Causal discovery for linear cyclic models with latent variables

Antti Hyttinen[1], Frederick Eberhardt[2], and Patrik O. Hoyer[1,3]
[1] HIIT / Dept. of Computer Science, University of Helsinki, Finland
[2] Dept. of Philosophy, Washington University in St Louis, MO, USA
[3] CSAIL, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

We consider the problem of identifying the causal relationships among a set of variables in the presence of both feedback loops and unmeasured confounders. This is a challenging task which, for full identification, typically requires the use of randomized experiments. For linear systems, Eberhardt et al (2010) recently provided a procedure for integrating data from several experiments, and gave a corresponding, but demanding, identifiability condition. In this paper we (i) characterize the underdetermination of the model when the identifiability condition is not fully satisfied, (ii) show that their algorithm is complete with regard to the search space and the assumptions, and (iii) extend the procedure to incorporate the common assumption of faithfulness, and any prior knowledge. The resulting method typically resolves much additional structure and often yields full identification with many fewer experiments. We demonstrate our procedure using simulated data, and apply it to the protein signaling dataset of Sachs et al (2005).

1 Introduction

Researchers are frequently interested in discovering the causal relationships among some given set of variables under study. Such relationships are often represented as directed graphs, in which the variables constitute the nodes of the graph, and a directed edge from one variable x_i to another variable x_j indicates that x_i is a *direct cause* of x_j relative to that set of variables. Since causal relations are not directly observable they must be inferred from available experimental or passive observational data. Several algorithms have been developed that discover as much as possible about such causal relations from passive observational data. One of the difficulties these algorithms confront is the almost inevitable underdetermination of the true causal structure. This problem is exacerbated when there are unmeasured (latent) common causes of the set of variables under consideration, or when there are feedback loops. Consequently, constraints are typically placed on the search space the algorithms consider: The 'FCI' algorithm (Spirtes et al., 2000) only considers *acyclic* causal structures but allows latent variables, while the 'CCD' algorithm of Richardson (1996) can handle cyclic causal systems but does not allow for latents. Even with these restrictions, both algorithms can at best return equivalence classes of causal graphs.

Thus, it is common to turn to experimental data. While randomized experiments break confounding and feedback loops, they pose different challenges. Given that experiments are often costly, how can we identify the causal structure from as few experiments as possible? How can we integrate the data from several existing experiments to yield as much information as possible about the causal relationships among the variables? In this paper, we show how to efficiently perform such causal discovery in *linear* models from a combination of observational and experimental data, while allowing *both* feedback loops and confounding hidden variables.

We consider a standard class of models known as linear non-recursive structural equation models with correlated disturbances (Bollen, 1989). Specifically, let $\mathbf{V} = \{x_1, \ldots, x_N\}$ denote the set of observed variables. Arranging these variables into the vector $\mathbf{x}$, the linear model is given by

$$\mathbf{x} := \mathbf{B}\mathbf{x} + \mathbf{e}, \tag{1}$$

where each element b_{ji} of $\mathbf{B}$ gives the *direct effect* from x_i to x_j, also denoted $b(x_i \rightarrow x_j)$, and

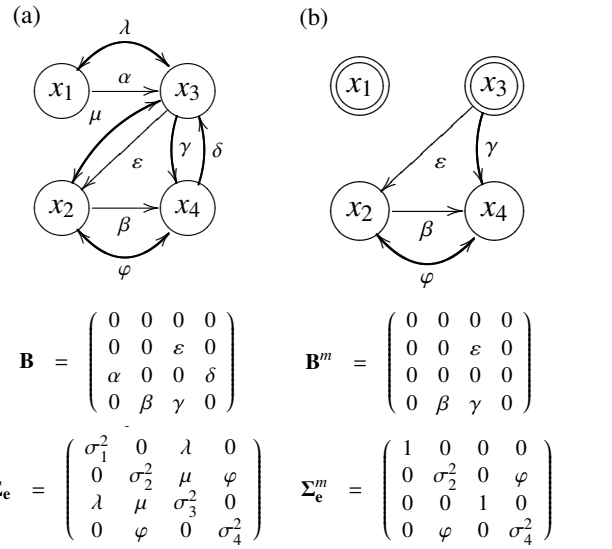

Figure 1: (a) Example model. (b) Manipulated model, corresponding to an experiment $(\mathbf{J}_m, \mathbf{U}_m)$ where $\mathbf{J}_m = \{x_1, x_3\}$ and $\mathbf{U}_m = \{x_2, x_4\}$. Disturbance variables are not shown.

the random vector $\mathbf{e}$ contains zero-mean *disturbance* (*error*) variables with a covariance matrix $\mathbf{\Sigma_e} = E\{\mathbf{e}\mathbf{e}^T\}$. An example model is given in Figure 1a.

An experiment $\mathcal{E}_m = (\mathbf{J}_m, \mathbf{U}_m)$ divides $\mathbf{V}$ into two mutually exclusive and exhaustive sets $\mathbf{J}_m$ and $\mathbf{U}_m$. $\mathbf{J}_m$ contains the variables subject to an intervention in $\mathcal{E}_m$ and $\mathbf{U}_m$ contains the variables that are passively observed in that experiment. In such an experiment, all variables $x_i \in \mathbf{J}_m$ are independently and simultaneously randomized. In terms of the directed graph, this is represented by cutting all edges *into* any such variable (Pearl, 2000). In terms of the parameters, we thus have a *manipulated* model $(\mathbf{B}^m, \mathbf{\Sigma}_\mathbf{e}^m)$, where $\mathbf{B}^m$ is equal to $\mathbf{B}$ except that all rows corresponding to such randomized variables are set to zero, and $\mathbf{\Sigma}_\mathbf{e}^m$ equals $\mathbf{\Sigma_e}$ but with all rows and columns corresponding to randomized variables set to zero, except for the corresponding diagonal element which is set to equal one due to the fixed variance of the randomization. See Figure 1b.

If the variables cannot be ordered such that the corresponding $\mathbf{B}$ is lower-triangular we have a truly non-recursive system that *cannot* be represented as a directed *acyclic* graph (DAG). If $\mathbf{\Sigma_e}$ has non-zero off-diagonal entries the system is said to exhibit confounding due to latent variables. In each experiment $\mathcal{E}_m$ the data are generated such that a random sample of disturbance vectors $\mathbf{e}$ are drawn with (manipulated) covariance $\mathbf{\Sigma}_\mathbf{e}^m$, and we observe the vectors $\mathbf{x}$ (and hence their covariance $\mathbf{\Sigma}_\mathbf{x}^m$) generated (at equilibrium) from the model with (manipulated) coefficient matrix $\mathbf{B}^m$. For the feedback system to reach equilibrium, the absolute values of all eigenvalues of $\mathbf{B}^m$ must be smaller than one.[1] A passive observational dataset is obtained in an 'experiment' in which $\mathbf{J}_m = \emptyset$ and $\mathbf{U}_m = \mathbf{V}$.

In an experiment in which $x_i \in \mathbf{J}_m$ and $x_j \in \mathbf{U}_m$, the *experimental effect* of x_i on x_j, denoted $t(x_i \rightsquigarrow x_j \,||\, \mathbf{J}_m)$, is defined as the covariance of x_i and x_j in this experiment, i.e. $\mathbf{\Sigma}_\mathbf{x}^m[i, j]$. This is equal to the sum of the strenghts of all uncut directed paths from x_i to x_j, where the strength of a path is simply the product of the edge coefficients (direct effects) on that path.[2]

Our task is to devise a sequence of experiments $(\mathcal{E}_1, \ldots, \mathcal{E}_M)$, and corresponding estimation procedure, that fully identifies the parameter matrices $\mathbf{B}$ and $\mathbf{\Sigma_e}$, in the sense that the estimates are consistent (converge to the true values in the infinite sample limit). Alternatively or in addition, for a fixed set of experiments one would like to recover as many as possible of these parameters. Note that if all but one variable is randomized in an experiment (i.e. $\mathbf{J}_m = \mathbf{V} \setminus \{x_j\}$) one can consistently estimate all direct effects $b(x_i \rightarrow x_j)$, $\forall i \neq j$, since in this experiment the direct effects equal the experimental effects. Thus one solution to identify $\mathbf{B}$ consists of $M = N$ such experiments each intervening on $N - 1$ variables. If in addition a passive observational dataset were available, one can obtain a consistent estimate of $\mathbf{\Sigma_e}$ from the identity $\mathbf{\Sigma_x} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Sigma_e}(\mathbf{I} - \mathbf{B})^{-T}$, where $\mathbf{\Sigma_x}$ is the covariance of $\mathbf{x}$ in a passive observational dataset.

Can we get by with fewer experiments? Recently, Eberhardt et al (2010) provided a procedure that identifies the full matrix $\mathbf{B}$ if and only if the following *pair condition* holds for each ordered variable pair $(x_i, x_j) \in \mathbf{V} \times \mathbf{V}$, with $i \neq j$: there is an experiment $\mathcal{E}_m = (\mathbf{J}_m, \mathbf{U}_m)$ in the sequence in which $x_i \in \mathbf{J}_m$ and $x_j \in \mathbf{U}_m$.

[1] As in (Eberhardt et al., 2010) we assume that this condition is satisfied for all possible manipulations of the $\mathbf{B}$-matrix.

[2] Note that this sum has an infinite number of terms when the model is cyclic.

However, several questions were left unanswered in their study. First, if the pair condition is not satisfied for all ordered pairs, which direct effects are identified and which are not? Second, is it possible that some alternative procedure might identify the full model even when for some pairs the condition is not satisfied? Finally, satisfying the pair condition for all ordered pairs is a very high bar for the identifiability of the underlying causal structure, as it requires that each variable must be subject to at least one intervention at some point in the sequence of experiments. For any observed variable that is not subject to an intervention their algorithm can only discover the causal structure marginalized over that variable. Thus, can we make use of prior knowledge when available, or strengthen some of the assumptions, to avoid requiring the pair condition for all pairs? We answer these three questions in Sections 2–4, respectively. Then, in Section 5, we describe a simple adaptive procedure for selecting the sequence of experiments, while providing simulations in Section 6 and an application to the protein signaling dataset of Sachs et al (2005) in Section 7. Conclusions are given in Section 8.

2 Characterization of underdetermination

Eberhardt et al (2010) showed that if the pair condition (see Section 1) is not satisfied for all ordered pairs then their estimation procedure leaves some total effects undetermined, and hence some elements of the direct effects matrix $\mathbf{B}$ are undetermined as well. They then suggested a numerical heuristic to identify the set of edges that are not yet determined. Here we show how, using an alternative formulation of the procedure, we obtain a characterization of the remaining underdetermination in the direct effects.

From an experiment $\mathcal{E}_m = (\mathbf{J}_m, \mathbf{U}_m)$, with $x_i \in \mathbf{J}_m$ and $x_j \in \mathbf{U}_m$, Eberhardt et at (2010) showed that one can derive linear constraints on the *total effects* entailed by the model. For the purposes of the present paper, it is much more useful to work with the *direct effects*. We can similarly derive the following linear constraints expressing the experimental effects as a linear sum of direct effects:

$$t(x_i \rightsquigarrow x_j \,||\, \mathbf{J}_m) = \sum_{x_k \in \mathbf{U}_m \setminus x_j} t(x_i \rightsquigarrow x_k \,||\, \mathbf{J}_m)\, b(x_k \to x_j) + b(x_i \to x_j) \quad (2)$$

For instance, for the experiment of Figure 1b, with $i = 3$ and $j = 4$, we get $t(x_3 \rightsquigarrow x_4 \,||\, \mathbf{J}_m) = t(x_3 \rightsquigarrow x_2 \,||\, \mathbf{J}_m) b(x_2 \to x_4) + b(x_3 \to x_4)$, which is easily verified. This equation holds for cyclic as well as acyclic systems, and derives from the definition of the experimental effect from x_i to x_j (see Section 1). When grouping all directed paths from x_i to x_j according to the final edge into x_j, each such group represents another experimental effect obtainable from $\mathcal{E}_m$. Note that the experimental effects $t(x_i \rightsquigarrow x_j \,||\, \mathbf{J}_m)$ and $t(x_i \rightsquigarrow x_k \,||\, \mathbf{J}_m)$ are numerical quantities estimated from the experiments, and the unknowns are the direct effects $b(x_k \to x_j)$, $\forall x_k \in \mathbf{U}_m \setminus x_j$, and $b(x_i \to x_j)$.

This alternative representation immediately lends itself to the identification of the underdetermination in the direct effects. All linear equations of the form of equation 2 can be written into a matrix equation $\mathbf{Kb} = \mathbf{k}$, where the unknown vector $\mathbf{b}$ groups the elements of the unknown matrix $\mathbf{B}$. A given element of $\mathbf{b}$ (and hence of $\mathbf{B}$) is undetermined if and only if that element is involved in the nullspace of the constraint matrix $\mathbf{K}$.

The above characterization does not provide much of an understanding of the underdetermination in terms of the graph structure. Nevertheless, consider the following. Any direct effects $b(\bullet \to x_j)$ into x_j are only constrained by experiments in which $x_j \in \mathbf{U}_m$. That is, the direct effects occur only in constraints of this type and there is only one (linearly independent) such constraint for each ordered pair $(\bullet, x_j)$ that the pair condition is satisfied for. Thus, in the general case, when the pair condition is not satisfied for a particular pair (x_i, x_j) then the entire j:th row of $\mathbf{B}$ is undetermined. Conversely, since the direct effects into x_j are the only direct effects that enter into these types of constraints, it follows that if the pair condition is satisfied for all pairs $(\bullet, x_j)$, then $n-1$ constraints can be determined and the row in $\mathbf{B}$ specifying the direct effects into x_j is fully identified. Hence, to *guarantee* the identifiability of a given direct effect $b(x_i \to x_j)$, it is necessary to satisfy the pair condition for all ordered pairs (x_k, x_j) with $k \neq j$. Note that in particular

graphs it may be possible to identify a direct effect $b(x_i \rightarrow x_j)$ even when the above condition is not true. In all cases, our code package provides the user with an explicit characterization of which coefficients are determined and which are not, given the results of any provided set of experiments.

3 Completeness of the procedure

An important question concerns whether the procedure introduced by Eberhardt et al (2010) fully exploits all the available data. Each experiment $\mathcal{E}_m = (\mathbf{J}_m, \mathbf{U}_m)$ supplies a data covariance matrix $\mathbf{\Sigma}^m_{\mathbf{x}}$, in which each entry $\mathbf{\Sigma}^m_{\mathbf{x}}[i, j]$ specifies the covariance between x_i and x_j in the experiment $\mathcal{E}_m$. The procedure as described (and the related pair condition theorem the authors gave) is based exclusively on constraints due to the experimental effects $t(x_i \rightsquigarrow x_j \,\|\, \mathbf{J}_m) = \mathbf{\Sigma}^m_{\mathbf{x}}[i, j]$ where $x_i \in \mathbf{J}_m$ and $x_j \in \mathbf{U}_m$. These covariances only constitute part of the information contained in a data convariance matrix $\mathbf{\Sigma}^m_{\mathbf{x}}$. In particular, the covariances between non-intervened variables, $\mathbf{\Sigma}^m_{\mathbf{x}}[j, k]$ with $x_j, x_k \in \mathbf{U}_m$, were not utilized at all.[3] It is tempting to think that this additional source of information could provide further leverage to identify the causal structure, and thereby reduce the demands for identifiability. However, we have the following negative result:

Lemma 1. *Let the true model generating the data be* $(\mathbf{B}, \mathbf{\Sigma_e})$. *For each of the experiments* $(\mathcal{E}_m)_{m=1,...,M}$ *the obtained data covariance matrix is* $\mathbf{\Sigma}^m_{\mathbf{x}}$. *If there is a direct effects matrix* $\widehat{\mathbf{B}} \neq \mathbf{B}$ *such that for all* $(\mathcal{E}_m)_{m=1,...,M}$ *and all* $x_i \in \mathbf{J}_m$ *and* $x_j \in \mathbf{U}_m$ *it produces the same experimental effects* $t(x_i \rightsquigarrow x_j \,\|\, \mathbf{J}_m)$, *then the model* $(\widehat{\mathbf{B}}, \widehat{\mathbf{\Sigma}}_{\mathbf{e}})$ *with* $\widehat{\mathbf{\Sigma}}_{\mathbf{e}} = (\mathbf{I} - \widehat{\mathbf{B}})(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Sigma_e}(\mathbf{I} - \mathbf{B})^{-T}(\mathbf{I} - \widehat{\mathbf{B}})^T$ *has data covariance matrices* $\widehat{\mathbf{\Sigma}}^m_{\mathbf{x}} = \mathbf{\Sigma}^m_{\mathbf{x}}$ *for all* $m = 1, ..., M$.

Proof. The proofs for all results given in this paper are provided in online supplementary material at: `http://cs.helsinki.fi/u/ajhyttin/exp/`

When $\mathbf{B}$ is underdetermined Lemma 1 constitutes a constructive proof that any measure of the covariance between two non-intervened variables provides no additional help with the identifiability of $\mathbf{B}$ in the model space considered in (Eberhardt et al., 2010). Intuitively, this result is a consequence of the dependence of the covariances between non-intervened variables on the model's disturbance covariance matrix $\mathbf{\Sigma_e}$. The additional $(n^2 + n)/2$ unknown parameters of $\mathbf{\Sigma_e}$ swamp the gains these covariance measures provide. Lemma 1 implies that the pair condition theorem can be strengthened to state that the method of Eberhardt et al. (2010) is complete with regard to the information contained in the data covariance matrices for the search space they consider.

Theorem 1 (Completeness Theorem)**.** *Given the data covariance matrices from a sequence of experiments* $(\mathcal{E}_m)_{m=1,...,M}$ *over the variables in* $\mathbf{V}$, *all direct effects* $b(x_i \rightarrow x_j)$ *are identified if and only if the pair condition is satisfied for all ordered pairs of variables w.r.t. these experiments.*[4]

However, measures of the covariances between non-intervened variables *are necessary* to identify the disturbance covariance matrix $\mathbf{\Sigma_e}$ (specifying the latent variables). In the original procedure $\mathbf{\Sigma_e}$ was determined using measurements from an additional passive observational dataset (with $\mathbf{J}_m = \emptyset$). It can be shown that a much weaker condition, similar to the pair condition for experimental effects, is necessary and sufficient for the identification of $\mathbf{\Sigma_e}$, if $\mathbf{B}$ is already determined. We can thus state the following general theorem of model identifiability:

Theorem 2 (Model Identifiability Theorem)**.** *Given a sequence of experiments* $(\mathcal{E}_m)_{m=1,...,M}$ *over the variables in* $\mathbf{V}$ *the model* $(\mathbf{B}, \mathbf{\Sigma_e})$ *is fully identified if and only if for each ordered pair of variables* (x_i, x_j) *there is an experiment* $\mathcal{E}_b = (\mathbf{J}_b, \mathbf{U}_b)$ *with* $x_i \in \mathbf{J}_b$ *and* $x_j \in \mathbf{U}_b$ *and another experiment* $\mathcal{E}_e = (\mathbf{J}_e, \mathbf{U}_e)$ *with* $x_i, x_j \in \mathbf{U}_e$.

Thus, the good news is that the algorithm given by (Eberhardt et al., 2010) does as well as it possibly could with regard to identifiability. The bad news is that the generality of its search space implies that the conditions for identifiability are very demanding. Hence, in the following section we consider how the use of background knowledge or an additional assumption of faithfulness can help.

[3] Obviously, when two variables are both in $\mathbf{J}_m$, then the covariance between them in that experiment is zero by assumption, since simultaneous interventions are assumed to make the intervened variables independent.

[4] Note the inevitable limitation of identifiability with regard to self-loops discussed in (Eberhardt et al., 2010).

4 The faithfulness assumption

When two variables are independent one commonly assumes that they are not causally connected. However, this assumption is non-trivial, since it precludes, for example, cases where two variables are connected by two separate pathways that exactly cancel each other out. The two variables are then probabilistically independent, while they are still causally connected by a directed path.

For instance, in the model of Figure 2a, in an experiment where x_1 is randomized, and x_2 and x_3 are (passively) observed, x_1 and x_3 would be found marginally independent, but dependent conditional on x_2. Such an observation could have many possible alternative explanations, two of which are shown in Figures 2b and 2c. In such cases scientists do often make the assumption that an absence of a correlation is an indication of the absence of a causal connection, and hence favor explanations (b) and (c) over (a). In this section we introduce inference rules that take advantage of this intuition.

The structure of a model entails certain marginal and conditional independencies in the resulting distribution; these are characterized by the *Markov condition*, and Spirtes (1995) has shown that the familiar concept of d-separation specifies all and only the independencies entailed in all linear structural equation models, including cyclic (non-recursive) models. The intuition given above is then formalized in the *faithfulness* assumption, which states that all independencies in the population distribution are derived from the structure of the graph, rather than specific parameter values (Spirtes et al., 2000; Pearl, 2000).

For maximum generality, the algorithm in (Eberhardt et al., 2010) did not use the assumption of faithfulness. However, given the demanding identifiability conditions (see Section 1), it is worth investigating whether faithfulness might add substantial benefit when the pair condition is not satisfied for all ordered pairs of variables.

In general, causal discovery based on faithfulness proceeds in two steps. First, independence tests are used to detect the absence of edges between pairs of variables. Subsequently, the detected absences are used to 'orient' as many as possible of the remaining edges. We here employ an analogous approach.

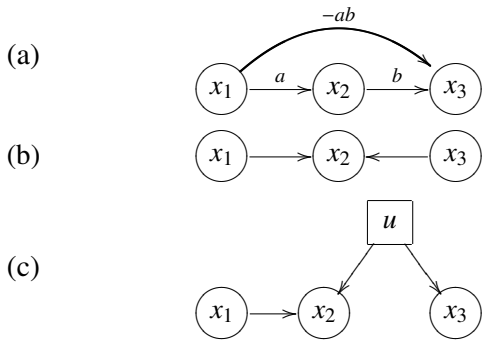

Figure 2: Example graphs. In (c) u is unmeasured.

First, we rely on the fact that if, in any experiment $\mathcal{E}_m = (\mathbf{J}_m, \mathbf{U}_m)$, two non-intervened variables $x_i, x_j \in \mathbf{U}_m$ are marginally or conditionally independent (with any conditioning set not including x_i and x_j), then by faithfulness $b(x_i \to x_j) = b(x_j \to x_i) = \mathbf{\Sigma_e}[i, j] = 0$. Similarly, if $x_i \in \mathbf{J}_m$ and $x_j \in \mathbf{U}_m$ are found marginally or conditionally independent, faithfulness requires that $b(x_i \to x_j) = 0$. (Note that while independencies imply the absence of edges, dependencies do not necessarily imply the presence of any edge between a given pair of edges.) In our implementation we run the statistically and computationally efficient schedule of independence tests suggested by the PC-algorithm (Spirtes et al., 2000) on the data from each experiment separately. Althought PC is designed for a different search space, any indepencies found are usable in our procedure as well. Any obtained constraints are termed *skeleton constraints*.

The orientation rules of the second step of the inference are more intricate. We cannot simply adopt the orientation rules from existing constraint-based algorithms since they only provide orientation rules for search spaces where the true causal structure either contains latent variables but no cycles (FCI, (Spirtes et al., 2000)) or contains cycles but no latent variables (CCD, (Richardson, 1996)). Since our model space contains both latent variables and cycles, and we have the advantage of experiments, different orientation rules are required. We employ the following two rules that take advantage of the orientation supplied by interventions. Any constraints thus obtained are termed *orientation constraints*:

(1) If in a given experiment we have $x_i \in \mathbf{J}_m$ and $x_j, x_k \in \mathbf{U}_m$, and $t(x_i \rightsquigarrow x_j \,||\, \mathbf{J}_m) \neq 0$ but $t(x_i \rightsquigarrow x_k \,||\, \mathbf{J}_m) = 0$, then $b(x_j \to x_k) = 0$ by faithful-

ness. For instance, in the model of Figure 1a, in an experiment with $\mathbf{J}_m = \{x_3\}$, we see an experimental effect from x_3 to x_2, but no experimental effect from x_3 to x_1. We would thus infer that $b(x_2 \to x_1) = 0$. Similarly we would infer that $b(x_4 \to x_1) = 0$. This rule is sound because by the antecedent there is a directed path from x_i to x_j so that, were there a non-zero direct effect $b(x_j \to x_k)$ it would follow that there would be a directed path from x_i to x_k, which for a faithful model would imply a non-zero experimental effect of x_i on x_k.

(2) Again, if we have $x_i \in \mathbf{J}_m$ and $x_j, x_k \in \mathbf{U}_m$, and observe $t(x_i \rightsquigarrow x_j \,||\, \mathbf{J}_m) \neq 0$, and in addition x_i is conditionally independent of x_k given x_j, then we infer that $b(x_k \to x_j) = 0$ and $\mathbf{\Sigma_e}[j,k] = 0$. The rule is correct under faithfulness because we must have a directed path from x_i to x_j, so if there existed a direct effect from x_k to x_j (or a confounder between the two) by faithfulness this would cause a dependence between x_i and x_k when conditioned on the collider x_j.

Since all the new constraints are (trivially) linear in the direct effects, they can be directly added to the set of constraints on the direct effects given by the experimental effects described in Section 2. The combined system can then be solved and the underdetermination characterized as before.

We note that the above rules clearly do not exhaust the inferences that could (potentially) be drawn by faithfulness. It is an open (and intriguing!) problem to devise a set of *complete* rules for causally insufficient, cyclic discovery.

Finally, if by domain knowledge we are guaranteed that x_i does not have a direct effect on x_j (with respect to $\mathbf{V}$), then we may naturally add the constraint $b(x_i \to x_j) = 0$. Such prior knowledge may be particularly useful for dense graphs or models which are close to unfaithful, when the faithfulness rules would not apply or would be unreliable.

5 Adaptive selection of experiments

While the form of the constraints obtained from the experimental effects (given in Section 2) can be predicted ahead of performing the experiments, constraints due to faithfulness come as an unexpected 'bonus': We cannot know ahead of time which independencies will be uncovered. Hence, to minimize the total number of experiments, one must react and adapt the sequence to newly discovered constraints.

We have found that a simple greedy selection procedure works well. As in the original procedure of Eberhardt et al (2010), we keep a list of which ordered pairs have the pair condition satisfied. However, in addition to pairs satisfied purely on the basis of the choice of previous experiments $\mathcal{E}_m$, we also treat any pair (x_i, x_j) as if it is satisfied whenever, using the characterization of underdetermination in Section 2, the coefficient $b(x_i \to x_j)$ is determined. This includes both coefficients directly determined by background knowledge or our faithfulness rules, as well as coefficients indirectly determined by the collection of all existing constraints. The next experiment is selected such that we maximize the number of ordered pairs for which the pair condition is guaranteed to be satisfied after the experiment, arbitrarily breaking ties. In the following simulations, we demonstrate that, for sparse graphs, this is an effective selection protocol.

6 Simulations

In this section, we describe a set of simulations on random graphs that we used to investigate the power provided by the faithfulness assumption.[5]

We generated a large number of random graphs over 10 variables, with sparsity ranging from zero edges up to 60 edges (out of 135 possible, counting both direct and confounding edges). The coefficients were drawn uniformly from $[0.3, 0.8]$ with random sign, and stability was examined by checking the eigenvalues of the resulting $\mathbf{B}$.

First, we study the theoretical limit behavior (infinite sample limit) of our procedure. In Figure 3a, we plot the average number of experiments needed to completely identify the model, as a function of the underlying model sparsity and the number of interventions per experiment. For one intervention per experiment (left panel), in the absence of faithfulness rules the full 10 experiments are needed regardless of sparsity, while for relatively sparse graphs on average a few experiments can be saved by utilizing faithfulness. When intervening on three variables per experiment (right panel), 7 experiments

[5] We encourage the interested reader to try out the method. A complete implementation (reproducing all the simulations) is available at: `http://cs.helsinki.fi/u/ajhyttin/exp/`

are needed in the basic case to satisfy the pair condition for all pairs, and significant savings can be obtained when using the faithfulness assumption. Meanwhile, Figure 3b shows the number of ordered pairs (a lower bound of the rank of the constraint matrix) satisfied after only three experiments, as a function of sparsity. It can be seen that for sparse graphs, most of the structure of the graph has already been discovered at this stage of the sequence of experiments.

Second, we look at finite sample behavior. Figure 4 shows the number of experiments used, as well as the resulting accuracy (linear correlation between estimated and true coefficients). In each experiment, 10,000 samples were used. We note the following: To guarantee high accuracy in dense graphs, the pair condition must be satisfied for all pairs based on the experimental setup alone (as in the 'no faithfulness' procedure). However, when the true model is sparse, significant savings in terms of the number of experiments are possible. Especially when intervening on several variables in each experiment, the full model is typically identified with high accuracy in just 4 experiments. Accuracy drops markedly for dense graphs, as the number and size of possible conditioning sets is so large that inevitably some dependent variables are mistakenly inferred to be independent, yielding large errors. These erroneous inferences cause the number of experiments to stay roughly constant as a function of the number of edges in the graph, in marked contrast to the infinite sample limit of Figure 3a.

7 Application to flow cytometry data

Finally, we applied the algorithm (with the faithfulness rules) to the flow cytometry data of Sachs et al (2005). In this data set only 4 of the 11 measured variables were manipulated with no changes made to the background conditions, see (Eberhardt et al, 2010) for details. This meant that the pair condition was satisfied for only 40 ordered pairs out of the total of 110. Together with the faithfulness rules, however, these experiments were enough to determine a majority of the direct effects in the model.

We ran the inference procedure with a variety of parameter settings (significance threshold for statistical dependence, using the full faithfulness rules or

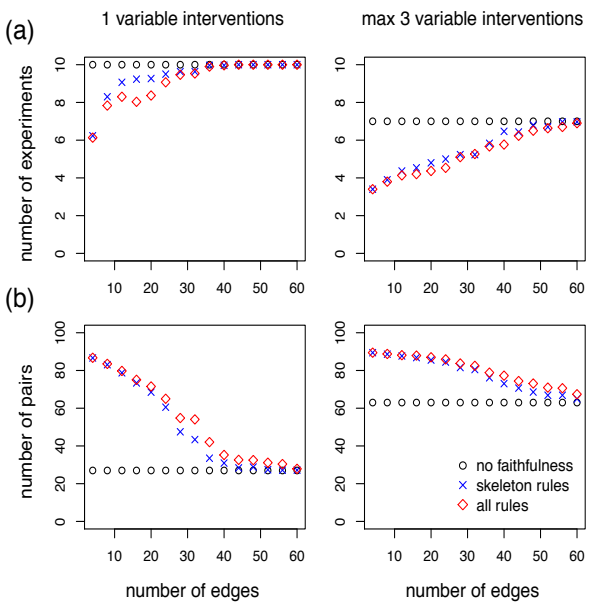

Figure 3: Performance of the procedure in the infinite-sample limit. (a) Number of experiments needed to identify the full model, and (b) amount of structure discovered after three experiments, as a function of the number of edges in the graph.

only the skeleton rules, threshold for detecting determined vs undetermined coefficients, etc). A typical result is shown in Figure 5. We emphasize that our method assumes linearity, while the true model is likely to be at least somewhat non-linear. Thus the main interest lies in the resulting structure, and possibly the signs of the direct effects. While there were differences in the inferred graphs, many features were common to all of our results.

In particular, we always find (a) the well known Raf→Mek→Erk pathway (and invariably, in addition, Mek seems to have a direct effect on Raf), (b) PKC influences (directly or indirectly) a number of targets, including Raf, PKA, and Jnk, and (c) a strong association between PIP2 and PIP3 (and these are sometimes though not always connected with Plcg). These features are quite compatible with the 'ground truth' (from the literature) model given by Sachs et al (2005). However, our procedure also suggests that many of the variables have effects *into* PKA, something not supported by their model. Finally, we note that our method quite often detects bidirectional relationships; at this point, we do not

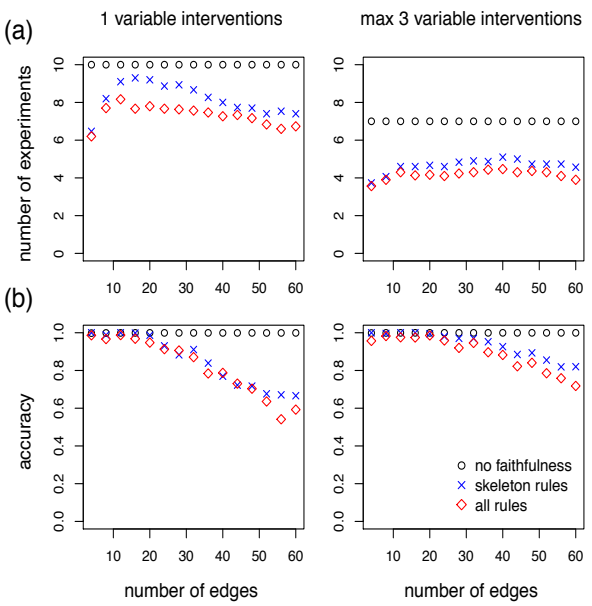

Figure 4: Results on simulated (sample) data. (a) Number of experiments needed to identify the model, and (b) accuracy (correlation between the estimates and the true values), as a function of the number of edges in the graph.

know whether this is due to the nature of our procedure or whether this is a true feature of the data.

8 Conclusions

The discovery procedure is relatively unique in the generality of the model space considered. While there exists a large body of work on learning *acyclic* causal structures with or without hidden variables, there is comparatively little on learning models involving feedback loops. Richardson (1996) gave a constraint-based discovery procedure for passive observational data, but did not allow for latent variables. More recently, both Schmidt and Murphy (2009) and Itani et al (2010) have introduced probabilistic models for cyclic structures involving discrete-valued variables, and given related discovery procedures. While all of these methods use somewhat different models and assumptions, ultimately they nevertheless all share the goal of elucidating causal structure among variables that are recurrently connected. A thorough empirical study, comparing the various methods both on simulations and on a number of real datasets, would be an important next step.

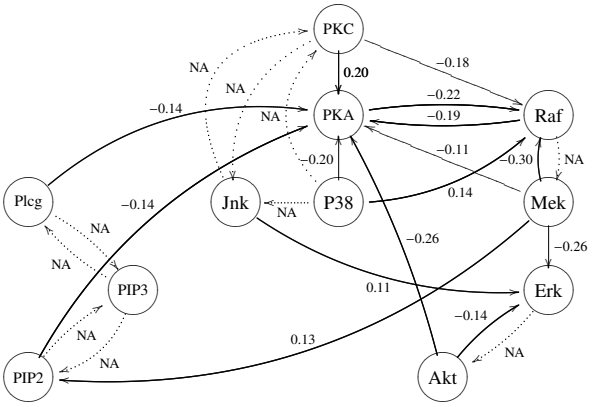

Figure 5: Protein interaction graph inferred from the dataset of Sachs et al (2005), with 'NA' denoting non-identified edge strengths (which could potentially be zero, hence these edges are plotted with dotted lines). Settings: significance threshold 0.05, only skeleton rules.

Acknowledgments

A.H. and P.O.H. were funded by Univ. of Helsinki Research Funds and the Academy of Finland.

References

K. A. Bollen. 1989. *Structural Equations with Latent Variables*. John Wiley & Sons.

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

S. Itani, M. Ohannessian, K. Sachs, G. P. Nolan, and M. A. Dahleh. 2010. Structure learning in causal cyclic networks. In *JMLR W&CP*, volume 6, pages 165–176.

J. Pearl. 2000. *Causality*. Oxford University Press.

T. Richardson. 1996. *Feedback Models: Interpretation and Discovery*. Ph.D. thesis, Carnegie Mellon.

K. Sachs, O. Perez, D. Pe'er, D.A. Lauffenburger, and G.P. Nolan. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529.

M. Schmidt and K. Murphy. 2009. Modeling discrete interventional data using directed cyclic graphical models. In *UAI '09*.

P. Spirtes, C. Glymour, and R. Scheines. 2000. *Causation, Prediction and Search*. MIT Press, 2 edition.

P. Spirtes. 1995. Directed cyclic graphical representation of feedback models. In *UAI'95*.