GLOBAL IDENTIFIABILITY OF LINEAR STRUCTURAL EQUATION MODELS

MATHIAS DRTON, RINA FOYGEL, AND SETH SULLIVANT

Abstract. Structural equation models are multivariate statistical models that are defined by specifying noisy functional relationships among random variables. We consider the classical case of linear relationships and additive Gaussian noise terms. We give a necessary and sufficient condition for global identifiability of the model in terms of a mixed graph encoding the linear structural equations and the correlation structure of the error terms. Global identifiability is understood to mean injectivity of the parametrization of the model and is fundamental in particular for applicability of standard statistical methodology.

1. Introduction

A mixed graph is a triple $G = (V, D, B)$ where V is a finite set of nodes and $D, B \subseteq V \times V$ are two sets of edges. The edges in D are directed, that is, $(i, j) \in D$ does not imply $(j, i) \in D$. We denote and draw such an edge as $i \to j$. The edges in B have no orientiation; they satisfy $(i, j) \in B$ if and only if $(j, i) \in B$. Following tradition in the field, we refer to these edges as bidirected and denote and draw them as $i \leftrightarrow j$. (In figures, we will draw bidirected edges also as dashed edges for better visual distinction.) We emphasize that in this setup the bidirected part (V, B) is always a simple graph, that is, at most one bidirected edge may join a pair of nodes. Moreover, neither the bidirected part (V, B) or the directed (V, D) contain loops, that is, $(i, i) \notin D \cup B$ for all $i \in V$. Finally, all considered mixed graphs are assumed to be acyclic, which means that the directed part (V, D) is an acyclic digraph, that is, a directed graph without directed cycles.

Enumerate the vertex set as $V = [m] := \{1, \ldots, m\}$. Since the graph is acyclic, its nodes can be ordered topologically such that $i \to j \in D$ only if $i < j$. Let $\mathbb{R}^D$ be the set of matrices $\Lambda = (\lambda_{ij}) \in \mathbb{R}^{m\times m}$ with $\lambda_{ij} = 0$ if $i \to j$ is not in D. Under a topological ordering of the nodes, all such matrices are strictly upper-triangular. Let $PD(m)$ be the cone of positive definite $m \times m$ matrices. Define $PD(B)$ to be the set of matrices $\Omega = (\omega_{ij}) \in PD(m)$ with $\omega_{ij} = 0$ if $i \neq j$ and $i \leftrightarrow j$ is not an edge in B. Finally, let I denote the identity matrix and write $\mathcal{N}_m(\mu, \Sigma)$ for the multivariate normal distribution with mean $\mu \in \mathbb{R}^m$ and covariance matrix Σ.

Definition 1. The linear structural equation model $\mathcal{M}(G)$ associated with an acyclic mixed graph $G = (V, D, B)$ is the family of multivariate normal distributions $\mathcal{N}_m(0, \Sigma)$ with

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

for $\Lambda \in \mathbb{R}^D$ and $\Omega \in PD(B)$.

Key words and phrases. Covariance matrix, Gaussian distribution, graphical model, multivariate normal distribution, parameter identification, structural equation model.

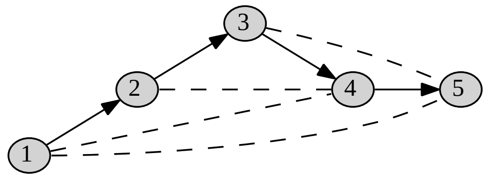

FIGURE 1. Acyclic mixed graph inducing a singular model.

If $\Lambda \in \mathbb{R}^D$, then $\det(I - \Lambda) = 1$, and thus the matrix $I - \Lambda$ is indeed always invertible. Moreover, the fact that $\det(I - \Lambda) = 1$ implies that the positive definite covariance matrix Σ is a polynomial function of the entries of Λ and Ω.

The set of parents of a node i, denoted $\mathrm{pa}(i)$, comprises the nodes j with $j \to i$ in D. The graphical model just defined is most naturally motivated in terms of a system of linear structural equations:

$$Y_j = \sum_{i \in \mathrm{pa}(j)} \lambda_{ij} Y_i + \varepsilon_j, \qquad j = 1, \ldots, m. \tag{1}$$

If $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)$ is a random vector following the multivariate normal distribution $\mathcal{N}(0, \Omega)$ then it is easily shown that the random vector $Y = (Y_1, \ldots, Y_m)$ is centered multivariate normal with covariance matrix $(I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}$.

Remark 1. Assuming centered distributions presents no loss of generality. An arbitrary mean vector could be incorporated by adding an intercept constant λ_{i0} to each equation in (1). The results discussed below would apply unchanged.

Linear structural equation models are ubiquitous in many applied fields, most notably in the social sciences where the models have a long tradition. Recent renewed interest in the models stems from their causal interpretability; compare [SGS00, Pea09]. While current research is often concerned with non-Gaussian generalizations of the models, there remain important open problems about the linear Gaussian models from Definition 1. These include the following fundamental problem, which concerns the global identifiability of the model parameters.

Question 1. *For which acyclic mixed graphs $G = (V, D, B)$ is the polynomial parametrization*

$$\phi_G : (\Lambda, \Omega) \mapsto (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}$$

an injective map from $\mathbb{R}^D \times PD(B)$ to the positive definite cone $PD(m)$?

Characterizing the graphs with injective parametrization is important because failure of injectivity can lead to failure of standard statistical methods. We briefly exemplify this issue for the models considered here and point the reader to [Drt09] and references therein for a more detailed discussion.

Example 1. Consider the graph $G = (V, D, B)$ from Figure 1. Let $\Lambda = (\lambda_{ij})$ be the matrix in $\mathbb{R}^D$ with

$$\lambda_{12} = 3, \; \lambda_{23} = -\frac{1}{2}, \; \lambda_{34} = \lambda_{45} = 1.$$

Let $\Omega = (\omega_{ij})$ be the matrix in $PD(B)$ with all diagonal entries equal to 2 and

$$\omega_{14} = \omega_{15} = \omega_{24} = \omega_{35} = 1.$$

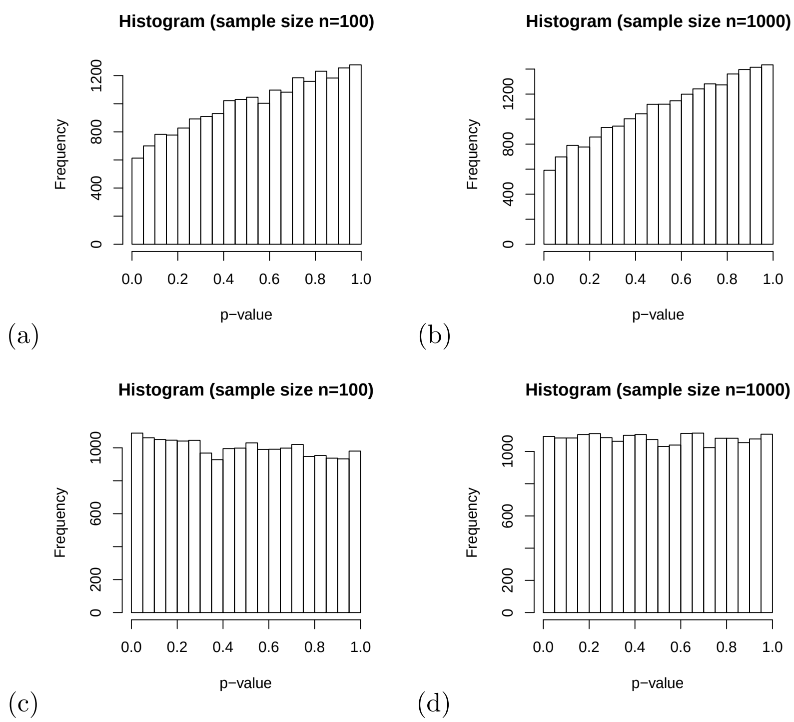

Figure 2. Histograms of p-values for a likelihood ratio test.

It can be shown that at the specified point (Λ, Ω) the map ϕ_G is not injective and the image of ϕ_G has a singularity. Suppose we use the likelihood ratio test for testing the model $\mathcal{M}(G)$ against the saturated alternative given by all multivariate normal distributions on $\mathbb{R}^m$. The standard procedure would compare the resulting likelihood ratio statistic to a chi-square distribution with two degrees of freedom. Figure 2 illustrates the problems with this procedure. What is plotted are histograms of p-values obtained from the chi-square approximation. Each histogram is based on simulation of 20,000 samples of size $n = 100$ or $n = 1000$. The samples underlying the two histograms in Figure 2(a)(b) are drawn from the multivariate normal distribution with covariance matrix $\Sigma = \phi_G(\Lambda, \Omega)$ for the above parameter choices. Many p-values being large, it is evident that the test is too conservative. For comparison, we repeat the simulations with $\lambda_{23} = 1/2$ and all other parameters unchanged. There is no identifiability failure in this second scenario, the image of ϕ_G is smooth in a neighborhood of the new covariance matrix and, as shown in Figure 2(c)(d), the expected uniform distribution for the p-values emerges in reasonable approximation. $\square$

Call a directed graph with at least two nodes an arborescence converging to node i if its edges form a spanning tree with a directed path from any node $j \neq i$ to i. In other words, i is the unique sink node. For a mixed graph $G = (V, D, B)$ and a subset of nodes $A \subset V$, let $D_A = D \cap (A \times A)$ be the set of directed edges with both endpoints in A. Similarly, let $B_A = B \cap (A \times A)$, and define the mixed subgraph

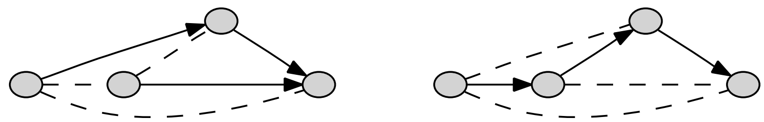

FIGURE 3. The two unlabeled graphs on four nodes with non-injective parametrization.

induced by A to be $G_A = (A, D_A, B_A)$. Our main result provides the following answer to Question 1.

Theorem 1. *The parametrization ϕ_G for an acyclic mixed graph $G = (V, D, B)$ fails to be injective if and only if there is an induced subgraph G_A, $A \subseteq V$, whose directed part (A, D_A) contains a converging arborescence and whose bidirected part (A, B_A) is connected. If ϕ_G is injective then its inverse is a rational map.*

A mixed graph $G = (V, D, B)$ is simple if there is at most one edge between any pair of nodes, that is, if $D \cap B = \emptyset$. Our theorem states in particular that only simple graphs may have an injective parametrization. Indeed, two edges $i \leftrightarrow j$ and $i \to j$ respectively connect and yield an arborescence in the subgraph $G_{\{i,j\}}$.

Corollary 1. *If G has at most three nodes then ϕ_G is injective if and only if G is simple. There are exactly two unlabeled simple acyclic mixed graphs on four nodes with ϕ_G not injective.*

Proof. An arborescence involving three nodes contains two edges. The bidirected part of a simple mixed graph can only be connected if there are two further edges. However, a simple graph with three nodes has at most three edges. The two examples on four nodes are shown in Figure 3. $\square$

As shown in the next lemma, it is easy to give a direct proof of the fact that only simple graphs can have an injective parametrization.

Lemma 1. *Suppose the map ϕ_G given by an acyclic mixed graph G is injective. Then G is a simple mixed graph, and ϕ_H is injective for any (not necessarily induced) subgraph H of G.*

Proof. If $H = (V', D', B')$ is a subgraph of $G = (V, D, B)$, that is, $V' \subseteq V$, $D' \subseteq D$ and $B' \subseteq B$, then ϕ_H is injective if and only if ϕ_G is injective at points that have all parameters λ_{ij} and ω_{ij} zero for edges $(i, j) \in D \setminus D'$ or $(i, j) \in B \setminus B'$. If G is not simple then it contains two edges $i \to j$ and $i \leftrightarrow j$. If $V = \{i, j\}$, then ϕ_G is not injective because it maps the 4-dimensional set $\mathbb{R}^D \times PD(B)$ to the 3-dimensional cone of positive definite 2×2 matrices. If $|V| > 2$, then the claim follows by passing to the subgraph induced by $\{i, j\}$. $\square$

The remainder of the paper is organized as follows. Section 2 reviews the connection of our work to the existing literature on identifiability of structural equation models. Section 3 lays out the natural stepwise approach to inversion of the parametrization ϕ_G, which uses the acyclic structure of the underlying graph. Necessity and sufficiency of the graphical condition from our main Theorem 1 are proven in Sections 4 and 5, respectively. In Section 6 we collect three lemmas used in the proof of sufficiency.

2. Prior work

Identifiability properties of structural equation models are a topic with a long history. A review of classical conditions, which do not take into account the finer graphical structure considered here, can be found for instance in the monograph [Bol89]. A more recent sufficient condition for global identifiability of the linear structural equation models from Definition 1 is due to [McD02, RS02]. It requires the presence of a bidirected edge $i \leftrightarrow j$ to imply the absence of directed paths from j to i (and from i to j). Following [RS02] we call an acyclic mixed graph with this property *ancestral*. It is clear that an ancestral mixed graph is simple. We revisit the result about ancestral graphs in Corollary 2 below.

Other recent work, such as [BP02], considers a weaker identifiability requirement for the model $\mathcal{M}(G)$ associated with an acyclic mixed graph $G = (V, D, B)$. For a pair of matrices $\Lambda_0 \in \mathbb{R}^D$ and $\Omega_0 \in PD(B)$, define the fiber

$$\mathcal{F}(\Lambda_0, \Omega_0) = \left\{(\Lambda, \Omega) \,:\, \phi_G(\Lambda, \Omega) = \phi_G(\Lambda_0, \Omega_0),\, \Lambda \in \mathbb{R}^D,\, \Omega \in PD(B)\right\}. \tag{2}$$

The map ϕ_G is injective if and only if all its fibers contain only a single point. If it holds instead that for generic choices of $\Lambda \in \mathbb{R}^D$ and $\Omega \in PD(B)$, the fiber $\mathcal{F}(\Lambda, \Omega)$ contains only the single point (Λ, Ω), then we say that the map ϕ_G is *generically injective* and the model $\mathcal{M}(G)$ is *generically identifiable*. Requiring a condition to hold for generic points means that the points at which the condition fails form a lower-dimensional algebraic subset. In particular, the condition holds for almost every point (in Lebesgue measure), and some authors thus also speak of an almost everywhere identifiable model. When the substantive interest is in all parameters of a model, generic identifiability constitutes a minimal requirement. However, generically but not globally identifiable models can present difficulties for statistical inference; recall Example 1 that treats a generically identifiable model.

The main theorem of [BP02], which we reprove in Corollary 3, states that ϕ_G is generically injective for every simple acyclic mixed graph G. The graph being simple, however, is far from necessary for generic injectivity of ϕ_G. A classical counterexample is the instrumental variable model based on the graph with edges $1 \to 2 \to 3$ and $2 \leftrightarrow 3$. For recent work on the topic see [Tia09] and references therein. To our knowledge, characterizing the mixed graphs G with generically injective parametrization ϕ_G remains an open problem.

The linear structural equation models $\mathcal{M}(G)$ considered in this paper are closely related to latent variable models known as semi-Markovian causal models. These non-parametric models are obtained by subdividing the bidirected edges, that is, each edge $i \leftrightarrow j$ is replaced by two directed edges $i \leftarrow u_{ij} \to j$, where u_{ij} is a new node. Each node u_{ij} added to the vertex set corresponds to a latent variable; compare also [RS02, Pea09]. The global identifiability problem for acyclic semi-Markovian causal models is solved in [SP06], using results from [Tia02]. This work is based on manipulating recursive density factorizations involving latent variables.

When restricting to normal distributions, the semi-Markovian latent variable model associated with an acyclic mixed graph $G = (V, D, B)$ may coincide with the model $\mathcal{M}(G)$ from Definition 1. For instance, if there are no directed edges ($D = \emptyset$), then the models are the same if and only if the bidirected part (V, B) is a forest of trees [DY10, Corollary 3.4]. When the models agree the global identifiability of the non-parametric semi-Markovian model implies global identifiability of the Gaussian linear structural equation model $\mathcal{M}(G)$. When the models disagree, however,

$\mathcal{M}(G)$ is larger than the Gaussian restriction of the semi-Markovian model. Therefore, further arguments are required to establish sufficient conditions for global identifiability of $\mathcal{M}(G)$. Moreover, the existing counterexamples to identifiability of semi-Markovian models involve binary variables and thus cannot be used to prove necessity of an identifiability condition for the Gaussian models $\mathcal{M}(G)$. Interestingly, however, our graphical condition from Theorem 1, which we first found by experimentation with computer algebra software, coincides with that of [SP06]. A reader familiar with the work in [Tia02] will also recognize similarities between the higher-level structure of the proofs given there and those in Section 5 of this paper. We conclude that in graphical terms, the Gaussian models from Definition 1 are just as difficult to identify as the non-parametric semi-Markovian causal models.

3. Stepwise inversion

Suppose $G = (V, D, B)$ is an acyclic mixed graph with vertex set $V = [m]$. The map ϕ_G is injective if all its fibers contain only a single point; recall the definition of a fiber in (2). Let $\Sigma = \phi_G(\Lambda_0, \Omega_0)$ for two matrices $\Lambda_0 \in \mathbb{R}^D$ and $\Omega_0 \in PD(B)$. This section describes how to find points (Λ, Ω) in the fiber $\mathcal{F}(\Lambda_0, \Omega_0)$. In particular, we show that the following algebraic criterion can be used to decide whether the map ϕ_G is injective. The lemma is proven after we describe a natural inversion approach that uses the acyclic structure of the graph G in a stepwise manner.

For each $i \leq m-1$, let $P(i) = \mathrm{pa}(i+1)$ be the parents of node $i+1$, and $S(i) = \{j \leq i : j \leftrightarrow i+1 \in B\}$ the siblings of $i+1$. (In other related work, the nodes incident to a bidirected edge $i \leftrightarrow j$ have also been called "spouses" of each other but we find "siblings" to be natural terminology given that a common parent to the two nodes is introduced when subdividing the edge as discussed in Section 2.)

Lemma 2. *Suppose $G = (V, D, B)$ is an acyclic mixed graph with its nodes labeled in a topological order. Then the parametrization ϕ_G is injective if and only if the rank condition*

$$\operatorname{rank}\left(\Omega_{[i]\setminus S(i),[i]} (I-\Lambda)^{-1}_{[i],P(i)} \right) = |P(i)|$$

holds for all nodes $i = 1, \ldots, m-1$ and all pairs $\Lambda \in \mathbb{R}^D$ and $\Omega \in PD(B)$.

Remark 2. In this paper, matrix inversion is always given higher priority than an operation of forming a submatrix. For any invertible matrix M and index sets A, B, the matrix $M^{-1}_{A,B} = (M^{-1})_{A,B}$ is thus the $A \times B$ submatrix of the inverse of M.

Computing points (Λ, Ω) in the fiber $\mathcal{F}(\Lambda_0, \Omega_0)$ means solving the polynomial equation system given by the matrix equation

$$\Sigma = (I-\Lambda)^{-T} \Omega (I-\Lambda)^{-1}. \tag{3}$$

For topologically ordered nodes, (3) implies that $\sigma_{11} = \omega_{11}$ and that the first column in the strictly upper-triangular matrix Λ contains only zeros. Hence, these are uniquely determined for all matrices in the fiber.

Let $i \geq 1$, and assume that we know the $[i] \times [i]$ submatrices of Λ and Ω of a solution to the equation in (3). Partition off the $(i+1)$-st row and column of the submatrices

$$(I-\Lambda)_{[i+1],[i+1]} = \begin{pmatrix} \Gamma & -\lambda \\ 0 & 1 \end{pmatrix}, \qquad \Omega_{[i+1],[i+1]} = \begin{pmatrix} \Psi & \omega \\ \omega^T & \omega_{i+1,i+1} \end{pmatrix}.$$

The matrices Γ and Ψ are known, $\lambda_{[i]\setminus P(i)} = 0$ and $\omega_{[i]\setminus S(i)} = 0$. The inverse of $I-\Lambda$ can be written as a block matrix as

$$(I-\Lambda)^{-1}_{[i+1],[i+1]} = \begin{pmatrix} \Gamma^{-1} & \Gamma^{-1}\lambda \\ 0 & 1 \end{pmatrix}. \tag{4}$$

In this notation, the part of equation (3) that pertains to the $[i+1]\times[i+1]$ submatrix of Σ is

$$\begin{pmatrix} \Sigma_{[i],[i]} & \Sigma_{[i],\{i+1\}} \\ & \sigma_{i+1,i+1} \end{pmatrix} = \begin{pmatrix} \Gamma^{-T}\Psi\Gamma^{-1} & \Gamma^{-T}\Psi\Gamma^{-1}\lambda + \Gamma^{-T}\omega \\ & \omega_{i+1,i+1} + \lambda^T\Gamma^{-T}\Psi\Gamma^{-1}\lambda + 2\omega^T\Gamma^{-1}\lambda \end{pmatrix},$$

where only the upper-triangular parts of the symmetric matrices are shown. Hence, given the values of Γ and Ψ, the choice of λ and ω is unique if and only if the equation

$$\Sigma_{[i],\{i+1\}} = \Gamma^{-T}\Psi\Gamma^{-1}\cdot\lambda + \Gamma^{-T}\cdot\omega \tag{5}$$

has a unique solution. Clearly, any feasible choice of a solution (λ,ω) to the equation in (5) leads to a unique solution $\omega_{i+1,i+1}$ via the equation

$$\sigma_{i+1,i+1} = \omega_{i+1,i+1} + \lambda^T\Gamma^{-T}\Psi\Gamma^{-1}\lambda + 2\omega^T\Gamma^{-1}\lambda. \tag{6}$$

Since $\lambda_{[i]\setminus P(i)} = 0$ and $\omega_{[i]\setminus S(i)} = 0$, equation (5) can be rewritten as

$$\Sigma_{[i],\{i+1\}} = (\Gamma^{-T}\Psi\Gamma^{-1}_{[i],P(i)})\cdot\lambda_{P(i)} + (\Gamma^{-1}_{S(i),[i]})^T\cdot\omega_{S(i)}.$$

It has a unique solution if and only if the matrix

$$\begin{bmatrix} \Gamma^{-T}\Psi\Gamma^{-1}_{[i],P(i)} & (\Gamma^{-1}_{S(i),[i]})^T \end{bmatrix}$$

has full column rank $|P(i)|+|S(i)|$. The matrix Γ is invertible because it is upper-triangular with ones along the diagonal. Thus the condition is equivalent to

$$\Gamma^T\begin{bmatrix} \Gamma^{-T}\Psi\Gamma^{-1}_{[i],P(i)} & (\Gamma^{-1}_{S(i),[i]})^T \end{bmatrix} = \begin{bmatrix} \Psi\Gamma^{-1}_{[i],P(i)} & I_{[i],S(i)} \end{bmatrix}$$

having full column rank. The second block is part of an identity matrix. We deduce that the condition is equivalent to requiring that $\Psi_{[i]\setminus S(i),[i]}\Gamma^{-1}_{[i],P(i)}$, the submatrix obtained by removing the rows and columns with index in $S(i)$, has rank $|P(i)|$. Note that

$$\Psi_{[i]\setminus S(i),[i]}\Gamma^{-1}_{[i],P(i)} = \Omega_{[i]\setminus S(i),[i]}(I-\Lambda)^{-1}_{[i],P(i)}$$

is the matrix appearing in Lemma 2.

Proof of Lemma 2. Consider a feasible pair (Λ,Ω). If the rank condition for this pair holds for all nodes $i=1,\dots,m-1$, then it follows from the stepwise inversion procedure described above that the fiber $\mathcal{F}(\Lambda,\Omega)$ contains only the single point (Λ,Ω). Therefore, the rank condition holding for all nodes and all matrix pairs implies that all fibers are singletons, or in other words, that the map ϕ_G is injective.

Conversely, assume that the rank condition fails for some node $i\le m-1$ and matrix pair (Λ,Ω). If $i=m-1$, then the considered fiber $\mathcal{F}(\Lambda,\Omega)$ is positive-dimensional, and ϕ_G not injective. If $i<m-1$, then it follows analogously that the parametrization ϕ_H for the induced subgraph $H=G_{[i+1]}$ is not injective. By Lemma 1, ϕ_G cannot be injective either. □

If the rank condition in Lemma 2 holds at a particular pair (Λ,Ω), then the fiber $\mathcal{F}(\Lambda,\Omega)$ contains only the pair (Λ,Ω). However, the converse is false in general, that is, failure of the rank condition at a particular pair (Λ,Ω) need not imply that

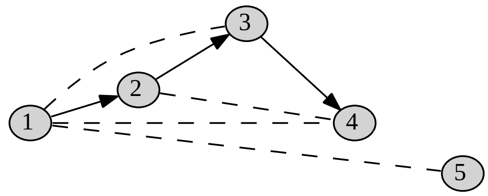

Figure 4. Graph with non-injective parametrization (see Example 2).

the fiber $\mathcal{F}(\Lambda, \Omega)$ contains more than one point. This may occur even for a simple acyclic mixed graph.

Example 2. Consider the graph in Figure 4, set $\lambda_{12} = \lambda_{23} = \lambda_{34} = 1$, and choose the positive definite matrix

$$\Omega = \begin{pmatrix} 2 & 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 3 & 0 \\ -1 & 0 & 0 & 0 & 3 \end{pmatrix}.$$

The rank condition for this pair (Λ, Ω) fails at node $i = 3$. Nevertheless, the fiber $\mathcal{F}(\Lambda, \Omega)$ is equal to $\{(\Lambda, \Omega)\}$. If we set $\omega_{15} = 0$, however, then $\mathcal{F}(\Lambda, \Omega)$ becomes one-dimensional. Using terminology from econometrics/causality, the variable corresponding to node 5 behaves like an "instrument;" compare for instance [Pea09]. $\square$

In order to prepare for arguments turning the algebraic condition from Lemma 2 into a graphical one, we detail the structure of the inverse $(I - \Lambda)^{-1}$ for a matrix $\Lambda \in \mathbb{R}^D$. Recall that the matrix $I - \Lambda$ is upper-triangular with ones along the diagonal and has entries equal to $-\lambda_{ij}$ for the directed edges $i \to j \in D$. Let $\mathcal{P}(i, j)$ denote the set of directed paths from i to j in the considered graph.

Lemma 3. *The entries of the inverse* $(I - \Lambda)^{-1}$ *are*

$$(I - \Lambda)^{-1}_{ij} = \sum_{\pi \in \mathcal{P}(i,j)} \prod_{k \to l \in \pi} \lambda_{kl}, \qquad i, j \in [m].$$

Proof. This well-known fact can be shown by induction on the matrix size m and using the partitioning in (4) under a topological ordering of the nodes. $\square$

Note that adopting the usual definition that takes an empty sum to be zero and an empty product to be one, the formula in Lemma 3 states that $(I - \Lambda)^{-1}_{ij} = 0$ if $i \neq j$ and $\mathcal{P}(i, j) = \emptyset$, and it states that $(I - \Lambda)^{-1}_{ii} = 1$ because $\mathcal{P}(i, i)$ contains only a trivial path without edges.

Corollary 2. *If the acyclic mixed graph* G *is ancestral then the parametrization* ϕ_G *is injective.*

Proof. Recall that if $G = (V, D, B)$ is ancestral and $i \leftrightarrow j$ is a bidirected edge in G, then there is no directed path from i to j or j to i. Suppose $V = [m]$ is topologically ordered, and let i be some node smaller than m. Pick a node $j \in S(i)$. Then there may not exist a directed path from j to a node in $P(i)$. It follows that

$$\Omega_{[i]\setminus S(i),[i]}(I-\Lambda)^{-1}_{[i],P(i)} = \Omega_{[i]\setminus S(i),[i]\setminus S(i)}(I-\Lambda)^{-1}_{[i]\setminus S(i),P(i)}.$$

The latter matrix is the product of a principal and thus positive definite submatrix of Ω and a matrix that contains the $P(i)\times P(i)$ identity matrix. It follows that this product has full column rank $|P(i)|$ for all feasible pairs (Λ,Ω) and all nodes $i\le m-1$. By Lemma 2, ϕ_G is injective. □

If the acyclic mixed graph G is simple, then $P(i)\subseteq [i]\setminus S(i)$ for all nodes $i\le m-1$. Hence, the matrix product appearing in the rank condition always has at least as many rows as columns. The next generic identifiability result follows immediately; recall the definitions in Section 2.

Corollary 3. *If $G=(V,D,B)$ is a simple acyclic mixed graph, then the map ϕ_G is generically injective.*

Proof. We need to show that for generic choices of $\Lambda\in\mathbb{R}^D$ and $\Omega\in PD(B)$, the fiber $\mathcal{F}(\Lambda,\Omega)$ is equal to the singleton $\{(\Lambda,\Omega)\}$. Set $\Lambda=0$ and choose Ω to be the identity matrix. Then each of the matrix products

$$\Omega_{[i]\setminus S(i),[i]}(I-\Lambda)^{-1}_{[i],P(i)}, \qquad i=1,\dots,m-1, \tag{7}$$

has the identity matrix as $P(i)\times P(i)$ submatrix. The rank condition from Lemma 2 thus holds for all $i\le m-1$. Since the matrices in (7) have polynomial entries, existence of a single pair (Λ,Ω) at which the $m-1$ matrices in (7) have full column rank implies that the set of pairs (Λ,Ω) for which at least one of the matrices fails to have full column rank is a lower-dimensional algebraic set; compare [CLO07, Chapter 9] for background on such algebraic arguments. □

4. Necessity of the graphical condition for identifiability

We now prove that the graphical condition in Theorem 1, which states that there be no induced subgraph whose directed part contains a converging arborescence and whose bidirected part is connected, is necessary for the parametrization ϕ_G to be injective. By Lemma 1, it suffices to consider a graph whose directed part is a converging arborescence and whose bidirected part is a spanning tree. In light of Lemma 2, the necessity of the graphical condition in Theorem 1 then follows from the following result.

Proposition 1. *Let $G=(V,D,B)$ be an acyclic mixed graph with topologically ordered vertex set $V=[m+1]$. If (V,D) is an arborescence converging to $m+1$ and (V,B) is a spanning tree, then there exists a pair of matrices $\Lambda\in\mathbb{R}^D$ and $\Omega\in PD(B)$ with*

$$\mathrm{kernel}\left(\Omega_{[m]\setminus S(m),[m]}(I-\Lambda)^{-1}_{[m],P(m)}\right)\neq\{0\}.$$

Let $\mathcal{L}(\Lambda)\subseteq\mathbb{R}^m$ be the column span of $(I-\Lambda)^{-1}_{[m],P(m)}$. We formulate a first lemma that we will use to prove Proposition 1.

Lemma 4. *If $V=[m+1]$ and (V,D) is an arborescence converging to node $m+1$, then the union of the linear spaces $\mathcal{L}(\Lambda)$ for all $\Lambda\in\mathbb{R}^D$ contains the torus $(\mathbb{R}^*)^m=(\mathbb{R}\setminus\{0\})^m$ of vectors with all coordinates non-zero.*

Proof. In the arborescence, there is a unique path $\pi(i)$ from any vertex $i \in [m] \setminus P(m)$ to the sink node $m+1$. Let $k(i)$ be the unique node in $P(m)$ that lies on this path. Let $\Lambda \in \mathbb{R}^D$ and $\alpha \in \mathbb{R}^{|P(m)|}$, and define the vector

$$\beta(\Lambda, \alpha) = (I - \Lambda)^{-1}_{[m],P(m)} \alpha \in \mathbb{R}^m.$$

Since the principal submatrix $(I-\Lambda)^{-1}_{P(m),P(m)}$ is an identity matrix, $\beta(\Lambda,\alpha)_i = \alpha_i$ for all $i \in P(m)$. For $i \in [m] \setminus P(m)$, we use Lemma 3 to obtain

$$\beta(\Lambda, \alpha)_i = \alpha_{k(i)} \prod_{j \to l \in \pi(i)} \lambda_{jl} = \lambda_{ij} \beta(\Lambda, \alpha)_j, \tag{8}$$

where $i \to j \in G$ is the unique edge originating from i.

Let x be any vector in $(\mathbb{R}^*)^m$. Our claim states that there exist a matrix $\Lambda \in \mathbb{R}^D$ and vector α such that $x = \beta(\Lambda, \alpha)$. Clearly, α has to be equal to the subvector $x_{P(m)}$. The associated unique choice of Λ is obtained by recursively solving for the entries λ_{ij} using the relationship in (8). □

Let $R(m) = [m] \setminus S(m)$ be the "rest" of the nodes. We are left with the problem of finding a matrix $\Omega \in PD(B)$ for which some vector in $(\mathbb{R}^*)^m$ lies in the kernel of the submatrix

$$\Omega_{R(m),[m]} = \begin{bmatrix} \Omega_{R(m),R(m)} & \Omega_{R(m),S(m)} \end{bmatrix}.$$

We note that if $\Omega_{R(m),R(m)}$ and $\Omega_{R(m),S(m)}$ have the required zeros and $\Omega_{R(m),R(m)}$ is positive definite, then there is a completion to a positive definite matrix $\Omega \in PD(B)$. Proposition 1 now follows by combining Lemma 4 with the next result.

Lemma 5. *If (V,B) is a tree on $V = [m+1]$, then there exists a matrix $\Omega \in PD(B)$ such that the vector $\mathbf{1} = (1, \ldots, 1)^T$ is in the kernel of the submatrix $\Omega_{R(m),[m]}$.*

Proof. Let T be the set of all nodes in $R(m)$ that are connected to some node in $S(m)$ by an edge in B. If $\Omega \in PD(B)$, then the submatrix $\Omega_{R(m),S(m)}$ has only zero entries in rows indexed by nodes $i \in R(m) \setminus T$. If $i \in T$, then the i-th row of $\Omega_{R(m),S(m)}$ has at least one entry that is not constrained to zero and may take any real value. Hence, we can choose a matrix $\Omega_{R(m),S(m)}$ that has row sum

$$\sum_{j \in S(m)} \omega_{ij} = \begin{cases} -1 & \text{if } i \in T, \\ 0 & \text{if } i \in R(m) \setminus T. \end{cases} \tag{9}$$

Let $H = (R(m), B_{R(m)})$ be the induced subgraph of G on vertex set $R(m)$. The Laplacian of H, $L(H) = (l_{ij})$, is the symmetric $R(m) \times R(m)$ matrix whose diagonal entries are the degrees of the nodes in H and whose off-diagonal entries l_{ij} are equal to -1 if $i \leftrightarrow j$ is an edge in H and 0 otherwise. The Laplacian is well-known to be positive semidefinite with all row sums zero. For a subset $C \subset [m]$, let $\mathbf{1}_C \in \mathbb{R}^m$ be the vector with entries equal to one at indices in C and zero elsewhere. The kernel of $L(H)$ is the direct sum of the linear spaces spanned by the vectors $\mathbf{1}_C$ for the connected components C of the graph H; compare [Chu97, Chapter 1].

Let $D_T = (d_{ij})$ be the diagonal matrix that has diagonal entry $d_{ii} = 1$ if $i \in T$ and $d_{ii} = 0$ otherwise. Both $L(H)$ and D_T are positive semidefinite matrices and thus the kernel of $L(H) + D_T$ is equal to $\ker L(H) \cap \ker D_T$. Since (V,B) is a connected graph, each connected component of H contains a node in T. Therefore, none of the vectors $\mathbf{1}_C$ are in the kernel of D_T, where C ranges over all connected components of H. This implies that the $\ker(L(H) + D_T) = \{0\}$, and hence this matrix is positive definite.

Let Ω be any matrix in $PD(B)$ whose submatrix $\Omega_{R(m),S(m)}$ satisfies (9) and whose principal submatrix $\Omega_{R(m),R(m)}$ is the positive definite matrix $L(H) + D_T$. The matrix $\Omega \in PD(B)$ has the desired property because

$$\Omega_{R(m),[m]}\mathbf{1} = (L(H) + D_T)\mathbf{1} + \Omega_{R(m),S(m)}\mathbf{1} = \mathbf{1}_T - \mathbf{1}_T = 0.$$

Such matrices exist because we can choose $\Omega_{S(m),S(m)}$ to be, for instance, a diagonal matrix with very large diagonal entries. Principal minors of Ω that are not submatrices of $\Omega_{R(m),R(m)}$ will be dominated by these diagonal entries and hence be positive. All other principal minors are positive since $\Omega_{R(m),R(m)} = L(H) + D_T$ was shown to be positive definite. □

5. Sufficiency of the graphical condition for identifiability

In this section we prove that the graphical condition in Theorem 1, which states that there be no induced subgraph whose directed part contains a converging arborescence and whose bidirected part is connected, is sufficient for the parametrization ϕ_G to be injective. Proposition 4 below shows that if ϕ_G is not injective and G does not contain an induced subgraph with both a converging arborescence and a bidirected spanning tree, then there is a subgraph G' with fewer nodes such that $\phi_{G'}$ still fails to be injective. The sufficiency of the graphical condition then follows immediately. To see this, note that a graph G with non-injective parametrization ϕ_G must contain some minimal induced subgraph G' with non-injective $\phi_{G'}$. Applying the contrapositive of Proposition 4 to G', we conclude that the directed part of G' contains a converging arborescence and the bidirected part of G' is connected.

In preparing for the proof of Proposition 4, we first treat the case when there is no arborescence; this gives Proposition 2. The case when there is no bidirected spanning tree is treated in Proposition 3. In either case, we reduce a given graph $G = (V, D, B)$ to the subgraph G_W induced by a subset $W \subsetneq V$. We use the notation $\tilde{\Lambda}$, $\tilde{\Omega}$, $\tilde{P}(i)$, $\tilde{S}(i)$, $\tilde{\mathcal{P}}(i,j)$ to denote the counterparts to Λ, Ω, $P(i)$, $S(i)$, and $\mathcal{P}(i,j)$, when performing this reduction of G to G_W.

Proposition 2. *Let $G = (V, D, B)$ be an acyclic mixed graph with topologically ordered vertex set $V = [m+1]$, with some $\Lambda \in \mathbb{R}^D$, $\Omega \in PD(B)$, and nonzero $\alpha \in \mathbb{R}^{|P(m)|}$, such that*

$$\Omega_{[m]\setminus S(m),[m]}(I - \Lambda)^{-1}_{[m],P(m)}\alpha = 0.$$

Suppose the directed part of G does not contain an arborescence converging to $m+1$. Let A be the set of nodes $i \le m$ with some path of directed edges from i to $m+1$, and $W = A \cup \{m+1\}$. Then $W \subsetneq V$ and ϕ_{G_W} is not injective.

Proof. Since G does not have a converging arborescence, $A \subsetneq [m]$ and $W \subsetneq V$.

Denote the induced subgraph as $G_W = (W, \tilde{D}, \tilde{B})$. Let $\tilde{\Lambda} = \Lambda_{W,W} \in \mathbb{R}^{\tilde{D}}$ and $\tilde{\Omega} = \Omega_{W,W} \in PD(\tilde{B})$. Note that $P(m) \subseteq A$ by definition, and so $\tilde{P}(m) = P(m)$. Suppose $j \in P(m)$. Then for each $i \in [m]\setminus A$, $\mathcal{P}(i,j) = \emptyset$ by definition, and so $(I-\Lambda)^{-1}_{ij} = 0$ by Lemma 3. For each $i \in A$, and for any path $i \to v_1 \to \cdots \to v_k \to j$ in G, each intermediate vertex $v_1, \ldots, v_k$ is in A by definition of A (since there is an edge $j \to m+1$). Therefore, $\tilde{\mathcal{P}}(i,j) = \mathcal{P}(i,j)$, and it follows that $(I-\tilde{\Lambda})^{-1}_{ij} = (I-\Lambda)^{-1}_{ij}$. In other words, when the nodes outside of W are removed from G, the remaining entries of $(I-\Lambda)^{-1}$ are unchanged, while the removed entries in the columns indexed by $P(m) = \tilde{P}(m)$ are all zero. We obtain that

$$\sum_{i\in A} \tilde{\Omega}_{A\setminus \tilde{S}(m),i}(I-\tilde{\Lambda})^{-1}_{i,\tilde{P}(m)}\alpha = \sum_{i\in A} \Omega_{A\setminus S(m),i}(I-\Lambda)^{-1}_{i,P(m)}\alpha = \sum_{i\in [m]} \Omega_{A\setminus S(m),i}(I-\Lambda)^{-1}_{i,P(m)}\alpha = \Omega_{A\setminus S(m),[m]}(I-\Lambda)^{-1}_{[m],P(m)}\alpha.$$

By assumption, the last quantity is zero. By Lemma 2, ϕ_{G_W} is not injective. □

We next prove a similar proposition for graphs whose bidirected part is not connected. The proof uses Lemmas 6 and 8, which are derived in Section 6.

Proposition 3. *Let $G = (V, D, B)$ be an acyclic mixed graph with topologically ordered vertex set $V = [m+1]$, with some $\Lambda \in \mathbb{R}^D$, $\Omega \in PD(B)$, and nonzero $\alpha \in \mathbb{R}^{|P(m)|}$, such that*

$$\Omega_{[m]\setminus S(m),[m]}(I-\Lambda)^{-1}_{[m],P(m)}\alpha = 0.$$

Suppose the bidirected part of G is not connected. Let A be the set of nodes $i \le m$ with some path of bidirected edges from i to $m+1$, and $W = A \cup \{m+1\}$. Then $W \subsetneq V$ and ϕ_{G_W} is not injective.

Proof. Since the bidirected part is not connected, $A \subsetneq [m]$ and $W \subsetneq V$.

Denote the induced subgraph as $G_W = (W, \tilde{D}, \tilde{B})$. Let $\tilde{\Lambda} = \Lambda_{W,W} \in \mathbb{R}^{\tilde{D}}$ and $\tilde{\Omega} = \Omega_{W,W} \in PD(\tilde{B})$. If $i \in S(m)$, then it holds trivially that $i \in A$ and thus $\tilde{S}(m) = S(m)$. By Lemma 8 below,

$$\begin{aligned}\tilde{\Omega}_{A\setminus \tilde{S}(m),A}(I-\tilde{\Lambda})^{-1}_{A,\tilde{P}(m)}\alpha_{\tilde{P}(m)} &= \tilde{\Omega}_{A\setminus S(m),A}(I-\Lambda)^{-1}_{A,P(m)}\alpha \\ &= \tilde{\Omega}_{A\setminus S(m),[m]}(I-\Lambda)^{-1}_{[m],P(m)}\alpha - \tilde{\Omega}_{A\setminus S(m),[m]\setminus A}(I-\Lambda)^{-1}_{[m]\setminus A,P(m)}\alpha.\end{aligned}$$

By hypothesis, the first term is zero. By Lemma 6 below, $(I-\Lambda)^{-1}_{[m]\setminus A,P(m)}\alpha = 0$, and so the second term is zero. Therefore,

$$\tilde{\Omega}_{A\setminus S(m),A}(I-\tilde{\Lambda})^{-1}_{A,\tilde{P}(m)}\alpha_{\tilde{P}(m)} = 0.$$

It remains to be shown that $\alpha_{\tilde{P}(m)} \neq 0$. Suppose instead that $\alpha_{\tilde{P}(m)} = 0$. Then, using Lemma 6, we obtain that

$$\begin{aligned}0 &= (I-\Lambda)^{-1}_{[m]\setminus A,P(m)}\alpha \\ &= (I-\Lambda)^{-1}_{[m]\setminus A,\tilde{P}(m)}\alpha_{\tilde{P}(m)} + (I-\Lambda)^{-1}_{[m]\setminus A,P(m)\setminus \tilde{P}(m)}\alpha_{P(m)\setminus \tilde{P}(m)} \\ &= 0 + (I-\Lambda)^{-1}_{[m]\setminus A,P(m)\setminus \tilde{P}(m)}\alpha_{P(m)\setminus \tilde{P}(m)}.\end{aligned}$$

However, $(I-\Lambda)^{-1}_{[m]\setminus A,P(m)\setminus \tilde{P}(m)}$ is a submatrix of $(I-\Lambda)^{-1}_{[m]\setminus A,[m]\setminus A}$, which is a full rank matrix as it is upper triangular with ones on the diagonal. Therefore, $(I-\Lambda)^{-1}_{[m]\setminus A,P(m)\setminus \tilde{P}(m)}$ is full rank, and so $\alpha_{P(m)\setminus \tilde{P}(m)} = 0$. It follows that $\alpha = 0$, which is a contradiction. We conclude that $\alpha_{\tilde{P}(m)} \neq 0$ and, by Lemma 2, that ϕ_{G_W} is not injective. □

Proposition 4. *Let $G = (V, D, B)$ be an acyclic mixed graph with topologically ordered vertex set $V = [m+1]$, such that the parametrization ϕ_G is not injective. If either the directed part of G does not contain an arborescence converging to $m+1$,* *or the bidirected part of G is not connected, then there is some proper induced subgraph G_W of G for which the parametrization ϕ_{G_W} is not injective.*

Proof. From Lemma 2, for some $i \leq m$,

$$\operatorname{rank}\left(\Omega_{[i]\setminus S(i),[i]}(I-\Lambda)^{-1}_{[i],P(i)}\right) < |P(i)|. \tag{10}$$

Suppose $i < m$. Take $W = [i+1]$, and denote the induced subgraph as $G_W = (W, \tilde{D}, \tilde{B})$. It holds trivially that $\tilde{\Lambda} := \Lambda_{[i+1],[i+1]} \in \mathbb{R}^{\tilde{D}}$ and $\tilde{\Omega} := \Omega_{[i+1],[i+1]} \in PD(\tilde{B})$, and furthermore $(I-\tilde{\Lambda})^{-1} = (I-\Lambda)^{-1}_{[i+1],[i+1]}$. It is then clear that, by Lemma 2, ϕ_{G_W} is not injective.

Next suppose instead that (10) is true for $i = m$. If the directed part of G does not contain an arborescence converging to $m+1$, then apply Proposition 2 to produce a proper induced subgraph G_W with ϕ_{G_W} non-injective. If instead the bidirected part of G is not connected, then apply Proposition 3 to produce a proper induced subgraph G_W with ϕ_{G_W} non-injective.

In all cases, we have constructed a subset $W \subsetneq V$ with ϕ_{G_W} not injective. □

6. Proofs of lemmas in Section 5

Lemma 6. *Let G, Λ, Ω, α, and A be as in the statement of Proposition 3. Then $(I-\Lambda)^{-1}_{[m]\setminus A,P(m)}\alpha = 0$.*

Proof. If $i \in [m]\setminus A$ and $j \in A$, then, by definition of A, it holds that $\Omega_{i,j} = 0$. Therefore, $\Omega_{[m]\setminus A,A} = 0$ and we obtain that

$$\Omega_{[m]\setminus A,[m]\setminus A}(I-\Lambda)^{-1}_{[m]\setminus A,P(m)}\alpha = \Omega_{[m]\setminus A,[m]}(I-\Lambda)^{-1}_{[m],P(m)}\alpha = 0.$$

For the last equality, observe that $[m]\setminus A \subset [m]\setminus S(i)$ since $S(i) \subset A$. Since $\Omega_{[m]\setminus A,[m]\setminus A}$ is positive definite, the claim follows. □

For a directed path π in the graph G, we write $\pi \not\subset G_A$ to indicate that not all the nodes of π lie in A. Also, by convention, $\mathcal{P}(j,j)$ is a singleton set containing the trivial path at j; in this case π has no edges and we define $\prod_{a\to b\in\pi}\lambda_{ab} = 1$.

Lemma 7. *Let G, Λ, Ω, α, and A be as in the statement of Proposition 3. Then for every $i \leq m$,*

$$\sum_{k\in P(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k),\pi\not\subset G_A} \prod_{a\to b\in\pi} \lambda_{ab} \right) = 0.$$

Proof. First, we prove the claim for $i \notin A$. Working from Lemma 6, we have that

$$0 = (I-\Lambda)^{-1}_{i,P(m)}\alpha = \sum_{k\in P(m)} (I-\Lambda)^{-1}_{ik}\alpha_k = \sum_{k\in P(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k)} \prod_{a\to b\in\pi} \lambda_{ab} \right). \tag{11}$$

Since $i \notin A$, any path $\pi \in \mathcal{P}(i,k)$ for any k necessarily satisfies $\pi \not\subset G_A$. Hence, we can rewrite (11) as

$$\sum_{k\in P(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k),\pi\not\subset G_A} \prod_{a\to b\in\pi} \lambda_{ab} \right) = 0.$$

Next we address the case $i \in A$. Inducting on i in decreasing order, we may assume that the claim holds for all $j \in \{i+1, i+2, \ldots, m\}$. (As a base case, we can set $i = m$ because, by the assumed topological order, $\mathcal{P}(m,k) = \emptyset$ for all nodes $k < m$.) The quantity claimed to be vanishing is

$$\sum_{k\in P(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k),\pi\not\subset G_A} \prod_{a\to b\in\pi} \lambda_{ab} \right) = \sum_{k\in P(m)} \alpha_k \left[\sum_{j:i\to j} \left(\sum_{\pi'\in\mathcal{P}(j,k),\pi'\not\subset G_A} \lambda_{ij} \prod_{a\to b\in\pi'} \lambda_{ab} \right) \right]. \tag{12}$$

This last equality is obtained by splitting any path $\pi = i \to v_1 \to \cdots \to v_n \to k$ into $i \to j := v_1$ and $\pi' = j \to v_2 \to \cdots \to v_n \to k$. Since we assume $i \in A$, it holds that $\pi \not\subset G_A$ if and only if $\pi' \not\subset G_A$. Interchanging the order of the summations in (12), we obtain that

$$\begin{aligned} \sum_{k\in P(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k),\pi\not\subset G_A} \prod_{a\to b\in\pi} \lambda_{ab} \right) &= \sum_{j:i\to j} \left[\sum_{k\in P(m)} \alpha_k \left(\sum_{\pi'\in\mathcal{P}(j,k),\pi'\not\subset G_A} \lambda_{ij} \prod_{a\to b\in\pi'} \lambda_{ab} \right) \right] \\ &= \sum_{j:i\to j} \lambda_{ij} \left[\sum_{k\in P(m)} \alpha_k \left(\sum_{\pi'\in\mathcal{P}(j,k),\pi'\not\subset G_A} \prod_{a\to b\in\pi'} \lambda_{ab} \right) \right]. \end{aligned}$$

Working with a topologically ordered set of nodes, the presence of an edge $i \to j$ implies $i < j$. The inductive hypothesis thus yields that

$$\sum_{k\in P(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k),\pi\not\subset G_A} \prod_{a\to b\in\pi} \lambda_{ab} \right) = \sum_{j:i\to j} \lambda_{ij} \cdot 0 = 0,$$

which completes the inductive step and the proof of the lemma. □

Lemma 8. *Let G, Λ, Ω, α, and A be as in the statement of Proposition 3. Then for all $i \in A$,*

$$(I - \tilde{\Lambda})^{-1}_{i,\tilde{P}(m)} \alpha_{\tilde{P}(m)} = (I - \Lambda)^{-1}_{i,P(m)} \alpha.$$

Proof. The right hand side of the claimed equation can be rewritten as

$$(I-\Lambda)^{-1}_{i,P(m)}\alpha = \sum_{k\in P(m)} (I-\Lambda)^{-1}_{ik}\alpha_k = \sum_{k\in P(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k)} \prod_{a\to b\in\pi} \lambda_{ab} \right) =$$
$$\sum_{k\in P(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k),\pi\subset G_A} \prod_{a\to b\in\pi} \lambda_{ab} \right) + \sum_{k\in P(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k),\pi\not\subset G_A} \prod_{a\to b\in\pi} \lambda_{ab} \right).$$

By Lemma 7, the second sum is equal to zero. Note also that if $k \in P(m)\backslash A$, then there is no path $\pi \in \mathcal{P}(i,k)$ with $\pi \subset G_A$. Therefore, the first sum can be indexed over $k \in \tilde{P}(m)$. We thus obtain that, as claimed,

$$(I-\Lambda)^{-1}_{i,P(m)}\alpha = \sum_{k\in\tilde{P}(m)} \alpha_k \left(\sum_{\pi\in\mathcal{P}(i,k),\pi\subset G_A} \prod_{a\to b\in\pi} \lambda_{ab} \right)$$
$$= \sum_{k\in\tilde{P}(m)} \alpha_k (I-\tilde{\Lambda})^{-1}_{ik} = (I-\tilde{\Lambda})^{-1}_{i,\tilde{P}(m)}\alpha_{\tilde{P}(m)}. \qquad \square$$

Acknowledgments

Mathias Drton was supported by the NSF under Grant No. DMS-0746265 and by an Alfred P. Sloan Fellowship. Seth Sullivant was supported by the NSF under grant No. DMS-0840795 and by the David and Lucille Packard Foundation.

References

[Bol89] Kenneth 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. MR996025 (90k:62001)

[BP02] Carlos Brito and Judea Pearl, *A new identification condition for recursive models with correlated errors*, Struct. Equ. Model. **9** (2002), no. 4, 459–474. MR1930449

[Chu97] Fan R. K. Chung, *Spectral graph theory*, CBMS Regional Conference Series in Mathematics, vol. 92, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997. MR1421568 (97k:58183)

[CLO07] David Cox, John Little, and Donal O'Shea, *Ideals, varieties, and algorithms*, third ed., Undergraduate Texts in Mathematics, Springer, New York, 2007, An introduction to computational algebraic geometry and commutative algebra. MR2290010 (2007h:13036)

[Drt09] Mathias Drton, *Likelihood ratio tests and singularities*, Ann. Statist. **37** (2009), no. 2, 979–1012. MR2502658

[DY10] Mathias Drton and Josephine Yu, *On a parametrization of positive semidefinite matrices with zeros*, 2010, arXiv:1001.3195.

[McD02] Roderick P. McDonald, *What can we learn from the path equations?: identifiability, constraints, equivalence*, Psychometrika **67** (2002), no. 2, 225–249. MR1986335

[Pea09] Judea Pearl, *Causality*, second ed., Cambridge University Press, Cambridge, 2009, Models, reasoning, and inference. MR2548166

[RS02] Thomas Richardson and Peter Spirtes, *Ancestral graph Markov models*, Ann. Statist. **30** (2002), no. 4, 962–1030. MR1926166 (2003h:60017)

[SGS00] Peter Spirtes, Clark Glymour, and Richard Scheines, *Causation, prediction, and search*, second ed., Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2000, With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper and Thomas Richardson, A Bradford Book. MR1815675 (2001j:62009)

[SP06] Ilya Shpitser and Judea Pearl, *Identification of joint interventional distributions in recursive semi-Markovian causal models*, Proceedings of the Twenty-First National Conference on Artificial Intelligence (Menlo Park, CA), AAAI Press, 2006, pp. 1219–1226.

[Tia02] Jin Tian, *Studies in causal reasoning and learning*, Ph.D. thesis, UCLA Computer Science Department, 2002.

[Tia09] ———, *Parameter identification in a class of linear structural equation models*, Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) (Pasadena, California), 2009, pp. 1970–1975.

Department of Statistics, The University of Chicago, Chicago, IL, U.S.A.
E-mail address: drton@uchicago.edu, rina@uchicago.edu

Department of Mathematics, North Carolina State University, Raleigh, NC, U.S.A.
E-mail address: smsulli2@ncsu.edu