to* J *is a set of treks* $\{t_1, \ldots, t_k\}$ *such that* I *is precisely the set of starting vertices of the* t_l *and* J *is precisely the set of end vertices of the* t_l*. We write* $T : I \to J$*. The map* $\pi : I \to J$ *that sends the starting vertex of each trek to its end vertex is a bijection, and we define the sign of* T *as* $\operatorname{sgn}(T) := \operatorname{sgn}(\pi)$*. The weight of* T *is* $w(T) := \prod_{l=1}^{k} w(t_l)$*.*

Definition 2.3. *A* sided intersection *between treks* t *and* t' *is a vertex where either the up parts of* t *and* t' *meet or the down parts of* t *and* t' *meet. We say that a trek system has no sided intersections if there is no sided intersection between any two of its treks.*

We have the following formula for subdeterminants of Σ.

Proposition 2.4 ([6]). *For* $I, J \subseteq V$ *of the same cardinality we have*

$$(*) \qquad \det \Sigma[I, J] = \sum_{T: I \to J \text{ without sided intersections}} \operatorname{sgn}(T)\, \mathrm{wt}(T).$$

The proof is an application of tail swapping as in the classical Lindström–Gessel–Viennot Lemma [3]. We will see another instance of tail swapping in Section 3. In [6] the proposition is used to give a combinatorial criterion, generalising d-separation, for the determinant to be identically zero on $\mathbb{R}^D \times \mathbb{R}^V_{>0}$. Furthermore, in [1] it is shown that the sum above is cancellation-free: if two trek systems $I \to J$ have the same weight, then they have the same sign. Moreover, it is shown there that the coefficient of each monomial is plus or minus a power of 2.

All of these results—the formula $(*)$ of course, but also the cancellation-freeness and the power-of-two phenomenon—persist when we specialise Ω to the identity matrix, as we did in Section 1 and as we do again in Section 3. Indeed, if $T : I \to J$ is a trek system without sided intersection, then the tops of the treks in T can be recovered from the specialisation of $w(T)$ as follows: m is a top if and only if either

(1) at least one a_{mj} appears in $w(T)$ and no a_{im} appears in $w(T)$; or else
(2) $m \in I \cap J$ and $w(T)$ contains no a_{mj} and no a_{im} (then some trek is $((m), (m))$).

Action by diagonal matrices. Let $d = \operatorname{diag}((d_i)_{i\in V})$ where the d_i are in $\mathbb{R}_{>0}$. Then

$$d\Sigma d = (d(I-A)^{-T}d^{-1})\cdot(d\Omega d)\cdot(d^{-1}(I-A)^{-1}d) = (I-A')^{-T}\Omega'(I-A')^{-1}$$

where $\Omega' = d\Omega d$ and where $A' = d^{-1}Ad$ has the same zero pattern as A. Hence, the group $(\mathbb{R}_{>0})^V$ acts on the parameter space $\mathbb{R}^D \times \mathbb{R}^V_{>0}$ and on the space of covariance matrices in such a manner that the map $(a,\omega) \mapsto \Sigma$ is equivariant. This implies that for any $I, J \subseteq V$ of equal cardinality the hypersurface in $\mathbb{R}^D \times \mathbb{R}^V_{>0}$ defined by $\det \Sigma[I,J] = 0$ is stable under this action.

Alternatively, this can be read off from $(*)$: scaling each a_{ij} with $d_i^{-1}d_j$ and ω_m with d_m^2, the weight of each trek from a vertex $i \in I$ to a vertex $j \in J$ gets scaled by $d_i d_j$, and therefore $\det \Sigma[I,J]$ scales with $\left(\prod_{i\in I} d_i\right)\cdot\left(\prod_{j\in J} d_j\right)$.

Define $f_\Omega := \det \Sigma[I,J]$ and let f be obtained from f_Ω by specialising Ω to the identity matrix. Let H_f be the hypersurface in $\mathbb{R}^D$ defined by f and let H_{f_Ω} be the hypersurface defined by f_Ω in $\mathbb{R}^D \times \mathbb{R}^V_{>0}$.

Proposition 2.5. *As semi-algebraic sets, H_{f_Ω} is isomorphic to $H_f \times \mathbb{R}^V_{>0}$. In particular, H_{f_Ω} is nonsingular if and only if H_f is.*

Proof. By the discussion above, the map

$$(a,d) \mapsto \left(\left(a_{ij}\cdot\frac{d_j}{d_i}\right)_{i\to j}, (d_m^2)_m\right)$$

maps $H_f \times \mathbb{R}^V_{>0}$ into H_{f_Ω}. The inverse is given by

$$(a',\omega) \mapsto \left(\left(a'_{ij}\cdot\frac{\sqrt{\omega_i}}{\sqrt{\omega_j}}\right)_{i\to j}, (\sqrt{\omega_m})_m\right).$$

Both maps are morphisms of semi-algebraic sets. □

3. Proof of the theorem

We retain the notation of Section 1; in particular, $\epsilon \sim \mathcal{N}(0,I)$, $f = \det \Sigma[S+i_0, S+j_0]$ and $H_f \subseteq \mathbb{R}^D$ is the hypersurface defined by f. In this section, we treat the a_{ij} as variables and our computations take place in the polynomial ring $\mathbb{R}[a_{ij} \mid (i,j) \in D]$. Let $\mathcal{J}$ be the ideal in this ring generated by all partial derivatives $\partial f/\partial a_{ij}$ of f.

Lemma 3.1. *For $s \in S$ and $j \in V$ with $s \to j$ the variable a_{sj} does not appear in f.*

Proof. Let $T : S + i_0 \to S + j_0$ be a trek system without sided intersection. If the arrow $s \to j$ were used in the up (respectively, down) part of some trek t in T, then t would have a sided intersection with the trek starting (respectively, ending) at s. So that arrow is not used and the conclusion follows from $(*)$. □

As a consequence, in the remaining discussion we may and will replace D by $D \setminus S \times V$, so that *G has no arrows going out of S*.

Lemma 3.2. *Suppose that G has no outgoing arrows from elements of S. For $s \in S$ with $i_0 \to s$ the variable $a_{i_0 s}$ appears at most linearly in f and its coefficient equals $\pm\det \Sigma[S+i_0, S+j_0-s+i_0]$. In particular, $\det \Sigma[S+i_0, S+j_0-s+i_0] \in \mathcal{J}$.*

Proof. If a trek t in a trek system $T : S+i_0 \to S+j_0$ without sided intersection uses the edge $i_0 \to s$, then it does so in its down part—indeed, in its up part it would yield a sided intersection with the trek starting at i_0. In particular, the variable $a_{i_0 s}$ appears only linearly in f. Furthermore, t ends in s, or else t would have a sided intersection with the trek ending at s.

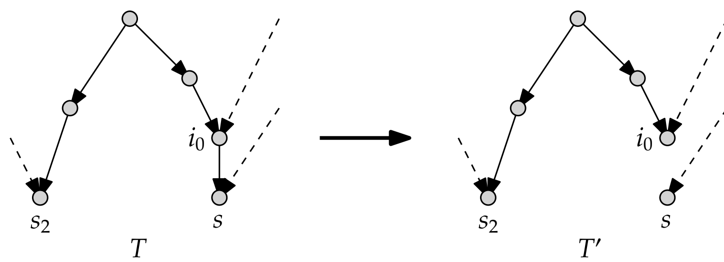

FIGURE 1. Proof of Lemma 3.2. We suggestively draw the arrows in up parts of treks as pointing in the south-west direction and arrows in down parts as pointing in the south-east direction—of course, this is not always possible!

So if we remove from t the arrow $i_0 \to s$, then we obtain a trek system $T' : S+i_0 \to S+j_0-s+i_0$ without sided intersection (Figure 1).

Conversely, if we have any trek system $T' : S+i_0 \to S+j_0-s+i_0$ without sided intersection, then no trek in it passes through s on its way down, because s has no outgoing arrows. Hence, adding the arrow $i_0 \to s$ to the trek t' in T' ending in i_0 yields a trek system $S+i_0 \to S+j_0$ without sided intersection.

Hence the map $T \mapsto T'$ gives a bijection between the terms in (the trek system expansion of) f divisible by $a_{i_0 s}$ and the terms in $\det \Sigma[S+i_0, S+j_0+i_0-s]$. Furthermore, $\operatorname{sgn}(T)$ equals $\pm \operatorname{sgn}(T')$, where the sign is the sign of the bijection $S+j_0-s+i_0 \to S+j_0$ that is the identity on $S+j_0-s$ and sends i_0 to s; in particular, this sign does not depend on T. $\square$

LEMMA 3.3. *Assume that $i_0 \to j_0$. The variable $a_{i_0 j_0}$ appears at most linearly in f and its coefficient equals $\pm(\det \Sigma[S+i_0, S+i_0]-g)$ where*

$$g = \sum_{T'' : S+i_0 \to S+i_0} \operatorname{sgn}(T'') w(T'') \tag{$**$}$$

is the sum over all trek systems $T'' : S+i_0 \to S+i_0$ without sided intersection of which one trek contains j_0 in its down part. *In particular,* $\det \Sigma[S+i_0, S+i_0]-g \in \mathcal{J}$.

Proof. If a trek t in a trek system $T : S+i_0 \to S+j_0$ without sided intersection uses the edge $i_0 \to j_0$, then it does so on its way down: on its way up it would yield a sided intersection with the trek starting at i_0. In particular, the variable $a_{i_0 j_0}$ appears only linearly in f.

Furthermore, t ends in j_0, or else it would have a sided intersection with the trek ending at j_0. So if we remove from t the arrow $i_0 \to j_0$, then we obtain a trek system $T'' : S+i_0 \to S+i_0$ without sided intersection (Figure 2). Also, $\operatorname{sgn}(T)$ equals $\operatorname{sgn}(T'')$ times the sign of the bijection $S+i_0 \to S+j_0$ that is the identity on S and maps i_0 to j_0; this will determines the sign $\pm$ in the lemma.

Conversely, given a trek system $T'' : S+i_0 \to S+i_0$ without sided intersection, we may try and add the arrow $i_0 \to j_0$ to the trek ending in i_0. The resulting trek system has no sided intersection if and only if no trek of T'' passes j_0 on its way down. The remaining T'' must be therefore be subtracted as in the lemma. $\square$

For $s \in S$ below j_0 define $p_{j_0,s} := \sum_{P : j_0 \to s} w(P)$, the sum of the weights of all directed paths in G from j_0 to s.

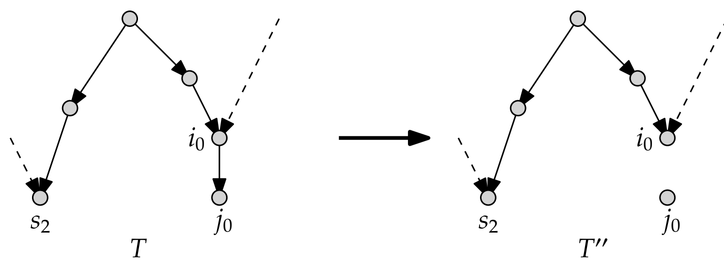

Figure 2. Proof of Lemma 3.3.

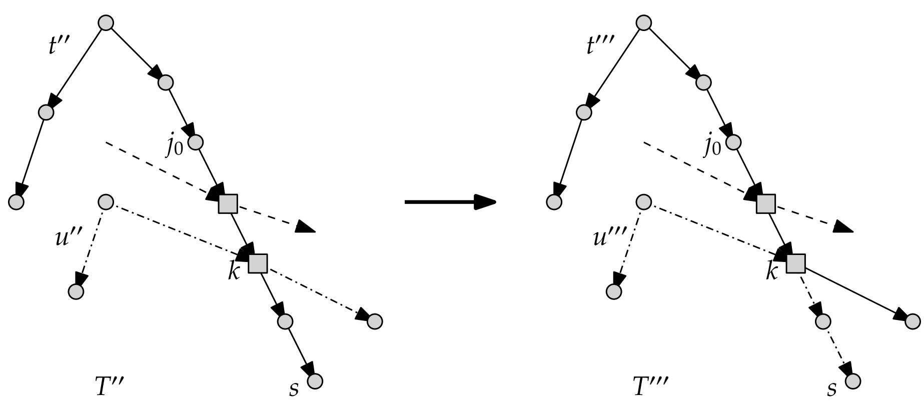

Figure 3. The tail swapping argument of Lemma 3.4. The sided intersections of t'' with other treks are depicted as square vertices.

Lemma 3.4. *The element g from $(**)$ satisfies*

$$g = \sum_{s \in S \text{ under } j_0} \operatorname{sgn}(\pi_s) \det \Sigma[S + i_0, S + i_0 - s + j_0] \cdot p_{j_0,s}$$

where $\pi_s : S + i_0 - s + j_0 \to S + i_0$ is the identity on $S + i_0 - s$ and sends j_0 to s.

Proof. Let $T' : S + i_0 \to S + i_0 - s + j_0$ be a trek system without sided intersection and let t' be the trek of T' ending in j_0. Appending to t' any path from j_0 down to s yields a trek system $T'' : S + i_0 \to S + i_0$ with sign $\operatorname{sgn}(T'') = \operatorname{sgn}(T') \operatorname{sgn}(\pi_s)$. In this manner, precisely those trek systems $T'' : S + i_0 \to S + i_0$ arise for which

(1) a unique trek t'' of T'' passes j_0 on its way down, and
(2) every sided intersection of T'' is between t'' and some other trek of T'' on their way down, and happens at a vertex below j_0.

So the left-hand side of the equation in the lemma equals $\sum_{T'' : S+i_0 \to S+i_0} \operatorname{sgn}(T'') \times w(T'')$ where W'' runs over the trek systems with properties (1) and (2). The right-hand side is the sub-sum over all T'' without any sided intersection. We construct a sign-changing involution on the remaining T'', as follows.

Let k be the *lowest* vertex on the down part of t'' that lies on the down part of some other trek $u'' \neq t''$ of T''. Swapping the parts of t'' and u'' below k yields treks t''' and u''' that still meet at k. Let T''' be the trek system obtained from T'' by replacing t'' with t''' and u'' with u''' (Figure 3).

The trek system T''' satisfies (1): t''' is its unique trek that passes j_0 on its way down. As for (2): the sided intersections between t''' and other treks are precisely the

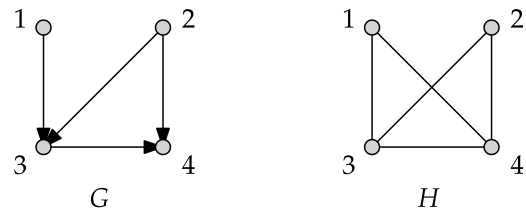

Figure 4. The graphs in Example 4.1.

sided intersections between t'' and other treks, so they happen below j_0. Also, u''' cannot have sided intersections with treks other than t''', because those would have come from a sided intersection between t''' and another trek happening below k—this is where the choice of k matters. Furthermore, $T''' \setminus \{t''', u'''\} = T'' \setminus \{t'', u''\}$, so there are no sided intersections between these treks. This shows that T''' satisfies (2). Also, the map $T'' \to T'''$ is an involution, since k is the last intersection of the down part of t''' with any down part of a trek in T'''. Since $\operatorname{sgn}(T''') = -\operatorname{sgn}(T'')$, this shows that the terms on the left-hand side that do not appear in the right-hand side cancel out. $\square$

Proof of the theorem. We claim that the zero set of $\mathcal{J}$ in $\mathbb{R}^D$ is empty. By Lemma 3.1 we may delete from G all outgoing arrows from elements of S without changing f. Since $i_0 \to j_0$, by Lemma 3.3 we have $\det \Sigma[S + i_0, S + i_0] - g \in \mathcal{J}$. The identity in Lemma 3.4 expresses g as a linear combination of the determinants in Lemma 3.2 where s runs over the elements of S below j_0. By assumption, for each of these s we have $i_0 \to s$, so Lemma 3.2 implies that $g \in \mathcal{J}$. Hence $\det \Sigma[S + i_0, S + i_0] \in \mathcal{J}$. But for any set of real parameters $a \in \mathbb{R}^D$ the matrix $\Sigma[S + i_0, S + i_0]$ is positive definite, hence has a nonzero determinant. This proves the claim. $\square$

4. A modest implication for the PC algorithm

In the edge-removal part of the PC algorithm [5] for learning G, in each step we have an undirected graph H whose edge set, if no error has occurred so far, contains that of G. Using the sample covariance matrix, a partial correlation $\operatorname{corr}(i_0, j_0|S)$ is then computed for some triple i_0, j_0, S such that there is an edge $i_0 - j_0$ in H and such that S is contained in the H-neighbours of i_0 or in the H-neighbours of j_0. Before this step all partial correlations with sets S' of cardinality smaller than that of S have already been checked. If the absolute value of the partial correlation is less than some prescribed λ, then the edge $i_0 - j_0$ is removed from H.

Our theorem suggests that it might be advantageous to perform this check first for sets S contained in the *intersection* of the neighbourhoods of i_0 and j_0 in H. Then, *if all the edges between i_0, j_0, S present in H are also present in the DAG G* (with some orientation), one readily checks that the conditions of the theorem are satisfied. Hence the volume of $\operatorname{Tube}(\lambda)$ is proportional to λ, and the region in the parameter space $\mathbb{R}^D$ of G where we would erroneously delete $i_0 - j_0$ in this step is small.

There are two obvious issues with this. First, in general it will not suffice to check S in the intersection of the neighbourhoods of i_0 and j_0. And second, the condition that all of those edges are indeed present in G is rather strong. To make better use of our theorem, one might want to develop a version of the PC algorithm where orientation steps are intertwined with the edge-deletion steps.

We conclude with two examples.

Example 4.1. Singular partial correlation hypersurfaces cannot be avoided in the edge removal step of the PC algorithm: let G be as in Figure 4, taken from [4, Example 4.8]. In the beginning, the PC algorithm finds all nonconditional independencies (so with $S = \varnothing$), and hence removes the edge $1-2$ to arrive at the graph H on the right. If the algorithm next chooses to consider the edge $1 - 4$, then it will delete this edge after finding that $1, 4$ are independent given 3. However, by symmetry of H it is equally likely that it will first consider the edge $1 - 3$. In [4] it is shown that the partial correlation f with $i_0 = 1, j_0 = 3$ and $S = \{4\}$ has a singular hypersurface $H_f \subseteq \mathbb{R}^D$ and that the corresponding $\mathrm{Tube}(\lambda)$ of bad parameter values is fatter.

Example 4.2. The paper [4] also discusses mathematical interpretations of existing heuristics in statistics. In particular, [4, Problem 6.2] discusses a volume inequality that would confirm the belief that *"collider-stratification bias tends to attenuate when it arises from more extended paths"*.

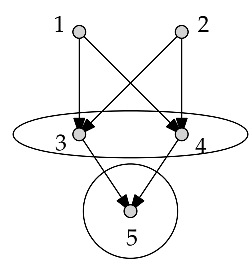

Figure 5. The graph from Example 4.2.

For Figure 5 their conjecture says that

$$\mathrm{Vol}(\{\lambda : |\mathrm{corr}(1, 2|5)| \leqslant \lambda\}) \geqslant \mathrm{Vol}\{\lambda : |\mathrm{corr}(1, 2|3, 4)| \leqslant \lambda\}.$$

The paper does not explicitly say with respect to which measure Vol is defined. If this were true for all measures, then $\mathrm{corr}(1, 2|5) \leqslant \mathrm{corr}(1, 2|3, 4)$. This is certainly not true in general: taking $a_{13} = -3, a_{14} = -2, a_{23} = 8, a_{24} = 10, a_{35} = 2, a_{45} = 0$ yields $\mathrm{corr}(1, 2|5)^2 = 1024/1189 > 88/105 = \mathrm{corr}(1, 2|3, 4)^2$. So formulating this statistical belief as a precise mathematical conjecture remains a challenge.

References

[1] Jan Draisma, Seth Sullivant, and Kelli Talaska, *Positivity for Gaussian graphical models*, Adv. Appl. Math. **50** (2013), no. 5, 661–674.

[2] Mathias Drton, Bernd Sturmfels, and Seth Sullivant, *Lectures on algebraic statistics*, Oberwolfach Seminars, vol. 39, Birkhäuser, 2009.

[3] Ira Gessel and Gérard Viennot, *Binomial determinants, paths, and hook length formulae*, Adv. Math. **58** (1985), 300–321.

[4] Shaowei Lin, Caroline Uhler, Bernd Sturmfels, and Peter Bühlmann, *Hypersurfaces and their singularities in partial correlation testing*, Found. Comput. Math. **14** (2014), no. 5, 1079–1116.

[5] Peter Spirtes, Clark Glymour, and Richard Scheines, *Causation, prediction, and search. With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper and Thomas Richardson.*, 2nd ed., MIT Press, 2001.

[6] Seth Sullivant, Kelli Talaska, and Jan Draisma, *Trek separation for Gaussian graphical models*, Ann. Stat. **38** (2010), no. 3, 1665–1685.

[7] Sewall Wright, *The method of path coefficients*, Ann. Math. Stat. **5** (1934), 161–215.

Jan Draisma, Universität Bern, Mathematisches Institut, Sidlerstrasse 5, 3012 Bern (Switzerland), and Eindhoven University of Technology, Department of Mathematics and Computer Science, P.O. Box 513, 5600 MB Eindhoven (The Netherlands)
E-mail : jan.draisma@math.unibe.ch