Algebraic Methods of Classifying Directed Graphical Models

Hajir Roozbehani[a], Yury Polyanskiy[b]

[a]*Department of Aeronautics and Astronautics at MIT. e-mail:* hajir@mit.edu
[b]*Department of Electrical Engineering and Computer Science,MIT, Cambridge, MA 02139 USA. e-mail:* yp@mit.edu

Abstract

Directed acyclic graphical models (DAGs) are often used to describe common structural properties in a family of probability distributions. This paper addresses the question of classifying DAGs up to an isomorphism. By considering Gaussian densities, the question reduces to verifying equality of certain algebraic varieties. A question of computing equations for these varieties has been previously raised in the literature. Here it is shown that the most natural method adds spurious components with singular principal minors, proving a conjecture of Sullivant. This characterization is used to establish an algebraic criterion for isomorphism, and to provide a randomized algorithm for checking that criterion. Results are applied to produce a list of the isomorphism classes of tree models on 4,5, and 6 nodes. Finally, some evidence is provided to show that projectivized DAG varieties contain useful information in the sense that their relative embedding is closely related to efficient inference.

1. Introduction

Consider two directed graphical models (or directed acyclic graphs, DAGs) on random variables (A, B, C):

$$A \to B \to C \qquad B \leftarrow A \to C \tag{1}$$

(See [1] for background on graphical models.) In this paper, we will say that these two models are *isomorphic* (as graphical models). Roughly, this means that after relabeling ($A \leftrightarrow B$), the two resulting models describe the same collection of joint distributions $P_{A,B,C}$. Note that the so defined isomorphism

The research was supported by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370 and the NSF CAREER award under grant agreement CCF-12-53205.

notion is weaker than the (directed) graph isomorphism: the graphs in (1) are not isomorphic.

On the other hand, there does not exist any relabeling making (1) equivalent to

$$B \to C \leftarrow A \tag{2}$$

In fact, a simple exercise in d-separation criterion shows that (1) and (2) list all possible isomorphism classes of directed *tree* models on three variables. However, note that the above DAGs are all isomorphic as undirected graphs.

The goal of this paper is to provide (computational) answer to: *What are the isomorphism classes of directed graphical models on n nodes?*

Note that when variables (A, B, C) are jointly Gaussian and zero-mean, then conditions such as (1) can be stated as algebraic constraints on the covariance matrix:

$$\mathbb{E}[AB]\mathbb{E}[BC] = \mathbb{E}[AC]\mathbb{E}[B^2]. \tag{3}$$

This suggests that checking isomorphism of models can be carried out via algebraic methods. Indeed, one needs to recall (see [2]) that graphical models equality can be tested by restricting to Gaussian random variables.

In this paper, we associate with every DAG two subsets of covariance matrices:

- all non-singular covariance matrices satisfying DAG constraints (denoted $\mathrm{loc}(G) \cap \Sigma^{++}$ below)
- all covariance matrices satisfying DAG constraints (denoted $\mathrm{loc}(G)$ below)

We give an *analytic* result: while $\mathrm{loc}(G)$ is not necessarily (Euclidean) closed, closures of both sets coincide.

Next, we switch to the *algebraic* part. Due to the analytic fact above, much simpler equations for non-singular matrices can be used to completely characterize the Zariski closure of $\mathrm{loc}(G)$ (denoted X_G below). Aesthetically pleasing is the fact that X_G is always an irreducible complex variety (affine and rational). Furthermore, two graphical models G and G' define the same set of conditional independence constraints if and only if $X_G = X_{G'}$.

For large graphs it is important to reduce the number of equations needed to describe X_G. The natural set of equations (denoted I_G below) turns out to be too small: its solution set $V(I_G)$ contains X_G and a number of spurious components. We show how to get rid of these spurious components, proving that

$$X_G = V((I_G : \theta_0^m)),$$

where θ_0 is an explicit polynomial (and establishing Conjecture 3.3 of Sullivant [3]). This provides a convenient method for computing X_G. After these preparations, we give our main result: isomorphism question $G \stackrel{?}{\sim} G'$ is equivalent to comparing intersections of X_G and $X_{G'}$ with a certain invariant variety. We give a randomized algorithm for this and apply it to provide a list of isomorphism classes on 4,5, and 6 nodes.

The question of DAG isomorphism does not seem to have appeared elsewhere, though the closely related question of DAG equivalence (or Markov equivalence [4]) is well-studied. As mentioned in [5], the natural space to work with when doing model selection or averaging over DAGs is that of their equivalence classes. In practice, the number of DAGs in an equivalence class encountered during model selection can be large. Some examples are presented in Table 4 of [5] where, for instance, the learning algorithm discovers equivalence classes on 402 nodes with more than 7×10^{21} members. However, one should keep in mind that not all equivalence classes are as large. In fact, Steinsky [6] showed, by a recursive method, that roughly 7 percent of equivalence classes of graphs with 500 nodes or less consist only of one element. Recent results [7] indicate that this ratio is valid asymptotically as well. Such classes appear to be the most common type of equivalences classes of DAGs [8]. In general, it is expected, based on observations on small graphs, that the ratio of equivalence classes to DAGs be around 0.27 [8][9]. Nevertheless, some equivalence classes (such as those that appear in [5]) are quite large. This has motivated the need to represent DAGs, and among the representatives that are relevant in this regard are the essential graphs[1] [4] and the characteristic imsets [12]. Both these methods have a combinatorial flavor and this work provides an algebraic alternative. The word algebraic here means commutative-algebraic, unlike in [12]. We remark that the two mentioned methods can also be applied to solve the isomorphism problem. For instance, using the results in [4] one can reduce DAG isomorphism to the isomorphism of certain directed multi-graphs. In fact, this gives a sense of the inherent computational difficulty involved in working with the isomorphism class of a DAG.

While the notion of Markov equivalence makes sense in the setting of [5], there are situations where it is natural to want to work with the isomorphism class of a DAG– the Markov equivalence class modulo permutations of variables. For instance, the recent results in [13] imply that there is a precise sense in which the isomorphism classes of all large graphs that admit efficient inference are related to graphs that look like the (unlabeled) trees listed in Figure 1, but are far from the complete DAGs. An exact description of such models, however, is problematic by the subsequent results in [13]. It thus appears reasonable to find good ways to approximate them, and for that we resort to the family of projective varieties.

It is also important to mention that the idea of associating an algebraic variety to a conditional independence (CI) model has been previously explored in a number of publications, among which we will discuss [14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. Some of our preparatory propositions can be found in the literature in slightly weaker forms and we attempt to give references. The main novelties are:

[1]The essential graphs were originally known as completed patterns and were introduced in [10] as the maximal invariants associated to equivalence classes of DAGs. They were also studied in [11] under the name maximally oriented graphs.

- We essentially leverage the directed-graph structure of the model (as opposed to general CI models) to infer stronger algebraic claims. In particular, our treatment is base independent – although for readability we present results for the varieties over $\mathbb{C}$.
- We present a computational procedure for answering the isomorphism question.

1.1. Preliminaries

Directed acyclic graphical models are constraints imposed on a set of probability distributions:

Definition 1. *A directed acyclic graphical model (DAG) G/k is the data:*

- *A set of indices $[n] := \{1, \cdots, n\}$ that are nodes of a directed acyclic graph. We frequently assume the nodes to be topologically sorted, i.e., $i < j$ whenever there is a path in the graph from i to j.*
- *A list $\mathcal{M}_G$ of* imposed *(a.k.a local Markov) relations*

$$i \perp\!\!\!\perp \mathbf{nd}(i) | \mathbf{pa}(i) \tag{4}$$

  where $\mathbf{pa}(i)$ denotes the set of parents of $i \in [n]$ and $\mathbf{nd}(i)$ is the set of non-descendants of i in the directed graph.

- *A subset $\mathcal{M}_G^{topo}$ of topologically sorted local Markov relations:*

$$i \perp\!\!\!\perp \mathbf{nd}(i) \cap \{j : j < i\} | \mathbf{pa}(i) \tag{5}$$

- *A set of G-*compatible *joint probability distributions*

$$\mathrm{Loc}(G) := \{P_X |\, I \perp\!\!\!\perp J | K \in \mathcal{M}_G \Rightarrow X_I \perp\!\!\!\perp X_J | X_K\},$$

  where X is a k^n-valued random variable[2]*.*

- *A set of* implied *relations*

$$\mathcal{C}_G := \cap_{P_X \in \mathrm{Loc}(G)} \{I \perp\!\!\!\perp J | K \text{ s.t } X_I \perp\!\!\!\perp X_J | X_K\}.$$

Given a collection of such models, it is often of interest to find representatives for their isomorphism classes (see also [14, 15])– these are models that have the same compatible distributions modulo labelings of variables:

[2]We mostly work with $k = \mathbb{R}$ or $\mathbb{C}$. Since $\mathbb{R}$ and $\mathbb{C}$ are measurably isomorphic, it does not matter which one we pick. We write G/k if we need to emphasize the base field k.

Definition 2. *Let $\mathcal{Q}$ be a permutation invariant family of distributions. Two DAGs G, G' are called $\mathcal{Q}$-equivalent if*

$$\mathrm{Loc}(G) \cap \mathcal{Q} = \mathrm{Loc}(G') \cap \mathcal{Q}.$$

When $\mathcal{Q}$ is the set of all distributions, we call such models equivalent. Likewise, two DAGs G, G' are called $\mathcal{Q}$-isomorphic if

$$p_{X_1 \cdots X_n} \in \mathrm{Loc}(G) \cap \mathcal{Q} \iff p_{X_{\pi(1)} \cdots X_{\pi(n)}} \in \mathrm{Loc}(G') \cap \mathcal{Q}$$

for some permutation π of indices. When $\mathcal{Q}$ is the set of all distributions, we call such models isomorphic and denote the isomorphism class of G by $[G]$.

We shall mainly focus on characterizing isomorphism classes of DAGs. A related question is that of understanding the structure of conditional independence constraints – see for the case of discrete random variables [16, 17, 15], positive discrete random variables [18], non-singular Gaussians [14], and general Gaussians [19].

Let $H = H_1 \times \cdots \times H_n$ be a product measure space endowed with the σ-algebra $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$. We assume that H_i is measurably isomorphic to $\mathbb{R}$ and that $\mathcal{H}_i$ is a Borel σ-algebra for all i. The next property, factorization, relies on a digraph structure and pertains only to DAGs:

Definition 3. *A probability measure P defined on $(H, \mathcal{H})$ is said to factorize w.r.t a DAG if it can be written as*

$$P(\mathcal{A}) = \int_{\mathcal{A}} \prod_{i}^{n} K_{i|\mathrm{pa}(i)}(dx_i | x_{\mathrm{pa}(i)}) \quad \forall \mathcal{A} \in \mathcal{H},$$

where $K_{i|\mathrm{pa}(i)}$'s are conditional probability kernels (which exist by [24, Theorem 2.7]) and $K_{i|\mathrm{pa}(i)}(dx_i|x_{\mathrm{pa}(i)}) = \mu_i(dx_i)$ if i has no parents in G.

Given a DAG G, we denote by $\mathrm{Fac}(G)$ the set of distributions that factorize w.r.t G. It is known (c.f. Section II.A) that

$$\mathrm{Fac}(G) = \mathrm{Loc}(G).$$

This means that two DAGs are equal (isomorphic) in the above sense if and only if they factorize the same set of distributions (modulo the labeling of the variables).

1.2. Notation

- $\mathcal{N}$ is the set of real valued Gaussians
- $\mathcal{N}^+$ is the non-singular subset of $\mathcal{N}$.
- $\Sigma = [\sigma_{ij}]$ is the affine space $\mathbb{C}^{\binom{n+1}{2}}$ of Hermitian $n \times n$ matrices.
- Σ^+ is the positive semi-definite (PSD) subset of Σ.

- Σ^{++} is the positive definite (PD) subset of Σ.
- $\Sigma^{\bullet}$ is the subset of matrices in Σ with non-zero principal minors[3].
- $\hat{\Sigma}$ is the subset of Σ consisting of matrices with ones along the diagonal. We also set $\hat{\Sigma}^{+} \triangleq \hat{\Sigma} \cap \Sigma^{+}$, $\hat{\Sigma}^{++} \triangleq \hat{\Sigma} \cap \Sigma^{++}$, and $\hat{\Sigma}^{\bullet} \triangleq \hat{\Sigma} \cap \Sigma^{\bullet}$.
- $\mathrm{loc}(G)$ is the set of covariance matrices in $\mathrm{Loc}(G) \cap \mathcal{N}$.
- f_G is the rational parametrization defined in II.B.
- Given $S \subset \Sigma$, $[S]$ and $[S]_Z$ are its standard and Zariski closures[4], respectively.
- Given $S \subset \Sigma$, $I(S)$ is the ideal of polynomials that vanish on $[S]_Z$.
- Given an ideal I, the associated algebraic set is given by

$$V(I) = \{x \in \mathbb{C}^n | f(x) = 0 \quad \forall f \in I\}.$$

- $X_G \triangleq [\mathrm{loc}(G)]_Z$, $\mathfrak{p}_G \triangleq I(X_G)$, $\hat{X}_G \triangleq [\mathrm{loc}(G) \cap \hat{\Sigma}]_Z$.
- Y_G is the closure of $\hat{X}_G$ inside $\mathbb{P}^{\binom{n}{2}}$.

1.3. Overview of main results

Our main purpose is to show that the computational tools in algebra are relevant for addressing the following problem:

Problem 1. *Given two DAGs, determine if they are isomorphic.*

Our starting point is to show that isomorphism and $\mathcal{N}^{+}$-isomorphism are equivalent for DAGs (see Section II.C). It is well known that checking $\mathcal{N}^{+}$-equivalence reduces to checking equality of algebraic subsets inside the positive definite cone (see for instance [20, 21, 22, 3]). This follows from the next proposition:

Proposition 1 (Lemma 2.8 in [25])**.** *Let $X \sim N(\mu, \sigma)$ be an m-dimensional Gaussian vector and $A, B, C \subset [m]$ be pairwise disjoint index sets. Then $X_A \perp\!\!\!\perp X_B | X_C$ if and only if the submatrix $\sigma_{AC,BC}$ has rank equal to the rank of σ_{CC}. Moreover, $X_A \perp\!\!\!\perp X_B | X_C$ if and only if $X_a \perp\!\!\!\perp X_b | X_C$ for all $a \in A$ and $b \in B$.*

Remark 1. *Note that the rank constraint is equivalent to vanishing of the minor $|\sigma_{AC',BC'}|$ for a maximal $C' \subset C$ such that $X_{C'}$ is non-singular* [5]*.*

[3] Note that $\Sigma^{\bullet}$ is Zariski open, while Σ^{+}, Σ^{++} are described by inequalities.

[4] The closure is always taken inside the affine complex space.

[5] A vector random variable is said to be non-singular if its distribution admits a density w.r.t. product Lebesgue measure.

Proposition 1 enables us to think algebraically and/or geometrically when deciding Gaussian equivalence. Indeed, it states that $\mathrm{loc}(G) \cap \Sigma^{++}$ can be identified with the positive definite subset of the real solutions to the polynomial equations generated by the implied relations in G. Working with such subsets, however, is not convenient from a computational point of view. This motivates the next problem:

Problem 2. *Give an algebraic description of* $\mathrm{loc}(G)$.

Let J_G be the ideal generated by the minors $|\sigma_{iK,jK}|$ of the implied relations $i \perp\!\!\!\perp j | K \in \mathcal{C}_G$ inside $\mathbb{C}[\Sigma]$. Similarly, the minors of imposed relations of G generate an ideal $I_G \subset J_G$ in $\mathbb{C}[\Sigma]$. Note that this ideal coincides with that generated by the toposorted imposed relations. The corresponding ideals generated inside $\mathbb{C}[\hat{\Sigma}]$ are denoted by $\hat{I}_G, \hat{J}_G$. With the established notation, for example, Proposition 1 implies

$$V(I_G) \cap \Sigma^{++} \cap \mathbb{R}^{\binom{n+1}{2}} = \mathrm{loc}(G) \cap \Sigma^{++}. \tag{6}$$

We address the above problem by identifying X_G with an irreducible component of $V(I_G)$. It is a curious fact that the points in $V(I_G) \cap \hat{\Sigma}^{++}$ correspond to covariances of circularly symmetric Gaussians that satisfy the CI constraints of $G/\mathbb{C}$. Thus if we work with complex Gaussians, we may avoid intersecting with the reals in (6).

In Section II, we first prove some geometric results, which can be summarized in the following diagram

$$\begin{array}{ccccc} \mathrm{Im} f_G \cap \Sigma^+ = \mathrm{loc}(G) & \subset & [\mathrm{loc}(G) \cap \Sigma^{++}] & \subset & X_G \\ \cup\!\!\!| & & \cap\!\!\!| & & \| \\ \mathrm{loc}(G) \cap \Sigma^{++} & & V(I_G) & & [V(I_G) \cap \Sigma^{\bullet}]_Z \end{array}$$

The same inclusions hold if we replace (I_G, Σ) with $(\hat{I}_G, \hat{\Sigma})$, $\Sigma^{\bullet}$ with Σ^{++}, or I_G with J_G.

It is known that $[\mathrm{loc}(G) \cap \Sigma^{++}]_Z$ is a complex irreducible rational algebraic variety, cf. [3]. Here we further show that it coincides with X_G and characterize $\mathfrak{p}_G = I(X_G)$ in two different ways: as the saturated ideal of I_G at $\theta_0 = \prod_{A \subset [n]} (|\Sigma_{AA}|)$ (Conjecture 3.3 in [3]), and as the unique minimal prime of I_G contained in the maximal ideal $\mathfrak{m}_I$ at the identity. We thus have the following relations inside $\mathbb{C}[\Sigma]$:

$$\begin{aligned} I_G \subset J_G \subset S^{-1} J_G \cap \mathbb{C}[\Sigma] &= S^{-1} I_G \cap \mathbb{C}[\Sigma] \\ = I(\mathrm{loc}(G) \cap \Sigma^{++}) &= \mathfrak{p}_G \subset \mathfrak{m}_I, \end{aligned}$$

where $S = \{\theta_0^n | n > 0\}$. One can replace $(\mathrm{loc}(G), \Sigma)$ with $(\mathrm{loc}(G) \cap \hat{\Sigma}, \hat{\Sigma})$ in the above. We note that the above relations hold verbatim over $\mathbb{Z}[\Sigma]$ and other base rings. Our main statement, shown in 2.5, is that two DAGs G, G' are isomorphic if and only if

$$S^{-1} I_G \cap \mathbb{C}[\Sigma]^{\Pi} = S^{-1} I_{G'} \cap \mathbb{C}[\Sigma]^{\Pi}.$$

We use the above results to provide a randomized algorithm for testing DAG isomorphism in 2.7. Section III concerns the special case of directed tree models. In particular, we show that $\hat{I}_T$ is a prime ideal for a tree model T and hence $\hat{I}_T = I(\mathrm{loc}(T) \cap \hat{\Sigma})$. This is analogous to primality of J_T, the ideal of implied relations, shown in [3] (see Corollary 2.4 and Theorem 5.8). We thus have that two directed tree models T and T' are equal if and only if $\hat{I}_T = \hat{I}_{T'}$. Moreover, we use our randomized algorithm to list the isomorphism classes of directed tree models for $n = 4, 5$, and 6 nodes. The number of isomorphism classes of directed tree models found by our procedure is $1, 1, 2, 5, 14, 42, 142, ...$ for $n \geq 1$. Curiously, the first 6 numbers are Catalan but the 7th is not.

2. Main results

2.1. Factorization and local Markov properties

In this section we show that isomorphic DAGs factorize the same set of probability distributions modulo the labeling of the variables. A theorem of Lauritzon (see [1, Theorem 3.27]) says that a non-singular measure satisfies the local Markov property if and only if its density factorizes. Let $(H, \mathcal{H})$ be as in Definition 3. One can further state:

Proposition 2. *Let G be a DAG and P a probability measure defined on $(H, \mathcal{H})$. The following are equivalent:*

1. *P factorizes w.r.t to G.*
2. *P satisfies all imposed constraints (4) w.r.t G.*
3. *P satisfies topologically sorted constraints (5) w.r.t. G.*

In particular, $\mathrm{Fac}(G) = \mathrm{Loc}(G)$.

Proof. $1 \implies 2 \implies 3$ are obvious. We show $3 \implies 1$. The main step is to show that if $1 \perp\!\!\!\perp 3|2$, then there exists a conditional probability kernel $K_{3|2}$ such that for all $\mathcal{A} \in \mathcal{H}_3$ we have

$$\int K_{3|12}(\mathcal{A}|x_1, x_2) dP_{X_1X_2} = \int K_{3|2}(\mathcal{A}|x_1) dP_{X_1X_2}.$$

The general result then follows from this by induction. Note that the inductive step requires P to satisfy only the toposorted constraints of G. Without loss of generality, let $K_{3|12}$ and $K_{1|2}$ be regular branches of conditional probabilities. Set

$$K_{3|2}(\mathcal{A}|x_2) := \int K_{3|12}(\mathcal{A}|x_1, x_2) K_{1|2}(dx_1|x_2).$$

We want to prove that $K_{3|2}$ is a regular branch of conditional probabilities $P_{X_3|X_2}$. Clearly, $K_{3|2}(\cdot|x_2)$ is a probability measure for all x_2. Now fix $\mathcal{A} \in \mathcal{H}_3$. By [24, Theorem I.6.3], $K_{3|2}(\mathcal{A}|\cdot)$ is measurable as well, hence, it is a regular branch of conditional probabilities.

Claim: $K_{3|2} = K_{3|12}$. Suppose this is not the case. Then there exists $\epsilon > 0$ such that

$$\mathcal{A} = \{K_{3|12}(\mathcal{L}|x_1, x_2) > K_{3|2}(\mathcal{L}|x_2) + \epsilon\}$$

has non zero probability for some $\mathcal{L} \in \mathcal{H}_3$. Let $\mathcal{F}_I$ be the σ-algebra generated by X_I. By the local Markov property

$$\begin{aligned}\mathbb{E}[1_{\mathcal{L}}1_{\mathcal{A}}] &= \mathbb{E}[\mathbb{E}_{\mathcal{F}_{1,2}}[1_{\mathcal{L}}1_{\mathcal{A}}]] = \mathbb{E}[1_{\mathcal{A}}\mathbb{E}_{\mathcal{F}_2}[1_{\mathcal{L}}]] \\ &= \int_{\mathcal{A}} K_{3|2}(\mathcal{L}|x_2)dP_{X_1X_2}.\end{aligned}$$

Now by direct computation

$$\begin{aligned}\mathbb{E}[1_{\mathcal{L}}1_{\mathcal{A}}] &= \int_{\mathcal{A}} K_{3|12}(\mathcal{L}|x_1, x_2)dP_{X_1X_2} \\ &> \int_{\mathcal{A}} K_{3|2}(\mathcal{L}|x_2)dP_{X_1X_2} + \epsilon\mathbb{P}[\mathcal{A}] \\ &= \mathbb{E}[1_{\mathcal{L}}1_{\mathcal{A}}] + \epsilon\mathbb{P}[\mathcal{A}],\end{aligned}$$

which is a contradiction. This completes the proof. ■

2.2. Weak limits of factorable Gaussians

This section provides a characterization of the singular distributions in $\mathrm{loc}(G)$ as the weak limit of sequences in $\mathrm{loc}(G) \cap \Sigma^{++}$. Note that since $(H, \mathcal{H})$ is a topological space, weak convergence $P_{X_n} \xrightarrow{w} P_X$ is well-defined .

We note that in the case of Gaussians $X_n \sim N(0, \sigma_n), X \sim N(0, \sigma)$, $P_{X_n} \xrightarrow{w} P_X$ is equivalent to $\sigma_n \to \sigma$ in the standard metric.To characterize $\mathrm{loc}(G) \cap \partial\mathcal{N}^+$, we shall find it useful to work with the parametrization

$$X_i = \sum_{j<i} \alpha_{ij} X_j + \omega_i Z_i, \tag{7}$$

where Z_i's are independent standard Gaussians. Suppose that $\alpha_{ij} = 0$ for all $(i, j) \notin E$, where E denotes the set of (directed) edges of G. Then this parametrization gives a polynomial map $f_G : \mathbb{R}^{|E|+n} \mapsto \mathbb{R}^{\binom{n+1}{2}}$, sending $\{\alpha_{ij}, \omega_i\}$ to $\mathrm{cov}(X)$. Indeed, starting from (7), one can write

$$\sigma_{ik} = \sum_{j<i} \alpha_{ij}\sigma_{jk} + \omega_i\gamma_{ik}$$

where $\gamma_{ik} = \mathrm{Cov}(Z_i, X_k)$, $\sigma_{ik} = \mathrm{Cov}(X_i, X_k)$. Note that $\gamma_{ik} = 0$ for $k < i$. With this notation, we can write

$$\gamma_{ki} = \sum_{j>i} \alpha_{ij}\gamma_{kj} + \omega_i\delta_{ik} = \sum_{j>i} \gamma_{kj}\alpha^*_{ji} + \omega_i\delta_{ik}.$$

Set $\Gamma := [\gamma_{ij}], A := [\alpha_{ij}], \Omega := [\omega_{ii}], \Sigma := [\sigma_{ij}]$. We can write the above equations in matrix form:

$$\Sigma = A\Sigma + \Omega\Gamma, \quad \Gamma = \Gamma A^* + \Omega.$$

Hence,

$$\Sigma = (I - A)^{-1}\Omega^2(I - A^*)^{-1}.$$

The image of f_G is Zariski dense in $[\mathrm{loc}(G) \cap \Sigma^{++}]_Z$:

Proposition 3 (*Proposition 2.5 in [3]*)**.** *Let G be a DAG and E be its set of edges. Then $[\mathrm{loc}(G) \cap \Sigma^{++}]_Z$ is a rational affine irreducible variety of dimension $n + |E|$.*

The next Proposition shows that

$$X_G = [\mathrm{loc}(G) \cap \Sigma^{++}]_Z.$$

Proposition 4. *Let G be a DAG. Then*

(a) $\mathrm{loc}(G) \cap \Sigma^{++}$ *is dense in* $\mathrm{loc}(G)$.

(b) $\mathrm{loc}(G) \cap \hat{\Sigma}^{++}$ *is dense in* $\mathrm{loc}(G) \cap \hat{\Sigma}$.

Proof. For part (a), we want to show $\mathrm{loc}(G) \subset [\mathrm{Im} f_G \cap \Sigma^{++}]$. Given a Gaussian $X \in \mathrm{Loc}(G)$, start with a representation

$$X_i = \sum_{j<i} \alpha_{ij} X_j + \omega_i Z_i,$$

where Z_i's are i.i.d. Gaussians. We need to show that there exists a G-compatible representation $\{\alpha'_{ij}, \omega_i\}$ for X, i.e., $\alpha'_{ij} = 0$ for all adjacent nodes i, j in G. We may assume by induction that the α_{ij}'s are G-compatible for $i, j < n$. Now write

$$X_n = X_{\mathbf{pa}} + X_{\mathbf{npa}} + \omega_n Z_n,$$

where

$$X_{\mathbf{pa}} = \sum_{i \in \mathbf{pa}(n)} \alpha_{in} X_i, \; X_{\mathbf{npa}} = \sum_{i \in ([n-1] \setminus \mathbf{pa}(n)} \alpha_{in} X_i.$$

Note that, for general random variables A, B, C, we have

$$A \perp\!\!\!\perp (B + C)|C \iff A \perp\!\!\!\perp B|C.$$

It thus follows that

$$X_{\mathbf{npa}(n)} \perp\!\!\!\perp X_{\mathbf{npa}(n)} + \omega_n Z_n | X_{\mathbf{pa}(n)}.$$

Now observe that for independent random variables A, Z

$$A \perp\!\!\!\perp A + Z \implies \mathbb{E}[(A + Z)A] = \mathbb{E}[A + Z]\mathbb{E}[A],$$

which implies $\mathbb{E}[A^2] = \mathbb{E}[A]^2$. In particular, if A is a Gaussian then it must be a constant. It follows from this observation that $X_{\mathbf{npa}(n)}$ is a linear function of $X_{\mathbf{pa}(n)}$, say $X_{\mathbf{npa}(n)} = cX_{\mathbf{pa}(n)}$. The G-compatible α'_{in}'s are then obtained by setting $\alpha'_{in} = (1 + c)\alpha_{in}$ if $i \in \mathbf{pa}(n)$ and $\alpha'_{in} = 0$ otherwise.

For part (b), given $\sigma \in \mathrm{loc}(G) \cap \hat{\Sigma}^{++}$, we can find a sequence σ_n in $\mathrm{loc}(G) \cap \Sigma^{++}$ that converges to σ. Pass to a subsequence with $\sigma_{ii} \neq 0$ for all i. Normalize the coordinates (using the defining equations) along σ_{ii}'s to obtain a sequence in $\mathrm{loc}(G) \cap \hat{\Sigma}^{++}$. Normalization is continuous around σ. Thus, the new sequence convergences to σ as well. ■

Remark 2. *The above proof also shows that* $\mathrm{loc}(G) = \mathrm{Im} f_G \cap \Sigma^+$.

This proposition shows that X_G contains all G-factorable Gaussians. There are, however, (singular) covariances on X_G that are not G-compatible. In other words, unlike independence, conditional independence is not preserved under weak limits as shown in the following example.

Example 1 ($\mathrm{loc}(G)$ is not closed)**.** *Let* $X_n \sim N(0,1)$, $W_n \sim N(0,1)$ *be independent Gaussians. Set* $Z_n = X_n$ *and* $Y_n = \frac{1}{n} X_n + \frac{\sqrt{n^2-1}}{n} W_n$. *Then* $X_n \perp\!\!\!\perp Z_n | Y_n, W_n$ *for all* n *and* $P_{X_n,Y_n,Z_n,W_n} \xrightarrow{w} P_{X,Y,X,W}$ *with* $X \sim N(0,1), W \sim N(0,1)$ *and* $Y = W$. *However,*

$$X \not\perp\!\!\!\perp X | W.$$

Thus the closure of $\mathrm{loc}(G) \cap \hat{\Sigma}^{++}$ *strictly contains* $\mathrm{loc}(G) \cap \hat{\Sigma}$.

Remark 3. *In general, the weak convergence of the joint* $P_{X^{(n)}} \xrightarrow{w} P_X$ *does not imply that of the conditional kernels* $P_{X_i^{(n)}|X_j^{(n)}} \xrightarrow{w} P_{X_i|X_j}$. *If the latter conditions are also satisfied, then conditional independence is preserved at the (weak) limit*[6]*.*

2.3. DAG isomorphism

The next result states that isomorphism of DAGs can be decided inside $\mathcal{N}^+$[7]:

Proposition 5 (Theorem 5.1 in [2])**.** *Let* G, G' *be DAGs. Then the following are equivalent:*

(a) G *and* G' *are equal.*

(b) G *and* G' *are* $\mathcal{N}$*-equal.*

(c) G *and* G' *are* $\mathcal{N}^+$*-equal.*

This property is also known as the faithfulness of Gaussians in the statistics literature (cf. [26]). Let us point out that, in general, $\mathcal{N}^+$-isomorphic models are not $\mathcal{N}$-isomorphic as shown in the next example.

[6] This follows directly from the lower semi-continuity of divergence.

[7] This statement generalizes to the class of chain graphs. See [2] for details.

Example 2. *Consider the models*

$$G_1 : 1 \perp\!\!\!\perp 3|2 \quad \& \quad 1 \perp\!\!\!\perp 2|3 \quad \& \quad 2 \perp\!\!\!\perp 3|1$$
$$G_2 : 1 \perp\!\!\!\perp 2 \perp\!\!\!\perp 3$$

A non-singular Gaussian belongs to the first model if and only if it belongs to the second model. However, a Gaussian $X_1 = X_2 = X_3$ is only compatible with the first model.

2.4. DAG varieties and ideals

Here we provide some algebraic and geometric descriptions for $\mathrm{loc}(G)$:

Theorem 1. *Let G be a DAG and let $\theta_0 = \prod_{A \subset [n]} (|\Sigma_{AA}|)$.*

(a) There is a Zariski closed subset B_G so that

$$V(I_G) = X_G \cup B_G$$

where $B_G \subseteq V(\theta_0) = \{\theta_0 = 0\}$.

(b) Let $\mathfrak{p}_G = I(\mathrm{loc}(G))$ so that $X_G = V(\mathfrak{p}_G)$. Then $\mathfrak{p}_G$ is a prime ideal obtained by saturating I_G

$$\mathfrak{p}_G = S^{-1} I_G \cap \mathbb{C}[\Sigma] \tag{8}$$

at the multiplicatively closed set $S = \{\theta_0^n, n = 1, \ldots\}$.

(c) X_G is smooth inside $\Sigma^{\bullet}$.

Remark 4.

(a) In Theorem 1b, we can replace I_G with J_G.

(b) It follows that $V(I_G)$ and $V(J_G)$ do not miss a single G-compatible Gaussian, but can add some bad components to the boundary $\partial\Sigma^{++}$. Theorem 1b states this in algebraic terms and provides a proof of Conjecture 3.3 in [3]. Theorem 8 in [23] gives an analogous result for the implied ideals of discrete random variables.

(c) In Theorem 1b, one can replace θ_0 with the product of principal minors $|\sigma_{KK}|$ where K appears as a conditional set in some imposed relation $i \perp\!\!\!\perp j|K$.

(d) There are many equivalent ways to recover $\mathfrak{p}_G$ from I_G besides (8)*. Indeed, (e.g. [27, Chapter 4]) we have*

$$\mathfrak{p}_G = (I_G : \theta_0^m)$$

for all m sufficiently large. Another characterization is from primary decomposition of I_G:

$$I_G = \mathfrak{p}_G \cap \mathfrak{q}_1 \cdots \mathfrak{q}_r \,,$$

where $\mathfrak{p}_G$ is the unique component that is contained in maximal ideal $\mathfrak{m}_x$ corresponding to covariance matrix x with non-singular principal minors (e.g. identity).

(e) *One can ask if Theorem 1 generalizes, i.e., if* $[V(I_G) \cap \Sigma^{++}]_Z = [V(I_G) \cap \Sigma^{\bullet}]_Z$ *for any conditional independence model* G*. This is equivalent to asking if some component of* $V(I_G)$ *can intersect* $\Sigma^{\bullet}$ *but avoid* Σ^{++}*. This question appears in [28], and remains open so far as we know.*

Proof. Let us consider the ring $S^{-1}\mathbb{C}[\Sigma]$ and if ij is an edge in G or if $i = j$, call σ_{ij} an edge variable and the rest are non-edge variables. Denote by Σ_{edge} the subspace of $\Sigma = [\sigma_{ij}]$ corresponding to the edge variables. Note that the ideal $S^{-1}I_G$ has one generator g_{ij} for every non-edge variable σ_{ij}. This generator corresponds to the constraint $i \perp\!\!\!\perp j | K$ where $i < j$ and $K \subset \{k : k < j\}$ are parents of j and we have

$$g_{ij} = \sigma_{ij}|\sigma_{KK}| - h_{ij}\,.$$

Here $|\sigma_{KK}|$ and h_{ij} are polynomials in $\{\sigma_{a,b}, a \leq b < j\}$. Now introduce a lexicographic ordering on pairs (i,j)[8] and among all non-edge variables entering into polynomial on the right consider the maximal one – denote it $\sigma_{i',j'}$. This variable has its corresponding generator:

$$g_{i'j'} = \sigma_{i'j'}|\sigma_{K'K'}| - h_{i'j'}.$$

Thus multiplying by a suitable power of $|\sigma_{K'K'}|$ we can write

$$|\sigma_{K'K'}|^r g_{ij} = |\sigma_{K'K'}|^r(\sigma_{ij}|\sigma_{KK}| - h_{ij})$$

and now every occurrence of $|\sigma_{K'K'}|\sigma_{i'j'}$ we replace with $g_{i'j'} + h_{i'j'}$. In the end we obtain

$$|\sigma_{K'K'}|^r g_{ij} = \sigma_{ij}u'_{ij} - h'_{ij}\,,$$

where the expression on the right no longer contains $\sigma_{i',j'}$ or any larger (w.r.t. ordering of pairs) variable. Thus, repeating similar steps in the end we obtain expression:

$$\gamma_{ij}g_{ij} = \sigma_{ij}\tilde{u}_{ij} - \tilde{h}_{ij}$$

where we have:

1. $\gamma_{ij} = \gamma_{ij}(\sigma)$ is a polynomial with $\mathbb{Z}$-coefficients.
2. γ_{ij} and $\tilde{u}_{ij}$ are units in $S^{-1}\mathbb{C}[\Sigma]$ (equivalently, they divide θ_0^m for some large enough m).
3. $\tilde{u}_{ij}$ and $\tilde{h}_{ij}$ are both polynomials with $\mathbb{Z}$-coefficients in *edge* variables $\sigma_{a,b}$ with $a \leq b < j$.

Consequently, since γ_{ij}'s are units we have

$$S^{-1}I_G = (\sigma_{ij}\tilde{u}_{ij} - \tilde{h}_{ij}, (i,j)\text{– non-edge}) \tag{9}$$

[8]For instance, take $(i,j) < (i',j')$ if $i < i'$ or $i = i'$ and $j < j'$ in the topological sort.

Note that on one hand $\mathfrak{p}_G = I(X_G)$ contains I_G. On the other hand, by Proposition 3 and (9) any minimal prime above $S^{-1}I_G$ has codimension equal to $S^{-1}\mathfrak{p}_G$. Thus, if we show that $S^{-1}I_G$ is prime we must have

$$S^{-1}I_G = S^{-1}\mathfrak{p}_G$$

and after intersecting with $\mathbb{C}[\Sigma]$ conclusion (8) and the rest of the theorem follow.

To that end let $g = g(\sigma_{\text{edge}})$ be the product of all $\tilde{u}_{ij}$. Let $\mathring{\sigma}$ be the identity matrix

$$\mathring{\sigma}_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

and let

$$\varphi^* : \mathbb{C}[\Sigma] \to \mathbb{C}[\Sigma_{\text{edge}}][1/g]$$

be the ring map associated to the rational map $\Sigma_{\text{edge}} \xrightarrow{\varphi} \Sigma$ given by

$$\varphi : \sigma_{\text{edge}} \mapsto \left(\sigma_{\text{edge}}, \frac{\tilde{h}_{ij}(\sigma_{\text{edge}})}{\tilde{u}_{ij}(\sigma_{\text{edge}})}, (i,j)\text{– non-edge} \right). \tag{10}$$

First note that $\mathring{\sigma}$ is a $\mathbb{C}$-point of $\mathbb{C}[\Sigma_{\text{edge}}][1/g]$ since

$$\theta_0(\mathring{\sigma}) = 1,\ g|\theta_0^m \implies g(\mathring{\sigma}) \neq 0.$$

Thus we can form a commutative diagram

$$\begin{array}{ccc} \mathbb{C}[\Sigma] & \xrightarrow{\varphi^*} & \mathbb{C}[\Sigma_{\text{edge}}][1/g] \\ & {\scriptstyle \text{ev}_{\mathring{\sigma}}} \searrow \quad \swarrow {\scriptstyle \text{ev}_{\mathring{\sigma}_{\text{edge}}}} & \\ & \mathbb{C} & \end{array}$$

Chasing θ_0 from $\mathbb{C}[\Sigma]$ to $\mathbb{C}$ in two different ways gives $\theta_0 \notin \ker \varphi^*$, and thus $\varphi^*(\theta_0) = h/g^m$ for $h \neq 0$. Localizing $\mathbb{C}[\Sigma_{\text{edge}}][1/g]$ at h gives a diagram

$$\begin{array}{ccccc} & & \overset{\varphi^*_{gh}}{\frown} & & \\ \mathbb{C}[\Sigma] & \xrightarrow{\varphi^*} & \mathbb{C}[\Sigma_{\text{edge}}][1/g] & \xrightarrow{\varphi^*_h} & \mathbb{C}[\Sigma_{\text{edge}}][1/gh] \end{array}$$

Note that φ^*_{gh} sends θ_0 to a unit. Hence, by the universal property of localization (see [27, Proposition 3.1]), it extends to a map

$$\varphi^{*,e}_{gh} : S^{-1}\mathbb{C}[\Sigma] \to \mathbb{C}[\Sigma_{\text{edge}}][1/gh] \tag{11}$$

$$\sigma_{ij} \mapsto \varphi^*(\sigma_{ij}), 1/\theta_0 \to g^m/h \tag{12}$$

that is onto and has $S^{-1}I_G$ as kernel. To verify the latter claim, take s in the kernel of (11) and write it as $s = p_{\text{edge}} + \sum_{ij} q_{ij}.(\sigma_{ij} - \tilde{h}_{ij}/\tilde{u}_{ij})$ where

$q_{ij} \in S^{-1}\mathbb{C}[\Sigma]$ with (i,j)–non-edge, and p_{edge} is a polynomial in σ_{edge}. This can be done by virtue of the binomial theorem:

$$\sigma_{ij}^n = (\frac{\tilde{h}_{ij}}{\tilde{u}_{ij}} - \frac{\tilde{h}_{ij}}{\tilde{u}_{ij}} + \sigma_{ij})^n = (\frac{\tilde{h}_{ij}}{\tilde{u}_{ij}})^n + n(\frac{\tilde{h}_{ij}}{\tilde{u}_{ij}})^{n-1}(\sigma_{ij} - \frac{\tilde{h}_{ij}}{\tilde{u}_{ij}}) + \cdots$$

Then $\varphi^*(s) = p_{\text{edge}}$, and thus $\varphi^*(s) = 0$ gives $p_{\text{edge}} = 0$, that is $s \in S^{-1}I_G$. The reverse inclusion is obvious.

This establishes isomorphism of rings

$$S^{-1}\mathbb{C}[\Sigma]/S^{-1}I_G = \mathbb{C}[\Sigma_{\text{edge}}][1/gh],$$

which implies that $S^{-1}I_G$ is prime, and that each local ring of $X_G \cap \Sigma^\bullet$ is regular (since all local rings of $\mathbb{C}[\Sigma_{\text{edge}}][1/gh]$ are regular).

Geometrically our proof corresponds to constructing a birational isomorphism:

$$\Sigma_{\text{edge}} \supset \mathcal{U} \xrightarrow{\varphi|_{\mathcal{U}}} V(I_G) \cap \Sigma^\bullet$$
$$\nwarrow \quad \swarrow$$
$$\pi$$

where $\mathcal{U} = D(gh)$ is a distinguished open, $\varphi|_{\mathcal{U}}$ is obtained by restriction of the map given in (10)(φ is regular on $\mathcal{U}$), and π is the projection from Σ to Σ_{edge}. ■

Alternatively, we can choose to prove our results over $\mathbb{C}[\hat{\Sigma}]$. Recall that $\hat{X}_G$ is defined as $[\text{loc}(G) \cap \hat{\Sigma}^{++}]_Z$.

Theorem 2. *Let G be a DAG and let $\hat{\theta}_0 = \prod_{A \subset [n]}(|\hat{\Sigma}_{AA}|)$. With the notation of Theorem 1,*

(a) $\hat{X}_G$ is an affine rational (irreducible) variety of dimension $|E|$ and $\hat{X}_G = X_G \cap \hat{\Sigma}$.

(b) We have

$$V(\hat{I}_G) = \hat{X}_G \cup \hat{B}_G,$$

where $\hat{B}_G \subset \{\hat{\theta}_0 = 0\}$.

(c) The prime ideal $\hat{p}_G \triangleq I(\hat{X}_G)$ is obtained by saturating $\hat{I}_G$

$$\hat{\mathfrak{p}}_G = \hat{S}^{-1}\hat{I}_G \cap \mathbb{C}[\hat{\Sigma}] \tag{13}$$

at the multiplicatively closed set $\hat{S} = \{\hat{\theta}_0^n, n = 1, \ldots\}$.

(d) $\hat{X}_G$ is smooth inside $\hat{\Sigma}^\bullet$.

Proof. We first claim that $X_G \cap \hat{\Sigma}$ is an affine rational irreducible variety that is smooth inside $\hat{\Sigma}^{\bullet}$. Consider the sequence

$$S^{-1}\mathbb{C}[\Sigma] \overset{\varphi^{*,e}_{gh}}{\to} \mathbb{C}[\Sigma_{\text{edge}}][1/gh] \overset{\text{ev}}{\to} \mathbb{C}[\hat{\Sigma}_{\text{edge}}][1/\hat{g}\hat{h}], \tag{14}$$

where $\varphi^{*,e}_{gh}$ is the extended map in (11) and $\hat{g} = g|_{\hat{\Sigma}}, \hat{h} = h|_{\hat{\Sigma}}$, and ev is the evaluation map $\sigma_{ii} = 1$. Observe that

$$X_G \cap \hat{\Sigma} = V((I_G : \theta_0^m)) \cap V(\ker(\text{ev})^c) = V((I_G : \theta_0^m) + \ker(\text{ev})^c) = V(\hat{I}_G : \hat{\theta}_0^m).$$

Also note that

$$(\hat{I}_G : \hat{\theta}_0^m) = \hat{S}^{-1}\hat{I}_G \cap \mathbb{C}[\hat{\Sigma}].$$

Let us show that $\hat{S}^{-1}\hat{I}_G$ is prime of codimension $|E|$. Note that $\varphi^{*,e}_{gh}$ is identity on the generators of $\ker(\text{ev})^c$, i.e., $\varphi^{*,e}_{gh}(\ker(\text{ev})^c) = \ker(\text{ev})$. This implies that $\ker(\text{ev}\varphi^{*,e}_{gh}) \supset \ker(\varphi^{*,e}_{gh}) + \ker(\text{ev})^c$. The reverse inclusion $\ker(\text{ev}\varphi^{*,e}_{gh}) \subset \ker(\varphi^{*,e}_{gh}) + \ker(\text{ev})^c$ is obvious. Since both maps in (14) are onto, we have an isomorphism of integral domains

$$S^{-1}\mathbb{C}[\Sigma]/(S^{-1}I_G + \ker(\text{ev})^c) = \mathbb{C}[\hat{\Sigma}]/\hat{S}^{-1}\hat{I}_G = \mathbb{C}[\hat{\Sigma}_{\text{edge}}][1/\hat{g}\hat{h}], \tag{15}$$

where we used the fact that $\ker(\varphi^{*,e}_{gh}) = S^{-1}I_G$, shown in the proof of Theorem 1. This proves the primality of $\hat{S}^{-1}\hat{I}_G$, hence, $X_G \cap \hat{\Sigma}$ is irreducible of dimension $|E|$. This further implies that the map (15) is induced by the restriction of the rational map $\varphi|_{\hat{\Sigma}}$. In particular, this restriction is an isomorphism inside $\hat{\Sigma}^{\bullet}$, proving the above claim.

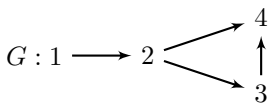
Now the proof reduces to shwoing that $\hat{X}_G = X_G \cap \hat{\Sigma}$. We know that $X_G \cap \hat{\Sigma}$ is irreducible by the above discussion, and that it contains $\hat{X}_G$ by Proposition 4. Thus the theorem follows if we show that $\hat{X}_G$ has dimension $|E|$. Recall from Proposition 1 and construction of φ that $\text{loc}(G) \cap \hat{\Sigma}^{++}$ can be obtained as the intersection of the image of φ with the real subset of $\hat{\Sigma}^{++}$, i.e., it has the structure of a real differentiable manifold of dimension $|E|$. Together with Proposition 4 and [29, Proposition 2.8.14], this implies that

$$\dim(\mathbb{R}[\hat{\Sigma}]/I_{\mathbb{R}}(\hat{X}_G)) = |E|,$$

where $I_{\mathbb{R}}(\hat{X}_G)$ is the ideal of polynomials with real coefficients that vanish on $\hat{X}_G$. Since the set $\text{loc}(G) \cap \hat{\Sigma}$ is stable under conjugation, we have $I(\hat{X}_G) = I_{\mathbb{R}}(\hat{X}_G)^e$. Then the going up theorem implies that $\hat{X}_G$ has dimension $|E|$ as desired. ■

The next example shows how Theorem 1 can be used to construct $\mathfrak{p}_G$ from I_G:

Example 3. *Consider the DAG*



The ideal of imposed relations is generated by relations $1 \perp\!\!\!\perp 3|2$ *and* $4 \perp\!\!\!\perp 1|(2,3)$*:*

$$I_G = \langle |\sigma_{12,23}|, |\sigma_{123,423}| \rangle.$$

It has primary components

$$I_{G,1} = \langle \sigma_{12}\sigma_{23} - \sigma_{13}\sigma_{22}, \sigma_{12}\sigma_{24} - \sigma_{14}\sigma_{22}, \sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23} \rangle$$

and

$$I_{G,2} = \langle \sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}, \sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}, \sigma_{22}\sigma_{33} - \sigma_{23}^2 \rangle.$$

It can be seen that only $I_{G,1}$ *intersects* $\Sigma^\bullet$*. We thus have* $\mathfrak{p}_G = I_{G,1}$*. Furthermore,* $I_{G,1}$ *is the unique ideal contained in the maximal ideal at the identity of* Σ*, and is also equal to the saturation of* I_G *at* $f = \sigma_{22}(\sigma_{22}\sigma_{33} - \sigma_{23}^2)$*. The ideal of implied relations is generated by relations* $1 \perp\!\!\!\perp 3|2, 1 \perp\!\!\!\perp 4|2, 1 \perp\!\!\!\perp 4|(2,3)$*, and* $1 \perp\!\!\!\perp 3|(2,4)$*:*

$$J_G = \langle |\sigma_{12,23}|, |\sigma_{12,42}|, |\sigma_{123,423}|, |\sigma_{124,324}| \rangle.$$

It has primary components

$$J_{G,1} = \langle \sigma_{13}\sigma_{22} - \sigma_{12}\sigma_{23}, \sigma_{14}\sigma_{22} - \sigma_{12}\sigma_{24}, \sigma_{14}\sigma_{23} - \sigma_{13}\sigma_{24} \rangle$$

and

$$J_{G,2} = \langle \sigma_{12}, \sigma_{22}, \sigma_{24}, \sigma_{23}^2 \rangle.$$

Again one can check that $J_{G,1} = I_{G,1} = S^{-1}J_G \cap \mathbb{C}[\Sigma]$ *is the unique component that is contained in the maximal ideal at the identity. Finally, we can see that*

$$S^{-1}\hat{I}_G \cap \mathbb{C}[\hat{\Sigma}] = \langle \hat{\sigma}_{12}\hat{\sigma}_{23} - \hat{\sigma}_{13}, \hat{\sigma}_{12}\hat{\sigma}_{24} - \hat{\sigma}_{14}, \hat{\sigma}_{13}\hat{\sigma}_{24} - \hat{\sigma}_{14}\hat{\sigma}_{23} \rangle$$

is the unique irreducible component of $V(\hat{I}_G)$ *that contains the origin of* $\hat{\Sigma}$*.*

Let us point out that X_G need not be smooth outside of $\Sigma^\bullet$ as the next example shows:

Example 4. *Let* G *be the Markov chain* $1 \to 2 \to 3$*. Then* $X_G = V(\sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23})$ *is a cone and has a singularity at the origin.*

2.5. Algebraic representation

Here, we put together the results of the previous sections to give an algebraic criteria for testing isomorphism of graphical models. We start by a result on equivalence of DAGs:

Proposition 6. *Let* G, G' *be DAGs. Then* G *is equal to* G' *if and only if* $X_G = X_{G'}$*, or equivalently, if and only if* $\hat{X}_G = \hat{X}_{G'}$*.*

Proof. If G is equal to G', then $\mathrm{loc}(G) = \mathrm{loc}(G')$ by Proposition 5, and thus $X_G = X_{G'}$. Conversely, if $X_G = X_{G'}$, then $V(I_G) \cap \Sigma^\bullet = V(I_{G'}) \cap \Sigma^\bullet$ by Theorem 1a. This implies that $V(I_G) \cap \Sigma^{++} = V(I_{G'}) \cap \Sigma^{++}$ and by (6) that $\mathrm{loc}(G) \cap \Sigma^{++} = \mathrm{loc}(G') \cap \Sigma^{++}$. Then by Proposition 5 we have $G = G'$. The last part of the assertion follows by a similar reasoning. ∎

In what follows, $\Pi = \{\pi_s\}_{s \in S_n}$ is the permutation group with induced action on $\mathbb{C}[\Sigma]$: $\pi_s(f(\sigma_{ij})) = f(\sigma_{s(i)s(j)})$ where $s \in S_n$ is a permutation of indices. The invariant subring $\{f \in \mathbb{C}[\Sigma] \mid f \circ \pi_s = f \ \forall s\}$ is denoted by $\mathbb{C}[\Sigma]^{\Pi}$. We can now state our main result:

Theorem 3. *Let G, G' be DAGs and S be as in Theorem 1. Then $G \sim G'$ if and only if*

$$S^{-1}I_G \cap \mathbb{C}[\Sigma]^{\Pi} = S^{-1}I_{G'} \cap \mathbb{C}[\Sigma]^{\Pi}. \tag{16}$$

Proof. By Proposition 6, DAGs G and G' are isomorphic iff there exists $\pi \in \Pi$ such that

$$X_{\pi(G)} = X_{G'}.$$

By Theorem 1b this in turn is equivalent to

$$\pi(X_G) = X_{G'}. \tag{17}$$

To see this note that

$$X_{\pi(G)} = V(I_{\pi(G)} : \theta_0^m) = V(\pi(I_G : \theta_0^m))) = \pi(X_G).$$

Furthermore, since S in Theorem 1 is Π-invariant, (17) is equivalent to

$$\pi(S^{-1}I_G) \cap \mathbb{C}[\Sigma] = S^{-1}I_{G'} \cap \mathbb{C}[\Sigma]\,. \tag{18}$$

Finally, because of primality of ideals in (18), the existence of π satisfying (18) is equivalent to (19) by [27, Exercise 5.13]. ■

As before, we can work over $\mathbb{C}[\hat{\Sigma}]$.

Theorem 4. *: Let G, G' be DAGs and $\hat{S}$ be as in Theorem 2. Then G and G' are isomorphic if and only if*

$$\hat{S}^{-1}\hat{I}_G \cap \mathbb{C}[\hat{\Sigma}]^{\Pi} = \hat{S}^{-1}\hat{I}_{G'} \cap \mathbb{C}[\hat{\Sigma}]^{\Pi}. \tag{19}$$

Proof. Recall that $X_{\pi(G)} = \pi(X_G)$. Thus by Proposition 6 we have

$$\pi(X_G) = X_{G'} \iff \pi(\hat{X}_G) = \hat{X}_{G'}.$$

The rest of the proof is analogous to that of Theorem 3 and is omitted. ■

We remark that $\hat{S}^{-1}\hat{I}_G$ can be replaced with $\hat{J}_G$ in the above. The presentation of the results in this form is a matter of convenience. Indeed, there is a simple way to generate $\hat{I}_G$:

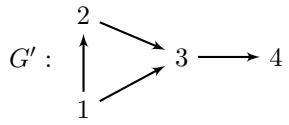
- Traverse the graph in the order of the topological sort and set $\hat{I}_i := \sum_{j<i, j \notin K} \langle |\hat{\sigma}_{iK,jK}| \rangle$ with $K := \mathbf{pa}(i)$ for all i.

- Output $\hat{I}_G = \sum_i \hat{I}_i$.

However, extracting the implied relations of a DAG requires more work.

We also point out that the extra components of $V(\hat{I}_G)$ (or $V(I_G)$) are not invariant across the isomorphism class of G:

Example 5. *The DAG*



is isomorphic to the DAG G in Example 3, but $\hat{I}_{G'}$ is a prime ideal. Note however that $J_{G'} = J_{\pi(G)}$ where π is the permutation (14)(23)*. Thus it is necessary to compute the saturation ideal in (19). We shall see, however, in section 2.7 that one can avoid computing the saturation ideal, and more importantly, the subring intersection by a probabilistic procedure.*

2.6. Changing the ground ring

We briefly note that none of our algebraic arguments depend on the choice of $\mathbb{C}$ as ground ring. In particular, instead of X_G we could have considered

$$X'_G = \mathrm{Spec}(\mathbb{Z}[\Sigma]/\mathfrak{q}_G)$$

where $\mathfrak{q}_G = S^{-1}I_G \cap \mathbb{Z}[\Sigma]$. Then we still have that X'_G is an integral, rational scheme over $\mathbb{Z}$, smooth at every point in $D(\theta_0) = \{\theta_0 \neq 0\}$. Two graphs $G \sim G'$ are isomorphic if and only if

$$S^{-1}I_G \cap \mathbb{Z}[\Sigma]^\Pi = S^{-1}I_{G'} \cap \mathbb{Z}[\Sigma]^\Pi.$$

Furthermore, we can identify X_G as the base-change of X'_G to $\mathbb{C}$: $X_G = X'_G \times_{\mathrm{Spec}(\mathbb{Z})} \mathrm{Spec}(\mathbb{C})$. To verify this, one merely needs to check that $\mathfrak{q}_G^e = \mathfrak{p}_G$. Indeed I_G has a Gröbner basis (w.r.t any term order) consisting of generators with $\mathbb{Z}$-coefficients (run Buchberger's algorithm and clear denominators at the end). One can thus see, by variable elimination, that $\mathfrak{p}_G = S^{-1}I_G \cap \mathbb{C}[\Sigma]$ can be generated by polynomials in $\mathbb{Z}[\Sigma]$. Other convenient choices of rings are $\mathbb{Q}$ and $\mathbb{R}$.

2.7. Randomized algorithm

Theorem 4 shows that testing isomorphism amounts to subring intersection. Computing this intersection is difficult since there is no easy description available for generating invariants of $\mathbb{C}[\hat{\Sigma}]^\Pi$. Another computational difficulty is that of computing the saturation ideal. This operation does not scale well with the number of nodes in the model. Here we give a randomized algorithm that avoids computing both the intersection and saturation ideals.

Algorithm 1 $\mathrm{ISODAG_m}$

function $\mathrm{ISODAG_m}(G,\, G')$
Sort G and G' topologically
Initialize ISO $\leftarrow$ **true**, $r \leftarrow 1$
while ISO **and** $r \leq m$ **do**
  Sample $z_G^r, z_{G'}^r$ respectively from $\hat{\varphi}_* P_G$,$\hat{\varphi}_* P_{G'}$ as follows:

  (i) Sample edge variables $\hat{\sigma}^r_{\mathrm{edge}}$ of G from P_G

  (ii) **for** $i := 2$ to n **do**
    Solve the (linear) toposorted imposed relations

$$|\hat{\sigma}^r_{iK,jK}| = 0, \quad K := \mathbf{pa}(i)$$

    for each non-edge variable $\hat{\sigma}^r_{ij}$, $j < i$

  (iii) **end for**

  (iv) $z_G^r \leftarrow (\hat{\sigma}^r_{\mathrm{edge}}, \hat{\sigma}^r_{\mathrm{non-edge}})$

  (v) Repeat for G'

  if $\Pi(z_G^r) \cap V(\hat{I}_{G'}) = \emptyset$ **or** $\Pi(z_{G'}^r) \cap V(\hat{I}_G) = \emptyset$ **then**
    ISO $\leftarrow$ **false**
  end if
  $r \leftarrow r + 1$
end while
return ISO

Let φ be the rational map in (10) and denote by $\hat{\varphi}$ its restriction $\varphi|_{\hat{\Sigma}}$. Let $\mathcal{Y} := \mathbb{F}_q^{|E|}$ and define

$$\mathcal{U} := \{y \in \mathcal{Y} : \hat{g}(y) \neq 0, \hat{h}(y) \neq 0\},$$

where $\hat{g}, \hat{h}$ are as in (14). Note that

$$|\mathcal{U}| \geq (q-d)q^{|E|-1}$$

with $d := \deg(\hat{g}) + \deg(\hat{h})$.

Now construct a random matrix with uniform distribution P_G on the finite set $\mathcal{U}$. This can be realized, for instance, by transforming a uniformly distributed matrix on $\mathcal{Y}$ through a kernel $P_{Y|X} : \mathcal{Y} \to \mathcal{Y} \cup \{\emptyset\}$ such that $P_{Y|X}(x|x) = 1$ if $x \in \mathcal{U}$ and $P_{Y|X}(\emptyset|x) = 1$ otherwise.Let $\hat{\varphi}_* P_G$ be the push-forward of P_G under $\hat{\varphi}$ and Z_G^i's be independent random variables with common distribution $\hat{\varphi}_* P_G$.

Given DAGs G, G', the algorithm $\mathrm{ISODAG_m}$ described above constructs m realizations $z_G^i, z_{G'}^i$ from $Z_G^i, Z_{G'}^i$. It then declares G and G' to be isomorphic

if and only if for each $i \leq m$, there is some permutation π such that both $\pi(z^i_G) \in V(\hat{I}_{G'})$ and $z^i_{G'} \in V(\pi(\hat{I}_G))$ hold. The latter conditions amount, respectively, to checking $f'_j(\pi(z^i_G)) = 0$ and $\pi(f_k)(z^i_{G'}) = 0$ for all generators f'_j of $\hat{I}_{G'}$ and f_k of $\hat{I}_G$. The next theorem shows that the probability of failure of the algorithm can be made arbitrarily small:

Theorem 5. *Let G be a DAG on n nodes and E be its set edges. Let Z^i_G be as in above and set $d := \deg(\hat{g}) + \deg(\hat{h})$. If $G \sim G'$, then*

$$\mathbb{P}[\mathrm{ISODAG_m}(G, G') = \text{yes}] = 1.$$

If $G \not\sim G'$, then

$$\mathbb{P}[\mathrm{ISODAG_m}(G, G') = \text{yes}] \leq (n! \frac{n + 2d - 1}{q - d})^m.$$

Proof. By Theorem 4, the algorithm outputs yes with probability 1 if G is isomorphic to G'. Now suppose $\mathrm{loc}(G) \not\subset \mathrm{loc}(G')$. We need to upper bound the probability that a realization of Z^i_G lands on $\hat{X}_{G'}$. Note that Z^i_G takes values on $\hat{X}_G \cap \hat{\Sigma}^\bullet$. It follows from Proposition 6 that a point on $\hat{X}_{G'} \cap \hat{\Sigma}^\bullet$ satisfies at least one relation of the form $|\hat{\sigma}_{iK,jK}|$ where $i \perp\!\!\!\perp j | K \notin \mathcal{M}^{topo}_G$. We can pullback the intersection $\hat{X}_G \cap \hat{\Sigma}^\bullet \cap \{|\hat{\sigma}_{iK,jK}| = 0\}$ to $\mathcal{U}$ via the embedding $\hat{\varphi}$. This corresponds to replacing the non-edge variables (w.r.t G) of $|\sigma_{iK,jK}|$ with rational functions of edge variables as in (10). Let f_{ijK} be the resulting rational function. Then the number of points on the intersection is upper bounded by the number of solutions in $\mathcal{Y}$ of $\{f_{ijK} = 0\}$. Note that the minor has total degree at most $n - 1$ and the largest power of any monomial appearing in the expansion of such minor is at most 2. Thus clearing the denominators in f_{ijK} gives a non-zero polynomial of degree at most $(n + 2d - 1)$ in at most $|E|$ variables. Such a polynomial has at most $(n+2d-1)q^{|E|-1}$ roots in $\mathcal{Y}$. On the other hand, $\mathcal{U}$ has at least $(q - d)q^{|E|-1}$ points. The union bound proves the result. ∎

If G and G' are two DAGs on n nodes and E edges, it is easy to verify that the algorithm can decide equivalence of G and G' in time $O(n^3|E^c|)$. To test if a sampled point lies on $\hat{X}_G$, one needs to solve a sequence of $|E^c|$ linear equations. Each equation involves one missing edge and one minor of size at most $(n - 1) \times (n - 1)$, which requires $O(n^3)$ operations to compute (note that we can do the arithmetics over rationals). This is comparable, in terms of complexity, with $O(n^4|E|)$ operations needed in the essential graph method (see [4]).

3. Directed tree models

We study the case of directed tree models in this section. The main property is the following:

Proposition 7. *Let T be a directed tree model. Then $\hat{I}_T$ is prime and $V(I_T) \cap \hat{\Sigma} = [\mathrm{loc}(T) \cap \hat{\Sigma}]_Z$. Furthermore, $\hat{X}_T$ is smooth everywhere.*

This is analogous to primality of J_T shown in [3] (see Corollary 2.4 and Theorem 5.8). A direct consequence is that two tree models are isomorphic if and only if $\hat{I}_T \cap \mathbb{C}[\hat{\Sigma}]^{\Pi} = \hat{I}_{T'} \cap \mathbb{C}[\hat{\Sigma}]^{\Pi}$. It also follows from the proposition that $\hat{I}_T = \hat{J}_T$. In II.E, we give a procedure to generate $\hat{I}_T$. To prove primality of $\hat{I}_T$, we introduce a second procedure that uses lower degree generators by modifying the first step:

- Set $L_0 = \emptyset$.
- Traverse the graph in order of the sort and set $K_i := \mathbf{pa}(i)$ for all i.
- For all $j < i$, let $K' \in K_i$ be the smallest subset that d-separates j and i. Set $L_i = L_{i-1} \cup \{f | f = \langle |\sigma_{iK',jK'}| \rangle\}$.
- Output $\hat{I}'_G := \sum_{f \in L_n} \langle f \rangle$.

Let us verify $\hat{I}_T = \hat{I}'_T$ for a tree model T. Suppose the claim holds on $n-1$ nodes and let the first procedure reach node n. Set $K = \mathrm{pa}(i)$ and let k be the separator of nodes i and n for some $i < n$. The generated polynomial is

$$|\sigma_{iK,nK}| = \begin{vmatrix} \sigma_{in} & \sigma_{iK} \\ \sigma_{Kn} & \sigma_{KK} \end{vmatrix}.$$

This can be expanded as

$$|\sigma_{iK,nK}| = \begin{vmatrix} \sigma_{in} & \sigma_{ik} \\ \sigma_{kn} & 1 \end{vmatrix} + \sum_{j \in \{K-k\}} \sigma_{ij} f_j + \sum_{j,j' \in K} \sigma_{jj'} f_{jj'},$$

which is in the ideal of the second procedure since i is not connected to $K-k$ and the parents of n are not connected either. Conversely, take the minor

$$\begin{vmatrix} \sigma_{in} & \sigma_{ik} \\ \sigma_{kn} & 1 \end{vmatrix}$$

produced by the second procedure and note that $i \perp\!\!\!\perp K-k$ and $k' \perp\!\!\!\perp k''$ for distinct $k', k'' \in K$. By the inductive hypothesis, the corresponding linear forms are in I_T. Using the above expansion for $|\sigma_{iK,nK}|$ one can see that this reduced minor is in the ideal of the first procedure as well. Thus the two ideals are equal when T is a tree. The first procedure is easy to implement and is what we use to generate the ideals. The second procedure is easy to analyze and has useful properties that we exploit later. For instance, it shows that the ideal of imposed relations of a tree model is generated by quadratic polynomials of type $\sigma_{ij} - \sigma_{ik}\sigma_{kj}$ or linear forms σ_{ij}. We note that the equality of $\hat{I}_T$ and $\hat{I}'_T$ does not generalize to all DAGs. For instance, in the case of Example 3, $\hat{I}_T$ is not prime whereas $\hat{I}'_T$ is prime for all DAGs with $n \leq 4$ nodes.

We are now ready to prove that the ideal of the imposed relations is prime for tree models:

Proof of Proposition 7. The generators of $\hat{I}_T$ are of the form either $\hat{\sigma}_{ik} - \hat{\sigma}_{ij}\hat{\sigma}_{jk}$ or $\hat{\sigma}_{ij}$. Then the rational map (11) in the proof of Theorem 1 can be taken to be a polynomial map. In other words, $\hat{S}^{-1}\hat{I}_T \cap \mathbb{C}[\hat{\Sigma}] = \hat{I}_T$. It further follows that $\hat{X}_T$ is smooth. ■

One can further show that for a certain lexicographic order the generators of $\hat{I}_G$ form a Gröbner basis. Set $d_{ij} = \min_{\pi \in \mathcal{D}(i,j|\emptyset)} |\pi|$, that is d_{ij} is the length of the shortest d-path from i to j. Note that if $i \perp\!\!\!\perp j|k$ is an imposed relation, then $d_{ij} > 1$. Order the variables as follows: $\sigma_{ij} \succ \sigma_{i'j'}$ if $d_{ij} > d_{i'j'}$ or $d_{ij} = d_{i'j'}$ but $(i,j) \succ_{\mathbb{Z}^2} (i'j')$, where $\succ_{\mathbb{Z}^2}$ is any order on $\mathbb{Z}^2$. We also set $\sigma_{ij} \succ 1$ for all variables.
Let $\alpha = (\alpha_{ij})$ be a vector. Denote by σ^α the monomial $\prod \sigma_{ij}^{\alpha_{ij}}$. Now define the relation $\sigma^\alpha \succ_{\text{dag}} \sigma^\beta$ if the first non-zero coordinate of $\alpha - \beta$ is positive. Note that $\succ_{\text{dag}}$ is a lexicographic order.

The above order has a pleasant property: the leading monomial of quadratic relations $\sigma_{ij} - \sigma_{ik}\sigma_{kj}$ generated by CI relations of T is always the linear form σ_{ij}. We use this to prove that the quadratic and linear imposed relations used to generate $\hat{I}'_T$ form a Gröbner basis w.r.t to this order:

Proposition 8. *The generators of $\hat{I}'_T$ form a Gröbner basis w.r.t the $\succ_{\text{dag}}$ oder.*

Proof. Note that $\hat{I}'_G$ is generated by linear forms and quadratic relations of the form $\sigma_{ij} - \sigma_{ik}\sigma_{kj}$. Given two such polynomials, $f = \sigma_{ij} - \sigma_{ik}\sigma_{kj}$ and $f' = \sigma_{i'j'} - \sigma_{i'k'}\sigma_{k'j'}$, we can check that the leading terms are linear, and hence, the resulting S-polynomial

$$S(f, f') = \sigma_{ik}\sigma_{kj}\sigma_{i'j'} - \sigma_{i'k'}\sigma_{k'j'}\sigma_{ij},$$

reduces to zero w.r.t $\{\sigma_{ij} - \sigma_{ik}\sigma_{kj}, \sigma_{i'j'} - \sigma_{i'k'}\sigma_{k'j'}\}$. ■

It was asked in [3] (see Conjecture 5.9) if there exists a Gröbner basis consisting of square free terms of degree one and two for J_T. The above proposition shows that this is the case for $\hat{I}_T$ (or $\hat{J}_T$).

Finally, using the procedure in section 2.7, we list the isomorphism classes of trees on $4, 5$, and 6 nodes. See Figure 1. Our computations are done in Magma.

4. Marginalization

Exact inference on a DAG G is the problem of extracting marginals of a G-compatible distribution. In this section, we establish a connection between the obstructions to embeddings of Y_G and obstructions to efficient inference on G.

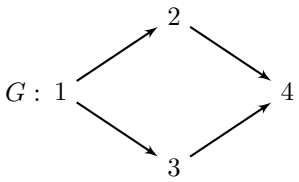
Definition 4. *1. Given a DAG G, we define marginalization of G w.r.t N, denoted by $\text{El}_N(G)$, to be the conditional independence model defined by the implied relations in G*

$$i \perp\!\!\!\perp j|K$$

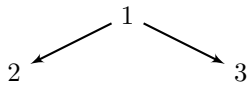
such that $\{i, j, K\} \cap N = \emptyset$.

2. *Let $e = |N|$. Given a DAG G, define the extension G^e of G w.r.t N to be the DAG on $n + e$ nodes whose implied relations are the same as G.*

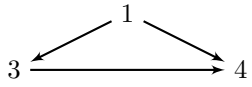
Take for instance the DAG



To extract the marginal on nodes 1 and 3, one needs to eliminate nodes 2 and 4. A well known problem is that the order of elimination matters. For instance, eliminating node 4 leads to a "sparse" graph

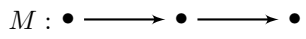


whereas eliminating node 2 gives a "dense" graph



A fact is that the computational effort in sequential eliminations is controlled by the sparsity of the subgraphs that appear in the process. The problem is that there is no good way of finding the right elimination order. What is worse is that, by the recent results in [13], it is even hard to say if such an order exists or not. The difficulty of inference in a DAG is controlled by the so called tree-widths of its underlying graph. This is hard to compute for large graphs, and it thus makes sense to settle for an easier question: *what are some necessary conditions for a good elimination order to exist?*

One observation is that having small tree-width implies that, under some suitable ordering, the graphs that appear in the elimination process have small tree-width as well, and one can take this as a measure of efficiency for inference. In the above example, we see that the marginals of interest in G are supported on the (unlabeled) Markov chain:



However, the same cannot be said about the complete graph on four nodes. It is thus useful to know how simple the space of models that support the marginals of a certain DAG can be. This gives us a notion of complexity of marginals in DAGs and it is clear that this notion depends on the class of the model, and not on how the model is represented as a graph. We formalize this notion as follows:

Definition 5. *(a) Suppose that the nodes in M are a subset of the nodes in G. We say that M lies below G if*

$$\forall Q_{YX} \in \mathrm{Loc}(G) \quad \exists \quad P_X \in \mathrm{Loc}(M) \quad s.t. \quad Q_{YX}(\mathcal{A}) = \int_{\mathcal{A}} K_{Y|X} dP_X$$

Likewise, we say that G lies above M.

(b) M is said to be minimal w.r.t the above property if

$$\mathrm{Loc}(M') \not\subset \mathrm{Loc}(M)$$

for all M' that satisfy (a).

(c) The class $[M]$ is said to lie below $[G]$ if (a) holds after some relabeling of variables (in G or M).

In other words, M lies below G if the marginal of any G-compatible distribution on the subspace associated to M factorizes w.r.t it. We note that the elimination of a DAG need not be a DAG (see [30]), but in cases that it is, the notions of minimal model and elimination model coincide.

The main question of interest is to decide when $[M]$ can lie below $[G]$. It is clear that an enumerative approach to answer this question is problematic, even for a small graph M, as the number of possibilities grow exponentially with the size of the eliminated subset (which grows as M gets smaller). The next proposition gives a necessary condition.

Proposition 9. *Let M, G be DAGs and suppose $e := |V_G| - |V_M| \geq 0$. Then if $[M]$ lies below $[G]$, there exists an embedding*[9] *$Y_G \hookrightarrow Y_{M^e}$.*

Proof. If M lies below G, then its implied relations are satisfied by the marginals in $\mathrm{Loc}(G)$. This gives an inclusion $\mathrm{loc}(G) \subset \mathrm{loc}(M^e)$, which induces an embedding $\hat{X}_G \hookrightarrow \hat{X}_M$. This embedding extends to the projective closures. ■

Example 6. *Let G be the complete DAG on n nodes and M be a Markov chain on $m \leq n$ nodes. It is clear, from dimension considerations, that $[M]$ cannot appear below $[G]$.*

We use the following proposition later:

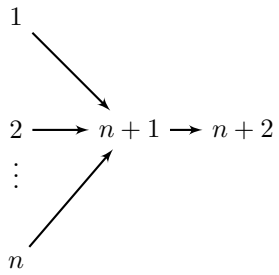
Proposition 10. *If there exists an embedding of $\mathbb{P}^n \times \mathbb{P}^1$ into $\mathbb{P}^m$, then $m \geq 2n+1$.*

Proof. This is perhaps easiest to see from Bezout's theorem, which states any two closed sub-varieties of $\mathbb{P}^m$ of complementary dimension must have a non-empty intersection. In $\mathbb{P}^n \times \mathbb{P}^1$, there are n-dimensional hyperplanes $\mathbb{P}^n \times \{pt\}$ that do no intersect, and this remains true after embedding. Thus each hyperplane must have codimension at least $n+1$ in the ambient projective space. ■

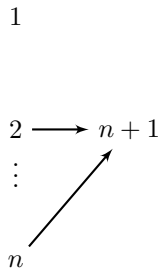
[9] By an embedding we mean a closed immersion.

This proposition shows that when $m = 1$ the Segre embedding $\mathbb{P}^n \times \mathbb{P}^m \hookrightarrow \mathbb{P}^{nm+n+m+1}$ uses the minimal target dimension[10]. This can be useful for our purposes.

Example 7. *Let G_n be the following family of DAGs*



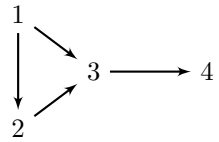
and M_n be a disconnected V-structure on $n+1$ nodes



One can check that $Y_{G_n} \simeq \mathbb{P}^n \times \mathbb{P}^1$ and that $Y_{M_n^e} \simeq \mathbb{P}^{2n}$. Then $[M_n]$ does not lie below $[G_n]$ by the above proposition. This example is tight, in the sense that $[M_{n-1}]$ does lie below $[G_n]$ as eliminating the nodes $n+1$ and $n+2$ makes G_n completely disconnected. In particular, G_2 does not lie above $\bullet \quad \bullet \to \bullet$.

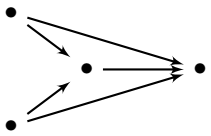
The examples above are special in that they all involve smooth projective varieties, and this is hardly an attribute of projective DAG varieties. Additional effort will be needed to deal with singularities.

Example 8. *Consider the DAG*



We want to show that for all n, $[G]$ does not lie above $[M]$ where $[M]$ is the class of a V-structure $\bullet \to \bullet \leftarrow \bullet$. The extension $[M^e]$ of a V-structure to four nodes is the DAG

[10] In the general case, there are embeddings $\mathbb{P}^n \times \mathbb{P}^m \hookrightarrow \mathbb{P}^{2(m+n)-1}$ and these can be shown to have the smallest target dimension.



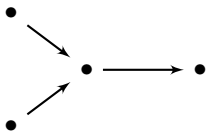
Suppose that $[M]$ is below $[G]$ to obtain an embedding $\iota : Y_G \hookrightarrow Y_{M^e}$. We note that Π_G is smooth everywhere except at the vertex of the affine cone over $\mathbb{P}^1 \times \mathbb{P}^1$. Blowing up $Y_{M^e} \simeq \mathbb{P}^5$ at $\iota(p)$ gives a commutative diagram:

$$\begin{array}{ccc} \mathrm{Bl}_p(Y_G) & \longrightarrow & \mathrm{Bl}_{\iota(p)}(Y_{M^e}) \\ \downarrow & & \downarrow \\ Y_G & \longrightarrow & Y_{M^e} \end{array}$$

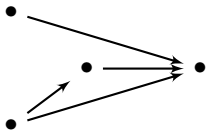
Since the property of being a closed immersion is stable under base change, it follows that the map $\mathrm{Bl}_p(Y_G) \to \mathrm{Bl}_{\iota(p)}(\mathbb{P}^5)$ is an embedding (see corollary II.7.15 in [31])We note that $\mathrm{Bl}_p(\Pi_G)$ is smooth, and has the structure of a $\mathbb{P}^1$-bundle over $\mathbb{P}^2 \times \mathbb{P}^1$. Likewise, the blow-up $\mathrm{Bl}_{\iota(p)}(\mathbb{P}^5)$ is a $\mathbb{P}^1$-bundle over $\mathbb{P}^4$. In other words, after blow ups, we get projective bundles over DAG varieties

$$\begin{array}{ccc} \mathrm{Bl}_p(Y_G) & \longrightarrow & \mathrm{Bl}_{\iota(p)}(Y_{M^e}) \\ \mathbb{P}^1\text{-}bundle \downarrow & & \downarrow \mathbb{P}^1\text{-}bundle \\ Y_{G'} & \longrightarrow & Y_{M'} \end{array}$$

where G' looks like



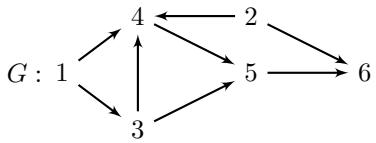
and M' is



One can check that $\mathrm{Pic}(\mathrm{Bl}_{\iota(p)}(Y_G)) = \mathbb{Z}^3$ while $\mathrm{Pic}(\mathrm{Bl}_p(Y_{M^e})) = \mathbb{Z}^2$. If an embedding exists, $\mathrm{Bl}(Y_G)$ must be isomorphic to a smooth codimension one subvariety in $\mathrm{Bl}_p(Y_{M^e})$, which has Picard group isomorphic to $\mathbb{Z}^2$ by the Lefschetz's hyperplane theorem. This example shows that if two Markov chain relations fit in a DAG, they cannot combine to produce a pure independence relation. The fact that the relations fit in a DAG is essential here.

It is common to encounter DAG varieties that do not have a simple geometric description as in the above families. One can work with numeric invariants in this case.

Example 9. *Let us check if the DAG*



lies above $M : \bullet \to \bullet \to \bullet$. *There is an embedding* $\iota : Y_G \hookrightarrow \mathbb{P}^{15}$ *defined by the ideal sheaf* $\mathcal{I}_G$. *If* $[G]$ *lies above* $[M]$, *then there must exist an irreducible quadric hyper-surface in* $\mathbb{P}^{15}$ *that contains the image of* ι. *This means that the map* $\Gamma(\mathbb{P}^{15}, \mathcal{O}_{\mathbb{P}^{15}}(2)) \to \Gamma(\mathbb{P}^{15}, \iota_* \mathcal{O}_{Y_G}(2))$ *has an irreducible quadratic polynomial in its kernel, which corresponds to a global section of* $\mathcal{I}_G(2)$. *Using Magma, we compute the Hilbert polynomial of* $\mathcal{I}_G$ *and conclude that* $h^0\mathcal{I}_G(1) = 2$ *and* $h^0\mathcal{I}_G(2) = 31$. *Thus there are 2 linearly independent hyperplanes*[11] *that contain the image of* ι. *These in turn give 31 reducible (and thus no irreducible) quadrics containing the image. This shows that* $[G]$ *does not lie above* $[M]$.

Proposition 11. *Let* M *be a Markov chain on three nodes and* G *be any DAG on* n *nodes. Set* $a_i := h^0(\mathcal{I}_G(i))$. *Then* $[G]$ *lies above* $[M]$ *if and only if*

$$a_2 - \frac{a_1}{2}(n^2 - n - a_1 + 3) > 0.$$

This section justifies the following questions: 1) *what are some invariants of DAG varieties that are useful for ruling out embeddings of the above type?* 2) *how are such invariants related to the combinatorial data of the DAG?* 3) *when can the numbers* a_i *be read from the Hilbert polynomial of* $\mathcal{I}_G$*?* We plan to come back to these questions in a future paper.

References

[1] S. L. Lauritzen, *Graphical Models*. Oxford University Press, 1996.

[2] M. Levitz, M. D. Perlman, and D. Madigan, "Separation and completeness properties for amp chain graph Markov models," *Ann. Stat.*, pp. 1751–1784, 2001.

[11] Two hyperplanes are said to be linearly independent if they are defined by linearly independent sections of $O(1)$.

[3] S. Sullivant, "Algebraic geometry of Gaussian Bayesian networks," *Adv. in Appl. Math.*, vol. 40, no. 4, pp. 482–513, 2008.

[4] S. A. Andersson, D. Madigan, and M. D. Perlman, "A characterization of Markov equivalence classes for acyclic digraphs," *Ann. Stat.*, vol. 25, no. 2, pp. 505–541, 1997.

[5] D. M. Chickering, "Learning equivalence classes of Bayesian-network structures," *J. Mach. Learn. Res.*, vol. 2, pp. 445–498, Mar. 2002. [Online]. Available: http://dx.doi.org/10.1162/153244302760200696

[6] B. Steinsky, "Enumeration of labelled chain graphs and labelled essential directed acyclic graphs," *Discrete Mathematics*, vol. 270, no. 1, pp. 267–278, 2003.

[7] S. Wagner, "Asymptotic enumeration of extensional acyclic digraphs," *Algorithmica*, vol. 66, no. 4, pp. 829–847, 2013.

[8] S. B. Gillispie and M. D. Perlman, "Enumerating markov equivalence classes of acyclic digraph dels," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2001, pp. 171–177.

[9] J. M. Peña, "Approximate counting of graphical models via mcmc revisited," in *Advances in Artificial Intelligence*. Springer, 2013, pp. 383–392.

[10] T. Verma and J. Pearl, "Equivalence and synthesis of causal models," in *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, ser. UAI '90. New York, NY, USA: Elsevier Science Inc., 1991, pp. 255–270. [Online]. Available: http://dl.acm.org/citation.cfm?id=647233.719736

[11] C. Meek, "Causal inference and causal explanation with background knowledge," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1995, pp. 403–410.

[12] M. Studenỳ, R. Hemmecke, and S. Lindner, "Characteristic imset: a simple algebraic representative of a Bayesian network structure," in *Proceedings of the 5th European workshop on probabilistic graphical models*, 2010, pp. 257–264.

[13] J. Kwisthout, H. L. Bodlaender, and L. C. van der Gaag, "The necessity of bounded treewidth for efficient inference in bayesian networks." in *ECAI*, vol. 215, 2010, pp. 237–242.

[14] R. Lněnička and F. Matúš, "On Gaussian conditional independence structures," *Kybernetika*, vol. 43, no. 3, pp. 327–342, 2007.

[15] F. Matús and M. Studenỳ, "Conditional independences among four random variables I," *Combin. Prob. Comput.*, vol. 4, pp. 269–278, 1995.

[16] F. Matús, "Conditional independences among four random variables II," *Combin. Prob. Comput.*, vol. 4, pp. 407–418, 1995.

[17] F. Matúš, "Conditional independences among four random variables III: final conclusion," *Combin. Prob. Comput.*, vol. 8, no. 03, pp. 269–276, 1999.

[18] P. Šimecek, "A short note on discrete representability of independence models," in *Proccedings of the European Workshop on Probabilistic Graphical Models*. Citeseer, 2006, pp. 287–292.

[19] ——, "Gaussian representation of independence models over four random variables," in *COMPSTAT conference*, 2006.

[20] M. Drton, B. Sturmfels, and S. Sullivant, *Lectures on Algebraic Statistics*, ser. Oberwolfach Seminars. Birkhäuser Basel, 2008.

[21] M. Drton and H. Xiao, "Smoothness of Gaussian conditional independence models," *Algebraic Methods in Statistics and Probability II*, vol. 516, pp. 155–177, 2009.

[22] F. Matúš, "Conditional independence structures examined via minors," *Ann. Math. Artif. Intell.*, vol. 21, no. 1, 1997.

[23] L. Garcia, M. Stillman, and B. Sturmfels, "Algebraic geometry of Bayesian networks," *J. of Symb. Comput.*, vol. 39, no. 3, pp. 331–355, 2005.

[24] E. Çinlar, *Probability and Stochastics*. Springer Science+ Business Media, 2011, vol. 261.

[25] M. Studeny, *Probabilistic Conditional Independence Structures*. Springer, 2005.

[26] C. Uhler, G. Raskutti, P. Bühlmann, and B. Yu, "Geometry of the faithfulness assumption in causal inference," *Ann. Stat.*, vol. 41, no. 2, pp. 436–463, 2013.

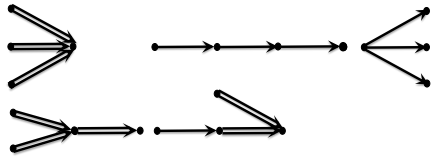
[27] M. F. Atiyah and I. G. Macdonald, *Introduction to Commutative Algebra*. Springer, 1969, vol. 2.

[28] F. Matúš, "Conditional independences in gaussian vectors and rings of polynomials," in *Conditionals, Information, and Inference*. Springer, 2005, pp. 152–161.

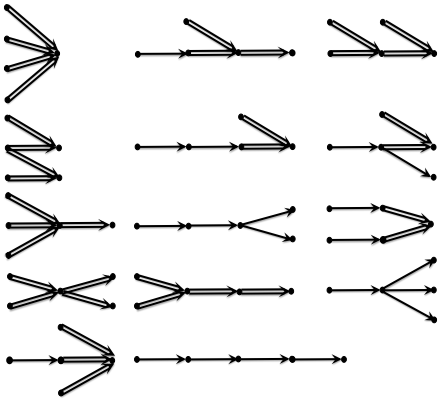
[29] J. Bochnak, M. Coste, and M.-F. Roy, *Real Algebraic Geometry*. Springer Berlin, 1998, vol. 95.

[30] T. Richardson and P. Spirtes, "Ancestral graph markov models," *Annals of Statistics*, pp. 962–1030, 2002.

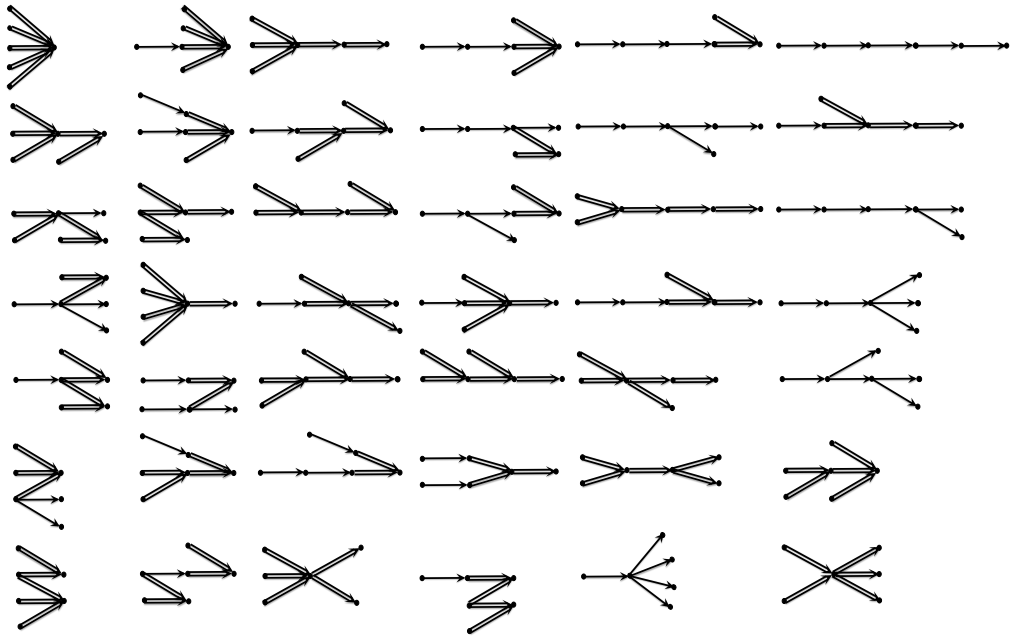
[31] R. Hartshorne, *Algebraic Geometry*. Springer, 1977, vol. 52.



(a) $n = 4$



(b) $n = 5$



(c) $n = 6$

Figure 1: Isomorphism classes of directed tree models on $4, 5$, and 6 nodes. The double arrows are essential.