Calculating Tetrad Constraints Implied by Directed Acyclic Graphs

by

Peter Spirtes

October 1990

Report CMU-PHIL-16

Philosophy
Methodology
Logic

Pittsburgh, Pennsylvania 15213-3890

1. Introduction

1.1. The Problem

In many disciplines, it is normally difficult or impossible (either for practical or ethical reasons) to carry out experiments, but it is possible to gather statistical data. Drawing causal conclusions from statistical data without experimentation is very difficult in part because of the enormous number of different possible causal models that can be constructed among a set of variables. Consider a set of n variables. Among each pair of variables A and B there are four possible causal relations: A causes B, B causes A, A causes B and B causes A, and neither causes the other. Hence there are $4^{\binom{n}{2}}$ possible causal models among n variables. For six variables, there are 1,073,741,824 different possible causal models. For twelve variables there are approximately 5.4 x 10^{39} different possible causal models.

Of course, often many of the possible causal models for a set of variables can be eliminated on substantive grounds. But in the social sciences, unlike the natural sciences, there often is no well established general theory that severely limits the number of causal models compatible with what we know about a domain. The sort of fragmentary knowledge about a domain available in the social sciences can drastically limit the number of plausible alternative causal explanations, but still leave an enormous number of possible causal models compatible with all prior substantive knowledge. Suppose (as is usually *not* the case), that the variables are totally ordered by time, so that for each pair of variables the only possibilities were that the earlier one caused the later one, or it didn't. In that case, for six variables there are still 32,768 causal models compatible with the time order. For twelve variables, there are approximately 7.4 x 10^{19} causal models compatible with the time order. Even in those cases where an initial model is already known or hypothesized, and the researcher is simply searching for *additional* causal relations among the variables, the number of possibilities can easily run into the tens of thousands.

Until recently, there have been few tools that would aid a researcher in systematically searching these enormous spaces of possibilities. Factor analysis provides a search procedure with well-known limitations. Several modules (in LISREL VI and EQS) have recently been offered for searching portions of the space of possibilities, but there are sound theoretical reasons for believing that they do not implement reliable search procedures, and recent Monte Carlo simulation tests confirm this (See [5]).

TETRAD II is a computer program that uses graph-theoretic methods to aid in systematically and reliably searching large spaces of a particular kind of causal model (Statistical Linear Causal Theoriesor LCTs). The TETRAD II procedures exploit a directed graph representation of conditional independence relations in linear probability distributions. In **Discovering Causal Structure** and subsequent papers, Glymour, Scheines, Spirtes and Kelly proved that various constraints on the covariance matrix implied by a given linear probability distribution corresponded to easily calculated features of their directed graph representation.

Recent work by Pearl and his colleagues has examined the relation between directed acyclic graphs (DAGs) and conditional independence relations in a broader class of probability distributions. This work has show how to characterize the conditional independence relations entailed by a certain class of probabilistic models in graph-theoretic terms. I will show that Pearl's graph-theoretic characterization of conditional independence applies to LCTs. This has the following useful consequences:

1. It shows how to easily calculate from the directed graph representation all and only the partial correlations of all order (among non-error variables) that are implied to vanish by the graph structure.

2. There are efficient algorithms for constructing DAGs that represent the causal structureof a process that generated a probability distribution P (as long as all common cause of the measured variables are themselves measured) that take as input the conditional independence relations true of P. These algorithms are directly applicable to LCTs using vanishing partial correlations as input.

3. In LCT's, vanishing tetrad differences (useful for the construction of causal models when latent variables are present) are equivalent to a certain set of vanishing partial correlations.

4. A sufficient condition for a set of vanishing tetrad differences to require the introduction of latent variables in order to be representable by the graph of a LCT can be given in terms of sets of vanishing partial correlations.

I will not pursue the second point in this technical report. The algorithm is described in [6].

1.2. LCTS and Graphs

The domain of application of TETRAD II is a class of statistical models that we call LCTs (Statistical Linear Causal Theories). Special cases of LCTs include regression models, path analytic models, and factor analytic models. Such models are used in psychology, sociology, epidemiology, political science, biology, engineering and educational research.

Basically, a LCT consists of a set of random variables over a probability space, a graph such that there is an edge from random variable A to random variable B iff A is a direct cause of B, a joint probability distribution for the exogenous random variables such that the exogenous variables are pairwise independent, and for each endogenous random variable a linear equation relating it to its direct causal parents.

The linear equations are expressed in a canonical form called "structural equations" in which one variable is set equal to a linear combination of all of its *direct* causes. One example of a set of structural equations relating T, v, w, x, y, and z, is:

$$v = aT + e_v$$

$$w = bT + fv + gy + e_w$$
$$x = cT + e_x$$
$$y = dT + e_y$$
$$z = eT + e_z$$

In this example, the variables v - z are measured variables, but T is a "latent construct" that is not directly measured. Each structural equation includes a unique unmeasured random variable e_i, called the "error term" . It is usually assumed that no variance is zero.

The model can be parameterized by a vector θ that includes the non-zero linear coefficients and the variance/covariance matrix for the exogenous variables. A variable is **exogenous** just in case it is a cause and not an effect. All disturbance terms are exogenous.

The directed graph G in a LCT is just a set of random variables V and a binary relation E on V. The E relation may be represented by a set of ordered pairs. In the pictures used to represent directed graphs, an arrow (a directed edge) from A to B represents the ordered pair <A,B>. For example, suppose the following set of ordered pairs represents our directed graph.

$$\{\langle T,v\rangle,\langle T,w\rangle,\langle T,x\rangle,\langle T,y\rangle,\langle T,z\rangle,\langle v,w\rangle,\langle y,w\rangle\langle e_v,v\rangle,\langle e_w,w\rangle,\langle e_x,x\rangle,\langle e_y,y\rangle,\langle e_z,z\rangle\}$$

The picture that corresponds to this set is shown in Fig. 1. (By convention, a random variables whose name begins with a capital letter represents an unmeasured or "latent" random variable, and a random variable which is not an error variable whose name begins with a lower case letter represents a measured random variable.)

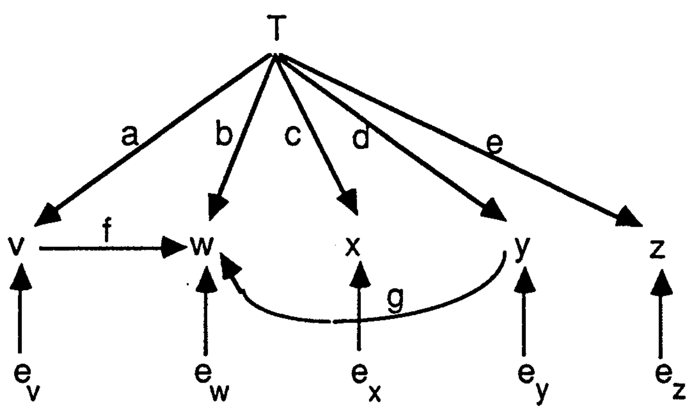


Fig. 1.2.1

In the linear equations representing a LCT, each dependent variable is a linear function of its immediate causes. Since there is an edge from A to B iff A is an immediate cause of B we can label the edge from A to B by the coefficient of A in the linear equation for B. The label of a path is simply the product of the labels of edges in the path or trek. The labels of an edge from an error variable by convention are set to 1.

Using these conventions, we can simply read the structural equations in a quantitative causal model from a labeled directed graph. Fig. 1.2.1 illustrates this point.

One fact that is crucial to the operation of TETRAD II is that the graph that is not labelled by the linear coefficients by itself, quite apart from the actual linear coefficients or the distribution of the independent variables, can imply constraints on the population covariance matrix.

I will consider constraints in the form of vanishing tetrad differences. For any four distinct variables, there are three possible vanishing tetrad differences, any two of which are independent. For x, y, z, and w they are:

$$\gamma_{wx}\gamma_{yz} - \gamma_{wy}\gamma_{xz} = 0$$
$$\gamma_{wy}\gamma_{xz} - \gamma_{wz}\gamma_{yx} = 0$$
$$\gamma_{wz}\gamma_{yx} - \gamma_{wx}\gamma_{yz} = 0$$

where γ_{wx} represents the covariance between w and x. ($\sigma^2{}_x$ represents the variance of x, and ρ_{wx} represents the correlation of w and x).

The model in Fig. 1.2.1 implies the following four equations.

$$\gamma_{vw} = (ab + adg)\sigma^2{}_T + f(a^2\sigma^2{}_T + \sigma^2{}_{ev})$$
$$\gamma_{xy} = cd\sigma^2{}_T$$
$$\gamma_{wy} = (bd + afd)\sigma^2{}_T + g(d^2\sigma^2{}_T + \sigma^2{}_{ey}) + fad\sigma^2{}_T$$
$$\gamma_{vx} = ac\sigma^2{}_T$$

Thus, the tetrad difference $\gamma_{vw}\gamma_{xy} - \gamma_{wy}\gamma_{vx} = fcd\sigma^2{}_{ev}\sigma^2{}_T - acg\sigma^2{}_T\sigma^2{}_{ey}$.

This tetrad difference might or might not be equal to zero depending upon the values of the model's free parameters. For example, if $f = g = \sigma^2{}_{ev} = \sigma^2{}_{ey} = a = d = 1$, then the difference will be equal to zero. On the other hand, if $f = g = \sigma^2{}_{ev} = \sigma^2{}_{ey} = a = 1$, and $d = 1/2$, then the difference will not be equal to zero. (We assume that all variances are not equal to zero.)

Now consider the tetrad difference $\gamma_{vx}\gamma_{yz} - \gamma_{vy}\gamma_{xz}$.

$$\gamma_{vx} = ac\sigma^2{}_T$$
$$\gamma_{yz} = de\sigma^2{}_T$$
$$\gamma_{vy} = ad\sigma^2{}_T$$
$$\gamma_{xz} = ce\sigma^2{}_T$$

In this case $\gamma_{vx}\gamma_{yz} - \gamma_{vy}\gamma_{xz} = (ac)(de)\sigma^4{}_T - (ad)(ce)\sigma^4{}_T = 0$, regardless of the values of a, c, d, e, or $\sigma^4{}_T$. When a tetrad difference is constrained to vanish for all values of the linear coefficients and for all variances of the independent variables in model M, we say that the vanishing tetrad difference is **linearly implied by M**. A graph G completely determines for every LCT M containing G whether or not M strongly implies a vanishing tetrad difference T. If every LCT M containing G strongly implies T then we say that the tetrad difference is **linearly implied by G.**

In this paper, we will assume that all LCT's are acyclic, unless explicitly stated otherwise.

1.3. The Methodological Principles and LCTs

A LCT in which the form of the equations has been specified, but not the actual values of the linear coefficients or the variances of the independent variables is called a **causal model,** and the linear coefficients and variances of the independent variables are called **free parameters.**

In the context of LCTS, the methodological principles that TETRAD II is based upon are:

Explanatory Principle: Other things being equal, prefer a causal model that implies for all values of its free parameters a vanishing tetrad difference T judged to hold in the population over a causal model that implies T only for particular values of its free parameters.

Falsification Principle: Other things being equal, prefer a causal model that doesn't imply for all possible values of its free parameters a vanishing tetrad difference T judged not to hold in the population over a causal model that does imply T for all possible values of its free parameters.

Simplicity Principle: Other things being equal, prefer a causal model with more degrees of freedom.

TETRAD II uses these principles to evaluate causal models; in effect it search for the causal model that provides the best explanation of the vanishing tetrad differences (and other constraints on the correlation matrix that are judged to hold in the population. Then the causal model is given to an estimation package in order to estimate the values of the free parameters. In order to use these principles to help researchers search a large space of causal models three things are needed:

1. A fast algorithm for determining the vanishing tetrad constraints linearly implied by a graph.

2. A statistical test for judging when a given tetrad difference vanishes in the population.

3. An efficient search strategy.

Each of these components is described in [4].

The algorithm for determining the vanishing tetrad constraints linearly implied by a LCT is based upon the fact that the unlabeled graph of a LCT by itself determines when a vanishing tetrad difference is linearly implied. In order to state the graphical condition that determines whether or not a LCT strongly implies that a tetrad difference vanishes, several definitions are needed.

A **digraph <R,E>** is an ordered pair where R is a set, and E (the set of edges) is a set of ordered pairs of distinct members of R. A **directed path of length n** in a digraph <R,E> is an ordered n+1-tuple of vertices $<v_1, \ldots, v_{n+1}>$ where for $1 \leq i \leq n$, $<v_i, v_{i+1}>$ is an edge in E. An **empty path** is a path with exactly one vertex in the sequence. An **acyclic** path is a directed path that contains each vertex in R at most once. A trek t(i,j) between two distinct vertices v_i and v_j is a pair of acyclic paths from some vertex u to v_i and v_j respectively that intersect only at u. Given a trek t(i,j) between i and j, **i(t(i,j))** will denote the path in t(i,j) from the source of t(i,j)to i and **j(t(i,j))** will denote the path in t(i,j) from the source of t(i,j) to j.

The following simple test for the existence of vanishing tetrad differences was suggested by Clark Glymour. In order to state the theorem on which the test is based, the following definition is needed.

If all l(t(k,l)) and all j(t(i,j)) intersect at a vertex P, then P is an **lj(t(i,j),t(k,l)) choke point.** Similarly, if all l(t(k,l)) and all j(t(i,j)) intersect at a vertex P, and all l(t(i,l)) and all j(t(j,k)) also intersect at P, then P is a **lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point.**

Theorem A: In a LCT G, there exists an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point or an ik(t(i,j),t(k,l),t(i,l),t(j,k)) choke point iff G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$. (Theorem 2.7.2 in text.)

1.4. Directed Acyclic Graphs

The primary goal of Pearl's research was to find a graphical representation of probabilistic models that would allow for a fast algorithm for updating probabilities conditional on new evidence that could be implemented on a parallel processor computer. He uses directed acyclic graphs (called DAGs) to encode independence constraints among random variables with a joint distribution. (The details of this encoding are described in the following section.) He defines a graph-theoretic relation - d-separability - such that given three disjoint sets of random variables X, Y, and Z, if X is independent of Y conditional on Z in P just when X is d-separated from Y by Z in G, then **G is a perfect map of P**, or **P is perfectly represented by G.**

Unfortunately, it is difficult to characterize the set of probability distributions that can be perfectly represented by a DAG. There are no known complete first order axiomatizations of conditional independence in probability distributions, nor of the independence relations encodable in DAGs. There are several known restrictions on the class of probability distributions perfectly representable by DAGs. For example, a distribution P is perfectly representable by some DAG only if for each triple of sets of variables X,Y, and Z, X is conditionally independent of Y given Z iff every x in X and every y in Y is conditionally independent given Z; there are distributions in which this is not the case. (However we will prove in the next section that every LCT that does not imply a conditional independence relation by virtue of the values of its linear coefficients and variances of independent variables is perfectly represented by some DAG.) Unlike LCT's, DAGs can perfectly represent probability distributions in which the variables are not linearly related.

Theorem B: In a LCT S = <G, (Ω,f,P), X, L>, if x and z are distinct non-error variables, and Y is a set of non-error variables not including x and z, then Y d-separates x and z if and only if $\rho_{xz.Y}$ is linearly implied to vanish. (Theorem 2.2.4 in text.)

1.5. Introduction of Latent Variables

One of the most difficult questions facing someone constructing a statistical model is when to introduce latent (unmeasured) variables. Other things being equal, latent variables should not be introduced into a model unless there is no model containing just the observed variables that provides a good explanation of the observed constraints on the covariance matrix. Using the fact that Pearl's characterization of conditional independence can be applid to LCT's,We will state conditions under which a LCT containing only a given set of variables cannot both perfectly represent the a given set of conditional independence constraints and strongly imply a given set of vanishing tetrad differences. If S is the set of measured variables, and no LCT containing just the variables in S can perfectly represent the observed conditional independence constraints and strongly imply the tetrad differences judged to vanish in the population, but some model with latent variables can, then that is one justification for introducing latent variables.

Theorem C: A n acyclic LCT G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ only if there is a (possibly empty) set q of random variables in G such that for each probability distribution perfectly represented by G, either ρ_{ij} or $\rho_{kl}= 0$, and ρ_{il} or $\rho_{jk}= 0$, or there exists a (possible empty) set q such that $\rho_{ij.q} = \rho_{kl.q} = \rho_{il.q} = \rho_{jk.q} = 0$. (Theorem 2.7.3 in text.)

2. Appendix

In this sectionWe state a number of definitions that are needed in the following proofs.

2.1. LCTS

Definition 2.1.1: Given an ordered n-tuple $N = <c_1, \ldots, c_n>$, an object **o is in N** iff $o = c_i$ for some i between 1 and n inclusive. I shall also write that $o \varepsilon N$.

This notation is somewhat ambiguous since ε is also used to mean set membership, but the context will always make it clear which use of ε is intended.

Definition 2.1.2: A **digraph** is an ordered pair <R,E>, where R is a set of **vertices** and E is a set of **edges**. Each edge is an ordered pair of distinct elements of R. The first element in an edge is called the **tail**, and the second element is called the **head**. An edge with a tail v_i and a head v_j is an edge **from** v_i to v_j; it is also said that the edge is **out of** v_i and **into** v_j. v_i is **adjacent** to v_j iff there is an edge from v_i to v_j. (Note that I am treating adjacency as an asymmetric relation.) **Adj(i)** is the set of all variables adjacent to i. The **indegree** of a vertex v is equal to the number of distinct edges into v; the **outdegree** of a vertex v is equal to the number of distinct edges out of v.

Definition 2.1.3: A **directed path of length n** in a digraph <R,E> is an ordered n+1-tuple of vertices $<v_1, \ldots, v_{n+1}>$ where for $1 \leq i \leq n$, $<v_i, v_{i+1}>$ is an edge in E. The path is said to **contain** edge $<v_i, v_{i+1}>$. The first vertex in the path is called the **source** of the path; the last vertex in the path is called the **sink** of the path. The path is said to **connect** the source to the sink. Two paths **intersect** iff they have a a vertex in common; any such common vertex is a **point of intersection**. A **cycle** is a path of at least length 1 in which the source equals the sink. A path **contains a cycle** iff it has a subpath which is a cycle. An **open path** is a path with no cyclic subpaths. A digraph is **acyclic** if and only if every path in the graph is open. A path with one vertex is an **empty path**. If path p is equal to $<v_1, \ldots, v_n>$ and path q is equal to $<v_n, \ldots, v_{n+m}>$, then **the concatenation of p and q** is equal to $<v_1, \ldots, v_n, \ldots, v_{n+m}>$ and is denoted by p&q.

Note that empty paths are the only paths that contains no edges. Also the concatenation of p with an empty path is p, and the concatenation of an empty path with p is p. The single vertex in an empty path is both its source and its sink.

Definition 2.1.4: A trek t(i,j) between two distinct vertices v_i and v_j is a pair of open paths from some vertex u to v_i and v_j respectively that intersect only at u. The source of the paths in the trek is called the **source** of the trek. v_i and v_j are called the **termini of the trek**. Given a trek t(i,j) between i and j, **i(t(i,j))** will denote the path in t(i,j) from the source of t(i,j) to i and **j(t(i,j))** will denote the path in t(i,j) from the source of t(i,j) to j.

One of the paths in a trek may be an empty path. However, since the termini of a trek are distinct, only one path in a trek can be empty.

Definition 2.1.5: Let a stochastic linear causal theory (LCT)) be <<R,E>, (Ω,f,P), X, L> where

(Ω,f,P) is a probability space.

<R,E> is a labelled digraph. R is a set of random variables over (Ω,f,P). The variables in R have a joint distribution. Every variable in R has a non-zero variance, and partial correlations of all orders between all variables exist. E is a set of directed edges between variables in R.

X is a consistent set of independent homogeneous linear equations in random variables in R. For each i in R of positive indegree there is an equation in X of the form

$$i = \sum_{j \in Adj(i)} a_{ij} \, j$$

where each a_{ij} is a non-zero real number and each j is in R. This implies that each vertex i in R of positive indegree can be expressed as a linear function of all and only the vertices adjacent to i. There are no other equations in X. A non-zero value of a_{ij} is the **equation coefficient** of j in the equation for i.

If vertices (random variables) u and v are exogenous, then u and v are pairwise statistically independent.

L is a function with domain E such that L(e) = a_{ij} iff head(e) = i and tail(e) = j. L(e) will be called the **label of e**. By extension, the product of labels of edges in any path or trek p will be denoted by L(p), and L(p) will be called the **label of p**. The label of an empty path is fixed at 1. There is a subset of R called the **error variables**. Each error variable is of indegree 0 and outdegree 1. For every i in R of indegree ≠ 0 there is exactly one error variable with an edge into i. Each error variable is independent of the set of all of the other error variables. I assume that the partial correlations of all orders involving only non-error variables exist.

The pairwise independence of exogenous variables (which I will call marginal independence) does not imply that for any disjoint triple of sets of exogenous variables X, Y, and Z, that X and Y are independent conditional on Z. This latter property I will call joint independence of the exogenous variables. Note that the variance of any endogenous variable y conditional on any set of variables that does not contain the error variable of y is not equal to zero.

Intuitively, a LCF is a LCT in which the linear coefficients C and the variances of the exogenous variables V are real variables instead of constants.

Let a **stochastic linear causal form (LCF)** be <<R,E>, C, V, L> where

<R,E> is a labelled digraph. There is a subset of R called the **error variables**. Each error variable is of indegree 0 and outdegree 1. For every i in R of indegree ≠ 0 there is exactly one error variable with an edge into i.

C is the set of c_{ij}, where there is an edge from i to j in <R,E>, and c_{ij} is a variable that ranges over the real numbers.

V is the set of variables σ^2_i, where i is an exogenous variable in <R,E> and σ^2_i is a variable that ranges over the real numbers.

L is a function with domain E such that $L(e) = c_{ij}$ iff head(e) = i and tail(e) = j. L(e) will be called the **label of e.** By extension, the product of labels of edges in any path or trek p will be denoted by L(p), and L(p) will be called the **label of p**. The label of an empty path is fixed at 1.

A LCT S is an instance of a LCF F if and only if the graph of S is isomorphic to the graph of F.

A **polynomial in real variables $<x_1,...,x_n>$** is a finite sum of terms each of which consists of the product of a real number with a product of variables in $<x_1,...,x_n>$, each of which is raised to some positive integral power. A **polynomial equation in real variables $<x_1,...,x_n>$** is a polynomial in $<x_1,...,x_n>$ equated to zero.

A tuple of possible values $<a_1,...,a_n>$ for the real variables $<x_1,...,x_n>$ can be considered to be a point in the Euclidean space $\mathbb{R}^n$ with the usual Euclidean metric.

A statistic **X is equivalent to a polynomial in the coefficents and variances of the exogenous variables** if and only if for each LCF F = <<R,E>, C, V, L>, and in every LCT S that is an instance of F, there is a polynomial in the variables in C and V such that X is equal to the result of substituting the linear coefficients of S in as values for the corresponding variables in C, and the variances of the exogenous variables in S as values for the corresponding variables in V.

Definition 2.1.6: In a LCT S, a variable i is **independent** iff i has zero indegree (i.e. there are no edges directed into it). Otherwise i is **dependent.**

Note that the *property* of independence is completely distinct from the *relation* of statistical independence. The context will make clear in which of these senses the term is used.

In what follows, each reference to paths, treks, covariances, etc. will be assumed to refer to objects in some *acyclic* LCT S (i.e. the graph in the LCT is acyclic).

Definition 2.1.7: Given a joint distribution P over random variables x, y, z, and w, a **vanishing tetrad difference among four distinct variables x, y ,z, w** is any of

$$\gamma_{xy}\gamma_{zw} - \gamma_{xw}\gamma_{yz} = 0$$

$$\gamma_{xw}\gamma_{yz} - \gamma_{xz}\gamma_{yw} = 0$$

$$\gamma_{xz}\gamma_{yw} - \gamma_{xy}\gamma_{zw} = 0$$

or an equivalent equation, where γ_{xy} is the covariance of x and y. In what follows, I will consider only vanishing tetrad differences between non-error variables.

Definition 2.1.8: Given a graph G, **P(i,j)** is the set of all paths from i to j. **T(i,j)** is the set of all treks between i and j.

I will adopt the following conventions. "w.l.g." abbreviates "without loss of generality", "r.h.s." abbreviates "right hand side", and "l.h.s." abbreviates "left hand side". The set of independent variables in the theory is represented by "I". The sum over an empty set is equal to 0 and the product over an empty set is 1. Any lower case letter except t indexed by two variables i and j (such as p(i,j)) will represent a path in P(i,j). t(i,j) will represent a trek in T(i,j). s(t(i,j)) represents the source of the trek t(i,j).

Definition 2.1.9: The **distributed form** of an expression or equation E is the result of carrying out every multiplication, but no additions, subtractions, or divisions in E.

If there are no divisions in an equation then its distributed form is a sum of monomials. For example, the distributed form of the equation $u = (a + b)(c + d)v$ is $u = acv + adv + bcv + bdv$.

Definition 2.1.10: In a LCF T, if an expression is equal to ce, where c is a non-zero constant, and e is a product of equation coefficients raised to positive integral powers, then e is the **equation coefficient factor**(e.c.f.) of ce, and c is the **constant factor** of (c.f.) ce.

Definition 2.1.11: In a graph G, a path p of length n is an **initial segment** of path q of length m iff $m \geq n$, and for $1 \leq i \leq n+1$, the i^{th} vertex of q equals the i^{th} vertex of p.

A path p of length n is a **final segment** of path q of length m, iff $m \geq n$, and for $1 \leq i \leq n+1$, the i^{th} vertex of p equals the $(m-n+i)^{th}$ vertex of p.

A path p of length n is a **proper initial segment** of path q of length m iff p is an initial segment of q and $p \neq q$.

A path p of length n is a **proper final segment** of path q of length m iff p is a final segment of q and $p \neq q$.

I will now state a number of lemmas that will be used in subsequent proofs, but whose proofs are obvious.

Lemma 2.1.12: In a graph G, if p(u,i) is an acyclic path, and x is a vertex on p(u,i), then there is a unique initial segment of p(u,i) from u to x.

Henceforth, when there is a path p(u,i) in a proof, and a vertex x on p(u,i), p(u,x) will refer to the unique initial segment of p(u,i) from u to i, and p(x,i) will refer to the unique final segment of p(u,i) from x to i.

Definition 2.1.13: In a graph G, the **last point of intersection of p(u,i) with p(v,j)** is the last vertex on p(u,i) that is also on p(v,j).

Lemma 2.1.1: If S is an acyclic LCT, the last point of intersection of p(u,i) with p(v,j) equals the last point of intersection of p(v,j) with p(u,i).

(Note that lemma 2.1.1 is not true in cyclic LCTs.)

2.2. LCTS and DAGS

The following definitions are based on those in [3].

Definition 2.2.1: An **undirected path from v_1 to v_n** in an acyclic graph G = $<V,E>$ is an ordered n-tuple of vertices $<v_1,v_2,\ldots v_{n-1},v_n>$ such that each vertex occurs only once, and for each pair of vertices v_k and v_{k+1}, either the edge $<v_k,v_{k+1}>$ is in E, or the edge $<v_{k+1},v_k>$ is in E. If, in an undirected path U = $<v_1,v_2,\ldots v_{n-1},v_n>$ there is a vertex v_k where the edges $<v_{k-1},v_k>$ is in E, and $<v_{k+1},v_k>$ is in E is then **v_k contains a direction-reversal in U**.

Note that directed paths are a special case of undirected paths. If the order of the endpoints of an undirected path doesn't matter, I will also sometimes say that an undirected path is between v_1 and v_n.

Definition 2.2.2: A **set of vertices X is d-separated from a set of vertices Y by a set of vertices Z** in a graph G = $<V,E>$ iff there is no undirected path U between a variable in X and a variable in Y such that

a. for every vertex v_k that contains a direction-reversal in U, there is a directed path from v_k to some variable in Z, and

b. for every vertex v_k that does not contain a direction-reversal in U, v_k is not in Z.

Definition 2.2.3: In a probability distribution P over a given set of random variables R, if X, Y, and Z are disjoint sets of random variables in R, **$I(X,Y,Z)_P$** iff X is independent of Z given Y in P. If the context makes it clear what P is, I will just write I(X,Y,Z). An acyclic graph **G is an I-map of probability distribution P** iff for every X, Y, and Z that are disjoint sets of random variables in R, if X is d-separated from Y by Z in G then $I(X,Z,Y)_P$. An acyclic graph **G is a minimal I-map of probability distribution P** iff G is an I-map of P, and no subgraph of G is an I-map of P. An acyclic graph **G is a D-map of probability distribution P** iff for every X, Y, and Z that are disjoint sets of random variables in R, if X is not d-separated from Y by Z in G then $\sim I(X, Z,Y)_P$. An acyclic graph **G is a perfect map of probability distribution P** iff for every X, Y, and Z that are disjoint sets of random variables in R, X is d-separated from Y by Z in G iff $I(X, Z,Y)_P$. However, when I-map, D-map, and perfect map are applied to the graph in a LCT, the quantifiers in the definitions apply only to sets of *non-error* variables.

Lemma 2.2.1: In an acyclic graph G, every undirected path p = $<v_1,v_2,\ldots v_{n-1},v_n>$ without direction-reversals contains a vertex v_k such that $<v_k,\ldots,v_1>$ and $<v_k,\ldots,v_n>$ are directed subpaths of p that intersect only at v_k.

Hence, corresponding to each undirected path p = $<v_1,v_2,...v_{n-1},v_n>$ without direction reversals is a trek t = $<<v_k,...,v_1>,<v_k,...,v_n>>$.

Lemma 2.2.2: In an acyclic directed graph G, for every trek $<<v_1,...,v_n>,<v_1,...,v_m>>$, the concatenation of $<v_n,...,v_1>$ with $<v_1,...,v_m>$ is an undirected path from v_n to v_m without direction-reversals.

Definition 2.2.4: Let P be a probability distribution defined over a set of random variables **U** = $\{x_1,x_2,...,x_n\}$ of elements, and let O be an ordering $(x_1,x_2,...,x_i,...)$ of the elements of **U**. a Markov boundary $\mathbf{B}_i$ of x_i with respect to the set $\mathbf{U}_{(i)} = \{x_1,x_2,...,x_{i-1}\}$ is a minimal set satisfying $\mathbf{B}_i \subseteq \mathbf{U}_{(i)}$ and $I(x_i,\mathbf{B}_i,\mathbf{U}_{(i)} - \mathbf{B}_i)$.The **boundary strata** of P relative to O is an ordered set of subsets of **U**, $(\mathbf{B}_1,\mathbf{B}_2,...,\mathbf{B}_i)$ such that each $\mathbf{B}_i$ is a Markov boundary of x_i with respect to $\mathbf{U}_{(i)}$. The DAG created by designating each $\mathbf{B}_i$ as parents of vertex x_i is called a **boundary DAG** of P relative to d.

The following theorem is taken from [3].

Theorem 2.2.1: If P is a probability distribution, and D is a boundary DAG of M relative to any ordering d, then D is a minimal I-map of P.

I will say that an acyclic graph has error variables if every vertex of indegree not equal to 0 has an edge into it from a vertex of indegree 0 and outdegree 1. If each independent random variable in a LCT S is normally distributed, then the joint distribution of the set of all random variables in the LCT is a multi-variate normal distribution. I will say the random variables in such a LCT have a linear multi-variate normal distribution. The next series of lemmas demonstrate that every acyclic graph with error variables is a perfect map of some LCT S in which the joint distribution Q of the random variables in S is a linear multi-variate normal distribution.

Lemma 2.2.3: If S = $<G, (\Omega,f,P), X, L>$ is an acyclic LCT in which the exogenous variables are jointly independent, then G is a minimal I-map of P.
Proof. I will show that G is a boundary DAG of P relative to some ordering d. It will then follow from Theorem2.2.1 that $<R,E>$ is a minimal I-map of Q.

Order the variables in R in any order such that for all x and y, if x is the tail of an arrow to y, then x precedes y in the ordering.

IIn order to show that G is a boundary DAG of P, it suffices to show that if the set of parents of x = x_i, is a set **B**, then **B** is a Markov boundary of x.

Let P_x be the set of parents of x in G. The values of variables in P_x completely determine the value of x. Hence, the distribution of x given P_x is the same as the distribution of x given any superset of its parents. Hence x is conditionally independent of all other variables given its parents.

Now I will show that there does not exist a proper subset A of x's parents such that x is independent of P_x - A given A. Suppose that such a subset exists. If x is independent of P_x - A given A, it follows that with probability 1, a linear combination of the values of P_x - A cancel (since they do not contribute to the value of x). This implies that with probability 1, some variable y in P_x - A is a linear combination of (P_x - A) - {y}. Hence the variance of y conditional on (P_x - A) - {y} is equal to zero, even though (P_x - A) - {y} does not contain the error variable of y. This is a contradiction, because it follows from the definition of a LCT that the variance of an endogenous variable conditional on any set of variables that does not contain its error variable is non-zero. ∴

Lemma 2.2.4: If a polynomial equation Q in real variables $<x_1,...,x_n>$ is not an identity, then for every solution *a* of Q, and for every $\varepsilon > 0$ there is a non-solution *b* of E such that $|b - a| < \varepsilon$.
Proof. The proof is by induction on the number n of variables in Q.

Base case: If n = 1, then there are only a finite number of solutions of Q. It follows that for every solution *a* of Q, and for every $\varepsilon > 0$ there is a non-solution *b* of Q such that $|b - a| < \varepsilon$.

Induction case: Suppose that Q is a polynomial equation in $<x_1,...,x_n>$, Q is not an identity, and the lemma is true for n-1. Take an arbitrary solution $<a_1,...,a_n>$ of Q. Transform Q into a polynomial equation Q' in x_n by fixing the values of the variables $<x_1,...,x_{n-1}>$ at $<a_1,...,a_{n-1}>$. There are two cases.

In the first case, Q' is not an identity. In that case, there are only a finite number of solutions to P'. Hence, there is a non-solution of Q' whose distance from a_n is $< \varepsilon$. Let a'_n be this non-solution of Q'. $a' = <a_1,...a_{n-1},a'_n>$ is a non-solution of Q, and $|a - a'| < \varepsilon$.

In the second case, Q' is an identity. Rewrite Q so that it is of the form

$$\sum_m Q_m x_n^m$$

where each Q_m is a polynomial in at most $x_1,...,x_{n-1}$.

For each m, the equation $Q_m = 0$ is a polynomial equation in less than n variables. If Q' is an identity, then when terms of the same power of x_n are added together, the coefficient of each power of x_n is zero. This implies that $<a_1,...,a_{n-1}>$ is a solution to $Q_m = 0$ for each m. If, for each m, $Q_m = 0$ is an identity, then so is Q; hence for some m, $Q_m = 0$ is not an identity. For this value of m, by the induction hypothesis, there is a non-solution $<a'_1,...,a'_{n-1}>$ to $Q_m = 0$ that is less than distance ε from $<a_1,...,a_{n-1}>$. If $<a'_1,...,a'_{n-1}>$ is substituted for $<x_1,...,x_{n-1}>$in Q, the resulting polynomial equation in x_n is not an identity. This reduces to the first case. ∴

Lemma 2.2.5: γ_{ij} is equivalent to a polynomial in the linear coefficients and variances of the independent variables.
Proof. Let a 0-generation random variable be an independent variable, and an n-generation random variable be such that the highest generation random variable that it contains in its structural equation is an n-1 generation variable. Let the generation of γ_{ij} equal the maximum of the generations of i and j. The proof is by induction on the generation of γ_{ij}.

Base Case. If the generation of γ_{ij} is zero, then $\gamma_{ij} = \sigma^2{}_i$ if i = j, and 0 otherwise. In either case, the equation is trivially a polynomial in the linear coefficients and variances of the independent variables.

Induction Case. Suppose that the lemma is true for n-1 generation covariances, and γ_{ij} is an n generation covariance. If

$$i = \sum_n a_n x_n \quad \text{and} \quad j = \sum_n b_n x_n$$

then

$$\gamma_{ij} = \sum_n a_n b_n \sigma_{x_n}{}^2 + \sum_{r,s} \sum_{r<s} a_r b_s \gamma_{x_r x_s}$$

By the induction hypothesis, $\gamma_{x_r x_s}$ and $\sigma_{x_n}{}^2$ are equivalent to polynomials in the linear coefficients and variances of the independent variables, (since $\sigma_{x_n}{}^2$ is just the covariance of x_n and x_n.) Hence, γ_{ij} is equivalent to a polynomial in the linear coefficients and variances of the independent variables. $\therefore$

Lemma 2.2.6: $\rho_{ij.X} = 0$ is equivalent to a polynomial equation in the linear coefficients and variances of the independent variables.

Proof. I will prove more generally that a polynomial equation in partial covariances is equivalent to a polynomial equation in the linear coefficients and variances of the independent variables. If X contains n distinct variables, then say $\rho_{ij.X}$ is a **partial correlation of order n**. Let the **pc-order** (partial correlation order) **of a polynomial in partial covariances** be the highest order of any partial covariance appearing in the polynomial. The proof is by induction on the pc-order of the polynomials.

Base Case. If polynomial P is of pc-order 0, then by lemma 2.2.5, P is equivalent to a polynomial equation in the linear coefficients and variances of the independent variables.

Induction Case. Suppose that the lemma is true for polynomials of pc-order n-1, and P be a polynomial of pc-order n. The recursion formula for partial covariances is

$$\gamma_{ij.Y \cup r} = \gamma_{ij.Y} - \frac{\gamma_{ir.Y}\gamma_{jr.Y}}{\gamma_{rr.Y}}$$

Form P' by using this recursion formula to replace each covariance of pc-order n appearing in P by an algebraic combination of covariances of pc-order n-1. Form P" by multiplying P' by the lowest common denominator of all of the terms in P', producing a polynomial of pc-order n-1. By the induction hypothesis, P" is equivalent to a polynomial equation in the linear coefficients and variances of the independent variables. Hence, a polynomial equation in partial covariances is equivalent to a polynomial equation in the linear coefficients and variances of the independent variables.

By definition,

$$\rho_{ij.X} = \frac{\gamma_{ij.X}}{\sqrt{\gamma_{ii.X}}\sqrt{\gamma_{jj.X}}}$$

so $\rho_{ij.X} = 0$ iff $\gamma_{ij.X} = 0$. Since the latter is a polynomial equation in partial covariances, it is equivalent to a polynomial equation in the linear coefficients and variances of the independent variables. It follows that the former is also equivalent to a polynomial equation in the linear coefficients and variances of the independent variables. ∴

Lemma 2.2.7: If G' is a subgraph of G, and there is some LCT S' containing G' such that $\rho_{ij.Z} \neq 0$ in S', then there is some LCT S containing G such that $\rho_{ij.Z} \neq 0$ in S.
Proof. By lemma 2.2.6 in G' $\rho_{ij.Z} = 0$ is equivalent to a polynomial equation P' in the linear coefficients and variances of independent variables in G'. Since there is some LCT S' containing G' such that $\rho_{ij.Z} \neq 0$ in S', P' is not an identity.

Similarly, in G $\rho_{ij.Z} = 0$ is equivalent to a polynomial equation P in the linear coefficients and variances of independent variables in G. The difference between P and P' (if any) is that P contains every term in P', as well as some additional terms. Let us say that two terms in a polynomial equation are **like terms** if they contain the same variables raised to the same powers. Each of the terms that are in P but not P' contain some linear coefficient that does not appear in any term in P'; hence each of the additional terms in P is not like any term in P'.

If P were an identity, then the sum of the coefficients of like terms in P would be equal to zero. Since P' is not an identity, there are like terms in P' such that the sum of their coefficients is not zero. These same like terms appear in P. Furthermore, since the only additional terms in P that are not in P' are not like any term in P', it follows that if the sum of the coefficients of like terms in P' is not zero, then the sum of the coefficients of the same like terms in P is not identically zero. Hence P is not identically zero, and there is some LCT S containing G such that $\rho_{ij.Z} \neq 0$ in S. ∴

The next lemma states that given a set Z of partial correlations and a graph G, if it is possible to construct a set S of LCTs containing G such that each z in Z fails to vanish for some one of the LCTS in S, then it is possible to construct a single LCT containing G such that all of the z in Z fail to vanish.

Lemma 2.2.8: Given a set of partial correlations Z and a graph G, if for all z in Z there exists a LCT S' containing G such that $z \neq 0$ in S', then there exists a single LCT S such that for all z in Z, $z \neq 0$ in S.
Proof. The proof is by induction on the cardinality of Z.

Base Case: If the only member of Z is z_1, then by assumption there is a LCT S containing G such that $z_1 = 0$.

Induction Case: Suppose that the lemma is true for each set of cardinality n-1, Z is of cardinality n, and for each z_i in Z, there is a LCT S' such that $z_i \neq 0$ in S'. By the induction hypothesis, there is a LCT S such that $z_i \neq 0$, $i \leq 1 \leq n-1$. Let V be a set values for the linear coefficients and variances of independent variables such that $z_i \neq 0$, $i \leq 1 \leq n-1$. The valuation V either makes z_n equal to zero or it doesn't. If it doesn't, then the proof is done. If it does, I will show how to perturb V by a small amount to make $z_n \neq 0$, while keeping each $z_i \neq 0$, $i \leq 1 \leq n-1$.

By lemma 2.2.6, each of the partial correlations in z_i in Z is equivalent to a polynomial p_i in the linear coefficients and the variances of independent variables in G. Suppose that the smallest non-zero value for any of the p_i under the valuation V is δ. By lemma 2.2.4, for arbitrarily small ε there is a non-solution V' to $z_n = 0$ within distance ε of V. Choose an ε small enough so that the largest possible change in any of the p_i is less than δ. For the valuation V' then $z_i \neq 0$, $i \leq 1 \leq n$. $\therefore$

Lemma 2.2.9: For every acyclic graph <R,E> with error variables, there is a LCT S, such that Q, the joint distribution of the random variables is linear multi-variate normal and the exogenous variables are jointly independent, and <R,E> is a D-map of Q.

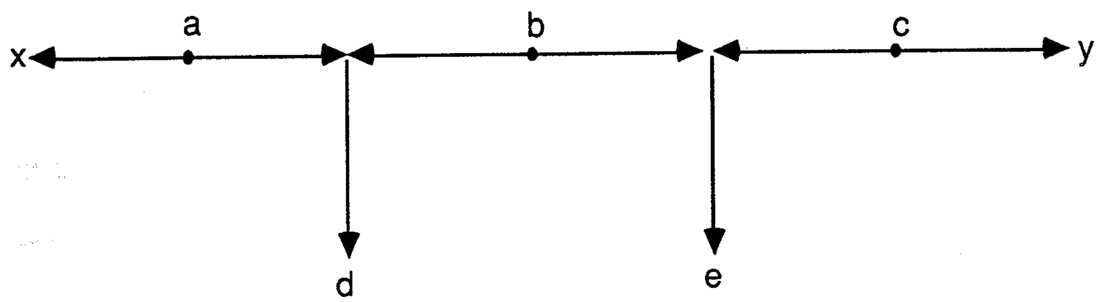
Proof. In order to show that <R,E> is a D-map of Q, I must show that for all disjoint sets of variables X, Y, and Z, if X and Y are not d-separated in <R,E>, then $\sim I(X,Z,Y)_Q$. In a linear multi-variate normal distribution, if X, Y, and Z are disjoint sets of variables, then $I(X,Z,Y)_Q$ iff $I(x,Z,y)_Q$ for each x in X and y in Y; similarly if X, Y, and Z are disjoint sets of variables then Z d-separates X and Y iff Z d-separates x in X and y in Y. Hence, I need consider only dependency statements of the form $\sim I(x,Z,y)_Q$, where x and y are individual variables. Also in a linear multi-variate normal distribution, $\rho_{xy.Z} = 0$ iff $I(x,Z,y)_Q$. So it suffices to prove that there is a LCT S such that for each x, y, and Z in <R,E> such that x and y are not d-separated by Z in <R,E>, $\rho_{xy.Z} \neq 0$ in S. The proof is by induction. I assume that in all of the LCTs constructed, the independent random variables are normally distributed.

Base Case. If Z is empty, then by lemma 2.2.1 and definition 2.2.2, x and y are not d-separated iff there is a trek connecting them. Form a subgraph G' and a sub-LCT S' containing G', such that there is exactly one trek between x and y. It was proved in [2] that in this case, the covariance between x and y is equal to the product of the labels of the edges in the trek (the linear coefficients) times the variance of the source of the trek. If each of these quantities is non-zero, so is the covariance, and also the correlation in S'. By lemma 2.2.7 if ρ_{xy} is not identically zero in S' it is also not identically zero in some LCT S containing <R,E>.

Induction Case. Suppose that for each x, y, and A of cardinality less than n such that x and y are *not* d-separated by A in <R,E>, there is a LCT S such that $\rho_{xy.A} \neq 0$ in S, and that Z is of cardinality n. Suppose that x and y are not d-separated by Z in <R,E>. It follows that there is an undirected path U between x and y such that every vertex without a direction reversal is not in Z, and every vertex v_i on U with a direction reversal is the source of a directed path U_i from v_i to a variable in Z. Form a subgraph G', and sub-LCT S' that contains G', such that G' contains only the undirected path U, one directed path U_i from each vertex v_i on U with a direction reversal, the vertices in those paths, and the vertices in Z. Shorten each U_i so that it contains only one variable in Z. Finally, if two variables v_n and v_m that have direction-reversals in U are the sources of directed paths D_n and D_m to the same variable v_o, replace the segment $<v_n,...,v_m>$ in U by the the undirected segment $<v_m,...,v_o,...,v_n>$, the concatenation of D_n and D_m. Now for each distinct vertex in U that has a direction reversal, G' contains a directed path to a distinct variable in Z. There are two cases.

In the first case, U contains no vertices with a direction reversal, and hence no vertices in Z. By lemma 2.2.1 there is a trek between x and y that contains no vertices in Z. Let r be an arbitrary vertex in Z, and W = Z - {r}. There is a trek between x and y that contains no vertices in W. It follows that W does not d-separate x and y, so by the induction hypothesis, there is a LCT containing G' such that $\rho_{xy.W} \neq 0$. Since by construction there are no undirected paths from x to r or y to r, it follows that in G' that $\rho_{xr.W} = 0$ and $\rho_{yr.W} = 0$. By the recursion formula for partial correlation, $\rho_{xy.Z} = 0$ iff $\rho_{xy.W} = \rho_{xr.W} \times \rho_{yr.W}$. But $\rho_{xy.W}$ is non-zero in S', and $\rho_{xr.W} \times \rho_{yr.W}$ is zero in S'. Hence $\rho_{xy.Z} \neq 0$ in S'. By lemma 2.2.7, there is some LCT S containing <R,E> such that $\rho_{xy.Z} \neq 0$ in S.

In the second case, U contains vertices with direction reversals, but every vertex without a direction-reversal is not in Z. See Fig. 2.2.1.



Z = {d,e}

Fig. 2.2.1

Let e be the vertex that is the sink of the directed path from the direction-reversal closest to y, and W = Z - {e}. Since by construction there is a trek between y and e that does not contain any variables in W, y and e are not d-separated by W. There is also an undirected path from x to e such that every vertex that does not contain a direction-reversal is not in W, and every vertex that does contain a direction-reversal has a descendant in W. Hence x and e are not d-separated by W. By the induction hypothesis, there is a LCT S' containing G' such that $\rho_{xe.W} \neq 0$, and $\rho_{ye.W} \neq 0$.

On the other hand, since path U was constructed so that each vertex that contained a direction-reversal had only one descendant in Z, and W does not contain e, x and y are d-separated by W. Hence $\rho_{xy.W} = 0$ in S'.

$\rho_{xy.Z} = 0$ iff $\rho_{xy.W} = \rho_{xe.W} \times \rho_{ye.W}$. Since $\rho_{xy.W} = 0$, while $\rho_{xe.W} \times \rho_{ye.W} \neq 0$, $\rho_{xy.Z} \neq 0$ in S'. By lemma 2.2.7, there is a LCT S containing <R,E> such that $\rho_{xy.Z} \neq 0$.

Since for each triple x, y, Z not d-separated in <R,E> there is a LCT S' containing <R,E> such that $\rho_{xy.Z} \neq 0$, by lemma 2.2.8 there is a LCT S containing <R,E> such that for each triple x, y, Z not d-separated in <R,E>, $\rho_{xy.Z} \neq 0$. Since the LCTs constructed in lemmas 2.2.7 and 2.2.8 don't change the normality of the independent variables, the joint distribution of the random variables in S is a linear multi-variate normal distribution. Hence there is a LCT S such that Q is a linear multi-variate normal distribution and <R,E> is a D-map of the Q. $\therefore$

Theorem 2.2.2: For every acyclic graph <R,E> with error variables, there is a LCT S containing <R,E> with a joint distribution Q of random variables that is linear multi-variate normal such that <R,E> is a perfect map of Q.

Proof. This follows immediately from lemmas 2.2.9 and 2.2.3. $\therefore$

The next theorem states that the d-separabilty relations between sets of non-error variables can be determined from a subgraph that does not include error terms.

Theorem 2.2.3: In an acyclic LCT S with graph G, let G' be the subgraph of G consisting of all of the non-error variables. Given three disjoint sets X, Y, and Z of non-error variables, X is d-separated from Y by Z in G iff X is d-separated from Y by Z in G'.
Proof. If an error variable occurs on an undirected path, then that error variable is either the source or the sink of the undirected path. Hence, error variables do not occur on any undirected path between non-error variables. It follows that the undirected paths in G and G' between non-error variables are exactly the same. The theorem then follows from the definition of d-separability. $\therefore$

Lemma 2.2.10: In a LCT S = <G, (Ω,f,P), X, L> if Y d-separates x and z, then $\rho_{xz.Y}$ is linearly implied to vanish.
Proof. Suppose Y d-separates x and z. The values of the partial correlations in P are completely determined by the values of the linear coefficients and the variances of the independent variables. Consider a multi-variate normal distribution P' in the LCT with the same linear coefficients and the same variances of independent variables as S, but in which the independent variables are normally distributed and jointly independent. By lemma 2.2.3 G is a minimal I-map of P', and because Y d-separates x and z, I(x,Y,z) in P'. Because P' is a multi-variate normal distribution, I(x,Y,z) if and only $\rho_{xz.Y} = 0$. It follows that $\rho_{xz.Y} = 0$ in P', and hence $\rho_{xz.Y} = 0$ in P. $\therefore$

Theorem 2.2.4: In a LCT S = <G, (Ω,f,P), X, L>, if x and z are distinct non-error variables, and Y is a set of non-error variables not including x and z, then Y d-separates x and z if and only if $\rho_{xz.Y}$ is linearly implied to vanish.
Proof. The only if clause follows from Lemma 2.2.10.

The if clause follows from lemma 2.2.9. By lemma 2.2.9 there is a LCT S such that Q, the joint distribution of the random variables is linear multi-variate normal, and G is a D-map of Q. In S, if x and z are not d-separated by Y, then x and z are not independent given Y, and $\rho_{xz.Y} \neq 0$. Hence if x and z are not d-separated by Y, $\rho_{xz.Y}$ is not linearly implied to vanish. $\therefore$

Corollary 2.2.1: In a LCT S = <G, (Ω,f,P), X, L> in which the exogenous variables are jointly independent, if x and z are distinct non-error variables, and Y is a set of non-error variables not including x and z, if $\rho_{xz.Y}$ is linearly implied to vanish then I(x,Y,z).
Proof. By Theorem 2.2.4, if $\rho_{xz.Y}$ is linearly implied to vanish then Y d-separates x and z in G. By lemma 2.2.3, G is a minimal I-map of P. Hence if Y d-separates x and z in G, I(x,Y,z). $\therefore$

Corollary 2.2.2: In a LCT S = <G, (Ω,f,P), X, L>, if G is a perfect map of P, x and z are distinct non-error variables, and Y is a set of non-error variables not including x and z, $\rho_{xz.Y}$ is linearly implied to vanish if and only if I(x,Y,z).
Proof. By Theorem 2.2.4, $\rho_{xz.Y}$ is linearly implied to vanish if and only if Y d-separates x and z in G. Since G is a perfect map of P, Y d-separates x and z if and only if I(x,Y,z). Hence $\rho_{xz.Y}$ is linearly implied to vanish if and only if I(x,Y,z). $\therefore$

2.3. The Path Form of a Covariance

In this section, I will prove a number of theorems that show how to express the covariance between i and j in terms of the labels of paths from i to j.

Definition 2.3.1: An **independent equation for a dependent variable j in a LCF** is an equation implied by X in which the random variables which appear on the r.h.s. are independent and have a non-zero coefficient, and each random variable occurs at most once on the r.h.s.

${}^{I}a_{ij}$ is the coefficient of j in the independent equation for i.

Lemma 2.3.1: In a LCF S, if j is an independent variable, then

$$ {}^{I}a_{ij} = \sum_{p \in P(i,j)} L(p) $$

Proof. This is a special case of Mason's rule for calculating the "total effect" of a variable j on a variable i. See [2]. ∴

The following two lemmas show how to calculate the variance of random variables and covariances between random variables in terms of the covariances between other random variables. The proofs of these lemmas can be found in [1].

Lemma 2.3.2: If Q is a set of random variables with a joint probability distribution and

$$ y = \sum_{i \in Q} a_{yi}\, i $$

and

$$ z = \sum_{j \in Q} a_{zj}\, j $$

then

$$ \gamma_{yz} = \sum_{i \in Q} \sum_{j \in Q} a_{yi} b_{zj} \gamma_{ij} $$

Lemma 2.3.2: If Q is a set of random variables with a joint probability distribution and

$$ y = \sum_{i \in Q} a_{yi}\, i $$

then

$$ \sigma^2{}_y = \sum_{i \in Q} \sum_{j \in Q} a_{yi} a_{yj} \gamma_{ij} $$

Definition 2.3.2: In a LCF S, U_X is the set of all independent variables that are the source of an open path to x. (Note that if x is independent then $x \varepsilon U_x$ since there is an empty path from every vertex to itself.) U_{xy} is $U_x \cap U_y$. I is the set of all independent variables in S.

Lemma 2.3.4: In a LCF S, if

$$y = \sum_{i \varepsilon I} {}^{I}a_{yi}\, i$$

and

$$z = \sum_{i \varepsilon I} {}^{I}a_{zi}\, i$$

then

$$\gamma_{yz} = \sum_{i \varepsilon U_{yz}} {}^{I}a_{yi}\, {}^{I}a_{zi}\, \sigma^2{}_i$$

Proof. I is a set of independent variables. It follows that γ_{ij} is equal to 0 if $i \neq j$, and γ_{ij} is equal to $\sigma^2{}_i$ if $i = j$. Substituting these values for γ_{ij} into the r.h.s. of the equation for γ_{yz} in lemma 2.3.2 shows that

(1)
$$\gamma_{yz} = \sum_{i \varepsilon I} {}^{I}a_{yi}\, {}^{I}a_{zi}\, \sigma^2{}_i$$

If i is in I, but i is not in U_{yz} then there is no pair of open paths from i to y and z. By lemma 2.3.1, if there is no pair of open paths from i to y and z, then the coefficient of i in the independent equation for either y or z is zero. So, the only non-zero terms in equation 1 are for $i \varepsilon U_{yz}$. ∴

Lemma 2.3.5: In a LCF S, if

$$y = \sum_{i \varepsilon I} {}^{I}a_{yi}\, i$$

then

$$\sigma^2{}_y = \sum_{i \varepsilon U_y} {}^{I}a^2_{yi}\sigma^2{}_i$$

Proof. I is a set of independent random variables. It follows that γ_{ij} is equal to 0 if $i \neq j$, and γ_{ij} is equal to $\sigma^2{}_i$ if $i = j$. Substituting these values for γ_{ij} into the r.h.s. of the equation for $\sigma^2{}_y$ in lemma 2.3.3 proves that

(2)
$$\sigma^2{}_y = \sum_{i \varepsilon I} {}^{I}a^2_{yi}\sigma^2{}_i$$

If i is in I, but i is not in U_y, then there is no open path, and hence no path, from i to y. It follows from lemma 2.3.1 that a_{yi} is zero. Hence the only non-zero terms in equation 2 come from $i \varepsilon U_y$.

∴

Lemma 2.3.6: In a LCF S

$$\gamma_{ij} = \sum_{k \,\varepsilon\, U_{ij}} \left(\sum_{p \,\varepsilon\, P_{ki}} \sum_{p' \,\varepsilon\, P_{kj}} L(P) L(p') \sigma_k^2 \right)$$

Proof. This follows immediately from lemmas 2.3.1 and 2.3.4. ∴

Lemma 2.3.7: In a LCF S

$$\sigma_i^2 = \sum_{k \,\varepsilon\, U_i} \left(\left(\sum_{p \,\varepsilon\, P_{ki}} L(P) \right)^2 \sigma_k^2 \right)$$

Proof. This follows immediately from lemmas 2.3.1 and 2.3.5. ∴

Lemma 2.3.8: In an acyclic graph, for all variables y and z, if $y \neq z$ and p and p' are two intersecting paths with sinks y and z respectively then there is a trek between y and z that consists of subpaths of p and p'.

Proof. Since p and p' intersect, they have a last point of intersection x. Let the source of the trek to be constructed be x. p(x,y) and p(x,z) do not intersect anywhere except at x. Since $y \neq z$, one of p(x,y) and p(x,z) is not empty. Hence <p(x,y),p(x,z)> is a trek. ∴

Definition 2.3.3: In an acyclic graph, p(u,i) and p(u,j) contain trek t iff i(t(i,j)) is a final segment of p(u,i) and j(t(i,j)) is a final segment of p(u,j).

Lemma 2.3.9: In an acyclic graph, if p(u,i) and p(u,j) contain both t(i,j) and t'(i,j), then t(i,j) = t'(i,j).

Proof. In an acyclic graph, there is a unique last point of intersection of p(u,i) and p(u,j), and unique final segments of p and p' whose source is the last point of intersection of p(u,i) and p(u,j). ∴

Definition 2.3.4: In a LCF S, the **path form of a product of covariances** $\gamma_{ij}\gamma_{kl}$ is the distributed form of

$$\left(\sum_{u \,\varepsilon\, U_{ij}} \left(\sum_{p \,\varepsilon\, P_{ui}} \sum_{p' \,\varepsilon\, P_{uj}} L(p) L(p') \sigma^2{}_u \right) \right) \left(\sum_{v \,\varepsilon\, U_{kl}} \left(\sum_{p'' \,\varepsilon\, P_{vk}} \sum_{p''' \,\varepsilon\, P_{vl}} L(p'') L(p''') \sigma^2{}_v \right) \right)$$

$\gamma_{ij}\gamma_{kl} - \gamma_{il}\gamma_{jk}$ **is in path form** iff both terms are in path form.

Henceforth, I will assume that all variances, covariances, products of covariances, and tetrad differences are in path form unless otherwise stated.

2.4. Vanishing Tetrad Differences: Necessary Conditions

Definition 2.4.1: In a LCF S, two expressions are **identically equal** iff they are equal for all values of the equation coefficients. A vanishing tetrad difference is **linearly implied by a LCF** S iff it is implied for all values of its linear coefficients and all non-zero variances of the independent variables. A vanishing tetrad difference is **linearly implied by a DAG** G iff it is implied by every probability distribution of which G is a perfect map. A vanishing tetrad difference is **linearly implied by a LCT** S if and only if it is linearly implied by the graph G of S.

I will adopt the following terminology. Suppose that m is a monomial in the path form of a product of covariances $\gamma_{ij}\gamma_{kl}$. By definition, m is of the form $L(p(u,i))L(p(u,j))L(p(v,k))L(p(v,l))\sigma^2{}_u\sigma^2{}_v$. Let the paths associated with m be the ordered quadruple $<p(u,i),p(u,j),p(v,k),p(v,l)>$. There is a one-to-one correspondence between monomials in the path form of a product of covariances, and such ordered quadruples. I will consider monomials m and m' that have equal values (or are even identically equal values) to be distinct monomials if their associated paths are different (i.e. the monomials may contain the same number of occurrences of the same edge labels, but in different orders.) Note that under this criterion of identity of monomials, no monomial appears twice in the path form of a product of covariances or tetrad difference. Henceforth when I consider sets of monomials appearing in some expression, I will do so under the assumption that each monomial occurs at most once in the expression (although distinct monomials that have identically equal values may occur in the expression). I will say that a monomial m contains a path or trek x if its associated quadruple contains x. For the sake of brevity I will use "implied" to mean "linearly implied" unless explicitly stated otherwise.

Lemma 2.4.1: A tetrad difference $\gamma_{ij}\gamma_{kl} - \gamma_{il}\gamma_{jk}$ is not linearly implied to vanish by a LCF S if there is a monomial m in the path form of $\gamma_{ij}\gamma_{kl}$ such that every monomial m' in the path form of $\gamma_{il}\gamma_{jk}$ contains an edge not in m.

Proof. Suppose that there is a monomial m in the path form of $\gamma_{ij}\gamma_{kl}$ such that every monomial m' in the path form of $\gamma_{il}\gamma_{jk}$ contains an edge not in m. Set every variable not in m to be zero. Then $\gamma_{il}\gamma_{jk}$ is zero since every monomial in $\gamma_{il}\gamma_{jk}$ contains a variable not in m. Set every variable in m to be positive. Then every non-zero monomial in the path form of $\gamma_{ij}\gamma_{kl}$ is positive, since the e.c.f of each non-zero monomial is positive, and the c.f. of each non-zero monomial is positive. $\gamma_{ij}\gamma_{kl}$ is not zero since every monomial in it is either 0 or positive, and some are positive. Hence the tetrad difference is not linearly implied to vanish. ∴

Lemma 2.4.2: In a LCF S, if the paths in a monomial m in the path form of a tetrad difference have different sources than the paths in a monomial m', then m contains some variable not in m'.

Proof. Each of the sources of the paths in m and m' are independent variables, and it is not the case that all of the paths in m or m' are empty. Let {i,j} be the sources of the paths in m, and {k,z} be the sources of the paths in m' and suppose that $\{i,j\} \neq \{k,z\}$. Suppose w.l.g. that $i \neq k$. Since i, k, and z are independent i does not occur on any paths with source k or z. m contains at least one edge x out of i. Since i does not occur on any path with source k or z, x does not occur on any path with source k or z. Hence m contains a variable (the label of x) that does not occur in m'. ∴

Definition 2.4.2: In a LCF S, **e(s)** is equal to s if s is an independent variable, and it is equal to the error variable into s if s is not an independent variable.

Lemma 2.4.3: In a LCF S, if there exist $t(i,j) \varepsilon T(i,j)$ and $t(k,l) \varepsilon T(k,l)$ such that $i(t(i,j)) \cap k(t(k,l)) = \varnothing$, $j(t(i,j)) \cap l(t(k,l)) = \varnothing$, and $i(t(i,j)) \cap l(t(k,l)) = \varnothing$, then there exists a monomial m in $\gamma_{ij}\gamma_{kl}$ such that every monomial m' in $\gamma_{il}\gamma_{jk}$ contains an edge not in m.

Proof. Let s be the source of t(i,j) and s' be the source of t(k,l). (Note that since i(t(i,j)) does not intersect l(t(k,l)), the source of t(i,j) does not equal the source of t(k,l), and hence e(s) does not equal e(s'). See Fig. 2.4.1. Let m = L(p(e(s),i))L(p(e(s),j))L(p(e(s'),k))L(p(e(s'),l)). m is the coefficient of a monomial in $\gamma_{ij}\gamma_{kl}$ (the full monomial also contains a factor equal to the product of the variances of the sources of the treks in m.)

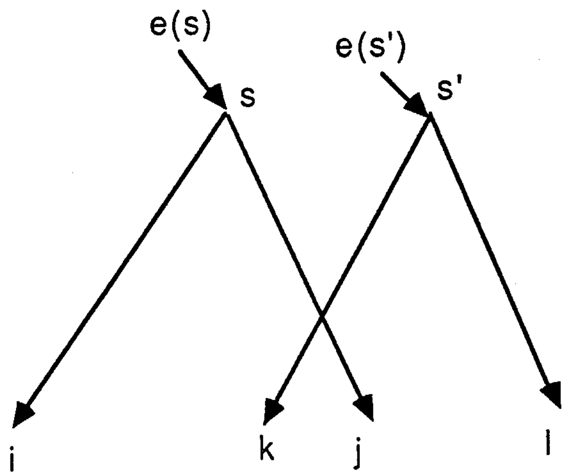


Fig. 2.4.1

Suppose there is a monomial m' in $\gamma_{il}\gamma_{jk}$ whose associated paths contain only edges in m. m' contains the product of the labels of edges in a trek t(i,l). Let the source of t(i,l) be s". If s" ≠ s and s" ≠ s', then e(s") ≠ e(s) and e(s") ≠ e(s'). Since e(s") is an independent variable, and the only independent variables in m are e(s) and e(s'), if e(s") ≠ e(s) and e(s") ≠ e(s'), then t(i,l) contains an edge not in m. Suppose then w.l.g. that s" = s. There is a path p(s,l) containing edges only in m. Since j(t(i,j)) ∩ l(t(k,l)) = ∅, and i(t(i,j)) ∩ l(t(k,l)), the only path in m that contains l is l(t(k,l)). Hence p(s,l) intersects l(t(k,l)) at some vertex. The only two paths in m with source s are i(t(i,j)) and j(t(i,j)), and neither of them intersects l(t(k,l). Hence one of them intersects some other paths that in turn intersects l(t(k,l)). The only other path in m that intersects l(t(k,l)) is k(t(k,l)). So p(s,l) intersects k(t(k,l)). Since the last point of intersection of l(t(k,l)) and k(t(k,l)) is s', p(s,l) intersects k(t(k,l)) at or before s'. But the only paths with source s in m are j(t(i,j)) and i(t(i,j)), and neither of them intersects k(t(k,l)) at or before s'. Hence, there is no path from s to l containing only edges in m. Similarly it can be shown that there is no path from s' to i containing only edges in m. Hence m' contains an edge not in m. ∎

Theorem 2.4.1: In a LCF S, if there exists a t(i,j) ε T(i,j) and t(k,l ε T(k,l) such that

i(t(i,j) ∩ k(t(k,l)) = ∅, and

l(t(k,l)) ∩ j(t(i,j)) = ∅,

or there exists a t(i,l) ε T(i,l) and t(j,k) ε T(j,k) such that

i(t(i,l)) ∩ k(t(j,k)) = ∅, and

$l(t(i,l)) \cap j(t(j,k)) = \varnothing$,

then S does not strongly imply that $\gamma_{ij}\gamma_{kl} - \gamma_{il}\gamma_{jk}$ vanishes.

Proof. Suppose w.l.g. that $i(t(i,j)) \cap k(t(k,l)) = \varnothing$, and $l(t(k,l)) \cap j(t(i,j)) = \varnothing$. There are four cases: either $i(t(i,j)) \cap l(t(k,l)) = \varnothing$ and $j(t(i,j)) \cap k(t(k,l)) = \varnothing$, $i(t(i,j)) \cap l(t(k,l)) = \varnothing$ and $j(t(i,j)) \cap k(t(k,l)) \neq \varnothing$, $i(t(i,j)) \cap l(t(k,l)) \neq \varnothing$ and $j(t(i,j)) \cap k(t(k,l)) = \varnothing$, or $i(t(i,j)) \cap l(t(k,l)) \neq \varnothing$ and $j(t(i,j)) \cap k(t(k,l)) \neq \varnothing$.

In the first three cases, by lemma 2.4.3 there exists a monomial m in $\gamma_{ij}\gamma_{kl}$ such that every m' in $\gamma_{il}\gamma_{jk}$ contains an edge not in m.

In the fourth case, let x be the last point of intersection of $i(t(i,j))$ and $l(t(k,l))$, and y be the last point of intersection of $j(t(i,j))$ and $k(t(k,l))$. x is not the source of either trek, since otherwise $i(t(i,j)) \cap k(t(k,l))$ or $j(t(i,j)) \cap l(t(k,l))$. Similarly, y is not the source of either trek. $\langle p(x,i),p(x,l)\rangle$ is a trek $t(i,l)$ between i and l, by lemma 2.3.8. Similarly, $\langle p(y,j),p(y,k)\rangle$ form a trek $t(j,k)$. (See Fig. 2.4.2.)

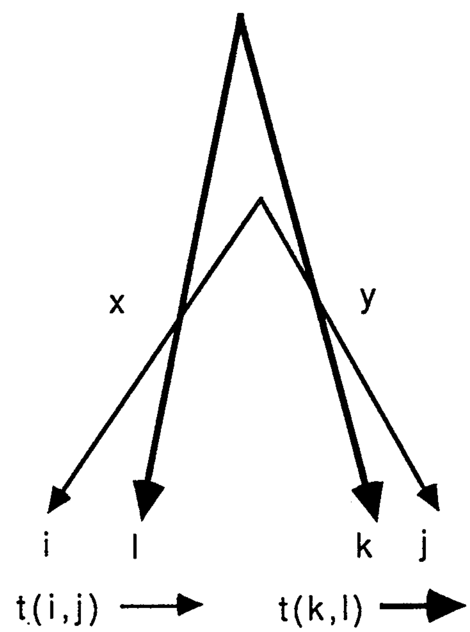


Fig. 2.4.2

Now I will show that $t(i,l) \cap t(j,k) = \varnothing$. $i(t(i,l)) \cap j(t(j,k)) = \varnothing$ since $i(t(i,l))$ is a proper subpath of $i(t(i,j))$ and $j(t(j,k))$ is a proper subpath of $j(t(i,j))$, and the last point of intersection of $i(t(i,j))$ and $j(t(i,j))$ is the source of $t(i,j)$. $i(t(i,l)) \cap k(t(j,k)) = \varnothing$, since $i(t(i,l))$ is a subpath of $i(t(i,j))$ and $k(t(j,k))$ is a subpath of $k(t(k,l))$, and $i(t(i,j)) \cap k(t(k,l)) = \varnothing$ by hypothesis. For similar reasons, $l(t(i,l)) \cap j(t(j,k)) = \varnothing$, and $l(t(i,l)) \cap k(t(j,k)) = \varnothing$. It follows from lemma 2.4.3 there exists a monomial m in $\gamma_{il}\gamma_{jk}$ such that every m' in $\gamma_{ij}\gamma_{kl}$ contains an edge not in m.

Since there exists a monomial m in $\gamma_{il}\gamma_{jk}$ such that every m' in $\gamma_{ij}\gamma_{kl}$ contains an edge not in m, by lemma 2.4.1 $\gamma_{ij}\gamma_{kl} - \gamma_{il}\gamma_{jk}$ is not linearly implied. $\therefore$

2.5. Vanishing Tetrads Differences Imply The Existence of a Choke Point

A vanishing tetrad difference is a constraint upon the covariances of four pairs of variables: <i,j>, <k,l>, <i,l> and <j,k>. Roughly speaking, a choke point for such a foursome of variable pairs is a point where all of the treks between i and j intersect all of the treks between k and l, and all of the treks between i and l intersect all of the treks between j and k. (A more precise definition is given later.) In this section, I will prove that in a LCF G, the existence of such a choke point is a necessary condition for the corresponding tetrad difference to vanish in distributions perfectly represented by G. I will prove this by showing that the existence of a choke point in G is equivalent to a condition that has already been proved to be a necessary condition for S to strongly imply a vanishing tetrad difference; namely, the trek intersection condition described in Theorem 2.4.1. Unfortunately, this proof is long and tedious because there are many different ways in which a choke point can fail to exist, depending upon which treks are assumed to intersect and which treks are assumed not to intersect. In each case I show that the non-existence of a choke point implies the violation of the necessary condition described in Theorem 2.4.1.

Two strategies are employed in the proofs. The first is to show that the assumptions about which treks intersect and don't intersect lead to contradictions. The second is to show that it is possible to construct a pair of treks t'(i,j) and t'(k,l) such that i(t'(i,j)) and k(t'(k,l)) don't intersect, and j(t'(i,j)) and l(t'(k,l)) don't intersect, or to construct a pair of treks t'(i,l) and t'(j,k) such that i(t'(i,l)) and k(t'(j,k)) don't intersect, and j(t'(j,k)) and l(t'(i,l)) don't intersect. In either case, by theorem 2.4.1, it follows that $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk}$ is not linearly implied by G.

In general, when constructing a trek t(i,j) I will speak as if it suffices to show how to construct a pair of (acyclic) paths p and p' from a common source s to sinks i and j respectively, without showing that the pair of paths constructed do not intersect. This is because even if p and p' do not form a trek because they intersect each other at some vertex other than s, I have shown in lemma 2.3.8 that subpaths of p and p' do form a trek, and the existence of the subpaths of p and p' is enough for our purposes. I am generally interested in showing that particular pairs of trek branches fail to intersect. If p_1 and p_2 fail to intersect, then subpaths of p_1 and p_2 also fail to intersect. Hence, if the goal is to show that trek branches t and t' fail to intersect, it suffices to show that p_1 and p_2 fail to intersect, even if t and t' are actually equal to subpaths of p_1 and p_2 respectively.

Let S be a set of vertices, and $P_k(S)$ be the set of all paths with sink k and a source in S. Let p(s,i) be a path from s to i. Let x_n be the n^{th} vertex on p(s,i) such that some path in $P_k(S)$ intersects it. Let the set of sources of paths in $P_k(S)$ whose first point of intersection with p(s,i) is x_n be S_n. Let the last vertex in p(s,i) that is the first intersection of some path in $P_k(S)$ with p(s,i) be x_{max}. Note that x_{max} is not necessarily the last point of intersection of some path in $P_k(S)$ with p(s,i); it is merely the last of the first points of intersection. See Fig. 2.5.1.

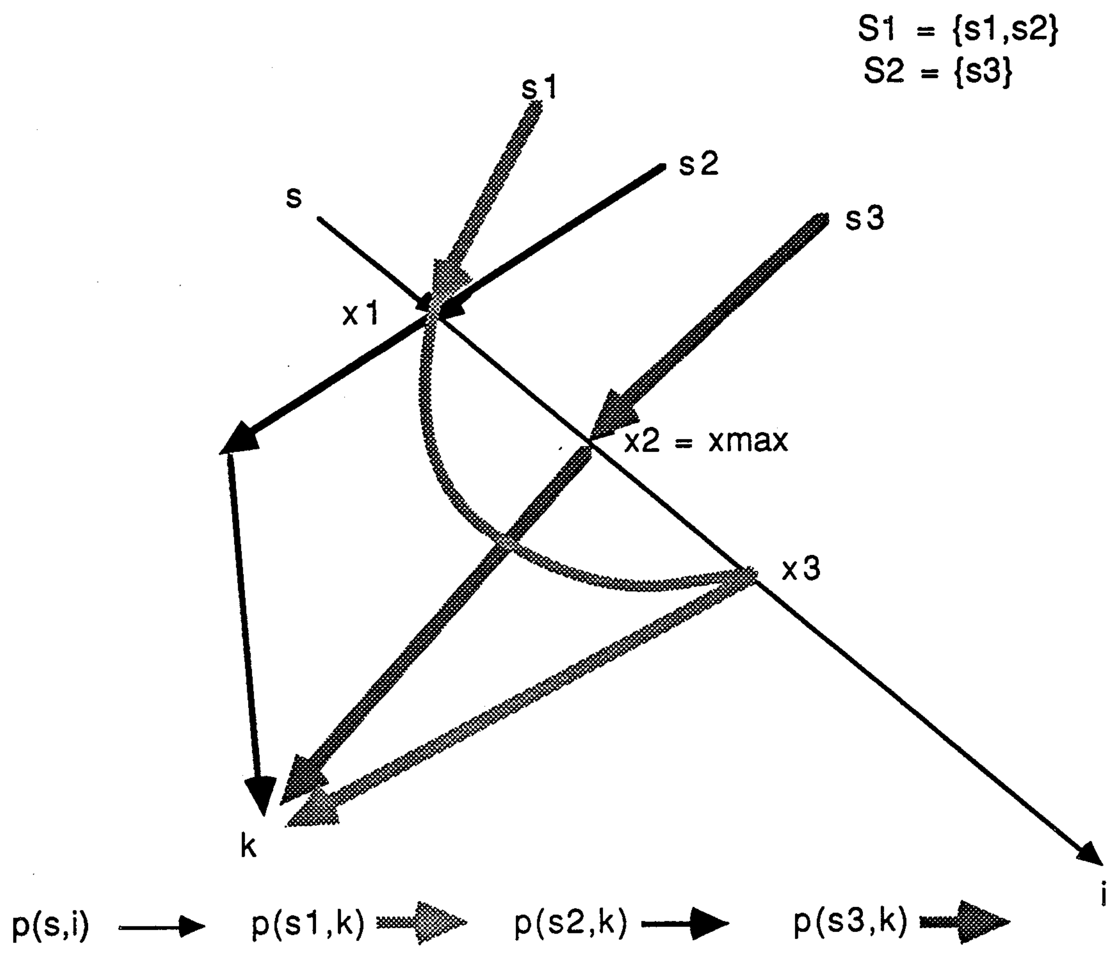


Fig. 2.5.1

Lemma 2.5.2: In an acyclic graph G, if p(s,i) is a path, and $P_k(S)$ is the set of all paths to k from a given set of sources S, and there does not exist a vertex Z such that all of the paths in $P_k(S)$ intersect p(s,i) at Z, then there is a pair of paths, p and p', with the following properties:

s is the source of p, and,

p' has a source in S, and

either p has sink i and p' has sink k or p has sink k and p' has sink i, and

p does not intersect p'.

Proof. If there is a path p' in $P_k(S)$ that does not intersect p(s,i) the proof is done. Assume then that every path in $P_k(S)$ intersects p(s,i). Let s" be the source of a path in S_{max} (the set of all sources of paths in $P_k(S)$ whose first intersection with p(s,i) is x_{max}.) The proof is by induction on the number of distinct vertices in which the paths in $P_k(S)$ intersect p(s,i).

Base Case: Suppose the antecedent is true. The paths in $P_k(S)$ intersect p(s,i) in two distinct vertices. There is a path p(s',k) that does not intersect p(s,i) at x_2 (= x_{max}), since otherwise all paths in $P_k(S)$ would intersect x_2, contrary to our hypothesis. In addition, p(s',k) does not intersect p(s,i) at any vertex prior to x_1, since otherwise the paths in $P_k(S)$ would intersect p(s,i) at more than two distinct vertices, contrary to our hypothesis. Similarly, there is a path p(s",k) that intersects p(s,i) only at x_2.

Let p(x_1,k) be a final segment of p(s',k) and p(s",x_2) an initial segment of p(s",k). There are two cases.

1. p(x1,k) does not intersect p(s",x_2). See Fig. 2.5.2. Let p(s,x1) be an initial segment of p(s,i), p(x_2,i) be a final segment of p(s,i), p = p(s,x_1) &p(x_1,k) and p' = p(s",x_2)&p(x_2,i). p and p' do not intersect for the following reasons:

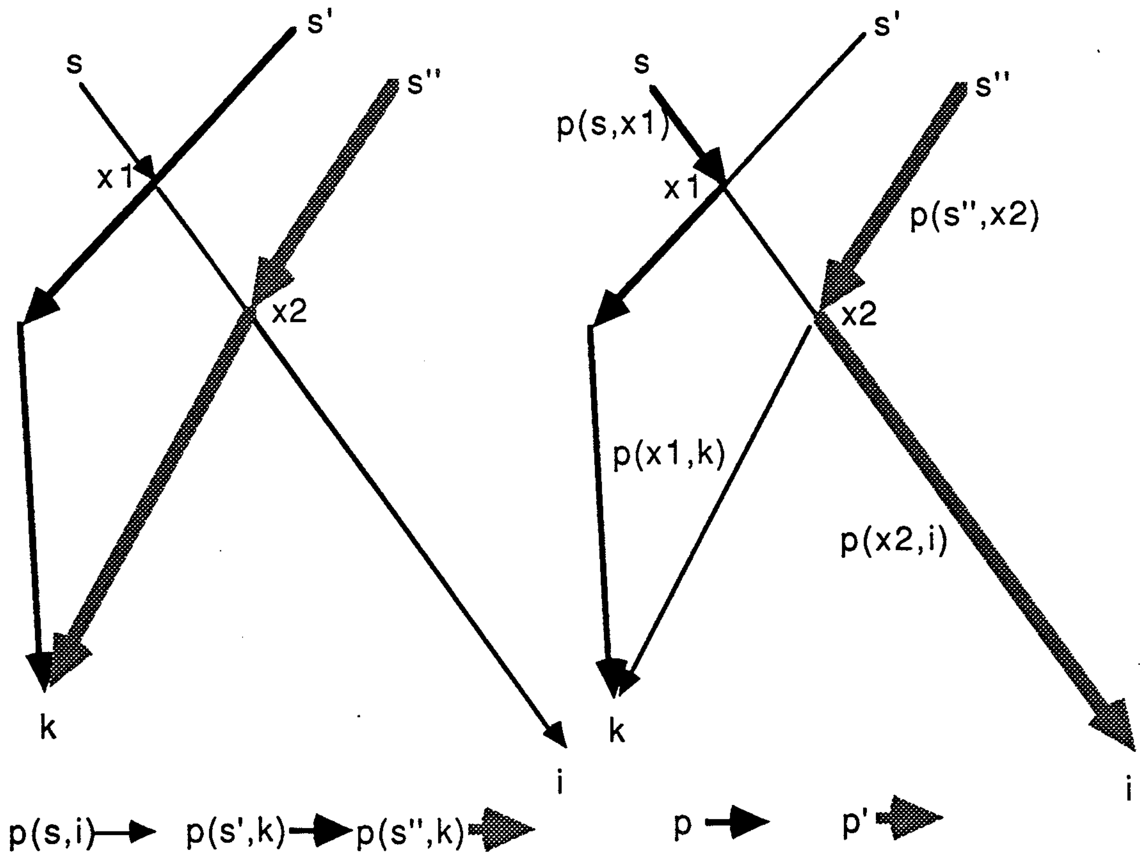


Fig. 2.5.2

a. p(s,x_1) does not intersect p(s",x_2). p(s",x_2) is a subpath of p(s",k), which, by hypothesis intersects p(s,i) only at x_2. Since x_2 occurs after x_1 on p(s,i), x_2 does not occur on p(s,x_1).

b. p(s,x_1) does not intersect p(x_2,i). p(s,x_1) and p(x_2,i) are both subpaths of p(s,i), G is acyclic, and by hypothesis x_1 occurs before x_2.

c. p(x_1,k) does not intersect p(s",x_2) by hypothesis.

d. $p(x_1,k)$ does not intersect $p(x_2,i)$. $p(x_1,k)$ is a subpath of $p(s',k)$ and $p(x_2,i)$ is a subpath of $p(s,i)$; by hypothesis $p(s',k)$ intersects $p(s,i)$ only at x_1, which does not occur on $p(x_2,i)$.

2. $p(x_1,k)$ does intersect $p(s'',x_2)$ at y. See Fig. 2.5.3. Let $p(s'',y)$ be an initial segment of $p(s'',k)$, $p(y,k)$ be a final segment of $p(s',k)$, $p = p(s,i)$ and $p' = p(s'',y)\&p(y,k)$. p and p' do not intersect for the following reasons:

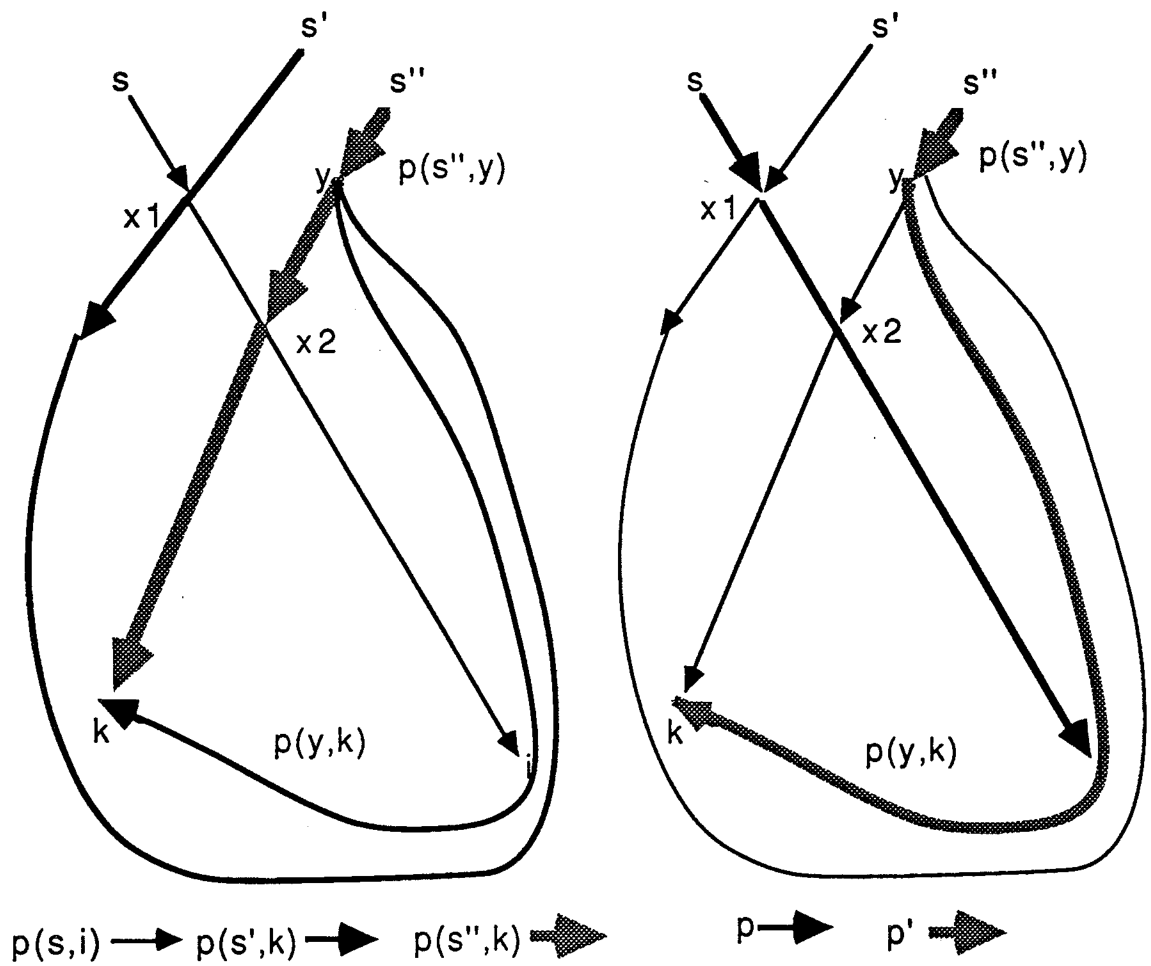


Fig. 2.5.3

a. $p(s,i)$ does not intersect $p(s'',y)$. $y \neq x_2$ since $p(x_1,k)$ intersect $p(s,i)$ only at x_1. Also, G is acyclic, y is prior to x_2 on $p(s'',k)$, and x_2 is the first point of intersection of $p(s'',k)$ with $p(s,i)$.

b. $p(s,i)$ does not intersect $p(y,k)$. y is on $p(s'',k)$ which does not contain x_1; hence y is not equal to x_1.It follows that $p(y,k)$ does not contain x_1, since y occurs after x_1 on $p(s',k)$, and $p(s',k)$ by hypothesis intersects $p(s,i)$ only at x_1. It follows that $p(y,k)$ does not intersect $p(s,i)$ at all.

Induction Case: Assume that the antecedent is true, and that the theorem is true for all $m < n$. If there is a path in $P_k(S)$ that does not intersect p(s,i), the proof is done. Suppose then that every path in $P_k(S)$ intersects p(s,i) and that the set of paths in $P_k(S)$ intersects p(s,i) at exactly n distinct vertices. Let $p(x_{max},i)$ be a final segment of p(s,i). Since not every path in $P_k(S)$ intersects p(s,i) at x_{max}, there is a point of intersection prior to x_{max} on p(s,i). Hence the number of distinct points of intersection of the paths in $P_k(S)$ with $p(x_{max},i)$ is less than n. By the induction hypothesis, there is a path p_1 with source x_{max} and a path p1' with a source s' in the sources of $P_k(S)$, such that one of p1 and p1' has a sink i, the other has sink k, and p1 and p1' do not intersect. Suppose w.l.g. that p_1 has sink i and p_1' has sink k. Since p_1' does not contain x_{max}, its first point of intersection with p(s,i) is some vertex x_r, which occurs on p(s,i) before x_{max} (by definition of x_{max}.) Let $p(x_r,k)$ be a final segment of p1', p(s",k) be a path in $P_k(S)$ whose first point of intersection with p(s,i) is x_{max}, and p(s", x_{max}) an initial segment of p(s",k). There are two cases.

1. $p(x_r,k)$ does not intersect p(s", x_{max}). Let p = $p(s,x_r)$&$p(x_r,k)$ and p' = p(s", x_{max})&p_1. p and p' do not intersect for reasons analogous to those in case 1 of the base case (with x_r substituted for x_1, and x_{max} substituted for x_2; see Fig. 2.5.2.)

2. $p(x_r,k)$ does intersect p(s",x_{max}), and the last point of intersection is y. $y \neq x_{max}$ because it lies on $p(x_r,k)$ and $p(x_r,k)$ does not contain x_{max}. Let p(y,k) be a final segment of $p(x_r,k)$. There are two cases.

   a. p(y,k) intersects p(s,x_{max}) and the first point of intersection is z. Let p(s",y) be an initial segment of p(s",x_{max}), p(y,z) an initial segment of p(y,k), and p(s,z) an initial segment of p(s,i). $z \neq x_{max}$ because p(y,k) does not intersect x_{max}. (See Fig. 2.5.4.)

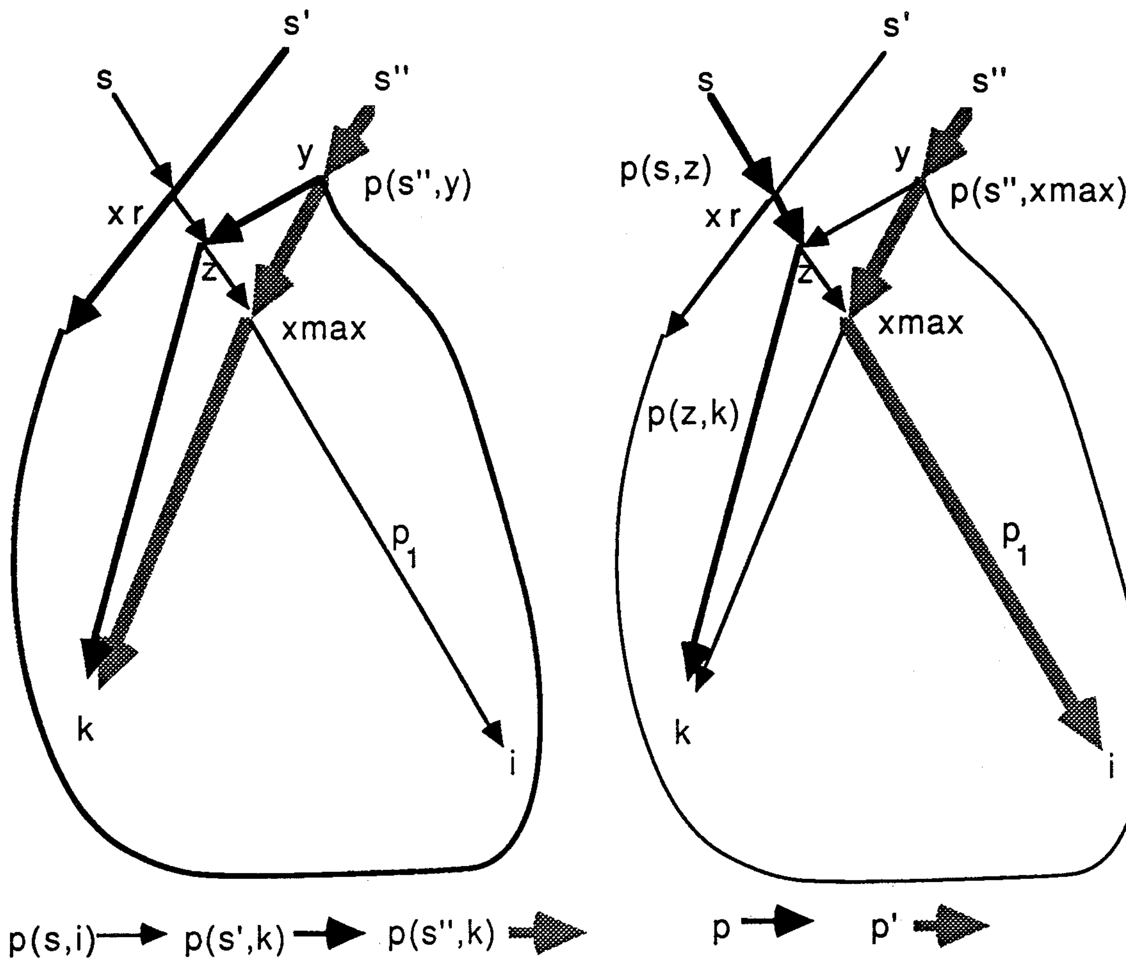


Fig. 2.5.4

I will now prove z is not after x_{max}. Consider the path p(s",y)&p(y,z). p(s",y) does not intersect p(s,i) because y occurs before x_{max}, p(s",y) is an initial segment of p(s",k) and the first point of intersection of p(s,i) and p(s",k) is x_{max}. The first point of intersection of p(y,z) and p(s,i) is z, since p(y,z) is an initial segment of p(y,k) and z is the first point of intersection of p(y,k) and p(s,i). Hence the first point of intersection of p(s",y)&p(y,z) with p(s,i) is z. p(s",y)&p(y,z) is an initial segment of a path from s" to k that is in $P_k(S)$. It follows that there is a path in $P_k(S)$ whose first point of intersection with p(s,i) is z. If z is after x_{max}, then there is a path in $P_k(S)$ whose first point of intersection with p(s,i) is after x_{max}, contrary to the definition of x_{max}.

Let p = p(s,z)&p(z,k) and p' = p(s",x_{max})&p_1. p(s,z) does not intersect p(s",x_{max}) since p(s",x_{max}) is an initial segment of p(s",k) and p(s,z) is an initial segment of p(s,i) and the first point of intersection of p(s,i) and p(s",k) is x_{max}. p(s,z) does not intersect p_1 (which has source x_{max}) since z occurs before x_{max} and the graph is acyclic. p(z,k) does not intersect p_1 since p(z,k) is a subpath of p_1' that does not intersect p_1 by construction. p(z,k) does not intersect p(s",x_{max}) since p(z,k) is a final segment of p(x_r,k), z is after y, and y is the last point of intersection of $p(x_r,k)$ and $p(s'',x_{max})$.

b. $p(y,k)$ does not intersect $p(s,x_{max})$. (This is similar to part 2 of the Base case, with x_{max} substituted for x2. See Fig. 2.5.3.) Let $p' = p(s'',y)\&p(y,k)$ and $p = p(s,x_{max})\&p_1$. I have already shown that $p(s'',y)$ does not intersect p(s,i) and $p(s,x_{max})$ is an initial segment of p(s,i). p(s",y) does not intersect p_1 because y is before x_{max}, and the graph is acyclic. p(y,k) does not intersect $p(s,x_{max})$ by hypothesis, and p(y,k) does not intersect p_1 because it is a subpath of p_1' that does not intersect p_1 by construction.

∎

Definition 2.5.1: In an acyclic directed graph G, if all l(t(k,l)) and all j(t(i,j)) intersect at a vertex P, then P is an **lj(t(i,j),t(k,l)) choke point**. Similarly, if all l(t(k,l)) and all j(t(i,j)) intersect at a vertex P, and all l(t(i,l)) and all j(t(j,k)) also intersect at P, then P is a **lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point.**

Lemma 2.5.3: In an acyclic graph G, if there is no lj(t(i,j),t(k,l)) choke point, then either there is a trek t'(k,l) such that there is no vertex p' that occurs in the intersection of all j(t(i,j)) with l(t'(k,l)), or there is a trek t'(i,j) such that there is no vertex p' that occurs in the intersection of all l(t(kl)) with j(t'(i,j))

Proof. Suppose that the lemma is false. Then, for each trek t'(k,l) there is a non-empty set of points P(t'(k,l)) such that every point in P(t'(k,l)) is in the intersection of all j(t(i,j)) with l(t'(k,l)). Similarly, for each trek t'(i,j) there is a non-empty set of points P(t'(i,j)) such that every point in P(t'(i,j)) is in the intersection of all l(t(k,l)) with j(t'(i,j)). Every j(t(i,j)) contains every vertex in $\bigcup_{t(k,l) \in T(k,l)} P(t(k,l))$ (since every j(t(i,j)) intersects each l(t'(k,l)) at some vertex in P(t'(k,l))), and every vertex in $\bigcup_{t(i,j) \in T(i,j)} P(t(i,j))$ occurs on some trek l(t'(k,l)). Similarly, every l(t(k,l)) contains every vertex in $\bigcup_{t(i,j) \in T(i,j)} P(t(i,j))$. Furthermore, for every vertex in $\bigcup_{t(k,l) \in T(k,l)} P(t(k,l))$ there is some l(t'(k,l)) that does not contain it (else all j(t(i,j)) and all l(t(k,l)) intersect at a single vertex), and some l(t"(k,l)) that does contain it. Similarly, for every vertex in $\bigcup_{t(i,j) \in T(i,j)} P(t(i,j))$ there is some j(t'(i,j)) that does not contain it and some j(t"(i,j)) that does contain it.

Since every vertex in $\bigcup_{t(k,l) \in T(k,l)} P(t(k,l))$ occurs on every j(t(i,j)), they can be ordered by the order of their occurrence on j(t(i,j)); similarly every vertex in $\bigcup_{t(i,j) \in T(i,j)} P(t(i,j))$ can be ordered. By the antecedent of the lemma, there are at least two vertices in each of $\bigcup_{t(k,l)\,\varepsilon\,T(k,l)} P(t(k,l))$ and

$\bigcup_{t(i,j)\,\varepsilon\,T(i,j)} P(t(i,j))$.

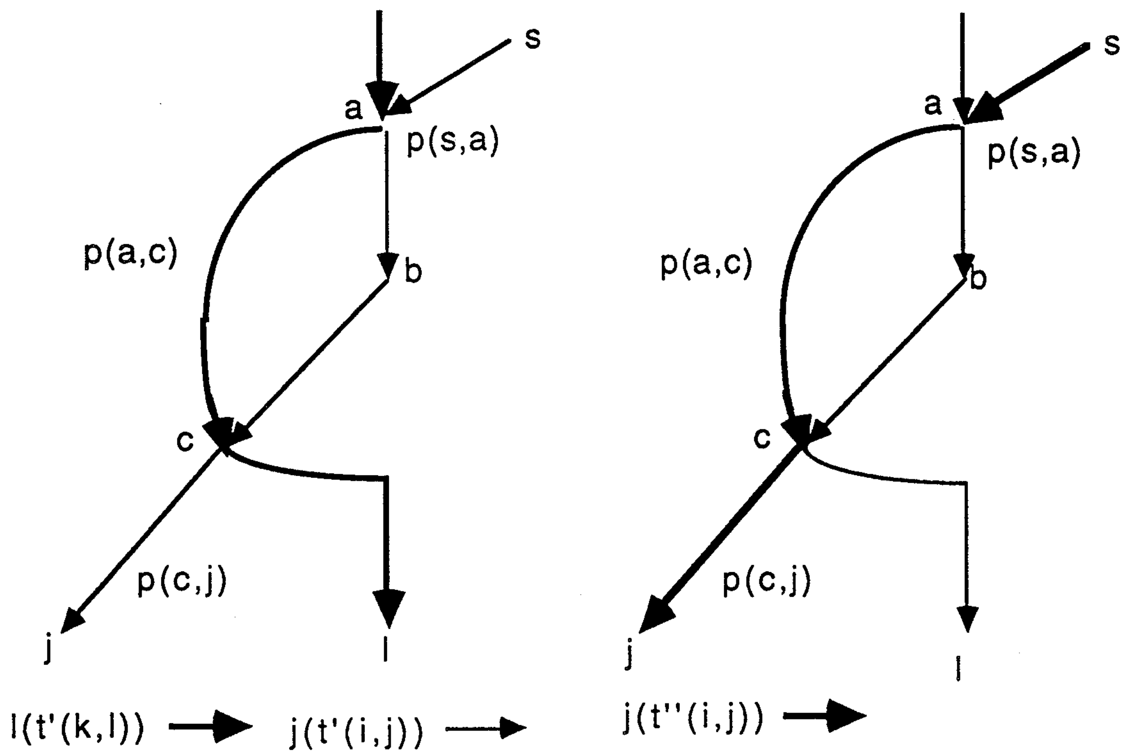


Fig. 2.5.5

See Fig. 2.5.5. Let a be the first vertex in $\bigcup_{t(i,j)\,\varepsilon\,T(i,j)} P(t(i,j))$ and b be the first vertex in

$\bigcup_{t(k,l)\,\varepsilon\,T(k,l)} P(t(k,l))$. Suppose w.l.g. that a is before b. There exists an l(t'(k,l)) that contains a (since every l(t(k,l)) contains a), that does not contain b, but that does contain some vertex c ($\neq$ b)

in $\bigcup_{t(k,l)\,\varepsilon\,T(k,l)} P(t(k,l))$.

There is also a j(t'(i,j)) that contains a. Let s be the source of t'(i,j), p(s,a) an initial segment of j(t'(i,j)), p(a,c) a segment of l(t'(k,l)), and p(c,j) a final segment of j(t'(i,j)). Let j(t"(i,j)) = p(s,a)&p(a,c)&p(c,j), and i(t"(i,j)) = i(t'(i,j)). j(t"(i,j)) does not contain b for the following reasons:

a. p(s,a) does not contain b because a occurs before b.

b. p(a,c) does not contain b because it is a segment of l(t'(k,l)) which does not contain b.

c. p(c,j) does not contain b because it is a segment of j(t'(i,j)), and since b is the first vertex in

$$\bigcup_{t(k,l)\,\varepsilon\, T(k,l)} P(t(k,l))$$ it occurs before c on j(t'(i,j)).

But this contradicts the fact that for every t(i,j), j(t(i,j)) contains b. ∴

Lemma 2.5.4. In an acyclic graph G, if there is no ik(t(i,j),t(k,l)) choke point, then either there is a trek t'(k,l) such that there is no vertex p' that occurs in the intersection of all i(t(i,j)) with k(t'(k,l)), or there is a trek t'(i,j) such that there is no vertex p' that occurs in the intersection of all k(t(kl)) with i(t'(i,j))

Proof. The proof of lemma 2.5.4 is the same as that of lemma 2.5.3 with i, j, k, l permuted.

Lemma 2.5.5: In an acyclic LCF G, if there is a trek t'(k,l) such that there is no vertex p' that occurs in the intersection of all j(t(i,j)) with l(t'(k,l)), then either there are treks t"(i,j) and t"(k,l) such that j(t"(i,j)) does not intersect l(t"(k,l)) or $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk}$ is not linearly implied by G.

Proof. Let s be the source of t'(k,l), and S be the set of sources of treks between i and j. By lemma 2.5.2 it is possible to construct a pair of paths p and p', with sources s and s' (in S), and sinks j and l, such that p and p' do not intersect. There are two cases.

1. If p is a path from s to l, and p' is a path from s' to j, then the following treks can be formed. (See Fig. 2.5.6.) j(t"(i,j)) = p', i(t"(i,j)) = i(t'(i,j)), k(t"(k,l)) = k(t'(k,l)), and l(t"(k,l)) = p. By construction p does not intersect p'; hence j(t"(i,j)) does not intersect l(t"(k,l)).

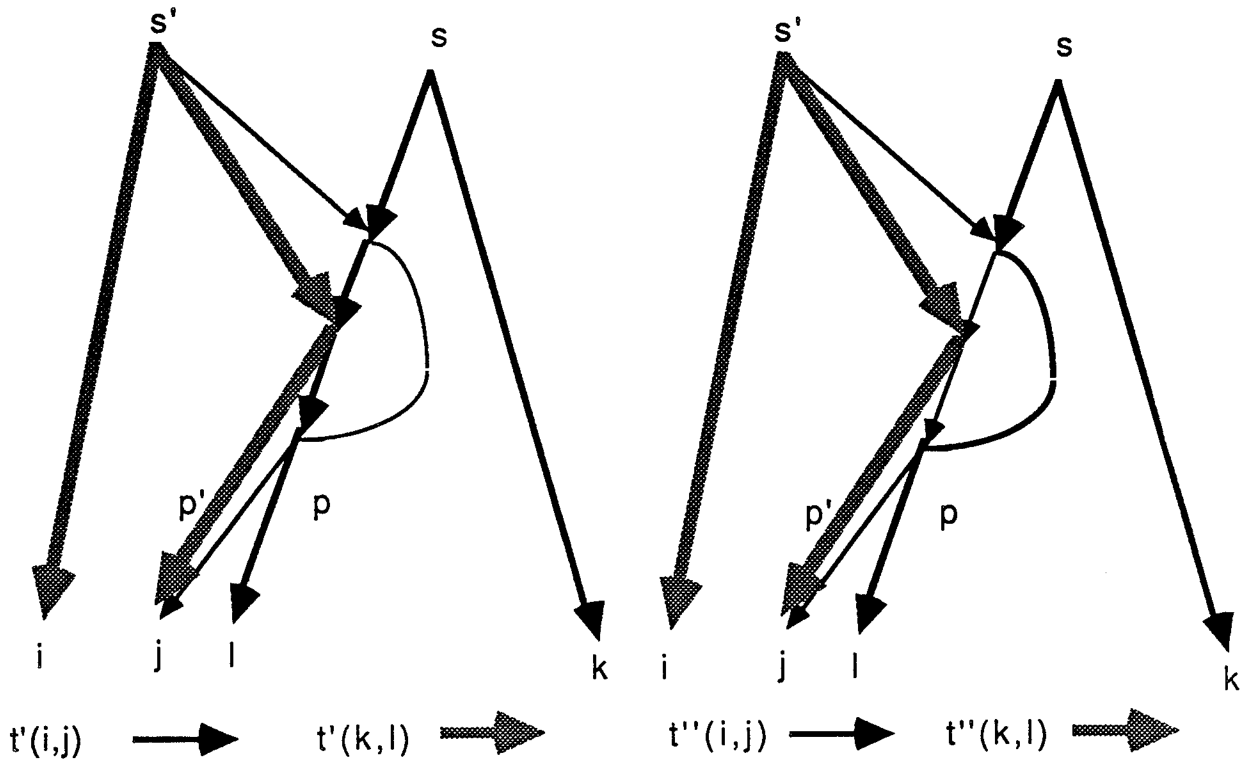


Fig. 2.5.6.

2. If p is a path from s to j, and p' is a path from s' to l, there are two cases.

   a. k(t'(k,l)) intersects i(t'(i,j)), and the first vertex of intersection is y. Let p(s,y) be an initial segment of k(t'(k,l)), p(y,k) a final segment of k(t'(k,l)), p(s',y) an initial segment of i(t'(i,j)), p(y,i) a final segment of i(t'(i,j)), j(t"(i,j)) = p, i(t"(i,j)) = p(s,y)&p(y,i), k(t"(k,l)) = p(s',y)&p(y,k), and l(t"(k,l)) = p'. (See Fig. 2.5.7.) By construction, j(t"(i,j)) and l(t"(k,l)) do not intersect.

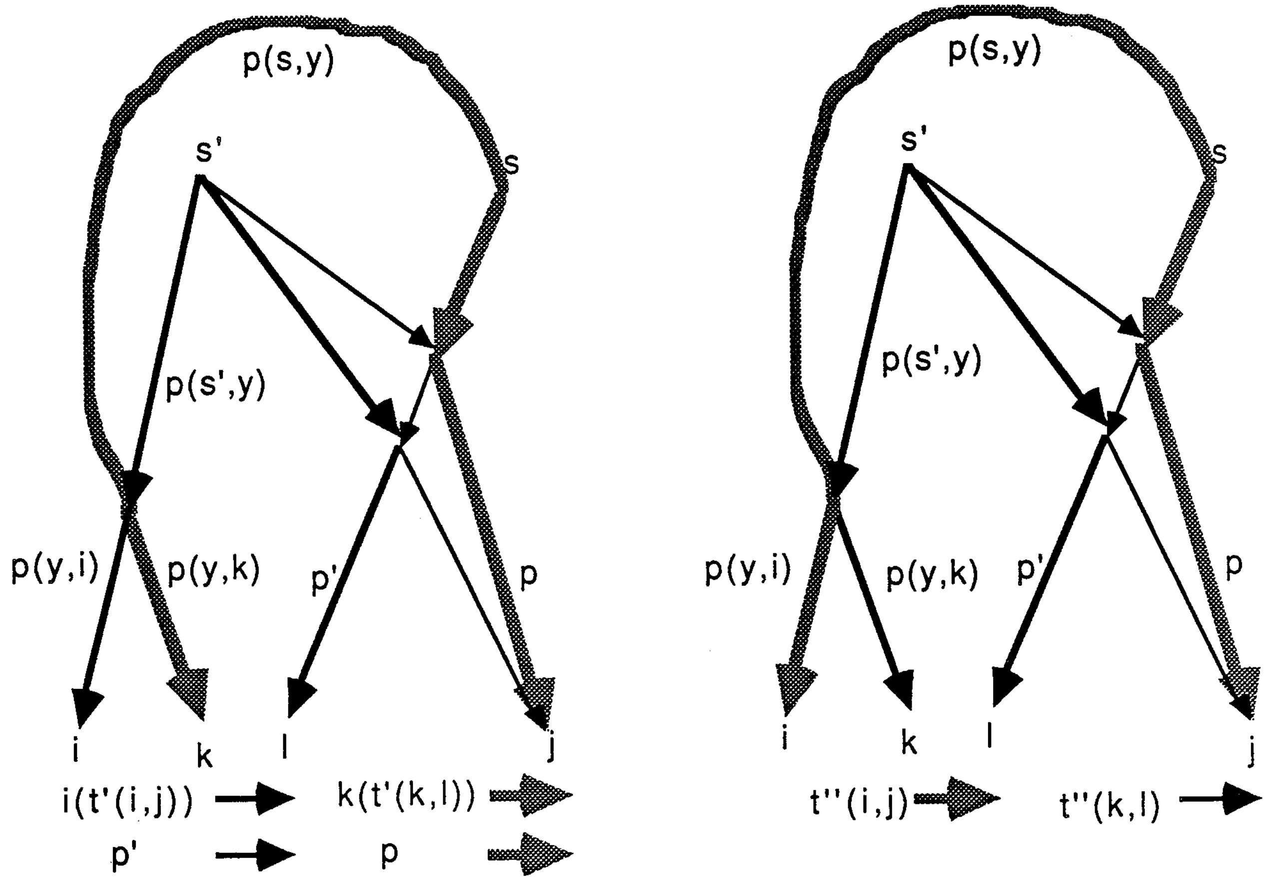


Fig. 2.5.7

b. If k(t'(k,l)) does not intersect i(t'(i,j)), the following treks can be formed. (See Fig. 2.5.8.) i(t'(i,l)) = i(t'(i,j)), l(t'(i,l)) = p', j(t'(j,k)) = p, and k(t'(j,k)) = k(t'(k,l)). By hypothesis, k(t'(j,k)) does not intersect i(t'(i,l)). By construction, l(t'(i,l)) does not intersect j(t'(j,k)). Hence by theorem 2.4.1, $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk}$ is not linearly implied by G. $\therefore$

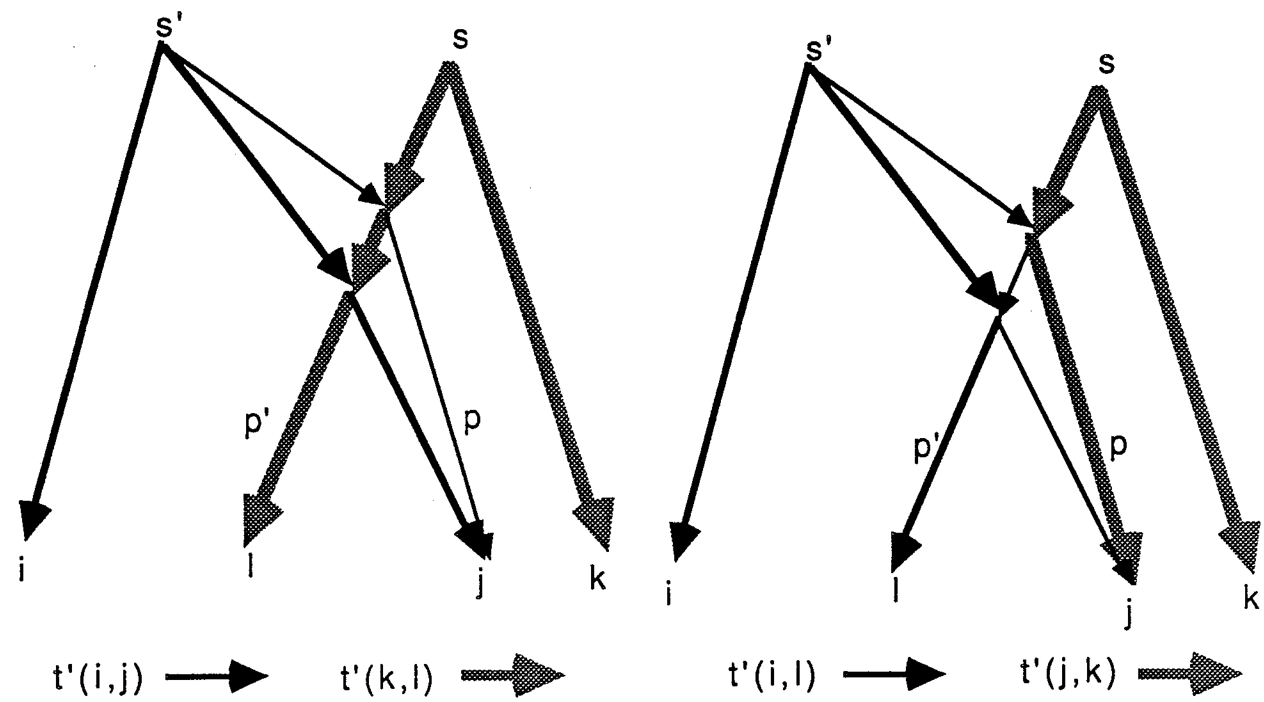


Fig. 2.5.8

Lemma 2.5.6: In an acyclic LCF G, if there is a trek t'(i,j) such that there is no vertex p' that occurs in the intersection of all l(t(k,l)) with j(t'(i,j)), then either there are treks t"(i,j) and t"(k,l) such that j(t"(i,j)) does not intersect l(t"(k,l)) or $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

Lemma 2.5.7: In an acyclic LCF G, if there is a trek t'(i,j) such that there is no vertex p' that occurs in the intersection of all k(t(k,l)) with i(t'(i,j)), then either there are treks t"(i,j) and t"(k,l) such that i(t"(i,j)) does not intersect k(t"(k,l)) or $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

Lemma 2.5.8: In an acyclic LCF G, if there is a trek t'(k,l) such that there is no vertex p' that occurs in the intersection of all i(t(i,j)) with k(t'(k,l)), then either there are treks t"(i,j) and t"(k,l) such that i(t"(i,j)) does not intersect k(t"(k,l)) or $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

The proofs of lemmas 2.5.6, 2.5.7, and 2.5.8 can all be obtained from the proof of lemma 2.5.5 by permuting i, j, k, and l.

Lemma 2.5.9: In an acyclic LCF G, if there is no lj(t(i,j),t(k,l)) choke point, and there is no ik(t(i,j),t(k,l)) choke vertex, then there exist treks t'(i,j) ,t'(k,l), t"(i,j), and t"(k,l) such that i(t'(i,j)) does not intersect k(t'(k,l)) and j(t"(i,j)) does not intersect l(t"(k,l)), or $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

Proof. This follows directly from lemmas 2.5.3 through 2.5.8. ∴

Lemma 2.5.10: In an acyclic LCF G, if there is no lj(t(i,j),t(k,l)) choke point, and there is no ik(t(i,j),t(k,l)) choke point, then $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

Proof. Assume that there is no lj(t(i,j),t(k,l)) choke point, and there is no ik(t(i,j),t(k,l)) choke point.

By lemma 2.5.9 either $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G or there exist treks t'(i,j), t'(k,l), t"(i,j), and t"(k,l) such that i(t'(i,j)) does not intersect k(t'(k,l)) and j(t"(i,j)) does not intersect l(t"(k,l)). If $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G, the proof is done. Assume then that there exist treks t'(i,j), t'(k,l), t"(i,j), and t"(k,l) such that i(t'(i,j)) does not intersect k(t'(k,l)) and j(t"(i,j)) does not intersect l(t"(k,l)). There are three cases.

1. Suppose for all t(i,j), t(k,l), j(t(i,j)) intersects l(t'(k,l)) at each vertex in a non-empty set of vertices P', and all l(t(k,l)) intersects j(t'(i,j)) at each vertex in a non-empty set of vertices P. Hence, all l(t(k,l)) contain every vertex in P and all j(t(i,j)) contain every vertex in P'. Since there is no lj(t(i,j),t(k,l)) choke point, there is no vertex Z such that for all t(i,j) and all t(k,l), Z occurs in the intersection of l(t(i,j)) and j(t(i,j)). Hence P and P' do not intersect.

Let a be the first vertex in P, and b be the first vertex in P'. Suppose w.l.g. that a occurs before b. Let s'(i,j) be the source of t'(i,j), s'(k,l) the source of t'(k,l) and s"(i,j) the source of t"(i,j), and s"(k,l) the source of t"(k,l). l(t"(k,l)) contains a (since all l(t(k,l)) contain a), and j(t"(i,j)) contains b, (since all j(t(i,j)) contain b.) There are two cases.

a. Suppose k(t"(k,l)) does not intersect i(t"(i,j)). Then, since k(t"(k,l)) does not intersect i(t"(i,j)) and j(t"(k,l)) does not intersect l(t"(k,l)), by theorem 2.4.1, $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

b. Suppose k(t"(k,l)) does intersect i(t"(i,j)) at a vertex x. (See Fig. 2.5.9.) Let p(s"(i,j),x) be an initial segment of i(t"(i,j)), p(x,k) a final segment of l(t"(k,l)). Let p(s"(i,j),b) be an initial segment of j(t"(k,l)) and p(b,l) be a final segment of l(t'(k,l)). Form the trek k(t'"(k,l)) = p(s"(i,j),x)&p(x,k), and l(t'"(k,l)) = p(s"(i,j),b)&p(b,l). p(s"(i,j),b) does not contain a, since it is a subpath of j(t"(i,j)) which does not intersect l(t"(k,l)), which does contain a. p(b,l) does not contain a, since a occurs after b. Hence l(t'"(k,l)) does not contain a; but this is a contradiction.

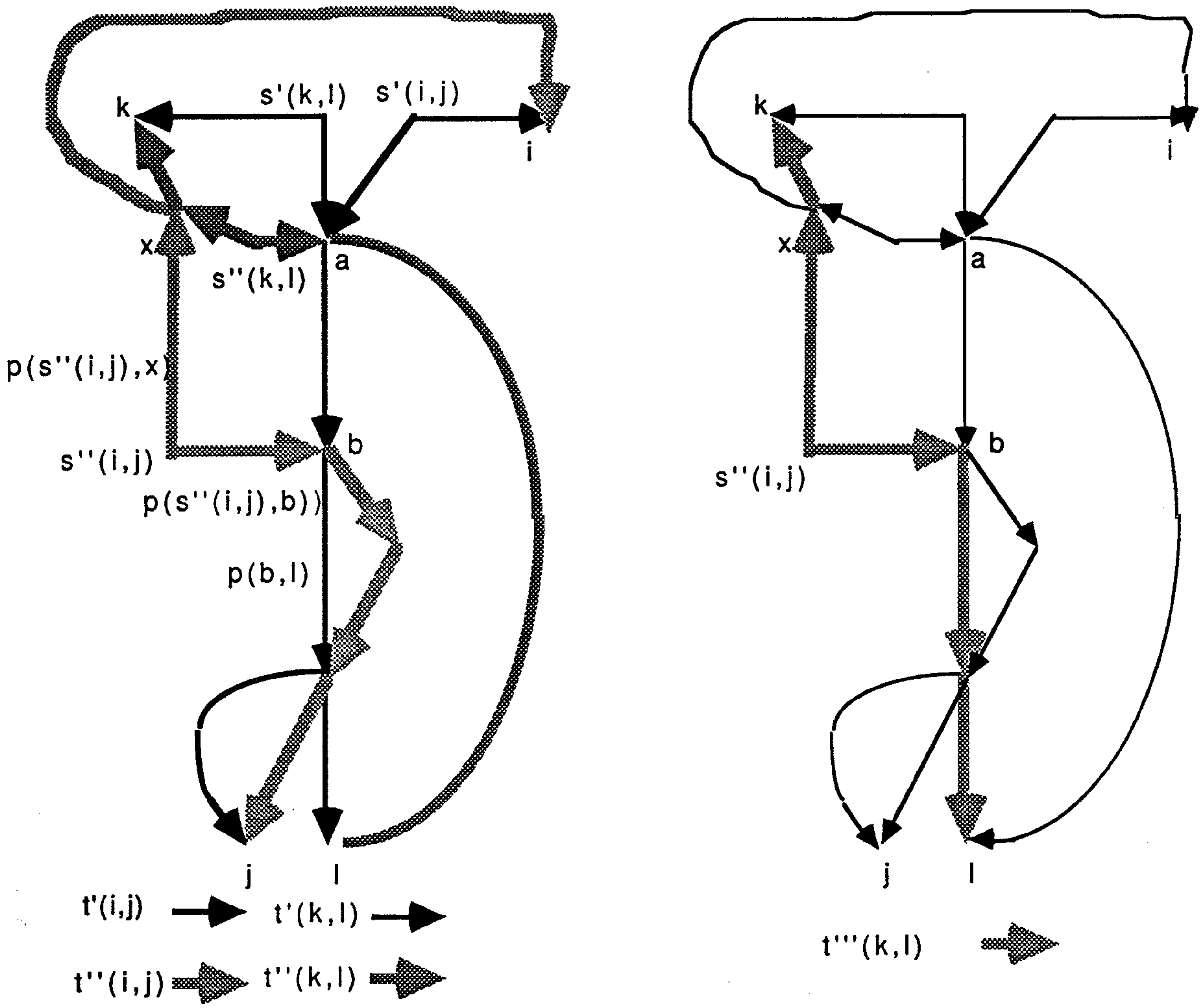


Fig. 2.5.9

2. All l(t(k,l)) intersect j(t'(i,j)), but not at a single vertex, or all j(t(i,j)) intersect l(t'(k,l)) but not at a single vertex. Assume w.l.g. that the latter is the case. Let s' be the source of t'(i,j) and s be the source of t(k,l). Let S be the set of sources of treks between i and j. By lemma 2.5.2, it is possible to form two paths p(s",l) and p(s,j) or p(s",j) and p(s,l) that don't intersect, where s" is in S. Assume that it is possible to form the paths p(s",l) and p(s,j) that don't intersect. (If the paths that don't intersect are p(s",j) and p(s,l) the proof is the same except that the indices are permuted.) Let t"(i,j) be a trek with source s" (See Fig. 2.5.10.) Let the first point of intersection of i(t"(i,j)) with i(t'(i,j)) be r. There are two cases.

a. i(t"(i,j)) does not intersect k(t'(k,l)) before it intersects i(t'(i,j)) at r. (See Fig. 2.5.10.) Let p(r,i) be a final segment of i(t'(i,j)) and p(s",r) be an initial segment of i(t"(i,j)). Let i(t'(i,l)) = p(s",r)&p(r,i), l(t'(i,l)) = p(s",l), j(t'(j,k)) = p(s,j) and k(t'(j,k)) = k(t'(k,l)). p(s",r) and p(r,i) do not intersect k(t'(k,l)) by hypothesis. By theorem 2.4.1 $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

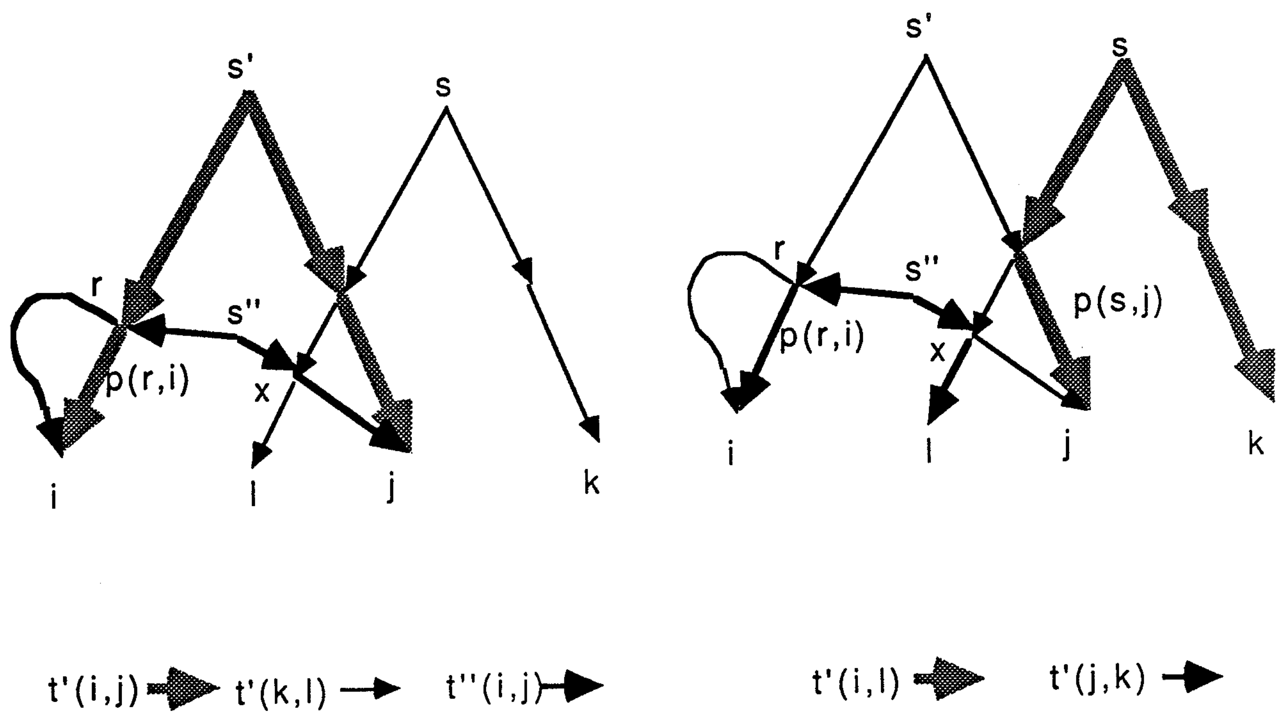


Fig. 2.5.10

b. i(t"(i,j)) does intersect k(t'(k,l)) before it intersects i(t'(i,j)), and the first point of intersection is q. Let p(q,k) be a final segment of k(t'(k,l)) and p(s",q) be an initial segment of i(t"(i,j)). Let y be the first point of intersection of p(s,j) and j(t'(i,j)), and p(s',y) be an initial segment of j(t'(i,j)). There are two cases.

1. p(s",l) intersects p(s',y) and the first point of intersection is z. Let p(s',z) be an initial segment of j(t'(i,j)), p(z,l) be a final segment of p(s",l), l(t'(i,l)) = p(s',z)&p(z,l), i(t'(i,l)) = i(t'(i,j)), j(t'(j,k)) = p(s,j), and k(t'(j,k)) = k(t'(k,l)). (See Fig. 2.5.11.)

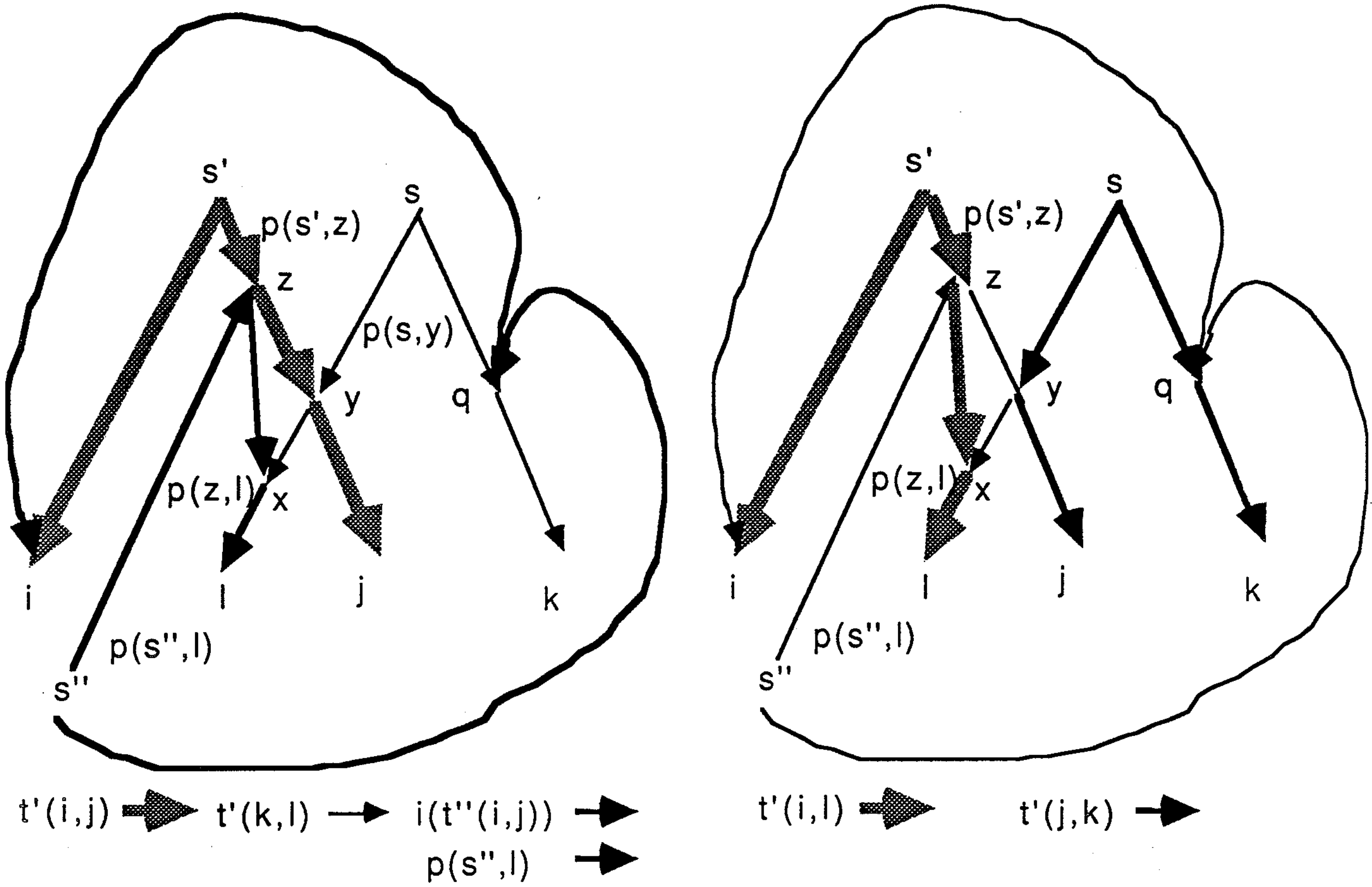


Fig. 2.5.11

k(t'(j,k)) does not intersect i(t'(i,l)) by hypothesis. j(t'(j,k)) does not intersect l(t'(i,l)) for the following reasons. p(s',z) does not intersect p(s,j) because p(s',z) is a subpath of j(t'(i,j)), z is before y, and the first point of intersection of j(t'(i,j)) and p(s,j) is y. p(z,l) does not intersect p(s,j) because it is a subpath of p(s",l) which does not intersect p(s,j) by construction. By theorem 2.4.1. $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

2. p(s",l) does not intersect p(s',y). Let l(t"(k,l)) = p(s",l), k(t"(k,l)) = p(s",q)&p(q,k), i(t'"(i,j)) = i(t'(i,j)), and j(t'"(i,j)) = p(s',y)&p(y,j). (See Fig. 2.5.12.) k(t"(k,l)) does not intersect i(t'"(i,j)) for the following reasons. p(s",q) does not intersect i(t'(i,j)) since p(s",q) is an initial segment of i(t"(i,j)), and q occurs before the first point of intersection of i(t"(i,j)) and i(t'(i,j)). p(q,k) does not intersect i(t'(i,j)) because it is a final segment of k(t'(k,l)), which does not intersect i(t'(i,j)) by hypothesis. l(t"(k,l)) does not intersect j(t'"(i,j)) for the following reasons. p(s',y) does not intersect p(s",l) by hypothesis, and p(y,j) is a subpath of p(s,j) which does not intersect p(s",l) by construction. By theorem 2.4.1 $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

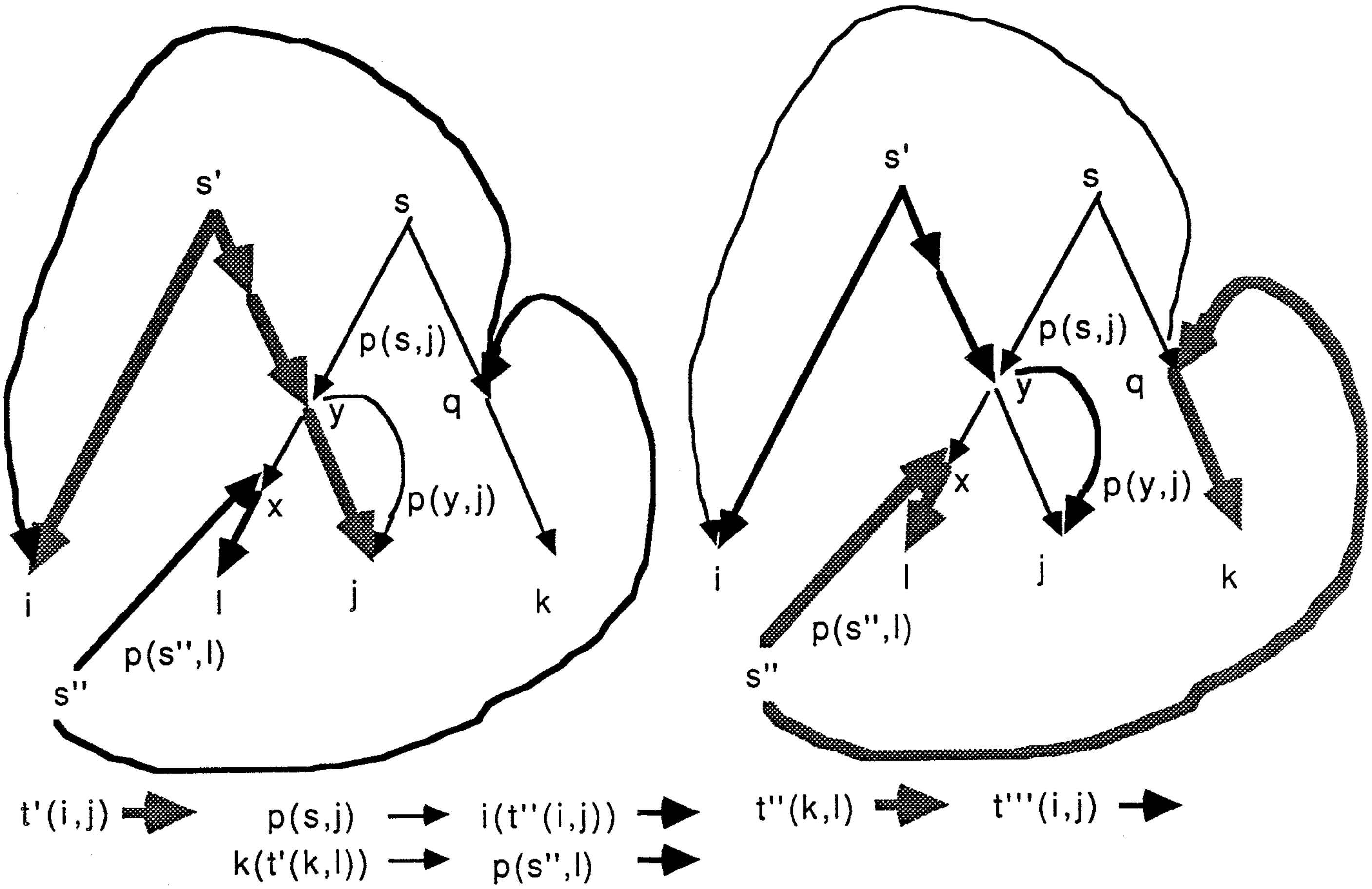


Fig. 2.5.12

3. Either there is an l(t''(k,l)) that does not intersect j(t'(i,j)) or there is a j(t''(i,j)) that does not intersect l(t'(k,l)). Assume w.l.g. that j(t''(i,j)) with source s''(i,j) does not intersect l(t'(k,l)). There are two cases.

a. Suppose that i(t''(i,j)) does not intersect k(t'(k,l)) before it intersects i(t'(i,j)) at vertex x. See Fig. 2.5.13.

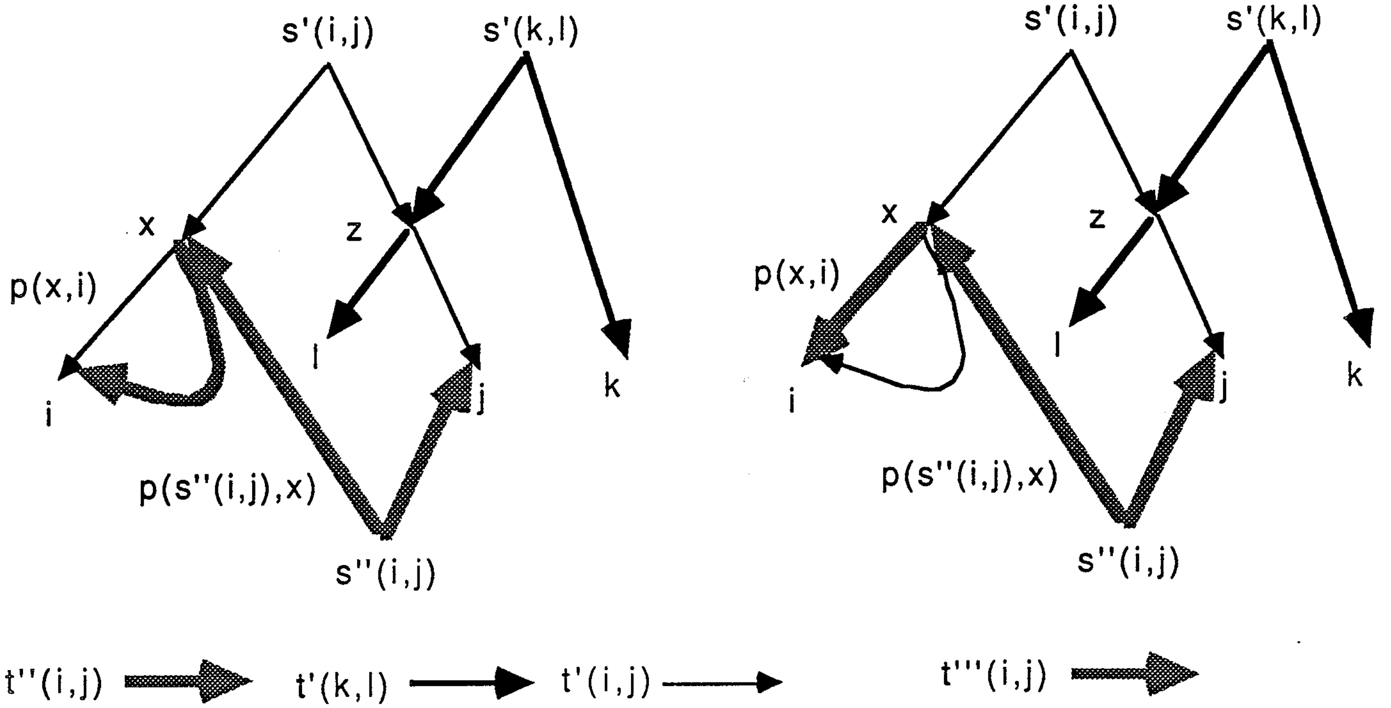


Fig. 2.5.13

Let p(x,i) be a final segment of i(t'(i,j)) and p(s"(i,j),x) be an initial segment of i(t"(i,j)). The trek t'"(i,j) can be formed as follows. j(t'"(i,j)) = j(t"(i,j)) and i(t'"(i,j)) = p(s"(i,j),x)&p(x,i). p(s"(i,j),x) does not intersect k(t'(k,l)) because by hypothesis x occurs on i(t"(i,j)) before it intersects k(t'(k,l)). p(x,i) does not intersect k(t'(k,l)) because it is a subpath of i(t'(i,j)) which does not intersect k(t'(k,l)) by hypothesis. Hence i(t'"(i,j)) does not intersect k(t'(k,l)). j(t'"(i,j)) = j(t"(i,j)) does not intersect l(t'(k,l)) by hypothesis. By theorem 2.4.1, $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.

b. Suppose i(t"(i,j)) intersects k(t'(i,j)) at y before it intersects i(t'(i,j)) at x. Let z be the first point of intersection of j(t'(i,j)) and l(t'(k,l)). (If no such vertex exists, then j(t'(i,j)) and l(t'(k,l)) do not intersect, i(t'(i,j)) and k(t'(k,l)) do not not intersect by hypothesis, and by theorem 2.4.1 $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.). Let p(s'(i,j),z) be an initial segment of i(t'(i,j)), and p(z,l) be a final segment of l(t'(k,l)). There are two cases.

1. Suppose that j(t"(i,j)) does not intersect p(s'(i,j),z). See Fig. 2.5.14.

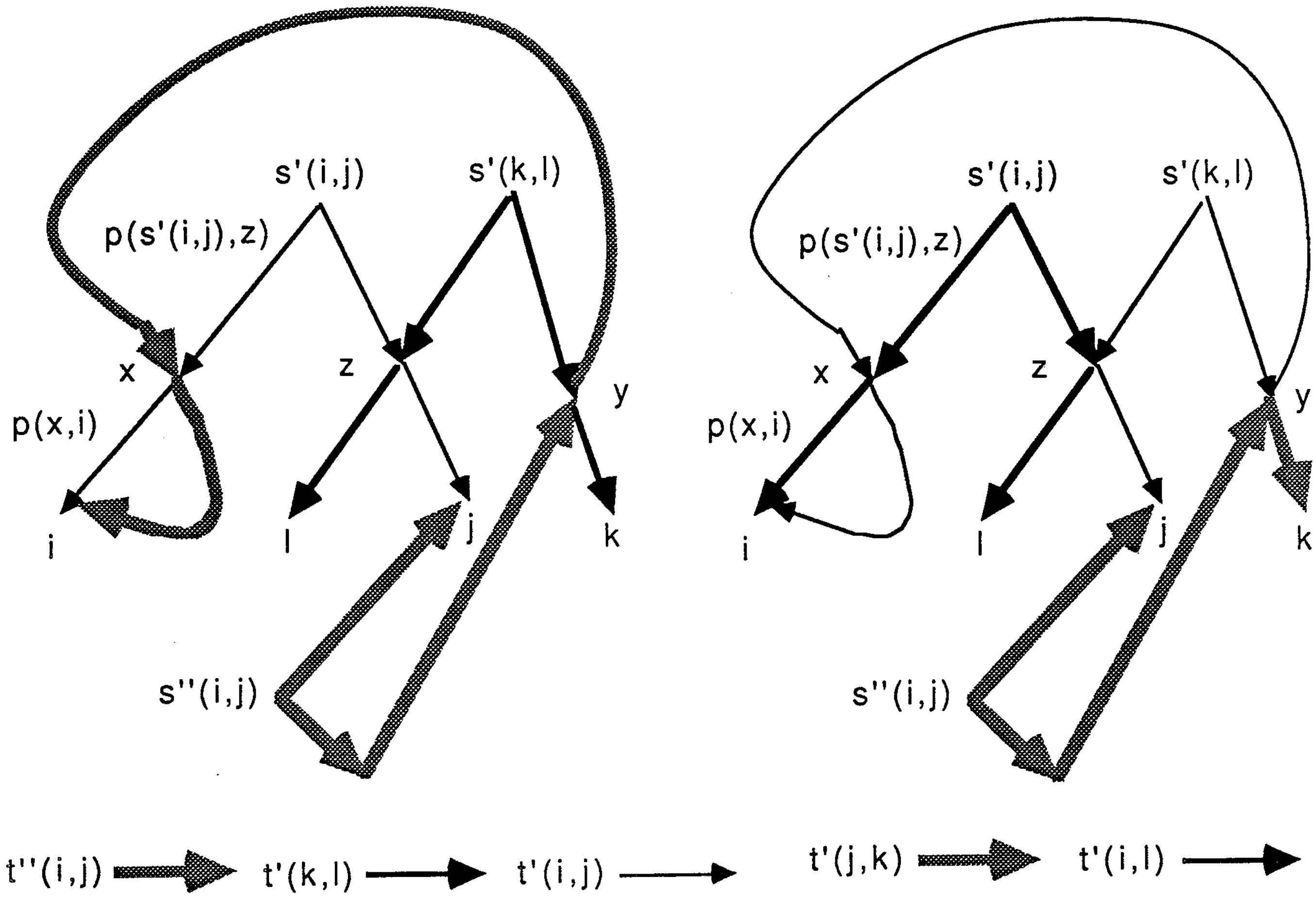


Fig. 2.5.14

Let p(y,k) be a final segment of k(t'(k,l)) and p(s''(i,j),y) be an initial segment of i(t''(i,j)). Let j(t'(j,k)) = j(t''(i,j)), k(t'(j,k)) = p(s'(i,j),y)&p(y,k), i(t'(i,l)) = i(t'(i,j)), l(t''(i,l)) = p(s'(i,j),z) & p(z,l). i(t'(i,l)) and k(t'(j,k)) do not intersect for the following reasons. i(t'(i,l)) does not intersect p(s''(i,j),y) because by hypothesis, i(t''(i,j)) intersects k(t'(k,l)) at y before it intersects i(t'(i,j)). i(t'(i,l)) does not intersect p(y,k) because i(t'(i,l)) = i(t'(i,j)) and p(y,k) is a subpath of k(t'(k,l)), which does not intersect i(t'(i,j)) by hypothesis. j(t'(j,k)) does not intersect l(t'(i,l)) for the following reasons. j(t'(j,k)) does not intersect p(s'(i,j),z) because j(t'(j,k)) = j(t''(i,j)), which does not intersect p(s'(i,j),z) by hypothesis. j(t'(j,k)) does not intersect p(z,l) because j(t'(j,k)) = j(t''(i,j)) which does not intersect l(t'(k,l)) (which contains p(z,l)) by hypothesis.

2. Suppose that j(t''(i,j)) does intersect p(s'(i,j),z) and the first point of intersection is r. (See Fig. 2.5.15.) Let p(s'(i,j),r) be an initial segment of j(t'(i,j)) and p(r,j) be a final segment of j(t''(i,j)). Let i(t'''(i,j)) = i(t'(i,j)) and j(t'''(i,j)) = p(s'(i,j),r)&p(r,j). i(t'''(i,j)) does not intersect k(t'(k,l)) by hypothesis. j(t'''(i,j)) does not intersect l(t'(k,l)) for the following reasons. p(s'(i,j),r) does not intersect l(t'(k,l)) since r is before z on j(t'(i,j)), and the first point of intersection of j(t'(i,j)) with l(t'(k,l)) is z. p(r,j) does not intersect l(t'(k,l)) because it is a subpath of j(t''(i,j)) which does not intersect l(t'(k,l)) by hypothesis.

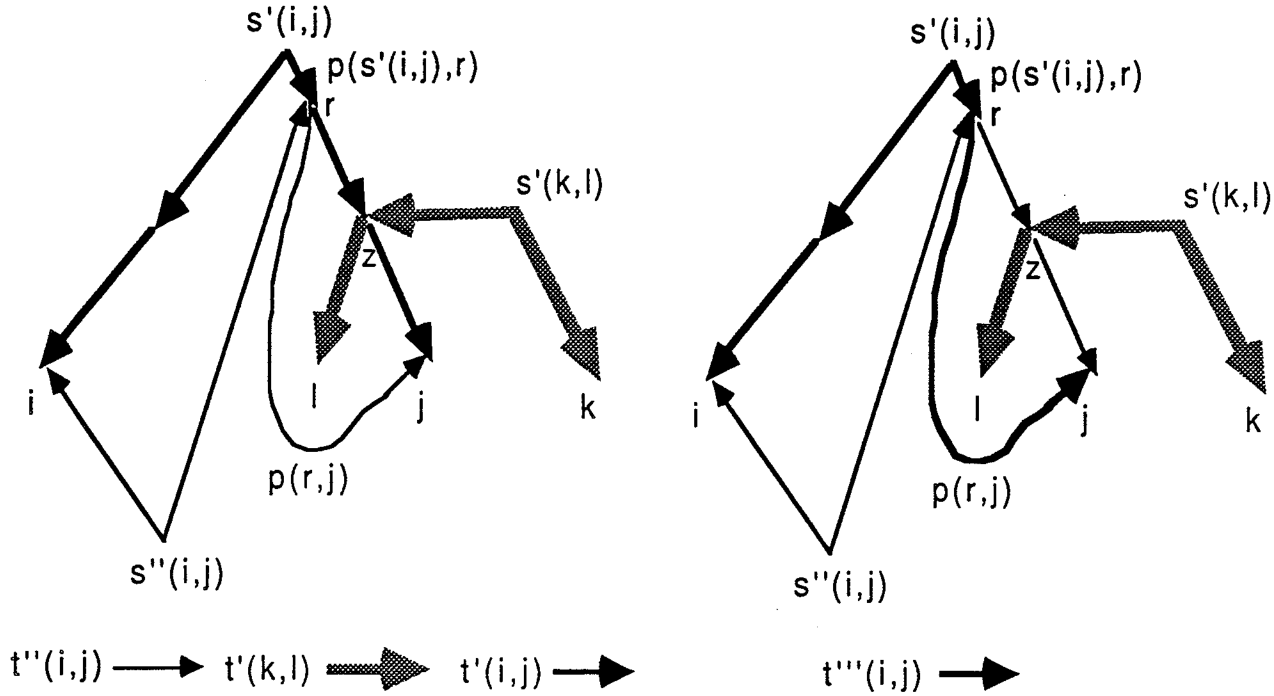


Fig. 2.5.15

∎

Lemma 2.5.11: In an acyclic LCF G, if there is no lj(t(i,l),t(j,k)) choke point, and there is no ik(t(i,l),t(j,k)) choke point, then $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G.
Proof. The proof is the same as that of lemma 2.5.10, with the indices permuted.

Lemma 2.5.12: In an acyclic LCF G, if G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$, then either there is an lj(t(i,j),t(k,l)) choke point and an lj(t(i,l),t(j,k)) choke point, or there is an ik(t(i,j),t(k,l)) choke point and an ik(t(i,l),t(j,k)) choke point.
Proof. Assume that G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$. By lemmas 2.5.10 and 2.5.11, if G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ then either there is an lj(t(i,j),t(k,l)) choke point or an ik(t(i,j),t(k,l)) choke point, and there is either an lj(t(i,l),t(j,k)) choke point or an ik(t(i,l),t(j,k)) choke point. If there is an lj(t(i,j),t(k,l)) choke point and an lj(t(i,l),t(j,k)) choke point, or there is an ik(t(i,j),t(k,l)) choke point and an ik(t(i,l),t(j,k)) choke point, the proof is done. Suppose then that there is an lj(t(i,j),t(k,l)) choke point and an ik(t(i,l),t(j,k)) choke point, but no ik(t(i,j),t(k,l)) choke point and no lj(t(i,l),t(j,k)) choke point. (The case where there is an lj(t(i,l),t(j,k)) choke point and an ik(t(i,j),t(k,l)) choke point, but no lj(t(i,j),t(k,l)) choke point and no ik(t(i,l),t(j,k)) choke point is essentially the same, with the indices permuted.)

By lemmas 2.5.3 through 2.5.8, if there is no lj(t(i,l),t(j,k)) choke point, then either there is a pair of treks t'(i,l) and t'(j,k) such that l(t'(i,l)) does not intersect j(t'(j,k)) or $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ is not linearly implied by G. Since the latter possibility contradicts our hypothesis, assume that there there is a pair of treks t'(i,l) and t'(j,k) such that l(t'(i,l)) does not intersect j(t'(j,k)). There are two cases.

If i(t'(i,l)) does not intersect k(t'(j,k)) then by theorem 2.4.1., G does not strongly imply $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$, contrary to our hypothesis. Suppose then that i(t'(i,l)) does intersect k(t'(j,k)) at a vertex y. (See Fig. 2.5.16.)

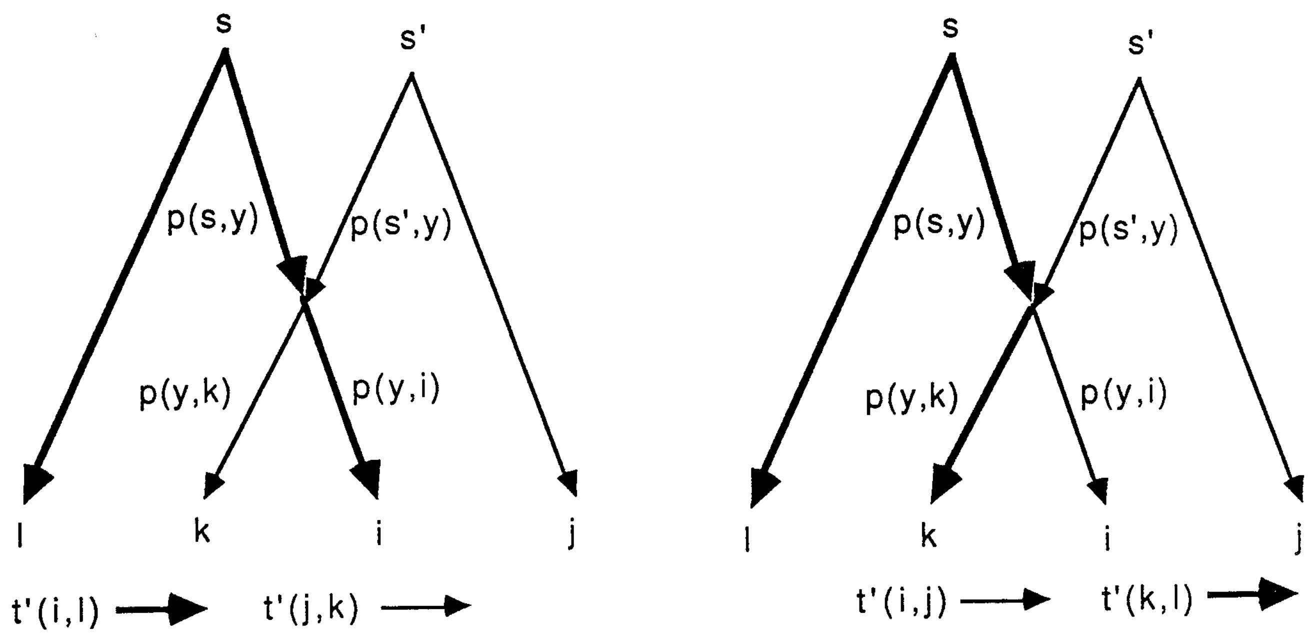


Fig. 2.5.16

Let s be the source of t'(i,l), s' the source of t'(j,k), p(s,y) an initial segment of i(t'(i,l)), p(y,k) a final segment of k(t'(j,k)), p(s',y) an initial segment of k(t'(j,k)), p(y,i) a final segment of i(t'(i,l)), i(t'(i,j)) = p(s',y)&p(y,i), j(t'(i,j)) = j(t'(j,k)), k(t'(k,l)) = p(s,y)&p(y,k), and l(t'(k,l)) = l(t'(i,l)). But since j(t'(i,j)) = j(t'(j,k)) does not intersect l(t'(k,l)) = l(t'(i,l)), there is no lj(t(i,j),t(k,l)) choke point, contrary to our hypothesis. ∴

Lemma 2.5.13: In an acyclic LCF G, if G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$, then either there is an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point, or there is an ik(t(i,j),t(k,l),t(i,l),t(j,k)) choke point.

Proof. Assume that G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$. By lemma 2.5.12, either there is an lj(t(i,j),t(k,l)) choke point and an lj(t(i,l),t(j,k)) choke point, or there is an ik(t(i,j),t(k,l)) choke point and an ik(t(i,l),t(j,k)) choke point. Suppose w.l.g. that the former is the case. If some lj(t(i,j),t(k,l)) choke point is also an lj(t(i,l),t(j,k)) choke point, the proof is done. Suppose then that no lj(t(i,j),t(k,l)) choke point is also an lj(t(i,l),t(j,k)) choke point. Let C be an lj(t(i,j)),t(k,l)) choke point. By hypothesis C is not an lj(t(i,l),t(j,k)) choke point, so there exist a pair of treks t'(i,l) and t'(j,k) with sources s and s' respectively, such that l(t'(i,l)) and j(t'(j,k)) do not intersect at C. (See Fig. 2.5.17.)

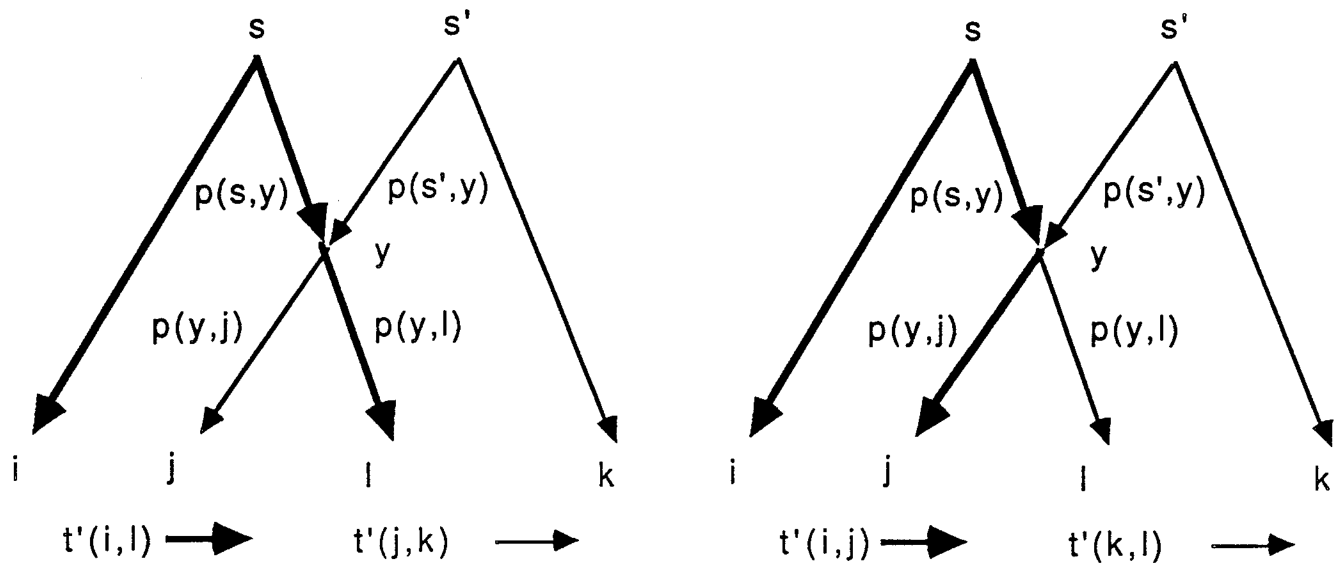


Fig.2.5.17

Hence there is at most one occurrence of C in the pair of paths l(t'(i,l)) and j(t'(j,k)). Since there is an lj(t(i,l),t(j,k)) choke point, l(t'(i,l)) and j(t'(j,k)) intersect at a point y. Let p(s,y) be an initial segment of l(t'(i,l)), p(y,j) be a final segment of j(t'(j,k)), p(s',y) an initial segment of j(t'(j,k)), p(y,l) a final segment of l(t'(i,l)), i(t'(i,j)) = i(t'(i,l)), j(t'(i,j)) = p(s,y)&p(y,j), k(t'(k,l)) = k(t'(j,k)) and l(t'(k,l)) = p(s',y)&p(y,l). Since l(t'(k,l)) and j(t'(i,j)) are rearrangements of the vertices in j(t'(j,k)) and l(t'(i,l)), the number of occurrences of any vertex in l(t'(k,l)) and j(t'(i,j)) is less than or equal to the number of occurrences of that vertex in j(t'(j,k)) and l(t'(i,l)). Since C occurs at most once in j(t'(j,k)) and l(t'(i,l)), it occurs at most once in l(t'(k,l)) and j(t'(i,j)). Hence l(t'(k,l)) and j(t'(i,j)) do not intersect at C, contrary to the hypothesis that C is an lj(t(i,j),t(k,l)) choke point. ∴

2.6. Vanishing Partial Correlations Imply Vanishing Tetrad Differences

Theorem 2.6.1: For any probability distribution over a set of random variables V, if there exists a subset p of V such that $\rho_{ij.p}\rho_{kl.p} - \rho_{il.p}\rho_{jk.p} = 0$, and for all variables u in p and all subsets v of p not containing u, either $\rho_{iu.v} = 0$ and $\rho_{ku.v} = 0$, or $\rho_{ju.v} = 0$ and $\rho_{lu.v} = 0$, then $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$.
Proof. The proof is by induction on the cardinality of p.

Base Case: Suppose the cardinality of p is zero. Then $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ follows trivially from $\rho_{ij.p}\rho_{kl.p} - \rho_{il.p}\rho_{jk.p} = 0$.

Induction Case: Suppose that the lemma is true for all sets of cardinality n or less. Let p have cardinality n+1. Assume that $\rho_{ij.p}\rho_{kl.p} - \rho_{il.p}\rho_{jk.p} = 0$.

Let y be a variable in p, and p' = p - {y}. Since $\rho_{ij.p}\rho_{kl.p} = \rho_{il.p}\rho_{jk.p}$, by the recursion formula for partial correlation,

$$\left(\frac{\rho_{ij.p'} - \rho_{iy.p'}\rho_{jy.p'}}{\left(\sqrt{1-\rho_{iy.p'}^2}\right)\left(\sqrt{1-\rho_{jy.p'}^2}\right)}\right)\left(\frac{\rho_{kl.p'} - \rho_{ky.p'}\rho_{ly.p'}}{\left(\sqrt{1-\rho_{ky.p'}^2}\right)\left(\sqrt{1-\rho_{ly.p'}^2}\right)}\right) = \left(\frac{\rho_{il.p'} - \rho_{iy.p'}\rho_{ly.p'}}{\left(\sqrt{1-\rho_{iy.p'}^2}\right)\left(\sqrt{1-\rho_{ly.p'}^2}\right)}\right)\left(\frac{\rho_{jk.p'} - \rho_{jy.p'}\rho_{ky.p'}}{\left(\sqrt{1-\rho_{jy.p'}^2}\right)\left(\sqrt{1-\rho_{ky.p'}^2}\right)}\right)$$

The denominator of the l.h.s. equals the denominator of the r.h.s., so the numerator of the l.h.s. equals the numerator of the r.h.s. Expanding the numerators of each side,

$$\rho_{ij.p'}\rho_{kl.p'} - \rho_{ij.p'}\rho_{ky.p'}\rho_{ly.p'} - \rho_{kl.p'}\rho_{iy.p'}\rho_{jy.p'} - \rho_{iy.p'}\rho_{jy.p'}\rho_{ky.p'}\rho_{ly.p'} =$$

$$\rho_{il.p'}\rho_{jk.p'} - \rho_{il.p'}\rho_{jy.p'}\rho_{ky.p'} - \rho_{jk.p'}\rho_{iy.p'}\rho_{ly.p'} - \rho_{iy.p'}\rho_{jy.p'}\rho_{ky.p'}\rho_{ly.p'}$$

The fourth terms on both sides are equal. By hypothesis, either $\rho_{iy.p'} = \rho_{ky.p'} = 0$, or $\rho_{jy.p'} = \rho_{ly.p'} = 0$. In either case, the second and third terms on each side are equal to zero. It follows that $\rho_{ij.p'}\rho_{kl.p'} - \rho_{il.p'}\rho_{jk.p'} = 0$. Since p' has one less member than p, by the induction hypothesis, $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$. ∎

2.7. The Existence of a Choke Point Implies Vanishing Tetrad Differences

Theorem 2.7.1: In an acyclic LCF G, if there exists an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point or an ik(t(i,j),t(k,l),t(i,l),t(j,k)) choke point, then G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$.

Proof. Suppose w.l.g. that x is an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point. There are two cases.

First consider the case where there is no trek between at least one of the pairs i and j, and k and l, and there is no trek between at least one of the pairs i and l, and j and k. It follows that at least one of ρ_{ij} and ρ_{kl} equals 0, and at least one of ρ_{il} and ρ_{jk} is equal to zero. Hence $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$.

Next suppose w.l.g. that there are treks t'(i,j) and t'(k,l). I will prove that $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ by proving that there exists a set q' of variables such that $\rho_{ij.q'}\rho_{kl.q'} - \rho_{il.q'}\rho_{jk.q'} = 0$, and for all variables u in q' and all subsets v of q' not containing u, either $\rho_{iu.v} = 0$ and $\rho_{ku.v} = 0$, or $\rho_{ju.v} = 0$ and $\rho_{lu.v} = 0$, and applying theorem 2.6.1.

Let q = {sources of treks between x and j or x and l}. Since x is on j(t'(i,j)) and l(t'(k,l)), and by definition the sink of j(t'(i,j)) is j, and the sink of l(t'(k,l)) is l, there are directed paths p(x,j) and p(x,l); hence x is in q. I will now demonstrate that I(i,q,j) by showing that i and j are d-separated by q. I will show that i and j are d-separated by q by showing that every undirected path between i and j either contains a vertex v with a direction-reversal that is not the source of a directed path from v to any vertex in q, or it contains some vertex in q that does not have a direction-reversal.

Consider first the undirected paths between i and j without direction-reversals. If there is an undirected path between i and j that does not contain x, by lemma 2.2.1 there is a trek between i and j that does not contain x. But, every t(i,j) contains x, since x is a choke point. Hence, there does not exist an undirected path between i and j without direction-reversals that does not contain x. Since x is in q, every undirected path that does not contain a direction-reversal contains a vertex in q.

Consider now undirected paths between i and j that contain direction-reversals. If some vertex w with a direction-reversal is not the source of a directed path from w to some vertex in q, the proof is done. Suppose then that every vertex w with a direction-reversal is the source of a directed path from w to some vertex in q. Consider w.l.g. an arbitrary undirected path p(j,i) from j to i. Let z be the first vertex on p(j,i) that has a direction-reversal. By hypothesis, there is a directed path p(z,u) where u is a vertex in q. Since the undirected path from j to z does not contain any direction-reversals, by lemma 1.2.1 there is a vertex s that is the source of a pair of directed paths p(s,j) and p(s,z). Since z has an edge directed into it, $s \neq z$. There are two cases.

a. s = j. See Fig. 2.7.1. There is a directed path p(j,z). There is a directed paths p(z,u). Since u is the source of a trek between x and j, there is a directed path p(u,x). I have already shown that there is a directed path p(x,j). Hence there is a cyclic path p(j,z)&p(z,u)&p(u,x)&p(x,j).

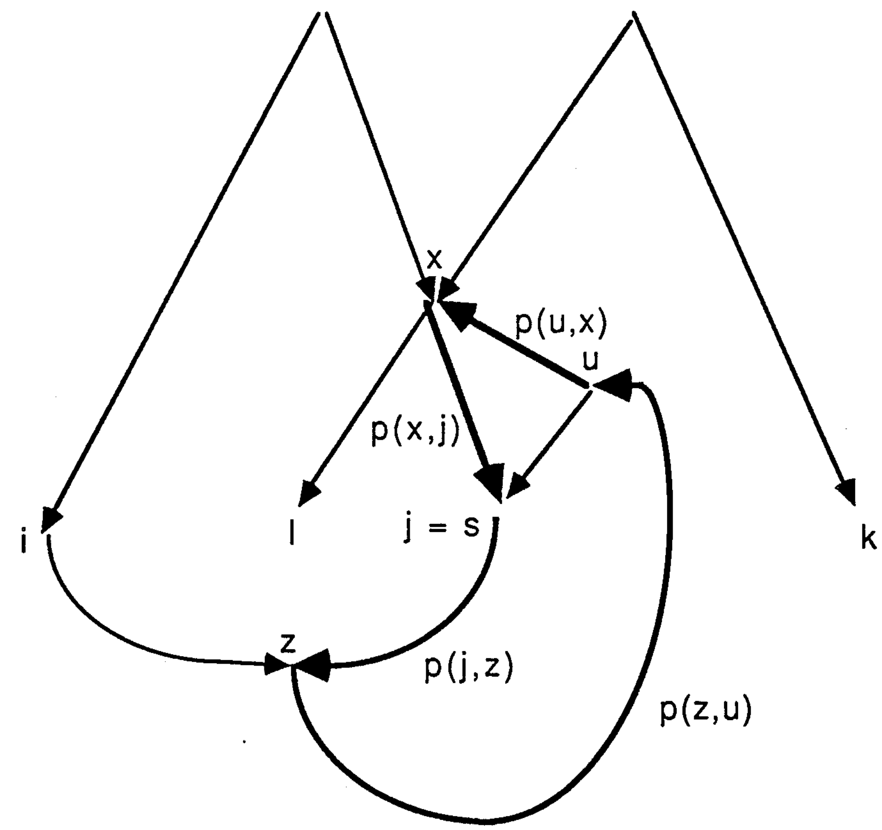


Fig. 2.7.1

b. $s \neq j$. See Fig. 2.7.2. There are directed path p(s,j), and a directed path p(s,z)&p(z,u)&p(u,x). By lemma 2.3.8 there is a trek t'(j,x) with source r, where r is the last point of intersection of p(s,j) and p(s,z)&p(z,u)&p(u,x), and j(t'(j,x)) is a subpath of p(s,j). Since r is on p(s,j), and s occurs before z on p(j,i), r occurs before z on p(j,i). Hence there is no direction-reversal at r in p(j,i). Also, r is in q, since it is the source of a trek between x and j. The undirected path p(j,i) contains a vertex in q that does not have a direction-reversal.

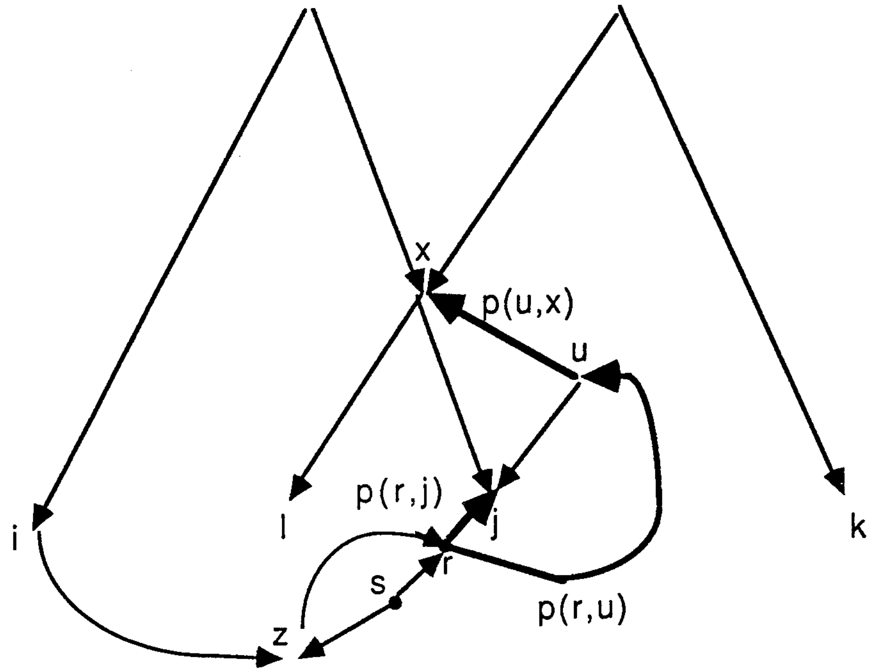


Fig. 2.7.2

In either case q d-separates x and y, so I(i,q,j). Similarly, it can be shown that I(k,q,l), I(i,q,l), I(j,q,k). It follows that $\rho_{ij.q} = 0$, $\rho_{kl.q} = 0$, $\rho_{il.q} = 0$, and $\rho_{jk.q} = 0$. Let q' = q - {x}. By the recursion formula for partial correlation, $\rho_{ij.q'} = \rho_{ix.q'}\rho_{jx.q'}$, $\rho_{kl.q'} = \rho_{kx.q'}\rho_{lx.q'}$, $\rho_{il.q'} = \rho_{ix.q'}\rho_{lx.q'}$, and $\rho_{jk.q'} = \rho_{jx.q'}\rho_{kx.q'}$. Hence $\rho_{ij.q'}\,\rho_{kl.q'} = \rho_{ix.q'}\rho_{jx.q'}\rho_{kx.q'}\rho_{lx.q'} = \rho_{ix.q'}\rho_{lx.q'}\rho_{jx.q'}\rho_{kx.q'} = \rho_{il.q'}\,\rho_{jk.q'}$.

I will next demonstrate that for each variable u in q', and each subset v of q' not containing u, I(i,v,u), by showing that i and u are d-separated by v. I will show that i and u are d-separated by v by showing that every undirected path between i and u either contains a vertex w with a direction-reversal that is not the source of a directed path from w to any vertex in v, or it contains some vertex in v that does not have a direction-reversal.

For u in q', consider an arbitrary undirected path p(i,u) that contains direction-reversals. Let z be the first point of p(i,u) after i that has a direction-reversal, and p(i,z) be an initial segment of p(i,u). If z is not the source of a path to some vertex r in q', then i and u are d-separated by q', and the proof is done. Suppose then that there is a directed path p(z,r) to some r in q'. Since p(i,z) contains no direction-reversals, by lemma 1.2.1 there is a vertex s on p(i,z) that is the source of directed paths p(s,i) and p(s,z). Hence s is the source of directed paths to i and r, p(s,i) and p(s,r) = p(s,z)&p(z,r) respectively. (If p(i,u) is an undirected path that contains no direction-reversals, then it still follows that by lemma 1.2.1 there is a vertex s on p(i,u) that is the source of directed paths p(s,i) and p(s,u).) r is either the source of a trek between x and j or x and l. Suppose w.l.g. that r is the source of a trek between x and j. Then r is the source of a directed path p(r,j) and a directed path p(r,x). r does not equal x by hypothesis. Hence p(r,j) does not contain x, since p(r,j) is a branch of a trek between j and x, and the two branches of the trek intersect only at r. p(s,r) does not contain x, else there is a cycle. Either p(s,i) contains x or it doesn't.

a. If p(s,i) does not contain x, then there are a pair of paths, p(s,i) and p(s,z)&p(z,r)&p(r,j) that do not contain x. See Fig. 2.7.3. Hence there is a trek between i and j that does not contain x. This contradicts the assumption that x is an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point.

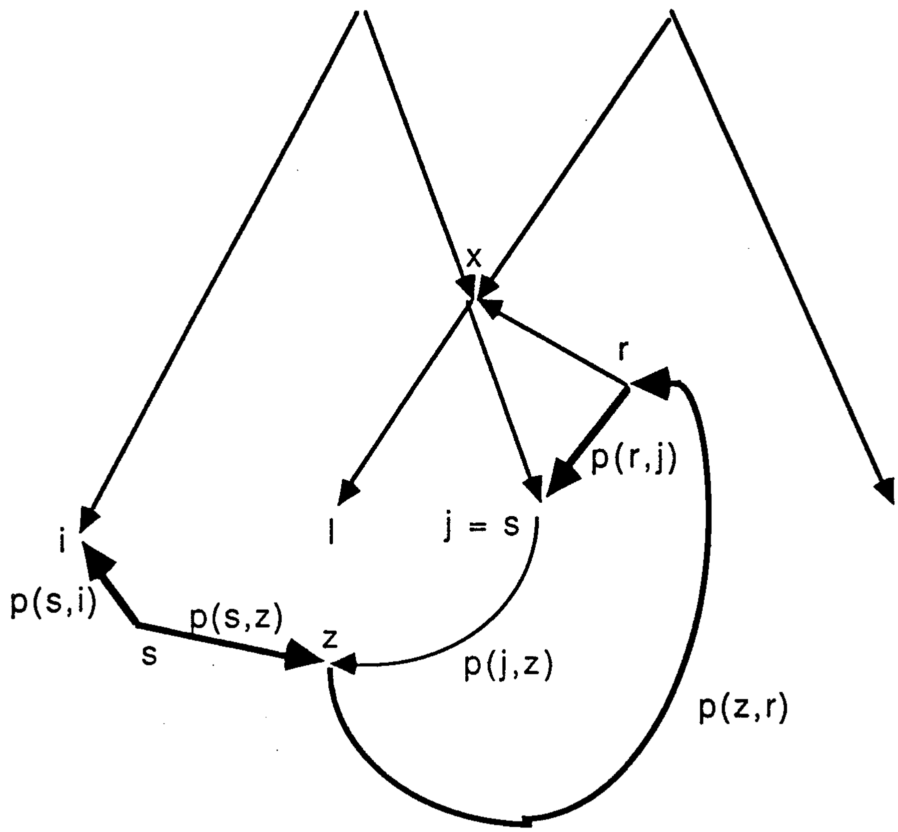


Fig. 2.7.3

b. Suppose p(s,i) does contain x. See Fig. 2.7.4. Then there is a directed path p(s,x) that is a subpath of p(s,i). There is also a directed path p(s,r) = p(s,z)&p(z,r). Hence there is a trek between x and r whose source s' lies on p(s,x), which is a subpath of p(s,i). s' is in q' since it is the source of a trek between x and r. s' lies on p(s,x) which is a subpath of p(s,i). p(s,i) does not have a direction-reversal at s', since s' occurs before z, which is the first direction-reversal. Hence p(s,i) contains a vertex in q' that does not have a direction-reversal.

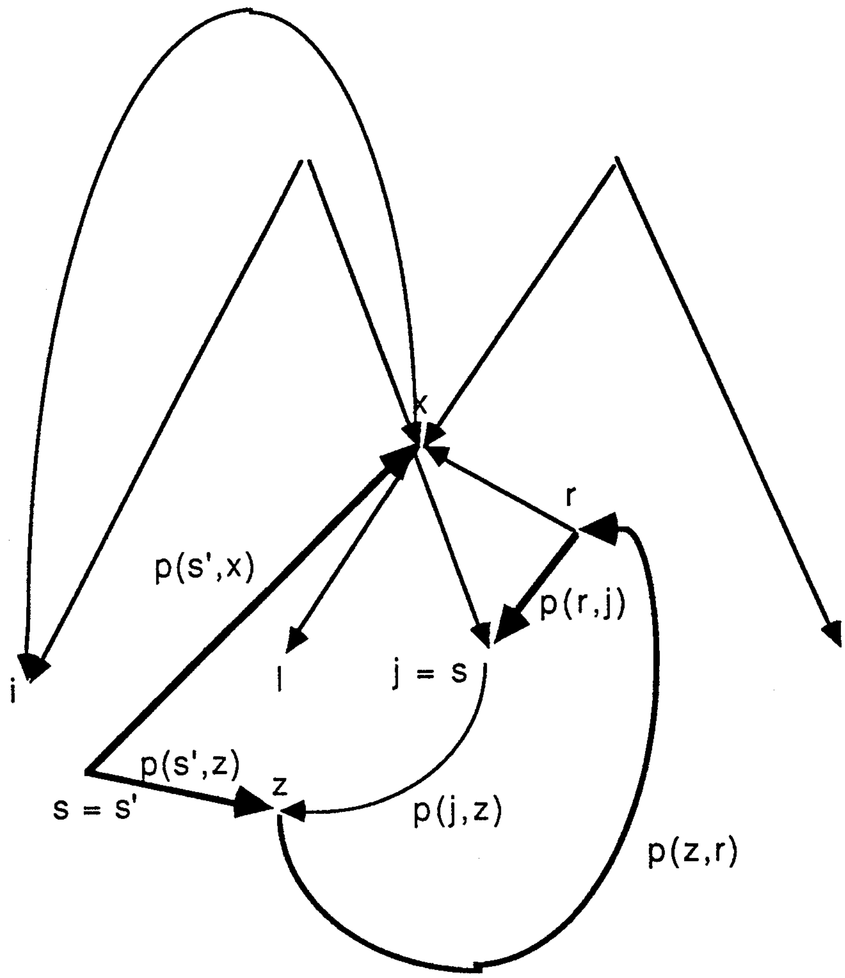


Fig. 2.7.4

By theorem 2.6.1, $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$. ∴

Theorem 2.7.2: In an acyclic LCF G, there exists an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point or an ik(t(i,j),t(k,l),t(i,l),t(j,k)) choke point iff G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$.
Proof. This follows directly from lemma 2.5.13 and theorem 2.7.1. ∴

Corollary 2.7.1: If an acyclic LCF G' is a subgraph of an acyclic LCF G, and G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$, then G' strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$.
Proof. If G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$, then by theorem 2.7.1 G has either an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point or an ik(t(i,j),t(k,l),t(i,l),t(j,k)) choke point. If G has either an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point or an ik(t(i,j),t(k,l),t(i,l),t(j,k)) choke point, then G' has either an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point or an ik(t(i,j),t(k,l),t(i,l),t(j,k)) choke point. By theorem 2.7.1, G' strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$. ∴

Theorem 2.7.3: A n acyclic LCF G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$ only if there is a (possibly empty) set q of random variables in G, either ρ_{ij} or $\rho_{kl} = 0$, and ρ_{il} or $\rho_{jk} = 0$, or there exists a (possible empty) set q such that $\rho_{ij.q} = \rho_{kl.q} = \rho_{il.q} = \rho_{jk.q} = 0$.

Proof. By theorem 2.7.2, if G strongly implies $\rho_{ij}\rho_{kl} - \rho_{il}\rho_{jk} = 0$, then there is either an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point or an ik(t(i,j),t(k,l),t(i,l),t(j,k)) choke point in G. In the proof of theorem 2.7.1 I demonstrated that the existence of an lj(t(i,j),t(k,l),t(i,l),t(j,k)) choke point or an ik(t(i,j),t(k,l),t(i,l),t(j,k)) choke point linearly implied by G then either either ρ_{ij} or $\rho_{kl} = 0$, and ρ_{il} or $\rho_{jk}=$ 0 there exists a set v of random variables such that $\rho_{ij.v} = 0$, $\rho_{kl,v} = 0$, $\rho_{il.v}= 0$, and $\rho_{jk.v} = 0$. ∴

Bibliography

[1] Freund, J., and Walpole, R., *Mathematical Statistics*, Prentice-Hall, Inc., Englewood Cliffs NJ,1980.

[2] Glymour, C. Scheines, R., Spirtes, P., and Kelly, K. , *Discovering Causal Structure.* Academic Press, San Diego, CA, 1987.

[3] Pearl, J.,*Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufman Publishers, San Meteo CA, 1988.

[4] Scheines, R., Spirtes, P., and Glymour, C., "A Qualitative Approach to Causal Modeling", in *Qualitative Issues in Simulation Modeling*, edited by Paul Fishwick, Springer-Verlag, 1989.

[5] Spirtes, P., Scheines, R. and Glymour, C. (1990). "Simulation Studies of the Reliability of Computer Aided Model Specification Using the TETRAD II, EQS and LISREL Programs." *Sociological Methods and Research*, **19**, 3-66.

[6] Spirtes, P., Glymour, C., and Scheines, R. "An Algorithm for Fast Recovery of Sparse Causal Graphs." *Social Science Computer Review*, in press.