Principled selection of impure measures for consistent learning of linear latent variable models

Ricardo Silva
Gatsby Computational Neuroscience Unit
University College London
rbas@gatsby.ucl.ac.uk

Abstract

In previous work, we have developed a principled way of learning the causal structure of linear latent variable models (Silva et al., 2006). However, we have considered the case for models with *pure measures* only. Pure measures are observed variables that measure no more than one latent variable. This paper presents theoretical extensions that justify the selection of some types of impure measures, allowing us to discover hidden variables that could not be identified in the previous case.

1 CONTRIBUTION

Linear latent variable models are graphical models where the distribution is Markov with respect to a directed acyclic graph (DAG), and each variable X_i is a linear combination of its parents $\{X_{i(1)}, X_{i(2)}, \ldots, X_{i(p_i)}\}$ with additive noise:

$$X_i = \lambda_{i1} X_{i(1)} + \lambda_{i2} X_{i(2)} + \cdots + \lambda_{i(p_i)} X_{i(p)} + \epsilon_i \quad (1)$$

We will consider as parameters the set of linear coefficients $\{\lambda\}$ and the covariance matrix of error terms $\{\epsilon\}$, which are assumed to have mean zero. In this work, we do not make use of higher-order moments for the error terms (which would not be informative anyway in the case of Gaussian variables). This is particularly relevant if sample sizes are small and higher-order moments cannot be estimated reliably.

These models, under a causal interpretation, allow for the prediction of effects of interventions (Spirtes et al., 2000). In particular, we consider the case where no observed variable is a cause of any latent variable. Bollen (1989) provides numerous examples of problems of this kind. Because of this assumption, we can define the *measurement model* of the latent variable graph as the subgraph that indicates which observed variables *measure* (i.e., are children of) which hidden variables. Figure 1 depicts a latent variable graph. The measurement model can be interpreted as the graph obtained when we ignore edges between hidden variables.

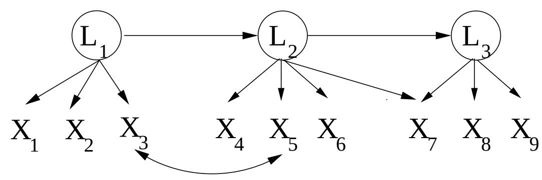

Figure 1: A latent variable graph. Hidden variables are enclosed in circles. The bi-directed edge denotes the presence of other hidden common causes for X_3 and X_5 (Silva et al., 2006; Richardson, 2003). X_3, X_5 and X_7 are impure because they measure more than one hidden variable in this system.

Our causal inference problem consists on discovering graphical structure that explains the observational data. Although we assume that the true model is a linear latent variable model where no observed variable is a cause of a latent variable, we do not want to require any other knowledge concerning the underlying structure (i.e., which hidden variables exist and which observed variables measure which latents). Such knowledge is certainly helpful, if available, but it is not desirable to make it mandatory.

In causal discovery, a fundamental challenge is the *identification problem*: given observational data for the observable variables only, how to reconstruct the set of causal models that could have generated this data? In latent variable modeling, the potential number of models is infinite in the worst-case (Spirtes et al., 2000). However, some problems will have partially identifiable structure. Figure 2 illustrates a case where the subgraph with L_1 and a few of its measures is identifiable.

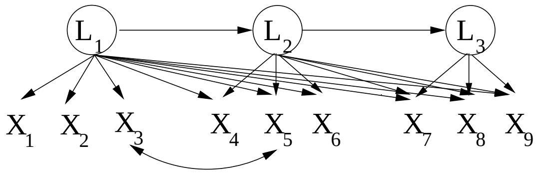

Figure 2: If our data is generated by a causal graph as the one above, there will be multiple choices of parameter values corresponding to the same marginal distribution for $\{X_1, \ldots, X_9\}$. Such parameters could be explained by many different structures. We could, however, discover identifiable structure for a submodel containing, for instance, $\{L_1, X_1, X_2, X_3, X_4\}$.

In previous work (Silva et al., 2006), we presented an algorithm, consistent with probability 1 in the limit of infinite data, which is able to recover substructures where: i. each latent has at least three observed measures (children); ii. each measure is *pure*, in the sense that they measure no more than one latent in the resulting structure. For example, in Figure 1, X_3, X_5 and X_7 are not pure measures, since they measure more than one latent[1]. All other observed variables are pure measures.

This means there are problems where hidden variables cannot be discovered by such an approach, because they fail to have three pure measures. Figure 3 illustrates one such case. If we choose to include features X_1, X_2, X_3 in our model, then L_1 can be identified, but not L_2: X_5 cannot be selected, since this would imply both X_3 and X_5 being impure, which means not enough pure measures would be left to measure L_2 (only X_4, X_6).

It is hard to avoid the need for at least three measures for each latent variable using no more than the second moments of the data. A treatment of such a condition is beyond the scope of this paper. However, the need for at least three *pure* measures can be relaxed, as hinted at the conclusion of Silva et al. (2006). Principled ways for learning impure structures is the contribution of this paper.

This feature selection problem is a problem of including impure measures (and the respective latents they measure) in a model by identifying its causal connections to other variables, as well the causal connections of their latent parents. An impure measure X_i will only be included in our learned graphical structure if it fulfills the following desiderata:

1. an impure measure X_i can be selected only if it is possible to characterize all causal models that explain the correlation of X_i and the other selected measures;

2. an impure measure X_i can be selected only if it can be used to obtain a consistent estimate of the covariance matrix of the latent variables in the discovered model;

[1]We only count latents with more than one observed child among the variables present in the submodel.

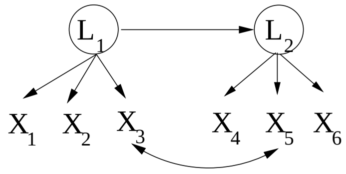

Figure 3: Although in the model above there is no pure measurement model for both L_1 and L_2 with at least three measures per latent, the impure model is still identifiable, and so is the linear coefficient corresponding to edge $L_1 \to L_2$.

This second desideratum is necessary if one wants to learn the causal model among latent variables. For instance, one should not include all observed variables of Figure 2, because the causal effect of L_2 into L_3 cannot be identified (Silva et al., 2006)[2].

2 MAIN RESULTS

We describe theoretical results for identification of causal structure with impure measures. These results are testable conditions that can be verified with the observable data[3].

First, a piece of notation. For a given set of four random variables, $\{W, X, Y, Z\}$, we say that the predicate $T(W, X, Y, Z)$ is true if and only if:

$$\sigma_{WX}\sigma_{YZ} = \sigma_{WY}\sigma_{XZ} = \sigma_{WZ}\sigma_{XY}$$
$$\sigma_{WX} \times \sigma_{YZ} \times \sigma_{WY} \times \sigma_{XZ} \times \sigma_{WZ} \times \sigma_{XY} \neq 0$$

where σ_{XY} is the covariance of variables X and Y. These type of constraints are also known as *tetrad constraints.* Standard statistical tests for such constraints exist (Spirtes et al., 2000; Silva et al., 2006).

Although not intuitive, this result implies the existence of one variable L conditioned on which $\{W, X, Y, Z\}$ are independent (Silva et al., 2006, Lemma 9). If there

[2]In fact, only L_1 can be identified by our procedure, resulting in a measurement model for one latent variable only.

[3]We also assume the faithfulness condition, common to other methods for inferring causal structure (Spirtes et al., 2000).

is no observed variable with this property, which can be tested with partial correlation constraints (Spirtes et al., 2000), then L has to be a latent variable. For the rest of this paper, we will assume that variables corresponding to constraints $T(\cdot, \cdot, \cdot, \cdot)$ are all hidden variables.

For example, $T(X_1, X_2, X_3, X_4)$ holds in all models depicted in the previous Figures. $T(X_1, X_2, X_3, X_5)$ does not hold. Also, $T(X_6, X_7, X_8, X_9)$ does not hold in Figure 1, since L_2 defines a path between X_6 and X_7 that does not include L_3.

We will describe results that capture two large classes of impure models. Each result has implications both in the measurement model (how is the impure variable causally connected to other variables?) and in the structure of latent variables (can I have a linear latent variable model with such a variable that allows me to estimate the covariance among latents?).

2.1 Case 1

Our first main result allows for the discovery of structures such as the one depicted in Figure 3. In this case, two variables are measuring some hidden common cause that is not measured by any other observed variable in the system (hence the simplified notation with bi-directed edges). Models with such a relationship are easily found in the literature of latent variable models (Bollen, 1989; Silva and Ghahramani, 2006).

Given six random variables $\{A, B, C, X, Y, Z\}$ assumed to measure unknown hidden variables within a linear latent variable model, the following lemma holds:

Lemma 1 *If*

$$\begin{array}{c} T(A,B,C,X) = T(A,B,C,Y) = true \\ T(A,X,Y,Z) = T(B,X,Y,Z) = true \\ T(A,B,X,Y) = false \\ T(A,B,C,Z) = T(X,Y,Z,C) = false \end{array}$$

and all entries in the covariance matrix of $\{A, B, C, X, Y, Z\}$ are non-zero, then

- *there is a latent T_1 in the true model that d-separates $\{A, B, C\}$;*
- *there is a latent T_2 in the true model that d-separates $\{X, Y, Z\}$;*
- *$T_1 \neq T_2$;*
- *T_1 (as well as T_2) d-separates every pair in $\{A, B, C, T_1\} \times \{X, Y, Z, T_2\}$, except (C, Z);*
- *at most one element of $\{A, B, C\}$ shares a hidden common cause with T_1, at most one element of $\{X, Y, Z\}$ shares a hidden common cause with T_2, and at most one element of $\{A, B, X, Y\}$ shares a hidden common cause with its respective latent parent in $\{T_1, T_2\}$;*
- *C is not a cause of Z is vice-versa, but they have extra hidden common causes in the true model that are different from T_1, T_2;*

The proof of this result and all other results here described are given in the appendix.

Notice that the graph in Figure 3 is a graph that corresponds to the description given in Lemma 1. However, this two-latent graph cannot be distinguished from the graph in Figure 4. Nevertheless, at most one measure might have an extra confounder factor between itself and the respective latent in the two-latent graph.

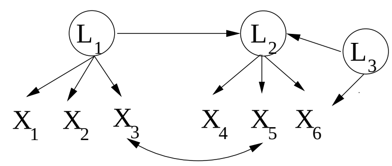

Figure 4: The following graph is undistinguishable from the graph Figure 3 according to our procedure.

Let $G_{2L}(A, B, C, X, Y, Z)$ be the latent variable graph constructed as follows:

- G_{2L} has two latent variables, T_1 and T_2, where T_1 is a parent of $\{A, B, C\}$ and T_2 is a parent of $\{X, Y, Z\}$;
- G_{2L} contains bi-directed edges $T_1 \leftrightarrow T_2$ and $C \leftrightarrow Z$;

Notice that this is not a DAG anymore, but what is sometimes known as *directed mixed graph* (DMG) (Richardson, 2003; Silva and Ghahramani, 2006). This notation is useful because we do not want to represent which other latents are hidden common causes of $\{T_1, T_2\}$ or $\{C, Z\}$. Parameterization is analogous to the DAG case, except that the pairs of error terms $\{\epsilon_{T_1}, \epsilon_{T_2}\}$, $\{\epsilon_C, \epsilon_Z\}$ now have non-zero covariances (Bollen, 1989).

The relevance of this construction is given by the following lemma:

Lemma 2 *Let $\mathbf{S} = \{A, B, C, X, Y, Z\}$ be a set of random variables with covariance matrix $\Sigma_\mathbf{S}$ generated by an unknown linear latent variable model. If $\{A, B, C, X, Y, Z\}$ satisfies the conditions of Lemma 1, then there is an unique Gaussian latent variable DMG model (up to the scale and sign of the latents) defined by $G_{2L}(A, B, C, X, Y, Z)$ such that the marginal covariance for $\mathbf{S}$ according to this model equals $\Sigma_\mathbf{S}$. In this case, the covariance of T_1 and T_2 defined by this model equals the covariance of the respective latents in the true model containing $\mathbf{S}$.*

The implication is that any consistent estimator for $\Sigma_{\mathbf{S}}$, as defined by this model, will provide a consistent estimator for the covariance of the represented latents.

2.2 Case 2

The second case we consider is simpler, and motivated by Figure 5. Although here there is a pure submodel where all latents are present and contain three pure measures, keeping variable Z in the model allows for more robust estimates of the covariance between latent variables, since more data becomes available. This represents a class of problems where an observed variable measures more than one latent present in the model (as opposed to Case 1, where the other parents of an observed variable are not represented explicitly).

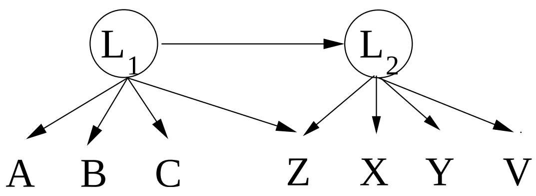

Figure 5: Keeping impure measure Z allows for a more robust estimation of the covariance between latents.

As before, start by assuming that a set of random variables $\{A, B, C, X, Y, Z, V\}$ are observed variables within a linear latent variable model.

Lemma 3 *If*

$$T(A, B, C, K) = true, \text{ for } K \in \{X, Y, V, Z\}$$
$$T(K, X, Y, V) = true, \text{ for } K \in \{A, B, C\}$$
$$T(K_1, K_2, K_3, K_4) = false, \text{ for}$$
$$\{K_1, K_2\} \subset \{A, B, C\}, \{K_3, K_4\} \subset \{X, Y, V\}$$
$$T(X, Y, Z, V) = true$$
$$T(A, X, Y, Z) = T(A, B, C, X) = false$$

and all entries in the covariance matrix of $\{A, B, C, X, Y, Z, V\}$ *are non-zero, then*

- *there is a latent* T_1 *in the true model that d-separates* $\{A, B, C\}$*;*
- *there is a latent* T_2 *in the true model that d-separates* $\{X, Y, Z, V\}$*;*
- $T_1 \neq T_2$*;*
- T_1 *d-separates every pair in* $\{A, B, C, Z, T_1\} \times \{X, Y, V, T_2\}$*, except* (Z, T_2)*;*
- T_2 *d-separates every pair in* $\{A, B, C, T_1\} \times \{X, Y, Z, V, T_2\}$*, except* (Z, T_1)*;*
- *at most one element of* $\{A, B, C, Z\}$ *shares a hidden common cause with* T_1*, at most one element of* $\{X, Y, Z, V\}$ *shares a hidden common cause with* T_2*, and at most one element of* $\{A, B, C, X, Y, V\}$ *shares a hidden common cause with its respective latent ancestor* $\{T_1, T_2\}$*;*

Again, this allows for the identification of a graph such as the one in Figure 5 up to the possibility of having one measure sharing a hidden common cause with its respective latent parent, similar to Figure 4. On top of that, Z might share another common hidden cause with its parents.

Regarding estimating the covariance between latents, we also have an analogous result. Let $G_{2L'}$ be the graph constructed as follows: $G_{2L'}$ has two latent variables, T_1 and T_2, where T_1 is a parent of $\{A, B, C, Z\}$ and T_2 is a parent of $\{X, Y, Z, V\}$. This leads to the following result:

Lemma 4 *Let* $\mathbf{S} = \{A, B, C, X, Y, Z, V\}$ *be a set of random variables with covariance matrix* $\Sigma_{\mathbf{S}}$ *generated by an unknown linear latent variable model. If* $\mathbf{S}$ *satisfies the conditions of Lemma 3, then there is an unique Gaussian latent variable DAG model (up to the scale and sign of the latents) defined by* $G_{2L'}(A, B, C, X, Y, Z, V)$ *such that the marginal covariance for* $\mathbf{S}$ *according to this model equals* $\Sigma_{\mathbf{S}}$*. In this case, the covariance of* T_1 *and* T_2 *defined by this model equals the covariance of the respective latents in the true model containing* $\mathbf{S}$*.*

3 CONCLUSION

Feature selection is also an identifiability problem, as illustrated by our results. Although the identification results refer to pairs of latents, they can be combined to infer models with several latents, in the same spirit of Silva et al. (2006). In a future work, we plan to extend the algorithm described by Silva et al. (2006) to include these latest results, including robust implementations less sensitive to statistical mistakes. This can be done in principle by combining the Bayesian approaches of Silva and Scheines (2006) and Silva and Ghahramani (2006).

References

K. Bollen. *Structural Equation Models with Latent Variables.* John Wiley & Sons, 1989.

T. Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30:145–157, 2003.

R. Silva and Z. Ghahramani. Bayesian inference for Gaussian mixed graph models. *22th UAI*, 2006.

R. Silva and R. Scheines. Bayesian learning of measurement and structural models. *23rd ICML*, 2006.

R. Silva, R. Scheines, C. Glymour, and P. Spirtes. Learning the structure of linear latent variable models. *JMLR*, 7:191–246, 2006.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search.* Cambridge U. Press, 2000.

APPENDIX: PROOFS

The following proofs uses definitions such as choke points, treks and d-separation. For definitions and examples of such concepts, see Appendix B of Silva et al. (2006).

Lemma 1 *If*

$$T(A, B, C, X) = T(A, B, C, Y) = true$$
$$T(A, X, Y, Z) = T(B, X, Y, Z) = true$$
$$T(A, B, X, Y) = false$$
$$T(A, B, C, Z) = T(X, Y, Z, C) = false$$

and all entries in the covariance matrix of $\{A, B, C, X, Y, Z\}$ *are non-zero, then*

- *there is a latent* T_1 *in the true model that d-separates* $\{A, B, C\}$*;*
- *there is a latent* T_2 *in the true model that d-separates* $\{X, Y, Z\}$*;*
- $T_1 \neq T_2$*;*
- T_1 *(as well as* T_2*) d-separates every pair in* $\{A, B, C, T_1\} \times \{X, Y, Z, T_2\}$*, except* (C, Z)*;*
- *at most one element of* $\{A, B, C\}$ *shares a hidden common cause with* T_1*, at most one element of* $\{X, Y, Z\}$ *shares a hidden common cause with* T_2*, and at most one element of* $\{A, B, X, Y\}$ *shares a hidden common cause with its respective latent parent in* $\{T_1, T_2\}$*;*
- C *is not a cause of* Z *is vice-versa, but they have extra hidden common causes in the true model that are different from* T_1, T_2*;*

Proof: By Lemma 9 of Silva et al. (2006) and the hypothesis $T(A, B, C, X) = true, T(A, X, Y, Z) = true$, it follows that there is a single latent T_1 that d-separates $\{A, B, C\}$, and a single latent T_2 that d-separates $\{X, Y, Z\}$ (since for simplicity we are assuming in this paper that no observed variable d-separates other pairs of observed variables, T_1 and T_2 are latents).

By adding the hypothesis $T(A, B, X, Y) = false$, we have that $T_1 \neq T_2$. One can show that by contradiction, starting by assuming $T_1 = T_2$ and then showing that would imply $T(A, B, X, Y) = true$. The argument is as follows. Assume $T_1 = T_2 = T$. By the Tetrad Representation Theorem (Spirtes et al., 2000; Silva et al., 2006) and the hypothesis, all treks connecting $\{A, B, X, Y\}$ pass through T, at least one trek exist between each pair (due to the faithfulness condition and the non-zero covariances between such variables), and at most one variable in $\{A, B, X, Y\}$ is connected to T by a trek into T. This implies $T(A, B, X, Y) = true$, contrary to our hypothesis.

Because $T(A, B, C, Z) = T(X, Y, Z, C) = false$, neither T_1 nor T_2 can d-separate C and Z. T_1 d-separates A from T_2, or otherwise there would be a trek connecting A to either X or Y that does not go through T_1 (by concatenating a trek $A - T_2$ with a trek $T_2 - X$, assuming, without loss of generalization, that X is not connected to T_2 by a trek into T_2). By analogy and symmetry, the other d-separations stated in the theorem hold.

C cannot be a cause of Z, because this implies a trek $T_1 - C \rightarrow \cdots \rightarrow Z$ that does not include T_2 (since T_2 is not a descendant of C by assumption): this trek contradicts the fact that T_2 d-separates $\{B, X, Y, Z\}$, since it creates trek $B - T_1 - C \rightarrow \cdots \rightarrow Z$. By symmetry, Z is not a cause of C. Because they are not d-separated by either T_1 nor T_2, other hidden common causes of C and Z should exist.

At most one element of $\{A, B, X, Y\}$ shares a hidden common cause with its respective latent parent in $\{T_1, T_2\}$. This proof is analogous to the proof of Theorem 15 in (Silva et al., 2006). $\square$.

Lemma 2 *Let* $\mathbf{S} = \{A, B, C, X, Y, Z\}$ *be a set of random variables with covariance matrix* $\Sigma_{\mathbf{S}}$ *generated by an unknown linear latent variable model. If* $\{A, B, C, X, Y, Z\}$ *satisfies the conditions of Lemma 1, then there is an unique Gaussian latent variable DMG model (up to the scale and sign of the latents) defined by* $G_{2L}(A, B, C, X, Y, Z)$ *such that the marginal covariance for* $\mathbf{S}$ *according to this model equals* $\Sigma_{\mathbf{S}}$*. In this case, the covariance of* T_1 *and* T_2 *defined by this model equals the covariance of the respective latents in the true model containing* $\mathbf{S}$*.*

Proof: Let the parameters of the linear latent variable model Markov to G_{2L} to be as follows:

$$A = \lambda_A T_1 + \epsilon_A$$
$$B = \lambda_B T_1 + \epsilon_B$$
$$C = \lambda_C T_1 + \epsilon_C$$
$$X = \lambda_X T_2 + \epsilon_X$$
$$Y = \lambda_Y T_2 + \epsilon_Y$$
$$Z = \lambda_Z T_2 + \epsilon_Z$$

and let Σ_ϵ be the covariance matrix of error terms, with non-diagonal elements set to zero, except the entries $\sigma_{\epsilon_C \epsilon_Z} = \sigma_{\epsilon_Z \epsilon_C}$. Finally, the last parameter is the covariance matrix $\Sigma_{\mathbf{T}}$ of $\{T_1, T_2\}$.

From the previous lemma, T_1 and T_2 correspond to true latents in the true model. To determine the scale and sign of such latents without loss of generalization, we will set $\lambda_A = \lambda_X = 1$ (Bollen, 1989). Set the parameter $\Sigma_{\mathbf{T}}$ to be the same as the population covariance of $\{T_1, T_2\}$ in the true model. Now, set $\lambda_B = \sigma_{BT_1}\sigma^2_{T_1}$, where σ^2_V is the variance of V, and these variances are taken from the population distribution of our random variables in the true model. Set $\sigma^2_{\epsilon_B}$ to be the "regression residual variance," $\sigma^2_{\epsilon_B} = \sigma^2_B - \lambda^2_B\sigma^2_{T_1}$. Set all coefficients and variances in the parameterization in an analogous way. Set the off-diagonal entry $\sigma_{\epsilon_C\epsilon_Z}$ to $\sigma_{CZ} - \lambda_C\lambda_Z\sigma_{T_1T_2}$.

Because the variances of the latents in this assignment correspond to the true variances, and by construction of the given parameterization, the marginal variance for B will correspond to its true variance. The same will hold for all observed variables in this model. Moreover, because T_1 d-separates A and X, in the true model we have that $\sigma_{AX} = \sigma_{AT_1}\sigma_{XT_2}\sigma_{T_1T_2}/(\sigma^2_{T_1}\sigma^2_{T_2})$. According to our paramerization, the same covariance is given by $\lambda_A\lambda_X\sigma_{T_1T_2} = \sigma_{AT_1}\sigma_{XT_2}\sigma_{T_1T_2}/(\sigma^2_{T_1}\sigma^2_{T_2})$, which corresponds to the true covariance. The argument is analogous for all other pairs, with the exception of (C, Z). In this case, however, we have the freedom to set $\sigma_{\epsilon_C\epsilon_Z}$ to obtain the actual population covariance of σ_{CZ}.

This solution is unique because the model is identifiable: it is well-known in the literature of structural equation models (Bollen, 1989) that the one-factor model $\{T_1 \rightarrow A,\ T_1 \rightarrow B,\ T_1 \rightarrow C\}$ has an unique choice of parameter values corresponding to the marginal covariance of $\{A, B, C\}$ (after setting $\lambda_A = 1$). The same is true for the T_2 factor. There is an unique choice for the covariance of T_1 and T_2 from any pair in $(A, B) \times (X, Y)$. Because all parameters but $\sigma_{\epsilon_C\epsilon_Z}$ are now fixed, this will imply a single possible value for $\sigma_{\epsilon_C\epsilon_Z}$ that will correspond to matrix $\Sigma_{\mathbf{S}}$. $\square$.

Lemma 3 *If*

$$
\begin{gathered}
T(A,B,C,K) = true,\ for\ K \in \{X,Y,V,Z\} \\
T(K,X,Y,V) = true,\ for\ K \in \{A,B,C\} \\
T(K_1,K_2,K_3,K_4) = false,\ for \\
\{K_1,K_2\} \subset \{A,B,C\}, \{K_3,K_4\} \subset \{X,Y,V\} \\
T(X,Y,Z,V) = true \\
T(A,X,Y,Z) = T(A,B,C,X) = false
\end{gathered}
$$

and all entries in the covariance matrix of $\{A, B, C, X, Y, Z, V\}$ *are non-zero, then*

- *there is a latent* T_1 *in the true model that d-separates* $\{A, B, C\}$*;*
- *there is a latent* T_2 *in the true model that d-separates* $\{X, Y, Z, V\}$*;*
- $T_1 \neq T_2$*;*
- T_1 *d-separates every pair in* $\{A, B, C, Z, T_1\} \times \{X, Y, V, T_2\}$*, except* (Z, T_2)*;*
- T_2 *d-separates every pair in* $\{A, B, C, T_1\} \times \{X, Y, Z, V, T_2\}$*, except* (Z, T_1)*;*
- *at most one element of* $\{A, B, C, Z\}$ *shares a hidden common cause with* T_1*, at most one element of* $\{X, Y, Z, V\}$ *shares a hidden common cause with* T_2*, and at most one element of* $\{A, B, C, X, Y, V\}$ *shares a hidden common cause with its respective latent ancestor* $\{T_1, T_2\}$*;*

Proof: Most of the results in the theorem follow directly from Theorem 15 of (Silva et al., 2006). We are left to show that T_1 does not d-separate Z from T_2, and that T_2 does not d-separate Z from T_1. We will prove that T_2 cannot d-separate Z from T_1. By symmetry, this proves the other result.

Suppose, for the sake of contradiction, that T_2 d-separates Z from T_1. Since T_1 lies on all treks connecting Z and A (a consequence of the Tetrad Representation Theorem and hypothesis $T(A, B, C, K = Z)$), it will follow that T_2 lies on all treks connecting Z and A.

Suppose first there are treks connecting A to T_2 and Z to T_2 that are both into T_2, as illustrated in Figure 6(a). If T_1 is a descendant of T_2, then A and Z will be d-connected given T_1, contrary to our hypothesis. If T_1 is not a descendant of T_2, there will a trek connecting T_1 and T_2 that is into T_2 (Figure 6(a)). But then T_2 does not d-separate Z from T_1.

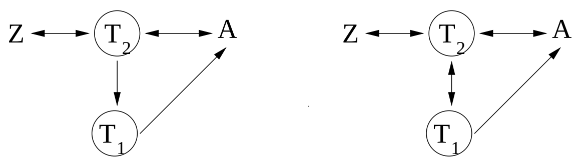

Figure 6: Treks connecting Z and A through T_1 and T_2.

Therefore, there is no trek connecting A to T_2 and Z to T_2 that are both in T_2. This fact, and the fact that $T(A, X, Y, V) = T(X, Y, Z, V) = true$, implies that at most one element of $\{A, X, Y, Z\}$ is connected to T_2 by a trek into T_2. The hypothesis and this result imply that T_2 d-separates all elements in $\{A, X, Y, Z\}$, which implies $T(A, X, Y, Z) = true$. Contradiction. $\square$

Lemma 4 *Let* $\mathbf{S} = \{A, B, C, X, Y, Z, V\}$ *be a set of random variables with covariance matrix* $\Sigma_{\mathbf{S}}$ *generated by an unknown linear latent variable model. If* $\mathbf{S}$ *satisfies the conditions of Lemma 3, then there is an unique Gaussian latent variable DAG model (up to the scale and sign of the latents) defined by* $G_{2L'}(A, B, C, X, Y, Z, V)$ *such that the marginal covariance for* $\mathbf{S}$ *according to this model equals* $\Sigma_{\mathbf{S}}$*. In this case, the covariance of* T_1 *and* T_2 *defined by this* *model equals the covariance of the respective latents in the true model containing* **S**.

Proof: This proof is similar to the proof of Lemma 2. The main different is that Z has two linear coefficients associated with it:

$$Z = \lambda_{Z1}T_1 + \lambda_{Z2}T_2 + \epsilon_Z \quad (2)$$

The definition of $\{\lambda_{Z1}, \lambda_{Z2}, \sigma^2_{\epsilon_Z}\}$ is analogous: they are the parameters obtained by the linear regression of Z on $\{T_1, T_2\}$. Because of the local separations (T_1 d-separating all elements of $\{A, B, C, Z\}$ and T_2 d-separating all elements in $\{X, Y, Z\}$), the covariance Σ_ϵ is allowed to be diagonal.

The uniqueness of this solution can also be obtained through standard results of structural equations models: parameters for $\{T_1 \rightarrow A$, $T_1 \rightarrow B$, $T_1 \rightarrow C\}$ are readily identifiable. The covariance of the latents is readily obtained from $\{T_1 \rightarrow A$, $T_1 \rightarrow B$, $T_1 \rightarrow C$, $T_2 \rightarrow X$, $T_2 \rightarrow Y$, $T_2 \rightarrow V$, $T_1 \leftrightarrow T_2\}$. Finally, parameters for the local equation (2) follow from the previously identified parameters. $\square$