Directed Cyclic Graphical Representations of Feedback Models[1]

by

Peter Spirtes

1. Introduction

The introduction of statistical models represented by directed acyclic graphs (DAGs) has proved fruitful in the construction of expert systems, in allowing efficient updating algorithms that take advantage of conditional independence relations (Pearl, 1988, Lauritzen et. al., 1993), and in inferring causal structure from conditional independence relations (Spirtes and Glymour, 1991, Spirtes, Glymour and Scheines, 1993, Pearl and Verma, 1991, Cooper, 1992). As a framework for representing the combination of causal and statistical hypotheses, DAG models have shed light on a number of issues in statistics ranging from Simpson's Paradox to experimental design (Spirtes, Glymour and Scheines, 1993). The relations of DAGs with statistical constraints, and the equivalence and distinguishability properties of DAG models, are now well understood, and their characterization and computation involves three properties connecting graphical structure and probability distributions: (i) a local

---

[1] Research for this paper was supported by the National Science Foundation through grant 9102169 and the Navy Personnel Research and Development Center and the Office of Naval Research through contract number N00014-93-1-0568. I am indebted to Clark

directed Markov property, (ii) a global directed Markov property, (iii) and factorizations of joint densities according to the structure of a graph (Lauritizen, et al., 1990).

Recursive structural equation models are one kind of DAG model. However, non-recursive structural equation models are not DAG models, and are instead naturally represented by directeed *cyclic* graphs in which a finite series of edges representing influence leads from a vertex representing a variable back to that same vertex. Such graphs have been used to model feedback systems in electrical engineering (Mason, 1953, 1956), and to represent economic processes (Haavelmo, 1943, Goldberger, 1973). In contrast to the acyclic case, almost nothing general is known about how directed cyclic graphs (DCGs) represent conditional independence constraints, or about their equivalence or identifiability properties, or about characterizing classes of DCGs from conditional independence relations or other statistical constraints. This paper addresses the first of these problems, which is a prerequisite for the others. The issues turn on how the relations among properties (i), (ii) and (iii) essential to the acyclic case generalize--or more typically fail to generalize--to directed cyclic graphs and associated families of distributions. It will be shown that when DCGs are interpreted by analogy with DAGs as representing functional dependencies with independently distributed noises or "error terms," the equivalence of the fundamental global and local Markov conditions characteristc of DAGs no longer holds, even in linear systems, and in non-linear systems both Markov properties may fail. For linear systems associated with DCGs with independent errors or noises, a characterisation of conditional independence constraints is obtained, and it is shown that the result

---

Glymour, Richard Scheines, Christopher Meek, Thomas Richardson, and Marek Druzdel

generalizes in a natural way to systems in which the error variables or noises are statistically dependent. For non-linear systems with independent errors a sufficient condition for conditional independence of variables in associated distributions is obtained.

The remainder of this paper is organized as follows: Section 2 defines relevant mathematical ideas and gives some necessary technical results on DAGs and DCGs. Section 3 obtains results for non-recursive linear structural equations models. Section 4 treats non-linear models of the same kind.

2. Directed Graphs

I place sets of variables and defined terms in boldface, and individual variables in italics. A **directed graph** is an ordered pair of a finite set of vertices **V**, and a set of directed edges **E**. A directed edge from A to B is an ordered pair of distinct vertices $<A,B>$ in **V** in which A is the **tail** of the edge and B is the **head**; the edge is **out of** A and **into** B, and A is **parent** of B and B is a **child** of A. A sequence of distinct edges $<E_1,...,E_n>$ in G is an **undirected path** if and only if there exists a sequence of vertices $<V_1,...,V_{n+1}>$ such that for $1 \leq i \leq n$ either $<V_i,V_{i+1}> = E_i$ or $<V_{i+1},V_i> = E_i$. A path U is **acyclic** if no vertex occurring on an edge in the path occurs more than once. A sequence of distinct edges $<E_1,...,E_n>$ in G is a **directed path** if and only if there exists a sequence of vertices $<V_1,...,V_{n+1}>$ such that for $1 \leq i \leq n$ $<V_i,V_{i+1}> = E_i$. If there is an acyclic directed path from A to B or $B = A$ then A is an **ancestor** of B, and B is a **descendant** of A. A directed graph is **acyclic** if and only if it contains no directed cyclic paths.[2]

---

for helpful conversations.

[2]An undirected path is often defined as a sequence of vertices rather than a sequence of

A **directed acyclic graph** (DAG) G with a set of vertices **V** can be given two distinct interpretations. On the one hand, such graphs can be used to represent causal relations between variables, where an edge from A to B in G means that A is a direct cause of B relative to **V**. A **causal graph** is a DAG given such an interpretation. On the other hand, a DAG with a set of vertices **V** can also represent a set of probability measures over **V**. Following the terminology of Lauritzen et. al. (1990) say that a probability measure over a set of variables **V** satisfies the **local directed Markov property** for a DAG G with vertices **V** if and only if for every W in **V**, W is independent of **V**\(**Descendants**(W,G) $\cup$ **Parents**(W,G)) given **Parents**(W,G), where **Parents**(W,G) is the set of parents of W in G, and **Descendants**(W,G) is the set of descendants of W in G. A DAG G **represents** the set of probability measures which satisfy the local directed Markov property for G. The use of DAGs to simultaneously represent a set of causal hypotheses and a family of probability distributions extends back to the path diagrams introduced by Sewell Wright (1934). Variants of DAG models were introduced in the 1980's in Wermuth (1980), Wermuth and Lauritzen (1983), Kiiveri, Speed, and Carlin (1984), Kiiveri and Speed (1982), and Pearl (1988). [3]

Lauritzen et. al. also define a **global directed Markov property** that is equivalent to the local directed Markov property for DAGs. Several

---

edges. The two definitions are essentially equivalent for acyclic directed graphs, because a pair of vertices can be identified with a unique edge in the graph. However, a cyclic graph may contain more than one edge between a pair of vertices. In that case it is no longer possible to identify a pair of vertices with a unique edge.

[3]It is often the case that some further restrictions are placed on the set of distributions represented by a DAG. For example, one could also require the Minimality Condition, i.e. that for any distribution P represented by G, P does not satisfy the local directed Markov Condition for any proper subgraph of G. This condition, and others are discussed in Pearl(1988) and Spirtes, Glymour, and Scheines(1993). We will not consider such further restrictions here.

preliminary notions are required. Let $\mathbf{An}(\mathbf{X},G)$ be the set of ancestors of members of $\mathbf{X}$ in G. Let $G(\mathbf{X})$ be the subgraph of G that contains only vertices in $\mathbf{X}$, with an edge from A to B in $\mathbf{X}$ if and only if there is an edge from A to B in G. G^M **moralizes** a directed graph G if and only if G^M is an undirected graph with the same vertices as G, and a pair of vertices X and Y are adjacent in G^M if and only if either X and Y are adjacent in G, or they have a common child in G. In an undirected graph G, $\mathbf{X}$ is separated from $\mathbf{Y}$ given $\mathbf{Z}$ if and only if every undirected path between a member of $\mathbf{X}$ and a member of $\mathbf{Y}$ contains a member of $\mathbf{Z}$. If $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ are disjoint sets of variables, $\mathbf{X}$ and $\mathbf{Y}$ are **d-separated** given $\mathbf{Z}$ in a directed graph G just when $\mathbf{X}$ and $\mathbf{Y}$ are separated given $\mathbf{Z}$ in $G^M(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G))$. The relation defined here was stated in Lauritzen, et. al. (1990). "d-separation" is a graphical relation introduced by Pearl (1986). Since Lauritzen et. al. (1990) proved that their graphical relation is equivalent to Pearl's for acyclic graphs, and the proof is readily extended to the cyclic case, I will also use "d-separation" to refer to the graphical relation just described. Now the definition: A probability measure over $\mathbf{V}$ satisfies the **global directed Markov property** for DAG G if and only if for any three disjoint sets of variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ included in $\mathbf{V}$, if $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$, then $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$. Lauritzen et. al. (1990) shows that the global and local directed Markov properties are equivalent in DAGs, even when the probability distributions represented have no density function. In section 2, I show that the local and global directed Markov properties are not equivalent for cyclic directed graphs.

The following lemmas relate the global directed Markov property to factorizations of a density function. Denote a density function over $\mathbf{V}$ by $f(\mathbf{V})$, where for any subset $\mathbf{X}$ of $\mathbf{V}$, $f(\mathbf{X})$ denotes the marginal of $f(\mathbf{V})$. If $f(\mathbf{V})$ is the density function for a probability measure over a set of variables **V**, say that *f*(**V**) **factors according to directed graph** *G* with vertices **V** if and only if for every subset **X** of **V**,

$$f(\mathbf{An}(\mathbf{X},G)) = \prod_{V \in \mathbf{An}(\mathbf{X},G)} g_V(V, \mathbf{Parents}(V,G))$$

Equation 1

where g_V is a non-negative function. The following result was proved in Lauritzen et. al. (1990).

Lemma 1: If **V** is a set of random variables with a probability measure *P* that has a density function *f*(**V**), then *f*(**V**) factors according to DAG *G* if and only if *P* satisfies the global directed Markov property for *G*.

As in the case of acyclic graphs, the existence of a factorization according to a cyclic directed graph *G* does entail that a measure satisfies the global directed Markov property for *G*. The proof given in Lauritzen et. al. (1990) for the acyclic case carries over essentially unchanged for the cyclic case.

Lemma 2: If **V** is a set of random variables with a probability measure *P* that has a density function *f*(**V**) and *f*(**V**) factors according to directed (cyclic or acyclic) graph *G*, then *P* satisfies the global directed Markov property for *G*.

However, unlike the case of acyclic graphs, if a probability measure over a set of variable **V** satisfies the global directed Markov property for cyclic graph *G* and has a density function *f*(**V**), it does not follow that *f*(**V**) factors according to *G*.

The following weaker result relating factorization of densities and the global directed Markov property does hold for both cyclic and acyclic directed graphs.

Lemma 3: If **V** is a set of random variables with a probability measure P that has a positive density function $f(\mathbf{V})$, and P satisfies the global directed Markov property for directed (cyclic or acyclic) graph G, then $f(\mathbf{V})$ factors according to G.

3. Non-recursive Linear Structural Equation Models

The problem considered in this section is to investigate the generalization of the Markov properties to linear, non-recursive structural equation models. First we must relate the social scientific terminology to graphical representations, and clarify the questions.

Linear structural equation models (which, following the terminology of Bollen (1989), will be referred to as linear SEMs) can also be represented as directed graph models. In a linear SEM the random variables are divided into two disjoint sets, the error terms and the non-error terms. Corresponding to each non-error random variable V is a unique error term ε_V. A linear SEM contains a set of linear equations in which each non-error random variable V is written as a linear function of other non-error random variables and ε_V. A linear SEM also specifies a joint distribution over the error terms. So, for example, the following is a linear SEM, where a and b are real constants, ε_X ,ε_Y, and ε_Z are jointly independent "error terms", and X, Y, Z, are random variables:

$$X = a\ Y + \varepsilon_X$$
$$Y = b\ \ Z + \varepsilon_Y$$
$$Z = \varepsilon_Z$$

ε_X , ε_Y , ε_Z are jointly independent and normally distributed

Equation 2

The directed graph of a linear SEM with uncorrelated errors is written with the convention that an edge does not appear if and only if the corresponding entry in the coefficient matrix is zero; the graph does not contain the error terms. Figure 1 is the DAG that represents the SEM shown above. A linear SEM is **recursive** if and only if its directed graph is acyclic.

$$Z \longrightarrow Y \longrightarrow X$$

Figure 1: Example of Recursive SEM

Initially I will consider only linear SEMs in which the error terms are jointly independent, but we will see that in the linear case in an important sense nothing is lost by this restriction: a linear SEM with dependent errors generates the same restrictions on the covariance matrix as does some linear SEM with extra variables and independent errors. Further, such an SEM with extra variables can always be found with the same graphical structure on the original variables as obtain in the original graph.

A linear SEM containing disjoint sets of variables **X**, **Y**, and **Z** **linearly entails** that **X** is independent of **Y** given **Z** if and only if **X** is independent of **Y** given **Z** for all values of the non-zero linear coefficients and all distributions of the exogenous variables in which they have positive variances. Let $\rho_{XY.\mathbf{Z}}$ be the partial correlation of X and Y given **Z**. A linear SEM containing X, Y, and **Z**, where $X \neq Y$ and X and Y are not in **Z**, **linearly entails** that $\rho_{XY.\mathbf{Z}} = 0$ if and only $\rho_{XY.\mathbf{Z}} = 0$ for all values of the non-zero linear coefficients and all distributions of the exogenous variables in which they have positive variances and in which $\rho_{XY.\mathbf{Z}}$ is defined. It follows from Kiiveri and Speed (1982) that if the error terms are jointly independent, then any distribution that forms a linear, recursive SEM with a directed graph G satisfies the local directed Markov property for G. One can therefore apply d-separation to the DAG in a linear, recursive SEM to compute the conditional independencies and zero partial correlations it linearly entails. The d-separation relation provides a polynomial (in the number of vertices) time algorithm for deciding whether a given vanishing partial correlation is linearly entailed by a DAG.

Linear non-recursive structural equation models (linear SEMs) are commonly used in the econometrics literature to represent feedback processes that have reached equilibrium.[4] Corresponding to a set of non-recursive linear equations is a cyclic graph, as the following example from Whittaker (1990) illustrates.

$$X_1 = \varepsilon_{X1}$$
$$X_2 = \varepsilon_{X2}$$
$$X_3 = \beta_{31}X_1 + \beta_{34}X_4 + \varepsilon_{X3}$$
$$X_4 = \beta_{42}X_2 + \beta_{43}X_3 + \varepsilon_{X4}$$

ε_{X1}, ε_{X2}, ε_{X3}, ε_{X4} are jointly independent and normally distributed

Equation 3

[4]Cox and Wermuth (1993), Wermuth and Lauritzen(1990) and (indirectly) Frydenberg(1990) consider a class of non-recursive linear models they call *block recursive*. The block recursive models overlap the class of SEMs, but they are neither properly included in that class, nor properly include it. Frydenberg (1990) presents necessary and sufficient conditions for the equivalence of two block recursive models.

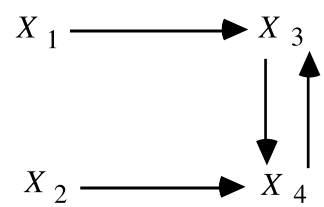


Figure 2: Example of Non-recursive SEM

In DAGs the global directed Markov property entails the local directed Markov property, because a variable V is d-separated from its non-parental non-descendants given its parents. This is not always the case in cyclic graphs. For example, in figure 4, X_4 is not d-separated from its non-parental non-descendant X_1 given its parents X_2 and X_3, so the local directed Markov property does not hold.[5]

Theorem 1: The probability measure P of a linear SEM L (recursive or non-recursive) with jointly independent error terms satisfies the global directed Markov property for the directed (cyclic or acyclic) graph G of L, i.e. if **X**, **Y**, and **Z** are disjoint sets of variables in G and **X** is d-separated from **Y** given **Z** in G, then **X** and **Y** are independent given **Z** in P.[6]

[5] Note that this use of cyclic directed graphs to represent feedback processes represents an extension of the causal interpretation of directed graphs. The causal structure corresponding to Figure 2 is described by an infinite acyclic directed graph containing each variable indexed by time. The cyclic graph can be viewed as a compact representation of such a causal graph. I am indebted to C. Glymour for pointing out that the local Markov condition fails in Whittaker's model. Indeed, there is *no* acyclic graph (even with additional variables) that linearly entails all and only conditional independence relations linearly entailed by Figure 2, although Thomas Richardson has pointed out that the directed cyclic graph of Figure 2 is equivalent to one in which in the edges from X_1 to X_3 and X_2 to X_4 are replaced, respectively, by edges from X_1 to X_4 and from X_2 to X_3

[6] This theorem has been independently proved by Jan Koster of the Erasmus University Rotterdam, in a paper which has not yet been published but has been submitted to Statistical Science.

Theorem 2: In a linear SEM L with jointly independent error terms and directed (cyclic or acyclic) graph G containing disjoint sets of variables $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, if $\mathbf{X}$ is not d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ then L does not linearly entail that $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$.

Applying Theorems 1 and 2 to the directed graph in Figure 2, only two conditional independence relations are entailed: X_1 is independent of X_2, and X_1 is independent of X_2 given X_3 and X_4. .

Theorem 3: In a linear SEM L with jointly independent error terms and (cyclic or acyclic) directed graph G containing X, Y and $\mathbf{Z}$, where $X \neq Y$ and $\mathbf{Z}$ does not contain X or Y, X is d-separated from Y given $\mathbf{Z}$ if and only if L linearly entails that $\rho_{XY.\mathbf{Z}} = 0$.

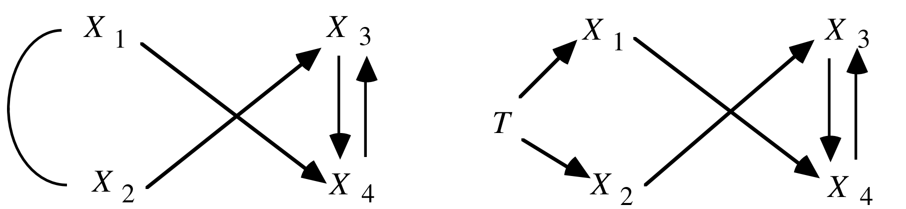
As in the acyclic case, d-separation provides a polynomial time procedure for deciding whether cyclic graphs entail a conditonal independence or vanishing partial correlation

Theorem 3 can be used to relax the restriction that the error terms in an a linear SEM L be jointly independent. If ε_X and ε_Y are not independent in linear SEM L, there is a linear SEM L' with independent error terms such that the marginal distribution of L' over the variables in L has the same covariance matrix as L. Form the graph G' of L' from the graph G of L in the following way. Add a latent variable T to G, and add edges from T to X and Y. In L', modify the equation for X by making it a linear functions of the parents of X (including T) in G', and replace ε_X by ε'_X; modify the equation for Y in an analogous way. There always exist linear coefficients and distributions over T and the new error terms such that the marginal covariance matrix for L' is equal to the covariance matrix of L, and ε'_X and

ε'_Y are independent. The process can be repeated for each pair of variables with correlated errors in L. Hence the zero partial correlations entailed by L can be derived by applying Theorem 3 to the graph of L'. Figure 3 illustrates this process. The set of variables **V** in the graph on the left is $\{X_1, X_2, X_3, X_4\}$. The graph on the left correlates the errors between X_1 and X_2 (indicated by the undirected edges between them.) The graph on the right has no correlated errors, but does have a latent variable T that is a parent of X_1 and X_2. The two graphs linearly entail the same zero partial correlations involving only variables in **V** (in this case they both entail no non-trivial zero partial correlations).

$X_3 = a \times X_2 + b \times X_4 + \varepsilon_3$	$X_3 = a \times X_2 + b \times X_4 + \varepsilon_3$
$X_4 = c \times X_1 + d \times X_3 + \varepsilon_4$	$X_4 = c \times X_1 + d \times X_3 + \varepsilon_4$
$X_1 = \varepsilon_1$	$X_1 = e \times T + \varepsilon'_1$
$X_2 = \varepsilon_2$	$X_2 = f \times T + \varepsilon'_2$
ε_1 and ε_2 correlated	ε'_1 and ε'_2 uncorrelated

Equation 4



Graph with Correlated Error **Graph without Correlated Error and Same Partial Correlations Over V**

Figure 3: Correlated Errors

3. Non-linear Structural Equation Models

A linear SEM is a special case of a SEM in which the equations relating a given variable to other variables and a unique error term need not be linear. In a SEM the random variables are divided into two disjoint sets, the error terms and the non-error terms. Corresponding to each non-error random variable V is a unique error term ε_V. A SEM contains a set of equations in which each non-error random variable V is written as a measureable function of other non-error random variables and ε_V.. The convention is that in the directed graph of a SEM there is an edge from A to B if and only if B is an argument in the function for A. As in the linear case, I will still assume that density functions exist for both the probabilty measure over the error terms and the non-error terms, that each non-error term V can also be written as a function of the error terms of its ancestors in G, that each ε_V is a function of V and its parents in G (which will be the case if the errors are additive or multiplicative), and that the Jacobean of the transformation between the error terms and the non-error terms is well-defined. Call such a set of equations and its associated graph a **pseudo-indeterministic** SEM (because the equations are actually deterministic if the unmeasured error terms are included, but appear indeterministic when the error terms are not measured.) A directed graph G **pseudo-indeterministically entails** that $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$ if and only if in every pseudo-indeterministic SEM with graph G, $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$.

This section establishes that d-separation again provides a fast algorithm for deciding whether a DAG pseudo-indeterministically entails a conditional independence relation, but in a cyclic directed graph d-separation may not pseudo-indeterministically entails a conditional independence relation. Instead, a different condition, yielding a polynomial time algorithm, is found to suffice for a cyclic direected graph to pseudo-indeterministically entail a conditional independence relation.

By Theorem 2, d-separation is a necessary condition for a conditional independence claim to be entailed by an SEM. The following remarks show d-separation is also sufficient for acyclic SEMs, but not for cyclic SEMS.

Theorem 4: If G is a DAG containing disjoint sets of variables **X**, **Y** and **Z**, **X** is d-separated from **Y** given **Z** if and only L pseudo-indeterministically entails that **X** is independent of **Y** given **Z**.

The following example gives a concrete illustration that there is a cyclic graph G in which X is d-separated from Y given $\{Z,W\}$, but G does not pseudo-indeterminstically entail that X is independent of Y given $\{Z,W\}$.

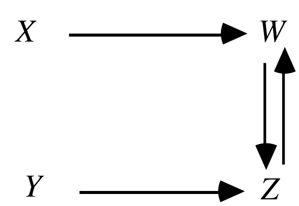


Figure 4: Graph G

$$X = \varepsilon_X$$
$$Y = \varepsilon_Y$$
$$Z = W \times Y + \varepsilon_Z$$
$$W = Z \times X + \varepsilon_Y$$

$\varepsilon_{X,}$ $\varepsilon_{X,}$ $\varepsilon_{X,}$ ε_X with independent standard normal distributions

Equation 5

The transformation from ε_X, ε_Y ε_Z, ε_W to X, Y, Z, W is 1-1 except where

$\varepsilon_X \times \varepsilon_Y = 1$ because

$$X = \varepsilon_X$$
$$Y = \varepsilon_Y$$
$$Z = \frac{\varepsilon_W \times \varepsilon_Y + \varepsilon_Z}{1-(\varepsilon_X \times \varepsilon_Y)}$$
$$W = \frac{\varepsilon_Z \times \varepsilon_X + \varepsilon_W}{1-(\varepsilon_X \times \varepsilon_Y)}$$

Equation 6

The Jacobean of the transformation from the ε's is $1/(1 + X\times Y)$. Hence, transforming the joint normal density of the ε's yields

$$f(X,Y,Z,W) =$$
$$\frac{1}{4\pi^2}\left(\exp(\frac{-x^2}{2})\times\exp(\frac{-y^2}{2})\times\exp(\frac{-(z-w\times y)^2}{2})\times\exp(\frac{-(w-z\times x)^2}{2})\right)\times$$
$$\left|\frac{1}{1+(X\times Y)}\right|$$

Equation 7

X is not independent of Y given $\{Z,W\}$ in this distribution because it is not possible to factor it into a product of terms, no one of which contains both X and Y.

However, it is possible to modify the graphical representation of the functional relations in such a way that d-separation applied to the new graph does entail conditional independence. In a directed graph G, a **cycle** is a cyclic directed path, in which each vertex occurs on exactly two edges. A set of cycles $\mathbf{C}$ is a **cyclegroup** if and only if it is a smallest set of cycles such that for each cycle C_1 in $\mathbf{C}$, $\mathbf{C}$ contains the transitive closure of all of the cycles intersecting C_1, i.e. it contains all of the cycles that intersect C_1, all of the cycles that intersect cycles that intersect C_1, etc. For example, in figure 7, there are two distinct cyclegroups: the first is $\{C_1,C_2,C_3\}$, and the second is $\{C_4, C_5\}$.

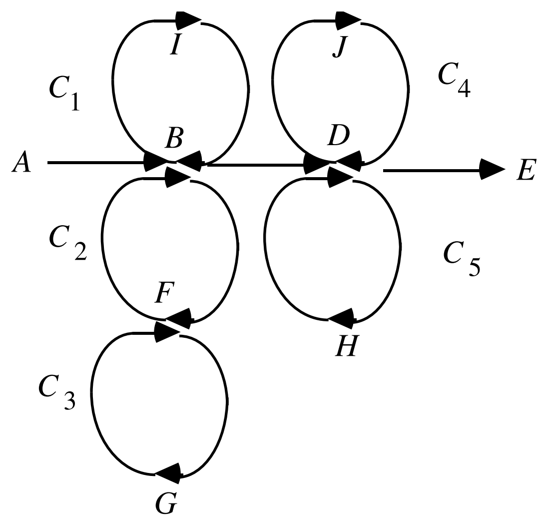


Figure 5: Cyclegroups

Let the set of all cycles in G be **Cycles**(G). If a vertex V or an edge $<V,W>$ occurs in some set **C** of cycles, for brevity write $V \in$ **C** or $<V,W> \in$ **C** respectively, although strictly speaking neither a vertex nor an edge is a member of a set of cycles. Form the **collapsed graph** G' from G by the following operations on each cyclegroup:

1. remove all of the edges between members of the cyclegroup;
2. arbitrarily number the vertices in the cyclegroup;
3. add an edge from each lower number vertex to each higher number vertex;
4. for each parent A of a member of the cyclegroup that is not itself in the cyclegroup, add an edge from A to each member of the cyclegroup.

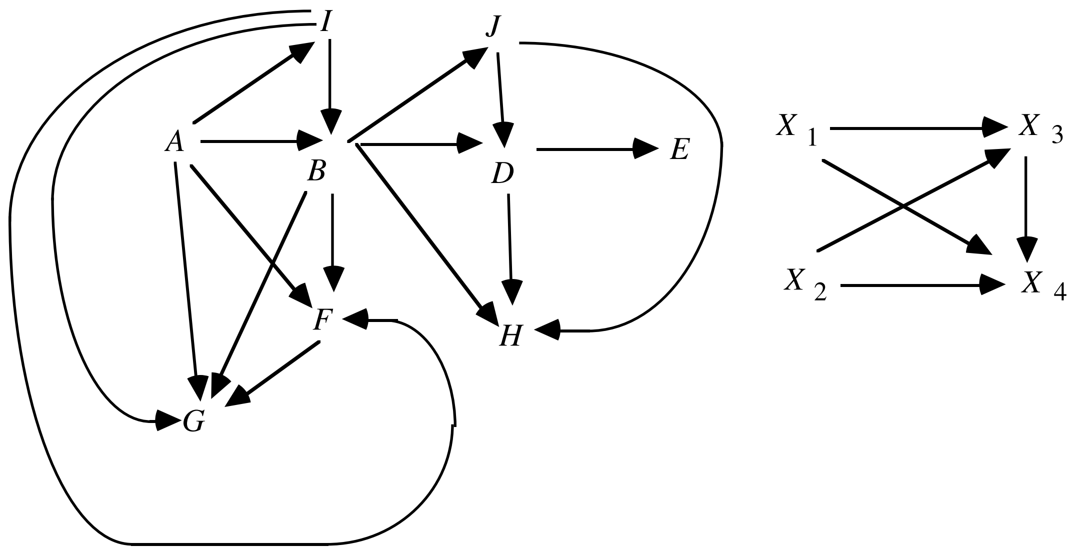
(The procedure does not define a unique collapsed graph due to the arbitrariness of the numbering, but since all of the collapsed graphs share the same d-separation relations, it does not matter.) Note that even if *G* is a cyclic graph, the collapsed graph is acyclic. The collapsed graph can be generated in polynomial time.

Theorem 5: In an SEM with directed graph *G* (cyclic or acyclic) and collapsed graph *G*' containing disjoint sets of variables **X**, **Y** and **Z**, if **X** is d-separated from **Y** given **Z** in *G*' then the SEM entails that **X** is independent of **Y** given **Z**.

A collapsed graph for the graph in figure 7 is shown in figure 8a, and a collapsed graph for the graph in figure 4 is shown in figure 8b.

I do not know whether the follow conjecture holds:

Conjecture:Let *G* (cyclic or acyclic) have collapsed graph *G*' containing disjoint sets of variables **X**, **Y** and **Z**. If *G* pseudo-indeterministically entails that **X** is independent of **Y** given **Z**, then in *G'* **X** is d-separated from **Y** given **Z**.



(a) (b)

Figure 6: Collapsed Graphs

5. Conclusion

These results raise a number of interesting questions whose answers may be of practical importance. Under what conditions, for example, are their results about conditional independence comparable to the equivalence of vanishing partial correlations in models with depndent errors and latent variable models with independent errors? There are polynomial algorithms (Verma and Pearl, 1990, Frydenberg, 1990) for determining when two arbitrary directed acyclic graphs entail the same set of conditional independence relations. Is there a polynomial algorithm for determining when two arbitrary directed graphs (cyclic or acyclic) linearly entail the same set of conditional independence relations? There are polynomial algorithms (Spirtes and Verma, 1992) for determining when two arbitrary directed acyclic graphs entail the same set of conditional independence relations over a common subset of variable **O**. Is there a polynomial algorithm for determining when two arbitrary directed graphs (cyclic or acyclic) linearly entail the same set of conditional independence relations over a common subset of variables **O**? Assuming Markov properties hold and completely characterize the condiitonal independence facts in distributions considered, there are correct polynomial algorithms for inferring features of (sparse) directed acyclic graphs from a probability distribution when there are no latent common causes (see Spirtes and Glymour, 1991, Cooper and Herskovitz, 1992). Are there comparable correct, polynomial algorithms for inferring features of directed graphs (cyclic or acyclic) from a probability distribution when there are no latent common causes? There are similarly correct, but not polynomial, algorithms for inferring features of directed acyclic graphs from a probability distribution even when there may be latent common causes (see Spirtes, 1992 and Spirtes, Glymour and Scheines, 1993). Are there comparable algorithms for inferring features of directed graphs (cyclic or acyclic) from a probability distribution even when there may be latent common causes?

References

Bollen, K., 1989, Structural Equations with Latent Variables. (Wiley, New York).

Frydenberg, M., 1990, The chain graph Markov property,*Scandinvaian Journal of Statistics,* **17**, 333-353.

Geiger, D., and Pearl, J., 1988, Logical and Algorithmic properties of Conditional Independence. Technical Report R-97, Cognitive Systems Laboratory, University of California, Los Angeles.

Goldberger, A., Duncan, O. (eds.), 1973, Structural Equation Models in the Social Sciences (Seminar Press, New York).

Haavelmo, T., 1943, The statistical implications of a system of simultaneous equations, *Econometrica*, **11**, 1-12.

Kiiveri, H. and Speed, T.,1982, Structural analysis of multivariate data: A review, *Sociological Methodology*, Leinhardt, S. (ed.). Jossey-Bass, San Francisco.

Kiiveri, H., Speed, T., and Carlin, J., 1984, Recursive causal models, *Journal of the Australian Mathematical Society*, **36**, 30-52.

Lauritzen, S., Dawid, A., Larsen, B., Leimer, H.,1990, Independence properties of directed Markov fields, *Networks*, **20**, 491-505.

Mason, S., 1953, Feedback theory-some properties of signal flow graphs, *Proceedings of the IRE*, **41**.

Mason, S., 1956, Feedback theory-further properties of signal flow graphs, *Proceedings of the IRE*, **44**.

Pearl, J.,1986, Fusion, propagation, and structuring in belief networks, *Artificial Intelligence* **29**, 241-88.

Pearl, J., 1988,. **Probabilistic Reasoning in Intelligent Systems**, (Morgan Kaufman: San Mateo, CA).

Pearl, J. and Verma, T. (1991). A theory of inferred causation, in **Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference** (Morgan Kaufmann, San Mateo, CA).

Spirtes, P. and Glymour, C., 1990, Causal Structure Among Measured Variables Preserved with Unmeasured Variables. Technical Report CMU-LCL-90-5, Laboratory for Computational Linguistics, Carnegie Mellon University.

Spirtes, P., and Glymour, C., 1991, An algorithm for fast recovery of sparse causal graphs, *Social Science Computer Review,* **9**, 62-72.

Spirtes, P., Glymour, C., and Scheines, R., 1993, **Causation, Prediction, and Search,** (Springer-Verlag Lecture Notes in Statistics 81, New York).

Wermuth, N.,1980, Linear recursive equations, covariance selection and path analysis, *Journal of the American Statistical Association*, **75**, 963-972.

Wermuth, N. and Lauritzen, S.,1983, Graphical and recursive models for contingency tables, *Biometrika*, **72**, 537-552.

Wermuth, N. and Lauritzen, S., 1990,. On substantive research hypotheses, conditional independence graphs and graphical chain models, *Journal of the Royal Statistical Society, Series B*, **52**, 21-50.

Whittaker, J.,1990, **Graphical Models in Applied Multivariate Statistics** (Wiley, New York).

Wright, S. (1934). The method of path coefficients, *Annals of Mathematical Statistics* **5**, 161-215.

Proofs for Referees

Lemma 3: If $\mathbf{V}$ is a set of random variables with a probability measure P that has a positive density function $f(\mathbf{V})$, and P satisfies the global directed Markov property for directed (cyclic or acyclic) graph G, then $f(\mathbf{V})$ factors according to G.

Proof. Assume that probability measure over $\mathbf{V}$ satisfies the global directed Markov property for directed (cyclic or acyclic) graph G. I will now show that for any disjoint sets of variables $\mathbf{R}$, $\mathbf{S}$, and $\mathbf{T}$ included in $\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G)$, if $\mathbf{R}$ and $\mathbf{S}$ are separated given $\mathbf{T}$ in $G^M(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G))$, then $\mathbf{R}$ and $\mathbf{S}$ are independent given $\mathbf{T}$. If $\mathbf{R}$, $\mathbf{S}$, and $\mathbf{T}$ are included in $\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G)$, then $\mathbf{An}(\mathbf{R} \cup \mathbf{S} \cup \mathbf{T},G)$ is included in $\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G)$. Any pair of vertices A and B adjacent in $G^M(\mathbf{An}(\mathbf{R} \cup \mathbf{S} \cup \mathbf{T},G))$ is also adjacent in $G^M(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G))$ because $G(\mathbf{An}(\mathbf{R} \cup \mathbf{S} \cup \mathbf{T},G))$ is a subgraph of $G(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G))$. Hence $G^M(\mathbf{An}(\mathbf{R} \cup \mathbf{S} \cup \mathbf{T},G))$ is a subgraph of $G^M(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G))$. It follows that if $\mathbf{R}$ and $\mathbf{S}$ are separated given $\mathbf{T}$ in $G^M(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G))$ they are also separated in $G^M(\mathbf{An}(\mathbf{R} \cup \mathbf{S} \cup \mathbf{T},G))$. But by the global directed Markov property, if $\mathbf{R}$ and $\mathbf{S}$ are separated given $\mathbf{T}$ in $G^M(\mathbf{An}(\mathbf{R} \cup \mathbf{S} \cup \mathbf{T},G))$ then $\mathbf{R}$ and $\mathbf{S}$ are independent given $\mathbf{T}$. It follows from the Hammersly-Clifford Theorem that the density function $f(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G))$ can be factored as

$$f(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G)) = \prod_{V \in \mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z},G)} g_V(V,\mathbf{Parents}(V,G))$$

Equation 8

where each g_V is a positive function, i.e., the density function factors according to G. $\therefore$

Theorem 1: The probability measure P of a linear SEM L (recursive or non-recursive) with jointly independent error terms satisfies the global directed Markov property for the directed (cyclic or acyclic) graph G of L, i.e. if **X**, **Y**, and **Z** are disjoint sets of variables in G and **X** is d-separated from **Y** given **Z** in G, then **X** and **Y** are independent given **Z** in P.

Proof. Let $\mathbf{Err}(\mathbf{X})$ be the set of error terms corresponding to a set of non-error variables **X**. In order to distinguish the density function for **V** from the density function for the error terms we will use $f_{\mathbf{V}}$ to represent the density function (including marginal densities) for the latter and $f_{\mathbf{Err}}$ to represent the density function of the former. If **V** is the set of variables in G, then by hypothesis,

$$f_{\mathbf{Err}}(\mathbf{Err}(\mathbf{V})) = \prod_{\varepsilon \in \mathbf{Err}(\mathbf{V})} f_{\mathbf{Err}}(\varepsilon)$$

Equation 9

It is possible to integrate out the error terms not in $\mathbf{Err}(\mathbf{An}(\mathbf{X},G))$ and obtain

$$f_{\mathbf{Err}}(\mathbf{Err}(\mathbf{An}(\mathbf{X},G))) = \prod_{\varepsilon \in \mathbf{Err}(\mathbf{An}(\mathbf{X},G))} f_{\mathbf{Err}}(\varepsilon)$$

Equation 10

Because for each variable X in **V**, X is a linear function of its parents in G plus a unique error term ε_X, it follows that ε_X is a linear function g_X of X and the parents of X in G. Hence $\mathbf{Err}(\mathbf{An}(\mathbf{X},G))$ is a function of $\mathbf{An}(\mathbf{X},G)$. Following Haavelmo (1943) it is possible to derive the density function for the set of variables $\mathbf{An}(\mathbf{X},G)$ by replacing each ε_X in $f_{\mathbf{Err}}(\varepsilon_X)$ by $g_X(X,\mathbf{Parents}(X))$ and multiplying by the absolute value of the Jacobean:

$$f_{\mathbf{V}}(\mathbf{An}(\mathbf{X},G)) = \prod_{X \in \mathbf{An}(\mathbf{X},G)} f_{\mathbf{Err}}(g_X(X,\mathbf{Parents}(X,G))) \times |J|$$

Equation 11

where J is the Jacobian of the transformation. Because the transformation is linear, the Jacobian is a constant. All of the terms in the multiplication are non-negative because they are either a density function or a positive constant. It follows from lemma 1 that if **X** and **Y** are d-separated given **Z** then **X** and **Y** are independent given **Z**. $\therefore$

Lemma 4: In a directed graph G with vertices **V**, if **X**, **Y**, and **Z** are disjoint subsets of **V**, and **X** is d-connected to **Y** given **Z** in G, then **X** is d-connected to **Y** given **Z** in an acyclic directed subgraph of G.

Proof. I will use the sense of d-connection defined in Pearl (1988) which Lauritzen et. al. (1990) proved equivalent to their sense of d-connection for acyclic graphs. The proof of the equivalence given by Lauritzen et. al can easily be extended to cyclic graphs. Vertex X is a **collider** on an acyclic undirected path U in directed graph G if and only if there are two adjacent edges on U directed into X. According to Pearl's definition, for three disjoint sets **X**, **Y**, and **Z**, **X** and **Y** are **d-separated** given **Z** in G if and only if there is no acyclic undirected path U from a member of **X** to a member of **Y** such that every non-collider on U is not in **Z**, and every collider on U has a descendant in **Z**. For three disjoint sets **X**, **Y**, and **Z**, **X** and **Y** are **d-connected** given **Z** in G if and only if **X** and **Y** are not **d-separated** given **Z**

Suppose that U is an undirected path that d-connects X and Y given $\mathbf{Z}$, and C is a collider on U. Let *length*(C,$\mathbf{Z}$) be 0 if C is a member of $\mathbf{Z}$, or the length of a shortest directed path from C to a member of $\mathbf{Z}$. Let *size*(U) equal the number of collider on U plus the sum over all colliders C on U of *length*(C,$\mathbf{Z}$). U is a **minimal path** that d-connects X and Y given $\mathbf{Z}$, if there is no other path U' that d-connects X and Y given $\mathbf{Z}$ such that *size*(U') < *size*(U). If there is a path that d-connects X and Y given $\mathbf{Z}$ there is at least one minimal path that d-connects X and Y given $\mathbf{Z}$.

Suppose $\mathbf{X}$ is d-connected to $\mathbf{Y}$ given $\mathbf{Z}$. Then for some X in $\mathbf{X}$ and Y in $\mathbf{Y}$, X is d-connected to Y given $\mathbf{Z}$ by some minimal path U in G. First I will show that no shortest acyclic directed path D_i from a collider C_i on U to a member of $\mathbf{Z}$ intersects U except at C_i. Suppose this is false. I will show that it follows that there is a path U' that d-connects X and Y given $\mathbf{Z}$ such that *size*(U') < *size*(U), contrary to the assumption that U is minimal. See figure 9.

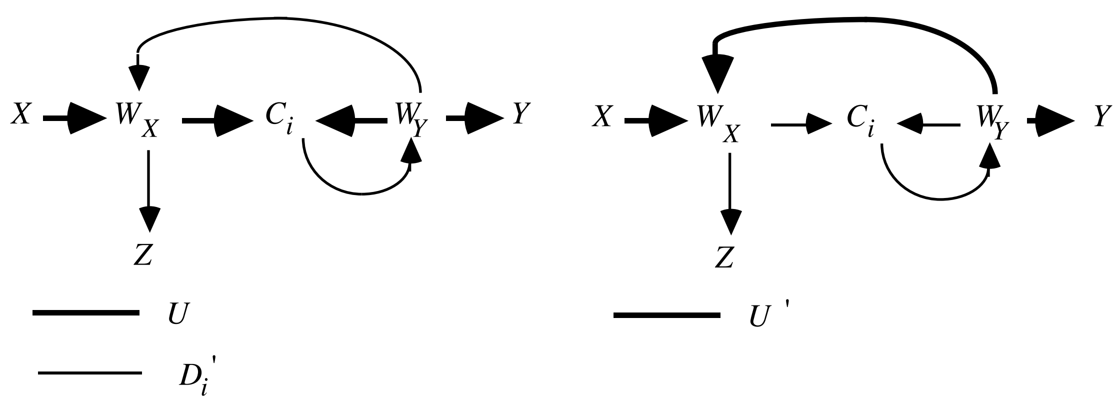


Figure 7

Form the path U' in the following way. If D_i intersects U at a vertex other than C_i then let W_X be the vertex on D_i and U that is closest to X on U, and W_Y be the vertex on D_i and U that is closest to Y on U. Suppose without loss of generality that W_X is after W_Y on D_i. Let U' be the concatenation of $U(X,W_X)$, $D_i(W_Y,W_X)$, and $U(W_Y,Y)$ (where $U(X,W_X)$ denotes the subpath of U between X and W_X.) It is now easy to show that U' d-connects X and Y given $\mathbf{Z}$, and $size(U') < size(U)$ because U' contains no more colliders than U and a shortest directed path from W_X to a member of $\mathbf{Z}$ is shorter than D_i. Hence U is not minimal, contrary to the assumption.

Next, I will show that if U is minimal, then it does not contain a pair of colliders C and D such that a shortest directed path from C to a member of $\mathbf{Z}$ intersects a shortest path from D to a member of $\mathbf{Z}$. Suppose this is false. See figure 10.

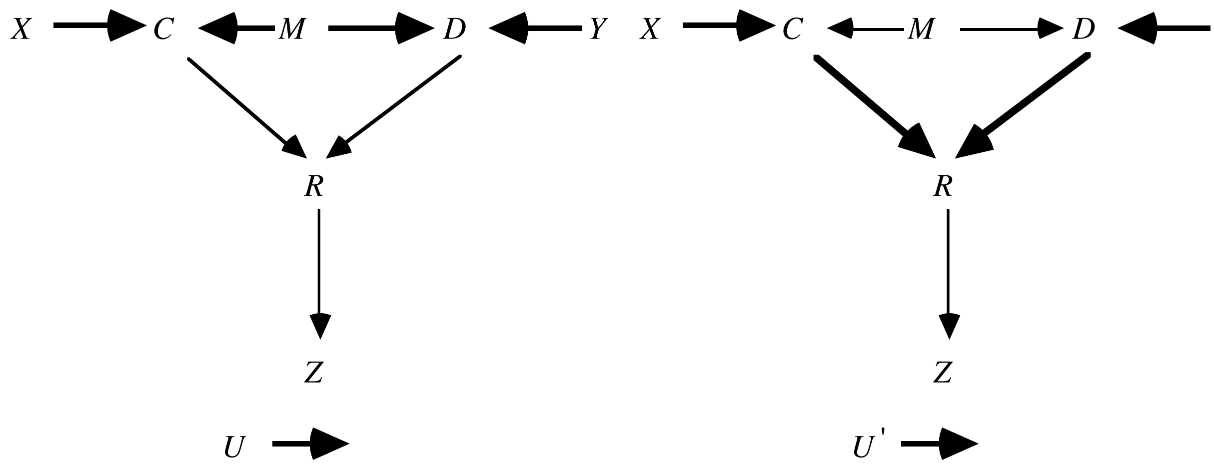


Figure 8

Let D_1 be a shortest directed acyclic path from C to a member of $\mathbf{Z}$ that intersects D_2, a shortest directed acyclic path from D to a member of $\mathbf{Z}$. Let the vertex on D_1 closest to C that is also on D_2 be R. Let U' be the concatenation of $U(X,C)$, $D_1(C,R)$, $D_2(C,R)$, and $U(Y,D)$. It is now easy to show that U' d-connects X and Y given $\mathbf{Z}$ and $size(U') < size(U)$ because U'

contains fewer colliders than U. Hence U is not minimal, contrary to the assumption.

For each collider C on a minimal path U that d-connects X and Y given $\mathbf{Z}$, a shortest directed path from C to a member of $\mathbf{Z}$ does not intersect U except at C, and does not intersect a shortest directed path from any other collider D to a member of $\mathbf{Z}$. It follows that the subgraph consisting of U and a shortest directed acyclic path from each collider on U to a member of $\mathbf{Z}$ is acyclic. $\therefore$

Theorem 2: In a linear SEM L with jointly independent error terms and directed (cyclic or acyclic) graph G containing disjoint sets of variables $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, if $\mathbf{X}$ is not d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ then L does not linearly entail that $\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$.

Proof. Suppose then that $\mathbf{X}$ is not d-separated from $\mathbf{Y}$ given $\mathbf{Z}$. By lemma 4, if $\mathbf{X}$ is not d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ in a cyclic graph G, then there is some acyclic subgraph G' of G in which $\mathbf{X}$ is not d-separated from $\mathbf{Y}$ given $\mathbf{Z}$. Geiger and Pearl (1988) have shown that if $\mathbf{X}$ is not d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ in a DAG, then there is some distribution represented by the DAG in which $\mathbf{X}$ is not independent of $\mathbf{Y}$ given $\mathbf{Z}$, and it has been shown (Spirtes, Glymour and Scheines, 1993) that there is in particular a linear normal distribution P in which $\mathbf{X}$ is not independent of $\mathbf{Y}$ given $\mathbf{Z}$. If P satisfies the global directed Markov property for G' it also satisfies it for G because every d-connecting path in G' is a d-connecting path in G. Hence there is some linear normal distribution represented by G in which $\mathbf{X}$ is not independent of $\mathbf{Y}$ given $\mathbf{Z}$. $\therefore$

Theorem 3: In a linear SEM L with jointly independent error terms and (cyclic or acyclic) directed graph G containing X, Y and $\mathbf{Z}$, where $X \neq Y$

and **Z** does not contain X or Y, X is d-separated from Y given **Z** if and only L linearly entails that $\rho_{XY.\mathbf{Z}} = 0$.

Proof. (This proof for cyclic or acyclic graphs is based on the proof for acyclic graphs in Verma and Pearl, 1990.) Let L' be a linear SEM with the same directed graph G and that is the same as L except that the exogenous variables are jointly normally distributed with the same variances as the corresponding variables in L. By theorems 1 and 2, L' linearly entails that X is independent of Y given **Z** if and only if X is d-separated from Y given **Z** in G. Hence for all values of the linear coefficients and all joint normal distributions over the exogenous variables in which the exogenous variables have positive variance and $\rho_{XY.\mathbf{Z}}$ exists, $\rho_{XY.\mathbf{Z}} = 0$ if and only if X is d-separated from Y given **Z** in G. Because the value of a partial correlation in a linear SEM depends only on the values of the linear coefficients and the variances of the exogenous variables, L' linearly entails $\rho_{XY.\mathbf{Z}} = 0$ if and only if X is d-separated from Y given **Z** in G and hence L also linearly entails that $\rho_{XY.\mathbf{Z}} = 0$ if and only if X is d-separated from Y given **Z** in G. $\therefore$

Note that for an SEM with graph G, if $V \neq X$, then $\partial \varepsilon_V / \partial X$ is non-zero only if there is an edge from X to V in G (because ε_V is a function only of V and V's parents in G.) Associate with each non-zero partial derivative $\partial \varepsilon_V / \partial X$ the edge from X to V in G. A product of partial derivatives form a **loop** in G if and only if the corresponding edges form a cycle in G. Two loops **intersect** if and only if their corresponding cycles intersect.

Let $J_{\mathbf{Err}(\mathbf{V})\text{->}\mathbf{V}}$ be the Jacobean of the transformation from **Err**(**V**) to **V**, and $J_{\mathbf{V}\text{->}\mathbf{Err}(\mathbf{V})}$ be the Jacobean of the transformation from **V** to **Err**(**V**). A product of partial derivatives S occurring in a term T in $J_{\mathbf{Err(V)\text{->}V}}$ is **minimally sufficient** in T if for each variable occurring in S, all of its occurrences in T are in S, and no subset of S has this property. For example, in

$$\frac{\partial\varepsilon_W}{\partial X}\times\frac{\partial\varepsilon_X}{\partial Y}\times\frac{\partial\varepsilon_Y}{\partial W}\times\frac{\partial\varepsilon_U}{\partial U}\times\frac{\partial\varepsilon_V}{\partial V}$$

Equation 12

the three minimally sufficient products are

$$\frac{\partial\varepsilon_W}{\partial X}\times\frac{\partial\varepsilon_X}{\partial Y}\times\frac{\partial\varepsilon_Y}{\partial W},\quad \frac{\partial\varepsilon_U}{\partial U},\ \text{and}\ \frac{\partial\varepsilon_V}{\partial V}$$

Equation 13

$J_{\mathbf{Err(V)\text{->}V}}$ is equal to $1/J_{\mathbf{V\text{->}Err(V)}}$, but it turns out to simplify the proofs if at intermediate stages we work with $J_{\mathbf{V\text{->}Err(V)}}$ than if we work with $J_{\mathbf{Err(V)\text{->}V}}$. $J_{\mathbf{V\text{->}Err(V)}}$ is the determinant of a matrix in which the element in the i^{th} row and j^{th} column is $\partial\varepsilon_{V_i} / \partial V_j$.

X is an **ancestral set** for a directed graph G with vertices **V** if and only if **X** = **An**(**Y**,G) for some **Y** included in **V**.

Theorem 4: In an acylic graph G containing disjoint sets of variables **X**, **Y** and **Z**, G pseudo-indeterministically entails that **X** is d-separated from **Y** given **Z** if and only L entails that **X** is independent of **Y** given **Z**.

Proof. The first part of the proof is essentially the same as that of Theorems 1 and 2, and shows that

$$f(\mathbf{An}(\mathbf{X},G)) = \prod_{X \in \mathbf{An}(\mathbf{X},G)} f(g_X(X,\mathbf{Parents}(X,G)) \times |J|$$

Equation 14

In an acyclic graph, the Jacobian of the transformation is a single term consisting of the product of the terms along the diagonal of the transformation matrix:

$$J = \prod_{V \in \mathbf{An}(\mathbf{X},G)} \frac{\partial \varepsilon_V}{\partial V} = \prod_{V \in \mathbf{An}(\mathbf{X},G)} m_V(V,\mathbf{Parents}(V,G))$$

Equation 15

(This is because for an acyclic graph the transformation matrix can be arranged so that it is lower triangular.) Each term $\partial\varepsilon_V / \partial V$ is some function m_V of V and its parents, because ε_V is a function of V and its parents. Hence by lemma 1, if **X** and **Y** are d-separated given **Z**, then **X** and **Y** are independent given **Z**.

Suppose that **X** and **Y** are not d-separated given **Z**. Then by Theorem 2, there is a linear SEM in which **X** and **Y** are not independent given **Z**. Since a linear SEM is a special case of an SEM, there is an SEM in which **X** and **Y** are not independent given **Z**. $\therefore$

Lemma 5: In an SEM with directed graph G with vertices **V**, if **X** is an ancestral set for G, then each minimally sufficient product of terms occurring T of $J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})}$ that is non-zero is either a loop in G(**X**), or $\partial\varepsilon_V / \partial V$ for V in **X**.

Proof. Each term in $J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})}$ is a product of partial derivatives in the transformation matrix, one from each row, and one from each column, times a variable that is either equal to 1 or -1. Hence each variable in **X** appears exactly once in the numerator of some partial derivative in the term, and exactly once in the denominator of some partial derivative in the term. If $\partial\varepsilon_V/\partial V$ occurs in *T*, it is minimally sufficient.

Suppose then that some minimally sufficient product of partial derivatives *S* occurring in *T* is not equal to $\partial\varepsilon_V/\partial V$ for any *V* in **X**. Then *S* does not contain $\partial\varepsilon_V/\partial V$ for *V* in **X**, because otherwise it would not be minimally sufficient. Hence each partial derivative in *S* is of the form $\partial\varepsilon_V/\partial Y$ where $V \neq Y$. Such a term is non-zero only if there is an edge from *Y* to *V* in *G*. Because *V* and *Y* are both in ancestral set **X**, if there is an edge from *Y* to *V* in *G*, then there is an edge from *Y* to *V* in *G*(**X**). Since all of the occurrences of the variables in *S* are in *S*, each variable occurs once in the numerator and once in the denominator of a partial derivative in *S*; so in *G*(**X**) there is a path in which all of the variables in *S* occur once at the head of an edge and once at the tail. It follows that there is a cycle in *G*(**X**) that corresponds to the product of partial derivatives in *S*. $\therefore$

A **cycleset** is a set of non-intersecting cycles. Let **Cycleset**(*G*) be the set of all cyclesets in *G*. Let **Vertices**(**C**) be the set of vertices occuring in a cycleset **C**.

Lemma 6: In an SEM with directed graph *G* with vertices **V**, if **X** is an ancestral set for *G* then

$$J_{\mathbf{X}->\mathbf{Err}(\mathbf{X})} = \sum_{\mathbf{C}\in\mathbf{Cycleset}(G(\mathbf{X}))} d(\mathbf{C}) \times \left(\prod_{<W,Y>\,\in\,\mathbf{C}} \frac{\partial\varepsilon_Y}{\partial W} \right) \left(\prod_{V\,\in\mathbf{X}\setminus\,\mathbf{Vertices}(\mathbf{C})} \frac{\partial\varepsilon_V}{\partial V} \right)$$

Equation 16

where in each term $d(\mathbf{C})$ is either equal to either 1 or -1.

Proof. For each $\mathbf{C}$ that is a set of loops in $G(\mathbf{X})$ that do not intersect, let

$$g(\mathbf{C}) = d(\mathbf{C}) \times \left(\prod_{<W,Y> \in \mathbf{C}} \frac{\partial \varepsilon_Y}{\partial W} \right) \left(\prod_{V \in \mathbf{X} \setminus \textbf{Vertices}(\mathbf{C})} \frac{\partial \varepsilon_V}{\partial V} \right)$$

Equation 17

I will show that that for each cycleset $\mathbf{C}$ in $G(\mathbf{X})$ that $g(\mathbf{C})$ is a term in $J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})}$, every non-zero term in $J_{\mathbf{V}\text{->}\mathbf{Err}(\mathbf{V})}$ is equal to $g(\mathbf{C})$ for some cycleset $\mathbf{C}$ in $G(\mathbf{X})$, and if $\mathbf{C_1}$ and $\mathbf{C_2}$ are distinct cyclesets then $g(\mathbf{C_1}) \neq g(\mathbf{C_2})$.

For each $\mathbf{C}$, a variable occurs once in the denominator of a partial derivative in $g(\mathbf{C})$, and once in the numerator of partial derivative in $g(\mathbf{C})$. Hence one partial derivative from each row and each column of the transformation matrix occurs in $g(\mathbf{C})$. But every product of partial derivatives which consists of one partial derivative from each column and each row of the transformation matrix is a term in $J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})}$ (because $J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})}$ is the determinant of the transformation matrix). Hence $g(\mathbf{C})$ is a term in $J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})}$.

Let $\mathbf{C_1}$ be a set of cycles such that no pair of cycles in $\mathbf{C_1}$ intersect, and similarly for $\mathbf{C_2}$. Suppose that $\mathbf{C_1} \neq \mathbf{C_2}$; then $g(\mathbf{C_1}) \neq g(\mathbf{C_2})$ unless there is some way to rearrange the edges in $\mathbf{C_1}$ into the cycles in $\mathbf{C_2}$. But because no pair of cycles in $\mathbf{C_1}$ intersect, each vertex that appears in $\mathbf{C_1}$ occurs in exactly two edges, once as the head, and once as the tail. Hence the edges in $\mathbf{C_1}$ cannot be rearranged into the loops in $\mathbf{C_2}$, and $g(\mathbf{C_1}) \neq g(\mathbf{C_2})$.

By lemma 5, each minimally sufficient product of terms occurring in T of $J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})}$ is either a loop or $\partial\varepsilon_V/\partial V$ for V in $\mathbf{X}$. By definition, the variables in distinct minimally sufficient product of terms do not overlap. Hence T consists of a product of non-intersecting minimally sufficient products of terms. Hence, for every non-zero term T in $J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})}$ there is a cycleset $\mathbf{C}$ such that $T = g(\mathbf{C})$. $\therefore$

Let **Cyclegroup**(G) be the set of all cyclegroups in G. If $\mathbf{C}$ is a cyclegroup in G, let **Cycleset**($\mathbf{C}$) be the set of all cyclesets included in $\mathbf{C}$.

Lemma 7: In an SEM with directed graph G, if $\mathbf{X}$ is an ancestral set for G, then

$$J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})} = \left(\prod_{V \notin \mathbf{Cycles}(G(\mathbf{X}))} \frac{\partial\varepsilon_V}{\partial V}\right) \times \left(\prod_{\mathbf{C} \in \mathbf{Cyclegroup}(G(\mathbf{X}))} \left(\sum_{\mathbf{D} \in \mathbf{Cycleset}(\mathbf{C})} \left(d(\mathbf{D}) \times \left(\prod_{V \in \mathbf{C} \setminus \mathbf{D}} \frac{\partial\varepsilon_V}{\partial V}\right)\left(\prod_{<W,Y> \in \mathbf{D}} \frac{\partial\varepsilon_Y}{\partial W}\right)\right)\right)\right)$$

Equation 18

where $d(\mathbf{D})$ is a variable equal either to 1 or -1.

Proof. By lemma 6,

$$J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})} = \sum_{\mathbf{C} \in \mathbf{Cycleset}(G(\mathbf{X}))} d(\mathbf{C}) \times \left(\prod_{<W,Y> \in \mathbf{C}} \frac{\partial\varepsilon_Y}{\partial W}\right)\left(\prod_{V \in \mathbf{X} \setminus \mathbf{Vertices}(\mathbf{C})} \frac{\partial\varepsilon_V}{\partial V}\right)$$

Equation 19

If V is not in a cycle in $G(\mathbf{X})$ then it is not in any cycleset. Hence, by lemma 5, every occurrence of V in each non-zero term in $J_{\mathbf{X\text{->}Err(X)}}$ is of the form $\partial\varepsilon_V/\partial V$. Hence it is possible to factor

$$\left(\prod_{V \notin \mathbf{Cycles}(G(\mathbf{X}))} \frac{\partial \varepsilon_V}{\partial V}\right)$$

Equation 20

from each non-zero term in the previous equation, because if V does not occur in a cycle, it does not occur in any cycleset. This leads to

$$J_{\mathbf{X\text{->}Err(X)}} =$$

$$\left(\prod_{V \notin \mathbf{Cycles}(G(\mathbf{X}))} \frac{\partial \varepsilon_V}{\partial V}\right)\left(\sum_{\mathbf{C} \in \mathbf{Cycleset}(G(\mathbf{X}))} d(\mathbf{C}) \times \left(\prod_{<W,Y> \in \mathbf{C}} \frac{\partial \varepsilon_Y}{\partial W}\right)\left(\prod_{V \in \mathbf{Cycles}(G(\mathbf{X})) \setminus \mathbf{Vertices}(\mathbf{C})} \frac{\partial \varepsilon_V}{\partial V}\right)\right)$$

Equation 21

The set of cyclegroups in G partitions the set of cycles in G. Hence each cycleset in G can be partitioned into a set of cyclesets, where each cycleset contains only cycles from the same cyclegroup. In addition, suppose that **C** is a set of cyclesets, where each cycleset in **C** contains cycles from only one cyclegroup, and each pair of cyclesets in **C** contains cycles from different cyclegroups. Then the union of any two cyclesets in **C** is also a cycleset. Hence

$$\sum_{\mathbf{C}\in\mathbf{Cycleset}(G(\mathbf{X}))} d(\mathbf{D})\times\left(\prod_{<W,Y>\in\mathbf{C}}\frac{\partial\varepsilon_Y}{\partial W}\right)\left(\prod_{V\in\mathbf{Cycles}(G(\mathbf{X}))\setminus\mathbf{C}}\frac{\partial\varepsilon_V}{\partial V}\right)=$$

$$\prod_{\mathbf{C}\in\mathbf{Cyclegroup}(G(\mathbf{X}))}\left(\sum_{\mathbf{D}\in\mathbf{Cycleset}(\mathbf{C})} d(\mathbf{D})\times\left(\prod_{V\in\mathbf{C}\setminus\mathbf{D}}\frac{\partial\varepsilon_V}{\partial V}\right)\left(\prod_{<W,Y>\in\mathbf{D}}\frac{\partial\varepsilon_Y}{\partial W}\right)\right)$$

Equation 22

$\therefore$

Lemma 8: For an SEM with directed graph G with vertices $\mathbf{V}$, if $\mathbf{X}$ is an ancestral set for G then

$$f_{\mathbf{V}}(\mathbf{X})=$$

$$\left(\prod_{V\notin\mathbf{Cycles}(G(\mathbf{X}))} g_V(V,\mathbf{Parents}(V,G(\mathbf{X}))\right)\times\left(\prod_{\mathbf{C}\in\mathbf{Cyclegroup}(G(\mathbf{X}))} g_{\mathbf{C}}(\mathbf{C},\mathbf{Parents}(\mathbf{C},G(\mathbf{X}))\right)$$

Equation 23

where each g is a non-negative function.

Proof. The transformed density function of $\mathbf{Err}(\mathbf{X})$ is equal to

$$\left(\prod_{X\in\mathbf{X}} f_{\mathbf{Err}}(h_X(X,\mathbf{Parents}(X,G(\mathbf{X})))\right)\times\left|J_{\mathbf{Err}(\mathbf{X})->\mathbf{X}}\right|$$

Equation 24

where $\varepsilon_X = h_X(X,\mathbf{Parents}(X,G(\mathbf{X}))$. By lemma 7,

$$J_{\mathbf{X}->\mathbf{Err}(\mathbf{X})} =$$
$$\left(\prod_{V \notin \mathbf{Cycles}(G(\mathbf{X}))} \frac{\partial \varepsilon_V}{\partial V}\right) \times$$
$$\left(\prod_{\mathbf{C} \in \mathbf{Cyclegroup}(G(\mathbf{X}))} \left(\sum_{\mathbf{D} \in \mathbf{Cycleset}(\mathbf{C})} \left(d(\mathbf{D}) \times \left(\prod_{V \in \mathbf{C} \setminus \mathbf{D}} \frac{\partial \varepsilon_V}{\partial V}\right)\left(\prod_{<W,Y> \in \mathbf{D}} \frac{\partial \varepsilon_Y}{\partial W}\right)\right)\right)\right)$$

Equation 25

Each term in

$$\left(\prod_{V \notin \mathbf{Cycles}(G(\mathbf{X}))} \frac{\partial \varepsilon_V}{\partial V}\right)$$

Equation 26

is a function of V and **Parents**(V,G(**X**)). Each term in

$$\prod_{\mathbf{C} \in \mathbf{Cyclegroup}(G(\mathbf{X}))} \left(\sum_{\mathbf{D} \in \mathbf{Cycleset}(\mathbf{C})} \left(d(\mathbf{D}) \times \left(\prod_{V \in \mathbf{C} \setminus \mathbf{D}} \frac{\partial \varepsilon_V}{\partial V}\right)\left(\prod_{<W,Y> \in \mathbf{D}} \frac{\partial \varepsilon_Y}{\partial W}\right)\right)\right)$$

Equation 27

contains only error terms associated with variables in **C**, and hence is a function of **C** and **Parents**(**C**,G(**X**)). Hence, there exist functions m_V such that

$$J_{\mathbf{X}->\mathbf{Err}(\mathbf{X})} =$$

$$\left(\prod_{V \notin \mathbf{Cycles}(G(\mathbf{X}))} m_V(V, \mathbf{Parents}(V, G(\mathbf{X}))\right) \times \left(\prod_{C \in \mathbf{Cyclegroup}(G(\mathbf{X}))} m_{\mathbf{C}}(\mathbf{C}, \mathbf{Parents}(\mathbf{C}, G(\mathbf{X})))\right)$$

Equation 28

Because $J_{\mathbf{Err}(\mathbf{X})\text{->}\mathbf{X}} = 1/J_{\mathbf{X}\text{->}\mathbf{Err}(\mathbf{X})}$, $J_{\mathbf{Err}(\mathbf{X})\text{->}\mathbf{X}}$ can also be factored as in Equation 28. Combining this with Equation 24, there exist functions non-negative functions g_V such that

$$f_{\mathbf{V}}(\mathbf{X}) = \left(\prod_{V \notin \mathbf{Cycles}(G(\mathbf{X}))} g_V(V, \mathbf{Parents}(V, G(\mathbf{X})) \right) \times \left(\prod_{\mathbf{C} \in \mathbf{Cyclegroup}(G(\mathbf{X}))} g_{\mathbf{C}}(\mathbf{C}, \mathbf{Parents}(\mathbf{C}, G(\mathbf{X}))) \right)$$

Equation 29

∴

Theorem 5: In an SEM L with directed (cyclic or acyclic) graph G with vertices **V** and collapsed graph G' containing disjoint sets of variables **X**, **Y** and **Z**, if **X** is d-separated from **Y** given **Z** in G' then **X** is independent of **Y** given **Z**.

Proof. By lemma 8

$$f_{\mathbf{V}}(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, G)) =$$

$$\left(\prod_{V \notin \mathbf{Cycles}(G(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, G)))} g_V(V, \mathbf{Parents}(V, G(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, G)) \right) \times$$

$$\left(\prod_{\mathbf{C} \in \mathbf{Cyclegroup}(G(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, G)))} g_{\mathbf{C}}(\mathbf{C}, \mathbf{Parents}(\mathbf{C}, G(\mathbf{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, G)))) \right)$$

Equation 30

This is a factorization according to the collapsed graph G', and hence by lemma 1, for three disjoint sets of variables **X**, **Y** and **Z**, if **X** and **Y** are d-separated given **Z** in G', then **X** and **Y** are independent given **Z**. ∴