Testable Implications of Linear Structural Equation Models

Bryant Chen
University of California, Los Angeles
Computer Science Department
Los Angeles, CA, 90095-1596, USA
bryantc@cs.ucla.edu

Jin Tian
Iowa State University
Computer Science Department
Ames, IA 50011
jtian@iastate.edu

Judea Pearl
University of California, Los Angeles
Computer Science Department
Los Angeles, CA, 90095-1596, USA
judea@cs.ucla.edu

Abstract

In causal inference, all methods of model learning rely on testable implications, namely, properties of the joint distribution that are dictated by the model structure. These constraints, if not satisfied in the data, allow us to reject or modify the model. Most common methods of testing a linear structural equation model (SEM) rely on the likelihood ratio or chi-square test which simultaneously tests all of the restrictions implied by the model. Local constraints, on the other hand, offer increased power (Bollen and Pearl 2013; McDonald 2002) and, in the case of failure, provide the modeler with insight for revising the model specification. One strategy of uncovering local constraints in linear SEMs is to search for overidentified path coefficients. While these overidentifying constraints are well known, no method has been given for systematically discovering them. In this paper, we extend the half-trek criterion of (Foygel, Draisma, and Drton 2012) to identify a larger set of structural coefficients and use it to systematically discover overidentifying constraints. Still open is the question of whether our algorithm is complete.

Introduction

Many researchers, particularly in economics, psychology, and the social sciences, use structural equation models (SEMs) to describe the causal and statistical relationships between a set of variables, predict the effects of interventions and policies, and to estimate parameters of interest. This qualitative causal information (i.e. exclusion and independence restrictions (Pearl 2009)), which can be encoded using a graph, imply a set of constraints on the probability distribution over the underlying variables. These constraints can be used to test the model and reject it when they are not consistent with data.

In the case of linear SEMs, the most common method of testing a model is a likelihood ratio or chi-square test that compares the covariance matrix implied by the model to that of the population covariance matrix (Bollen 1989; Shipley 1997). While this test simultaneously tests all of the restrictions implied by the model, failure does not provide the modeler with information about which aspect of the model needs to be revised. Additionally, if the model is very large and complex, it is possible that a global chi-square test will not reject the model even when a crucial testable implication is violated. In contrast, if the testable implications are enumerated and tested individually, the power of each test is greater than that of a global test (Bollen and Pearl 2013; McDonald 2002), and, in the case of failure, the researcher knows exactly which constraint was violated. Finally, in order to use the global chi-square test, it is necessary to know the degrees of freedom. For models where all of the free parameters are identifiable, the degrees of freedom is $df = \frac{p(p+1)}{2} - n$ where p is the number of variables and n is the number of free parameters (Bollen 1989; Browne 1984). However, in cases where when one or more free parameters are not identifiable (the model is underidentified), this equation no longer holds. Instead, the degrees of freedom is equivalent to the number of equality constraints on the covariance matrix. (See discussion on SEMNET forum with subject heading, "On Degrees of Freedom".) Better understanding how to obtain and count these equality constraints provides insights into the validity of the global chi-square test.

There are a number of methods for discovering local equality constraints that can be applied to a linear structural equation model. It is well known that conditional independence relationships can be easily read from the causal graph using d-separation (Pearl 2009), and (Kang and Tian 2009) gave a procedure that enumerates a set of conditional independences that imply all others. Additionally, a tetrad is the difference in the product of pairs of covariances (e.g. $\sigma_{12}\sigma_{34} - \sigma_{13}\sigma_{24}$) and the structure of a linear SEM typically implies that some of the tetrads vanish while others do not (Bollen and Pearl 2013; Spearman 1904).

Overidentifying constraints, the subject of this paper, represent another strategy for obtaining local constraints in linear models. These constraints are obtained when there are at least two minimal sets of logically independent assumptions in the model that are sufficient for identifying a model coefficient, and the identified expressions for the coefficient are distinct functions of the covariance matrix (Pearl 2001; 2004). In this case, an equality constraint is obtained by

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

equating the two identified expressions for the coefficient[1,2]. When this constraint holds in the data, the overidentified coefficient has the additional benefit of "robustness" (Pearl 2004). (Brito 2004) gave a sufficient condition for overidentification. Additionally, some of the non-independence constraints described by (McDonald 2002) are equivalent to overidentifying constraints. Finally, the Sargan test, also known as the Hansen or J-Test, relies on overidentification to check the validity of an instrumental variable (Sargan 1958; Hansen 1982). However, to our knowledge, no algorithm has been given for the systematic listing of overidentifying constraints. With this goal in mind, we modify the identifiability algorithm of (Foygel, Draisma, and Drton 2012) in order to discover and list overidentifying constraints.

It is well known that Wright's rules allow us to equate the covariance of any two variables as a polynomial over the model parameters (Wright 1921). (Brito and Pearl 2002a) recognized that these polynomials are linear over a subset of the coefficients, thereby reducing the problem of identification to analysis of a system of linear equations and providing the basis for the "G-Criterion" of identifiability. (Brito 2004) also noted that in some cases more linear equations could be obtained than needed for identification leading to the previously mentioned condition for overidentification. (Foygel, Draisma, and Drton 2012) generalized the G-Criterion, calling their condition the "half-trek criterion" and gave an algorithm that determines whether a SEM, as a whole, is identifiable. This algorithm can be modified in a straightforward manner to give the identified expressions for model coefficients. In this paper, we extend this (modified) half-trek algorithm to allow identification of possibly more coefficients in underidentified models (models where at least one coefficient is not identifiable)[3] and use it to list overidentifying constraints. These overidentifying constraints can then be used in conjunction with conditional independence constraints to test local aspects of the model's structure. Additionally, they could potentially be incorporated into constraint-based causal discovery algorithms.

Related Work

In addition to the work discussed in the introduction, the identification problem has also been studied extensively by econometricians and social scientists (Fisher 1966; Bowden and Turkington 1990; Bekker, Merckens, and Wansbeek 1994; Rigdon 1995). More recently, the problem has been addressed by the AI community using graphical modeling techniques. The previously mentioned work by (Brito and Pearl 2002a; Brito 2004; Foygel, Draisma, and Drton 2012) as well as (Brito and Pearl ; 2002b) developed graphical criteria for identification based on Wright's equations, while other work by (Tian 2005; 2007; 2009) used partial regression equations instead.

Non-conditional independence constraints have also been explored by the AI community in the context of non-parametric causal models. They were first noted by (Verma and Pearl 1990) while (Tian and Pearl 2002) and (Shpitser and Pearl 2008) developed algorithms for systematically discovering such constraints using the causal graph.

Preliminaries

We will use $r_{YX.Z}$ to represent the regression coefficient of Y on X given Z. Similarly, we will denote the covariance of X on Y given Z as $\sigma_{YX.Z}$. Throughout the paper, we also assume without loss of generality that model variables have been standardized to mean zero and variance one.

A linear structural equation model consists of a set of equations of the form,

$$x_i = pa_i^t \beta_i + \epsilon_i$$

where pa_i (connoting *parents*) are the set of variables that directly determine the value of X_i, β_i is a vector of coefficients that convey the strength of the causal relationships, and ϵ_i represents errors due to omitted or latent variables. We assume that ϵ_i is normally distributed.

We can also represent the equations in matrix form:

$$\mathbf{X} = \mathbf{X}^t \Lambda + \epsilon,$$

where $\mathbf{X} = [x_i]$, Λ is a matrix containing the coefficients of the model with $\Lambda_{ij} = 0$ when X_i is not a cause of X_j, and $\epsilon = [\epsilon_1, \epsilon_2, ..., \epsilon_n]^t$.

The causal graph or path diagram of an SEM is a graph, $G = (V, D, B)$, where V are vertices, D directed edges, and B bidirected edges. The vertices represent model variables. Edges represent the direction of causality, and for each equation, $x_i = pa_i^t \beta_i + \epsilon_i$, edges are drawn from the variables in pa_i to x_i. Each edge, therefore, is associated with a coefficient in the SEM, which we will often refer to as its path coefficient. The error terms, ϵ_i, are not represented in the graph. However, a bidirected edge between two variables indicates that their corresponding error terms may be statistically dependent while the lack of a bidirected edge indicates that the error terms are independent.

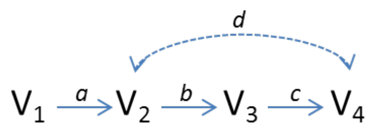
If an edge, called (X, Y), exists from X to Y then we say that X is a parent of Y. The set of parents of Y in a graph G is denoted $Pa_G(Y)$. Additionally, we call Y the head of (X, Y) and X the tail. The set of tails for a set of edges, E, is denoted $Ta(E)$ while the set of heads is denoted $He(E)$. If there exists a sequence of directed edges from X to Y then we say that X is an ancestor of Y. The set of ancestors of Y is denoted $An_G(Y)$. Finally, the set of nodes connected to Y by a bidirected arc are called the siblings of Y or $Sib_G(Y)$. In cases where the graph in question is obvious, we may omit the subscript G.

[1]Some authors use the term "overidentifying constraint" to describe any equality constraint implied by the model. We use it to describe only cases when a model coefficient is overidentified.

[2]Parameters are often described as *overidentified* when they have "more than one solution" (MacCallum 1995) or are "determined from [the covariance matrix] in different ways" (Jöreskog et al. 1979). However, expressing a parameter in terms of the covariance matrix in more than one way does not necessarily mean that equating the two expressions actually constrains the covariance matrix. See (Pearl 2001) and (Pearl 2004) for additional explanation and examples.

[3]Currently, state of the art SEM software like LISREL, EQS, and MPlus are unable to identify any coefficients in underidentified models.

Figure 1: SEM Graph



A *path* from X to Y is a sequence of edges connecting the two vertices. A path may go either along or against the direction of the edges. A non-endpoint vertex W on a path is said to be a *collider* if the edges preceding and following W on the path both point to W, that is, $\rightarrow W \leftarrow$, $\leftrightarrow W \leftarrow$, $\rightarrow W \leftrightarrow$, or $\leftrightarrow W \leftrightarrow$. A vertex that is not a collider is a *non-collider*.

A path between X and Y is said to be *unblocked given a set* Z (possibly empty), with $X, Y \notin Z$ if:

1. every noncollider on the path is not in Z
2. every collider on the path is in $An(Z)$ (Pearl 2009)

If there is no unblocked path between X and Y given Z, then X and Y are said to be *d-separated* given Z (Pearl 2009). In this case, the model dictates that X and Y are independent given Z.

Obtaining Constraints via Overidentification of Path Coefficients

The correlation between two variables in an SEM can be easily expressed in terms of the path coefficients using the associated graph and Wright's path-tracing rules (Wright 1921; Pearl 2013). These expressions can then be used to identify path coefficients in terms of the covariance matrix when the coefficients are identifiable. Consider the following model and its associated graph (Figure 1):

$$\begin{aligned} V_1 &= \epsilon_1 \\ V_2 &= aV_1 + \epsilon_2 \\ V_3 &= bV_2 + \epsilon_3 \\ V_4 &= cV_3 + \epsilon_4 \\ \text{Cov}[\epsilon_2, \epsilon_4] &= d \end{aligned}$$

Using the single-door criterion criterion (Pearl 2009) and Wright's path-tracing rules, we identify two distinct expressions for c in terms of the covariance matrix: $c = r_{43.2} = \frac{\sigma_{43.2}}{1-\sigma_{32}^2}$ and $c = \frac{abc}{ab} = \frac{\sigma_{41}}{\sigma_{31}}$, where $r_{YX.Z}$ is the regression coefficient of Y on X given Z and $\sigma_{XY.Z}$ is the covariance of X and Y given Z. As a result, we obtain the following constraint: $\frac{\sigma_{43.2}}{1-\sigma_{32}^2} = \frac{\sigma_{41}}{\sigma_{21}\sigma_{32}}$.

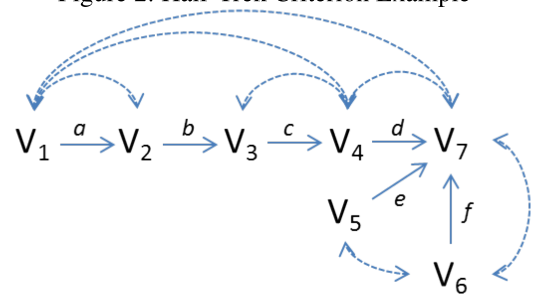
If violated, this constraint calls into question the lack of edge between V_1 and V_4.

Finding Constraints using an Extended Half-Trek Criterion

Identification

The half-trek criterion is a graphical condition that can be used to identify structural coefficients in an SEM, both recursive and non-recursive (Foygel, Draisma, and Drton 2012). In this section, we will paraphrase some preliminary definitions from (Foygel, Draisma, and Drton 2012) and present a generalization of the half-trek criterion that allows identifiability of potentially more coefficients in underidentified models.

Figure 2: Half-Trek Criterion Example



Definition 1. *(Foygel, Draisma, and Drton 2012) A* half-trek*, π, from X to Y is a path from X to Y that either begins with a bidirected arc and then continues with directed edges towards Y or is simply a directed path from X to Y.*

If there exists a half-trek from X to Y we say that Y is half-trek reachable from X. We denote the set of nodes that are reachable by half-trek from X, $htr(X)$.

Definition 2. *(Foygel, Draisma, and Drton 2012) For any half-trek, π, let* Right(π) *be the set of vertices in π that have an outgoing directed edge in π (as opposed to bidirected edge) union the last vertex in the trek. In other words, if the trek is a directed path then every vertex in the path is a member of Right(π). If the trek begins with a bidirected edge then every vertex other than the first vertex is a member of Right(π).*

Definition 3. *(Foygel, Draisma, and Drton 2012) A system of half-treks, $\pi_1, ..., \pi_n$, has* no sided intersection *if for all $\pi_i = \{\pi_1^i, ..., \pi_k^i\}, \pi_j = \{\pi_1^j, ..., \pi_m^j\}$, $\pi_1^i \neq \pi_1^j$ and Right(π_i)$\cap$Right(π_j)$= \emptyset$.*

The half-trek criterion is a sufficient graphical condition for identifiability of the path coefficients of a variable, v. If satisfied, we can obtain a set of linear equalities among the covariance matrix and the coefficients of v. Further, this set of equations is linearly independent with respect to the coefficients of v, and we can therefore use standard methods for solving a set of linearly independent equations to identify the expressions for coefficients of v. Here, we modify the half-trek criterion to identify *connected edge sets* (defined below), which are subsets of a variable's coefficients, rather than all of the variable's coefficients at once. As a result, an unidentifiable edge will inhibit identification of the edges in its connected edge set only and not all of v's coefficients. In this way, we increase the granularity of the criterion to determine identifiability of some variable coefficients even when they are not all identifiable.

Definition 4. *Let $Pa_1, Pa_2, ..., Pa_k$ be the unique partition of Pa(v) such that any two parents are placed in the same* *subset, Pa_i, whenever they are connected by an unblocked path. A* connected edge set *with head v is a set of edges from Pa_i to v for some $i \in \{1, 2, ..., k\}$.*

In Figure 2, there are two connected edge sets with head V_7. One is $\{(V_4, V_7)\}$ and the other is $\{(V_5, V_7), (V_6, V_7)\}$. V_4 has no path to other parents of V_7 while V_5 and V_6 are connected by a bidirected arc.

Definition 5. *(Edge Set Half-Trek Criterion) Let E be a connected edge set with head v. A set of variables Y satisfies the half-trek criterion with respect to E, if*

(i) $|Y| = |E|$

(ii) $Y \cap (v \cup Sib(v)) = \emptyset$ *and*

(iii) There is a system of half-treks with no sided intersection from Y to $Ta(E)$.

When it is clear from the context, we will simply refer to the edge set half-trek criterion as the half-trek criterion. An edge set, E, is identifiable if there exists a set, Y_E, that satisfies the half-trek criterion with respect to E. However, Y_E must consist only of "allowed" nodes. Intuitively, a node, y, is allowed for E if it is either not half-trek reachable from $He(E)$ or any of y's coefficients that lie on an unblocked path between $He(E) \cup Ta(E)$ and y are themselves identifiable. We define this notion formally below.

Definition 6. *Let ES be the set of connected edge sets in the causal graph, G. We say that a is an* HT-allowed *node for edge set E_v with head v if a is not half-trek reachable from v or both of the following conditions are satisfied:*

(i) There exists an ordering on ES, $\prec$, and a family of subsets (Y_E), one subset for each $E \prec E_v$, such that Y_E satisfies the half-trek criterion with respect to E and $E_i \prec E_j$ for $E_i, E_j \prec E_v$ whenever $He(E_i) \subseteq htr(He(E_j)) \cap Y_{E_j}$ and there exists an unblocked path between $Ta(E_i)$ and $He(E_j) \cup Ta(E_j)$.

(ii) The edge set of any edges belonging to a that lie on a half-trek from v to a are ordered before E_v

Let $CE(y, E)$ be the connected edge sets containing edges belonging to y that lie on an unblocked path from y to $He(E) \cup Ta(E)$. Now, define $\text{Allowed}(E, \text{IDEdgeSets}, G)$, used in Algorithms 1 and 2, as the set $(V \setminus htr(He(E))) \cup \{y | CE(y, E) \in \text{IDEdgeSets}\}$ for some set of connected edge sets, IDEdgeSets. Intuitively, $\text{Allowed}(E, \text{IDEdgeSets}, G)$ contains the set of nodes that have been determined to be allowable for E based on the edge sets that have been identified by the algorithm so far.

If Y is a set of allowed variables for E that satisfies the half-trek criterion with respect to E, we will say that Y is an *HT-admissible* set for E.

Theorem 1. *If a HT-admissible set for edge set E with head v exists then E is identifiable. Further, let $Y_E = \{y_1, ..., y_k\}$ be a HT-admissible set for E, $Ta(E) = \{p_1, ..., p_k\}$, and Σ be the covariance matrix of the model variables. Define $\mathbf{A}$ as*

$$\mathbf{A_{ij}} = \begin{cases} [(I - \Lambda)^T \Sigma]_{y_i, p_j}, & y_i \in htr(v), \\ \Sigma_{y_i, p_j}, & y_i \notin htr(v) \end{cases}$$

and $\mathbf{b}$ as

$$\mathbf{b_i} = \begin{cases} [(I - \Lambda)^T \Sigma]_{y_i, v}, & y_i \in htr(v), \\ \Sigma_{y_i, v}, & y_i \notin htr(v) \end{cases}$$

Then $\mathbf{A}$ is an invertible matrix and $\mathbf{A} \cdot \Lambda_{Ta(E), v} = \mathbf{b}$.

Proof. The proof for this theorem is similar to the proof of Theorem 1 (HTC-identifiability) in (Foygel, Draisma, and Drton 2012). Rather than giving the complete proof, we give some brief explanation for why our changes are valid. We made two significant changes to the half-trek criterion. First, we identify connected edge sets rather than the entire variable. Since different edge sets are unconnected, the paths from a half-trek admissible set, Y_E, to $v = He(E)$ travel only through the coefficients of E and no other coefficients of v. As a result, $\mathbf{A} \cdot \Lambda_{Ta(E), v} = \mathbf{b}$ is still valid. Additionally, $\mathbf{A}$ is still invertible due to the lack of sided intersection in Y_E.

Second, if $y \in Y_E$ is half-trek reachable from v, we do not require y to be fully identified but only the edges of y that lie on paths between $He(v) \cup Ta(v)$ and y. Let E_y be the edges of y and $E^c_{htr(v)} \subseteq E_y$ be the set of y's edges that *do not* lie on any half-trek from v to y. With respect to the matrix, $\mathbf{A}$, $(I - \Lambda)^T \Sigma)_{y, Ta(E)}$ is still obtainable since $\Sigma_{Ta(E^c_{htr(v)}), Ta(E)} = 0$. We do not need to identify the coefficients of $E^c_{htr(v)}$ since they will vanish from $\mathbf{A}$. Similarly, they vanish from $\mathbf{b}$ since $\Sigma_{Ta(E^c_{htr(v)}), v} = 0$. □

If a connected edge set E is identifiable using Theorem 1 then we say that E is *HT-identifiable*.

Using Figure 2 as an example, we consider the coefficients of the two connected edge sets with head V_7, $\{d\}$ and $\{e, f\}$. The coefficients, e and f, are not HT-identifiable, but d is. $\{V_3\}$ is an HT-admissible set for c even though $V_3 \in htr(V_7)$ since V_3's only coefficient, b, is identifiable using $\{V_2\}$. Therefore, each coefficient of V_3 that is reachable from V_7 is identifiable and it is allowed to be in the set Y_d. Since coefficients e and f are not HT-identifiable, the half-trek criterion of (Foygel, Draisma, and Drton 2012) simply states that the variable V_7 is not identifiable and fails to address the identifiability of d.

Finding a HT-admissible set for a connected edge set, E, with head, v, from a set of allowed nodes, A_E, can be accomplished by modifying the max-flow algorithm described in (Foygel, Draisma, and Drton 2012). First, we construct a graph, $G_f(E, A)$, with at most $2|V| + 2$ nodes and $3|V| + |D| + |B|$ edges, where D is the set of directed edges and B the set of bidirected edges in the original graph, G. The graph, $G_f(E, A)$, is constructed as follows:

First, $G_f(E, A)$ is comprised of three types of nodes:

(i) a source s and a sink t

(ii) a "left-hand copy" $L(a)$ for each $a \in A$

(iii) a "right-hand copy" $R(w)$ for each $w \in V$

The edges of $G_f(e, A)$ are given by the following:

(i) $s \to L(a)$ and $L(a) \to R(a)$ for each $a \in A$

(ii) $L(a) \to R(w)$ for each $a \leftrightarrow w \in B$

(iii) $R(w) \to R(u)$ for each $w \to u \in D$

(iv) $R(w) \to t$ for each $w \in Ta(E)$

Finally, all edges have capacity ∞, the source, s, and sink, t, have capacity ∞, and all other nodes have capacity 1. Intuitively, a flow from source to sink represents a half-trek in the original graph, G, while the node capacity of 1 ensures that there is no sided intersection. The only difference between our construction and that of (Foygel, Draisma, and Drton 2012) is that instead of attaching parents of v to the sink node in the max-flow graph, we attach only nodes belonging to $Ta(E)$ to the sink node.

Once the graph, $G_f(E, A)$, is constructed, we then determine the maximum flow using any of the standard maximum flow algorithms. The size of the maximum flow indicates the number of variables that satisfy conditions (ii) and (iii) of the Edge Set Half-Trek Criterion. Therefore, if the size of the flow is equal to $|E|$ then E is identifiable. Further, the "left-hand copies" of the allowed nodes with non-zero flow are the nodes that can be included in the HT-admissible set, Y_E. Let $\mathrm{MaxFlow}(G_f(E, A))$, to be used in Algorithms 1 and 2, return the "left-hand copies" of the allowed nodes with non-zero flow.

Theorem 2. *Given a causal graph, $G = (V, D, B)$, a connected edge set, E, and a subset of "allowed" nodes, $A \subseteq V \setminus (\{v\} \cup sib(v))$, there exists a set $Y \subseteq A$ satisfying the half-trek criterion with respect to E if and only if the flow network $G_f(E, A)$ has maximum flow equal to $|Ta(E)|$. Further, the nodes in $L(A)$ of the flow graph, $G_f(E, A)$ with non-zero flow satisfy conditions (ii) and (iii) of the Edge Set Half-Trek Criterion, and therefore, comprise a HT-admissible set for E if the maximum flow is equal to $|Ta(E)|$.*

The proof for this theorem is omitted since it is very similar to the proof for Theorem 6 in (Foygel, Draisma, and Drton 2012). Algorithm 1 gives a method for identifying coefficients in a linear SEM, even when the model is underidentified[4].

[4]For the case of acyclic graphs, the half-trek method can be extended in scope by first decomposing the distribution by c-components as described in (Tian 2005). As a result, if the graph is acyclic then the distribution should first be decomposed before applying the half-trek method. For the sake of brevity, we will not discuss decomposition in detail and instead refer the reader to (Foygel, Draisma, and Drton 2012) and (Tian 2005).

Algorithm 1 Identify

> **Input:** $G = (V, D, B), ES$
> **Initialize:** $\mathrm{IDEdgeSets} \leftarrow \emptyset$.
> **repeat**
>   **for each** E in $ES \setminus \mathrm{IDEdgeSets}$ **do**
> $A_E \leftarrow \mathrm{Allowed}(E, \mathrm{IDEdgeSets}, G)$
> $Y_E \leftarrow \mathrm{MaxFlow}(G_f(v, A))$
>   **if** $|Y_E| = |Ta(E)|$ **then**
>   Identify E using Theorem 1
> $\mathrm{IDEdgeSets} \leftarrow \mathrm{IDEdgeSets} \cup E$
>   **end if**
>   **end for**
> **until** $\mathrm{IDEdgeSets} = ES$ or no coefficients have been identified in the last iteration

---

Overidentifying Constraints

Once we have identified a connected edge set, E, using the half-trek criterion, we can obtain a constraint if there exists an alternative HT-admissible set for E, $Y_E^{'} \neq Y_E$, since $Y_E^{'}$ can then be used to find distinct expressions for some of the path coefficients associated with E. Further, these distinct expressions are derived from different aspects of the model structure and, therefore, are derived from logically independent sets of assumptions specified by the model. As a result, an equality constraint is obtained by setting both expressions for a given coefficient equal to one another. The following lemma provides a condition for when such a set exists.

Lemma 1. *Let the vertex set, Y_E, be an HT-admissible set for connected edge set, E, and let v be the head of E. An alternative HT-admissible set for E, $Y_E^{'} \neq Y_E$, exists if there exists a vertex, $w \notin Y_E$, such that*

1. *there exists a half-trek from w to $Ta(E)$,*
2. *$w \notin (v \cup Sib(v))$, and*
3. *if $w \in htr(v)$, the connected edge sets containing edges belonging to w that lie on unblocked paths between w and $He(E) \cup Ta(E)$ are HT-identifiable.*

Proof. Let $\pi_w^{'}$ be any half-trek from w to $Ta(E)$. Clearly, $\{\pi_w^{'}\} \cup \Pi$ has a sided intersection since there are more half-treks in $\{\pi_w^{'}\} \cup \Pi$ than vertices in $Ta(E)$. Now, let π be the first half-trek that $\pi_w^{'}$ intersects and x be the first vertex that they share. More precisely, if we order the vertices in $\pi_w^{'}$ so that w is the first vertex, $t \in Ta(E)$ the last, and the remaining vertices ordered according to the path from w to t then x is the first vertex in the set $(\cup_{i|\pi_i \in \Pi} \mathrm{Right}(\pi_i)) \cap \mathrm{Right}(\pi_w^{'})$ and π is the half-trek in Π such that $x \in \mathrm{Right}(\pi)$. Note that π is unique because Π has no sided intersection. Now, let π_w be the half-trek from w to $Ta(E)$ that is the same as $\pi_w^{'}$ until vertex x after which it follows the path of π. π_w is a half-trek from w to v and $\Pi^{'} = \Pi \setminus \{\pi\} \cup \{\pi_w\}$ has no sided intersection. Now if we let z be the first vertex in π then it follows that $Y_E^{'} = Y_E \setminus \{z\} \cup \{w\}$ satisfies the half-trek criterion with respect to E. □

Note that this proof also shows how to obtain the alternative set $Y_E^{'}$ once a variable w that satisfies Lemma 1 is found. In Figure 1, (V_3, V_4) can be identified using either $Y = \{V_3\}$, which yields $c = \sigma_{43.2}$ or $Y' = \{V_1\}$, which yields $c = \frac{\sigma_{41}}{\sigma_{31}}$. Both sets satisfy the half trek criterion and yield different expressions for c giving the same constraint as before, $\frac{\sigma_{43.2}}{1-\sigma_{32}^2} = \frac{\sigma_{41}}{\sigma_{31}}$.

While overidentifying constraints can be obtained by identifying the edge set E using different HT-admissible sets, there is actually a more direct way to obtain the constraint. Let $Ta(E) = \{p_1, ..., p_k\}$, and Σ be the covariance matrix of the model variables. Define $\mathbf{a_w}$ as

$$\mathbf{a_w} = \begin{cases} [(I-\Lambda)^T\Sigma]_{w,p_j}, & w \in htr(v), \\ \Sigma_{w,p_j}, & w \notin htr(v) \end{cases}$$

and $\mathbf{b_w}$ as

$$\mathbf{b_w} = \begin{cases} [(I-\Lambda)^T\Sigma]_{w,v}, & w \in htr(v), \\ \Sigma_{w,v}, & w \notin htr(v) \end{cases}$$

From Theorem 1, we have that $\mathbf{A} \cdot \Lambda_{Ta(E),v} = \mathbf{b}$. Similarly, $\mathbf{a_w}^T \cdot \Lambda_{Ta(E),v} = \mathbf{b_w}$. Now, $\mathbf{a_w} = \mathbf{d}^T \cdot \mathbf{A}$ for some vector of constants $\mathbf{d}$ since $\mathbf{A}$ has full rank. As a result, $\mathbf{b_w} = \mathbf{d}^T \cdot \mathbf{b}$, giving us the constraint $\mathbf{a_w}^T\mathbf{A}^{-1}\mathbf{b} = \mathbf{b_w}$.

Clearly, if we are able to find additional variables that satisfy the conditions given above for w, we will obtain additional constraints. Constraints obtained using the above method may or may not be conditional independence constraints. We refer to constraints identified using the half-trek criterion as *HT-overidentifying constraints*. Algorithm 2 gives a procedure for systematically discovering HT-overidentifying constraints.

We use Figure 2 as an illustrative example of Algorithm 2. First, b is identifiable using $Y_b = \{V_2\}$, and V_1 satisfies the conditions of Lemma 1 giving the constraint $\sigma_{23} = \frac{\sigma_{13}}{\sigma_{12}}$, which is equivalent to the conditional independence constraint that V_1 is independent of V_3 given V_2. Next, d is identifiable using $Y_d = \{V_3\}$, which then allows a to be identifiable using $\{V_7\}$. Since a is identifiable, $\{V_2\}$ is HT-admissible for c, allowing c to be identified. $\{V_2\}$ also satisfies the conditions of Lemma 1 for d giving the constraint:

$$\frac{-\sigma_{23}\sigma_{27} + \sigma_{37}}{-\sigma_{23}\sigma_{24} + \sigma_{34}} = \frac{-\frac{\sigma_{14}\sigma_{23}\sigma_{27} - \sigma_{14}\sigma_{37} + \sigma_{17}}{\sigma_{24}\sigma_{23}\sigma_{27} - \sigma_{24}\sigma_{37} + \sigma_{27}}\sigma_{17} + \sigma_{27}}{-\frac{\sigma_{14}\sigma_{23}\sigma_{27} - \sigma_{14}\sigma_{37} + \sigma_{17}}{\sigma_{24}\sigma_{23}\sigma_{27} - \sigma_{24}\sigma_{37} + \sigma_{27}}\sigma_{14} + \sigma_{24}}.$$

Note that d is not identifiable using other known methods and, as a result, this constraint is not obtainable using the methods of (Brito 2004) or (McDonald 2002). Lastly, a is also overidentified using $\{V_4\}$ yielding the constraint:

$$\frac{\sigma_{14}\sigma_{23}\sigma_{27} - \sigma_{14}\sigma_{37} + \sigma_{17}}{\sigma_{24}\sigma_{23}\sigma_{27} - \sigma_{24}\sigma_{37} + \sigma_{27}} = \frac{-c\sigma_{23} + \sigma_{24}}{-c\sigma_{13} + \sigma_{14}}$$

where

$$c = \frac{\frac{-\sigma_{14}\sigma_{23}\sigma_{27} + \sigma_{14}\sigma_{37} - \sigma_{17}}{\sigma_{24}\sigma_{23}\sigma_{27} - \sigma_{24}\sigma_{37} + \sigma_{27}}\sigma_{14} + \sigma_{24}}{\frac{-\sigma_{14}\sigma_{23}\sigma_{27} + \sigma_{14}\sigma_{37} - \sigma_{17}}{\sigma_{24}\sigma_{23}\sigma_{27} - \sigma_{24}\sigma_{37} + \sigma_{27}}\sigma_{13} + \sigma_{23}}$$

Finally, suppose an overidentifying constraint is violated. How can the modeler use this information to revise his or her assumptions? Consider a connected edge set E with head v and a HT-admissible set Y_E. If a variable w satisfies the conditions of Lemma 1 then the set $Y_E \cup \{w\}$ must have at least one variable that is not a parent of v. Let $K = Y_E \setminus (Pa(v) \cup \{w\})$. If the overidentifying constraint, $\mathbf{b_w} = \mathbf{a_w} \cdot \mathbf{A^{-1}} \cdot \mathbf{b}$, is violated then the modeler should consider adding an edge, either directed, bidirected, or both, between K to v since doing so eliminates the constraint.

Algorithm 2 Find Constraints

```
Input: G = (V, D, B), ES
Initialize: IDEdgeSets ← ∅.
repeat
  for each E ∈ ES do
    A_E ← Allowed(E, IDEdgeSets, G)
    Y_E ← MaxFlow(G_f(E, A))
    if |Y_E| = |Ta(E)| then
      if E ∉ IDEdgeSets then
        Identify E using Theorem 1
        IDEdgeSets ← IDEdgeSets ∪ E
      end if
      for each w in A_E \ Y_E do
        if v ∈ htr(w) then
          Output constraint: b_w = a_w · A^{-1} · b
        end if
      end for
    end if
  end for
until one iteration after all edges are identified or no new
  edges have been identified in the last iteration
```

Conclusion

In this paper, we extend the half-trek criterion in order to identify additional coefficients when the model is underidentified. We then use our extended criterion to systematically discover overidentifying constraints. These local constraints can be used to test the model structure and potentially be incorporated into constraint-based discovery methods.

Acknowledgements

We thank the reviewers for their comments. Part of this research was conducted while Bryant Chen was working at the IBM Almaden Research Center. Finally, this research was supported in parts by grants from NSF #IIS-1249822 and #IIS-1302448 and ONR #N00014-13-1-0153 and #N00014-10-1-0933.

References

Bekker, P.; Merckens, A.; and Wansbeek, T. 1994. *Identification, Equivalent Models, and Computer Algebra*. Statistical Modeling and Decision Science. Academic Press.

Bollen, K. A., and Pearl, J. 2013. Eight myths about causality and structural equation models. In *Handbook of Causal Analysis for Social Research*. Springer. 301–328.

Bollen, K. 1989. *Structural Equations with Latent Variables*. New York: John Wiley.

Bowden, R., and Turkington, D. 1990. *Instrumental Variables*. Econometric Society Monographs. Cambridge University Press.

Brito, C., and Pearl, J. Generalized instrumental variables. In Darwiche, A., and Friedman, N., eds., *Uncertainty in Artificial Intelligence, Proceedings of the Eighteenth Conference*. 85–93.

Brito, C., and Pearl, J. 2002a. A graphical criterion for the identification of causal effects in linear models. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*. 533–538.

Brito, C., and Pearl, J. 2002b. A new identification condition for recursive models with correlated errors. *Journal Structural Equation Modeling* 9(4):459–474.

Brito, C. 2004. *Graphical methods for identification in structural equation models*. Ph.D. Dissertation, Computer Science Department, University of California, Los Angeles, CA.

Browne, M. W. 1984. Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology* 37(1):62–83.

Fisher, F. 1966. *The identification problem in econometrics*. Economics handbook series. McGraw-Hill.

Foygel, R.; Draisma, J.; and Drton, M. 2012. Half-trek criterion for generic identifiability of linear structural equation models. *The Annals of Statistics* 40(3):1682–1713.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society* 1029–1054.

Jöreskog, K. G.; Sörbom, D.; Magidson, J.; and Cooley, W. W. 1979. *Advances in factor analysis and structural equation models*. Abt Books Cambridge, MA.

Kang, C., and Tian, J. 2009. Markov properties for linear causal models with correlated errors. *The Journal of Machine Learning Research* 10:41–70.

MacCallum, R. C. 1995. Model specification: Procedures, strategies, and related issues.

McDonald, R. 2002. What can we learn from the path equations?: Identifiability constraints, equivalence. *Psychometrika* 67(2):225–249.

Pearl, J. 2001. Parameter identification: A new perspective. Technical Report R-276, Department of Computer Science, University of California, Los Angeles, CA.

Pearl, J. 2004. Robustness of causal claims. In Chickering, M., and Halpern, J., eds., *Proceedings of the Twentieth Conference Uncertainty in Artificial Intelligence*. Arlington, VA: AUAI Press. 446–453.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press, 2nd edition.

Pearl, J. 2013. Linear models: A useful microscope for causal analysis. *Journal of Causal Inference* 1(1):155–170.

Rigdon, E. E. 1995. A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research* 30(3):359–383.

Sargan, J. D. 1958. The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society* 393–415.

Shipley, B. 1997. An inferential test for structural equation models based on directed acyclic graphs and its nonparametric equivalents. Technical report, Department of Biology, University of Sherbrooke, Canada. Also in *Structural Equation Modelling*, 7:206–218, 2000.

Shpitser, I., and Pearl, J. 2008. Dormant independence. In *Proceedings of the Twenty-Third Conference on Artificial Intelligence*. 1081–1087.

Spearman, C. 1904. General intelligence, objectively determined and measured. *The American Journal of Psychology* 15(2):201–292.

Tian, J., and Pearl, J. 2002. On the testable implications of causal models with hidden variables. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, 519–527.

Tian, J. 2005. Identifying direct causal effects in linear models. In *Proceedings of the National Conference on Artificial Intelligence*, volume 20, 346.

Tian, J. 2007. A criterion for parameter identification in structural equation models. In *Proceedings of the Twenty-Third Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-07)*, 392–399.

Tian, J. 2009. Parameter identification in a class of linear structural equation models. In *Proceedings of the International Joint Conferences on Artificial Intelligence*, 1970–1975.

Verma, T., and Pearl, J. 1990. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, 220–227.

Wright, S. 1921. Correlation and causation. *Journal of Agricultural Research* 20:557–585.