Identification of Linear Latent Variable Model with Arbitrary Distribution

Zhengming Chen[1*], **Feng Xie**[2*], **Jie Qiao**[1,4*], **Zhifeng Hao**[1,5], **Kun Zhang**[3], **Ruichu Cai**[1,4†]

[1]School of Computer Science, Guangdong University of Technology, Guangzhou, China
[2]School of Mathematical Sciences, Peking University, Beijing, China
[3]Department of Philosophy, Carnegie Mellon University, Pittsburgh, USA
[4] Peng Cheng Laboratory, Shenzhen, Guangdong, China
[5] College of Science, Shantou University, Shantou, Guangdong, China
chenzhengming1103@gmail.com, xiefeng009@gmail.com, qiaojie.chn@gmail.com
zfhao@gdut.edu.cn, kunz1@cmu.edu, cairuichu@gdut.edu.cn

Abstract

An important problem across multiple disciplines is to infer and understand meaningful latent variables. One strategy commonly used is to model the measured variables in terms of the latent variables under suitable assumptions on the connectivity from the latents to the measured (known as measurement model). Furthermore, it might be even more interesting to discover the causal relations among the latent variables (known as structural model). Recently, some methods have been proposed to estimate the structural model by assuming that the noise terms in the measured and latent variables are non-Gaussian. However, they are not suitable when some of the noise terms become Gaussian. To bridge this gap, we investigate the problem of identification of the structural model with arbitrary noise distributions. We provide necessary and sufficient condition under which the structural model is identifiable: it is identifiable iff for each pair of adjacent latent variables L_x, L_y, (1) at least one of L_x and L_y has non-Gaussian noise, or (2) at least one of them has a non-Gaussian ancestor and is not d-separated from the non-Gaussian component of this ancestor by the common causes of L_x and L_y. This identifiability result relaxes the non-Gaussianity requirements to only a (hopefully small) subset of variables, and accordingly elegantly extends the application scope of the structural model. Based on the above identifiability result, we further propose a practical algorithm to learn the structural model. We verify the correctness of the identifiability result and the effectiveness of the proposed method through empirical studies.

Introduction

Discovering causal relations among latent variables is important in many domains, such as social science, climate science, and psychology. For example, to design a proper psychotherapy program, we need to understand the causal relations among role conflict and depersonalization, emotional exhaustion, and personal accomplishment (Byrne 2016). However, these variables of interest are usually unobserved, and we can only make use of measured variables generated by them, e.g., the response level from questionnaires for role conflict factor.

To discover the causal relationship among the latent variables, the linear latent variable model is introduced. In the linear latent variable model, researchers often use a two-phase framework to address this issue (Spirtes, Glymour, and Scheines 2000). It first finds the pure measurement model of the causal relations between the measured variables and their corresponding latent variables, and then infers the causal relationships between latent variables (structural model) by analyzing the measured variables. In the literature of measurement model, Silva et al. (2006) have shown that the pure measurement model can be fully identifiable under the purity assumption (each latent have at least three pure measured variables), and proposed a BPC algorithm to estimate it. Later, Kummerfeld et al. (2014) proposed a more efficient method, called FOFC, to estimate the pure measurement model. In the field of the structural model, there exist work such as (Silva et al. 2006), which proposed a MIMBuild algorithm to estimate the causal structure of latent variables given a pure measurement model. However, it can only output structures up to the Markov equivalence class for latent variables. Shimizu, Hoyer, and Hyvärinen (2009) showed that the causal relationships among latent factors are identifiable when the data are non-Gaussian. Following the non-Gaussian assumption, Cai et al. (2019) recently designed the so-called Triad constraints and proposed a more efficient method to infer the latent structure, and Xie et al. (2020) further proposed a generalized independent noise condition to address the case where there are multiple latent variables behind measured variables. Recently, Zeng et al. (2021) proposed the MD-LiNA to estimate the underlying causal structure among latent factors for multi-domain data.

In many real-world scenarios, however, the full non-Gaussianity assumption (all noise terms are non-Gaussian) may be violated. For example, none of the existing methods is able to uniquely identify the causal direction between

*These authors contributed equally.
†Corresponding author.

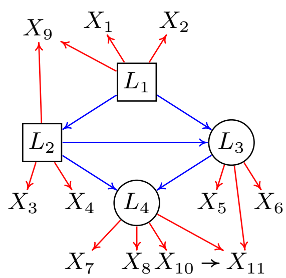


Figure 1: An example of linear latent variable model involving 4 latent variables and 11 observed variables. Here, the red edges form a measurement model, while the blue edges form a structural model. Moreover, the rectangles represent node with non-Gaussian noise, while the circles represent the nodes with Gaussian noise, i.e., $\varepsilon_{L_1}, \varepsilon_{L_2}$ are non-Gaussian and, $\varepsilon_{L_3}, \varepsilon_{L_4}$ are Gaussian.

the latent variables L_3, L_4 with Gaussian noise in Figure 1. However, we show that it is still possible to identify the causal direction between L_3, L_4 in this causal structure, as we will discuss in Section 4. Intuitively, although the noises of L_3, L_4 are Gaussian, there still exist some non-Gaussian contribution from L_1 and L_2, which might be helpful to further identify the causal direction among the variables with Gaussian noise. Moreover, it is still unclear whether the causal relationship is uniquely identifiable between the variables with partial non-Gaussian noise, i.e., only a subset of them have non-Gaussian noise. Therefore, to recover the causal relations among latent variables, it is crucial to understand when and where the causal relationship is identifiable, in order to estimate the structural model.

In this paper, we will discuss the problem of identification of structural model with arbitrary distribution. First, we prove that the causal direction between two latent variables is fully identifiable if at least one of them has a non-Gaussian noise term. Furthermore, inspired by the intuition about the transitivity of non-Gaussian noise, interestingly, the causal direction is actually identifiable if at least one of the two latent variables cannot be d-separated from a non-Gaussian ancestor of it by the common causes of the two latent variables. More details will be presented in section 4. Based on the proposed theoretical results, we proposed an statistically efficient algorithm for learning Linear Latent Causal Structure with Arbitrary Distribution (LLCS-AD). The contributions of this work are as follows. 1) We provide necessary and sufficient identification conditions of the structural model with the arbitrary distribution. 2) We develop a practical algorithm for learning the causal structure of latent variables with the arbitrary distribution. 3) We demonstrate that our algorithm, compared to existing ones, works clearly better in high dimensions.

Related Works

Most causal discovery approaches focus on the situation without latent variables (Spirtes, Glymour, and Scheines 2000; Pearl 2009; Peters, Janzing, and Schölkopf 2017). Roughly speaking, they can be divided into the following three categories: constraint-based methods, such as PC (Spirtes, Glymour, and Scheines 2000), IC (Pearl 2009) and their variants; score-based methods, such as GES (Chickering 2002); functional-based methods, such as, LiNGAM (Shimizu et al. 2006), ANM (Hoyer et al. 2009; Cai et al. 2018), PNL (Zhang and Hyvärinen 2009), and IGCI (Janzing et al. 2012). Moreover, some approaches investigate causal discovery under the mixed distribution among observed variables (Hoyer et al. 2008a). However, although these methods have been widely used in many fields, they may fail to identify the correct causal structure in cases with latent confounders. The reason is that they do not properly take into account the latent variables in the procedure, which could cause many practical issues (Zhang et al. 2018). Some methods generalize the traditional constraint-based methods to allow the existence of latent variables, including the FCI algorithm (Spirtes, Meek, and Richardson 1995; Colombo et al. 2012). By further introduce linear non-Gaussian assumption, causal structure can be identified in present of latent confounder by using overcomplete ICA (Hoyer et al. 2008b).

Problem Definition

In this paper, we focus on linear acyclic latent variable causal models. Here, we use $\mathbf{V} = \mathbf{X} \cup \mathbf{L}$ to denote the total set of variables, where $\mathbf{X} = \{X_1, X_2, ...X_m\}$ denotes the set of observed variables, and $\mathbf{L} = \{L_1, L_2, ...L_n\}$ denotes the set of latent variables. By incorporating a causal Directed Acyclic Graph (DAG), we assume that all variables $\mathbf{V}$ satisfy the following generating process: $V_i = \sum_{k(j)<k(i)} b_{ij} V_j + \varepsilon_{V_i}, i = \{1, 2, .., n + m\}$, where $k(i)$ denotes the k-th index in an arranged causal order in graph $\mathbf{G}$, such that no later variable causes any earlier variables, b_{ij} represents the causal strength from V_j to V_i, and ε_{V_i} is the independent and identically distributed noise variable such that $p(\varepsilon_{V_1}, ..., \varepsilon_{V_{n+m}}) = \prod_i p_i(\varepsilon_{V_i})$. We use $Pa(V_i) = \{V_j | V_j \rightarrow V_i\}$, $Ch(V_i) = \{V_j | V_i \rightarrow V_j\}$, $Anc(V_i) = \{V_j | V_j \rightsquigarrow V_i\}$, $Des(V_i) = \{V_j | V_i \rightsquigarrow V_j\}$, $Adj(V_i) = \{V_j | V_j - V_i\}$ to denote the set of parents, children, ancestors, descendants, and adjacent nodes of V_i, respectively.

Without loss of generality, we assume that all variables have a zero mean. Here we provide the definition of the Linear Latent Variable model as follows:

Definition 1 (Linear Latent Variable Model (LLVM) (Shimizu et al. 2006; Spirtes, Glymour, and Scheines 2000))**.** *A model satisfying the following assumptions is a linear latent variable measurement model. For brevity, we called it LLVM.*

A1. [Causal Markov property] Any variable is independent of its non-descendants in graph G conditional on any values of its parents in G.

A2. [Causal faithfulness property] There are only implied the condition independent constraints on the causal graph G.

A3. [Linear acyclic additive noise assumption] Each variable in $\mathbf{V}$ *is a linear function of its parents plus*

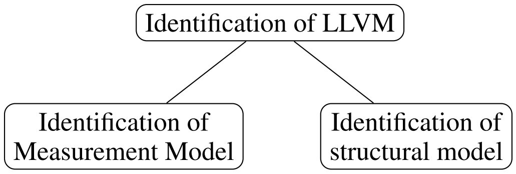


Figure 2: Identification of the Linear Latent Variable Model

an additive error term of positive finite variance, and the complete causal graph is acyclic.

One example of LLVM is given in Figure 1. To identify such an LLVM model, we categorize the problem into two folds as illustrated in Figure 2.

For the identification of the measurement model, recent results have shown that it can be uniquely identified if the measurement assumption and the 3-purity assumption hold.Then, given the learned measurement model, one is able to learn the causal structure among the latent variables, i.e., learning the structural model. In this work, we mainly focus on the identification of structural model given the known measurement model with arbitrary distribution. Moreover, given the measurement model, we only require the measurement assumption as well as a weaker 2-purity assumption, which are listed below:

A4. [Measurement Assumption] There is no observed variable being an ancestor of any latent variable.

A5. [2-Purity Assumption] Each latent variable L has at least 2 pure measured variables as children.

Note that purity assumption might be justified by some existing methods, e.g., an empty causal structure would occur by BPC algorithm if the purity assumption does not hold (Silva et al. 2006).

Therefore, given the measurement, we aim to develop the identification of structural model with A1-A5 holds.

Identification of Structural Model

In this section, we first briefly review the previous work on the identification of structural model. Then we will show the unsolved problem with arbitrary distributions, which is underdeveloped. Finally, we present the necessary and sufficient conditions that render the structural model identifiable.

Traditionally, Spirtes, Glymour, and Scheines (2000) and Silva et al. (2006) have proved that the causal structure among latent variables can be identified up to Markov equivalence class. It has been shown that by further assuming that each noise follows the non-Gaussian distribution, the structural model can be fully identifiable using the GIN condition (Xie et al. 2020). However, it is still unclear how to deal with the arbitrary distribution that allows the existence of both non-Gaussian and Gaussian noise. Thus, to complete the identification of the structural model, we will discuss the remaining two cases in this work, as given below.

- The causal direction between L_x and L_y in which only one latent variable has non-Gaussian noise.
- The causal direction between L_x and L_y in which two of the latent variables both have Gaussian noise.

Please note, that in these two cases, we mainly focus on the non-Gaussianity of the latent variables, because the measured variables of the non-Gaussian latent variables are always non-Gaussian, according to the Crámer decomposition theorem.

In the following, we assume that the Markov equivalence class G of structural model has already been identified.

Before we give the fundamental identifiability theorem about our work, we first need to introduce a Generalized Independent Noise (GIN) mechanism in LLVM that is able to capture the high order statistics in non-Gaussian data:

Definition 2 (GIN condition (Xie et al. 2020))**.** *Let $\mathbf{Y}$ and $\mathbf{Z}$ be two observed random vectors. Suppose the variables follow the linear non-Gaussian acyclic causal model. Define the surrogate-variable of $\mathbf{Y}$ relative to $\mathbf{Z}$, as*

$$E_{\mathbf{Y}||\mathbf{Z}} \coloneqq \omega^{\intercal}\mathbf{Y}, \tag{1}$$

where ω satisfies $\omega^{\intercal}\mathbb{E}[\mathbf{Y}\mathbf{Z}^{\intercal}] = 0$ and $\omega \neq 0$. We say that $(\mathbf{Z}, \mathbf{Y})$ follows GIN condition if and only if $E_{\mathbf{Y}||\mathbf{Z}}$ is independent from $\mathbf{Z}$.

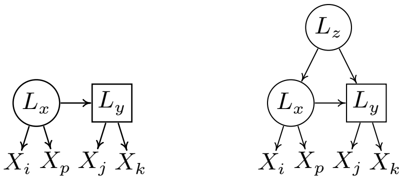
Remark 1. *Let $\mathbf{Y}$ and $\mathbf{Z}$ be two observed random vectors. Suppose the variables follow **Gaussian** distributions. Then $E_{\mathbf{Y}||\mathbf{Z}}$ is always statistically independent from $\mathbf{Z}$, i.e., $(\mathbf{Z}, \mathbf{Y})$ always follows GIN condition.*

Interestingly, GIN condition allows us to further identify the causal direction between latent variables in the equivalence class, if there are two variables whose noises are non-Gaussian. We provide the following example to show the benefit that the GIN condition brings.

Example 1. *As shown in Figure 1. The noise of L_1 and L_2 are non-Gaussian. Considering the causal direction $L_1 \to L_2$, we can construct the test $E_{(X_1,X_3)||X_2}$ using X_2 as the surrogate variable of L_1. Then we have $E_{(X_1,X_3)||X_2} \perp\!\!\!\perp X_2$, i.e., the GIN condition hold in the causal direction. Similarly, in the reverse direction, we can construct the test $E_{(X_1,X_3)||X_4}$ by placing X_2 with X_4, but we will have $E_{(X_1,X_3)||X_4} \not\perp\!\!\!\perp X_4$, i.e., the GIN condition does not hold in the reverse direction. Based on such asymmetry of independence, the causal direction can be identified. More details will be provided in appendix regarding how such independent property holds.*

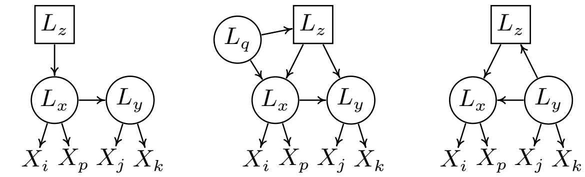
Given the GIN condition, we are able to answer the two questions above. In answering the first question – identification of causal relationship in which only one latent variable has non-Gaussian noise, we decompose the question into two typical structures as shown in Figure 3. We will prove that the direction between any two adjacent latent variables L_x, L_y in a given Markov equivalent class is identifiable if at least one of them has non-Gaussian noise. To do so, we will first provide identification of Figure 3 (a) and (b) in Lemma 1 and Lemma 2, respectively. Then we obtain the final result in Theorem 1. All proofs are given in the Supplementary Material.

Lemma 1. *For each pair of adjacent latent variables L_x, L_y in the Markov equivalence class G, if (1) there is no confounder and (2) at least one of noises ε_{L_x}, ε_{L_y} is non-Gaussian, then the causal direction between L_x and L_y is identifiable.*



(a) Without confounder (b) With confounder

Figure 3: Three types of structure such that only one latent variable has non-Gaussian noise (the rectangle one).



(a) Common ancestor that has non-Gaussian noise. (b) Common parent that has non-Gaussian noise. (c) Not common ancestor.

Figure 4: Three types of non-Gaussian transitivity structures. Rectangles indicate the nodes with non-Gaussian noise, while the circles represent the nodes with Gaussian noise.

Lemma 1 prove that for each pair of adjacent latent variables that has no confounder is identifiable if there are at least one of them has non-Gaussian noise as illustrated in Figure 3(a). The proof is based on the GIN condition. Note that the result still holds in the reverse direction. Next, we will discuss the identification in the case that has confounders.

Lemma 2. *For each pair of adjacent latent variables L_x, L_y in the Markov equivalence class G with a confounder set that influence L_x and L_y, the causal direction between L_x and L_y is identifiable if at least one of ε_{L_x} and ε_{L_y} is non-Gaussian.*

Lemma 2 shows that even there exist confounders between the two directly connected nodes as illustrated in Figure 3(b), the causal direction is still identifiable. The key difference from the Lemma 1 is that we need to consider the confounders as a conditional set, which means that we need to regress L_x, L_y against the confounders and then apply Lemma 1 on the residual–we then obtain Lemma 2. The remaining question is whether the confounder set can be found as a condition and whether it is possible to learn an incorrect causal direction. In the following theorem, we will show that it can always search a proper conditional set from $Adj(L_x) \cup Adj(L_y)$ such that the asymmetric of GIN condition holds. Furthermore, we also find that if the cause variable is non-Gaussian, we can not mistakenly identify the causal direction as given in Proposition 1.

Proposition 1. *For each pair of adjacent latent variables $L_x \to L_y$ in the Markov equivalence class G, the reverse direction will not be mistakenly identified as the causal direction if ε_{L_x} is non-Gaussian.*

In other words, we can always get rid of childen from non-Gaussian node when searching the conditional set, which will avoid some extreme cases (see more details in appendix). Thus, by combining Lemma 1 and 2, we have Theorem 1.

Theorem 1. *For each pair of adjacent latent variables L_x, L_y in the Markov equivalence class G, if at least one of the noise ε_{L_x} and ε_{L_y} is non-Gaussian, then the causal direction between L_x and L_y is identifiable.*

To further illustrate Theorem 1, we provide an example below.

Example 2. *Consider the causal relationship $L_2 \to L_3$ in Figure 1. Because there exists the confounder L_1, we will first need to search the conditional set to eliminate the influence of confounder. To do so, we first construct the adjacent set $Adj(L_2) \cup Adj(L_3) = \{L_1, L_4\}$. Then we can find a conditional set $S = \{L_1\}$ such that $E_{(X_7,X_3,X_1)||(X_4,X_2)} \perp\!\!\!\perp (X_4, X_2)$, $E_{(X_7,X_3,X_1)||(X_8,X_2)} \not\perp\!\!\!\perp (X_8, X_2)$ (Lemma 2) and $E_{(X_3,X_1)||X_2} \perp\!\!\!\perp X_2$. Note that L_4 will not be considered in the conditional set because $E_{(X_3,X_1)||X_7} \not\perp\!\!\!\perp X_7$ (proposition 1). Then, based on the above asymmetry, we conclude $L_2 \to L_3$.*

Theorem 1 has shown that if there are at least one of the latent variables has non-Gaussian noise, the edges around such a variable will be identifiable. However, we may ask whether two latent variables both of which have Gaussian noise enjoy a similar property. Thanks to the transitivity of non-Gaussianity in linear causal relations, we can use the information from the non-Gaussian ancestor to identify the causal direction.

Roughly speaking, the reason why we can identify the causal direction is that a non-Gaussian component transmits to its descendent. Thus, we categorize it into two cases: (1) common non-Gaussian ancestor, or (2) not common non-Gaussian ancestor as shown in Figure 4. Take Figure 4(a) as an example, in this common non-Gaussian ancestor structure, the non-Gaussainity from L_z can be absorbed by L_x such that the noise term of L_x can be rewritten as $\hat{\varepsilon}_{L_x} = \varepsilon_{L_x} + \beta\varepsilon_{L_z}$. Then based on Lemma 1, such causal direction is identifiable. We conclude the above analysis in Lemma 3.

Lemma 3. *For each pair of adjacent latent variables L_x, L_y in the Markov equivalence class G, the causal direction between L_x and L_y is identifiable, if (1) L_x, L_y have one common non-Gaussian ancestor $L_z \in \{Anc(L_x) \cap Anc(L_y)\}$, (2) and there are no confounder between L_x, L_y.*

Note that such result is similar to Lemma 1 as we can reduce to the previous cases thank to the transitivity. Next, in Lemma 4 we will further show that the causal direction is still identifiable if there exists confounder.

Lemma 4. *For each pair of adjacent latent variables L_x, L_y in the Markov equivalence class G with confounder set S that influence L_x and L_y, the causal direction between L_x and L_y is identifiable, if L_x, L_y has one common non-Gaussian ancestor $L_z \in \{Anc(L_x) \cap Anc(L_y)\}$, and L_z is not conditional independent of L_x, L_y given its confounder* set $S = \{Pa(L_x) \cap Pa(L_y)\}$, i.e., $L_z \not\perp\!\!\!\perp (L_x, L_y)|S$.*

The result in Lemma 4 is also similar to Lemma 2. That is, we will also need to find a proper conditional set from $Adj(L_x) \cup Adj(L_y)$ in order to identify the causal direction.

Next, in Lemma 5, we will discuss the cases that has common parent that has non-Gaussian noise, which is slightly different compared with the previous results as we require an additional confounder L_q on between L_x and L_z as shown in Figure 4(b).

Lemma 5. *For each pair of causal relationship $L_x \to L_y$ in the Markov equivalence class G, the causal direction $L_x \to L_y$ is identifiable, if (1) there exist one latent variable L_z with non-Gaussian noise such that $L_z \in Pa(L_x) \cap Pa(L_y)$ and there exist a confounder L_q of L_x, L_y such that $L_q \in Pa(L_x) \cap Pa(L_z)$.*

In summary, we have investigated the cases that have common non-Gaussian ancestor. In the following Lemma 6, we will further discuss the cases that only one variables is affected by variables that have non-Gaussian noise.

Lemma 6. *For each pair of adjacent latent variables L_x, L_y in the Markov equivalence class G, the causal direction between L_x and L_y is identifiable, if there exists one latent variable L_z with non-Gaussian noise such that (1) $L_z \in Anc(L_x)$, $L_z \notin Anc(L_y)$, or (2) $L_z \in Anc(L_y)$, $L_z \notin Anc(L_x)$.*

An example of Lemma 6 is illustrated in Figure 4(c), in which $L_z \in Anc(L_x)$ and $L_z \notin Anc(L_y)$, i.e., only L_z is affected by L_z. Then based on Lemma 6, the causal direction between L_x and L_y is identifiable.

To conclude, we provide a general condition for the identification of transitivity of non-Gaussian noise in the following theorem.

Theorem 2 (Transitivity of non-Gaussian noise)**.** *For each pair of adjacent latent variables L_x, L_y in the Markov equivalence class G, the causal direction between L_x and L_y is identifiable, if there exists a latent variable L_z that has non-Gaussian component ε such that (1) $L_z \in Anc(L_x) \cup Anc(L_y)$, and (2) ε is not conditional independent from $\{L_x, L_y\}$ given the confounder set $S \coloneqq Pa(L_x) \cap Pa(L_y)$, i.e., $\varepsilon \not\perp\!\!\!\perp (L_x, L_y)|S$.*

Intuitively, Theorem 2 shows that if we ensure the "transitivity" holds, i.e., $\varepsilon \not\perp\!\!\!\perp (L_x, L_y)|S$, where S is the confounder set of L_x, L_y, then the causal direction between L_x, L_y is identifiable. Moreover, we only require that there exists a latent variable L_z that has non-Gaussian component which means that ε_{L_z} could be Gaussian if it has a non-Gaussian ancestor. We give an example to illustrate Theorem 2 using the graph in Figure 1.

Example 3. *For $L_3 \to L_4$, we found that the common ancestor set of L_3, L_4, contains $\{L_1, L_2\}$ while the confounder set contains $\{L_2\}$. Because $\varepsilon_{L_1} \not\perp\!\!\!\perp (L_x, L_y)|L_2$, based on Theorem 2, we have the causal direction $L_3 \to L_4$ is identifiable. Specifically, we can search a subset from the common adjacent set $S = Adj(L_3) \cap Adj(L_4) = \{L_2\}$ such that in causal direction, we have $E_{(X_7,X_3,X_5)||(X_4,X_6)} \perp\!\!\!\perp (X_4, X_6)$ while in the reverse direction, we have $E_{(X_5,X_7,X_3)||(X_4,X_8)} \not\perp\!\!\!\perp (X_4, X_8)$ because of $E_{(X_5,X_7,X_3)||(X_4,X_8)} \not\perp\!\!\!\perp \varepsilon_{L_1}$. According to this asymmetry, the causal direction between L_3 and L_4 is identifiable.*

To sum up the above theoretical results, we develop necessary and sufficient condition for the identifiability of the structural model in the following theorem.

Theorem 3 (Identification of Structural Model)**.** *Suppose that assumptions A1–A5 hold, for each pair of adjacent latent variables L_x, L_y in the Markov equivalence class G, the causal direction between L_x and L_y is identifiable if and only if (1) at least one of latent variables L_x, L_y has non-Gaussian noise or (2) there exists a latent variable $L_z \in \{Anc(L_x) \cup Anc(L_y)\}$ that has non-Gaussian component ε such that ε is not conditional independent from $\{L_x, L_y\}$ given the confounder set $S = \{Pa(L_x) \cap Pa(L_y)\}$, i.e., $\varepsilon \not\perp\!\!\!\perp (L_x, L_y)|S$.*

An Algorithm for Learning Casual Structural with Arbitrary Distribution

In this section, we extend the above results to estimate the causal structure of latent variables. To this end, we propose a fusion algorithm to learn **L**inear **L**atent **C**ausal **S**tructure with **A**rbitrary **D**istributions (LLCS-AD). For notational convenience, we use notation $GIN(L_x, L_y)$ to show that $(\{X_2\}, \{X_1, Y_1\})$ satisfy GIN condition, i.e., $E_{(X_1,Y_1)||(X_2)} \perp\!\!\!\perp X_2$, where $\{X_1, X_2\}$ and Y_1 are the children of L_x and L_y, respectively. Furthermore, we use notation $GIN(L_x, L_y|L_z)$ to show that $(\{X_2, Z_2\}, \{X_1, Y_1, Z_1\})$ satisfy GIN condition, i.e., $E_{(X_1,Y_1,Z_1)||(X_2,Z_2)} \perp\!\!\!\perp \{X_2, Z_2\}$, where Z_1, Z_2 are the children of L_z.

Our method is outlined in Algorithm 1. We first learn the measurement model using the existing method, such as BPC (Line 1). Then, we construct the causal skeleton by employing PC-MIMBuild (Silva et al. 2006) (Line 2). The above procedures output a pattern (or Markov equivalence class) of the latent variable, namely the skeleton of latent variable. Next, we orient the undirected edges according to the Lemma 1 (Lines 3-9). Note that we can orient the non-Gaussian node and transitive non-Gaussian node if there do not exist any confounder.

In Line 10-18, we first enumerate all equivalence classes of partial DAG output by step 8. Then, we verify the GIN conditions according to Theorem 1 for each edge in every equivalence graph and reject the graph which does not satisfy the GIN conditions. In fact, the structure of the equivalence graph would be rejected if there exists non-Gaussian noise. Consequently, we could merge the equivalence class that can not be rejected. That is, for each equivalence G_i, if the direction of one edge in each equivalence class is consistent, we accept the direction. Otherwise, we reject the direction of the edge. In Line 19, we further orient the undirected edges according to Theorem 2 using the function $TransOrient(G, X)$. Lastly, in Line 20, we further orient the undirected edges by Meek rules if such edges exist.

The details of the function $TransOrient(G, X)$ are provided in Algorithm 2. As shown in algorithm, firstly, Lines 2-7 record all the unoriented edges with a non-Gaussian ancestor by testing the GIN condition violating; Secondly, in

Algorithm 1: LLCS-AD

Require: Data set $\mathbf{X} = \{X_1, \ldots, X_m\}$
Ensure: A partial DAG G for latent variable
1: $G \leftarrow$ measurement model by BPC algorithm on X;
2: $G \leftarrow$ skeleton of latent variables by PC-MIMBuild algorithm on X, G;
3: **for** each adjacent pair $L_x - L_y \in G$ **do**
4:   **if** $GIN(L_x, L_y)$ **then**
5:     orient $L_x \rightarrow L_y$ in G;
6:   **else if** $GIN(L_y, L_x)$ **then**
7:     orient $L_y \rightarrow L_x$ in G;
8:   **end if**
9: **end for**
10: $\mathbf{G} \leftarrow$ all equivalence classes of G;
11: **for** each $G_i \in \mathbf{G}$ **do**
12:   **for** each $L_x \in G_i$ **do**
13:     **if** $\neg GIN(L_x, Pa(L_x))$ **then**
14:       remove G_i from $\mathbf{G}$;
15:     **end if**
16:   **end for**
17: **end for**
18: merge the equivalence classes $\mathbf{G}$ to partial DAG G
19: $G \leftarrow TransOrient(G, X)$;
20: Orient the undirected edges by Meek rules (Meek 1995);
21: return G;

Lines 8-18, we select the candidate confounder set for each edge and test the asymmetry of the GIN condition to make the causal direction identifiable.

Theorem 4. *Suppose that assumptions A1–A5 hold. Given the large enough sample size, LLCS-AD asymptotically outputs the correct causal structure G.*

Complexity analysis: Given the learned skeleton, the worst case time complexity in Algorithm 1 (Line 3-21) is $\mathcal{O}(P + N \cdot N! + P)$, where N is the number of latent variables, and P is the number of edges of the latent variable skeleton graph. In such a case, the noises are all Gaussian, requiring to enumerate all d-separated equivalent classes, which is an extreme case. Thus, we also analyzed the case that all noises are non-Gaussian, and its worst case time complexity is $\mathcal{O}(P + \frac{N-1}{2}(N-1)!)$, which is much faster than the case that all noises are Gaussian.

Experiments

In this section, we verify the effectiveness of our proposed in both synthetic data and real-world data. In synthetic data, we will verify the theoretical results in terms of some noises are non-Gaussian in causal structure and the transitivity structure in the structural model. We further verify our method in a teacher's burnout study real-world dataset.

Synthetic Data

In the simulation studies, we conducted three different control experiments: (1) the sensitivity of sample size, (2) the ratio of non-Gaussian noise, and (3) the performance in the non-Gaussian transitivity case. And we controlled the experiments by traversing the controlled parameter while keeping other setting fixed as default. All data were generated from a random causal structure. For experiments (1) and (2), we controlled different levels of sample size, and ratio of non-Gaussian noise, ranging from $\{500, 1000, \mathbf{2000}\}$, $\{10\%, 30\%, 50\%, \mathbf{80\%}\}$, respectively. The default setting is marked as bold, and the number of latent variables was set to 8. For experiment (3), we fixed a non-Gaussian variable as the root of the causal structure, then randomly generated its Gaussian descendent, in which we controlled the number of latent variables range from $\{2, 4, 8, 12\}$. All non-Gaussian noises were following Uniform distribution $U(-2, 2)$.

All data were generated from the linear latent variable model in which each latent variable has three measured variables. The connection strength at each edge was sampled uniformly from $[-2, -0.5] \cup [0.5, 2]$. Because the data are non-Gaussian, the Hilbert-Schmidt Independence Criterion (HSIC) test (Gretton et al. 2005) was used as the independence test tool, and we set the significance level as $\alpha = 0.01$. Each experiment was repeated 10 times with randomly gen-

Algorithm 2: Orient by transitivity of non-Gaussian noise (ONG)

1: Function TransOrient(G, X)
2: $E \leftarrow \Phi$;
3: **for** each undirected edge $L_x - L_y \in G$ **do**
4:   **if** $\neg GIN(L_x, L_y)$ and $\neg GIN(L_y, L_x)$ **then**
5:     $E \leftarrow E \cup \{L_x - L_y\}$
6:   **end if**
7: **end for**
8: **for** each $L_x - L_y \in E$ **do**
9:   $\mathbf{L} \leftarrow Adj(L_x) \cap Adj(L_y)$;
10:   **for each** $\mathbf{L}' \subset \mathbf{L}$ **do**
11:     **if** $GIN(L_x, L_y|\mathbf{L}')$ and $\neg GIN(L_y, L_x|\mathbf{L}')$ **then**
12:       orient $L_x \rightarrow L_y$;
13:     **else if** $\neg GIN(L_x, L_y|\mathbf{L}')$ and $GIN(L_y, L_x|\mathbf{L}')$ **then**
14:       orient $L_y \rightarrow L_x$;
15:       break;
16:     **end if**
17:   **end for**
18: **end for**
19: **return** G

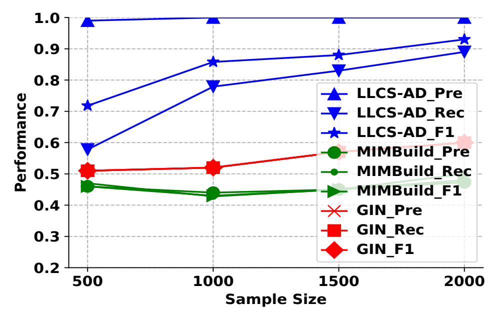


Figure 5: Sensitivity to sample size.

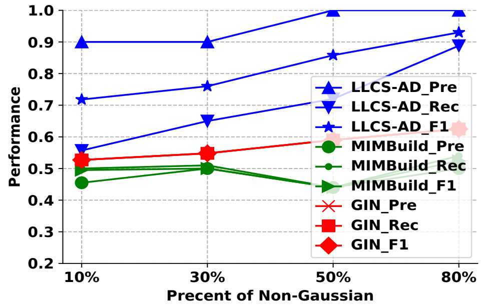


Figure 6: The ratio of non-Gaussianity.

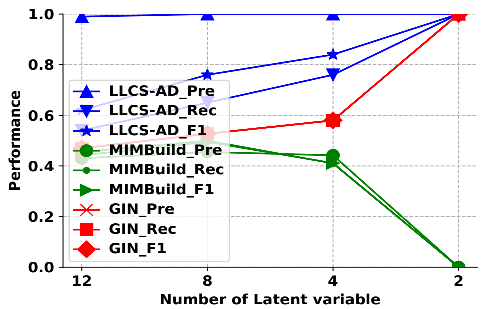


Figure 7: The performance of Transitivity.

erated data, and the results were averaged.

We compared our method with PC-MIMBuild (Silva et al. 2006) and GIN (Xie et al. 2020). Precision, Recall, and F1 score were used to evaluate the algorithms with a known ground truth containing all pure measured variable and latent variables. The sparseness of the generating DAG is $s = 2/(k-1)$, where k is the number of latent variables, such that the average indegree for each latent variable is two (Cui et al. 2018).

The simulation results are given in Figure 5, 6, and 7. Overall, our method, LLCS-AD, achieves the best performance in all cases. We can also see that GIN is better than MIMBuild. The reason is that GIN further utilize the non-Gaussianity. But GIN and MIMBuild still do not perform well as they can not deal with the arbitrary distribution. In addition, the precision of LLCS-AD is much higher than others. The reason is that LLCS-AD can tell whether an edge is identifiable or not in the arbitrary distribution while other methods can not.

Specifically, as shown in Figure 5, all methods are sensitives to the sample size, and we can see that the 1000 sample size is enough to obtain a good result. In Figure 6, as the ratio of non-Gaussianity variables increase, the performance of our method and GIN also increase. At the same time, MIMBuild is not sensitive to the ratio non-Gaussianity. The reason is that MINBuild does not consider the information from non-Gaussian. As shown in Figure 7, interestingly, GIN and LLCS-AD both correct when there has 2 latent variables. Recall that the root is non-Gaussian, so GIN can also identify the partial non-Gaussian noise cases. However, as the number of latent variables increase GIN become worse compared with our method, which show the effectiveness of LLCS-AD. Moreover, due to the randomly generated structure, there could exists more unidentifiable structure as the number of latent variables increase. Thus the performance could decrease as the number of latent variable increase, which also verify Theorem 2. However, the precision of our method is 1, which verifies the effectiveness of our method regarding the identifiability of the transitivity non-Gaussian component.

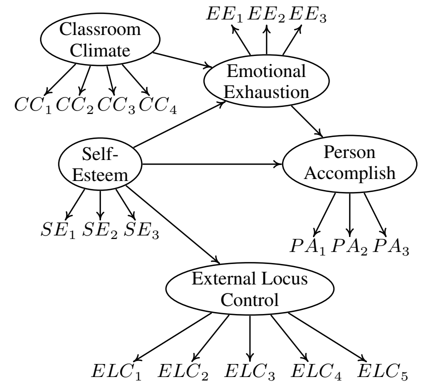


Figure 8: Result from LLCS-AD in the teacher's burnout study.

Real-World Data

In this section, we applied our method on a real-world dataset collected by Byrne (2016) that investigates the impact of organizational (self-esteem, classroom climate) and personality (self-esteem, external locus of control) on three facets of burnout in full-time elementary teachers. There are five latent variables with more than three pure measured variables for each latent variable. Figure 8, shows the output from LLCS-AD. We can see that the learned causal stricture is reasonable and is consistent with the conclusion given by Byrne (2016); Maslach, Jackson, and Leiter (1997).

Compared with the baseline methods, MIMBuild only obtain an Markov equivalence class if not using the meek rule while GIN outputted one incorrect causal order. To conclude, we have a more robustness results, which shows the effectiveness of our method.

Conclusion

We provided necessary and sufficient conditions for the identifiability of the structural model in linear latent variable model with arbitrary distribution. Based on the proposed theoretical results, we developed an algorithm for learning the linear latent variable model that allows the arbitrary distribution. Experimental results on simulation data and real-world data further verified the effectiveness of our algorithm. Future research along this line includes relaxing the purity assumptions in learning the measurement model in our algorithm and proposing more efficient learning algorithms for the linear latent variable model.

Acknowledgments

This research was supported in part by National Science Fund for Excellent Young Scholars (6212200101) and Natural Science Foundation of China (61876043, 61976052), and the China Postdoctoral Science Foundation (020M680225). Kun Zhang was supported in part by the National Institutes of Health (NIH) under Contract R01HL159805, by the NSF-Convergence Accelerator Track-D award #2134901, and by a grant from Apple Inc. The NIH or NSF is not responsible for the views reported in this article. We appreciate the comments from anonymous reviewers, which greatly helped to improve the paper.

References

Byrne, B. M. 2016. *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge Press, 3rd edition.

Cai, R.; Qiao, J.; Zhang, Z.; and Hao, Z. 2018. Self: structural equational likelihood framework for causal discovery. In *Thirty-second AAAI conference on artificial intelligence*.

Cai, R.; Xie, F.; Glymour, C.; Hao, Z.; and Zhang, K. 2019. Triad Constraints for Learning Causal Structure of Latent Variables. In *NeurIPS*, 12863–12872.

Chickering, D. M. 2002. Optimal structure identification with greedy search. *JMLR*, 3(Nov): 507–554.

Colombo, D.; Maathuis, M. H.; Kalisch, M.; and Richardson, T. S. 2012. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, 294–321.

Cui, R.; Groot, P.; Schauer, M.; and Heskes, T. 2018. Learning the Causal Structure of Copula Models with Latent Variables. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018*, 188–197. AUAI Press.

Gretton, A.; Herbrich, R.; Smola, A. J.; Bousquet, O.; and Schölkopf, B. 2005. Kernel Methods for Measuring Independence. *Journal of Machine Learning Research*, 6: 2075–2129.

Hoyer, P. O.; Hyvärinen, A.; Scheines, R.; Spirtes, P.; Ramsey, J.; Lacerda, G.; and Shimizu, S. 2008a. Causal discovery of linear acyclic models with arbitrary distributions. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, 282–289.

Hoyer, P. O.; Janzing, D.; Mooij, J. M.; Peters, J.; and Schölkopf, B. 2009. Nonlinear causal discovery with additive noise models. In *NeurIPS*, 689–696.

Hoyer, P. O.; Shimizu, S.; Kerminen, A. J.; and Palviainen, M. 2008b. Estimation of causal effects using linear non-Gaussian causal models with hidden variables. *IJAR*, 49(2): 362–378.

Janzing, D.; Mooij, J.; Zhang, K.; Lemeire, J.; Zscheischler, J.; Daniušis, P.; Steudel, B.; and Schölkopf, B. 2012. Information-geometric approach to inferring causal directions. *Artificial Intelligence*, 182: 1–31.

Kummerfeld, E.; Ramsey, J.; Yang, R.; Spirtes, P.; and Scheines, R. 2014. Causal clustering for 2-factor measurement models. In *Proceedings of the 2014th European Conference on Machine Learning and Knowledge Discovery in Databases-Volume Part II*, 34–49.

Maslach, C.; Jackson, S. E.; and Leiter, M. P. 1997. *Maslach burnout inventory*. Scarecrow Education Press.

Meek, C. 1995. Causal inference and causal explanation with background knowledge. In *UAI*, 403–410.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press, 2nd edition.

Peters, J.; Janzing, D.; and Schölkopf, B. 2017. *Elements of Causal Inference*. MIT Press.

Shimizu, S.; Hoyer, P. O.; and Hyvärinen, A. 2009. Estimation of linear non-Gaussian acyclic models for latent factors. *Neurocomputing*, 72(7-9): 2024–2027.

Shimizu, S.; Hoyer, P. O.; Hyvärinen, A.; and Kerminen, A. 2006. A linear non-Gaussian acyclic model for causal discovery. *JMLR*, 7(Oct): 2003–2030.

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

Spirtes, P.; Glymour, C.; and Scheines, R. 2000. *Causation, Prediction, and Search*. MIT press.

Spirtes, P.; Meek, C.; and Richardson, T. 1995. Causal inference in the presence of latent variables and selection bias. In *UAI*, 499–506. Morgan Kaufmann Publishers Inc.

Xie, F.; Cai, R.; Huang, B.; Glymour, C.; Hao, Z.; and Zhang, K. 2020. Generalized Independent Noise Conditionfor Estimating Latent Variable Causal Graphs. In *NeurIPS*, 14891–14902.

Zeng, Y.; Shimizu, S.; Cai, R.; Xie, F.; Yamamoto, M.; and Hao, Z. 2021. Causal discovery with multi-domain LiNGAM for latent factors. In *IJCAI*, 2097–2103.

Zhang, K.; and Hyvärinen, A. 2009. On the identifiability of the post-nonlinear causal model. In *UAI*, 647–655. AUAI Press.

Zhang, K.; Schölkopf, B.; Spirtes, P.; and Glymour, C. 2018. Learning causality and causality-related learning: some recent progress. *National science review*, 5(1): 26–29.