APPENDIX FOR "STRUCTURAL ESTIMATION OF PARTIALLY OBSERVED LINEAR NON-GAUSSIAN ACYCLIC MODEL: A PRACTICAL APPROACH WITH IDENTIFIABILITY"

CONTENTS

A Related work **2**

B When the PO-LiNGAM identifiability condition is violated **3**

B.1 Theoretical analysis 3

B.2 Experiments 4

C Details of identification algorithm **4**

C.1 Phase I 5

C.2 Phase II 5

C.3 Phase III 5

C.4 Supplementary algorithms for phase I 5

C.5 Check for multiple root atomic units 5

D Illustration of our algorithm **9**

E Computational complexity of our algorithm **10**

F Proofs **10**

F.1 Proof of Proposition 1 11

F.2 Proof of Theorem 1 11

F.3 Proof of Remark 1 12

F.4 Proof of Theorem 2 12

F.5 Proof of Remark 2 13

F.6 Proof of Corollary 1 13

F.7 Proof of Proposition 2 13

F.8 Proof of Theorem 3 13

F.9 Proof of Corollary 2 14

F.10 Proof of Theorem 4 14

F.11 Proof of Corollary 3 14

G More details on simulation experiments **14**

G.1 Data generation process and implementation 14

G.2 Evaluation metrics 15

G.3 More details on F1-score 15

G.4 More results of simulation experiments 15

H More details on the real-world data experiment **16**

I Notations and Terms **16**

A Related work

Inferring causal relationships from observed data is a popular field of statistics and artificial intelligence. As described in the introduction, from the scope of application, most methods can be divided into three areas: a) Infer the causal relations among measured variables with the assumption that there are no latent variables, for example, PC algorithm (Spirtes & Glymour, 1991), LiNGAM-based method (Shimizu et al., 2006), Greedy search (Chickering, 2002). b) Infer the causal relations among measured variables in the presence of latent variable interference, for example, FCI and its variants (Spirtes et al., 1995; Colombo et al., 2012; Claassen et al., 2013). c) Infer the causal relations among latent variables, for example, measurement models (Silva et al., 2006; Kummerfeld & Ramsey, 2016; Cai et al., 2019; Xie et al., 2020) and hierarchical models (Pearl, 1988; Zhang, 2004; Choi et al., 2011; Drton et al., 2017; Xie et al., 2022; Huang et al., 2022; Kong et al., 2023).

Depending on the different constraints and criteria used, most previous causal discovery methods can be divided into the following main categories:

1.) (Non-parametric setting) Conditional independence constraint: The PC algorithm (Spirtes & Glymour, 1991), as well as the FCI algorithm (Spirtes et al., 1995) and their variants (Colombo et al., 2012; Claassen et al., 2013), are prominent examples of non-parametric methods that infer causal relationships among measured variables by testing conditional independence. They do not care about whether the causal relationships are linear or not. However, these methods struggle with the latent variables and their causal relationships. Another limitation is that they yield multiple Markov equivalent structures, introducing ambiguity regarding the presence of confounders in the resulting graph.

2.) (Linear Gaussian setting) Tetrad condition and rank deficiency constraint: By investigating the sub-covariance matrix of measured variables, Tetrad condition (Spearman, 1928) allows to locate latent variables and infer the causal relationships among them in linear-Gaussian settings (Pearl, 1988; Silva et al., 2006; Kummerfeld & Ramsey, 2016). These methods generally assume that each latent variable possesses at least three pure measured variables as pure children, and each measured variable is exclusively a pure child of a single latent variable. Squires et al. (2022) presented a Tetrad-based method that allows latent variables to be influenced by measured variables. It relies on strong assumptions, such as the absence of edges between pairs of observed nodes or between pairs of latent nodes. The rank deficiency constraint is a general version of the Tetrad condition. Huang et al. (2022) use it to infer a latent hierarchical structure, which assumes k+1 pure children and k+1 neighbors for a set of latent variables with size k, and implicitly assumes all clustered children are pure children. Similar to the conditional independence constraint-based methods, they return a result that is asymptotic to the ground truth graph.

3.) (Linear non-Gaussian setting) Independent component analysis (ICA): Shimizu et al. (2006) leveraged non-Gaussianity of data and showed that a linear non-Gaussian acyclic model (LiNGAM) is identifiable based on ICA algorithm. By over-complete ICA, for example, Hoyer et al. (2008); Shimizu et al. (2009); Tashiro et al. (2014) learn the causal structure with latent variables. The limitation of over-complete ICA is that there are equivalence structures and the estimation is easy to fall into local optima. Recently, based on overcomplete ICA, Adams et al. (2021) established necessary and sufficient conditions for structure identifiability in both the linear non-Gaussian and the linear heterogeneous setting. However, it does not have a practical estimation approach and requires knowing the number of latent variables at the beginning.

4.) (Linear non-Gaussian setting) Independent noise, Triad constraint, and GIN condition: Shimizu et al. (2011) proposed DirectLiNGAM in 2011, which learns LiNGAM by replacing the ICA algorithm with regression and independence tests of noises. Triad constraint (Cai et al., 2019) is an extension of the independent noise condition for latent variables with the same assumption that the

noise terms are non-Gaussian, and is used to discover the structure of latent variables. GIN condition Xie et al. (2020) is a more general version that allows multiple parents behind multiple children. Xie et al. (2022) identify a latent hierarchical structure by using the GIN condition, which assumes one-factor clusters and measured variables to be descendants of latent variables.

5.) (Linear setting) Matrix decomposition and Optimization: Matrix decomposition techniques have been applied to model causal structures. Under specific conditions, the precision matrix can be decomposed into two components: a low-rank matrix that characterizes the causal relationships from latent variables to measured variables, and a sparse matrix that represents the relationships among measured variables (Chandrasekaran et al., 2010; 2011). By requiring triple measured variables than latent variables, Anandkumar et al. (2013) decompose the covariance matrix, and is capable of handling DAGs with effective depth one and multi-level Directed Acyclic Graphs (DAGs) (similar to latent hierarchical structures), with the help of low-order observable moments.

6.) Others: With the help of discrete latent variables, Kivva et al. (2021) proposed a mixture oracles-based method to identify the potentially nonlinear latent variable graph. Recently, Kong et al. (2023) learns a nonlinear latent hierarchical causal model by self-supervised representation learning. This innovative method accommodates general nonlinearity and multi-dimensional continuous variables. However, this approach still relies on strong assumptions regarding the data generation process and the estimation is easy to fall into local optima.

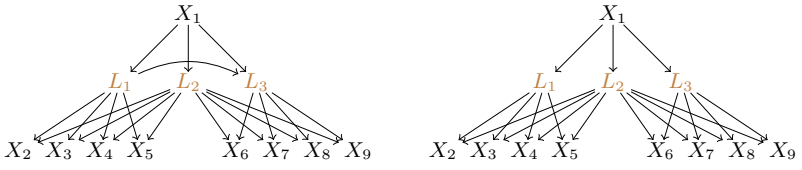
B When the PO-LiNGAM identifiability condition is violated

B.1 Theoretical analysis

PO-LiNGAM Identifiability Condition: a.) Any latent variable is in at least one atomic unit. b.) For any two atomic units with overlapping variables, their non-overlapping parts do not influence each other.

(i) Violation of identifiability condition(a): Our algorithm discovers new atomic units by clustering their individual pure children, and whether the number of pure children for latent variables satisfies the condition is tested during the discovery process, which is mainly achieved by Theorem 2, Remark 2, and Corollary 1. Meanwhile, please notice that our algorithm sequentially identifies the whole causal graph from leaves to roots. The higher-order causal structure can be inferred only after the lower-order causal structure are fully identified. As a consequence, the algorithm halts its inference of the higher-order causal structure when the latent set lacks a sufficient number of individual pure children to be located, so that it does not output spurious latent variables. In other word, since this condition is violated, it can not detect all latent variables, but we can still trust the discovered ones. The atomic units, whose causal orders are lower or equal to the unidentified latent set, and the causal relations among them are still identifiable.

(ii) Violation of identifiability condition(b): Let's consider the example in Figure 1(a), where two atomic units $\{\mathrm{L}_1, \mathrm{L}_2\}$ and $\{\mathrm{L}_2, \mathrm{L}_3\}$ are both caused by $\{\mathrm{X}_1\}$. Meanwhile, L_1 causes L_3, which violates identifiability condition(b). In our algorithm, for any two atomic units with overlapping variables, it ignores the influence of non-overlapping part of one atomic unit to that of another and infer their causal relationships separately, as illustrated in Remark 1 and Remark 2. Therefore, the discovered causal graph will ignore the causal relationships between non-overlapping parts of any two overlapping atomic units, as shown in Figure 1(b).

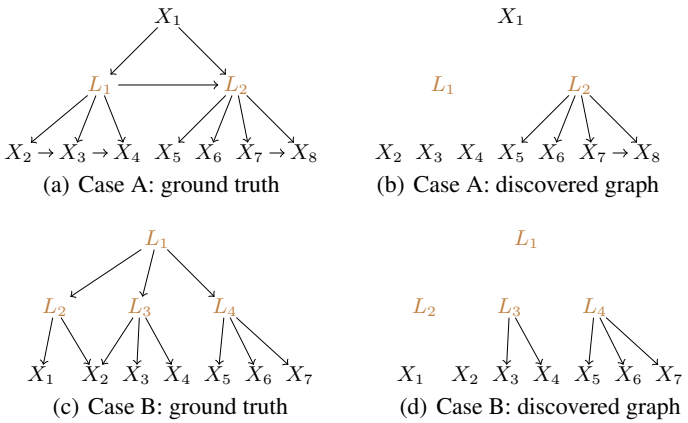


(a) Ground-truth graph (b) Discovered graph

Figure 1: An example of causal graphs where identifiability condition(b) is violated.

B.2 EXPERIMENTS

(i) To evaluate the performance of our algorithm under the violation of condition(a), we adjusted the causal graphs in Cases 2 and 3 of the simulation experiment, resulting in Case A and Case B of Figure 2. In Case A, L_1 lacks sufficient pure atomic unit children for discovery due to the causal chain $\mathrm{X}_2 \to \mathrm{X}_3 \to \mathrm{X}_4$. Nevertheless, the algorithm can still identify the causal structure of L_1 and its descendants, as shown in Figure 2(b). Moreover, in Case B, where L_2 cannot be discovered due to having only one pure atomic unit child, namely X_1, the algorithm still identifies the causal structure of L_3 and L_4, as shown in Figure 2(d). The corresponding average numerical result from ten executions is presented in Table 1. The high Correct Ordering Rate indicates that the discovered partial causal graph is highly accurate.



(a) Case A: ground truth (b) Case A: discovered graph

(c) Case B: ground truth (d) Case B: discovered graph

Figure 2: Causal graphs that violate identifiability condition(a), and the corresponding discovered results acquired by our method.

Table 1: The performance of our method on three causal graphs that violate the PO-LiNGAM identifiability condition. Case A and Case B violate the condition(a), as shown in Figure 2. Case C violates the condition(b), as shown in Figure 1.

	Case A			Case B			Case C		
	5k	10k	50k	5k	10k	50k	5k	10k	50k
Correct Ordering Rate ↑	**0.98**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.94**	**0.94**	**1.0**
Error Rate in Latent Variables ↓	0.45	0.5	0.5	0.53	0.5	0.5	0.66	0.52	0.33
F1-score ↑	0.55	0.56	0.56	0.60	0.62	0.62	0.54	0.66	0.85

(ii) To evaluate the performance of our algorithm under the violation of condition(b), we tested the algorithm with the ground-truth graph in Figure 1(a). The result discovered by our method is in Figure 1(b). The corresponding average numerical result from ten executions is presented in Case C of Table 1. The high Correct Ordering Rate indicates that the discovered partial causal graph is highly accurate.

C DETAILS OF IDENTIFICATION ALGORITHM

Our algorithm is iterative, with each iteration comprising three main phases, progressively uncovering the entire causal graph from leaf to root nodes. For each iteration, it first recursively identifies the leaf nodes and their parent nodes from the active atomic unit set $\mathbb{A}$ in the causal graph composed of atomic units of $\mathbb{A}$ and their latent confounders, and update $\mathbb{A}$ (**Phase I**). Then, we discover new atomic units by clustering their pure children in $\mathbb{A}$, and update $\mathbb{A}$ (**Phase II**). After that, we refine newly discovered atomic units by checking overlapping atomic units and decomposable 'atomic units', and update $\mathbb{A}$ (**Phase III**). At the end of an iteration, $\mathbb{A}$ is sent back to the Phase I and a new iteration starts until no more atomic units in $\mathbb{A}$ are updated.

C.1 PHASE I

The detailed algorithm for Phase I is in Algorithm 1.

C.2 PHASE II

The detailed algorithm for Phase II is in Algorithm 2.

C.3 PHASE III

The detailed algorithm for Phase III is in Algorithm 3.

C.4 SUPPLEMENTARY ALGORITHMS FOR PHASE I

To handle the situation explained at the end of section 3.2, we proposed two additional steps for Phase I, including Algorithm 4 (line 8, Alg. 1) and Algorithm 5 (line 14, Alg. 1).

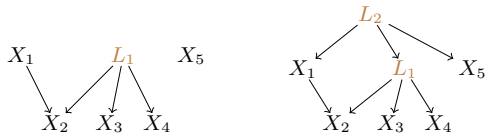
Small atomic units that can be completely covered by larger atomic units are sometimes indistinguishable from those larger atomic units due to the different causal order counting from the leaf nodes of the graph as well as the characterization of the GIN condition. The purpose of Algorithm 4 is to check this case by rechecking the number of parent variables of the related atomic unit children. If such a case happens, most children are re-added into $\mathbb{A}$ and treated the same as others in $\mathbb{A}$. For the special children that have partial parents not in $\mathbb{A}$, Algorithm 5 is used to relocate parents for them after the normal children are settled.

C.5 CHECK FOR MULTIPLE ROOT ATOMIC UNITS

Most causal discovery methods for handling latent variables Silva et al. (2006); Cai et al. (2019); Xie et al. (2020; 2022); Huang et al. (2022) implicitly assume that there are no unconditional independence between any two variables in the measured variable set $\mathbf{X}$. In other words, they assume that there is at most one root node in the causal graph, which may be a latent variable or a latent variable set with common children.

In practical situations, there may be more than one root node in a causal graph, not to mention that the input measured variables may form multiple disconnected subgraphs. In Figure 3(a), $\mathrm{X}_1 \ldots \mathrm{X}_4$ form a subgraph, while X_5 is not connected to any other variables. Furthermore, in the sub causal graph consisting of $\mathrm{X}_1 \ldots \mathrm{X}_4$, X_1 is unconditionally independent of both X_3 and X_4. In such a case, there are three root nodes: X_1, L_1, and X_5.

Our methods can handle the situation regardless of the number of root atomic units in the causal graph. At the end of the inference process, the algorithm tends to cluster all the true root atomic units (three or more) together and assigns them a common latent parent variable, as depicted in Figure 3(b). This occurs because when the active atomic unit set $\mathbb{A}$ contains only root atomic units, those root atomic units naturally satisfy the GIN condition and are unconditionally independent of each other without requiring any additional operations. It is considered that those root atomic units share a common latent parent variable with a value of 0. To verify the presence of multiple root atomic units in the causal graph, we can perform the independence tests. If the discovered root atomic unit comprises only one latent variable and lacks impure children, we should examine the unconditional independence of the pure atomic unit children of that discovered root atomic unit (For latent atomic units, their measured surrogate sets are used). If these pure children are found to be unconditionally independent of each other, then we can conclude that they are the true root atomic units of the causal graph. One the other hand, if the initially discovered causal graph has only two root atomic units, we can do the same unconditional independence test to check whether there is a common latent variable parent behind them.



(a) Ground-truth graph (b) Discovered graph

Figure 3: An example of causal structures where there are multiple root atomic units.

Algorithm 1: Phase I: IdentifyLeafNodesAndTheirParents

Input: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$
Output: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$

1 **repeat**
2 Select an atomic unit $\mathcal{U} \in \mathbb{A}$;
3 **for** *GrLen* $\leftarrow$ *1* **to** $|\mathbb{A} \backslash \mathcal{U}|$ **do**
4 **repeat**
5 Select atomic unit subset $\mathbb{P}$ from $\mathbb{A} \backslash \mathcal{U}$ such that $|\mathbb{P}| = GrLen$;
6 **if** $\mathcal{U}$ *with* $\mathbb{P}$ *satisfies Theorem 1 and Remark 1* **then**
7 $\mathbb{A} \leftarrow \mathbb{A} \backslash \mathcal{U}$, and update $\mathcal{G}$;
8 $(\mathcal{G}, \mathbb{A}) \leftarrow$ CheckSubAtomicUnits$(\mathcal{G}, \mathbb{A}, \mathbb{P})$;
9 Return to line 2;
10 **end**
11 **until** *all subsets with size GrLen in* $\mathbb{A} \backslash \mathcal{U}$ *selected*;
12 **end**
13 **until** *no more atomic units can be removed from* $\mathbb{A}$;
14 $(\mathcal{G}, \mathbb{A}) \leftarrow$ RefindChildrenForSubAtomicUnits$(\mathcal{G}, \mathbb{A})$;
15 **return** $\mathcal{G}, \mathbb{A}$

Algorithm 2: Phase II: DiscoverNewAtomicUnits

Input: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$
Output: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$

GrLen:= 2, $\mathbb{A}^{'}$:= $\mathbb{A}$, and NewAtomicUnitSet := $\emptyset$;
repeat
  ClusterList:= $\emptyset$;
  repeat
    Select subset $\mathbf{Y}$ from $\mathrm{MS}_b(\mathbb{A}^{'})$ such that $\|\mathbf{Y}\|$ = GrLen, $\mathbf{Y} \subseteq \mathrm{MS}_b(\mathbb{Y})$ and $\mathbb{Y} \geq 2$;
    if $\mathbf{Y}$ *satisfy the Theorem 2 and Remark 2* **then**
      Add $\mathbb{Y}$ into ClusterList;
    end
  until *all subset with size GrLen in* $\mathrm{MS}_b(\mathbb{A}^{'})$ *selected*;
  Gather the sets having overlapping atomic units with each other from the ClusterList as a set of sets and add it into another set $\mathbf{C}$, repeatedly and exhaustively;
  for *each* $\mathbf{C}_i \in \mathbf{C}$ **do**
    when GrLen= 2, **if** $\mathbf{C}_i$ contains only one set (bi-unit cluster) **then** identify pure children by Lemma 1 in Xie et al. (2022);
    Select one set $\in \mathbf{C}_i$, remove one atomic unit from it, get a base $\mathbb{B}$ with $\|\mathbb{B}\| \geq$ Grlen-1;
    PureChildSet = $\emptyset$; **for** each set $\mathbb{S} \in \mathbf{C}_i$ **if** $|\mathbb{S}| = |\mathbb{B}| + 1$ **do** add $\mathbb{S} \backslash \mathbb{B}$ into PureChildSet;
    for *each atomic unit* $\mathcal{B} \in \mathbb{B}$ **do**
      Select Grlen-1 atomic units from PureChildSet and patch $\mathcal{B}$ with them to form set $\mathbb{T}$;
    end
    if *bi-unit cluster is pure cluster, or* $|PureChild| \geq$ *Grlen-1 and all* $\mathbb{T}s \in \mathbf{C}_i$ **then**
      The atomic units in PureChildSet and $\mathbb{B}$ are all pure children;
      $\mathbb{A} \leftarrow \mathbb{A} \backslash \mathbb{T} \backslash$PureChildSet, and update $\mathcal{G}$;
      Create a new atomic unit $\mathcal{U}$ and save it in NewAtomicUnitSet;
    end
  end
  $\mathbb{A}^{'} \leftarrow \mathbb{A}^{'} \backslash$ClusterList, and GrLen $\leftarrow$ GrLen+1;
until *no more clusters are found*;
return $\mathcal{G}$, $\mathbb{A}$, and NewAtomicUnitSet

Algorithm 3: Phase III: RefineAtomicUnits

Input: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$, NewAtomicUnitSet
Output: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$

1 **repeat**
2 Select two atomic unit $\mathcal{U}_1, \mathcal{U}_2 \in \mathbb{A}$, where at least one in NewAtomicUnitSet;
3 **for** $m \leftarrow 1$ **to** $\|\mathcal{U}_1\|$ **do**
4 Select m measured surrogates from $\mathrm{MS}_b(\mathcal{U}_1)$, denoted as $\mathbf{U}_1$;
5 **if** *(*$\mathrm{MS}_a(\mathcal{U}_1 \cup \mathcal{U}_2)$, $\mathbf{U}_1 \cup \mathrm{MS}_b(\mathcal{U}_2)$*) follows GIN condition* **then**
6 $\mathcal{U}_1$ and $\mathcal{U}_2$ have $\|\mathcal{U}_1\| - \|\mathbf{U}_1\| + 1$ overlapping variables;
7 Update $\mathcal{G}$, and merge $\mathcal{U}_1$ with $\mathcal{U}_2$;
8 Return to line 2;
9 **end**
10 **end**
11 **until** *all binary atomic unit set selected*;
12 **for** *each atomic unit* $\mathcal{U} \in \mathbb{A}$ **do**
13 **for** each $\mathcal{S}_i \in \mathbb{A} \backslash \mathcal{U}$ **and** $\mathcal{S}_i \subseteq \mathcal{U}$ **do** add $\mathcal{S}_i$ into set $\mathbb{C}$;
14 **repeat**
15 Select a subset $\mathbb{S} \subseteq \mathbb{C}$;
16 **if** *atomic units in* $\mathbb{S}$ *do not overlap and* $\|\mathcal{U}\| = \|\mathbb{S}\|$ **then**
17 Move the children of $\mathcal{U}$ to below the atomic units in $\mathbb{S}$;
18 **if** $\mathbb{A}$ is not updated, **then** $\mathbb{A} \leftarrow \mathbb{A} \backslash \mathcal{U}$;
19 Update $\mathcal{G}$;
20 **end**
21 **until** *all subset in* $\mathbb{C}$ *selected*;
22 **end**
23 **return** $\mathcal{G}, \mathbb{A}$

Algorithm 4: CheckSubAtomicUnits

Input: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$, atomic unit set $\mathbb{P} \subseteq \mathbb{A}$
Output: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$

1 Let $\widehat{\mathbb{P}}$ be the union of $\mathbb{P}$ and the atomic units in $\mathbb{A}$ that overlap with atomic units in $\mathbb{P}$;
2 Let $\mathbb{U}$ contains the atomic units whose parents is a subset of $\widehat{\mathbb{P}}$;
3 GrLen:= 2, $\mathbb{U}^{'} := \mathbb{U}$;
4 **repeat**
5 ClusterList:= $\emptyset$;
6 **repeat**
7 Select subset $\mathbf{Y}$ from $\mathrm{MS}_b(\mathbb{U}^{'})$ such that $\|\mathbf{Y}\|$ = GrLen, $\mathbf{Y} \subseteq \mathrm{MS}_b(\mathbb{Y})$ and $\mathbb{Y} \geq 2$;
8 **if** $\mathbf{Y}$ *satisfy the Theorem 2 and Remark 2* **then**
9 Add $\mathbb{Y}$ into ClusterList;
10 **if** $\|\mathbf{Y}\| \leq$ *known* $\|\mathrm{Pa}(\mathbb{Y})\|$ **then**
11 **if** there are atomic units $\notin \mathbb{A}$, which has partial atomic unit parents in $\widehat{\mathbb{P}}$ and the rest atomic unit parents $\notin \mathbb{A}$, **then** mark them and the atomic units in $\mathbb{U}$ with matching symbols;
12 $\mathbb{A} \leftarrow (\mathbb{A} \backslash \widehat{\mathbb{P}}) \cup \mathbb{U}$, update $\mathcal{G}$, and go to line 18;
13 **end**
14 **end**
15 **until** *all subset with size GrLen in* $\mathrm{MS}_b(\mathbb{U}^{'})$ *selected*;
16 $\mathbb{U}^{'} \leftarrow \mathbb{U}^{'} \backslash$ClusterList, and GrLen $\leftarrow$ GrLen+1;
17 **until** *no more clusters are found*;
18 **return** $\mathcal{G}, \mathbb{A}$

Algorithm 5: RefineChildrenForSubAtomicUnits

Input: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$
Output: Partial causal graph $\mathcal{G}$, Active atomic unit set $\mathbb{A}$

for *each marked set of atomic units in* $\mathcal{G}$ **do**
  if *the atomic units in* $\mathbb{U}$ *are all relocated to be children of the smaller atomic units* **then**
    Relocate the parents for the atomic units that used to have partial atomic unit parents in $\widehat{\mathbb{P}}$ and the rest of atomic unit parents $\notin \mathbb{A}$, by Theorem 1 and Remark 1;
    Update $\mathcal{G}$;
  end
end
return $\mathcal{G}, \mathbb{A}$

D ILLUSTRATION OF OUR ALGORITHM

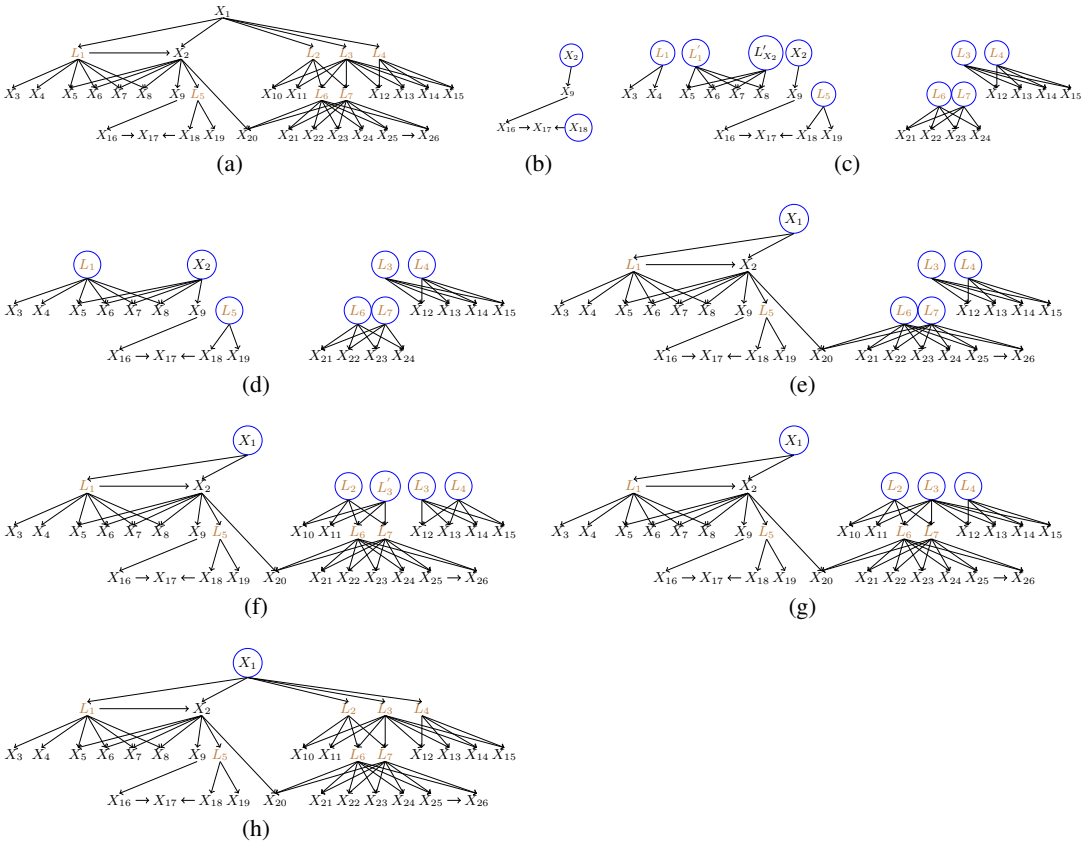


Figure 4: Illustration of the entire procedure of our algorithm. For each subgraph, the nodes in blue circles are active variables that are under investigation, and other measured variables for which causal relationships have not been identified are not shown for clarity. (a) Ground-truth structure. (b) Output after phase I of the first iteration of our algorithm. (c) Output after phase II of the first iteration. (d) Output after phase III of the first iteration. (e) Output after phase I of the second iteration. (f) Output after phase II of the second iteration. (g) Output after phase III of the second iteration. (h) Output after phase I of the third iteration, which has recovered the true causal structure.

In this section, we illustrate our algorithm with the ground-truth graph in Figure 4(a). We assume Oracle tests for GIN conditions. In this structure, $\mathrm{L}_1, \ldots, \mathrm{L}_7$ are latent variables and $\mathrm{X}_1, \ldots, \mathrm{X}_{26}$ are measured variables. We can see from the graph that measured variables are interspersed with

latent variables. The causal relationships among them are also complex. The estimation process is as follows:

i Initially, the active atomic unit set $\mathbb{A} = \{\{\mathrm{X}_1\}, \ldots, \{\mathrm{X}_{26}\}\}$ and graph $\mathcal{G} = \emptyset$.

ii During phase I of the first iteration, it first identifies $\mathrm{Pa}(\{\mathrm{X}_{17}\}) = \{\{\mathrm{X}_{16}\}, \{\mathrm{X}_{18}\}\}$, and remove $\{\mathrm{X}_{17}\}$ from the active set $\mathbb{A}$. After that, we sequentially identify $\mathrm{Pa}(\{\mathrm{X}_{16}\}) = \{\mathrm{X}_9\}$ and $\mathrm{Pa}(\{\mathrm{X}_9\}) = \{\mathrm{X}_2\}$, and update $\mathbb{A}$. The result after phase I of the first iteration is shown in Figure 4(b), and the current $\mathbb{A}$ = $\{\{\mathrm{X}_1\}, \ldots, \{\mathrm{X}_8\}, \{\mathrm{X}_{10}\}, \ldots, \{\mathrm{X}_{15}\}, \{\mathrm{X}_{18}\}, \ldots, \{\mathrm{X}_{26}\}\}$.

iii During phase II of the first iteration, we cluster the individual pure children in $\mathbb{A}$ to form new atomic units. It is worth noting that we do not know the true serial numbers of the latent variables. In the following, for the sake of clarity, we use the true serial numbers of latent variables to denote them. As shown in Figure 4(c), we introduce the new atomic unit $\{\mathrm{L}_1\}$ to be parent of individual pure children set $\{\{\mathrm{X}_3\}, \{\mathrm{X}_4\}\}$, the new atomic unit $\{\mathrm{L}'_1, \mathrm{L}'_{X_2}\}$ to be parent of individual pure children set $\{\{\mathrm{X}_5\}, \{\mathrm{X}_6\}, \{\mathrm{X}_7\}, \{\mathrm{X}_8\}\}$, the atomic unit $\{\mathrm{L}_3, \mathrm{L}_4\}$ to be parent of individual pure children set $\{\{\mathrm{X}_{12}\}, \{\mathrm{X}_{13}\}, \{\mathrm{X}_{14}\}, \{\mathrm{X}_{15}\}\}$, the atomic unit $\{\mathrm{L}_5\}$ to be parent of individual pure children set $\{\{\mathrm{X}_{18}\}, \{\mathrm{X}_{19}\}\}$, the atomic unit $\{\mathrm{L}_6, \mathrm{L}_7\}$ to be parent of individual pure children set $\{\{\mathrm{X}_{21}\}, \{\mathrm{X}_{22}\}, \{\mathrm{X}_{23}\}, \{\mathrm{X}_{24}\}\}$. We add the newly discovered atomic units into $\mathbb{A}$ and remove their individual pure children from $\mathbb{A}$.

iv During phase III of the first iteration, we find that the newly discovered atomic unit $\{\mathrm{L}'_1, \mathrm{L}'_{X_2}\}$ can be fully replaced by $\{\mathrm{L}_1\}$ and $\{\mathrm{X}_2\}$, as shown in Figure 4(d). We move the children of $\{\mathrm{L}'_1, \mathrm{L}'_{X_2}\}$ to $\{\mathrm{L}_1\}$ and $\{\mathrm{X}_2\}$, and remove $\{\mathrm{L}'_1, \mathrm{L}'_{X_2}\}$ from $\mathbb{A}$.

v Now, we start the second iteration. During phase I of the second iteration, we find $\mathrm{Pa}(\{\mathrm{L}_5\}) = \{\mathrm{X}_2\}$, $\mathrm{Pa}(\{\mathrm{X}_{20}\}) = \{\{\mathrm{X}_2\}, \{\mathrm{L}_6, \mathrm{L}_7\}\}$, $\mathrm{Pa}(\{\mathrm{X}_2\}) = \{\{\mathrm{L}_1\}, \{\mathrm{X}_1\}\}$, $\mathrm{Pa}(\{\mathrm{L}_1\}) = \{\mathrm{X}_1\}$, $\mathrm{Pa}(\{\mathrm{X}_{26}\}) = \{\{\mathrm{X}_{25}\}, \{\mathrm{L}_6, \mathrm{L}_7\}\}$, and $\mathrm{Pa}(\{\mathrm{X}_{25}\}) = \{\mathrm{L}_6, \mathrm{L}_7\}$, as shown in Figure 4(e).

vi In phase II of the second iteration, we introduce the new atomic unit $\{\mathrm{L}_2, \mathrm{L}'_3\}$ to be parent of individual pure children set $\{\{\mathrm{X}_{10}\}, \{\mathrm{X}_{11}\}, \{\mathrm{L}_6, \mathrm{L}_7\}\}$, as shown in Figure 4(f).

vii In phase III of the second iteration, we find that the newly discovered atomic unit $\{\mathrm{L}_2, \mathrm{L}'_3\}$ have one variable overlapping with atomic unit $\{\mathrm{L}_3, \mathrm{L}_4\}$, and merge them. The result is shown in Figure 4(g).

viii Now, we start the third iteration. During phase I of the third iteration, we find $\mathrm{Pa}(\{\mathrm{L}_2, \mathrm{L}_3\}) = \{\mathrm{X}_1\}$ and $\mathrm{Pa}(\{\mathrm{L}_3, \mathrm{L}_4\}) = \{\mathrm{X}_1\}$, as shown in Figure 4(h). After that, the current active atomic unit set contains only the measured variable$\{\mathrm{X}_1\}$, and the whole underlying causal graph is identified.

E COMPUTATIONAL COMPLEXITY OF OUR ALGORITHM

In this section, we analyze the complexity of our algorithm. Denote by n the number of active atomic units at the beginning of each phase. The complexity of phase I is upper bounded by $\mathcal{O}(n\sum_{k=1}^{n-1}\binom{n-1}{k})$. The phase II has the worst case complexity $\mathcal{O}(\sum_{k=2}^{n}\binom{n}{k})$. Denote by l the size of the largest atomic unit in the active atomic unit set. The complexity of phase III is upper bounded by $\mathcal{O}(l\binom{n}{2})$. The total complexity of the algorithm to discover the whole causal graph depends on the number of (both measured and latent) variables and the structural density of the causal graph, which determine how many iterations the algorithm needs to run.

F PROOFS

Before presenting the proofs of our results, we need a few more theorems and definitions derived from Xie et al. (2020; 2023).

Definition 6 (GIN condition (Xie et al., 2020)). *Let* $\mathbf{Z}$, $\mathbf{Y}$ *be sets of variables in a linear non-Gaussian acyclic causal model. We say that* $(\mathbf{Z}, \mathbf{Y})$ *follows GIN condition if and only if* $\omega^\intercal\mathbf{Y}$ *are statistically independent of* $\mathbf{Z}$*, where* ω *satisfies* $\omega^\intercal\mathbb{E}[\mathbf{Y}\mathbf{Z}^\intercal] = 0$ *and* $\omega \neq 0$.

In other words, $(\mathbf{Z}, \mathbf{Y})$ violates the GIN condition if and only if $E_{\mathbf{Y}||\mathbf{Z}}$ is dependent on $\mathbf{Z}$.

Recently, Xie et al. (2023) extended the original graphical criteria of the GIN condition with the help of trek and trek-separation (t-separation) (Sullivant et al., 2010). We next describe the notion of the trek and trek-separation criterion, which is more general than d-separation in linear causal models. After that, we show the graphical implication of the GIN condition in PO-LiNGAM, which helps to exploit the GIN condition to discover the causal graph.

Definition 7 (trek (Sullivant et al., 2010)). *A trek in $\mathcal{G}$ from $\mathbf{i}$ to $\mathbf{j}$ is an ordered pair of directed paths $(\mathbf{P_1}, \mathbf{P_2})$ where $\mathbf{P_1}$ has sink $\mathbf{i}$, $\mathbf{P_2}$ has sink $\mathbf{j}$, and both $\mathbf{P_1}$ and $\mathbf{P_2}$ have the same source $\mathbf{k}$. The common source $\mathbf{k}$ is called the top of the trek, denoted top$(\mathbf{P_1}, \mathbf{P_2})$. Note that one or both of $\mathbf{P_1}$ and $\mathbf{P_2}$ may consist of a single vertex, that is, a path with no edges.*

Definition 8 (t-separation (Sullivant et al., 2010)). *Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C_A}$, and $\mathbf{C_B}$ be four subsets of $\mathbf{V}$. We say the ordered pair ($\mathbf{C_A}$, $\mathbf{C_B}$) t-separates $\mathbf{A}$ from $\mathbf{B}$ if, for every trek (τ_1; τ_2) from a vertex in $\mathbf{A}$ to a vertex in $\mathbf{B}$, either τ_1 contains a vertex in $\mathbf{C_A}$ or τ_2 contains a vertex in $\mathbf{C_B}$.*

Theorem 5 (GIN Graphical Criteria in PO-LiNGAM). *Let $\mathbf{Y}$ and $\mathbf{Z}$ be two sets of measured variables of a partially observed linear non-Gaussian acyclic causal model (PO-LiNGAM). Assume the rank-faithfulness holds. $(\mathbf{Z}, \mathbf{Y})$ satisfies the* GIN *condition if and only if there exists a variable set $\mathcal{S}$ with $0 \leq Dim(\mathcal{S}) \leq min(Dim(\mathbf{Y}) - 1, Dim(\mathbf{Z}))$, such that 1) the order pair $(\emptyset, \mathcal{S})$ t-separates $\mathbf{Z}$ and $\mathbf{Y}$, and that 2) the covariance matrix of $\mathcal{S}$ and $\mathbf{Z}$ has rank $Dim(\mathcal{S})$, and so does that of $\mathcal{S}$ and $\mathbf{Y}$.*

Note: In this paper, $Dim(\mathcal{S}) = \|\mathcal{S}\|$ represents the number of variables in $\mathcal{S}$, regardless of whether $\mathcal{S}$ is a set of variables or a set of atomic units.

Roughly speaking, the conditions in this theorem can be interpreted in the following way: i.) a causally earlier subset (according to the causal order) of $\mathbf{Y}$ t-separates $\mathbf{Y}$ from $\mathbf{Z}$, and ii.) the linear transformation from that subset of the common causes to $\mathbf{Z}$ has full column rank. Based on the definition of d-separation, the first interpretation is that the causally earlier subset $\mathcal{S}$ (according to the causal order) of $\mathbf{Y}$ d-separates $\mathbf{Y}\backslash\mathcal{S}$ from $\mathbf{Z}\backslash\mathcal{S}$.

Proof. Recently, it has been shown in Xie et al. (2023) that the graphical criteria hold in a linear non-Gaussian acyclic causal model. In (Xie et al., 2020; 2022; 2023), they allowed part variables to be latent. In our paper, the latent variable can be the descendent or ancestor of measured variables, which does not affect the graphical criteria of GIN because linear causal models are transitive. □

F.1 Proof of Proposition 1

Proof. Let $\mathbf{V}$ be a set of variables satisfying PO-LiNGAM with no latent variables, and V be a variable in $\mathbf{V}$.

(i) Assume that V is a leaf variable of the causal graph, and $\mathbf{P}$ are its all parents. We know that a) $0 \leq \|\mathbf{P}\| \leq \min(\|\mathrm{V} \cup \mathbf{P}\| - 1, \|\mathbf{V}\backslash\mathrm{V}\|)$, and b) $\mathbf{P}$ is the minimal set so that $(\emptyset, \mathbf{P})$ t-separates $\mathbf{V}\backslash\mathrm{V}$ and $\mathrm{V} \cup \mathbf{P}$. Furthermore, the set of common components between $\mathrm{V} \cup \mathbf{P}$ and $\mathbf{V}\backslash\mathrm{V}$ is $\mathbf{P}$, so the covariance matrix of $\mathbf{P}$ and $\mathbf{V}\backslash\mathrm{V}$ has rank $\|\mathcal{S}\|$, and so does that of $\mathbf{P}$ and $\mathrm{V} \cup \mathbf{P}$. Therefore, $(\mathbf{V}\backslash\mathrm{V}, \mathrm{V} \cup \mathbf{P})$ follows the GIN condition and there is no $\tilde{\mathbf{P}} \subset \mathbf{P}$ such that $(\mathbf{V}\backslash\mathrm{V}, \mathrm{V} \cup \tilde{\mathbf{P}})$ follows the GIN condition.

(ii) Assume that $(\mathbf{V}\backslash\mathrm{V}, \mathrm{V} \cup \mathbf{P})$ follows the GIN condition and there is no $\tilde{\mathbf{P}} \subset \mathbf{P}$ such that $(\mathbf{V}\backslash\mathrm{V}, \mathrm{V} \cup \tilde{\mathbf{P}})$ follows the GIN condition. We know that $\mathbf{P}$ is the minimal set that $(\emptyset, \mathbf{P})$ t-separates $\mathbf{V}\backslash\mathrm{V}$ and $\mathrm{V} \cup \mathbf{P}$, and therefore, $\mathbf{P}$ is causal earlier than V. In a causal graph, the minimal causal earlier set that t-separates one variable from the other variables is the parent set of that variable.

Therefore, from (i) and (ii), the proposition is proved. □

F.2 Proof of Theorem 1

Proof. The proof of theorem 1 is similar to the proof of Proposition 1. Let $\mathbb{U}$ be a set of non-overlapping atomic units satisfying PO-LiNGAM with known $\mathrm{MS}_{a,b}$ for each atomic unit in $\mathbb{U}$, and $\mathcal{U}$ be an atomic unit in $\mathbb{U}$.

(i) Assume that $\mathcal{U}$ is a leaf atomic unit in the current causal graph, and $\mathbb{P}$ is the set of parent atomic units of $\mathcal{U}$. We know that a) $0 \leq \|\mathbb{P}\| \leq \min(\|\mathcal{U} \cup \mathbb{P}\| - 1, \|\mathbb{U} \backslash \mathcal{U}\|)$, and b) $\mathbb{P}$ is the minimal set so that $(\emptyset, \mathbb{P})$ t-separates $\mathbb{U} \backslash \mathcal{U}$ and $\mathcal{U} \cup \mathbb{P}$. Furthermore, the set of common components between $\mathcal{U} \cup \mathbb{P}$ and $\mathbb{U} \backslash \mathcal{U}$ is $\mathbb{P}$, so the covariance matrix of $\mathbb{P}$ and $\mathcal{U} \cup \mathbb{P}$ has rank $\|\mathbb{P}\|$, and so does that of $\mathbb{P}$ and $\mathbb{U} \backslash \mathcal{U}$. And because some of the atomic units may be latent, we use their measured surrogate variable set to represent them. Therefore, $\big(\mathrm{MS}_a(\mathbb{U} \backslash \mathcal{U}), \mathrm{MS}_b(\mathcal{U} \cup \mathbb{P})\big)$ follows the GIN condition and there is no $\tilde{\mathbb{P}} \subset \mathbb{P}$ such that $\big(\mathrm{MS}_a(\mathbb{U} \backslash \mathcal{U}), \mathrm{MS}_b(\mathcal{U} \cup \tilde{\mathbb{P}})\big)$ follows the the GIN condition. The same is true if we represent them using another set of measured surrogate variables.

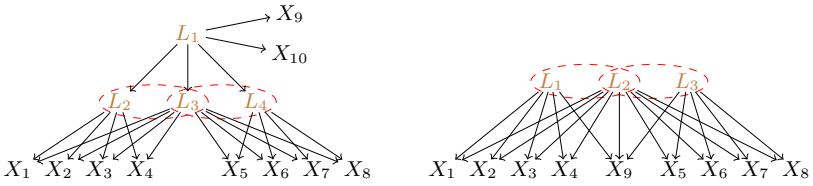
(ii) Assume that $\big(\mathrm{MS}_a(\mathbb{U} \backslash \mathcal{U}), \mathrm{MS}_b(\mathcal{U} \cup \mathbb{P})\big)$ follows the GIN condition and there is no $\tilde{\mathbb{P}} \subset \mathbb{P}$ such that $\big(\mathrm{MS}_a(\mathbb{U} \backslash \mathcal{U}), \mathrm{MS}_b(\mathcal{U} \cup \tilde{\mathbb{P}})\big)$ follows the the GIN condition. $\mathrm{MS}_a(\mathbb{U} \backslash \mathcal{U})$ contains a set of measured surrogate set of $\mathbb{U} \backslash \mathcal{U}$, $\mathrm{MS}_b(\mathcal{U} \cup \mathbb{P})$ contains another set of measured surrogate set of $\mathcal{U} \cup \mathbb{P}$. The common component of $\mathbb{U} \backslash \mathcal{U}$ and $\mathcal{U} \cup \mathbb{P}$ is $\mathbb{P}$. And, without $\mathrm{MS}_b(\mathbb{P})$, the GIN condition cannot be satisfied singly using $\mathrm{MS}_b(\mathcal{U})$ as the $\mathbf{Y}$ set. Thus, $\mathbb{P}$ is the minimal set that $(\emptyset, \mathbb{P})$ t-separates $\mathrm{MS}_a(\mathbb{U} \backslash \mathcal{U})$ and $\mathrm{MS}_b(\mathcal{U} \cup \mathbb{P})$, and therefore, $\mathbb{P}$ is causal earlier than $\mathcal{U}$. In the causal graph, the minimal causal earlier atomic unit set that t-separates one atomic unit from the other atomic units is the parent set of that atomic unit.

Therefore, from (i) and (ii), the theorem is proved. □

F.3 Proof of Remark 1

Proof. (i) The first term is to eliminate the effect of overlapping atomic unit sets when searching for parents, since for any two atomic units with overlapping variables, we don't have enough information to know whether the non-overlapping part of one atomic unit is the parent of the non-overlapping part of another atomic unit. As illustrated in case 1 of Figure 5, If we want to know the parent of the atomic unit $\{\mathrm{L}_2, \mathrm{L}_3\}$, we need remove the atomic unit $\{\mathrm{L}_3, \mathrm{L}_4\}$ from $\mathbb{U}$ and $\mathbb{P}$. Otherwise, both $\{\mathrm{L}_3, \mathrm{L}_4\}$ and $\{\mathrm{L}_1\}$ will be considered as the parents of $\{\mathrm{L}_2, \mathrm{L}_3\}$ rather than the true parent $\{\mathrm{L}_1\}$.

(ii) For the second term, let us consider the case 2 of Figure 5, where we want to find the parent of $\{\mathrm{X}_9\}$. $\mathbb{P} = \{\{\mathrm{L}_1, \mathrm{L}_2\}, \{\mathrm{L}_2, \mathrm{L}_3\}\}$ and $\mathcal{U} = \{\mathrm{X}_9\}$. There is no point in testing the GIN condition for $\big(\mathrm{MS}_a(\mathbb{U} \backslash \mathcal{U}), \mathrm{MS}_b(\mathcal{U} \cup \mathbb{P})\big)$, since even without $\mathrm{MS}_b(\mathcal{U})$ in $\mathbf{Y}$ set, the graph criteria of GIN condition is still satisfied, so is GIN condition. That is because those two atomic units $\{\mathrm{L}_1, \mathrm{L}_2\}$ and $\{\mathrm{L}_2, \mathrm{L}_3\}$ have overlapping L_2, and therefore the unrepeated variable set behind their measured surrogate variable has size 3 instead of 4. Thus, we need to find a subset of $\mathrm{MS}_b(\mathbb{P})$ so that the number of variables is equal to the true size $\|\mathbb{P}\|$. That is the maximal subset $\mathbf{S}$ of $\mathrm{MS}_b(\mathbb{P})$ that makes $\big(\mathrm{MS}_a(\mathbb{U} \backslash \mathcal{U}), \mathbf{S}\big)$ does not follow GIN condition. □



(a) Case 1 (b) Case 2

Figure 5: The illustrative examples for remark 1.

F.4 Proof of Theorem 2

The aim behind Theorem 2, Remark 2, and Corollary 1 is to find a sufficiently large number of pure children to support the discovery of new atomic units.

Proof. Let $\mathbb{U}$ be a set of non-overlapping atomic units satisfying PO-LiNGAM with known $\mathrm{MS}_{a,b}$ for each atomic unit in $\mathbb{U}$, and there is no such leaf atomic unit in $\mathbb{U}$ that its full parents are also in $\mathbb{U}$. Assume that $(\mathrm{MS}_a(\mathbb{U} \backslash \mathbb{Y}), \mathbf{Y})$ follows the GIN condition, and there is no subset $\tilde{\mathbf{Y}} \subset \mathbf{Y}$ such that $(\mathrm{MS}_a(\mathbb{U} \backslash \mathbb{Y}), \tilde{\mathbf{Y}})$ follows the GIN condition. It means that there is a variable set $\mathcal{S}$ with size

$\|\mathbf{Y}\| - 1$, which is causal earlier than $\mathbb{Y}$, and the order pair $(\emptyset, \mathcal{S})$ t-separates $\mathrm{MS}_a(\mathbb{U}\backslash\mathbb{Y})$ and $\mathbf{Y}$. Since we know in $\mathbb{U}$ there is no leaf atomic unit that its full parents are also in $\mathbb{U}$, $\mathcal{S}$ contains at least one latent variable which is the parent of $\mathbf{Y}$. However, $\mathcal{S}$ may also contain variables that are also in $\mathbf{Y}$. The third statement is to check out this situation by explicitly adding the variables in $\mathcal{P}$ into $\mathcal{S}$ and testing the GIN condition. Therefore, with these three statements, we can ensure that behind $\mathbb{Y}$ there is a parent variable set $\mathbf{P}$ involving at least one latent variable, which has size $\|\mathbf{Y}\| - 1$ and can d-separates $\mathbb{Y}$ from the atomic units of $\mathbb{U}\backslash(\mathbb{Y} \cup \mathbf{P})$. □

F.5 Proof of Remark 2

Proof. It is obvious that Remark 2 does not affect the correctness of Theorem 2 whether or not there are overlapping atomic units in $\mathbb{U}$. However, it is reasonable that the parents of the covered atomic unit must be a subset of or the same as the parents of the covering atomic unit. As illustrated in Figure 6, the atomic unit $\{\mathrm{L}_2\}$ is fully covered by $\{\mathrm{L}_2, \mathrm{L}_3\}$, and both $\{\mathrm{L}_2\}$ and $\{\mathrm{L}_2, \mathrm{L}_3\}$ have parent $\{\mathrm{L}_1\}$. After removing $\{\mathrm{L}_2\}$ from the active set, the atomic unit $\{\mathrm{L}_2, \mathrm{L}_3\}$ can be considered as an individual pure child for discovering the new atomic unit $\{\mathrm{L}_1\}$ according to Corollary 1. □

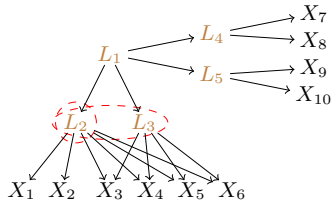


Figure 6: The illustrative examples for remark 2.

F.6 Proof of Corollary 1

Proof. If one atomic unit $\mathcal{Y}$ is replaceable in the $\mathbb{Y}$ that satisfies Theorem 2 with Remark 2, then we know that in the current causal graph composed of atomic units in $\mathbb{U}$ and their latent confounders, $\mathcal{Y}$ is not directly causally connected to the other atomic units in $\mathbb{Y}\backslash\mathcal{Y}$, except for their common total parent set. The reverse is also true. □

F.7 Proof of Proposition 2

Proof. We demonstrate the utility of Proposition 2 and prove it with the example in Figure 7. With our algorithm, at the beginning, the active set contains only measured variables. After Phase 1 of the first iteration, nothing is identified. During Phase 2 of the first iteration, we cluster the pure children to discover new atomic units. It is obvious that we can cluster $\{\mathrm{X}_8\}$ and $\{\mathrm{X}_9\}$ to form the atomic unit $\{\mathrm{L}_4\}$. However, when clustering for discovering the atomic unit $\{\mathrm{L}_1, \mathrm{L}_2\}$, the atomic unit $\{\mathrm{X}_2\}$ will also be clustered together with any two of $\{\{\mathrm{X}_3\}, \{\mathrm{X}_4\}, \{\mathrm{X}_5\}, \{\mathrm{X}_6\}\}$, and considered as a individual pure children of the atomic unit $\{\mathrm{L}_1, \mathrm{L}_2\}$, though $\{\mathrm{X}_2\}$ is actually the individual pure child of $\{\mathrm{L}_1\}$. That is because that atomic unit $\{\mathrm{L}_1, \mathrm{L}_2\}$ covers atomic unit $\{\mathrm{L}_1\}$ and when clustering for $\{\mathrm{L}_1, \mathrm{L}_2\}$, $\{\mathrm{L}_1\}$ does not have enough individual pure children to separate it from $\{\mathrm{L}_1, \mathrm{L}_2\}$. In the following iterations, $\{\mathrm{L}_3\}$ will be discovered and identified as a child of $\{\mathrm{L}_1, \mathrm{L}_2\}$. After that, $\{\mathrm{L}_1\}$ can be separated from $\{\mathrm{L}_1, \mathrm{L}_2\}$ by re-clustering the children of $\{\mathrm{L}_1, \mathrm{L}_2\}$. We know that $\{\mathrm{X}_2\}$ and $\{\mathrm{L}_3\}$ can be clustered, which shows the number of parent variables behind them is 1 instead of 2, and therefore, there is a small atomic unit covered by $\{\mathrm{L}_1, \mathrm{L}_2\}$. □

F.8 Proof of Theorem 3

Proof. Let $\mathcal{U}_1$ and $\mathcal{U}_2$ be two atomic units. Let $\mathbf{U}_1$ be part of variables in $\mathrm{MS}_b(\mathcal{U}_1)$.

(i) Assume that $\mathcal{U}_1$ and $\mathcal{U}_2$ are two discovered atomic units. $(\mathrm{MS}_a(\mathcal{U}_1 \cup \mathcal{U}_2), \mathrm{MS}_b(\mathcal{U}_1 \cup \mathcal{U}_2))$ can satisfy GIN condition only if $\mathcal{U}_1$ and $\mathcal{U}_2$ have overlapping variables. According to Theorem 5, $\mathcal{S} = \mathcal{U}_1 \cup \mathcal{U}_2$. $Dim(\mathcal{S})$ should be smaller or equal to $Dim(\mathcal{U}_1) + Dim(\mathcal{U}_2) - 1$ so that the graph

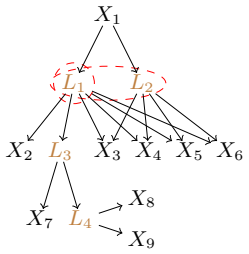


Figure 7: The illustrative examples for proposition 2.

criteria of the GIN condition is satisfied. It means $\mathcal{U}_1$ and $\mathcal{U}_2$ have overlapping variables. When $\mathbf{U}_1$ is the minimal set that $(\mathrm{MS}_a(\mathcal{U}_1 \cup \mathcal{U}_2), \mathbf{U}_1 \cup \mathrm{MS}_b(\mathcal{U}_2))$ follows the GIN condition, we know the true number $\|\mathcal{U}_1 \cup \mathcal{U}_2\|$ is $\|\mathcal{U}_2\| + \|\mathbf{U}_1\| - 1$. Therefore, we know that the two atomic units have $\|\mathcal{U}_1\| - \|\mathbf{U}_1\| + 1$ overlapping variables in total.

(ii) Assume that $\mathcal{U}_1$ and $\mathcal{U}_2$ have $\|\mathcal{U}_1\| - \|\mathbf{U}_1\| + 1$ overlapping variables. According to the graphical criteria of PO-LiNGAM (Theorem 5), we can know that $\mathbf{U}_1$ is the minimal set that $(\mathrm{MS}_a(\mathcal{U}_1 \cup \mathcal{U}_2), \mathbf{U}_1 \cup \mathrm{MS}_b(\mathcal{U}_2))$ follows the GIN condition. □

F.9 Proof of Corollary 2

Proof. The proof of Corollary 2 is obvious. For a set of atomic units $\mathbb{S}$ that do not overlap with each other and can be covered by atomic unit $\mathcal{U}$, $\|\mathcal{U}\| = \|\mathbb{S}\|$ is equivalent to the fact that each atomic unit in $\mathbb{S}$ is part of $\mathcal{U}$, and together they can form $\mathcal{U}$. □

F.10 Proof of Theorem 4

Proof. In section 3.1, we show that the causal structure among atomic units is identifiable by Theorem 1 and Remark 1. In section 3.2, we show that the new atomic units can be discovered by clustering their pure children with the help of Theorem 2, Remark 2, and Corollary 1. In section 3.3, we refine the discovered atomic units by Theorem 3 and Corollary 2.

The three sections are integrated into our iterative algorithm to discover the causal graph from the leaf to root nodes. For the possible problem with covered small atomic units during the iterative discovery process, we check and identify it through Proposition 2. Furthermore, at the end of discovery process, we check for multiple root atomic units by testing unconditional independence relationships. Therefore, the atomic units, their size, and the causal structure among them can be fully identified with our algorithm. It is obvious that if there are no direct causal relationships between variables within atomic units, the entire causal graph $\mathcal{G}$ is virtually identifiable. □

F.11 Proof of Corollary 3

Proof. Corollary 3 is similar to Theorem 4, but with stronger assumptions. If each latent variable has at least two pure variable children, each atomic unit we discovered will only contain one variable. Thus, we do not need to consider the problem that the discovered atomic units have overlapping variables. Similarly, we need to check whether there are multiple root atomic units at the end. Roughly speaking, with Theorem 1, Theorem 2, and Corollary 1, the entire causal graph $\mathcal{G}$ is fully identifiable. Furthermore, if there is no latent confounder, we need not consider the discovery of latent variables. The entire causal graph $\mathcal{G}$ is identifiable with only Proposition 1, and the result is the same as the typical LiNGAM discovery algorithm (Shimizu et al., 2006; 2011). □

G More details on simulation experiments

G.1 Data generation process and implementation

According to Eq. 1, we generated the causal strength $b_{i,j}$ uniformly from $[-2, -0.5] \cup [0.5, 2]$ and the non-Gaussian noise terms were generated from exponential distributions to the second power.

For BPC and IL^2H, Gaussian noise is used. HSIC-based independence tests (Zhang et al., 2018) were used to test the GIN condition.

G.2 Evaluation metrics

We adapted the evaluation metrics from Xie et al. (2022) and Huang et al. (2022) to evaluate the results. They are:

a.) Correct ordering rate: the number of correctly inferred causal orderings divided by the total number of inferred causal orderings.

b.) Error rate in latent variables: the absolute difference between the inferred number of latent variables and the number of latent variables in the true structure divided by the number of latent variables in the true structure.

c.) F1-score: $2\frac{\text{precision}*\text{recall}}{\text{precision}+\text{recall}}$ to measure the similarity between the inferred adjacency matrix and the true adjacency matrix.

G.3 More details on F1-score

In this paper, an adjacency matrix is a square matrix whose each element represents whether or not the variables in the graph are connected by edges with direction. For example, $Adj_{\mathcal{G}}(i,j) = 1$ represents there is a direct edge from i_{th} point to j_{th} point, and $Adj_{\mathcal{G}}(i,j) = 0$ represents there is no direct edge from i_{th} point to j_{th} point.

We used the F1-score to calculate the percentage of similarity between estimated and ground-truth adjacency matrices. To calculate the F1-score, we first calculate the precision = $\frac{\text{true positive}}{\text{total test positive}}$ that represents the number of correct inferred edges over number of total inferred edges, and the recall = $\frac{\text{true positive}}{\text{total true positive}}$ that represents the number of correct inferred edges over number of total ground-truth edges. F1-score = $2\frac{\text{precision}*\text{recall}}{\text{precision}+\text{recall}}$, which combines the precision and recall of the adjacency matrices by taking their harmonic mean, and can better represent the structural recovery rate.

In practice, the latent variable indices in the estimated graphs may not match those in the real graphs. To remove this ambiguity, similar to Huang et al. (2022), we permuted the latent variable indices in the estimated graphs and used the one that has the minimal difference from the true graph. In addition, if the estimated number of latent variables is smaller than the true number of latent variables, we add extra latent variables that do not have edges with others to $\mathcal{G}$. If the estimated number of latent variables is larger than the true number of latent variables, we find a subset of the latent variables in $\mathcal{G}$ that best matches the true one.

G.4 More results of simulation experiments

Here, we evaluate the running time for four simulation cases. Each case is executed ten times, and the results are averaged. The summarized results are reported in the Table 2.

Table 2: The averaged running times for various cases with different sample sizes.

	Case 1			Case 2			Case 3			Case 4		
	5k	10k	50k	5k	10k	50k	5k	10k	50k	5k	10k	50k
Time (s)	8	10	18	25	26	47	18	20	42	2026	2164	3132

To evaluate the performance of our algorithm when some noise is Gaussian-distributed, we modified the first two cases of the simulation experiment. In Case 1, we generate standard Gaussian noise for X_2 and X_4. In Case 2, we generate standard Gaussian noise for X_2 and X_5. Each case is executed ten times, and the results are averaged. The summarized results are reported in the Table 3. Although some of the noise is Gaussian-distributed, our method still recovers the true structures.

Table 3: The performance of our method on causal structure graphs in the presence of Gaussian-distributed noise.

	Case 1			Case 2		
	5k	10k	50k	5k	10k	50k
Correct Ordering Rate ↑	0.93	1.0	0.94	0.91	0.95	0.95
Error Rate in Latent Variables ↓	0.0	0.0	0.0	0.20	0.15	0.10
F1-score ↑	0.95	0.97	0.95	0.91	0.93	0.96

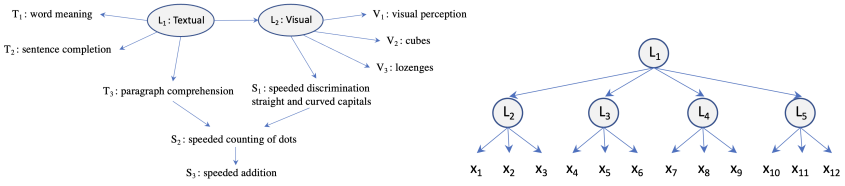
H MORE DETAILS ON THE REAL-WORLD DATA EXPERIMENT

The Holzinger and Swineford (1939) dataset consists of mental ability test scores of seventh- and eighth-grade children from two different schools (Pasteur and Grant-White). In the original dataset, there are scores for 26 tests. However, a smaller subset with 9 variables is more widely used in the literature, for example, in Jöreskog (1969). This dataset can be retrieved from R package lavaan (Rosseel, 2012). The variables used in our paper are given in Table 4, which can be broadly categorized into three dimensions: Visual, Textual, and Speeded.

Table 4: The details of 9 variables in 'Holzinger & Swineford 1939' dataset.

Visual	Visual perception test from Spearman VPT Part I
	Cubes, Simplification of Brighams Spatial Relations Test
	Lozenges from Thorndike-Shapes flipped over then identify target
Textual	Word Meaning Test
	Sentence Completion Test
	Paragraph Comprehension Test
Speeded	Speeded discrimation of straight and curved caps
	Speeded counting of dots in shapes
	Speeded addition test

The second-order factor analysis model dataset is a built-in dataset of the Mplus software (Muthén & Muthén, 2017). The ground truth is the same as in Figure 8(b). Each set of three measured variables has a corresponding first-order latent indicator, all of which are influenced by a second-order latent indicator. A detailed explanation of the dataset and model can be found in `http://www.statmodel.com/HTML_UG/chapter5V8.htm`.



(a) (b)

Figure 8: The outputs on (a) 'Holzinger & Swineford 1939' dataset and (b) Second-order factor analysis dataset. The nodes in circles are the latent indicators discovered from measured variables.

I NOTATIONS AND TERMS

Table 5: More explanations on Notions and Terms.

Notation / Term	Description	Initially Used
X	A measured variable.	First paragraph of Section 2.1
L	A latent variable.	First paragraph of Section 2.1
V	A variable, which is either measured or latent.	First paragraph of Section 2.1
$\mathbf{V}$	Boldfaced letter; A set of variables, each of which is either measured or latent.	First paragraph of Section 2.1
$\mathcal{V}, \mathcal{U}$	Script letter; An atomic unit, whose definition is in Definition 3.	Definition 3
$\mathbb{U}, \mathbb{Y}$	Blackboard bold letter; A set of atomic units.	Theorem 1
$\|\| * \|\|$ (e.g., $\mathbf{V}$, $\mathbb{U}$)	The number of variables in set * ($\mathbf{V}$, $\mathbf{U}$).	Definition 3
$\| * \|$ (e.g., $\mathbb{U}$)	The number of atomic units in set * ($\mathbb{U}$).	Algorithm 2
$\mathrm{Pa}(*)$ (e.g., $\mathrm{V}, \mathcal{U}$)	The parent of * ($\mathrm{V}, \mathcal{U}$).	First paragraph of Section 2.1
Pure atomic unit child	Each pure child itself is an atomic unit.	Definition 3
Surrogate variable of $\mathcal{U}$	Each descendant is a surrogate variable of $\mathcal{U}$.	Definition5
$\mathrm{MS}_{a,b}(\mathcal{U})$	The measured surrogate variable set a/b of $\mathcal{U}$.	Definition5
Decomposable 'atomic unit'	The discovered 'atomic unit' during phase II, which can be fully decomposed into several small true atomic units during phase III.	First paragraph of Section 3.3

REFERENCES

Jeffrey Adams, Niels Hansen, and Kun Zhang. Identification of partially observed linear causal models: Graphical conditions for the non-gaussian and heterogeneous cases. *Advances in Neural Information Processing Systems*, 34:22822–22833, 2021.

Animashree Anandkumar, Daniel Hsu, Adel Javanmard, and Sham Kakade. Learning linear bayesian networks with latent variables. In *International Conference on Machine Learning*, pp. 249–257. PMLR, 2013.

Ruichu Cai, Feng Xie, Clark Glymour, Zhifeng Hao, and Kun Zhang. Triad constraints for learning causal structure of latent variables. *Advances in neural information processing systems*, 32, 2019.

Venkat Chandrasekaran, Pablo A Parrilo, and Alan S Willsky. Latent variable graphical model selection via convex optimization. In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1610–1613. IEEE, 2010.

Venkat Chandrasekaran, Sujay Sanghavi, Pablo A Parrilo, and Alan S Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov):507–554, 2002.

Myung Jin Choi, Vincent YF Tan, Animashree Anandkumar, and Alan S Willsky. Learning latent tree graphical models. *Journal of Machine Learning Research*, 12:1771–1812, 2011.

Tom Claassen, Joris Mooij, and Tom Heskes. Learning sparse causal models is not np-hard. *arXiv preprint arXiv:1309.6824*, 2013.

Diego Colombo, Marloes H Maathuis, Markus Kalisch, and Thomas S Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pp. 294–321, 2012.

Mathias Drton, Shaowei Lin, Luca Weihs, and Piotr Zwiernik. Marginal likelihood and model selection for gaussian latent tree and forest models. 2017.

Patrik O Hoyer, Shohei Shimizu, Antti J Kerminen, and Markus Palviainen. Estimation of causal effects using linear non-gaussian causal models with hidden variables. *International Journal of Approximate Reasoning*, 49(2):362–378, 2008.

Biwei Huang, Charles Jia Han Low, Feng Xie, Clark Glymour, and Kun Zhang. Latent hierarchical causal structure discovery with rank constraints. *arXiv preprint arXiv:2210.01798*, 2022.

Karl G Jöreskog. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2):183–202, 1969.

Bohdan Kivva, Goutham Rajendran, Pradeep Ravikumar, and Bryon Aragam. Learning latent causal graphs via mixture oracles. *Advances in Neural Information Processing Systems*, 34:18087–18101, 2021.

Lingjing Kong, Biwei Huang, Feng Xie, Eric Xing, Yuejie Chi, and Kun Zhang. Identification of nonlinear latent hierarchical models. *arXiv preprint arXiv:2306.07916*, 2023.

Erich Kummerfeld and Joseph Ramsey. Causal clustering for 1-factor measurement models. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1655–1664, 2016.

Linda K Muthén and Bengt Muthén. *Mplus user's guide: Statistical analysis with latent variables, user's guide*. Muthén & Muthén, 2017.

Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.

Yves Rosseel. lavaan: An r package for structural equation modeling. *Journal of statistical software*, 48:1–36, 2012.

Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.

Shohei Shimizu, Patrik O Hoyer, and Aapo Hyvärinen. Estimation of linear non-gaussian acyclic models for latent factors. *Neurocomputing*, 72(7-9):2024–2027, 2009.

Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvarinen, Yoshinobu Kawahara, Takashi Washio, Patrik O Hoyer, Kenneth Bollen, and Patrik Hoyer. Directlingam: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research-JMLR*, 12(Apr):1225–1248, 2011.

Ricardo Silva, Richard Scheines, Clark Glymour, Peter Spirtes, and David Maxwell Chickering. Learning the structure of linear latent variable models. *Journal of Machine Learning Research*, 7 (2), 2006.

Charles Spearman. Pearson's contribution to the theory of two factors. *British Journal of Psychology*, 19(1):95, 1928.

Peter Spirtes and Clark Glymour. An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, 9(1):62–72, 1991.

Peter Spirtes, Christopher Meek, and Thomas Richardson. Causal inference in the presence of latent variables and selection bias. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, pp. 499–506, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1558603859.

Chandler Squires, Annie Yun, Eshaan Nichani, Raj Agrawal, and Caroline Uhler. Causal structure discovery between clusters of nodes induced by latent factors. In *Conference on Causal Learning and Reasoning*, pp. 669–687. PMLR, 2022.

Seth Sullivant, Kelli Talaska, and Jan Draisma. Trek separation for gaussian graphical models. 2010.

Tatsuya Tashiro, Shohei Shimizu, Aapo Hyvärinen, and Takashi Washio. Parcelingam: A causal ordering method robust against latent confounders. *Neural computation*, 26(1):57–83, 2014.

Feng Xie, Ruichu Cai, Biwei Huang, Clark Glymour, Zhifeng Hao, and Kun Zhang. Generalized independent noise condition for estimating latent variable causal graphs. *Advances in neural information processing systems*, 33:14891–14902, 2020.

Feng Xie, Biwei Huang, Zhengming Chen, Yangbo He, Zhi Geng, and Kun Zhang. Identification of linear non-gaussian latent hierarchical structure. In *International Conference on Machine Learning*, pp. 24370–24387. PMLR, 2022.

Feng Xie, Biwei Huang, Zhengming Chen, Ruichu Cai, Clark Glymour, Zhi Geng, and Kun Zhang. Generalized independent noise condition for estimating causal structure with latent variables. *arXiv preprint arXiv:2308.06718*, 2023.

Nevin L Zhang. Hierarchical latent class models for cluster analysis. *The Journal of Machine Learning Research*, 5:697–723, 2004.

Qinyi Zhang, Sarah Filippi, Arthur Gretton, and Dino Sejdinovic. Large-scale kernel methods for independence testing. *Statistics and Computing*, 28:113–130, 2018.