Supplementary Material for

'Causal Discovery of Linear Cyclic Models from Multiple Experimental Data Sets with Overlapping Variables'

Antti Hyttinen
HIIT & Dept. of Computer Science
University of Helsinki
Finland

Frederick Eberhardt
Dept. of Philosophy
Carnegie Mellon University
Pittsburgh, PA, USA

Patrik O. Hoyer
HIIT & Dept. of Computer Science
University of Helsinki
Finland

A Proof of Theorem 1

First we restate Assumption 1 here for convenience:

Assumption 1 (No unit cycles) *The sum-product of edge-coefficients on any subset of paths from a variable back to itself cannot sum up to exactly 1.*

Note that when we are considering sum-products in cyclic graphs, whenever needed, we may sum a divergent geometric series.[1]

We will use the following Lemma to prove the Theorem:

Lemma 1 *Given a model* $\mathbf{B}$ *satisfying Assumption 1, the matrix* $(\mathbf{I}-\mathbf{B})$ *is invertible.*

Proof of Lemma 1: Assume that $(\mathbf{I}-\mathbf{B})$ is not invertible. Thus, there exists a $\mathbf{v} \neq \mathbf{0}$ such that $(\mathbf{I}-\mathbf{B})\mathbf{v} = \mathbf{0}$, which can be rewritten as $\mathbf{B}\mathbf{v} = \mathbf{v}$. Without loss of generality assume that $\mathbf{v}_1 = 1$.

Let us denote by $\mathcal{M} := \mathcal{V} \setminus \{x_1\}$ the set of all variables except x_1. We can then write $\mathbf{B}\mathbf{v}$ in the following block form

$$\left[\begin{array}{cc} 0 & \mathbf{B}_{1,\mathcal{M}} \\ \mathbf{B}_{\mathcal{M},1} & \mathbf{B}_{\mathcal{M},\mathcal{M}} \end{array}\right] \left[\begin{array}{c} 1 \\ \mathbf{v}_{\mathcal{M}} \end{array}\right],$$

where the top-left element is 0 because diagonal elements of $\mathbf{B}$ can be assumed to be zero, and thus from $\mathbf{B}\mathbf{v} = \mathbf{v}$ we obtain

$$\left[\begin{array}{c} \mathbf{B}_{1,\mathcal{M}}\mathbf{v}_{\mathcal{M}} \\ \mathbf{B}_{\mathcal{M},1} + \mathbf{B}_{\mathcal{M},\mathcal{M}}\mathbf{v}_{\mathcal{M}} \end{array}\right] = \left[\begin{array}{c} 1 \\ \mathbf{v}_{\mathcal{M}} \end{array}\right]$$

Hence $\mathbf{B}_{1,\mathcal{M}}\mathbf{v}_{\mathcal{M}} = 1$ and $\mathbf{B}_{\mathcal{M},1} = (\mathbf{I}-\mathbf{B}_{\mathcal{M},\mathcal{M}})\mathbf{v}_{\mathcal{M}}$.

Consider two possible cases. First, if $(\mathbf{I}-\mathbf{B}_{\mathcal{M},\mathcal{M}})$ is invertible we can 'marginalize' the variables in $\mathcal{M}$, as discussed in (Hyttinen et al., 2012), to arrive at a linear cyclic model with a 1×1 direct effects matrix $\tilde{\mathbf{B}}$. The formula for marginalizing is

$$\tilde{\mathbf{B}} = \mathbf{B}_{1,1} + \mathbf{B}_{1,\mathcal{M}}(\mathbf{I}-\mathbf{B}_{\mathcal{M},\mathcal{M}})^{-1}\mathbf{B}_{\mathcal{M},1}$$

and using the previous identities we obtain

$$\begin{aligned} \tilde{\mathbf{B}} &= \mathbf{B}_{1,\mathcal{M}}(\mathbf{I}-\mathbf{B}_{\mathcal{M},\mathcal{M}})^{-1}(\mathbf{I}-\mathbf{B}_{\mathcal{M},\mathcal{M}})\mathbf{v}_{\mathcal{M}} \\ &= \mathbf{B}_{1,\mathcal{M}}\mathbf{v}_{\mathcal{M}} = [1] \end{aligned}$$

Thus, the sum-product of *all* paths from x_1 back to itself is equal to one, violating Assumption 1.

On the other hand, consider the case $(\mathbf{I}-\mathbf{B}_{\mathcal{M},\mathcal{M}})$ not invertible. In this case, we remove all edges entering or leaving x_1 from the original model, and thus consider only the submodel $\mathbf{B}_{\mathcal{M},\mathcal{M}}$. Given that $(\mathbf{I}-\mathbf{B}_{\mathcal{M},\mathcal{M}})$ is not invertible, we can use the same argument as above. If at any point in the recursion we encounter an invertible submatrix over which we can marginalize, by the above procedure we have shown that there is a subset of paths (removing all edges from variables considered earlier in the recursion, and considering all paths through the remaining set of edges) summing to exactly 1. If, on the other hand, we do *not* encounter an invertible $(\mathbf{I}-\mathbf{B}_{\mathcal{M},\mathcal{M}})$ at any point of the recursion, we will, at the latest, encounter one at the point where $\mathcal{M}$ is just a singleton, because in that case $\mathbf{B}_{\mathcal{M},\mathcal{M}}$ contains just the scalar 0 (remember that we are restricting to a zero diagonal), so $(\mathbf{I}-\mathbf{B}_{\mathcal{M},\mathcal{M}}) = [1]$ which is trivially invertible. Hence, at the latest, the 2×2 matrix of direct effects among the two 'last' variables considered will contain a cycle equal to 1.

Corollary 1 *A model which satisfies Assumption 1 is weakly stable under any manipulations.*

Proof of Corollary 1: Any given manipulation of the model direct effects matrix $\mathbf{B}$ simply amounts to removing (setting to zero) edges into those variables

[1] The sum $\sum_{k=0}^{\infty}(t(x_i \rightsquigarrow x_j||\mathcal{J}_j)t(x_j \rightsquigarrow x_i||\mathcal{J}_i))^k$ is evaluated to $1/(1-t(x_i \rightsquigarrow x_j||\mathcal{J}_j)t(x_j \rightsquigarrow x_i||\mathcal{J}_i))$ also when $|t(x_i \rightsquigarrow x_j||\mathcal{J}_j)t(x_j \rightsquigarrow x_i||\mathcal{J}_i)| > 1$ (Hardy, 1949).

that are intervened. Consider the submatrix of connections among the variables not intervened on, call it $\mathbf{B}_m$, and note that the remaining edges form a *subset* of the original set of edges. Assume for the moment that this submodel is not weakly stable, i.e. $(\mathbf{I} - \mathbf{B}_m)$ is not invertible. Thus, by Lemma 1, $\mathbf{B}_m$ does not satisfy Assumption 1. But given that the edges in $\mathbf{B}_m$ are a subset of those in the original $\mathbf{B}$, it follows that $\mathbf{B}$ also does not satisfy Assumption 1.

We are now ready to prove Theorem 1:

Theorem 1 (Sufficiency) *Given some set of experiments, a linear cyclic model with latent variables satisfying Assumption 1 is fully identified if the pair condition is satisfied for all ordered pairs* $(x_i, x_j) \in \mathcal{V} \times \mathcal{V}$, $x_i \neq x_j$.

Proof of Theorem 1:

As noted by Eberhardt et al. (2010) the system of linear equations on the total effects can be divided into n subsystems, each constraining the total effects from one variable to all the others. Consider the subsystem $\mathbf{A}\mathbf{t}_n = \mathbf{a}$ constraining the total effects $t(x_n \rightsquigarrow \bullet)$. Select $n-1$ equations such that the i:th equation of the system comes from an experiment with intervention set $\mathcal{J}_i$ satisfying the pair condition for the pair (n, i) and is of the form

$$t(x_n \rightsquigarrow x_i) \quad = \quad \sum_{x_j \in \mathcal{J}_i} t(x_n \rightsquigarrow x_j) t(x_j \rightsquigarrow x_u || \mathcal{J}_i) \quad (1)$$

with $t(x_n \rightsquigarrow x_n) := 1$. For example, for $n = 4$ the system $\mathbf{A}\mathbf{t}_n = \mathbf{a}$ would be the following:

$$\mathbf{A} = \begin{bmatrix} 1 & -t(x_2 \rightsquigarrow x_1 || \mathcal{J}_1) & -t(x_3 \rightsquigarrow x_1 || \mathcal{J}_1) \\ -t(x_1 \rightsquigarrow x_2 || \mathcal{J}_2) & 1 & -t(x_3 \rightsquigarrow x_2 || \mathcal{J}_2) \\ -t(x_1 \rightsquigarrow x_3 || \mathcal{J}_3) & -t(x_2 \rightsquigarrow x_3 || \mathcal{J}_3) & 1 \end{bmatrix}$$

$$\mathbf{t}_n = \begin{bmatrix} t(x_4 \rightsquigarrow x_1) \\ t(x_4 \rightsquigarrow x_2) \\ t(x_4 \rightsquigarrow x_3) \end{bmatrix}$$

$$\mathbf{a} = \begin{bmatrix} t(x_4 \rightsquigarrow x_1 || \mathcal{J}_1) \\ t(x_4 \rightsquigarrow x_2 || \mathcal{J}_2) \\ t(x_4 \rightsquigarrow x_3 || \mathcal{J}_3) \end{bmatrix}$$

where we will denote $t(x_i \rightsquigarrow x_j || \mathcal{J}_j) = 0$ if $x_i \notin \mathcal{J}_j$.

Now note that the matrix $\mathbf{A}$ resembles the $\mathbf{I} - \mathbf{B}$ matrix of a linear cyclic model $\mathbf{B}_p$ with $n-1$ variables, where the direct effects have been replaced by certain experimental effects measured in the experiments. We call matrix $\mathbf{B}_p = \mathbf{I} - \mathbf{A}$ the *path model* corresponding to the set of experiments constraining the total effects $t(x_n \rightsquigarrow \bullet)$. This path model is illustrated in Figure 1.

To prove the theorem, we need to show that $\mathbf{A} = \mathbf{I} - \mathbf{B}_p$ is always invertible, under Assumption 1 on $\mathbf{B}$. We will

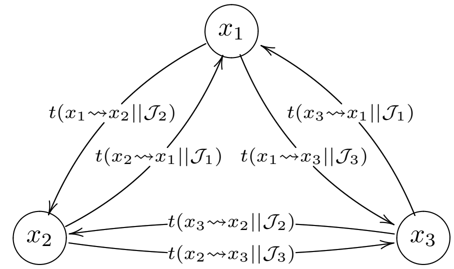

Figure 1: Path model. See text for details.

do this by showing that, if $\mathbf{A}$ is not invertible, there exists a subset of paths in the original model, from one variable back to itself, that sum to one.

Hence, assume that $\mathbf{I} - \mathbf{B}_p$ is not invertible. This implies, by Lemma 1, that there exists a subset of paths in the path model $\mathbf{B}_p$, from some variable back to itself, which sum to one. Without loss of generality, assume this variable is x_1. We need to show that the sum-product of any subset of paths, from x_1 back to itself, in $\mathbf{B}_p$ is equal to the sum-product of some subset of paths, from x_1 back to itself, in $\mathbf{B}$. This we do by creating a correspondence between paths in $\mathbf{B}_p$ and paths in $\mathbf{B}$; specifically, any path in $\mathbf{B}_p$ is the sum-product of a set of paths in $\mathbf{B}$, and two distinct paths q_1 and q_2 in $\mathbf{B}_p$ do not share any corresponding paths in $\mathbf{B}$, so there is no risk of counting any given path in $\mathbf{B}$ twice.

Definition: A path p in $\mathbf{B}$, from x_s to x_t, is said to *correspond* to a path q, from x_s to x_t, in $\mathbf{B}_p$ if

(a) each node in q can be mapped to an equivalently labeled node in p, without changing their relative ordering, and

(b) any nodes in p located between a mapped node x_a and a mapped node x_b (with $x_a \rightarrow x_b$ in q) belong to the set $\mathcal{V} \setminus \mathcal{J}_b$, i.e. to the set of nodes that do *not* point into x_b in $\mathbf{B}_p$

First, note that the strength of each edge in the path model $\mathbf{B}_p$ is the sum-product of its corresponding paths in the original model $\mathbf{B}$, as each entry is the corresponding experimental effect $t(x_i \rightsquigarrow x_j || \mathcal{J}_j)$, and the definition of this experimental effect is the sum-product of paths leaving x_i and ending in x_j without going through any nodes in $\mathcal{J}_j$.

Second, and finally, we prove that any path p in $\mathbf{B}$ can correspond to *at most* one path q in $\mathbf{B}_p$. To see this, we explicitly construct q in a deterministic way such that any deviation from the construction would result in a violation of the correspondence. Start at the last (target) node of p, call this node q_0. Mark this as the last (target) node of q. Next, consider the previous node in p. If this node points to q_0 in $\mathbf{B}_p$, then for correspondence this node has to be mapped to q, call it q_1. If, on the other hand, the node does *not* point to q_0, then this node *cannot* be mapped to q. Moving to the previous element in p, always check if it points to the latest mapped node in q, and map it to q if and only if it does. If, when finally arriving at the first element of p, it can be mapped to q, a valid corresponding path is found. If it cannot, no valid corresponding path exists. Because all decisions are 'forced', no other paths except that constructed can correspond to p.

Now, under the assumption that $\mathbf{A} = (\mathbf{I} - \mathbf{B}_p)$ is not invertible, $\mathbf{B}_p$ has at least one node x_i such that there exists a subset of paths in $\mathbf{B}_p$, from x_i back to itself, that sums to 1. Given that each path in $\mathbf{B}_p$ corresponds to the sum-product of a distinct set of paths in $\mathbf{B}$, this implies that there exists a subset of paths in $\mathbf{B}$ from x_i back to itself, which sum to one. This contradicts Assumption 1.

B Proof of Theorem 2

For convenience we restate the theorem here:

Theorem 2 (Worst Case Necessity) *Given any set of experiments that does* not *satisfy the pair condition for all ordered pairs of variables $(x_i, x_j) \in \mathcal{V} \times \mathcal{V}$, $x_i \neq x_j$, there exist two distinct linear cyclic models with latent variables satisfying Assumption 1 that are indistinguishable from those experiments.*

Proof:

Given that the set of experiments does not satisfy the pair condition for all pairs, there exists an ordered pair (x_i, x_j) that is not satisfied. The theorem is proven by constructing two distinct models $(\mathbf{B}, \mathbf{\Sigma_e})$ and $(\tilde{\mathbf{B}}, \tilde{\mathbf{\Sigma}}_\mathbf{e})$ that nevertheless yield the same data covariance matrices in all experiments not satisfying the pair (x_i, x_j).

Specifically, consider an n variable model $(\mathbf{B}, \mathbf{\Sigma_e})$ such that all elements of $\mathbf{B}$ are 0 except for element $\mathbf{B}[j,i] = 1$. Let $\mathbf{\Sigma_e} = \mathbf{I}$, the $n \times n$ identity matrix. On the other hand, consider a similar model $(\tilde{\mathbf{B}}, \tilde{\mathbf{\Sigma}}_\mathbf{e})$, with all elements of $\tilde{\mathbf{B}}$ equal to 0, and a disturbance covariance matrix $\tilde{\mathbf{\Sigma}}_\mathbf{e}$ given by

$$\tilde{\mathbf{\Sigma}}_\mathbf{e}[r,c] = \begin{cases} 2, & \text{if } r = c = j \\ 1, & \text{if } r = c \neq j \\ 1, & \text{if } r = i \text{ and } c = j \\ 1, & \text{if } r = j \text{ and } c = i \\ 0, & \text{otherwise} \end{cases}$$

In essence, the above construction creates a completely empty pair of models, except that in one there is a direct effect from x_i to x_j, and in the other there is a confounder between x_i and x_j.

For both models, in any experiment intervening on neither x_i nor x_j (i.e. both are passively observed), the variance of x_i is 1, the variance of x_j is 2, and the covariance between the two variables is 1. For both models, in any experiment intervening on x_j but not x_i, the two variables are uncorrelated with unit variances. The same applies to any experiment intervening on both x_i and x_j. Together, these handle all the cases, except those experiments intervening on x_i and observing x_j, which are not allowed because that pair is not satisfied. If either variable is latent in the experiment, no information concerning their relationship is learned. Hence, both models produce the same data covariance matrices in all experiments not satisfying the given pair condition, and the theorem is proven.

Note that the fact that we chose a model such that the variables x_i and x_j are unconnected to the other variables is not purely out of convenience; when there are more connections in the graph it opens up the possibility that variables which are latent in a given experiment can affect the covariances of the observed variables. In some cases, this allows for identifying the full model even when the pair condition is not satisfied for all pairs. For this reason, our theorem only discusses *worst case* necessity.

C Proof of Completeness

The proof of completeness given by Hyttinen et al. (2012) considering only fully observed experiments can be adapted to the scenario with overlapping experiments. The linear equations constraining the unknown total effects $\mathbf{T}$ obtained in experiment $\mathcal{E}_k = (\mathcal{J}_k, \mathcal{U}_k, \mathcal{L}_k)$ (see Equation 1) can be written in matrix form

$$\mathbf{E}^k_{\mathcal{U}_k\mathcal{J}_k}\mathbf{T}_{\mathcal{J}_k\mathcal{J}_k} = \mathbf{T}_{\mathcal{U}_k\mathcal{J}_k}$$

where

$$\mathbf{E}^k_{\mathcal{U}_k\mathcal{J}_k} = \begin{bmatrix} t(x_{j_1} \rightsquigarrow x_{u_1} || \mathcal{J}_k) & t(x_{j_2} \rightsquigarrow x_{u_1} || \mathcal{J}_k) & \cdots \\ t(x_{j_1} \rightsquigarrow x_{u_2} || \mathcal{J}_k) & t(x_{j_2} \rightsquigarrow x_{u_2} || \mathcal{J}_k) & \cdots \\ \vdots & \vdots & \end{bmatrix}$$

$$\mathbf{T}_{\mathcal{J}_k\mathcal{J}_k} = \begin{bmatrix} t(x_{j_1} \rightsquigarrow x_{j_1}) & t(x_{j_2} \rightsquigarrow x_{j_1}) & \cdots \\ t(x_{j_1} \rightsquigarrow x_{j_2}) & t(x_{j_2} \rightsquigarrow x_{j_2}) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

$$\mathbf{T}_{\mathcal{U}_k\mathcal{J}_k} = \begin{bmatrix} t(x_{j_1} \rightsquigarrow x_{u_1}) & t(x_{j_2} \rightsquigarrow x_{u_1}) & \cdots \\ t(x_{j_1} \rightsquigarrow x_{u_2}) & t(x_{j_2} \rightsquigarrow x_{u_2}) & \cdots \\ \vdots & \vdots & \end{bmatrix}$$

Here matrix $\mathbf{T}_{\mathcal{J}_k\mathcal{J}_k}$ is invertible (assuming weak stability of the model) as the total effects matrix of the model where variables $\mathcal{U}_k \cup \mathcal{L}_k$ are marginalized. Thus, the experimental effects produced by a model with total effects matrix $\mathbf{T}$ are simply:

$$\mathbf{E}^k_{\mathcal{U}_k\mathcal{J}_k} = \mathbf{T}_{\mathcal{U}_k\mathcal{J}_k}\mathbf{T}^{-1}_{\mathcal{J}_k\mathcal{J}_k}$$

Now, assume the true data generating model $(\mathbf{B}, \mathbf{\Sigma_e})$ with total effects matrix $\mathbf{T} = (\mathbf{I} - \mathbf{B})^{-1}$ has been observed in K overlapping experiments $\{\mathcal{E}_k\}_{k=1}^K$, producing experimental effects $\{\mathbf{E}^k_{\mathcal{U}_k\mathcal{J}_k}\}_{k=1}^K$. If the method of Eberhardt et al. (2010) has not identified the true causal structure, there is a direct effects matrix $\tilde{\mathbf{B}}$, distinct from $\mathbf{B}$, with total effects matrix $\tilde{\mathbf{T}} = (\mathbf{I} - \tilde{\mathbf{B}})^{-1}$ satisfying the linear equations on the total effects used by Eberhardt et al. (2010)

$$\mathbf{E}^k_{\mathcal{U}_k\mathcal{J}_k}\tilde{\mathbf{T}}_{\mathcal{J}_k\mathcal{J}_k} = \tilde{\mathbf{T}}_{\mathcal{U}_k\mathcal{J}_k}$$

for all $k = 1 \ldots K$. Then the direct effects matrix $\tilde{\mathbf{B}}$ produces the same experimental effects in all experiments $\{\mathcal{E}_k\}_{k=1}^K$:

$$\begin{aligned}\tilde{\mathbf{E}}^k_{\mathcal{U}_k\mathcal{J}_k} &= \tilde{\mathbf{T}}_{\mathcal{U}_k\mathcal{J}_k}\tilde{\mathbf{T}}^{-1}_{\mathcal{J}_k\mathcal{J}_k} \\ &= \mathbf{E}^k_{\mathcal{U}_k\mathcal{J}_k}\tilde{\mathbf{T}}_{\mathcal{J}_k\mathcal{J}_k}\tilde{\mathbf{T}}^{-1}_{\mathcal{J}_k\mathcal{J}_k} = \mathbf{E}^k_{\mathcal{U}_k\mathcal{J}_k}\end{aligned}$$

Lemma 14 of Hyttinen et al. (2012) shows that if a direct effects matrix $\tilde{\mathbf{B}}$ produces the same experimental effects in experiments $\{\mathcal{E}_k\}_{k=1}^K$ as the true model $(\mathbf{B}, \mathbf{\Sigma_e})$, then the model $(\tilde{\mathbf{B}}, \tilde{\mathbf{\Sigma}}_\mathbf{e})$ where

$$\tilde{\mathbf{\Sigma}}_\mathbf{e} := (\mathbf{I} - \tilde{\mathbf{B}})(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Sigma_e}(\mathbf{I} - \mathbf{B})^{-T}(\mathbf{I} - \tilde{\mathbf{B}})^T$$

produces also the same covariance matrices as the true model $(\mathbf{B}, \mathbf{\Sigma_e})$ in experiments $\{\mathcal{E}_k\}_{k=1}^K$. Thus the two models cannot be distinguished based on 2nd order statistics. A model not identified by Eberhardt et al. (2010) is inherently underdetermined by the experiments at hand.

D Inference Rules for Faithfulness

We repeat the definition of a minimal conditioning set from the main text. It is adapted from the definitions in Claassen and Heskes (2011).

For variables x and y and disjoint sets of variables C and D not containing x and y, we denote a minimal independence by

$$\begin{aligned}&x \perp\!\!\!\perp y \,|\, D \cup [C] \text{ whenever we have} \\ &\quad x \perp\!\!\!\perp y \,|\, D \cup C \text{ and } \forall C' \subsetneq C, x \not\perp\!\!\!\perp y \,|\, D \cup C',\end{aligned}$$

and a minimal dependence by

$$\begin{aligned}&x \not\perp\!\!\!\perp y \,|\, D \cup [C] \text{ whenever we have} \\ &\quad x \not\perp\!\!\!\perp y \,|\, D \cup C \text{ and } \forall C' \subsetneq C, x \perp\!\!\!\perp y \,|\, D \cup C'.\end{aligned}$$

In both cases D and C can be empty, although when C is empty, the statements become trivial.

For notational simplicity, we assume for any variable x and any set of variables W, $t(\bullet\!\rightsquigarrow\! x \,||\, W \cup \{x\}) = 0$. We also extend the notation of minimal conditioning sets to $x \perp\!\!\!\perp y \,|\, [C] \,||\, \mathcal{J}$ to mean that $x \perp\!\!\!\perp y | \, [C]$ holds in the manipulated distribution of an experiment that has $\mathcal{J}$ as its intervention set. We then have the following rules:

1: If $x \perp\!\!\!\perp y \,|\, [C] \,||\, \mathcal{J}$, then
(zero constraints)

- if $y \notin \mathcal{J}$, then $t(x \rightsquigarrow y \,||\, \mathcal{J} \cup C \cup x) = 0$
- if $x \notin \mathcal{J}$, then $t(y \rightsquigarrow x \,||\, \mathcal{J} \cup C \cup y) = 0$
- if $C = \{c\}$ and $x \in \mathcal{J}$ and $y, c \in \mathcal{U}$, then $t(y \rightsquigarrow c \,||\, \mathcal{J} \cup y) = 0$
- if $C = \{c\}$ and $y \in \mathcal{J}$ and $x, c \in \mathcal{U}$, then $t(x \rightsquigarrow c \,||\, \mathcal{J} \cup x) = 0$

(bilinear zero constraints)
for any $u \in \mathcal{L} \cup \mathcal{U} \setminus \{C, x, y\}$

- if $y \notin \mathcal{J}$, then
$t(x \rightsquigarrow u \,||\, \mathcal{J} \cup C \cup x) \times t(u \rightsquigarrow y \,||\, \mathcal{J} \cup C \cup u) = 0$
- if $x \notin \mathcal{J}$, then
$t(y \rightsquigarrow u \,||\, \mathcal{J} \cup C \cup y) \times t(u \rightsquigarrow x \,||\, \mathcal{J} \cup C \cup u) = 0$
- if $C = \{c\}$ and $x \in \mathcal{J}$ and $y, c \in \mathcal{U}$, then
$t(y \rightsquigarrow u \,||\, \mathcal{J} \cup y) \times t(u \rightsquigarrow c \,||\, \mathcal{J} \cup u) = 0$
- if $C = \{c\}$ and $y \in \mathcal{J}$ and $x, c \in \mathcal{U}$, then
$t(x \rightsquigarrow u \,||\, \mathcal{J} \cup x) \times t(u \rightsquigarrow c \,||\, \mathcal{J} \cup u) = 0$

for any $u \in \mathcal{V} \setminus \{C, x, y\}$

- if $x, y \notin \mathcal{J}$, then
$t(u \rightsquigarrow y \,||\, \mathcal{J} \cup C \cup u) \times t(u \rightsquigarrow x \,||\, \mathcal{J} \cup C \cup u) = 0$

2: If $x \perp\!\!\!\perp y \,|\, [C] \,||\, \mathcal{J}$ and $x \not\perp\!\!\!\perp y \,|\, C \cup [w] \,||\, \mathcal{J}$ for some $w \in \mathcal{U} \setminus \{C, x, y\}$, then
(zero constraints)

- if $y \notin \mathcal{J}$, then $t(w \rightsquigarrow y \,||\, \mathcal{J} \cup w) = 0$
- if $x \notin \mathcal{J}$, then $t(w \rightsquigarrow x \,||\, \mathcal{J} \cup w) = 0$
- $\forall c \in C \cap \mathcal{U}$, then $t(w \rightsquigarrow c \,||\, \mathcal{J} \cup w) = 0$

(bilinear zero constraints)
for any $u \in \mathcal{L} \cup \mathcal{U} \setminus \{C, x, y, w\}$

- if $y \notin \mathcal{J}$, then
  $t(w \rightsquigarrow u \,||\, \mathcal{J} \cup w) \times t(u \rightsquigarrow y \,||\, \mathcal{J} \cup u) = 0$
- if $x \notin \mathcal{J}$, then
  $t(w \rightsquigarrow u \,||\, \mathcal{J} \cup w) \times t(u \rightsquigarrow x \,||\, \mathcal{J} \cup u) = 0$
- $\forall c \in C \cap \mathcal{U}$, then
  $t(w \rightsquigarrow u \,||\, \mathcal{J} \cup w) \times t(u \rightsquigarrow c \,||\, \mathcal{J} \cup u) = 0$

3: If $x \perp\!\!\!\perp y \,|\, [C] \,||\, \mathcal{J}$ and $x \perp\!\!\!\perp y \,|\, C \cup w \,||\, \mathcal{J}$ for some $w \in \mathcal{U} \setminus \{C, x, y\}$, then
(bilinear constraint)

- $t(x \rightsquigarrow w \,||\, \mathcal{J} \cup C \cup x) \times t(y \rightsquigarrow w \,||\, \mathcal{J} \cup C \cup y) = 0$
  (note that x and y can be both in $\mathcal{J}$)

4: Expand these constraints to all supersets of the intervention sets.

$$t(x \rightsquigarrow y || \mathcal{J}) = 0 \quad \Rightarrow \quad t(x \rightsquigarrow y || \mathcal{J} \cup C) = 0$$

for all $C \subseteq \mathcal{V} \setminus \{y\}$.

We have omitted consequences of faithfulness that cannot be represented as a product of two experimental effects. In particular this includes equations that are polynomial in the experimental effects or that include terms that involve the correlation due to latent confounding. For example, if $x \perp\!\!\!\perp y$ in a passive observational data set, then, among other things, we could also conclude that there is no latent confounder of x and y that is not a variable in $\mathcal{V}$.

References

Claassen, T. and Heskes, T. (2011). A logical characterization of constraint-based causal discovery. In *UAI 2011*.

Eberhardt, F., Hoyer, P. O., and Scheines, R. (2010). Combining experiments to discover linear cyclic models with latent variables. In *AISTATS 2010*.

Hardy, G. H. (1949). *Divergent Series*. Oxford: Clarendon Press.

Hyttinen, A., Eberhardt, F., and Hoyer, P. O. (2012). Learning linear cyclic causal models with latent variables. Submitted. Available online from the authors' homepages.