

# Algebraic Properties of HTC-identifiable Graphs

Bohao Yao, Robin Evans

April 4, 2022

## Abstract

HTC-identifiable graphs are a large family of graphs known to be generically identifiable. In this paper, we explore some algebraic properties of linear structural equation models that can be represented by an HTC-identifiable graph. In particular, we prove that all mixed graphs are HTC-identifiable if and only if all the regression coefficients can be recovered from the covariance matrix using straightforward linear algebra operations on specified equations. Furthermore, given an HTC-identifiable graph, we provide a set of polynomials that generates the ideal that encompasses all the equality constraints of the corresponding graphical model on the cone of positive definite matrices. We further prove that, for a subset of HTC-identifiable graphs, this set of polynomials are the minimal generators of said ideal.

## 1 Introduction

It is often natural to model the joint distribution of a random vector  $X = (X_1, \dots, X_n)^T$  as a collection of noisy linear interdependencies. In particular, we may assume that each variable  $X_i$  can be expressed as a linear function of the remaining variables and a stochastic noise term  $\epsilon_i \sim N(0, \omega_{ii})$ ,

$$X_i = \sum_{j \in \text{pa}(i)} \lambda_{ji} X_j + \epsilon_i. \quad (1)$$

In matrix form, this system of structural equations could be rewritten as

$$X = \Lambda^T X + \epsilon \quad (2)$$

where  $\Lambda = (\lambda_{ij})$  is a matrix of coefficients  $\lambda_{ij}$ , and  $\epsilon = (\epsilon_i)$  is an error vector with a multivariate normal distribution and covariance  $\Omega = (\omega_{ij})$ . Models of this form are known as *linear structural equation models* (SEMs). Assuming  $(I - \Lambda)$  is invertible, we can rearrange (2) to obtain a covariance matrix for  $X$

$$\Sigma := \text{Cov}[X] = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}. \quad (3)$$

Background on SEMs can be found in [1, 11, 16]. In the present day, SEMs are applied in subject fields as varied as epidemiology, ecology, behavioural sciences, social sciences and economics [9, 10, 14, 15]. The properties of identifiability in SEMs is a topic with a long history. A review of the classical conditions of these properties can be found in [1]. We shall start with the definition of *global identifiability*.

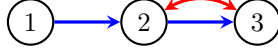


Figure 1: The instrumental variable model

**Definition 1.1.** Let  $\phi : \Theta \rightarrow N$  be a rational map defined everywhere on the parameter space  $\Theta$  into the natural parameter space  $N$  of an exponential family. The model  $\mathcal{M} = \text{im } \phi$  is said to be *globally identifiable* if  $\phi$  is a one-to-one map on  $\Theta$ .

Necessary and sufficient graphical conditions for linear SEMs to be globally identifiable were found by Drton et al. [5]. Other recent works, such as those by Brito and Pearl [2], consider the weaker identifiability requirement of *generic identifiability*.

**Definition 1.2.** Let  $\phi : \Theta \rightarrow N$  be a rational map defined everywhere on the parameter space  $\Theta$  into the natural parameter space  $N$  of an exponential family. The model  $\mathcal{M} = \text{im } \phi$  is said to be *generically identifiable* if  $\phi^{-1}(\phi(\theta)) = \{\theta\}$  for almost all  $\theta \in \Theta$  with respect to the Lebesgue measure.

From the definitions, we see that all globally identifiable models are also generically identifiable. However, not all generically identifiable models are also globally identifiable as we shall see in the following example from [4, Example 3.1].

**Example 1.3.** Consider the instrumental variable model in Figure 1. We can recover the observable  $\Lambda$  from the observed  $\Sigma$  using

$$\lambda_{12} = \frac{\sigma_{12}}{\sigma_{11}}, \quad \lambda_{23} = \frac{\sigma_{13}}{\sigma_{12}}.$$

Note that the denominator in the first equation is always positive since  $\Sigma$  is positive definite and the denominator in the second equation is zero if and only if  $\lambda_{12} = 0$ . Therefore, while the instrumental variable model is generically identifiable, it is not globally identifiable as it is not injective on the whole of  $\Theta$  (i.e. when  $\lambda_{12} \neq 0$ ).

The first class of graphs that were proven to be generically identifiable is the bow-free acyclic directed mixed graphs by Brito and Pearl [2]. Foygel et al. [7] later introduced a new identifiability criterion known as HTC-identifiability and proved that all HTC-identifiable graphs are also generically identifiable. As all acyclic bow-free graphs are HTC-identifiable, hence the result obtained by Foygel et al. could be seen as an extension to that of Brito and Pearl. At the time of writing, there is no known graphical criterion which is both necessary and sufficient for an SEM to be generically identifiable. We shall provide a rigorous definition of HTC-identifiability in the next section.

As a large class of generically identifiable graphs, HTC-identifiable graphs warrant a thorough study. In this paper, we will explore the algebraic structure of HTC-identifiable graphs by introducing an identifiability criterion which we call *linear identifiability*. Linearly identifiable graphs satisfy certain algebraic structures which could be used to recover  $\Lambda$  from  $\Sigma$  efficiently using linear algebra. We will then prove that linear identifiability and HTC-identifiability are equivalent:

**Theorem 3.17.** For any mixed graph  $G$ ,  $\mathcal{M}(G)$  is linearly identifiable if and only if  $\mathcal{M}(G)$  is HTC-identifiable.

Hence, for HTC-identifiable graphs, we have algorithms which are much more efficient in recovering graphical parameters,  $(\Lambda, \Omega)$ , relative to graphs which are not HTC-identifiable.

This paper is structured as follows: In Section 2, we will provide the preliminaries on structural equation models with a focus on treks and constraints. In Section 3, we will prove that linearly identifiable and HTC-identifiable are equivalent. In Section 4, we will exploit the algebraic properties of HTC-identifiable graphs we found in Section 3 to provide an efficient algorithm to recover the parameters,  $(\Lambda, \Omega)$ , of any HTC-identifiable graphs. We will use these parameters to find the model ideal  $J(G)$  of any HTC-identifiable graph  $G$ , where  $J(G)$  defines the model  $\mathcal{M}(G)$  in the positive definite cone. We will further prove that this ideal is minimal for a subset of HTC-identifiable graphs. We will discuss some complexity issues with our algorithm.

## 2 Preliminaries

Following the works by Wright [21], SEMs are often represented by graphical models. Here, each vertex corresponds to a random variable, each directed edge corresponds to the linear dependencies and each bidirected edge corresponds to possible correlations among noise terms.

A *directed mixed graph* is a triple  $G = (V, \mathcal{D}, \mathcal{B})$  where  $V$  is a set of vertices,  $\mathcal{D}$  is the set of directed edges ( $\rightarrow$ ) and  $\mathcal{B}$  is the set of bidirected edges ( $\leftrightarrow$ ). In this paper, we will only consider graphs without any loops, that is  $i \rightarrow i \notin \mathcal{D}$  and  $i \leftrightarrow i \notin \mathcal{B}$ . If  $i \rightarrow j \in \mathcal{D}$ , we say  $i$  is a *parent* of  $j$  and  $j$  is a *child* of  $i$ . If  $i \leftrightarrow j \in \mathcal{B}$ , we say that  $i$  and  $j$  are *siblings*. If a graph contains both  $i \rightarrow j \in \mathcal{D}$  and  $i \leftrightarrow j \in \mathcal{B}$ , then  $i$  is both a parent and a sibling of  $j$ . The sets of all parents and siblings of  $v \in V$  are denoted  $\text{pa}(v)$  and  $\text{sib}(v)$  respectively.

A *directed path* of length  $\ell$  is of the form  $v_0 \rightarrow v_1 \rightarrow \dots \rightarrow v_\ell$  where all vertices are distinct. Similarly, a *bidirected path* of length  $\ell$  is of the form  $v_0 \leftrightarrow v_1 \leftrightarrow \dots \leftrightarrow v_\ell$  where all vertices are distinct. A *directed cycle* is a directed path with  $v_0 = v_\ell$  and all other vertices  $v_j$  are distinct (for  $0 < j < \ell$ ). A directed mixed graph without any directed cycle is *acyclic*.

Let  $\mathbb{R}^{\mathcal{D}}$  be the set of real  $|V| \times |V|$  matrices  $\Lambda$  with support  $\mathcal{D}$ . That is,

$$\mathbb{R}^{\mathcal{D}} := \{\Lambda \in \mathbb{R}^{|V| \times |V|} : \lambda_{ij} = 0 \text{ if } i \rightarrow j \notin \mathcal{D}\}.$$

Let  $\mathbb{R}_{\text{reg}}^{\mathcal{D}}$  be the subset of matrices  $\Lambda \in \mathbb{R}^{\mathcal{D}}$  for which  $(I - \Lambda)$  is invertible. Let  $PD_V$  be the cone of positive definite  $|V| \times |V|$  matrices and let  $PD(\mathcal{B})$  be the subcone of matrices with support  $\mathcal{B}$ , that is

$$PD(\mathcal{B}) := \{\Omega \in PD_V : \Omega \text{ symmetric, } \omega_{ij} = 0 \text{ if } i \neq j \text{ and } i \leftrightarrow j \notin \mathcal{B}\}.$$

We define the linear structural equation model given by a directed mixed graph  $G = (V, \mathcal{D}, \mathcal{B})$  to be the family of all multivariate normal distributions on  $\mathbb{R}^{|V|}$  with the covariance matrix in the set

$$\mathcal{M}(G) := \{(I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1} : \Lambda \in \mathbb{R}_{\text{reg}}^{\mathcal{D}}, \Omega \in PD(\mathcal{B})\}.$$

### 2.1 Treks

A vertex  $i$  on a walk  $\pi$  is called a *collider on  $\pi$*  if the edges preceding and succeeding  $i$  on the walk  $\pi$  both have an arrowhead at  $i$  (i.e.  $\rightarrow i \leftarrow$ ,  $\rightarrow i \leftrightarrow$ ,  $\leftrightarrow i \leftarrow$ ,  $\leftrightarrow i \leftrightarrow$ ). A *trek* from  $i$  to  $j$  is a walk from  $i$  to  $j$  without any colliders. Therefore, all treks are of the form

$$v_\ell^L \leftarrow v_{\ell-1}^L \leftarrow \dots \leftarrow v_1^L \leftarrow v_0^L \leftrightarrow v_0^R \rightarrow v_1^R \rightarrow \dots \rightarrow v_{r-1}^R \rightarrow v_r^R$$

or

$$v_\ell^L \leftarrow v_{\ell-1}^L \leftarrow \dots \leftarrow v_1^L \leftarrow v_0^{LR} \rightarrow v_1^R \rightarrow \dots \rightarrow v_{r-1}^R \rightarrow v_r^R,$$

where  $v_\ell^L = i, v_r^R = j$ . For a trek  $\pi$ , let  $\pi^L$  denote the directed walk on the left hand side of the trek, containing all the vertices with superscripts  $L$ . For example, in the first case,  $\pi^L = v_\ell^L \leftarrow v_{\ell-1}^L \leftarrow \dots \leftarrow v_1^L \leftarrow v_0^L$ ; whereas in the second case,  $\pi^L = v_\ell^L \leftarrow v_{\ell-1}^L \leftarrow \dots \leftarrow v_1^L \leftarrow v_0^{LR}$ . Similarly, let  $\pi^R$  denote the directed walk containing all the vertices with superscripts  $R$ . A *half-trek* is a trek with  $|\pi^L| = 0$ . Note that the vertex  $v_0^{LR}$  in the second case is a part of both  $\pi^L$  and  $\pi^R$ . Let  $\text{htr}(v)$  be the set of nodes in  $V \setminus (\{v\} \cup \text{sib}(v))$  that can be reached from  $v$  via a half-trek. A set of treks  $\Pi = \{\pi_1, \pi_2, \dots\}$  is said to have no *sided intersection* if  $\pi_i^L \cap \pi_j^L = \pi_i^R \cap \pi_j^R = \emptyset$  for all  $i \neq j$ .

Let  $\mathcal{T}_{vw}$  be the set of all treks from  $v$  to  $w$ . For a trek  $\pi$  with no bidirected edges and a source  $i$ , the trek monomial is defined as

$$\pi(\Lambda, \Omega) = \omega_{ii} \prod_{x \rightarrow y \in \pi} \lambda_{xy}.$$

For a trek  $\pi$  with a bidirected edge  $i \leftrightarrow j$ , the trek monomial is defined as

$$\pi(\Lambda, \Omega) = \omega_{ij} \prod_{x \rightarrow y \in \pi} \lambda_{xy}.$$

**Theorem 2.1** (Trek rule). *The covariance matrix  $\Sigma$  for a mixed graph  $G$  is given by*

$$\sigma_{vw} = \sum_{\pi \in \mathcal{T}_{vw}} \pi(\Lambda, \Omega). \quad (4)$$

*Proof.* The proof originated from [22], and is obtained from expanding out the matrix in (3). If  $G$  is acyclic, the RHS of (3) and (4) agrees. If  $G$  is not acyclic, the RHS of (3) is a power series. Writing  $(I - \Lambda)^{-1} = I + \Lambda + \Lambda^2 + \dots$ , the RHS of (3) and (4) again agrees.  $\square$

**Definition 2.2.** A set of nodes  $Y \subset V$  satisfies the *half-trek criterion* with respect to  $v \in V$  if

1.  $|Y| = |\text{pa}(v)|$ ,
2.  $Y \cap (\{v\} \cup \text{sib}(v)) = \emptyset$ , and
3. there is a system of half-treks with no sided intersection from  $Y$  to  $\text{pa}(v)$ .

**Definition 2.3.** We say that  $\mathcal{M}(G)$  is *HTC-identifiable* if there exists a family of subsets  $\{Y_v : v \in V\}$  of the vertex set  $V$  such that for each node  $v \in V$ , the vertex set  $Y_v$  satisfies the half-trek criterion with respect to  $v$  and there is a total ordering such that  $w \prec v$  whenever  $w \in Y_v \cap \text{htr}(v)$ .

In the original paper, Foygel et al. classify graphs into three classes: HTC-identifiable, HTC-infinite-to-one and HTC-inconclusive. For the purpose of this paper, we will consider the two latter graph classes simply as ‘not HTC-identifiable’.

## 2.2 Constraints

Consider two graphs  $G$  and  $G'$ . If both  $G$  and  $G'$  are *directed acyclic graphs* (DAGs), mixed graphs without bidirected edges and directed cycles, then  $\mathcal{M}(G) = \mathcal{M}(G')$  if and only if  $G$  and  $G'$  have the same d-separation relations [8]. However, this is only helpful in situations where all the variables are observed. If  $G$  is a DAG with hidden variables, we can use the latent projection operation [12] to transform  $G$  into an *acyclic directed mixed graph* (ADMG). In this case, the conditional

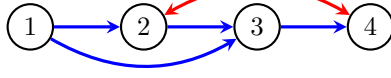


Figure 2: The Verma graph.

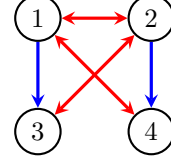


Figure 3: The ‘gadget’ graph.

independence relations on the observed vertices are no longer sufficient to describe  $\mathcal{M}(G)$  in general. The ideal generating these algebraic relations with equality constraints is often referred to as the *vanishing ideal*,

$$\mathcal{I}(G) = \langle f \in \mathbb{R}[\Sigma] \mid f(\Sigma) = 0 ; \forall \Sigma \in \mathcal{M}(G) \rangle.$$

**Example 2.4.** Consider the Verma graph in Figure 2. While there are no conditional independences involving only the observed variables  $X_1, X_2, X_3$  and  $X_4$ . However, we do have an equality constraint on the corresponding covariance matrix  $\Sigma$ , in the sense that  $\Sigma \in \mathcal{M}(G)$  only if

$$\begin{aligned} f_{\text{Verma}}(\Sigma) = & \sigma_{11}\sigma_{13}\sigma_{22}\sigma_{34} - \sigma_{12}^2\sigma_{13}\sigma_{34} - \sigma_{11}\sigma_{14}\sigma_{22}\sigma_{33} + \sigma_{12}^2\sigma_{14}\sigma_{33} \\ & - \sigma_{11}\sigma_{13}\sigma_{23}\sigma_{24} + \sigma_{11}\sigma_{14}\sigma_{23}^2 + \sigma_{12}\sigma_{13}^2\sigma_{24} - \sigma_{12}\sigma_{13}\sigma_{14}\sigma_{23} = 0. \end{aligned}$$

The graph decomposition result [18] leading to Tian’s algorithm [13, 19] provides a method to find constraints in non-parametric graphical models. For instance, the constraint in Example 2.4 can be seen as the independence between  $X_1$  and  $X_4$  after fixing  $X_2$  and  $X_3$  (i.e. changing random vertices 2 and 3 into a fixed vertex by removing incoming all edges into those vertices [13]). Though non-parametrically complete, Tian’s algorithm fails to find the following constraint on the ‘gadget’ graph in Figure 3:

$$\sigma_{11}\sigma_{22}\sigma_{34} - \sigma_{13}\sigma_{14}\sigma_{22} + \sigma_{13}\sigma_{12}\sigma_{24} - \sigma_{23}\sigma_{11}\sigma_{24} = 0.$$

Given any graph  $G$ , the vanishing ideal  $\mathcal{I}(G)$  is traditionally found using Gröbner basis [3] which is computationally expensive. Various attempts, such as that of Drton et al. [6], aim to find a more efficient algorithm for computing  $\mathcal{I}(G)$ .

In this paper, we are interested in exploiting the algebraic properties of HTC-identifiable graphs to find the set of polynomials that generate  $\mathcal{I}(G)$  for any HTC-identifiable graph  $G$ . To do this, we will first consider a subideal  $J(G)$  of  $\mathcal{I}(G)$ , which we call the *model ideal*, such that

$$V(J(G)) \cap PD_V = \mathcal{M}(G),$$

where  $V(I)$  refers to the *variety* defined by an ideal  $I$ ,

$$V(I) = \{a \in \mathbb{S}_{|V|} : f(a) = 0 \text{ for all } f \in I\}$$

for symmetric  $|V| \times |V|$  matrices  $\mathbb{S}_{|V|}$ . The model ideal  $J(G)$  gives a more practical description of  $\mathcal{M}(G)$  compared to the vanishing ideal  $\mathcal{I}(G)$  as  $\mathcal{M}(G) \subseteq PD_V$  and we are dealing with varieties in  $PD_V$ . Hence, we only have to check if the constraints in  $J(G)$  are satisfied to verify that  $\Sigma \in \mathcal{M}(G)$ .

**Example 2.5.** Consider the graph in Figure 4 studied in [17, Example 2.13]. Here, the conditional independences corresponds to the following statements:  $X_1 \perp\!\!\!\perp X_5$ ,  $X_3 \perp\!\!\!\perp X_5 \mid \{X_1, X_2\}$  and

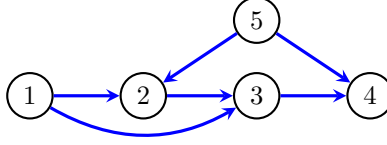


Figure 4: Directed Acyclic Graph Example

$\{X_1, X_2\} \perp\!\!\!\perp X_4 \mid \{X_3, X_5\}$ . Let  $C(G)$  be the conditional independence ideal generated by the determinantal cubics. In the space of  $\mathbb{R}(\Sigma)$ , the vanishing ideal  $\mathcal{I}(G)$  is generated by  $\mathcal{I}(G) = C(G) + \langle f_{\text{Verma}} \rangle$ . However, in the cone of positive definite matrices, we have  $f_{\text{Verma}} \in C(G)$ . Hence,  $J(G) = C(G)$ .

### 3 Linear Identifiability

In this section, we will first define linear identifiability and then explore the properties of linearly identifiable graphs. In particular, we will show that linearly identifiable graphs are equivalent to HTC-identifiable graphs.

First, notice that equation (3) can be rearranged into

$$(I - \Lambda)^T \Sigma (I - \Lambda) = \Omega. \quad (5)$$

By equating each entry in  $(I - \Lambda)^T \Sigma (I - \Lambda)$  that corresponds to a zero entry in  $\Omega$  (i.e. a missing bidirected edge), we obtain a system of equations in terms of the unknowns  $\lambda_{ij}$  only. Since  $\Sigma$  is a covariance matrix, it is symmetric. As any matrix congruent to a symmetric matrix is also symmetric,  $(I - \Lambda)^T \Sigma (I - \Lambda)$  is also a symmetric matrix. By assumption,  $\Omega$  is a positive definite covariance matrix, and is hence symmetric with strictly positive diagonal entries. Therefore, it suffices to equate only the strictly upper or strictly lower triangular entries of both matrices.

Define the matrices  $A = (a_{ij}) = (I - \Lambda)^T \Sigma$  and  $B = (b_{ij}) = (I - \Lambda)^T \Sigma (I - \Lambda)$ , where each entry is a polynomial in which the indeterminates are entries in both  $\Lambda$  and  $\Sigma$ . Expanding the matrix multiplication, we obtain

$$a_{ij} = \sigma_{ij} - \sum_{\ell \in \text{pa}(i)} \lambda_{\ell i} \sigma_{\ell j}, \quad (6)$$

$$b_{ij} = a_{ij} - \sum_{\ell \in \text{pa}(j)} \lambda_{\ell j} a_{i\ell}. \quad (7)$$

Observe that  $a_{ij} \in \mathbb{R}[\Sigma, \Lambda]$  is the polynomial obtained by summing all the covariances for half-treks from  $i$  to  $j$ , and  $b_{ij} \in \mathbb{R}[\Sigma, \Lambda]$  is the polynomial expression for the bidirected edge between  $i$  and  $j$ . Equating each entry of both matrices in (5), we obtain  $b_{ij} = \omega_{ij}$ . Therefore, for each missing bidirected edge between  $i$  and  $j$ , we have the equation  $b_{ij} = b_{ji} = 0$ . We immediately have a necessary condition for generic identifiability.

**Proposition 3.1.** *If  $G$  has more than  $\binom{|V|}{2}$  edges, then  $\mathcal{M}(G)$  is not generically identifiable.*

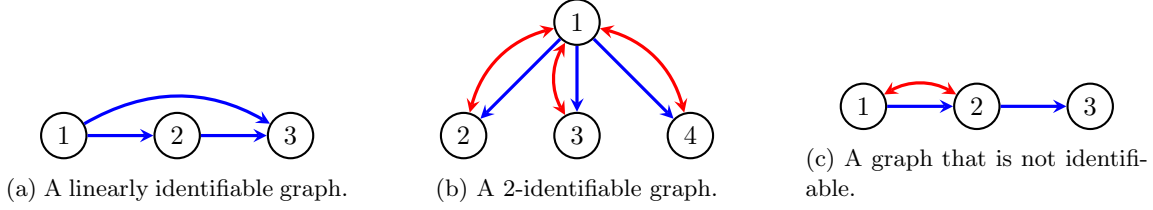


Figure 5

*Proof.* Each missing bidirected edge corresponds to an equation  $b_{ij} = 0$  while each directed edge corresponds to an unknown  $\lambda_{ij}$ . If we have more than  $\binom{|V|}{2}$  edges, then we have more unknowns than equations. Hence,  $\mathcal{M}(G)$  is not identifiable.  $\square$

We shall now provide the definition for a graphical model to be linearly identifiable.

**Definition 3.2.** Let  $G$  be a mixed graph. We say that  $\mathcal{M}(G)$  is *linearly identifiable* if for each vertex  $v$ , the system of  $m = |\text{pa}(v)|$  linearly independent equations of the form  $\{b_{jv} = 0 \mid j \leftrightarrow v \notin \mathcal{B}\}$  (defined in (7)) that is expressed as a polynomial (e.g. through substitution) for the  $m$  indeterminates  $\Lambda_{\text{pa}(v),v} := \{\lambda_{iv} \mid i \in \text{pa}(v)\}$ .

If a model is linearly identifiable, then we solve for  $\Lambda$  (as polynomials in  $\Sigma$ ) recursively using linear algebra. We can then recover  $\Omega$  using equation (5). Hence, all linearly identifiable graphs are generically identifiable. We shall demonstrate linear identifiability with a few examples.

**Example 3.3** (A linearly identifiable graph). Consider the saturated graph  $G$  with three vertices in Figure 5(a). In this graph, the missing bidirected edges give us the equations

$$\begin{aligned} b_{21} &= \sigma_{12} - \lambda_{12}\sigma_{11} = 0, \\ b_{31} &= \sigma_{13} - \lambda_{13}\sigma_{11} - \lambda_{23}\sigma_{12} = 0, \\ b_{32} &= \sigma_{23} - \lambda_{13}\sigma_{12} - \lambda_{23}\sigma_{22} - \lambda_{12}(\sigma_{13} - \lambda_{13}\sigma_{11} - \lambda_{23}\sigma_{12}) = 0. \end{aligned}$$

The vertex 2 has one parent and we have that the first equation is linear in  $\lambda_{12}$ . Solving this, we obtain  $\lambda_{12} = \sigma_{12}/\sigma_{11}$ . The vertex 3 has two parents. Substituting  $\lambda_{12}$  into the last two equations, we have a system of two linearly independent equations with unknowns  $\lambda_{13}, \lambda_{23}$  and coefficients in  $\Sigma$ . Hence,  $\mathcal{M}(G)$  is linearly solvable.

However, not all graphs have such a system of equations, which we shall see in next example.

**Example 3.4** (A graph that is not linearly identifiable). Consider the graph  $G$  in Figure 5(b) with three bows. The missing bidirected edges give us the equations

$$\begin{aligned} b_{32} &= \sigma_{23} - \lambda_{13}\sigma_{12} - \lambda_{12}(\sigma_{13} - \lambda_{13}\sigma_{11}) = 0, \\ b_{42} &= \sigma_{24} - \lambda_{14}\sigma_{12} - \lambda_{12}(\sigma_{14} - \lambda_{14}\sigma_{11}) = 0, \\ b_{43} &= \sigma_{34} - \lambda_{14}\sigma_{13} - \lambda_{13}(\sigma_{14} - \lambda_{14}\sigma_{11}) = 0. \end{aligned}$$

$\mathcal{M}(G)$  is not linearly identifiable as we cannot express any subset of  $\lambda_{12}, \lambda_{13}$  or  $\lambda_{14}$  linearly in terms of  $\Sigma$ . In fact, from the first two equations, we have

$$\lambda_{13} = \frac{\sigma_{23} - \lambda_{12}\sigma_{13}}{\sigma_{12} - \lambda_{12}\sigma_{11}} \quad \text{and} \quad \lambda_{14} = \frac{\sigma_{24} - \lambda_{12}\sigma_{14}}{\sigma_{12} - \lambda_{12}\sigma_{11}}.$$

Substituting these into the third equation, we obtain the quadratic equation  $a\lambda_{12}^2 + b\lambda_{12} + c = 0$  with

$$\begin{aligned} a &= \sigma_{11}^2\sigma_{34} - \sigma_{11}\sigma_{13}\sigma_{14}, \\ b &= 2\sigma_{12}\sigma_{13}\sigma_{14} - 2\sigma_{11}\sigma_{12}\sigma_{34}, \\ c &= \sigma_{12}^2\sigma_{34} - \sigma_{12}\sigma_{13}\sigma_{24} - \sigma_{12}\sigma_{14}\sigma_{23} + \sigma_{11}\sigma_{23}\sigma_{24}. \end{aligned}$$

Since  $b^2 - 4ac \neq 0$ , if  $\Sigma \in \mathbb{C}^{|V| \times |V|}$  then  $\mathcal{M}(G)$  is not generically identifiable. In fact, if  $\Sigma \in \mathbb{C}^{|V| \times |V|}$  and we have to solve a  $k$ th order equation without repeated roots, then  $\mathcal{M}(G)$  is  $k$ -identifiable.

However, determining whether a graph is linearly identifiable is not straightforward. For example, there are instances where the coefficient of one of the unknowns,  $\lambda_{ij}$ , is zero.

**Example 3.5.** Consider the graph in Figure 5(c). The missing bidirected edges correspond to the equations

$$b_{31} = \sigma_{13} - \lambda_{23}\sigma_{12} = 0, \tag{8}$$

$$b_{32} = \sigma_{23} - \lambda_{23}\sigma_{22} - \lambda_{12}(\sigma_{13} - \lambda_{23}\sigma_{12}) = 0. \tag{9}$$

At first glance, these equations might seem to suggest that  $\mathcal{M}(G)$  is linearly identifiable as we can use (8) to solve for  $\lambda_{23}$  and then substitute that solution into (9) to solve for  $\lambda_{12}$ . However, the coefficient of  $\lambda_{12}$  is precisely  $b_{31} = 0$ . Hence, we cannot find an equation that is linear in  $\lambda_{12}$ , so  $\mathcal{M}(G)$  is not linearly identifiable. Indeed, this graph is not even generically identifiable.

As linearly identifiable graphs are not straightforward to classify, we will first introduce a weaker definition that specifically excludes the case where the coefficients of the unknowns might be zero.

**Definition 3.6.** Let  $G$  be a mixed graph. We say that  $\mathcal{M}(G)$  is *quasi-linearly identifiable* if there exists a total ordering  $\prec$  such that for all vertex  $v$ , we can find a system of  $m = |\text{pa}(v)|$  linearly independent equations of the form  $b_{ij} = 0$  for the  $m$  indeterminates  $\Lambda_{\text{pa}(v),v} := \{\lambda_{iv} \mid i \in \text{pa}(v)\}$  where each equation is expressed as a polynomial (e.g. through substitution) in terms of  $\Lambda_{\text{pa}(v),v}$  with coefficients that are functions of  $\Sigma$  and  $\Lambda_{\text{pa}(k),k}$  where  $k \prec v$ .

The main difference between quasi-linearly identifiable models and linearly identifiable models is that in the former, we do not worry about the specific value of each  $\lambda_{ij}$  at each iteration even if it might cause the coefficients of future unknowns to be zero after substitution.

**Example 3.7.** In Example 3.5, vertex 1 satisfies the quasi-linear condition as  $|\text{pa}(1)| = 0$ , and a system of zero equations is trivial. Furthermore, the vertex 3 satisfies the quasi-linear condition as (8) is a system of  $|\text{pa}(3)| = 1$  equation with indeterminates in  $\Sigma$  that is linear in  $\{\lambda_{i3} \mid i \in \text{pa}(3)\}$ . Finally, the vertex 2 also satisfies the quasi-linear condition as (9) is a system of  $|\text{pa}(2)| = 1$  equation with indeterminates in  $\Sigma$  and  $\lambda_{j3}$ , where the vertex 3 also satisfies the quasi-linear condition, that is linear in  $\{\lambda_{i2} \mid i \in \text{pa}(2)\}$ . Hence, the graph in Example 3.5 is quasi-linearly identifiable.

If a quasi-linearly identifiable model is also linearly identifiable, we could solve for  $\lambda_{jk}$  symbolically for all  $j \in \text{pa}(k)$  and substitute this back into the equations  $b_{\ell v} = 0$  to obtain  $|\text{pa}(v)|$  linearly independent equations in  $\lambda_{iv}$ .



### 3.1 Properties of Quasi-linearly Identifiable Graphs

Consider the equation

$$\begin{aligned} b_{ij} &= a_{ij} - \sum_{k \in \text{pa}(j)} \lambda_{kj} a_{ik} \\ &= \sigma_{ij} - \sum_{\ell \in \text{pa}(i)} \lambda_{\ell i} \sigma_{\ell j} - \sum_{k \in \text{pa}(j)} \lambda_{kj} a_{ik}. \end{aligned} \quad (10)$$

Suppose that  $v$  is the first vertex that satisfies the quasi-linear condition. Then, we want to have a set of  $m = |\text{pa}(v)|$  linear equations of the form

$$\sigma_{ij} - \sum_{p \in \text{pa}(v)} \lambda_{pv} \sigma_{hv} = 0$$

for some values of  $i, j, h$ , so we can solve for  $\lambda_{pv}$  for all  $p \in \text{pa}(v)$ .

We see that this can be achieved by picking  $m$  equations of the form  $b_{vj} = 0$  (i.e. no bidirected edge between  $v$  and  $j$ ) such that the last summation in (10) is zero. In other words, for all  $k \in \text{pa}(j)$ , we have  $a_{vk} = 0$  (i.e. no half-trek from  $v$  to  $k$ ).

Writing this in terms of graph properties, we want to find a vertex  $v$  and a set of vertices  $S$  such that:

1.  $|S| = |\text{pa}(v)|$ ;
2.  $S \cap (\text{sib}(v) \cup \{v\}) = \emptyset$ ; and
3. for each vertex  $j \in S$ , for all  $k \in \text{pa}(j)$ , there are no half-treks from  $v$  to  $k$ .

We can now define the quasi-linear properties above for all vertices recursively using the following algorithm.

**Algorithm 1.**

**Input:** Mixed graph  $G$ .

**Output:** Set of recursively quasi-linear vertices

**Initialise:**  $Q = \emptyset$ ;

**while**  $\exists v \in V \setminus Q$  such that we can find a set of vertices  $S_v$  with properties

1.  $|S_v| = |\text{pa}(v)|$ ,

2.  $S_v \cap (\text{sib}(v) \cup \{v\}) = \emptyset$  and

3. For each vertex  $j \in S_v$ , either  $j \in Q$ , or for all  $k \in \text{pa}(j)$ , there are no half-treks from  $v$  to  $k$ .

**do**

$Q = Q \cup \{v\}$

**end**

**return**  $Q$ .

**Definition 3.8.** Suppose  $G$  is a mixed graph. We say that a vertex  $v$  is *recursively quasi-linear* in  $G$  if it is contained in the output of Algorithm 1.

**Lemma 3.9.** *Let  $v \in V$  be an arbitrary vertex. Suppose  $|S_v| = |\text{pa}(v)|$  and  $S_v \cap (\text{sib}(v) \cup \{v\}) = \emptyset$ . Then, for all  $j \in S_v$ , there are no half-treks from  $v$  to  $k$  for all  $k \in \text{pa}(j)$  if and only if there are no half-treks from  $v$  to  $j$ .*

*Proof.* ( $\Leftarrow$ ) Suppose we have no half-treks from  $v$  to  $j$ . Since  $k$  is a parent of  $j$ , we will not have a half-trek from  $v$  to  $k$  since otherwise we can append  $k \rightarrow j$  at the end of this half-trek to form a half-trek from  $v$  to  $j$ .

( $\Rightarrow$ ) Suppose that there are no half-treks between  $v$  and the parents of  $j$ . Since  $S_v \cap (\text{sib}(v) \cup \{v\}) = \emptyset$ ,  $v$  is not a sibling of  $j$ . Hence, there are no half-treks from  $v$  to  $j$ .  $\square$

**Theorem 3.10.** *Let  $G$  be a mixed graph. Then  $\mathcal{M}(G)$  is quasi-linearly identifiable if and only if we have recursively defined that all vertices are recursively quasi-linear, i.e. if Algorithm 1 outputs  $V$ .*

*Proof.* ( $\Leftarrow$ ) It suffices to show that if  $v$  is recursively quasi-linear, then we can find  $|\text{pa}(v)|$  equations that are linear in  $\{\lambda_{iv} \mid i \in \text{pa}(v)\}$ . Suppose that for each vertex  $v$  we can find such an  $S_v$  satisfying the conditions in Algorithm 1. From condition 2, for each  $j \in S_v$ , we have an equation of the form  $b_{vj} = 0$ . Condition 1 states that we can find  $|\text{pa}(v)|$  such equations. Expanding all such equations  $b_{vj} = 0$  into (10), if  $j \in S$  is not recursively quasi-linear, condition 3 guarantees that the last summand in (10) is zero. If  $j \in S$  is recursively quasi-linear, then by condition 3 the last summand in (10) is linear in  $\lambda_{\ell i}$  with indeterminates in  $\Sigma$  and  $\lambda_{kj}$ .

( $\Rightarrow$ ) Suppose that a vertex  $v$  is not recursively quasi-linear. If either of the first two conditions in Algorithm 1 fails, we no longer have  $|\text{pa}(v)|$  linearly independent equations in any subset of  $\{\lambda_{iv} \mid i \in \text{pa}(v)\}$ . Hence,  $\mathcal{M}(G)$  is not quasi-linearly identifiable. Now, if the third condition fails, then by definition we have at least one other vertex that is not recursively quasi-linear. So we must have at least one other vertex that fails to satisfy the first two conditions in Algorithm 1.  $\square$

**Example 3.11.** In Example 3.5, vertex 1 is recursively quasi-linear trivially. Furthermore, we have  $S_3 = \{1\}$  so the vertex 3 is also recursively quasi-linear. Moreover,  $S_2 = \{3\}$ , so vertex 2 is also recursively quasi-linear. Hence, the graph is quasi-linearly identifiable.

## 3.2 Properties of Linearly Identifiable Graphs

Previously, we found the necessary and sufficient conditions for a graph to be quasi-linearly identifiable. The following Theorem will determine when a quasi-linearly identifiable graph is also linearly identifiable.

**Theorem 3.12.** *Let  $G$  be a mixed graph such that  $\mathcal{M}(G)$  is quasi-linearly identifiable. Then  $\mathcal{M}(G)$  is linearly identifiable if and only if for each vertex  $v$ , we have a set of vertices  $S_v$  satisfying the half-trek criterion with respect to  $v$  in addition to the three conditions in Algorithm 1.*

*Proof.* Since  $\mathcal{M}(G)$  is quasi-linearly identifiable, we have a system of  $|\text{pa}(v)|$  linear equations in  $\{\lambda_{iv} \mid i \in \text{pa}(v)\}$  for each vertex  $v$ . We can rewrite the equations  $b_{sv} = a_{sv} - \sum_{\ell \in \text{pa}(v)} \lambda_{\ell v} a_{s\ell} = 0$  for some choice of  $s \in S_v$  into the following matrix equation

$$\begin{bmatrix} a_{s_1 p_1} & \cdots & a_{s_1 p_m} \\ \vdots & & \vdots \\ a_{s_m p_1} & \cdots & a_{s_m p_m} \end{bmatrix} \begin{bmatrix} \lambda_{p_1 v} \\ \vdots \\ \lambda_{p_m v} \end{bmatrix} = \begin{bmatrix} a_{s_1 v} \\ \vdots \\ a_{s_m v} \end{bmatrix}$$

where  $S_v = \{s_1, \dots, s_m\}$  and  $\text{pa}(v) = \{p_1, \dots, p_m\}$ . Note that the leftmost matrix is precisely  $A_{S_v, \text{pa}(v)}$ . Hence the system of equations are linearly independent if and only if  $A_{S_v, \text{pa}(v)}$  is of full rank.

Now, if  $A_{S_v, \text{pa}(v)}$  has full rank, by Lemma 3 of Foygel et al. [7],  $S_v$  satisfies the half-trek criterion. On the other hand, if  $S_v$  satisfies the half-trek criterion, then we have a system of half-treks with no sided intersection from  $S_v$  to  $\text{pa}(v)$ . Since each entry of  $A_{S_v, \text{pa}(v)}$  is a polynomial obtained from summing all the covariances for half-trek between  $s_i \in S_v$  and  $p_j \in \text{pa}(v)$ ,  $A_{S_v, \text{pa}(v)}$  has full rank.  $\square$

**Example 3.13.** In Example 3.5, we have that  $\mathcal{M}(G)$  is quasi-linearly identifiable. However,  $\mathcal{M}(G)$  is not linearly identifiable since  $S_2 = \{3\}$  but there are no half-treks from 3 to 2, therefore  $S_2$  does not satisfy the half-trek criterion with respect to 2.

We now have the following necessary and sufficient conditions for linear identifiability.

**Definition 3.14.** We say that a vertex  $v$  satisfies the *linear identifiability criterion* if there exists a set of vertices  $S_v$  such that:

- $S_v$  satisfies the half-trek criterion with respect to  $v$ ;
- for each vertex  $j \in S_v$ , either  $j$  satisfies the linear identifiability criterion or there are no half-treks from  $v$  to  $j$ .

**Proposition 3.15.** Let  $G$  be a mixed graph. Then  $\mathcal{M}(G)$  is linearly identifiable if and only if we can recursively check that all vertices in  $G$  satisfy the linear identifiability criterion.

*Proof.* From Lemma 3.9 and Theorem 3.10, a graphical model is quasi-linearly identifiable if and only if for each vertex  $v \in V$ , we can find a set of vertices  $S_v$  such that

- $|S_v| = |\text{pa}(v)|$ ,
- $S_v \cap (\text{sib}(v) \cup \{v\}) = \emptyset$  and
- for each vertex  $j \in S_v$ , either  $j$  is recursively quasi-linear or there are no half-treks from  $v$  to  $j$ .

From Theorem 3.12, a quasi-linearly identifiable graphical model is linearly identifiable if and only if the sets  $S_v$  defined earlier further satisfies the half-trek criterion with respect to  $v$  for each vertex  $v \in V$ .  $\square$

### 3.3 Connection to HTC-identifiable Graphs

Note that from Definition 2.3, we can rewrite the definition of HTC-identifiability as follows.

**Definition 3.16.** Let  $G$  be an HTC-identifiable graph. Then,  $\mathcal{M}(G)$  is HTC-identifiable if and only if for each vertex  $v$ , there is a total ordering  $\prec$  such that we can find a vertex set  $S_v$  satisfying both

1.  $S_v$  satisfies the half-trek criterion with respect to  $v$ , and
2. For each vertex  $j \in S_v$ , either  $j \prec v$  or there are no half-treks from  $v$  to  $j$ .

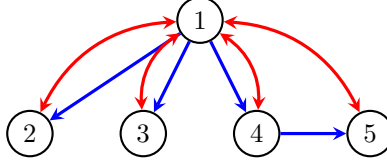


Figure 6: A generically identifiable graph that is not HTC-identifiable.

**Theorem 3.17.** *For any mixed graph  $G$ ,  $\mathcal{M}(G)$  is linearly identifiable if and only if  $\mathcal{M}(G)$  is HTC-identifiable.*

*Proof.* ( $\Rightarrow$ ) Define a total ordering on the vertex set where  $j \prec v$  if we defined  $j$  to satisfy the linear identifiability criterion before  $v$  in the recursion. Hence, the conditions in Definition 3.14 are the same as those in Definition 3.16.

( $\Leftarrow$ ) We shall proceed by induction. First, let 1 be the first vertex in the total ordering  $\prec$ . Then for each vertex  $j \in S_1$ , there are no half-treks from 1 to  $j$ , hence, 1 satisfies the linear identifiability criterion.

Now, suppose that vertices  $1 \prec \dots \prec k$  satisfy the linear identifiability criterion. We want to show that the  $k+1^{\text{st}}$  vertex in the total ordering  $\prec$  also satisfies the linear identifiability criterion. Since  $\mathcal{M}(G)$  is HTC-identifiable, there exists some  $S_{k+1}$  satisfying the half-trek criterion with respect to  $k+1$  and if  $j \in S_{k+1}$ , either  $j \prec k+1$  or there are no half-treks from  $v$  to  $j$ . But if  $j \prec k+1$ , by our induction hypothesis,  $j$  satisfies the linear identifiability criterion. Hence,  $k+1$  also satisfies the linear identifiability criterion.  $\square$

Linear identifiability can be used to explain the algebraic properties of HTC-identifiable graphs and graphs that are not HTC-identifiable.

**Example 3.18.** The graph in Figure 6 that is modified from Example 3.4 with an additional vertex added. This graph is not linearly identifiable as we cannot find an  $S_v$  satisfying both conditions of Definition 3.14 for any vertex  $v$  except for  $v = \{1\}$  as we can take  $S_1 = \emptyset$ . Hence, it is not HTC-identifiable. Suppose we have

$$\Lambda = \begin{bmatrix} 0 & 0.25 & 0.5 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \Omega = \begin{bmatrix} 1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 1 & 0 & 0 & 0 \\ 0.1 & 0 & 1 & 0 & 0 \\ 0.1 & 0 & 0 & 1 & 0 \\ 0.1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

As we have seen in Example 3.4, equating  $b_{32} = b_{42} = b_{43} = 0$ , we obtain a quadratic equation in  $\lambda_{12}$ . Solving this, we obtain  $\lambda_{12} = 0.25$  or  $0.45$ . If we equate  $b_{32} = b_{52} = b_{53} = 0$  and solve the corresponding quadratic equation, we also obtain  $\lambda_{12} = 0.25$  or  $0.45$ . However, equating  $b_{42} = b_{52} = b_{54} = 0$  and solving the corresponding quadratic equation, we obtain  $\lambda_{12} = 0.25$  or  $0.354$ . Since only  $\lambda_{12} = 0.25$  is a common solution, this graph is generically identifiable, even though it is not linearly identifiable.

However, graphs that are not HTC-identifiable may be decomposed, using Tian's decomposition [18], into mixed components that are HTC-identifiable. Hence, the parameters for each of the mixed components could be recovered using linear algebra.

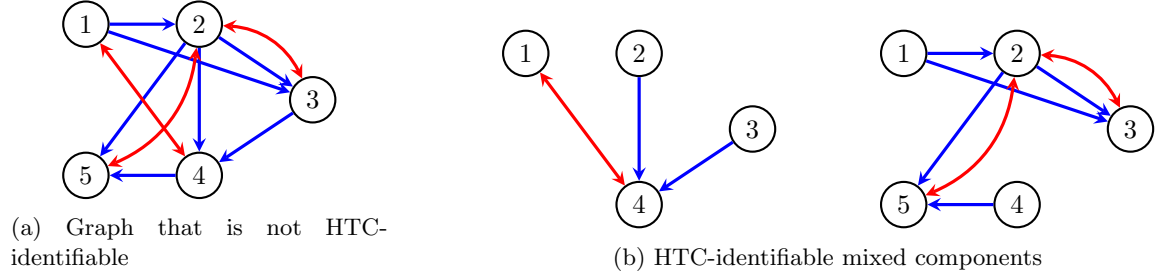


Figure 7: Graph that is not HTC-identifiable but can be decomposed into HTC-identifiable mixed components

**Example 3.19.** Consider the graph  $G = (V, \mathcal{D}, \mathcal{B})$  in Figure 7(a) which is not HTC-identifiable. Suppose we decompose  $G$  into its mixed components  $G_1 = (V_1, \mathcal{D}_1, \mathcal{B}_1)$  and  $G_2 = (V_2, \mathcal{D}_2, \mathcal{B}_2)$  shown in Figure 7(b). We see that both  $G_1$  and  $G_2$  are HTC-identifiable. Hence, we are able to use linear algebra to recover the parameters  $\Lambda_1$  and  $\Lambda_2$  of  $G_1$  and  $G_2$  respectively. However, for a directed edge  $i \rightarrow j \in \mathcal{D}$ , the value of  $\lambda_{ij}$  is identical to that of  $[\Lambda_k]_{ij}$  for which  $i \rightarrow j \in \mathcal{D}_k$ . So even if  $G$  is not HTC-identifiable, its parameters  $\Lambda$  could be recovered using linear algebra operations, but not from the system of equations  $\{b_{ij} = 0 \mid i \leftrightarrow j \notin \mathcal{B}\}$ .

## 4 Constraints in Linearly Identifiable Graphs

In this section, we will demonstrate some applications of our result by exploiting the algebraic properties of HTC-identifiable graphs to compute  $\Lambda$  and the *model ideal*  $J(G)$ . While algorithms for HTC-identifiable graphs for finding  $\Lambda$  already exist [7], our argument will be purely algebraic in nature. We will also prove that the generators we find for the model ideal are minimal for some subset of HTC-identifiable graphs.

### 4.1 Generators of the Model Ideal

Let  $v$  be a vertex in a linearly identifiable graph  $G$ . Suppose  $S_v = \{s_1, \dots, s_k\}$  and  $\text{pa}(v) = \{p_1, \dots, p_k\}$ . Then,

$$b_{s_i v} = a_{s_i v} - \sum_{\ell \in \text{pa}(v)} \lambda_{\ell v} a_{s_i \ell} = 0$$

for each  $1 \leq i \leq k$ . We can rewrite this into a matrix

$$\begin{bmatrix} a_{s_1 p_1} & \dots & a_{s_1 p_k} \\ \vdots & & \vdots \\ a_{s_k p_1} & \dots & a_{s_k p_k} \end{bmatrix} \begin{bmatrix} \lambda_{p_1 v} \\ \vdots \\ \lambda_{p_k v} \end{bmatrix} = \begin{bmatrix} a_{s_1 v} \\ \vdots \\ a_{s_k v} \end{bmatrix}. \quad (11)$$

By definition of linear identifiability, each of the  $a_{ij}$  can be expressed solely in terms of  $\Sigma$ , and the square matrix on the left is invertible. Since linearly identifiable graphs are also HTC-identifiable, we have a total ordering on the vertices of the graph. This leads us to an algorithm to recover  $\Lambda$ ,  $\Omega$  and the generators for the model ideal  $J(G)$ .

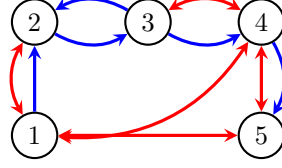


Figure 8: An HTC-identifiable graph with cycles.

**Algorithm 2.**

**Input:** HTC-identifiable graph  $G$ , the associated sets  $\{S_v \mid v \in V\}$  satisfying the half-trek criterion wrt  $v \in V$  and the total ordering  $1 \prec \dots \prec |V|$ .

**Output:** Symbolic values of all regression coefficients as a matrix  $\Lambda = (\lambda_{ij})$ .

**Initialise:**  $v = 1$ ;

**while**  $v \leq |V|$  **do**

**for**  $s_1, \dots, s_k \in S_v$  and  $p_1, \dots, p_k \in \text{pa}(v)$  **do**

        Solve for  $\lambda_{p_i v}$  for all  $1 \leq i \leq k$  using the equation:

$$\begin{bmatrix} \lambda_{p_1 v} \\ \vdots \\ \lambda_{p_k v} \end{bmatrix} = \begin{bmatrix} a_{s_1 p_1} & \dots & a_{s_1 p_k} \\ \vdots & & \vdots \\ a_{s_k p_1} & \dots & a_{s_k p_k} \end{bmatrix}^{-1} \begin{bmatrix} a_{s_1 v} \\ \vdots \\ a_{s_k v} \end{bmatrix},$$

        where  $a_{ij} = \sigma_{ij} - \sum_{\ell \in \text{pa}(i)} \lambda_{\ell i} \sigma_{\ell j}$ .

**end**

$v = v + 1$ .

**end**

**return**  $\Lambda = (\lambda_{ij})$ .

Given  $\Sigma$ , Algorithm 2 could also be used to compute the values of  $\Lambda$  numerically. After obtaining either the numeric or the symbolic values of  $\Lambda$ , we can compute the values for  $\Omega$  using the equation

$$\Omega = (I - \Lambda)^T \Sigma (I - \Lambda). \quad (12)$$

If both  $\Lambda$  and  $\Omega$  were computed symbolically, we could proceed to find the generators the model ideal.

**Example 4.1.** Consider the graph in Figure 8 with the total ordering  $1 \prec 3 \prec 5 \prec 2 \prec 4$  and sets  $S_1 = \emptyset, S_3 = \{1\}, S_5 = \{3\}, S_2 = \{3, 5\}, S_4 = \{2\}$ . Note that these sets satisfy the conditions in Definition 3.16.

Applying Algorithm 2, for the first vertex, there is nothing to do as  $\text{pa}(1) = \emptyset$  and hence we have no  $\lambda$ 's to solve for. For the next vertex, we have  $\lambda_{23} = a_{12}^{-1} a_{13} = \sigma_{12}^{-1} \sigma_{13}$ . Similarly,  $\lambda_{45} = a_{34}^{-1} a_{35} = (\sigma_{34} - \lambda_{23} \sigma_{24}) / (\sigma_{35} - \lambda_{23} \sigma_{25}) = (\sigma_{12} \sigma_{34} - \sigma_{13} \sigma_{24}) / (\sigma_{12} \sigma_{35} - \sigma_{13} \sigma_{25})$  and so forth.

**Theorem 4.2.** Suppose that  $\Lambda$  is a matrix of rational functions with indeterminates  $\Sigma$  found by Algorithm 2. For an acyclic HTC-identifiable graph  $G$ , the model ideal is generated by

$$\left\langle [(I - \Lambda)^T \Sigma (I - \Lambda)]_{ij} \mid i \leftrightarrow j \notin \mathcal{B}, i \notin S_j \text{ and } j \notin S_i \right\rangle.$$

For a cyclic HTC-identifiable graph, the model ideal is given by saturating the above ideal by  $\det(I - \Lambda)$ .

*Proof.* Equating entries in (5), we see that the equality constraints in the positive definite cone are generated by

$$\left\langle [(I - \Lambda)^T \Sigma (I - \Lambda)]_{ij} \mid i \leftrightarrow j \notin \mathcal{B} \right\rangle.$$

Hence,  $J(G) = \langle b_{ij} \mid i \leftrightarrow j \notin \mathcal{B} \rangle$ .

Now, from Algorithm 2, we used the system of equations  $\{b_{sv} = 0 \mid s \in S_v\}$  to solve for  $\lambda_{pv}$  for all  $p \in \text{pa}(v)$ . Hence if either  $i \in S_j$  or  $j \in S_i$ , substituting these  $\lambda$ 's back into the polynomial  $b_{ij}$  will result in a zero polynomial. Note that for all other choices of  $(i, j)$ , substituting  $\Lambda$  will not result in  $b_{ij}$  being a zero polynomial as the structural equations are linearly independent.

In cyclic graphs, we remove the infeasible solution  $\det(I - \Lambda) = 0$  by saturating the ideal.  $\square$

Theorem 4.2 could be interpreted as having each missing bidirected edge being either a part of the system of equations used in Algorithm 2 to solve for  $\Lambda$ , or is a generator for the model ideal  $J(G)$ . In particular, our choice for  $J(G)$  is generated by  $\binom{|V|}{2} - |\mathcal{D}| - |\mathcal{B}|$  polynomials. Note that the result of Theorem 4.2 is very similar to [20, Theorem 1] where it was proven that  $b_{ij} \in \mathcal{I}(G)$  if  $i \leftrightarrow j \notin \mathcal{B}, i \notin S_j$  and  $j \notin S_i$ . However, in Theorem 4.2, we showed that these polynomials in fact generate the model ideal.

**Example 4.3.** Let  $G$  be the Verma graph in Figure 2, we have  $S_v = \text{pa}(v)$  for all vertex  $v$  and

$$\Lambda = \begin{bmatrix} 0 & \lambda_{12} & \lambda_{13} & 0 \\ 0 & 0 & \lambda_{23} & 0 \\ 0 & 0 & 0 & \lambda_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \Omega = \begin{bmatrix} \omega_{11} & 0 & 0 & 0 \\ 0 & \omega_{22} & 0 & \omega_{24} \\ 0 & 0 & \omega_{33} & 0 \\ 0 & \omega_{24} & 0 & \omega_{44} \end{bmatrix}.$$

In the first iteration of our algorithm, we obtain  $\lambda_{12} = \sigma_{11}^{-1} \sigma_{12}$ . In the second iteration, we have

$$\begin{bmatrix} \lambda_{13} \\ \lambda_{23} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$$

where  $a_{ij} = \sigma_{ij} - \sum_{k \in \text{pa}(i)} \lambda_{ki} \sigma_{kj}$ . Solving this matrix equation, we obtain

$$\lambda_{13} = \frac{\sigma_{13 \cdot 2}}{\sigma_{11 \cdot 2}} \quad \text{and} \quad \lambda_{23} = \frac{\sigma_{23 \cdot 1}}{\sigma_{22 \cdot 1}}$$

where the conditional covariance matrix of a multivariate normal distribution is given by

$$\Sigma_{AC \cdot B} = \Sigma_{AC} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BC}.$$

In the last iteration, we have  $\lambda_{34} = a_{33}^{-1}a_{34} = \sigma_{34,12}/\sigma_{33,12}$ . Note that we could have computed  $\Omega$  symbolically by equation (5). Finally, since the only pair of non-adjacent vertices are 1 and 4, we have

$$\begin{aligned} J(G) = \langle b_{41} \rangle &= \left\langle \sigma_{14} - \sum_{k=1}^3 \lambda_{k4} \sigma_{1k} \right\rangle = \langle \sigma_{14} - \lambda_{34} \sigma_{13} \rangle \\ &= \langle \sigma_{11} \sigma_{13} \sigma_{22} \sigma_{34} - \sigma_{12}^2 \sigma_{13} \sigma_{34} - \sigma_{11} \sigma_{14} \sigma_{22} \sigma_{33} + \sigma_{12}^2 \sigma_{14} \sigma_{33} - \\ &\quad \sigma_{11} \sigma_{13} \sigma_{23} \sigma_{24} + \sigma_{11} \sigma_{14} \sigma_{23}^2 + \sigma_{12} \sigma_{13}^2 \sigma_{24} - \sigma_{12} \sigma_{13} \sigma_{14} \sigma_{23} \rangle \end{aligned}$$

which is exactly what we have in Example 2.4. Note that the second to last equality comes from the fact that both  $\lambda_{14} = \lambda_{24} = 0$ .

## 4.2 Minimality of the Generators

Now that we can find the generators of  $J(G)$  for any HTC-identifiable  $G$ , it is natural to ask if generators of the model ideal found in Theorem 4.2 are in fact the minimal generators in the following sense:

**Definition 4.4.** We say that the polynomials  $f_1, \dots, f_n$  are *minimal generators* for  $I$  if we have  $\langle f_1, \dots, f_n \rangle = I$  but  $\langle f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_n \rangle \neq I$  for any  $1 \leq i \leq n$ . In other words, the polynomials  $f_1, \dots, f_n$  generate  $I$  but if we remove any of the  $f_i$  for  $1 \leq i \leq n$ , the remaining polynomials no longer generate  $I$ .

First, we will define a subset of HTC-identifiability:

**Definition 4.5.** The sets  $\{S_v \mid v \in V\}$  in an HTC-identifiable graph has *no subset cycles* if it satisfies the criterion in Definition 3.16 and there do not exist vertices  $v_1, \dots, v_n$  with  $v_1 \in S_{v_2}, v_2 \in S_{v_3}, \dots, v_n \in S_{v_1}$ .

**Lemma 4.6.** Suppose that we have an HTC-identifiable graph  $G = (V, \mathcal{D}, \mathcal{B})$  with a collection of sets  $\{S_v \mid v \in V\}$  that has no subset cycles. Further suppose that there exist vertices  $i, j \in V$  such that  $i \leftrightarrow j \notin \mathcal{B}, i \notin S_j$  and  $j \notin S_i$ . Then, we can add the bidirected edge  $i \leftrightarrow j$  to obtain another HTC-identifiable graph with the sets  $\{S_v \mid v \in V\}$  unchanged.

*Proof.* Firstly, we check that each  $S_v$  still satisfies the half-trek criterion with respect to  $v$ .

1.  $|S_v| = |\text{pa}(v)|$  still holds as adding bidirected edges does not impact the number of parents of any vertex  $v$  while each  $S_v$  remains unchanged.
2. Since  $S_i \cap (\{i\} \cup \text{sib}(i)) = \emptyset$ , adding the bidirected edge  $i \leftrightarrow j$  where  $j \notin S_i$  preserves that equality. By symmetry,  $S_j \cap (\{j\} \cup \text{sib}(j)) = \emptyset$ . For any other vertices  $w$ , the equation  $S_w \cap (\{w\} \cup \text{sib}(w)) = \emptyset$  is preserved as  $\text{sib}(w)$  and  $S_w$  remains unchanged.
3. There still exists a system of half-treks with no sided intersection from  $S_v$  to  $\text{pa}(v)$ , as adding the bidirected edge does not impact half-treks.

Recall that a graph is HTC-identifiable if  $S_v$  satisfies the half-trek criterion with respect to each vertex  $v \in V$  and  $w \prec v$  whenever  $w \in S_v \cap \text{htr}(v)$ . For the second condition, by definition by the total ordering in HTC-identifiable graphs, it is equivalent to view the total ordering as a topological



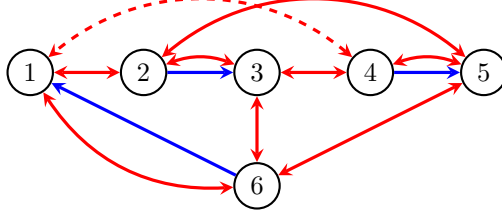


Figure 9: An HTC-identifiable graph that is not HTC-identifiable after adding the dotted edge.

ordering  $<$  on the directed graph  $G' = (V', \mathcal{D}', \mathcal{B}')$  with  $V' = V$ ,  $\mathcal{D}' = \{w \rightarrow v \mid w \in S_v \cap \text{htr}(v)\}$ ,  $\mathcal{B}' = \emptyset$  since  $w \prec v$  in  $G$  if and only if  $w < v$  in  $G'$ . If the sets  $S_v$  have no subset cycles,  $G'$  is a DAG and a topological ordering exists. Therefore, leaving  $S_v$  unchanged, the new graph satisfies both conditions of HTC-identifiability.  $\square$

**Theorem 4.7.** *If  $G$  is an HTC-identifiable graph with no subset cycles, then the generators of the model ideal  $J(G)$  found in Theorem 4.2 are minimal.*

*Proof.* Suppose that  $J(G) = \langle f_1, \dots, f_n \rangle$  is not a minimal generator of  $J(G)$ . Without loss of generality, suppose that the constraint  $f_1$  corresponding to the missing edge between  $v$  and  $w$  is redundant such that  $J(G) = \langle f_2, \dots, f_n \rangle$ . Now, consider the graph  $G'$  obtained by adding the bidirected edge  $v \leftrightarrow w$  to  $G$ . Note that  $G'$  is also an HTC-identifiable graph by Lemma 4.6, with  $J(G') = \langle f_2, \dots, f_n \rangle$ .

However, since the constraint  $f_1$  can be generated by polynomials generating  $J(G')$ , the bidirected edge  $v \leftrightarrow w$  in  $G'$  can only take one value  $\omega_{vw} = 0$ , which is a contradiction.  $\square$

**Corollary 4.8.** *If  $G$  is a bow-free acyclic graph, then the generators of the model ideal  $J(G)$  found in Theorem 4.2 are minimal.*

*Proof.* In a bow-free acyclic graph, we have  $S_v = \text{pa}(v)$ . Since the graph is acyclic, these sets contain no subset cycles. The result follows from Theorem 4.7.  $\square$

**Example 4.9.** If we have an HTC-identifiable graph that does not satisfy the conditions of Lemma 4.6, it is possible that it is no longer HTC-identifiable if we add a bidirected edge.

Consider the graph in Figure 9. Before adding the bidirected edge  $1 \leftrightarrow 4$ , we must have  $S_5 = \{3\}$ . Now, for  $S_3$ , we can have either  $S_3 = \{1\}$  or  $S_3 = \{5\}$ . However, if we pick the latter, we have  $3 \in S_5$  and  $5 \in S_3$  which is HTC-nonidentifiable [7, Theorem 2]. Similarly, we have either  $S_1 = \{3\}$  or  $S_1 = \{5\}$  but by picking the former we have  $3 \in S_1$  and  $1 \in S_3$  which is HTC-nonidentifiable. Hence, the only choice is  $S_1 = \{5\}$ ,  $S_3 = \{1\}$  and  $S_5 = \{3\}$ .

Now, we have  $1 \notin S_4$  and  $4 \notin S_1$ . Adding the edge  $1 \leftrightarrow 4$ , we obtain  $1 \in S_3 \cap \text{htr}(3)$ ,  $3 \in S_5 \cap \text{htr}(5)$  and  $5 \in S_1 \cap \text{htr}(1)$ . Hence, the resulting graph is no longer HTC-identifiable as we can no longer find a total ordering.

### 4.3 Time Complexity

In this section, we will show that if  $S_v$  is given for every  $v \in V$ , we are able to compute  $\Lambda$  and  $\Omega$  numerically in polynomial time. This is useful if we are working in graphs whose  $S_v$  is known or

easy to find, such as acyclic bow-free graphs where  $S_v = \text{pa}(v)$ .

### 4.3.1 Numerical Computations

We first introduce a naive bound for the numerical complexity of the above algorithm based on the number of vertices of  $G$ . We shall also assume that our numeric values of  $\Sigma$  do not result in any singular matrices in Algorithm 2.

**Theorem 4.10.** *Suppose that  $G$  is HTC-identifiable, then the complexity of finding  $S_v$  for each vertex  $v$  is at most  $O(|V|^2(|V| + r)^3)$  where  $r \leq |\mathcal{D}|/2$  is the number of reciprocal edge pairs in  $\mathcal{D}$ .*

*Proof.* See [7, Theorem 7].  $\square$

**Proposition 4.11.** *Suppose  $G$  is HTC-identifiable and  $S_v$  is known for each  $v \in V$ . Then the complexity of finding  $\Lambda$  and  $\Omega$  numerically is at most  $O(|V|^4)$ .*

*Proof.* The complexity of solving  $k$  equations with  $k$  unknowns using naive Gauss-Jordan elimination is  $O(k^3)$ . In our algorithm for finding the values of  $\lambda_{ij}$ 's, we first solve a linear equation with at most one unknown, then two linear equations with at most two unknowns and so on until we solved  $|V| - 1$  linear equations with at most  $|V| - 1$  unknowns. Since we this process is iterated  $|V|$  times, the complexity of finding the values of  $\lambda_{ij}$ 's is at most  $O(|V|^4)$ . Finally, since the complexity of matrix multiplication is  $O(|V|^3)$ , we can compute both  $\Lambda$  and  $\Omega$  in  $O(|V|^4)$ .  $\square$

In particular, the bound of  $O(|V|^4)$  is attained by a complete DAG. Furthermore, in graphs where  $S_v$  is known, Proposition 4.11 could be useful. Otherwise, if  $S_v$  is unknown, the complexity is at most  $O(|V|^2(|V| + r)^3)$ .

One might notice that each unknown  $\lambda_{ij}$  corresponds precisely to a directed edge in  $G$ . In a sparse graph where  $|\mathcal{D}|$  is small, it might be beneficial to express the complexity in terms of  $|\mathcal{D}|$  instead.

**Proposition 4.12.** *Suppose  $G$  is HTC-identifiable and  $S_v$  is known for each  $v \in V$ . Then the complexity of finding  $\Lambda$  numerically is at most  $O(|\mathcal{D}|^3)$ .*

*Proof.* This follows from the fact that we have at least  $|\mathcal{D}|$  equations with exactly  $|\mathcal{D}|$  unknowns. Observe that each time when we apply the Gauss-Jordan elimination, we can solve for any subset of the  $|\mathcal{D}|$  unknowns. Therefore, we obtain the following optimization problem:

$$\begin{aligned} & \underset{k_1, \dots, k_{|\mathcal{D}|}}{\text{maximise}} && k_1^3 + \dots + k_{|\mathcal{D}|}^3 \\ & \text{subject to} && k_1 + \dots + k_{|\mathcal{D}|} = |\mathcal{D}| \\ & \text{where} && k_1, \dots, k_{|\mathcal{D}|} \in \mathbb{Z}_{\geq 0}, \end{aligned}$$

where  $k_i$  is the number of unknowns in the  $i^{\text{th}}$  set of equations. Since the cubic term grows faster than the linear term, the maximum is obtained when  $k_i = |\mathcal{D}|$  for some  $i$  and  $k_j = 0$  for all  $j \neq i$ .  $\square$

**Corollary 4.13.** *Suppose  $G$  is an HTC-identifiable graph and  $S_v$  is known for each  $v \in V$ . Denote  $P = \max_{v \in V} |\text{pa}(v)|$ . Then the complexity of finding  $\Lambda$  numerically is at most  $O(|V|P^3)$ .*

*Proof.* We use the Gauss-Jordan elimination at most  $v$  times, solving for at most  $P$  unknowns each time.  $\square$

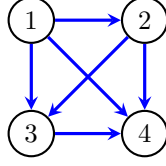


Figure 10: A complete DAG on 4 vertices.

### 4.3.2 Symbolic Complications

Now, we will show that our algorithm does not run symbolic calculations in polynomial time by providing a simple counter-example. Since finding the generators of  $J(G)$  requires symbolic computation, we can not compute  $J(G)$  in polynomial time in general.

Consider a complete DAG such as the graph shown in Figure 10. Using our algorithm, solving for  $\Lambda$  symbolically, we obtain

$$\lambda_{14} = \sigma_{14 \cdot 23} / \sigma_{11 \cdot 23}, \quad \lambda_{24} = \sigma_{24 \cdot 13} / \sigma_{22 \cdot 13}, \quad \lambda_{34} = \sigma_{34 \cdot 12} / \sigma_{33 \cdot 12}.$$

In particular, the regression coefficient  $\lambda_{ij}$  is precisely  $\sigma_{ij \cdot S} / \sigma_{ii \cdot S}$  in a DAG, where  $S = \text{pa}(j) \setminus \{i\}$  [11]. Recall that the conditional covariance matrix of a multivariate normal distribution is given by

$$\Sigma_{AA \cdot B} = \Sigma_{AA} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BA}.$$

Now, there are  $|B|!$  different terms in  $\det(\Sigma_{BB})$ . Since  $\sigma_{ij \cdot V} \propto \det(\Sigma_{-i, -j})$  which has  $(|V| - 1)!$  terms, the number of terms in  $\lambda_{ij}$  increases factorially with respect to  $|\text{pa}(i)|$ . As Algorithm 2 requires us to invert matrices with entries as functions of  $\lambda$ , we might not be able to solve for  $\Lambda$  symbolically in polynomial time. However, the expression may in principle factorise, and therefore we might only have to deal with  $O(\log(n!)) = O(n \log(n))$  terms.

## 5 Conclusion

In this paper, we have defined a subset of generically identifiable graphs called *linearly identifiable graphs*. Graphs are linearly identifiable if their model parameters can be recovered from the covariance matrix with straightforward linear algebra operations, given in Algorithm 2 and (12). We have also proven that a graph is linearly identifiable if and only if it is HTC-identifiable.

Furthermore, we have shown that, for some graphs, there is a bijection between the generators of the model ideal, generated by all equality constraints of  $\mathcal{M}(G) \cap PD_V$ , and the vertex pairs  $\{(i, j) \mid i \leftrightarrow j \notin \mathcal{B}, i \notin S_j \text{ and } j \notin S_i\}$ . In graphs where the collection of sets  $\{S_v \mid v \in V\}$  has no subset cycles, the generators of the vertex pairs are minimal. To prove this, we utilise Lemma 4.6, which fails to hold in all HTC-identifiable graphs as demonstrated in Example 4.9. However, we do conjecture that the generators we found are minimal for all HTC-identifiable graphs.

**Conjecture 1.** For any HTC-identifiable graph  $G$ , the generators of the model ideal  $J(G)$  found in Theorem 4.2 are minimal.

Finally, we have shown that the complexity to recover the parameters of an HTC-identifiable graph  $G$ , given that we know  $\{S_v \mid v \in V\}$  is at most  $O(|V|^4)$ . Meanwhile, it is known that the

complexity of finding  $S_v$  for each vertex  $v$  given an HTC-identifiable graph is at most  $O(|V|^2(|V| + r)^3)$  where  $r \leq |\mathcal{D}|/2$  is the number of reciprocal edge pairs in  $\mathcal{D}$ . Unfortunately, we do not have an efficient algorithm to find the sets  $S_v$  without subset cycles, if such a collection of sets exists.

## References

- [1] K. A. Bollen. *Structural equations with latent variables*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1989. A Wiley-Interscience Publication.
- [2] C. Brito and J. Pearl. A new identification condition for recursive models with correlated errors. *Structural Equation Modeling*, 9(4):459–474, 2002.
- [3] D. A. Cox, J. Little, and D. O’Shea. *Ideals, varieties, and algorithms*. Undergraduate Texts in Mathematics. Springer, Cham, fourth edition, 2015.
- [4] M. Drton. Algebraic problems in structural equation modeling. In *The 50th anniversary of Gröbner bases*, pages 35–86. Mathematical Society of Japan, 2018.
- [5] M. Drton, R. Foygel, S. Sullivant, et al. Global identifiability of linear structural equation models. *The Annals of Statistics*, 39(2):865–886, 2011.
- [6] M. Drton, E. Robeva, and L. Weihs. Nested covariance determinants and restricted trek separation in Gaussian graphical models. *Bernoulli*, 26(4):2503–2540, 2020.
- [7] R. Foygel, J. Draisma, M. Drton, et al. Half-trek criterion for generic identifiability of linear structural equation models. *The Annals of Statistics*, 40(3):1682–1713, 2012.
- [8] D. Geiger, T. Verma, and J. Pearl. Identifying independence in Bayesian networks. *Networks*, 20(5):507–534, 1990.
- [9] J. B. Grace. *Structural equation modeling and natural systems*. Cambridge University Press, 2006.
- [10] S. L. Hershberger. The growth of structural equation modeling: 1994-2001. *Structural Equation Modeling*, 10(1):35–46, 2003.
- [11] J. Pearl. *Causality: models, reasoning and inference*, volume 29. Springer, 2000.
- [12] J. Pearl and T. S. Verma. A statistical semantics for causation. *Statistics and Computing*, 2(2):91–95, 1992.
- [13] T. S. Richardson, R. J. Evans, J. M. Robins, and I. Shpitser. Nested Markov properties for acyclic directed mixed graphs. *arXiv preprint arXiv:1701.06686*, 2017.
- [14] K. J. Rothman and S. Greenland. Causation and causal inference in epidemiology. *American journal of public health*, 95(S1):S144–S150, 2005.
- [15] M. E. Sobel. Causal inference in the social sciences. *Journal of the American Statistical Association*, 95(450):647–651, 2000.

- [16] P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, Prediction, and Search*. Springer-Verlag, New York, NY, 1993.
- [17] S. Sullivant. Algebraic geometry of gaussian bayesian networks. *Advances in Applied Mathematics*, 40(4):482–513, 2008.
- [18] J. Tian and J. Pearl. A general identification condition for causal effects. In *AAAI*, pages 567–573, 2002.
- [19] J. Tian and J. Pearl. On the testable implications of causal models with hidden variables. In *Uncertainty in Artificial Intelligence*, pages 519–527, 2002.
- [20] T. van Ommen and J. M. Mooij. Algebraic equivalence of linear structural equation models. In *Uncertainty in Artificial Intelligence*, 2017.
- [21] S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20(7):557–585, 1921.
- [22] S. Wright. The method of path coefficients. *The Annals of Mathematical Statistics*, 5(3):161–215, 1934.