

MIT Open Access Articles

Algebraic methods of classifying directed graphical models

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Roozbehani, Hajir, and Yury Polyanskiy. "Algebraic Methods of Classifying Directed Graphical Models." 2014 IEEE International Symposium on Information Theory (June 2014).

As Published: <http://dx.doi.org/10.1109/ISIT.2014.6875189>

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: <http://hdl.handle.net/1721.1/100995>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Algebraic Methods of Classifying Directed Graphical Models

Hajir Roozbehani and Yury Polyanskiy

Abstract—In information theory, structural system constraints are frequently described in the form of a directed acyclic graphical model (DAG). 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 Sullivan. 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 and 5 nodes.

I. INTRODUCTION

Before formal treatment, we give a high level overview of this paper. Consider two directed graphical models (or directed acyclic graphs, DAGs) on random variables (A, B, C) :

$$A \rightarrow B \rightarrow C \quad B \leftarrow A \rightarrow C \quad (1)$$

(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 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

$$B \rightarrow C \leftarrow A \quad (2)$$

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. 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:

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

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

H.R. is with the Department of Aeronautics and Astronautics at MIT. e-mail: hajir@mit.edu Y.P. is with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139 USA. e-mail: yp@mit.edu.

This material is based upon work supported by the National Science Foundation CAREER award under grant agreement CCF-12-53205, and by Siemens through the MIT Energy Initiative.

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

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

We give an *analytic* result: while $\text{loc}(G)$ 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 $\text{loc}(G)$ (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 Sullivan [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 and 5 nodes.

The question of DAG isomorphism does not seem to have appeared elsewhere, though the closely related question of DAG equality (also known as 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. This has motivated the need to represent DAGs, and among the representatives that are relevant in this regard are the essential graphs [4] and the characteristic imsets [6]. Both these methods have a combinatorial flavor and this work provides an algebraic alternative. The word algebraic here means commutative-algebraic, unlike in [6].

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 [7]–[16]. Some of our preparatory

propositions can be found in the literature in slightly weaker forms and we attempt to give references. The main novelty is that we essentially leverage the directed-graph structure of the model (as opposed to general CI model) to infer stronger algebraic claims. In particular, our treatment is base independent – although for readability we present results for the varieties over \mathbb{C} .

A. Preliminaries

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

Definition 1: A directed acyclic graphical model (DAG) \mathcal{E}/k is the data:

- A set of indices $[n] := \{1, \dots, n\}$ 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 $\mathcal{M}_{\mathcal{E}}$ of *imposed* (a.k.a local Markov) relations

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

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 $\mathcal{M}_{\mathcal{E}}^{\text{topo}}$ of topologically sorted local Markov relations:

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

- A set of \mathcal{E} -compatible joint probability distributions

$\text{Loc}(\mathcal{E}) := \{P_X | I \perp\!\!\!\perp J | K \in \mathcal{M}_{\mathcal{E}} \Rightarrow X_I \perp\!\!\!\perp X_J | X_K\}$, where X is a k^n -valued random variable¹.

- A set of *implied* relations

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

Given a collection of such models, it is often of interest to find representatives for their isomorphism classes (see also [7], [8])– these are models that have the same compatible distributions modulo labelings of variables:

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

$$\text{Loc}(\mathcal{E}) \cap \mathcal{Q} = \text{Loc}(\mathcal{E}') \cap \mathcal{Q}.$$

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

$$p_{X_1 \dots X_n} \in \text{Loc}(\mathcal{E}) \cap \mathcal{Q} \iff p_{X_{\pi(1)} \dots X_{\pi(n)}} \in \text{Loc}(\mathcal{E}') \cap \mathcal{Q}$$

for some permutation π of indices. When \mathcal{Q} is the set of all distributions, we call such models isomorphic.

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 [8]–[10], positive discrete random variables [11], non-singular Gaussians [7], and general Gaussians [12].

Let $H = H_1 \times \dots \times H_n$ be a product measure space endowed with the σ -algebra $\mathcal{H} = \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n$. 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

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

where $K_{i|\mathbf{pa}(i)}$'s are conditional probability kernels (which exist by [17, Theorem 2.7]) and $K_{i|\mathbf{pa}(i)}(dx_i | x_{\mathbf{pa}(i)}) = \mu_i(dx_i)$ if i has no parents in G .

Given a DAG G , we denote by $\text{Fac}(G)$ the set of distributions that factorize w.r.t G . We show in Section II.A that

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

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).

B. 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 .
- Σ^\bullet is the subset of matrices in Σ with non-zero principal minors².
- $\hat{\Sigma}$ 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 $\hat{\Sigma}^\bullet \triangleq \hat{\Sigma} \cap \Sigma^\bullet$.
- $\text{loc}(G)$ is the set of covariance matrices in $\text{Loc}(G) \cap \mathcal{N}$.
- f_G is the rational parametrization defined in II.B.
- Given $S \subset \Sigma$, $[S]$ and $[S]_Z$ are its standard and Zariski closures³, 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 $V(I) = \{x \in \mathbb{C}^n | f(x) = 0 \quad \forall f \in I\}$.
- $X_G \triangleq [\text{loc}(G)]_Z$, $\mathfrak{p}_G \triangleq I(X_G)$, $\hat{X}_G \triangleq [\text{loc}(G) \cap \hat{\Sigma}]_Z$.

C. 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}^+ -equality reduces to checking equality of algebraic subsets inside the positive definite cone (see for instance [3], [13]–[15]). This follows from the next proposition:

¹We 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 \mathcal{E}/k if we need to emphasize the base field k .

²Note that Σ^\bullet is Zariski open, while Σ^+ , Σ^{++} are described by inequalities.

³The closure is always taken inside the affine complex space.

Proposition 1 (Lemma 2.8 in [18]): 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-singular⁴.

Proposition 1 enables us to think algebraically and/or geometrically when deciding Gaussian equality. Indeed, it states that $\text{loc}(G) \cap \Sigma^{++}$ 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 $\text{loc}(G)$.

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 topsorted 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

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

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

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

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

It is known that $[\text{loc}(G) \cap \Sigma^{++}]_Z$ 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 $\mathbb{C}[\Sigma]$:

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

where $S = \{\theta_0^n | n > 0\}$. One can replace $(\text{loc}(G), \Sigma)$ with $(\text{loc}(G) \cap \hat{\Sigma}, \hat{\Sigma})$ in the above. We note that the above

⁴A vector random variable is said to be non-singular if its distribution admits a density w.r.t. product Lebesgue measure.

relations hold verbatim over $\mathbb{Z}[\Sigma]$ and other base rings. Our main statement, shown in II-D, 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 II-E. We then use the algorithm to list the isomorphism classes of directed tree models for $n = 4$ and 5 nodes. We also include the list for $n = 6$ in the extended version of the paper. There, we further discuss some special properties of directed tree models. In particular, we show that \hat{I}_T is a prime ideal for a tree model T and hence $\hat{I}_T = I(\text{loc}(T) \cap \hat{\Sigma})$. This is analogous to primality of J_T , the ideal of implied relations, shown in [3] (see Corollary 2.4 and Theorem 5.8).

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.

II. MAIN RESULTS

A. 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 Lauritzen (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, $\text{Fac}(G) = \text{Loc}(G)$.

Proof: see [19]. ■

B. Weak limits of factorable Gaussians

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

To characterize $\text{loc}(G) \cap \partial \mathcal{N}^+$, we shall find it useful to work with the parametrization

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

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 $\text{cov}(X)$. Indeed, starting from (7), one can write

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

where $\gamma_{ik} = \text{Cov}(Z_i, X_k)$, $\sigma_{ik} = \text{Cov}(X_i, X_k)$. 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 $[\text{loc}(G) \cap \Sigma^{++}]_Z$:

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

The next Proposition shows that

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

Proposition 4: Let G be a DAG. Then

- (a) $\text{loc}(G) \cap \Sigma^{++}$ is dense in $\text{loc}(G)$.
- (b) $\text{loc}(G) \cap \hat{\Sigma}^{++}$ is dense in $\text{loc}(G) \cap \hat{\Sigma}$.

Proof: see [19]. ■

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 ($\text{loc}(G)$ 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 $\text{loc}(G) \cap \hat{\Sigma}^{++}$ strictly contains $\text{loc}(G) \cap \hat{\Sigma}$.

Remark 2: 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) limit⁵.

C. DAG varieties and ideals

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

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

- (a) There is a Zariski closed subset Y_G so that

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

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

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

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

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

- (c) X_G is smooth inside Σ^\bullet .

Remark 3:

- (a) In Theorem 1b, we can replace I_G with J_G . Analogous statements hold over $\mathbb{C}[\hat{\Sigma}]$ as shown in [19].

- (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 [16] 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. [20, 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 :

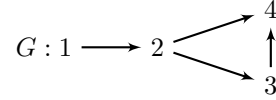
$$I_G = \mathfrak{p}_G \cap \mathfrak{q}_1 \cdots \mathfrak{q}_r,$$

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).

Proof: see [19] for details. ■

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

Example 2: 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 Σ^\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 $f = \sigma_{22}(\sigma_{22}\sigma_{33} - \sigma_{23}^2)$.

D. 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 equality of DAGs:

Proposition 5: 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: see [19]. ■

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 2: Let G, G' be DAGs and S be as in Theorem 1. Then G and G' are isomorphic if and only if

$$S^{-1}I_G \cap \mathbb{C}[\Sigma]^\Pi = S^{-1}I_{G'} \cap \mathbb{C}[\Sigma]^\Pi. \quad (9)$$

Proof: see [19]. ■

⁵This follows directly from the lower semi-continuity of divergence.

```

1: function ISODAGm( $G, G'$ )
2: Sort  $G$  and  $G'$  topologically
3: Initialize ISO  $\leftarrow$  true,  $r \leftarrow 1$ 
4: while ISO and  $r \leq m$  do
5:   Sample  $z_G^r, z_{G'}^r$  respectively from  $\hat{\varphi}_* P_G, \hat{\varphi}_* P_{G'}$  as follows:
6:     (i) Sample edge variables  $\hat{\sigma}_{\text{edge}}^r$  of  $G$  from  $P_G$ 
7:     (ii) for  $i := 2$  to  $n$  do
8:       Solve the (linear) toposorted imposed relations
9:         
$$|\hat{\sigma}_{iK,jK}^r| = 0, \quad K := \mathbf{pa}(i)$$

10:        for each non-edge variable  $\hat{\sigma}_{ij}^r, j < i$ 
11:     (iii) end for
12:     (iv)  $z_G^r \leftarrow (\hat{\sigma}_{\text{edge}}^r, \hat{\sigma}_{\text{non-edge}}^r)$ 
13:     (v) Repeat for  $G'$ 
14:   if  $\Pi(z_G^r) \cap V(\hat{I}_{G'}) = \emptyset$  or  $\Pi(z_{G'}^r) \cap V(\hat{I}_G) = \emptyset$  then
15:     ISO  $\leftarrow$  false
16:   end if
17:    $r \leftarrow r + 1$ 
18: end while
19: return ISO

```

We use the preceding results to give a randomized algorithm for testing DAG isomorphism. In [19], we associate with every DAG G a rational map (see proof of Theorem 1) $\varphi: \Sigma_{edge} \rightarrow \Sigma$ where φ is regular on a distinguished open $D(gh)$ and its image is dense in X_G . Here, g and h are certain polynomials in $\mathbb{C}[\Sigma_{edge}]$. Denote by $\hat{\varphi}: \hat{\Sigma}_{edge} \rightarrow \hat{X}_G$ the map $\varphi|_{\hat{\Sigma}}$. Let $\mathcal{Y} := \mathbb{F}_q^{[E]}$ and define

where $\hat{g} = g|_{\hat{\Sigma}}$, $\hat{h} = h|_{\hat{\Sigma}}$. Now construct a random matrix with uniform distribution P_G on the finite set \mathcal{U} . 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 $\hat{\varphi}_* P_G$.

Theorem 3: Let G be a DAG on n nodes and E be its set of edges. Let Z_G^i be as in above and set $d := \deg(\hat{g}) + \deg(\hat{h})$. If $G \sim G'$, then

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

REFERENCES

-
- (a) $n = 4$
- (b) $n = 5$

- [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] M. Studený, 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.
- [7] R. Lněnička and F. Matúš, “On Gaussian conditional independence structures,” *Kybernetika*, vol. 43, no. 3, pp. 327–342, 2007.
- [8] F. Matúš and M. Studený, “Conditional independences among four random variables I,” *Combin. Prob. Comput.*, vol. 4, pp. 269–278, 1995.
- [9] F. Matúš, “Conditional independences among four random variables II,” *Combin. Prob. Comput.*, vol. 4, pp. 407–418, 1995.
- [10] F. Matúš, “Conditional independences among four random variables III: final conclusion,” *Combin. Prob. Comput.*, vol. 8, no. 03, pp. 269–276, 1999.
- [11] P. Šimeček, “A short note on discrete representability of independence models,” in *Proceedings of the European Workshop on Probabilistic Graphical Models*. Citeseer, 2006, pp. 287–292.
- [12] —, “Gaussian representation of independence models over four random variables,” in *COMPSTAT conference*, 2006.
- [13] M. Drton, B. Sturmfels, and S. Sullivant, *Lectures on Algebraic Statistics*, ser. Oberwolfach Seminars. Birkhäuser Basel, 2008.
- [14] M. Drton and H. Xiao, “Smoothness of Gaussian conditional independence models,” *Algebraic Methods in Statistics and Probability II*, vol. 516, pp. 155–177, 2009.
- [15] F. Matúš, “Conditional independence structures examined via minors,” *Ann. Math. Artif. Intell.*, vol. 21, no. 1, 1997.
- [16] L. Garcia, M. Stillman, and B. Sturmfels, “Algebraic geometry of Bayesian networks,” *J. of Symb. Comput.*, vol. 39, no. 3, pp. 331–355, 2005.
- [17] E. Çinlar, *Probability and Stochastics*. Springer Science+ Business Media, 2011, vol. 261.
- [18] M. Studený, *Probabilistic Conditional Independence Structures*. Springer, 2005.
- [19] H. Roozbehani and Y. Polyanskiy, “Algebraic methods of classifying directed graphical models (extended),” *ArXiv preprint arXiv:1401.5551*, Jan. 2014.
- [20] M. F. Atiyah and I. G. Macdonald, *Introduction to Commutative Algebra*. Springer, 1969, vol. 2.