GENERIC IDENTIFIABILITY OF DIRECTED GAUSSIAN GRAPHICAL MODELS WITH ONE LATENT SOURCE

DENNIS LEUNG, MATHIAS DRTON, AND HISAYUKI HARA

ABSTRACT. We study parameter identifiability of directed Gaussian graphical models with one latent variable. In the scenario we consider, the latent variable is a confounder that forms a source node of the graph and is a parent to all other nodes, which correspond to the observed variables. We give a graphical condition that is sufficient for the Jacobian matrix of the parametrization map to be full rank, which entails that the parametrization is generically finite-to-one, a fact that is sometimes also referred to as local identifiability. We also derive a graphical condition that is necessary for such identifiability. Finally, we give a condition under which generic parameter identifiability can be determined from identifiability of a model associated with a subgraph. The power of these criteria is assessed via an exhaustive algebraic computational study on models with 4, 5, and 6 observable variables.

1. INTRODUCTION

In this paper we study parameter identifiability of directed Gaussian graphical models with one latent variable as a source. These models can be described as follows. Let $X_1, \dots, X_m$ be observable variables, and let L be a hidden variable, and suppose the variables are related by linear equations as

$$X_v = \sum_{w \neq v} \lambda_{wv} X_w + \delta_v L + \epsilon_v, \qquad v = 1, \dots, m,$$

where λ_{wv}, δ_v are real coefficients quantifying linear relationships, and the ϵ_v are independent mean zero Gaussian noise terms with variances $\omega_v > 0$. The latent variable L is assumed to be standard normal and independent of the noise terms ϵ_v. Letting $X = (X_1, \dots, X_m)^T$, $\epsilon = (\epsilon_1, \dots, \epsilon_m)^T$ and $\delta = (\delta_1, \dots, \delta_m)^T$, we may present the model in the vectorized form

$$X = \Lambda^T X + \delta L + \epsilon, \tag{1.1}$$

where Λ is the matrix (λ_{wv}) with $\lambda_{vv} = 0$ for all $v = 1, \dots, m$. We are then interested in specific models, in which for certain pairs of nodes $w \neq v$ the coefficient λ_{wv} is constrained to zero. In particular, we are interested in recursive models, that is, models in which the matrix Λ can be brought into strictly upper triangular form by permuting the indices of the variables (and thus the rows and columns of Λ). This implies that $I_m - \Lambda$ is invertible, where I_m is the $m \times m$ identity matrix. It follows that the observable variate vector X has a m-variate normal distribution $N_m(0, \Sigma)$ with covariance matrix

$$\Sigma = (I_m - \Lambda)^{-T} (\Omega + \delta\delta^T)(I_m - \Lambda)^{-1}, \tag{1.2}$$

Date: Draft from 19 January 2015.
2000 *Mathematics Subject Classification.* 62H05.

where Ω is the diagonal matrix with $(\Omega)_{vv} = \omega_v$.

A Gaussian latent variable model postulating recursive zero structure in the matrix Λ from (1.1) can be thought of as associated with a graph $G = (V, E)$ whose vertex set $V = \{1, \dots, m\}$ is the index set for the observable variables $X_1, \dots, X_m$. For two distinct nodes $w, v \in V$, the edge set E includes the directed edge (w, v), denoted as $w \to v$ if and only if the model includes λ_{wv} as a free parameter. When the model is recursive, the directed graph G is acyclic and following common terminology we refer to G as a DAG (for directed acyclic graph). In this paper, we will then always assume that the vertices are labelled in topological order, that is, we have $V = \{1, \dots, m\}$ and $w \to v \in E$ only if $w < v$.

To emphasize the presence of the latent variable L, one could equivalently represent the model by an extended DAG $\overline{G} = (\overline{V}, \overline{E})$ on $m+1$ nodes enumerated as $\overline{V} := \{0, 1, \dots, m\}$, where the node 0 corresponds to the latent variable L, and if $G = (V, E)$ is the graph on m nodes representing the model in the preceding paragraph, then $\overline{E} = E \cup \{0 \to v : v \in \{1, \dots, m\}\}$. The edges $0 \to v$ correspond to the coefficients δ_v.

For the DAG $G = (V, E)$, let

$$\mathbb{R}_E := \{\Lambda = (\lambda_{wv}) \in \mathbb{R}^{m \times m} : \lambda_{wv} = 0 \iff w \to v \notin E\}$$

be the linear space of coefficient matrices, and let diag_m^+ be the set of all $m \times m$ diagonal matrices with a positive diagonal.

Definition 1.1. The Gaussian one latent source model associated with a given DAG $G = (V, E)$, denoted as $\mathcal{N}_*(G)$, is the family of all m-variate normal distribution $N_m(0, \Sigma)$ with a covariance matrix of the form

$$\Sigma = (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1},$$

for $\Lambda \in \mathbb{R}_E$, $\Omega \in \mathrm{diag}_m^+$ and $\delta \in \mathbb{R}^m$.

The model $\mathcal{N}_*(G)$ has the parametrization map

$$\phi_G : (\Lambda, \Omega, \delta) \longmapsto (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1} \tag{1.3}$$

defined on the set $\Theta := \mathbb{R}_E \times \mathrm{diag}_m^+ \times \mathbb{R}^m$, which is an open subset of $\mathbb{R}^{2m+|E|}$, where $|E|$ denotes cardinality of the directed edge set E. Clearly, the image of ϕ_G is in PD_m, the cone of positive definite $m \times m$ matrices. Note that since G is acyclic, we have $(I_m - \Lambda)^{-1} = I_m + \Lambda^2 + \dots, \Lambda^{m-1}$ and ϕ_G is a polynomial map.

We assume the reader to be familiar with graphical models at the level of Lauritzen (1996) and Pearl (2009). In this paper we will derive graphical conditions on G that are sufficient/necessary for identifiability of the model $\mathcal{N}_*(G)$. We begin by clarifying what precisely we will mean by identifiability. The most stringent notion, namely that of global identifiability, requires ϕ_G to be injective on all of Θ. This is too stringent, however, as for any triple $(\Lambda, \Omega, \delta) \in \Theta$, $\phi_G(\Lambda, \Omega, \delta) = \phi_G(\Lambda, \Omega, -\delta)$, which implies that the fiber of $(\Lambda, \Omega, \delta)$, defined as

$$\mathcal{F}(\Lambda, \Omega, \delta) := \{(\Lambda', \Omega', \delta') \in \Theta : \phi_G(\Lambda, \Omega, \delta) = \phi_G(\Lambda', \Omega', \delta')\},$$

always has cardinality ≥ 2. We may account for this symmetry by requiring ϕ_G to be 2-to-1 on all of Θ but this is not enough as there are always some fibers that are infinite. For instance, it is easy to show that $|\mathcal{F}(\Lambda, \Omega, \delta)| = \infty$ when $\delta = 0$. Instead, it is natural to consider notions of generic identifiability. Specifically, our contributions will pertain to the notion of generic finite identifiability, as defined below, that only requires finite identification of parameters away from a fixed null set in Θ; here a null set is a set of Lebesgue measure zero.

All null sets appearing in our work are algebraic sets, where an algebraic set $A \subset \mathbb{R}^n$ is the set of common zeros of a collection of multivariate polynomials, i.e.,

$$A = \{a \in \mathbb{R}^n : f_i(a) = 0,\ i = 1, \dots, k\},$$

for $f_i \in \mathbb{R}[x_1, \dots, x_n]$, where $\mathbb{R}[x_1, \dots, x_n]$ is the ring of polynomials in n variables over $\mathbb{R}$. Note that A is a closed set. If all polynomials f_i are the zero polynomial then $A = \mathbb{R}^n$. Otherwise, A is a proper subset, $A \subsetneq \mathbb{R}^n$, and its dimension is then less than n. In particular, a proper algebraic subset of $\mathbb{R}^n$ has measure zero.

Definition 1.2. Let S be an open subset of $\mathbb{R}^n$, and let f be a map defined on S. Then f is said to be generically finite-to-one if there exists an algebraic set $\tilde{S} \subset \mathbb{R}^n$ with $\dim(\tilde{S}) < n$ such that for all $s \in S \backslash \tilde{S}$, the fiber

$$\mathcal{F}(s) := \{s' \in S : f(s') = f(s)\},$$

is finite. Otherwise, f is said to be generically infinite-to-one.

Definition 1.3. The model $\mathcal{N}_*(G)$ of a given DAG $G = (V, E)$ is generically finitely identifiable if its parametrization map ϕ_G defined on Θ is generically finite-to-one.

We will use the term "generic point" to refer to any point in the domain Θ that lies outside a fixed proper algebraic subset $\Xi \subset \Theta$, and a property is said to hold generically if it holds everywhere on Θ except Ξ. The following lemma is well-known, and its proof will be included in Section 6 for the sake of completeness. It gives, as an immediate corollary, a trivial necessary condition for generic finite identifiability.

Lemma 1.1. *Suppose $f : S \to \mathbb{R}^d$ is a polynomial map defined on an open set $S \subset \mathbb{R}^n$. The following statements are equivalent:*

(i) f is generically finite-to-one.

(ii) There exists a proper algebraic subset $\tilde{S} \subset \mathbb{R}^n$ such that the fibers of the restricted map $f|_{S \backslash \tilde{S}}$ are all finite, i.e. for any $s \in S \backslash \tilde{S}$,

$$\mathcal{F}_{S \backslash \tilde{S}}(s) := \{s' \in S \backslash \tilde{S} : f(s') = f(s)\}$$

has finite cardinality.

(iii) The Jacobian matrix of f is generically of full column rank.

Corollary 1.2. *Given a DAG $G = (V, E)$, a necessary condition for generic finite identifiability of its associated model $\mathcal{N}_*(G)$ is $\binom{m+1}{2} - 2m \geq |E|$.*

Proof of Corollary 1.2. The Jacobian matrix of ϕ_G is of dimension $\binom{m+1}{2} \times (|E| + 2m)$, and it is necessary that $\binom{m+1}{2} \geq |E| + 2m$ for it to have full column rank. □

We remark that the seemingly weaker statement (ii) in Lemma 1.1 will be used to prove our results in Section 4. In light of Corollary 1.2, for the rest of this paper we will restrict our attention to DAGs $G = (V, E)$ with $\binom{m+1}{2} - 2m \geq |E|$, otherwise the parametrization ϕ_G is generically infinite-to-one for sure.

One of our contributions is a sufficient graphical condition stated in Theorem 1.3 below. For $v \neq w \in V$, we will use $v - w$ or $w - v$ to denote the edge $(v, w) = (w, v)$ of an undirected graph on V. With slight abuse of notation, we may also use $v - w$ or $w - v$ to denote an edge $v \to w \in E$ when the directionality of edges in a DAG

$G = (V, E)$ is to be ignored. For any directed/undirected graph $G = (V, E)$, the complement of G, denoted as $G^c = (V, E^c)$, is the undirected graph on V with the undirected edge set $E^c = \{v — w : v — w \notin E\}$.

Theorem 1.3 (Sufficient condition for generic finite identifiability). *Given a DAG $G = (V, E)$, its associated model $\mathcal{N}_*(G)$ in Definition 1.1 is generically finitely identifiable if every connected component of G^c contains an odd cycle.*

Our approach to proving Theorem 1.3 also yields a necessary condition for generic finite identifiability. This condition can be stated in terms of two undirected graphs on the node set V, denoted $G_{|L,cov} = (V, E_{|L,cov})$ and $G_{con} = (V, E_{con})$, where $E_{|L,cov}$ captures the dependency of variable pairs after conditioning on the latent variable L, and E_{con} captures the dependency of variable pairs after conditioning on all other variables. From (1.1) it can be seen that $\Sigma_{|L} := (I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1}$ is the covariance matrix of X conditioning on L, hence $v—w \in E_{|L,cov}$ if and only if $(\Sigma_{|L})_{vw} \not\equiv 0$, and analogously $v—w \in E_{con}$ if and only if $(\Sigma_{|L}^{-1})_{vw} \not\equiv 0$. It is well known that these two undirected graphs can be induced with d-separation criteria applied to the extended DAG $\overline{G}$; see Drton et al. (2009, p. 73) for example.

Theorem 1.4 (Necessary condition for generic finite identifiability). *Given a DAG $G = (V, E)$, for the model $\mathcal{N}_*(G)$ to be generically finitely identifiable, it is necessary that the following two conditions both hold:*

(i) Let d_{con} be the number of connected components in the graph G^c_{con} that do not contain any odd cycle, then $|E_{con}| - |E| \geq d_{con}$.

(ii) Let d_{cov} be the number of connected components in the graph $G^c_{|L,cov}$ that do not contain any odd cycle, then $|E_{|L,cov}| - |E| \geq d_{cov}$.

On top of the closely related work of Stanghellini (1997) and Vicard (2000), identifiability of directed Gaussian models with one latent variable has been studied in Stanghellini and Wermuth (2005). Our models are special cases with the latent node being a common parent of all the observable nodes. We can readily adapt the sufficient graphical criteria given in Stanghellini and Wermuth (2005) for certifying that the model $\mathcal{N}_*(G)$ of a given DAG G is generically finitely identifiable with respect to Definition 1.3. Our sufficient condition stated in Theorem 1.3, however, is strictly stronger, in the sense that every DAG G satisfying the sufficient conditions in Stanghellini and Wermuth (2005) necessarily satisfies the condition in Theorem 1.3.

Section 2 will review Stanghellini and Wermuth (2005)'s work. In Section 3 we will prove Theorems 1.3 and 1.4. In fact, since the parametrization map in (1.3) is a multivariate polynomial, the generic finite identifiability of a given model is decidable by algebraic Gröbner basis computations with respect to a system of polynomial equations; refer to Cox et al. (2007). At the end of Section 3, we will compare our graphical criteria with these algebraic computational results for all models $\mathcal{N}_*(G)$ of DAGs G with $m = 4, 5, 6$ nodes. In Section 4, we will give results on situations where we can determine generic finite identifiability of a model $\mathcal{N}_*(G)$ based on knowledge about the generic finite identifiability of a model $\mathcal{N}_*(G')$, where G' is an induced subgraph of G. We will wrap up the paper with some discussions.

Before ending this section, it is worth mentioning how Markov equivalence plays its role in our problem. Recall that two DAGs defined on the same set of nodes are Markov equivalent if they have the same d-separations. The following theorem, which will be proved in Section 6, says that generic finite identifiability is a universal property within a Markov equivalent class of DAGs.

Theorem 1.5. *Suppose $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are two Markov equivalent DAGs on the same set of m nodes V. Then the model $\mathcal{N}_*(G_1)$ is generically finitely identifiable if and only if the same is true for $\mathcal{N}_*(G_2)$.*

2. Prior work

Stanghellini and Wermuth (2005) gave sufficient criteria for identifiability of a general directed Gaussian graphical model with one latent node that can be any node in the DAG. Assuming unit variance for the latent variable in any such model, given the covariance matrix of the observables, one can identify the covariance of any observable-latent variable pair if its corresponding DAG satisfies Stanghellini and Wermuth (2005)'s graphical criteria. When restricted to our subclass of models in Definition 1.1, this amounts to being able to identify the m-vector $(I_m - \Lambda^T)^{-1}\delta$, whose v-th component equals $\mathrm{cov}(X_v, L)$; refer to (1.1). As it turns out, by the following lemma the same criteria are also sufficient for identifying the parameter tuple $(\Lambda, \Omega, \delta)$ up to sign change in δ.

Lemma 2.1. *Given the system* (1.1)*, suppose that $\Sigma_{|L}$, the covariance matrix of X conditioning on L, is known, then the system of equations*

$$\Sigma_{|L} = (I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1}$$

can always be uniquely solved for,with solution (Λ, Ω) that is rational function of $\Sigma_{|L}$.

Proof. The proof is elementary, by considering the Cholesky decomposition of the matrix $\Sigma_{|L} = R^T R$, where R is the upper triangular Cholesky factor of $\Sigma_{|L}$. Then the parameters λ_{vw} and ω_v can be solved for by considering the matrix equation $R = \Omega^{1/2}(I_m - \Lambda)^{-1}$ column by column from right to left, where $\Omega^{1/2} = \mathrm{diag}(\sqrt{\omega_1}, \ldots, \sqrt{\omega_m})$. We leave the details to the readers. $\square$

Stanghellini and Wermuth (2005)'s result, with respect to generic finite identifiability, is then stated below.

Corollary 2.2 (Stanghellini and Wermuth (2005), adapted)**.** *Let $G = (V, E)$ be a DAG. The model $\mathcal{N}_*(G)$ is generically finitely identifiable if either*

(i) Every connected component of $G^c_{|L,cov} = (V, E^c_{|L,cov})$ has an odd cycle, or
(ii) Every connected component of $G^c_{con} = (V, E^c_{con})$ has an odd cycle.

Proof. Theorem 1 in Stanghellini and Wermuth (2005) gives either (i) or (ii) as a sufficient condition for identifying, up to sign, the m-vector $(I_m - \Lambda^T)^{-1}\delta$ of observable-latent covariances, when $\Sigma = \phi_G(\Lambda, \Omega, \delta)$ for an unknown generic point $(\Lambda, \Omega, \delta)$ in Θ. Immediately, this also means that we can uniquely solve for the matrix $(I_m - \Lambda^T)^{-1}\Omega(I_m - \Lambda)^{-1}$ in view of (1.2), and consequently Ω and Λ can be identified by Lemma 2.1. Knowing Λ, δ can be solved for up to sign by the previous knowledge of $(I_m - \Lambda^T)^{-1}\delta$. Hence (i) or (ii) is in fact a sufficient condition for generic finite-to-one-ness of ϕ_G. . $\square$

By how the graphs $G_{|L,cov}$ and G_{con} are induced from the DAG G as described in the Section 1, we know that our sufficient condition in Theorem 1.3 is strictly stronger than Stanghellini and Wermuth (2005)'s condition in Corollary 2.2 for generic finite identifiability.

Theorem 2.3. *A DAG G satisfying either one of the conditions in Corollary 2.2 necessarily satisfies the condition in Theorem 1.3. As a consequence, any model $\mathcal{N}_*(G)$ that is determined to be generically finitely identifiable using Corollary 2.2 can also be such determined by Theorem 1.3.*

Proof. Let $G = (V, E)$, $G_{|L,cov} = (V, E_{|L,cov})$ and $G_{con} = (V, E_{con})$. An edge $v \to w \in E$ also present itself as an undirected edge in both $E_{|L,cov}$ and E_{con}, hence by ignoring directionality G is a subgraph of both $G_{|L,cov}$ and G_{con}. It is then seen that G^c is a supergraph of both $G^c_{|L,cov}$ and G^c_{con}. As such, if every connected component of $G^c_{|L,cov}$, or G^c_{con}, contains an odd cycle, the same is true of G^c. □

3. Graphical criteria based on the Jacobian matrix of parametrization maps

We will prove Theorems 1.3 and 1.4 in this section. In particular, we will introduce other mappings whose generic finite-to-one-ness is equivalent to that of ϕ_G. The appropriate mathematical jargons to study these maps are that of semialgebraic sets/maps. A subset $A \subset \mathbb{R}^n$ is semialgebraic if $A = \cup_{i=1}^{s} \cap_{j=1}^{r_i} \{x \in \mathbb{R}^n : P_{ij}(x) \quad s_{ij} \quad 0\}$, where s_{ij} is one of the comparison operators $\{<, =, >\}$, $x = (x_1, \ldots, x_n)^T$ and $P_{ij}(x) \in \mathbb{R}[x_1, \ldots, x_n]$. All algebraic sets are semialgebraic, and the domain Θ of the parametrization ϕ_G is also semialgebraic. Let $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^{n'}$ be semialgebraic sets, a map $f : A \longrightarrow B$ is semialgebraic if the graph of f is a semialgebraic set in $\mathbb{R}^{n+n'}$. Polynomial and rational mappings are semialgebraic. A continuous bijective semialgebraic map $f : V \longrightarrow W$ is a semialgebraic homeomorphism if the inverse f^{-1} is semialgebraic and continuous.

Lemma 3.1. *Let $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^{n'}$ be semialgebraic sets such that $f : S \longrightarrow T$ is a semialgebraic homeomorphism. Suppose $g : T \longrightarrow \mathbb{R}^d$ is a polynomial map, then $g \circ f$ is generically finite-to-one iff g is generically finite-to-one.*

Proof. The proof is elementary by using three basic facts from real algebraic geometry: (i) The image of semialgebraic set under a semialgebraic map is semialgebraic. (ii) If $A \subset \mathbb{R}^n$ is a semialgebraic, then the Zariski closure of A, i.e., the smallest algebraic set in $\mathbb{R}^n$ containing A, has the same dimension as A. (iii) If S is a semialgebraic set and f is a semialgebraic homeomorphism defined on S, $\dim(S) = \dim(f(S))$. The interested readers are referred to Coste (2002). □

For a DAG $G = (V, E)$, $\phi_G = \tilde{\phi}_G \circ g$, where

$$\tilde{\phi}_G : (\Lambda, \Omega, \delta) \longmapsto (I - \Lambda^T)^{-1} \Omega (I - \Lambda)^{-1} + \delta\delta^T \tag{3.1}$$

and $g : (\Lambda, \Omega, \delta) \longmapsto (\Lambda, \Omega, (I - \Lambda^T)^{-1}\delta)$ are maps defined on Θ. Since g is a polynomial isomorphism (and hence a semialgebraic homeomorphism), by Lemma 3.1, ϕ_G is generically finite-to-one if and only if the map $\tilde{\phi}_G$ is.

One may also consider the function composition $inv \circ \phi_G$, where inv is simply the matrix inverting function such that $inv(\Sigma) = \Sigma^{-1}$ for any $m \times m$ invertible matrix. Trivially ϕ_G is generically finite-to-one if and only if $inv \circ \phi_G$ is so. It is clear that

$$inv \circ \phi_G(\Lambda, \Omega, \delta) = (I - \Lambda)(\Omega + \delta\delta^T)^{-1}(I - \Lambda^T)$$

for any triple $(\Lambda, \Omega, \delta) \in \Theta$. Suppose we define the function

$$\varphi_G : \quad (\Lambda, \Psi, \gamma) \longmapsto (I - \Lambda)(\Psi - \gamma\gamma^T)(I - \Lambda^T) \tag{3.2}$$

on the same domain Θ. Using the matrix identity $(\Omega + \delta\delta^T)^{-1} = (\Psi - \gamma\gamma^T)$ (Rao, 1973, p. 33), with $\Psi = \Omega^{-1} = \mathrm{diag}(\psi_1, \ldots, \psi_m)$, $\gamma = k\Psi\delta$ and $k = (1 + \delta^T\Psi\delta) > 0$, we obtain

$$inv \circ \phi_G = \varphi_G \circ \rho,$$

where $\rho : \Theta \longrightarrow \Theta$ is such that $\rho(\Lambda, \Omega, \delta) = (\Lambda, \Psi, \gamma)$ and $\rho(\Theta) = \Theta$. Since ρ is a bijective rational map with a continuous semialgebraic inverse, it is a semialgebraic homeomorphism, then again by Lemma 3.1, $inv \circ \phi_G$ is generically finite-to-one if and only if φ_G is so, and the same equivalence carries between ϕ_G and φ_G. Finally, by defining the map

$$\tilde{\varphi}_G : (\Lambda, \Psi, \gamma) \longmapsto (I - \Lambda)\Psi(I - \Lambda^T) - \gamma\gamma^T \tag{3.3}$$

and $h : (\Lambda, \Psi, \gamma) \longmapsto (\Lambda, \Psi, (I - \Lambda)\gamma)$, both on the domain Θ, it can be seen that $\varphi_G = \tilde{\varphi}_G \circ h$ and by an argument analogous to the one appearing right after (3.1), φ_G is generically finite-to-one if and only if $\tilde{\varphi}_G$ is. We summarize the above in the following lemma.

Lemma 3.2. *The map in* (1.3) *is generically finite-to-one iff any one of the maps in* (3.1), (3.2) *and* (3.3) *is generically finite-to-one.*

Let $J(\tilde{\varphi}_G)$ be the Jacobian matrix of the map $\tilde{\varphi}_G$ in (3.3). It will be examined to prove Theorem 1.3. In light of Lemmas 1.1 and 3.2, we will show that if the condition in Theorem 1.3 is satisfied by G, $J(\tilde{\varphi}_G)$ will be generically of full column rank. This will require the following lemma that is based on observations made in Vicard (2000).

Lemma 3.3. *Let $G = (V, E)$ be an undirected graph on m nodes. Suppose for each edge v — $w \in E$, there is a function f_{vw} that takes the variable pair (x_v, x_w) as input such that*

$$f_{vw}(x_v, x_w) = x_v x_w.$$

Let $x = (x_1, \ldots x_m)^T$. If d is the number of connected components of G that do not contain any odd cycles, then the $|E| \times m$ Jacobian matrix of the vector valued function $f(x) = \big(f_{vw}(x_v, x_w)\big)^T_{(v,w)\in E}$ generically has rank $m - d$.

Proof. It suffices to assume x only takes its value in the generic open set $\mathcal{X} := \{x : x_v \neq 0 \text{ for all } v \in V\}$ of $\mathbb{R}^m$. This assumption is made so that Lemma 1 in Vicard (2000) is applicable later without difficulty. Let J_f be the Jacobian matrix of the function f. By rank thoerem (Rudin, 1976, p.229), the dimension of the null space, $\mathrm{null}(J_f)$, of J_f is same as the dimension of the fibers of f. Since $\mathrm{rank}(J_f) = m - \mathrm{null}(J_f)$, it suffices to show that the fiber of f has dimension d.

Fix a point $y = (y_{vw})^T_{(v,w)\in E}$ in the image space $f(\mathcal{X}) = \{f(x) : x \in \mathcal{X}\}$. Consider the system of equations

$$y_{vw} = x_v x_w \text{ for } v \text{ — } w \in E. \tag{3.4}$$

Let $G_1 = (V_1, E_1), \ldots, G_k = (V_k, E_k)$ be the connected components of G, where $\dot\cup_{i=1}^k V_i = V$ and $\dot\bigcup_{i=1}^k E_i = E$, and $k' \leq k$ be the number of connected components containing two nodes at least. Without any loss of generality, we assume $G_{k'+1}, \ldots, G_k$

are all the connected components with only a single node. In this regard we can break (3.4) into k' disjoint subsystems, where each subsystem has the form

$$y_{vw} = x_v x_w \text{ for } v — w \in E_i \tag{3.5}$$

and exclusively involves the variables $\{x_v : v \in V_i\}$ for $i \in \{1, \ldots, k'\}$. For each $i = 1, \ldots, k'$, by Lemma 1 in Vicard (2000) and also the relevant discussion in the proof of Theorem 1 in the same paper, the solution set $\{x_v : v \in V_i\}$ either contains two points when G_i contains an odd cycle, or can be parametrized by a nonzero free variable in $\mathbb{R}$ when G_i does not. In the former case the dimension of the corresponding solution set is zero, whereas in the latter case the dimension of the corresponding solution set is one. In addition, each singleton component $G_i = (V_i, \varnothing)$ for $i = k'+1, \ldots, k$ provides an additional dimension, since the corresponding variables in x are not restricted by any equations. Hence the number of connected components G_i that do not contain any odd cycles gives the dimension of the solution set for the whole system (3.4). □

$\tilde{\varphi}_G$ maps the $2m + |E|$-dimensional space $\Theta := \mathbb{R}_E \times \text{diag}^+_m \times \mathbb{R}^m$ to the $\binom{m+1}{2}$-dimensional space of symmetric $m \times m$ matrices, thus $J(\tilde{\varphi}_G)$ is of dimension $\binom{m+1}{2} \times (2m + |E|)$. We will arrange the rows and columns of $J(\tilde{\varphi}_G)$ as described below. Define the set $N := \{(v, w) : v < w \text{ and } (v, w) \notin E\}$. These are known as the "non-edges" of the DAG G and we will equivalently denote $(v, w) \in N$ by $v \not\to w$. Also, define $D := \{(v, v) : v \in V\}$. Note that $D \cup E \cup N$ constitute the position coordinates of the upper triangular half of an $m \times m$ symmetric matrix. The rows of $J(\tilde{\varphi}_G)$ are indexed from top to down in the order: D, E and N. The columns of $J(\tilde{\varphi}_G)$ are indexed such that partial derivatives with respect to the free input variables in the triple (Λ, Ψ, γ) appear from left to right, in the order: Ψ, Λ and γ. Hence we have partitioned $J(\tilde{\varphi}_G)$ into 9 blocks:

$$J(\tilde{\varphi}_G) = \begin{array}{c} \\ D \\ E \\ N \end{array} \begin{array}{c} \begin{array}{ccc} \Psi & \Lambda & \gamma \end{array} \\ \begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \end{array}. \tag{3.6}$$

The following lemma is simply derived by taking derivatives of the function $\tilde{\varphi}_G$. Its proof appears in Section 6.

Lemma 3.4. *The Jacobian matrix $J(\tilde{\varphi}_G)$ is generically of full column rank provided that the submatrix $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is so.*

We now give the proof of Theorem 1.3.

Proof of Theorem 1.3. By Lemma 3.4 and Lemma 1.1, it suffices to show that $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is generically of full column rank. For each $v \not\to w \in N$,

$$[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vw} = \left[(I_m - \Lambda)\Psi(I_m - \Lambda^T)\right]_{vw} - \gamma_v \gamma_w. \tag{3.7}$$

Note that taking partial derivatives of these with respect to $\gamma = (\gamma_v)_{v \in \{1,\ldots,m\}}$ gives rise to the block $[J(\tilde{\varphi}_G)]_{N,\gamma}$ in (3.6). Ignoring the directionality of non-edges in N, $G^c = (V, N)$. Suppose we define the vector-valued function $f = (f_{vw})^T_{(v,w) \in N}$, and each component function f_{vw} takes as input the variable pairs (γ_v, γ_w) such that

$$f_{vw}(\gamma_v, \gamma_w) = \gamma_v \gamma_w.$$

By Lemma 3.3, the Jacobian matrix of f, $J(f)$, is generically of full column rank m, since all connected components of G^c contain an odd cycle.

Since for each $v \not\to w \in N$, the term $[(I_m - \Lambda)\Psi(I_m - \Lambda^T)]_{vw}$ in (3.7) does not involve any variables in γ, we can see that $[J(\tilde{\varphi}_G)]_{N,\gamma} = -J(f)$, hence $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is also generically of full column rank. □

We remark that, in fact, Theorem 1.3 can also be arrived at by working with the map $\tilde{\phi}_G$ in (3.1), by proving a lemma similar to Lemma 3.4 corresponding to its Jacobian matrix $J(\tilde{\phi}_G)$. We chose to work with $\tilde{\varphi}_G$ since, unlike $\tilde{\phi}_G$, the map $\tilde{\varphi}_G$ doesn't involve taking inverse of the matrix $I_m - \Lambda$, leading to a simpler derivation of the corresponding Jacobian matrix in the proof of Lemma 3.4. The proof of Theorem 1.4, nevertheless, will use both $\tilde{\varphi}_G$ and $\tilde{\phi}_G$.

Proof of Theorem 1.4. We will first prove the necessity of condition (i) by showing that if $|E_{con}| - |E| < d_{con}$, the Jacobian matrix $J(\tilde{\varphi}_G)$ will have row rank less than $2m + |E|$, and hence it cannot be of full column rank which implies the failure of generic finite identifiability by Lemma 1.1.

We partition the edge set N as $N = N_1 \,\dot{\cup}\, N_2$, where $N_1 = \{v \not\to w \in N : v - w \in E_{con}\}$, and $N_2 = N \backslash N_1$. Then we can further partition the submatrix $[J(\tilde{\varphi}_G)]_{N,\{\Psi,\Lambda,\gamma\}}$ into two block rows indexed by N_1 and N_2 as

$$[J(\tilde{\varphi}_G)]_{N,\{\Psi,\Lambda,\gamma\}} = \begin{array}{c c} & \begin{array}{ccc} \Psi & \Lambda & \gamma \end{array} \\ \begin{array}{c} N_1 \\ N_2 \end{array} & \begin{bmatrix} \cdots & \cdots & \cdots \\ 0 & 0 & \cdots \end{bmatrix} \end{array}.$$

Note that the block $[J(\tilde{\varphi}_G)]_{N_2,\{\Psi,\Lambda\}} = 0$ since for any $v \not\to w \in N_2$, $[\Sigma^{-1}_{|L}]_{vw} = 0$ and $(I - \Lambda)\Psi(I - \Lambda^T)$ has the same zeros structure as $\Sigma^{-1}_{|L}$. With this, it suffices to show that under the stated condition, the rank of $[J(\tilde{\varphi}_G)]_{N_2,\gamma}$ is $m - d_{con}$, for if it is true, then there exists a subset $N_2' \subset N_2$ with $|N_2'| = m - d_{con}$, such that row deleted matrix

$$[J(\tilde{\varphi}_G)]_{\{D,E,N_1,N_2'\},\{\Psi,\Lambda,\gamma\}}$$

has the same rank as the original Jacobian matrix $J(\tilde{\varphi}_G)$. Since the row deleted matrix $[J(\tilde{\varphi}_G)]_{\{D,E,N_1,N_2'\},\{\Psi,\Lambda,\gamma\}}$ has $2m + |E_{con}| - d_{con}$ rows, its rank, under the stated condition $|E_{con}| - |E| < d_{con}$, is less than $2m + |E|$. As a result, $J(\tilde{\varphi}_G)$ cannot be of full column rank.

Now it remains to show that $\operatorname{rank}([J(\tilde{\varphi}_G)]_{N_2,\gamma}) = m - d_{con}$. By ignoring directionality of the non-edges in N_2, $G^c_{con} = (V, N_2)$, since every edge in E can also be treated as an undirected edge in E_{con}. Now for $v - w \in N_2$, consider the functions

$$[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vw} = [(I_m - \Lambda)\Psi(I_m - \Lambda^T)]_{vw} - \gamma_v\gamma_w = -\gamma_v\gamma_w,$$

where the last equality follows from the fact that $[(I_m - \Lambda)\Psi(I_m - \Lambda^T)]_{vw} = 0$, as $v - w$ is not an edge in G_{con}. Hence

$$f(\gamma) := ([\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vw})^T_{(v,w)\in N_2}$$

can be treated as a vector valued function defined on the variable vector γ alone. By Lemma 3.3, the Jacobian of $-f$, $J(-f)$, is generically of rank $m - d_{con}$. Note that $[J(\tilde{\varphi}_G)]_{N_2,\gamma} = J(f)$, we are done with proving the necessity of (i).

The proof of (ii) follows the exact same argument as that of (i), by replacing (a) G_{con} with $G_{|L,cov}$, (b) d_{con} with d_{cov}, (c) $\tilde{\varphi}_G$ with $\tilde{\phi}_G$, (d) Ψ with Ω, (e) γ with

δ and (f) $J(\tilde{\varphi}_G)$ with $J(\tilde{\phi}_G)$, where $J(\tilde{\phi}_G)$ is partitioned as

$$J(\tilde{\phi}_G) = \begin{array}{c} \\ D \\ E \\ N \end{array} \begin{array}{c} \begin{array}{ccc} \Omega & \Lambda & \delta \end{array} \\ \begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \end{array} \tag{3.8}$$

similarly to (3.6). □

By Gröbner basis computations Cox et al. (2007), we have performed an exhaustive study on the generic finite identifiability of the $\mathcal{N}_*(G)$ models for all DAGs $G = (V, E)$ with $m = 4, 5, 6$ nodes. Of course, we omitted DAGs with $\binom{m}{2} - 2m < |E|$ since the parametrization map of their models are trivially infinite-to-one by Corollary 1.2. Via these algebraic calculations, we will always be able to tell whether a given model $\mathcal{N}_*(G)$ is generically finitely identifiable. We tested our conditions in Theorem 1.3 and Theorem 1.4 on a DAG in each isomorphic class with respect to relabeling of nodes, and the results are tabulated in Table 1.

Table 1. For each $m = 4, 5, 6$, the number of unlabeled DAGs $G = (V, E)$ with $\binom{m+1}{2} - 2m \geq |E|$ satisfying each classifying criterion is given.

m	4	5	6
Generically finite-to-one	5	95	3344
✓ SW	5	49	985
✓ OS	5	88	2957
✓ OS and SW	5	49	985
Generically ∞-to-one	1	20	552
X ON	1	20	361
Total # of DAGs	6	115	3896
✓ ON	5	95	3535

*OS = **O**ur **S**ufficient condition in Theorem 1.3

*ON = **O**ur **N**ecessary condition in Theorem 1.4

*SW = **S**tanghellini and **W**ermuth's sufficient condition in Corollary 2.2

*✓ = satisfy, X = not satisfy

For all $m = 4, 5, 6$, our sufficient condition in Theorem 1.3 is very successful in singling out DAGs that are determined to be generically finitely identifiable via Gröbner basis computations. For instance when $m = 6$, it is able to single 2957 out of 3344 such graphs, considerably more powerful than the sufficient condition in Corollary 2.2 of Stanghellini and Wermuth (2005) which only manages to single out 985 of those. Moreover, for each $m = 4, 5, 6$, the number of graphs that satisfy both OS and SW verifies the fact that our sufficient condition is strictly stronger than that of Stanghellini and Wermuth for generic finite identifiability.

Our necessary condition in Theorem 1.4 is also useful in singling out graphs that give generically infinite-to-one models. For instance, when $m = 6$, 361 out of 552 such graphs violate that condition.

4. SUBGRAPH EXTENSION

This section concerns results on how we can extend knowledge about identifiability of an induced subgraph to that of the original DAG. Recall the following standard graphical terms: For a given DAG $G = (V, E)$,

(i) $s \in V$ is called a sink node if there is no edge of the form $s \to s' \in E$,
(ii) $s \in V$ is called a source node if there is no edge of the form $s' \to s \in E$,
(iii) $pa(v) = \{w : w \to v \in E\}$ is the parent set of the node v
(iv) $ch(v) = \{w : v \to w \in E\}$ is the children set of the node v

The main result of this section is stated as follows:

Theorem 4.1. *Given a DAG $G = (V, E)$, if either*

(i) there exists a sink node $s \in V$ such that $pa(s) \neq V\backslash\{s\}$ and the model $\mathcal{N}_*(G')$ of the induced subgraph G' on $V\backslash\{s\}$ is generically finitely identifiable, or
(ii) there exists a source node $s \in V$ such that $ch(s) \neq V\backslash\{s\}$ and the model $\mathcal{N}_*(G')$ of the induced subgraph G' on $V\backslash\{s\}$ is generically finitely identifiable,

then the model $\mathcal{N}_*(G)$ is generically finitely identifiable.

Recall that in Table 1, for $m = 6$, $3344 - 2957 = 387$ DAGs not satisfying our sufficient condition in Theorem 1.3 give generically finitely identifiable models. With Theorem 4.1 we may ask: Given we have knowledge of the generic finite identifiabilty statuses of all DAGs with $m = 5$ nodes, how many of the 387 graphs above can we single out as generically finitely identifiable by applying Theorem 4.1? It turns out that 194 of them can have generic finite identifiability be told in this way.

Theorem 4.1 is obtained by studying the maps ϕ_G and φ_G in (1.3) and (3.2). Consider (1.3). For a DAG G and a generic point $(\Lambda_0, \Omega_0, \delta_0) \in \Theta$, ϕ_G is generically finite-to-one, if the equation

$$\phi_G(\Lambda_0, \Omega_0, \delta_0) = (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1}, \tag{4.1}$$

or equivalently,

$$(I_m - \Lambda^T)\phi_G(\Lambda_0, \Omega_0, \delta_0)(I_m - \Lambda) = \Omega + \delta\delta^T, \tag{4.2}$$

has finitely many solutions for $(\Lambda, \Omega, \delta)$ in Θ. Trivially, $(\Lambda, \Omega, \delta) = (\Lambda_0, \Omega_0, \delta_0)$ must satisfy (4.2). Generically, we can always assume the vector δ has no zero elements, in that case the matrix on the right hand side of (4.2) is known as a Spearman matrix.

Definition 4.1. A $m \times m$ matrix Υ is a Spearman matrix if it admits the structure $\Upsilon = \Omega + \delta\delta^T$, where Ω is a diagonal matrix with a positive diagonal and δ is a vector with no zero elements. Ω and $\delta\delta'$ are known as the diagonal and the rank-1 components of Υ respectively.

For our purpose, Definition 4.1 deviates slightly from the literature, which only requires Ω to have a non-negative diagonal entries and δ to be a non-zero vector. Any Speaman matrix is necessarily positive definite and symmetric. It is an elementary result that (for instance, Theorem 5.5 in Anderson and Rubin (1956)), up to sign change in δ, the pair (Ω, δ) can be uniquely recovered as a rational function of Υ, which justify our reference to Ω and $\delta\delta'$ as "the" diagonal and rank-1 components. The following version of Theorem 1 in Bekker and de Leeuw (1987) gives an alternative characterization for Spearman matrices.

Theorem 4.2. *For $m \geq 4$, a $m \times m$ positive definite and symmetric matrix $\Upsilon = (\upsilon_{ij})$ is a Spearman matrix if and only if, after sign changes of rows and corresponding columns, all its elements are positive and such that*

$$\upsilon_{ij}\upsilon_{kl} - \upsilon_{ik}\upsilon_{jl} \tag{4.3}$$

$$= \upsilon_{il}\upsilon_{jk} - \upsilon_{ik}\upsilon_{jl} \tag{4.4}$$

$$= \upsilon_{ij}\upsilon_{kl} - \upsilon_{il}\upsilon_{jk} = 0 \tag{4.5}$$

for $i < j < k < l$, and

$$\upsilon_{ii}\upsilon_{jk} - \upsilon_{ik}\upsilon_{ji} > 0 \tag{4.6}$$

for $i \neq j \neq k$.

Let $\mathrm{SPR}(m)$ denote the set of all $m \times m$ Spearman matrices. The polynomial expressions in (4.3), (4.4) and (4.5) are the 2×2 off-diagonal (i.e., does not involve any diagonal entries) minors of the matrix Υ, known as tetrads in the literature. Each quadruple $i < j < k < l$ will be referred to as the indices of a tetrad. Note that the tetrad in (4.5), $\upsilon_{ij}\upsilon_{kl} - \upsilon_{il}\sigma_{jk}$, is the difference between the tetrads in (4.3) and (4.4), so they are algebraically dependent. In general, a symmetric $m \times m$ matrix will have $2\binom{m}{4}$ algebraically independent tetrads. Vanishing tetrads are the defining constraints for a matrix to be in the Zariski closure of $\mathrm{SPR}(m)$ (Drton et al., 2007). For any given $m \times m$ matrix Υ, let's use $\mathrm{TETRADS}(\Upsilon)$ to denote the unique column vector of length $2\binom{m}{4}$ that enumerates all $2\binom{m}{4}$ algebraically independent tetrads, unique up to permutation of its entries and choices of independent tetrads.

For each triple $(\Lambda, \Omega, \delta)$ that solves (4.2), it must be true that

$$\mathrm{TETRADS}\left((I_m - \Lambda^T)\phi_G(\Lambda_0, \Omega_0, \delta_0)(I_m - \Lambda)\right) = 0 \tag{4.7}$$

by Theorem 4.2. Together with the uniqueness of the diagonal and rank-1 components for a Spearman matrix, if we can show only finitely many Λ's solve the system (4.7), then we have shown that the model $\mathcal{N}_*(G)$ is generically finitely identifiable. Our proof for (i) of Theorem 4.1 follows this approach.

Alternatively, with Lemma 3.2, by considering (3.2) we can also prove generic finite identifiability by showing that for a generic point $(\Lambda_0, \Psi_0, \gamma_0) \in \Theta$, the equation

$$\varphi_G\left(\Lambda_0, \Psi_0, \gamma_0\right) = (I_m - \Lambda)(\Psi - \gamma\gamma^T)(I_m - \Lambda^T), \tag{4.8}$$

or equivalently,

$$(I_m - \Lambda)^{-1}\varphi_G\left(\Lambda_0, \Psi_0, \gamma_0\right)(I_m - \Lambda^T)^{-1} = \Psi - \gamma\gamma^T \tag{4.9}$$

has finitely many solutions for (Λ, Ψ, γ) in Θ. We define our own notion of coSpearman matrix to describe the matrix on the right hand side of (4.9).

Definition 4.2. A $m \times m$ matrix Υ is a coSpearman matrix if it admits the structure $\Upsilon = \Psi - \gamma\gamma^T$, where Ψ is a diagonal matrix with a positive diagonal and γ is a vector with no zero elements. Ψ and $\gamma\gamma^T$ are known as the diagonal and the rank-1 components of Υ respectively.

Again, in the above definition, it is justified to refer to Ψ and $\gamma\gamma^T$ as "the" diagonal and rank-1 components of Υ respectively, since, up to sign change in γ, (Ψ, γ) can be uniquely recovered as a rational function of Υ (See for example Stanghellini (1997)). We also have the following theorem analogous to Theorem 4.2:

Theorem 4.3. *For $m \geq 4$, a $m \times m$ positive definite and symmetric matrix $\Upsilon = (\upsilon_{ij})$ is a coSpearman matrix if and only if, after sign changes of rows and corresponding columns, all its non-diagonal elements are negative and such that*

$$\upsilon_{ij}\upsilon_{kl} - \upsilon_{ik}\upsilon_{jl} \tag{4.10}$$

$$= \upsilon_{il}\upsilon_{jk} - \upsilon_{ik}\upsilon_{jl} \tag{4.11}$$

$$= \upsilon_{ij}\upsilon_{kl} - \upsilon_{il}\upsilon_{jk} = 0 \tag{4.12}$$

for $i < j < k < l$, and

$$\upsilon_{ii}\upsilon_{jk} - \upsilon_{ik}\upsilon_{jl} < 0 \tag{4.13}$$

for $i \neq j \neq k$.

Proof. The thoerem can be proved by mimicking the proof for Theorem 4.2 as in Bekker and de Leeuw (1987). □

Analogously, by the tetrad characterizations (4.10), (4.11) and (4.12), together with the uniqueness of the diagonal and rank-1 components, we can show generic finite identifiability by showing that the system of tetrad equations for the left hand side of (4.9),

$$\text{TETRADS}\Big((I_m - \Lambda)^{-1}\varphi_G\left(\Lambda_0, \Psi_0, \gamma_0\right)\left(I_m - \Lambda^T\right)^{-1}\Big) = 0, \tag{4.14}$$

admits only finitely many Λ as solutions.

Note that the finiteness of solutions in Λ for the system (4.7), or (4.14), is only a sufficient condition for the generic finite identifiabiility of $\mathcal{N}_*(G)$. It is not obvious that the above two systems necessarily have finitely many solutions when $\mathcal{N}_*(G)$ is generically finitely identifiable. The following lemma, which will be used in the proof of Theorem 4.1, claims that such converse implication must hold for the following two types of DAGs, whose generic finite identifiability of their models can be easily checked by Theorem 1.3. Let "$\subsetneq$" denote "being a proper subset of".

Lemma 4.4. *Let $G = (V, E)$ be a DAG.*

(i) Suppose $E \subsetneq \{(1,m), (2,m), \dots, (m-2,m), (m-1,m)\}$, then for a generic point $(\Lambda_0, \Omega_0, \delta_0) \in \Theta := \mathbb{R}_E \times \text{diag}_m^+ \times \mathbb{R}^m$ with $\Lambda_0 = (\lambda_{vw}^0)$, the system

$$TETRADS\Big((I_m - \Lambda^T)\phi_G\left(\Lambda_0, \Omega_0, \delta_0\right)\left(I_m - \Lambda\right)\Big) = 0$$

is linear in the variables $\lambda_{pa(m),m} := (\lambda_{vm})^T_{v\in pa(m)}$ and has the unique solution $\lambda^0_{pa(m),m} := (\lambda^0_{vm})^T_{v\in pa(m)}$.

(ii) Suppose $E \subsetneq \{(1,2), (1,3), \dots, (1,m-1), (1,m)\}$, then for a generic point $(\Lambda_0, \Psi_0, \gamma_0) \in \Theta := \mathbb{R}_E \times \text{diag}_m^+ \times \mathbb{R}^m$ with $\Lambda_0 = (\lambda_{vw}^0)$, the system

$$TETRADS\Big((I_m - \Lambda)^{-1}\varphi_G\left(\Lambda_0, \Psi_0, \gamma_0\right)\left(I_m - \Lambda^T\right)^{-1}\Big) = 0$$

is linear in the variables $\lambda_{1,ch(1)} := (\lambda_{1v})^T_{v\in ch(1)}$ and has the unique solution $\lambda^0_{1,ch(1)} := (\lambda^0_{1v})^T_{v\in ch(1)}$.

The proof of Lemma 4.4 will be deferred to the last section. We will now prove Theorem 4.1. Given a subset $S \subset V$ and the vector $x = (x_1, \dots, x_m)^T$, we will define x_S to be the subvector $(x_v)^T_{v \in S}$, and use the shorthand $[m] := \{1, \dots, m\}$ for any natural number m.

Proof of Theorem 4.1. We will first prove (i), which uses Lemma 4.4(i). The proof of (ii) will follow a similar pattern using Lemma 4.4(ii).

Without loss of generality, we may assume $s = m$ by giving the nodes a new topological order when necessary. Define the following two DAGs: Let $V_1 = V \backslash \{m\} = \{1, \dots, m-1\}$, $G_1 = (V_1, E_1)$ be the subgraph of G induced on the node set V_1, and $G_2 = (V, E \backslash E_1)$ be the graph on V obtained from G by taking away all the edges that do not have the sink node m as the head. As usual, let $\Theta := \mathbb{R}_E \times \mathrm{diag}^+_m \times \mathbb{R}^m$. We will construct a proper algebraic subset Ξ of Θ, such that the restricted map $\phi_G|_{\Theta \backslash \Xi}$ has finite fiber, i.e., for any $\theta \in \Theta \backslash \Xi$, the set

$$\mathcal{F}(\theta) := \{\theta' : \phi_G(\theta') = \phi_G(\theta) \text{ and } \theta' \in \Theta \backslash \Xi\}$$

is finite, and then (ii) of Lemma 1.1 applies.

Let $\Theta_1 := \mathbb{R}_{E_1} \times \mathrm{diag}^+_{m-1} \times \mathbb{R}^{m-1}$, the open set on which ϕ_{G_1} is defined. Since $\mathcal{N}_*(G_1)$ is generically finitely identifiable, there exists an algebraic subset $\Xi_1 \subset \Theta_1$ with $\dim(\Xi_1) < \dim(\Theta_1)$ such that $\phi_{G_1}|_{\Theta_1 \backslash \Xi_1}$ has finite fibers, by (ii) of Lemma 1.1. Extend Ξ_1 to a proper algebraic subset of Θ by defining

$$\widetilde{\Xi}_1 := \Xi_1 \times \mathbb{R}^{|E \backslash E_1|} \times \mathbb{R}^+ \times \mathbb{R},$$

where $\mathbb{R}^{|E \backslash E_1|}$, $\mathbb{R}^+$ and $\mathbb{R}$ accommodate the additional free variables $\lambda_{pa(m),m} := (\lambda_{vm})^T_{v \in pa(m)}$, ω_m and δ_m respectively.

On the other hand, if we define $\lambda_{E_1} := (\lambda_{vw})^T_{(v,w) \in E_1}$, any tetrad of the matrix $(I_m - \Lambda^T)\phi_G(\Lambda, \Omega, \delta)(I_m - \Lambda)$ with indices $i < j < k < m$ looks like

$$a\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right) \lambda_{m'm} - b\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right),$$

with $m' \in pa(m)$, and a, b being monomials in entries of λ_{E_1} and $\phi_G(\Lambda, \Omega, \delta)$. Hence, for $(\Lambda, \Omega, \delta) \in \Theta$, the part of the system

$$\mathrm{TETRADS}\left((I_m - \Lambda^T)\phi_G(\Lambda, \Omega, \delta)(I_m - \Lambda)\right) = 0$$

that involves the variables $\lambda_{pa(m),m}$ has the form

$$C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right) \lambda_{pa(m),m} = c\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right), \tag{4.15}$$

where C is a $2\binom{m-1}{3} \times |pa(m)|$ matrix and c is a $2\binom{m-1}{3}$-vector, both with elements that are monomial functions in $\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)$. By Lemma 4.4(i) and the assumption that $pa(m) \subsetneq V \backslash \{s\}$, C is of full rank when we set λ_{E_1} to be zero and pick a generic point $\lambda_{pa(m),m} \in \mathbb{R}^{|E \backslash E_1|}$, since then (4.15) will become the system of tetrad equations for the graph G_2. Hence C is generically of full rank on the space Θ, and we let $\widetilde{\Xi}_2$ be the proper algebraic subset of Θ such that $C\left(\lambda_{E_1}, \phi_G(\Lambda, \Omega, \delta)\right)$ is of full rank for any $(\Lambda, \Omega, \delta) \in \Theta \backslash \widetilde{\Xi}_2$. Define $\Xi := \widetilde{\Xi}_1 \cup \widetilde{\Xi}_2$, which is a proper algebraic subset of Θ.

Let $(\Lambda_0, \Omega_0, \delta_0)$ be a point in $\Theta \backslash \Xi$ and $\Sigma_0 = \phi_G(\Lambda_0, \Omega_0, \delta_0)$. It remains to show that the matrix equation system

$$\Sigma_0 = (I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1} \tag{4.16}$$

has only finitely many solutions in $(\Lambda, \Omega, \delta)$ over the set $\Theta \backslash \Xi$. The subsystem

$$\begin{aligned}
&(\Sigma_0)_{[m-1],[m-1]} \\
&= [(I_m - \Lambda^T)^{-1}(\Omega + \delta\delta^T)(I_m - \Lambda)^{-1}]_{[m-1],[m-1]} \\
&= (I_{m-1} - \Lambda^T_{[m-1],[m-1]})^{-1}(\Omega_{[m-1],[m-1]} + \delta_{[m-1]}\delta^T_{[m-1]})(I_{m-1} - \Lambda_{[m-1],[m-1]})^{-1},
\end{aligned}$$

where the second equality is true by m being a sink node, can at most have finitely many solutions in $(\lambda_{E_1}, \omega_1, \ldots, \omega_{m-1}, \delta_{[m-1]})$ over the set $\Theta \backslash \Xi$, in view of the fact that

$$\begin{aligned}
&(\Sigma_0)_{[m-1],[m-1]} \\
&= (I_{m-1} - \Lambda_0{}^T_{[m-1],[m-1]})^{-1}(\Omega_{0[m-1],[m-1]} + \delta_{0[m-1]}\delta_0{}^T_{[m-1]})(I_{m-1} - \Lambda_{0[m-1],[m-1]})^{-1},
\end{aligned}$$

$(\Lambda_{0[m-1],[m-1]}, \Omega_{0[m-1],[m-1]}, \delta_{0[m-1]}) \in \Theta_1 \backslash \Xi_1$ and $\phi_{G_1}|_{\Theta_1 \backslash \Xi_1}$ has finite fibers. Let $\mathcal{S}$ be this finite set of solutions in $(\lambda_{E_1}, \omega_1, \ldots, \omega_{m-1}, \delta_{[m-1]})$. Now it remains to show that for each $(\lambda_{E_1}, \omega_1, \ldots, \omega_{m-1}, \delta_{[m-1]}) \in \mathcal{S}$ that can be extended to a solution $(\Lambda, \Omega, \delta) \in \Theta \backslash \Xi$ for the whole system (4.16), there can only be finitely many $(\lambda^T_{pa(m),m}, \omega_m, \delta_m)^T$ for such an extension.

Let $s \in \mathcal{S}$ be extendable and $(s, \lambda^T_{pa(m),m}, \omega_m, \delta_m) \in \Theta \backslash \Xi$ be any solution of (4.16). In particular, the system of tetrad equations

$$\text{TETRADS}\left((I_m - \Lambda^T)\Sigma_0(I_m - \Lambda)\right) = 0$$

must be satisfied by $(s, \lambda^T_{pa(m),m}, \omega_m, \delta_m)$. Since $(s, \lambda^T_{pa(m),m}, \omega_m, \delta_m) \notin \tilde{\Xi}_2$, $\lambda_{pa(m),m}$ is uniquely determined by how $\tilde{\Xi}_2$ is defined. Finally, (ω_m, δ_m) is also uniquely determined, as they are part of the right hand side of

$$(I_m - \Lambda^T)\Sigma_0(I_m - \Lambda) = \Omega + \delta\delta^T,$$

where the left hand side, a Spearman matrix, has been uniquely determined. We are done with proving (i).

The proof of (ii) is exactly analogous and we will only give a sketch here. Instead of considering ϕ_G we will turn to φ_G, which is also defined on the domain $\Theta = \mathbb{R}_E \times \text{diag}^+_m \times \mathbb{R}^m$. Without loss of generality we will let $s = 1$. Let $G_1 = (V_1, E_1)$ be the subgraph induced on the nodes $V_1 = \{2, \ldots, m\}$ by G, and $G_2 = (V, E \backslash E_1)$, then φ_{G_1} is defined on the domain $\Theta_1 = \mathbb{R}_{E_1} \times \text{diag}^+_{m-1} \times \mathbb{R}^{m-1}$ and φ_{G_2} is defined on Θ. By assumption the $\mathcal{N}_*(G_1)$ is generically finitely identifiable, so there exists a proper algebraic subset $\Xi_1 \subset \Theta_1$ such that $\varphi_{G_1}|_{\Theta \backslash \Xi_1}$ has finite fibers, by (ii) of Lemma 1.1. Extend Ξ_1 to a proper algebraic subset $\tilde{\Xi}_1 \subset \Theta$ by defining

$$\tilde{\Xi}_1 := \Xi_1 \times \mathbb{R}^{|E \backslash E_1|} \times \mathbb{R}^+ \times \mathbb{R}.$$

On the other hand, by defining $\lambda_{E_1} := (\lambda_{vw})^T_{(v,w) \in E_1}$ for any $(\Lambda, \Psi, \gamma) \in \Theta$, the part of the system

$$\text{TETRADS}\left((I_m - \Lambda)^{-1}\varphi_G(\Lambda, \Psi, \gamma)(I_m - \Lambda^T)^{-1}\right) = 0$$

that involves the variables $\lambda_{1,ch(1)} := (\lambda_{1v})^T_{v \in ch(1)}$ has the form

$$C\left(\lambda_{E_1}, \varphi_G(\Lambda, \Psi, \gamma)\right)\lambda_{1,ch(1)} = c\left(\lambda_{E_1}, \varphi_G\left(\Lambda, \Psi, \gamma\right)\right), \tag{4.17}$$

where C is a $2\binom{m-1}{3} \times |ch(1)|$ matrix and c is a $2\binom{m-1}{3}$-vector, both have elements being monomial functions in Λ_{E_1} and $\varphi_G(\Lambda, \Psi, \gamma)$. By part (ii) of Lemma 4.4,

$C(\lambda_{E_1}, \varphi_G(\Lambda, \Psi, \gamma))$ is generically of full rank. Hence, we may find a algebraic subset $\widetilde{\Xi}_2 \subsetneq \Theta$ such that C is full rank for any $(\Lambda, \Psi, \gamma) \in \Theta \backslash \widetilde{\Xi}_2$.

Then we form $\Xi := \widetilde{\Xi}_1 \cup \widetilde{\Xi}_2$, a proper algebraic subset of Θ. By going through similar argument as the proof for (i), we can argue that $\varphi_G|_{\Theta \backslash \Xi}$ has finite fibers. Details are left to the readers. □

5. Discussion

Although our sufficient criterion in Theorem 1.3 is more powerful in certifying the identifiability status of a model than the sufficient criteria of Corollary 2.2 by Stanghellini and Wermuth (2005), nevertheless, it doesn't suggest a procedure to solve for the parameter values $(\Lambda, \Omega, \delta)$ for a given covariance matrix Σ of the observables. In addition, it doesn't provide the cardinality k of the finite fiber of a generic parameter in Θ, if ϕ_G is certified to be generically finite-to-one. Per our discussion in Section 1, the most intuitive guess for k is 2, where the source of non-injectivity of ϕ_G is due to sign change in δ alone. However, our exhaustive algebraic study on DAGs with $m = 4, 5, 6$ using Gröbner basis computations suggests that most generically finitely identifiable models $\mathcal{N}_*(G)$ have $k = 2$, but exceptions do exist. This matter can be further explored.

In this paper we have studied directed Gaussian graphical model with a common latent source that is a parent to all the observables. In particular, all the edges $\{0 \to v : v = 1, \dots, m\}$ are present in the extended DAG $\overline{G}$. On examining the proofs of Theorem 1.3 and Theorem 1.4, one can actually see that these criteria can be extended to models where some of the edges $\{0 \to v : v = 1, \dots, m\}$ are missing in the graph $\overline{G}$. Assuming all other assumptions hold as before, for such models, the parametrization in (1.3) remains the same, with the exception that the vector δ contains some predetermined zero elements, i.e. $\delta_v \equiv 0$ whenever $0 \to v$ is not a directed edge of $\overline{G}$. We will state these results here, leaving the proof, which shares the same spirit as the proof of Corollary 1 in Grzebyk et al. (2004), to the readers.

Theorem 5.1 (Sufficient condition for generic finite identifiability, missing factor loading edges)**.** *Given a DAG $G = (V, E)$. Suppose V' is a subset of V such that $\delta_v \equiv 0$ in the model* (1.1) *for every $v \in V'$, which corresponds to missing edges $\{0 \to v : v \in V'\}$ in the extended DAG $\overline{G}$. The parametrization map ϕ_G defined on $\mathbb{R}_E \times \mathrm{diag}_m^+ \times \mathbb{R}^{m-|V'|}$ is generically finite-to-one, provided that every connected component of $G^c|_{V \backslash V'}$, the subgraph of G^c induced on $V \backslash V'$, contains an odd cycle.*

Theorem 5.2 (Necessary condition for generic finite identifiability, missing factor loading edges)**.** *Given a DAG $G = (V, E)$. Let $\overline{G}$ and V' be as in Theorem 5.1. For the parametrization map ϕ_G to be generically finite-to-one on the domain $\mathbb{R}_E \times \mathrm{diag}_m^+ \times \mathbb{R}^{m-|V'|}$, it is necessary that the following two conditions both hold:*

- *(i) Let $\widetilde{G}^c_{con} = (V \backslash V', \widetilde{E}_{con})$ be the subgraph of G^c_{con} induced on $V \backslash V'$. If d_{con} is the number of connected components in the graph $\widetilde{G}^c_{con}$ that do not contain any odd cycle, then $|\widetilde{E}_{con}| - |E| \geq d_{con}$.*
- *(ii) Let $\widetilde{G}^c_{|L,cov} = (V \backslash V', \widetilde{E}_{|L,cov})$ be the subgraph of G^c_{con} induced on $V \backslash V'$. If d_{cov} is the number of connected components in the graph $\widetilde{G}^c_{|L,cov}$ that do not contain any odd cycle, then $|\widetilde{E}_{|L,cov}| - |E| \geq d_{cov}$.*

6. Proofs

Proof of Lemma 1.1. It suffices to assume $d \geq n$, otherwise J_f can never be of full column rank.

$(i) \Rightarrow (ii)$ is obvious. $(ii) \Rightarrow (iii)$ is also true, since if towards a contradiction $\text{Rank}(J_f) = r < n$ on a set of generic points $S' \subset S$, then for any point $s \in S'$, by the rank theorem Rudin (1976), there exists a neighborhood $\mathcal{O}_s \subset S'$ of s, such that the restricted map $f|_{\mathcal{O}_s}$ has fibers of dimension $n - r > 0$, contradicting (ii). It remains to show $(iii) \Rightarrow (i)$.

We will show the following stronger statement: The set of ensemble of all infinite fibers, i.e.

$$\mathbb{F}_f := \{s \in \mathbb{R}^n : \left|f^{-1}(\{f(s)\})\right| = \infty\}$$

is contained in a proper algebraic subset of $\mathbb{R}^n$.

It suffices to assume $n = d$, for WLOG, we can assume that $\pi \circ f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ has a generically full rank Jacobian, where π is the projection onto the first n coordinates. Then $\mathbb{F}_f \subset \mathbb{F}_{\pi \circ f}$.

So assume $n = d$. Let $Z := \{s \in S : \text{Det}(J_f)(s) = 0\}$, where $\text{Det}(J_f)$ is treated as a polynomial on $\mathbb{R}^n$.

Claim. For a given point $y \in \mathbb{R}^n$, suppose $f^{-1}(\{y\}) \subset Z^c$, then $f^{-1}(\{y\})$ is finite.

This claim will be proved at the end. The converse statement of the claim is that if $y \in \mathbb{R}^n$ is a point such that $f^{-1}(\{y\})$ is infinite, then $f^{-1}(\{y\}) \cap Z \neq \varnothing$. As such, $\mathbb{F}_f \subset f^{-1}(f(Z))$. The rest is to show that $f^{-1}(f(Z))$ is contained in a proper algebraic subset of S. To that end, it suffices to show that $f(Z) \subset \tilde{S}$ for some proper algebraic subset $\tilde{S} \subset \mathbb{R}^n$, then $f^{-1}(f(Z)) \subset f^{-1}(\tilde{S})$. Trivially, $f^{-1}(\tilde{S})$ is a proper algebraic subset since if $f^{-1}(\tilde{S}) = S$, it will violate that f has full rank Jacobian on $S \backslash Z$ by inverse function theorem.

$\tilde{S}$ can be taken to be the Zariski closure of $f(Z)$ by the following facts: (i) Z is an algebraic set of dimension less than n. (ii) By property of semialgebraic maps, the dimension of $f(Z)$ is less than that of Z. (iii) Zariski closure of $f(Z)$ has the same dimension as Z.

Finally, the claim can be proved by contradiction as follows. Suppose $f^{-1}(\{y\})$ is not finite. Since it is an algebraic set, it must be of positive dimension, say $k > 0$. Then there must be a local region of $f^{-1}(\{y\})$ that is diffeomorphic to $\mathbb{R}^k$, via a function $g : \mathbb{R}^k \to f^{-1}(\{y\})$ with full rank Jacobian. Then the function composition $f \circ g : \mathbb{R}^k \to \{y\}$ will have a Jacobian of positive rank, which is impossible since the image of $f \circ g$ is a single point. □

Proof. For $i = 1, 2$, let $\overline{G}_i = (\overline{V}, \overline{E}_i)$ be the extended DAG of G_i, i.e., $\overline{V} = \{0, 1, \ldots, m\}$, and $\overline{E}_i = E_i \cup \{0 \to v : v \in \{1, \ldots, m\}\}$. Recall that a directed edge $v \to w$ is covered if $pa(w) = v \cup pa(v)$, where $pa(\cdot)$ is the standard notation for parent set. Since the set of covered edges in E_i is same as the set of covered edges in $\overline{E}_i$, by Chickering (1995), $\overline{G}_1$ and $\overline{G}_2$ are also Markov equivalent. Let $\mathcal{N}(\overline{G}_i)$ be the set of all zero mean $(m+1)$-variate Gaussian distributions that obey the global Markov property of $\overline{G}_i$ with the variance of node 0, or equivalently, the variance of the latent variable L, equal to 1. Note that $\mathcal{N}(\overline{G}_1) = \mathcal{N}(\overline{G}_2)$ by Markov equivalence, as such we can define the set of $(m+1) \times (m+1)$ matrices

$$\mathcal{S} := \{\Sigma : N(0, \Sigma) \in \mathcal{N}(\overline{G}_i)\}$$

for $i = 1$ or 2. For any given matrix in $\mathcal{S}$, we will label its rows and columns by $0, 1, \ldots, m$ in that order, corresponding to the nodes of $\overline{V}$. Define the map $\Phi_{\overline{G}_i} : \Theta_i \longrightarrow \mathcal{S}$ such that

$$\Phi_{\overline{G}_i}\big((\Lambda, \Omega, \delta)\big) = (I_{m+1} - \overline{\Lambda}^T)^{-1}\overline{\Omega}(I_{m+1} - \overline{\Lambda})^{-1},$$

where $\Theta_i := \mathbb{R}_{E_i} \times \mathrm{diag}_m^+ \times \mathbb{R}^m$, $\overline{\Lambda}$ is a $(m+1) \times (m+1)$ matrix such that

$$\overline{\Lambda}_{vw} = \begin{cases} \delta_w & \text{if } v = 0, w = 1, \ldots, m \\ \Lambda_{vw} & \text{if } v, w = 1, \ldots, m \\ 0 & \text{otherwise} \end{cases}$$

and $\overline{\Omega}$ is a diagonal matrix with $\overline{\Omega}_{00} = 1$ and $\overline{\Omega}_{vv} = \omega_v$ for $v = 1, \ldots m$. Let $\pi : \mathcal{S} \longrightarrow \mathrm{PD}_m$ be the projection of a $(m+1) \times (m+1)$ matrix in $\mathcal{S}$ onto the lower right $m \times m$ submatrix, i.e.

$$\pi(\Sigma) = \Sigma_{\{1,\ldots,m\},\{1,\ldots,m\}}.$$

Then the parametrization map for the one factor model of G_i, ϕ_{G_i}, equals the function compositions of π and $\Phi_{\overline{G}_i}$, i.e.

$$\phi_{G_i} = \pi \circ \Phi_{\overline{G}_i}.$$

By Lemma 2.1, $\Phi_{\overline{G}_i}$ is bijective with a rational inverse, so it is semialgebraic homeomorphic. So for $i = 1, 2$, by Lemma 3.1, $\phi_{G_i} = \pi \circ \Phi_{\overline{G}_i}$ is generically finite-to-one if and only if π is, so we are done. □

Proof of Lemma 3.4. Each of the sets D, E and N are ordered: For D, $(v, v) < (w, w)$ if $v < w$, and for E or N, a pair $(v, w) < (v', w')$ if either (i) $v < v'$, or if (ii) $v = v'$ and $w < w'$. Without loss of generality, we assume that the elements of each block row/column index set are arranged as follows:

Ψ: ψ_v is arranged to the left of ψ_w if $v < w$
Λ: λ_{vw} is arranged to the left of λ_{ux} if $v \to w < u \to x$
γ: γ_v is arranged to the left of γ_w when $v < w$
D: (v, v) is listed before (w, w) if $v < w$
E: $v \to w$ is listed before $u \to x$ if $v \to w < u \to x$
N: $v \not\to w$ is listed before $u \not\to x$ if $v \not\to w < u \not\to x$

We will first outline the structure of $J(\tilde{\varphi}_G)$ row block by row block.

(a) "$[J(\tilde{\varphi}_G)]_{D,\{\Psi,\Lambda,\gamma\}}$": For a given pair $(v, v) \in D$,

$$[\tilde{\varphi}_G(\Lambda, \Psi, \gamma)]_{vv} = \psi_v + \left(\sum_{w: v \to w \in E} \psi_w \lambda_{vw}^2\right) - \gamma_v^2.$$

Hence,

$$[J(\tilde{\varphi}_G)]_{(v,v),\psi_w} = \begin{cases} 1 & \text{if } v = w \\ \lambda_{vw}^2 & \text{if } v \to w \in E \\ 0 & \text{otherwise} \end{cases}, \tag{6.1}$$

$$[J(\tilde{\varphi}_G)]_{(v,v),\lambda_{wu}} = \begin{cases} 2\lambda_{wu}\psi_u & \text{if } v = w \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

and

$$[J(\tilde{\varphi}_G)]_{(v,v),\gamma_u} = \begin{cases} -2\gamma_u & \text{if } v = u \\ 0 & \text{otherwise} \end{cases}. \tag{6.3}$$

(b) "$[J(\tilde{\varphi}_G)]_{E,\{\Psi,\Lambda,\gamma\}}$": For any $v \to w \in E$,

$$[\tilde{\varphi}_G(\Lambda,\Psi,\gamma)]_{vw} = -\lambda_{vw}\psi_w + \left(\sum_{u: \substack{v\to u\in E \\ w\to u\in E}} \lambda_{vu}\lambda_{wu}\psi_u \right) - \gamma_v\gamma_w.$$

Hence,

$$[J(\tilde{\varphi}_G)]_{v\to w,\psi_u} = \begin{cases} -\lambda_{vw} & \text{if } u = w \\ \lambda_{vu}\lambda_{wu} & \text{if } v \to u \in E \text{ and } w \to u \in E \\ 0 & \text{otherwise} \end{cases}, \tag{6.4}$$

$$[J(\tilde{\varphi}_G)]_{v\to w,\lambda_{ux}} = \begin{cases} -\psi_w & \text{if } v = u, w = x \\ \lambda_{wx}\psi_x & \text{if } u = v,\ u \to x \in E \text{ and } w \to x \in E \\ \lambda_{vx}\psi_x & \text{if } u = w,\ u \to x \in E \text{ and } v \to x \in E \\ 0 & \text{otherwise} \end{cases} \tag{6.5}$$

and

$$[J(\tilde{\varphi}_G)]_{v\to w,\gamma_u} = \begin{cases} -\gamma_w & \text{if } v = u \\ -\gamma_v & \text{if } w = u \\ 0 & \text{otherwise} \end{cases}. \tag{6.6}$$

(c) "$[J(\tilde{\varphi}_G)]_{N,\{\Psi,\Lambda,\gamma\}}$": For any $v \not\to w \in N$,

$$[\tilde{\varphi}_G(\Lambda,\Psi,\gamma)]_{vw} = \left(\sum_{u: \substack{v\to u\in E \\ w\to u\in E}} \lambda_{vu}\lambda_{wu}\psi_u \right) - \gamma_v\gamma_w. \tag{6.7}$$

Hence,

$$[J(\tilde{\varphi}_G)]_{v\not\to w,\psi_u} = \begin{cases} \lambda_{vu}\lambda_{wu} & \text{if } v \to u\ ,\ w \to u \in E \\ 0 & \text{otherwise} \end{cases}, \tag{6.8}$$

$$[J(\tilde{\varphi}_G)]_{v\not\to w,\lambda_{ux}} = \begin{cases} \lambda_{wx}\psi_x & \text{if } u = v,\ u \to x,\ w \to x \in E \\ \lambda_{vx}\psi_x & \text{if } u = w,\ u \to x \in E \text{ and } v \to x \in E \\ 0 & \text{otherwise} \end{cases} \tag{6.9}$$

and

$$[J(\tilde{\varphi}_G)]_{v\not\to w,\gamma_u} = \begin{cases} -\gamma_w & \text{if } v = u \\ -\gamma_v & \text{if } w = u \\ 0 & \text{otherwise} \end{cases}. \tag{6.10}$$

With a slight abuse of notation, let $|\Psi|$, $|\gamma|$, $|\Lambda|$ denote the number of free variables in Ψ, γ and Λ respectively. Considering that $|D| = |\Psi|$ and $|E| = |\Lambda|$, $|N| \geq |\gamma|$ since $J(\tilde{\varphi}_G)$ is a tall matrix. If $[J(\tilde{\varphi}_G)]_{N,\gamma}$ is generically of full column rank, there exists a subset $N' \subset N$ such that $|N'| = |\gamma|$ and $\operatorname{Det}(J(\tilde{\varphi}_G)_{N',\gamma})$ is a nonzero polynomial in the variables of γ, in consideration of (6.10). Now it suffices to show that the $(2m+|E|)\times(2m+|E|)$ square submatrix $[J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}$ is generically of full rank, which is equivalent to that generically, its determinant is nonzero. In particular, it suffices to show that $\mathrm{Det}\big([J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}\big)$ is a nonzero polynomial in free variables of (Λ,γ) when we assume that $\psi_1=\cdots=\psi_m=1$. Let P denote the set of all permutation functions from the set $D\cup E\cup N'$ to the set of free variables of Λ, Ψ and γ, by Leibniz's formula for determinant, we have

$$\begin{aligned}
&\mathrm{Det}\big([J(\tilde{\varphi}_G)]_{\{D,E,N'\},\{\Psi,\Lambda,\gamma\}}\big)\\
&=\sum_{\sigma\in P}\mathrm{sgn}(\sigma)\prod_{s\in D\cup E\cup N'}J(\tilde{\varphi}_G)_{s,\sigma(s)}\\
&=\mathrm{Det}\big(J(\tilde{\varphi}_G)_{N',\gamma}\big)\left[\sum_{\sigma\in\tilde{P}}\mathrm{sgn}(\sigma)\prod_{s\in D\cup E}J(\tilde{\varphi}_G)_{s,\sigma(s)}\right]\\
&\quad+\sum_{\sigma\in P\setminus\tilde{P}}\mathrm{sgn}(\sigma)\prod_{s\in D\cup E\cup N'}J(\tilde{\varphi}_G)_{s,\sigma(s)}\\
&=\mathrm{Det}\big(J(\tilde{\varphi}_G)_{N',\gamma}\big)+\sum_{\sigma\in P\setminus\tilde{P}}\mathrm{sgn}(\sigma)\prod_{s\in D\cup E\cup N'}J(\tilde{\varphi}_G)_{s,\sigma(s)},
\end{aligned}\tag{6.11}$$

where $\tilde{P}$ is the subset of all permutations in P such that each $\sigma\in\tilde{P}$ has the property that $\sigma((v,v))=\psi_v$ for all $(v,v)\in D$ and $\sigma((v,w))=\lambda_{vw}$ for all $(v,w)\in E$, i.e. any such σ fixes the positions of the diagonal of the submatrix $[J(\tilde{\varphi}_G)]_{\{D,E\},\{\Psi,\Lambda\}}$. The equality in (6.11) follows from (6.1), (6.5) and the fact that we have assumed $\psi_1=\cdots=\psi_m=1$. Since any $\sigma\in P\setminus\tilde{P}$ does not fix all the diagonal positions of $[J(\tilde{\varphi}_G)]_{\{D,E\},\{\Psi,\Lambda\}}$, by (6.1) and (6.5), every summand in the second term of (6.11) is a polynomial term involving free variables of Λ. Recall that $\mathrm{Det}\big(J(\tilde{\varphi}_G)_{N',\gamma}\big)$ is a nonzero polynomial only in free variables of γ, it cannot be cancelled by the second term in (6.11), and we are done. □

Proof for Lemma 4.4. We will first prove (i). Let $\Sigma_0=\phi_G(\Lambda_0,\Omega_0,\delta_0)$ and

$$S=(I_m-\Lambda^T)\Sigma_0(I_m-\Lambda)=(s_{ij}),\tag{6.12}$$

it is easily seen that for $1\le i,j\le m$, $i<j$,

$$\begin{aligned}
s_{ij}&=\sum_{1\le k,k'\le m}\lambda_{ki}[\Sigma_0]_{kk'}\lambda_{k'j}-\sum_{1\le k\le m}[\Sigma_0]_{ik}\lambda_{kj}-\sum_{1\le k\le m}\lambda_{ki}[\Sigma_0]_{kj}+[\Sigma_0]_{ij}\\
&=\begin{cases}-\sum_{(k,m)\in E}[\Sigma_0]_{ik}\lambda_{km}+[\Sigma_0]_{im} & \text{if } j=m\\ [\Sigma_0]_{ij} & \text{if } j<m\end{cases},
\end{aligned}$$

where the last equality follows from the fact that λ_{ij} are nonzero only when $(i,j)\in E$. Hence for any quadruple (i,j,k,l), with $1\le i<j<k<l\le m$, on the left hand side, the corresponding tetrad equation system

$$\begin{aligned}
s_{ij}s_{kl}-s_{ik}s_{jl}&=0\\
s_{il}s_{jk}-s_{ik}s_{jl}&=0
\end{aligned}$$

either has constant polynomials when $l\ne m$, or has degree 1 polynomials in the variables $\{\lambda_{vm}:(v,m)\in E\}$ when $l=m$. The entire system

$$\mathrm{TETRADS}(S)=0$$

is a thus a consistent linear system that can be represented as

$$A\lambda_{pa(m),m} = c, \tag{6.13}$$

where $\lambda_{pa(m),m} = (\lambda_{vm})^T_{v\in pa(m)}$ is the $|pa(m)|$-column vector of free Λ variables, A is a $2\binom{m-1}{3} \times |pa(m)|$ matrix and c is a $2\binom{m-1}{3}$-column vector, with both A and c depends on Σ_0. We are done with proving the claim on the linearity of the system.

To finish the proof we only need to show that (6.13) is uniquely solvable in $\lambda_{pa(m),m}$. By Theorem 1.3, $\mathcal{N}_*(G)$ is seen to be generically finitely identifiable. We will aim to contradict this fact if the system does not have a unique solution. Note that the solution set must be the translate of a linear subspace in $\mathbb{R}^{|E|}$. Let's denote this translated linear subspace as $\mathcal{S}$. Suppose towards a contradiction, $\mathcal{S}$ is of positive dimension. Note $\lambda^0_{pa(m),m} = (\lambda^0_{vm})^T_{v\in pa(m)} \in \mathcal{S}$ since it must solve the system. If we substitute $\Lambda = \Lambda_0$ into (6.12), we obtain

$$S_0 = (I_m - \Lambda_0^T)\Sigma_0(I_m - \Lambda_0) = (s^0_{ij}),$$

and in consideration of (4.6) in Theorem 4.2, it must be true that

$$s^0_{ii}s^0_{jk} - s^0_{ik}s^0_{ji} > 0 \text{ , for all } i \neq j \neq k. \tag{6.14}$$

Note the left hand side of (6.14) is a multivariate polynomial in the variables $\{\lambda^0_{vm} : v \in pa(m)\}$, we must be able to form a sufficiently small open ball $\mathcal{O}(\lambda^0)$ around $\lambda^0_{pa(m),m}$ in $\mathbb{R}^{|E|}$, such that for any $\Lambda = (\lambda_{vw}) \in \mathbb{R}_{|E|}$ with $\lambda_{pa(m),m} \in \mathcal{O}(\lambda^0)$,

$$s_{ii}s_{jk} - s_{ik}s_{ji} > 0 \text{ , for all } i \neq j \neq k,$$

where $S = (s_{ij})$ is as defined in (6.12). Since $\mathcal{S}$ is a translated linear space and $\lambda^0_{pa(m),m} \in \mathcal{S}$, $\mathcal{O}(\lambda^0) \cap \mathcal{S}$ is seen to be an infinite set in which every point $\lambda_{pa(m),m}$ leads to a matrix Λ that makes $(I_m - \Lambda^T)\Sigma_0(I_m - \Lambda)$ a Spearman matrix. Hence the system

$$(I_m - \Lambda^T)\Sigma_0(I_m - \Lambda) = \Omega + \delta\delta'$$

locally has infinitely many solutions, contradicting the generic finite identifiability of $\mathcal{N}_*(G)$.

We will sketch the proof of (ii) which is exactly analogous. We first let $\Upsilon_0 = \varphi_G(\Lambda_0, \Psi_0, \gamma_0)$ and define

$$\tilde{S} = (I_m - \Lambda)^{-1}\Upsilon_0(I_m - \Lambda^T)^{-1} = (\tilde{s}_{ij}). \tag{6.15}$$

On examining the structure of the matrix $\tilde{S}$ using the fact that $(I_m - \Lambda)^{-1} = I_m + \Lambda + \Lambda^2 \cdots + \Lambda^{m-1}$, it can be seen that system of tetrads

$$\text{TETRADS}(\tilde{S}) = \text{TETRADS}\big((I_m - \Lambda)^{-1}\Upsilon_0(I_m - \Lambda^T)^{-1}\big) = 0$$

is a linear system in the variables $\{\lambda_{1v} : v \in ch(1)\}$. Similar as before, using Theorem 1.3, we can prove by contradiction that the system can only have a unique solution in $\{\lambda_{1v} : v \in ch(1)\}$, in view of the inequality (4.13) in Theorem 4.3. We leave the details to the readers. $\square$

References

Anderson, T. W. and Rubin, H. (1956). "Statistical inference in factor analysis." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. V*, 111–150. University of California Press, Berkeley and Los Angeles.

Bekker, P. A. and de Leeuw, J. (1987). "The rank of reduced dispersion matrices." *Psychometrika*, 52(1): 125–135.

Chickering, D. M. (1995). "A transformational characterization of equivalent Bayesian network structures." In *Uncertainty in artificial intelligence (Montreal, PQ, 1995)*, 87–98. Morgan Kaufmann, San Francisco, CA.

Coste, M. (2002). "An introduction to semialgebraic geometry." *RAAG network school*, 145.

Cox, D., Little, J., and O'Shea, D. (2007). *Ideals, varieties, and algorithms*. Undergraduate Texts in Mathematics. Springer, New York, third edition. An introduction to computational algebraic geometry and commutative algebra.

Drton, M., Sturmfels, B., and Sullivant, S. (2007). "Algebraic factor analysis: tetrads, pentads and beyond." *Probab. Theory Related Fields*, 138(3-4): 463–493.

— (2009). *Lectures on algebraic statistics*, volume 39 of *Oberwolfach Seminars*. Birkhäuser Verlag, Basel.
URL `http://dx.doi.org/10.1007/978-3-7643-8905-5`

Grzebyk, M., Wild, P., and Chouanière, D. (2004). "On identification of multi-factor models with correlated residuals." *Biometrika*, 91(1): 141–151.

Lauritzen, S. L. (1996). *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York. Oxford Science Publications.

Pearl, J. (2009). *Causality*. Cambridge University Press, Cambridge, second edition. Models, reasoning, and inference.

Rao, C. R. (1973). *Linear statistical inference and its applications*. John Wiley & Sons, New York-London-Sydney, second edition. Wiley Series in Probability and Mathematical Statistics.

Rudin, W. (1976). *Principles of mathematical analysis*. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, third edition. International Series in Pure and Applied Mathematics.

Stanghellini, E. (1997). "Identification of a single-factor model using graphical Gaussian rules." *Biometrika*, 84(1): 241–244.

Stanghellini, E. and Wermuth, N. (2005). "On the identification of path analysis models with one hidden variable." *Biometrika*, 92(2): 337–350.

Vicard, P. (2000). "On the identification of a single-factor model with correlated residuals." *Biometrika*, 87(1): 199–205.

Department of Statistics, University of Washington, Seattle, WA, U.S.A.
E-mail address: `dmhleung@uw.edu`

Department of Statistics, University of Washington, Seattle, WA, U.S.A.
E-mail address: `md5@uw.edu`

Faculty of Economics, University of Niigata, Niigata, Japan
E-mail address: `hara@econ.niigata-u.ac.jp`