Learning Unknown Intervention Targets in Structural Causal Models from Heterogeneous Data

Yuqin Yang[*]
Georgia Institute of Technology

Saber Salehkaleybar[*]
École Polytechnique Fédérale de Lausanne

Negar Kiyavash
École Polytechnique Fédérale de Lausanne

Abstract

We study the problem of identifying the unknown intervention targets in structural causal models where we have access to heterogeneous data collected from multiple environments. The unknown intervention targets are the set of endogenous variables whose corresponding exogenous noises change across the environments. We propose a two-phase approach which in the first phase recovers the exogenous noises corresponding to unknown intervention targets whose distributions have changed across environments. In the second phase, the recovered noises are matched with the corresponding endogenous variables. For the recovery phase, we provide sufficient conditions for learning these exogenous noises up to some component-wise invertible transformation. For the matching phase, under the causal sufficiency assumption, we show that the proposed method uniquely identifies the intervention targets. In the presence of latent confounders, the intervention targets among the observed variables cannot be determined uniquely. We provide a candidate intervention target set which is a superset of the true intervention targets. Our approach improves upon the state of the art as the returned candidate set is always a subset of the target set returned by previous work. Moreover, we do not require restrictive assumptions such as linearity of the causal model or performing invariance tests to learn whether a distribution is changing across environments which could be highly sample inefficient. Our experimental results show the effectiveness of our proposed algorithm in practice.

1 Introduction

Causal relationships among a set of variables in a system can be modeled by a structural causal model (SCM) where each variable is a function of its direct causes and some exogenous noise. An intervention on a variable can be considered as modifying its causal mechanism, i.e., changing the conditional probability distribution of the intervened variable given its direct causes. In randomized control trials, randomized interventions on a target variable are utilized to estimate the causal effect of the target. However, in some applications, we may not have full control in terms of which variables are intervened on. For instance, in recovering causal protein-signaling networks from single-cell data [SPP+05, NSMV17], drugs are injected into cells to inhibit or activate some signaling proteins, and gene expression levels are measured. In these experiments, the intervention targets are unknown. Moreover, in some cases, an intervention is done by an unknown source and we must locate the source of the intervention in the system. As an example, microservices systems in cloud clusters are vulnerable to faults such as equipment failures or adversarial attacks. It is crucial to locate the root cause of faulty operation in the system by identifying the source of fault/intervention [AGM+21, BMBJ22]. In these examples, the collected data is often heterogeneous and is gathered

[*]Equal contribution. Correspondence to: yuqinyang@gatech.edu.

Accepted to the NeurIPS 2023 Workshop on Causal Representation Learning.

from multiple domains/environments where the causal mechanisms of some of the variables are changing across the environments.

In this paper, we consider the problem of learning the unknown intervention targets from a collection of interventional distributions obtained from multiple environments. This problem is closely related to learning an equivalence class of all causal graphs consistent with the collected interventional data. The latter problem has been studied in several work, and some of the proposed methods also provide information about the locations of intervention targets as a byproduct of the returned equivalence class. These previous methods have several drawbacks such as being limited to linear systems, requiring a huge number of conditional and invariance tests, or lacking the ability to handle latent confounders in the systems.

We propose Locating Intervention Target (LIT) algorithm which returns the observed variables that are intervention targets. LIT has two main phases: the recovery phase and the matching phase. In the recovery phase, through a contrastive-learning approach, the exogenous noises corresponding to intervention targets are recovered up to some permutation and component-wise invertible transformation[2]. In the matching phase, the recovered exogenous noises are matched to their corresponding observed variables (if any)[3] by performing conditional independence (CI) tests. The main contributions of this paper are:

- For the recovery phase, we provide identifiability results for recovering the exogenous noises whose distributions change across the environments. In particular, in nonlinear causal models with exogenous noises belonging to an exponential family, the recovery is possible under some mild invertibility assumption (Assumption 1(a)) for causally sufficient systems (i.e., there are no latent confounders). For systems with latent variables, under some further assumptions (Assumption 1(b)), we show that the recovery is still possible (Proposition 1).
- For the matching phase, we prove that LIT algorithm recovers the true intervention targets for causally sufficient systems using the recovered exogenous noises (Theorem 1). LIT algorithm requires only quadratic number of CI tests while previous work [JKSB20] requires exponential number of CI/invariance tests with respect to number of variables in the system. In the presence of latent confounders, we show that LIT algorithm returns a superset of true intervention targets and present a graphical characterization for the recovery output (Theorem 2). Unlike previous work, LIT algorithm allows for latent confounders to change across environments. Moreover, for the setting studied in the literature (i.e., all latent confounders are not changing across environments), our recovery output is more informative than the state-of-the-art.
- Experimental results show that LIT outperforms previous work in recovering intervention targets in the presence of latent confounders or when the underlying SCM is nonlinear.

2 Preliminaries

In this section, we present the notations used in the paper as well as some necessary background. Upper case letters denote random variables and bold letters indicate sets of random variables. For ease of notation, we also denote the vectorized form of a set of random variables by bold letters. We show the cardinality of set $\mathbf{X}$ by $|\mathbf{X}|$. We also denote the set $\{1, \cdots, n\}$ by $[n]$.

Structural Causal Models. A structural causal model (SCM) $\mathcal{M}$ is a 4-tuple $\langle \mathbf{N}, \mathbf{X}, \mathcal{F}, P(\mathbf{N}) \rangle$ where $\mathbf{N}$ is the set of exogenous noises and $\mathbf{X}$ is the set of endogenous variables. $\mathcal{F}$ represents a collection of functions $\mathcal{F} = \{f_i\}$ such that each endogenous variable $X_i \in \mathbf{X}$ is determined by $X_i := f_i(PA_i, N_i)$ where $PA_i \subseteq \mathbf{X}$ is the set of parents of X_i and $N_i \in \mathbf{N}$ is its corresponding exogenous noise. It is assumed that $\{N_i\}$ are jointly independent. In a given SCM, we may only observe a subset of endogenous variables. Thus, we partition $\mathbf{X}$ into two disjoint subsets $\mathbf{O}$ and $\mathbf{L}$, where $\mathbf{O}$ is the set of observed and $\mathbf{L}$ is the set of latent variables. Under the causal sufficiency assumption, we observe all the endogenous variables, i.e., $\mathbf{L} = \emptyset$.

The graph G of an SCM is constructed by considering one vertex for each X_i and drawing directed edges from each parent in PA_i to X_j. We assume that the graph G is a directed acyclic graph (DAG),

[2] In the paper, whenever we say that some exogenous noises can be recovered, it means that they are recovered up to some permutation and component-wise invertible transformation.

[3] It is noteworthy that a recovered exogenous noise may correspond to a latent variable. In that case, it should not be matched with any observed variable.

i.e., it contains no directed cycle. We say X_j is an ancestor of X_i if there exists a direct path from X_j to X_i. In graph G, we denote the set of ancestors and children of X_i by $An_G(X_i)$ and $Ch_G(X_i)$, respectively. We also consider each variable $X_i \in \mathbf{X}$ as its own ancestor. The CI relations can be read from the causal graph using a graphical criterion known as *d-separation* [Pea88]. For disjoint subsets of variables $\mathbf{U}, \mathbf{V}, \mathbf{W}$, we denote the CI relation of $\mathbf{U}$ from $\mathbf{V}$ given $\mathbf{W}$ by $\mathbf{U} \perp\!\!\!\perp \mathbf{V}|\mathbf{W}$. The analogous d-separation statement, $\mathbf{U}$ is d-separated from $\mathbf{V}$ given $\mathbf{W}$ in graph G, is written as $(\mathbf{U} \perp\!\!\!\perp \mathbf{V}|\mathbf{W})_G$. In the presence of latent confounders, the causal relationships are often represented by a maximal ancestral graph (MAG). See [RS02] for the definitions of MAGs and inducing paths.

Soft Intervention. We consider *soft* interventions on a subset of variables such as $\mathbf{W} \subseteq \mathbf{X}$ of the form obtained by replacing structural assignment $X_i := f_i(PA_i, N_i)$ with $X_i := f_i(PA_i, N_i')$ for all $X_i \in \mathbf{W}$. N_i' is the new exogenous noise corresponding to X_i. Note that in the definition of soft intervention, neither the set of parents nor causal mechanisms f_is change. In some applications, this operation is more realistic than *hard* interventions, where intervened variables are forced to take a fixed value [VSST22]. For instance, in molecular biology, the effect of added chemicals to a cell cannot be set to some constant value [EM07], or in control theory, for the task of system identification [Lju98], a mathematical model describing the underlying dynamical system is identified by applying certain inputs without changing the dynamics of the system.

3 Methodology

3.1 Problem definition

We consider a multi-environment setting comprised of D environments $\mathcal{E} = \{E_1, ..., E_D\}$. The underlying causal DAG and the functional mechanism for generating the variables from their parents remain the same across all environments while the distributions of exogenous noises may vary due to some unknown soft interventions. In particular, we have access to a collection of joint distributions over $\mathbf{O}$, $\mathcal{P} = \{p_1(\mathbf{O}), \cdots, p_D(\mathbf{O})\}$ from D environments. We also denote $p_i(\mathbf{N})$ as the joint distribution over the set of exogenous noises $\mathbf{N}$ in environment E_i. Let $\mathbf{T}$ be the set of variables whose exogenous noises are changing across environments, i.e., $\mathbf{T} := \{X_i | \exists d, d' \in [D], p_d(N_i) \neq p_{d'}(N_i), 1 \leq i \leq n\}$. These are the variables that are intervened on by some external stimuli and we seek to learn them. Let $\mathbf{N_T} := \{N_i | X_i \in \mathbf{T}\}$ be the set of exogenous noises whose distributions are changing across the environments, and $\mathbf{T_O} = \mathbf{T} \cap \mathbf{O}$ be the set of observed variables that are intervened on. Similarly, denote the set of intervention targets in the latent part by $\mathbf{T_L} = \mathbf{T} \cap \mathbf{L}$. Note that under causal sufficiency, $\mathbf{T_O} = \mathbf{T}$. Our goal is to locate interventions, i.e., recover the unknown observable targets of interventions $\mathbf{T_O}$ from merely the observational distributions $\mathcal{P}$ over the multiple environments.

In the following, we present our method for learning the intervention targets, which has two main phases: the recovery phase and the matching phase. In Section 3.2, we present the recovery phase, which is to recover the set of exogenous noises whose distributions are changing across the environments (up to some permutation and component-wise invertible transformations). Next, we present the matching phase in Section 3.3, where we match the recovered noises with the corresponding variables in $\mathbf{X}$ in order to learn $\mathbf{T_O}$.

3.2 Recovery phase

For a given SCM $\mathcal{M}$, due to the assumption that the causal graph is a DAG, each observed variable $X_i \in \mathbf{X}$ can be written as $X_i = g_i(\mathbf{N})$ where function g_i only depends on exogenous noises corresponding to the ancestors of X_i. We collect all these equations in the vector form $\mathbf{X} = \mathbf{g}_\mathcal{M}(\mathbf{N})$ where $\mathbf{g}_\mathcal{M} : \mathbb{R}^n \to \mathbb{R}^n$ and $n = |\mathbf{X}|$ is the number of variables in the system. We call function $\mathbf{g}_\mathcal{M}$, the "mixing function" of SCM $\mathcal{M}$.

As we have access to a collection of distributions in $\mathcal{P}$, we will exploit the heterogeneity in data to recover the exogenous noises in $\mathbf{N_T}$. Specifically, we utilize a contrastive learning approach. We observe auxiliary variable U indicating the index of the environment, and train a nonlinear regression model with universal approximation capability[4] for the supervised learning task (Please refer to

[4]Universal approximation capability refers to the ability to approximate any Borel measurable function to any desired degree of accuracy. See [HSW89] for more detail.

Appendix A.1 for more details). We consider a specific exponential family for the distributions of the exogenous noises (see Appendix A.1 for the definition of the exponential family and the assumptions on it). To ensure that the noises in $\mathbf{N_T}$ can be recovered, we require the following assumption:

Assumption 1. *For a given SCM $\mathcal{M}$, we assume that either: (a) the corresponding mixing function $\mathbf{g}_{\mathcal{M}}$ is invertible, or (b) there exists an invertible function $\tilde{\mathbf{g}} : \mathbb{R}^{|\mathbf{O}|} \to \mathbb{R}^{|\mathbf{O}|}$ such that $\tilde{\mathbf{g}}(\mathbf{O}) = (\mathbf{N_T}, \mathbf{V})$ where $\mathbf{V} \in \mathbb{R}^{|\mathbf{O}|-|\mathbf{T}|}$ is a random vector satisfying $\mathbf{V} \perp\!\!\!\perp U$ and $\mathbf{N_T} \perp\!\!\!\perp \mathbf{V}|U$.*

Assumption 1(a) is standard in contrastive learning algorithms under causal sufficiency assumption. It is satisfied for all acyclic linear SCMs and nonlinear additive noise models. To extend the results to the latent confounder setting, we added Assumption 1(b). In particular, vector $\mathbf{V}$ is the recovered part that is invariant across the environments and is independent of $\mathbf{N_T}$ given the index of the environment. It corresponds to a function of exogenous noises whose distributions do not change across the environments.

Proposition 1. *Assume that $\min(D-1, |\mathbf{O}|) \geq |\mathbf{T}|$ (Recall that D is the number of environments and $\mathbf{O}$ and $\mathbf{L}$ are the set of observed and latent variables in the system, respectively). By utilizing the contrastive-learning approach, the exogenous noises in $\mathbf{N_T}$ can be recovered up to some permutation and component-wise strictly monotonic transformations with measure one in the following two settings: 1) $\mathbf{L} = \emptyset$ under Assumption 1(a). 2) $\mathbf{L} \neq \emptyset$ under Assumption 1(b).*

Please refer to Appendix A for a detailed description of the contrastive learning approach and extra discussion about when Assumption 1 is satisfied.

3.3 Matching phase

Throughout the matching phase, we assume that the exogenous noises in $\mathbf{N_T}$ are recovered up to some permutation and component-wise invertible transformations. Denote the recovered noise corresponding to the noise $N_i \in \mathbf{N_T}$ as $\tilde{N}_i$ (i.e., $\tilde{N}_i$ is an invertible transformation of N_i), and denote the collection of all the recovered noises as $\tilde{\mathbf{N}}_{\mathbf{T}}$. Note that we cannot learn the correspondence of the recovered noises to the true noises due to permutation indeterminacy according to Proposition 1. In fact, the goal of the matching phase is to recover the mapping between the recovered noises in $\tilde{\mathbf{N}}_{\mathbf{T}}$ and the corresponding exogenous noises in $\mathbf{N_T}$.[5]

We show how to use $\tilde{\mathbf{N}}_{\mathbf{T}}$ from the recovery phase to recover the intervention target. In Section 3.3.1, we define a new notion of faithfulness assumption called T-faithfulness based on the *augmented graph*. In Section 3.3.2, we study causally sufficient models. We present the conditions and the algorithm for recovering the intervention target, and show that the intervention target set $\mathbf{T}$ can be uniquely identified with quadratic number of CI tests. We also study the model with latent confounders in Appendix B.2.

3.3.1 T-faithfulness assumption

For a given SCM $\mathcal{M}$ with the causal graph G and intervention targets $\mathbf{T}$, we construct an augmented graph $G_{\mathbf{T}}$ as follows. For each variable $X_i \in \mathbf{T_O}$ with corresponding exogenous noise N_i (recall that $\mathbf{T_O}$ and $\mathbf{T_L}$ are the sets of intervention targets in the observed and latent variables, respectively), we add vertex N_i and edge $N_i \to X_i$ to $G_{\mathbf{T}}$. Further, for each latent confounder $X_l \in \mathbf{T_L}$ with corresponding recovered noise N_l, we replace X_l with N_l since X_l can be recovered up to an invertible transformation. We denote the set of noises corresponding to the variables in $\mathbf{T_L}$ by $\mathbf{N_{T_L}}$. Following this construction, variables in $G_{\mathbf{T}}$ consist of all changing noises in $\mathbf{N_T}$, and all variables in $\mathbf{X} \setminus \mathbf{T_L}$.

It can be shown that the joint distribution $p(\mathbf{X} \setminus \mathbf{T_L}, \mathbf{N_T})$ satisfies Markov property with respect to graph $G_{\mathbf{T}}$ (see Appendix B.1). However, in order to infer the graphical properties of the augmented graph from only observed variables $\mathbf{O}$ and the recovered noises $\tilde{\mathbf{N}}_{\mathbf{T}}$, we need a form of faithfulness.

Assumption 2 (T-faithfulness)**.** *The model is T-faithful to the augmented graph $G_{\mathbf{T}}$, in the sense that for any noise $N_i \in \mathbf{N_T}$, observed variable $X_k \in \mathbf{O}$, and disjoint sets $\mathbf{W}_1 \subseteq \mathbf{O} \setminus \{X_k\}$,*

[5] Due to the permutation indeterminacy, there exists a one-to-one mapping σ that maps each noise in $\mathbf{N_T}$ to a distinct noise in $\tilde{\mathbf{N}}_{\mathbf{T}}$. For notation simplicity, we denote $\sigma(N_i) \in \tilde{\mathbf{N}}_{\mathbf{T}}$ as $\tilde{N}_i$ for each $N_i \in \mathbf{N_T}$. Since each noise $N_i \in \mathbf{N_T}$ corresponds to a variable $X_i \in \mathbf{T}$, the goal of the matching phase is to learn the inverse mapping σ^{-1} that maps the recovered noises to the variables in $\mathbf{T}$.

$\mathbf{W}_2 \subseteq \mathbf{N_T} \setminus \{N_i\}$*:* $(N_i \perp\!\!\!\perp X_k | \mathbf{W}_1, \mathbf{W}_2)_{G_\mathbf{T}}$ *if and only if* $\tilde{N}_i \perp\!\!\!\perp X_k | \mathbf{W}_1, \tilde{\mathbf{W}}_2$*, where* $\tilde{N}_i$ *is an arbitrary invertible transformation of* N_i*, and* $\tilde{\mathbf{W}}_2 = \{\tilde{N}_j | N_j \in \mathbf{W}_2\}$.

Assumption 2 implies that for any changing exogenous noise $N_i \in \mathbf{N_T}$ and observed variable X_j, the recovered noise $\tilde{N}_i$ is (marginally or conditionally) dependent on X_j if and only if N_i and X_j are d-connected in $G_\mathbf{T}$. Therefore, given observed variables and the recovered noises, we can construct the indicator set $\mathbf{I}_i := \{\tilde{N}_j \in \tilde{\mathbf{N}}_\mathbf{T} | \tilde{N}_j \not\perp\!\!\!\perp X_i\}$ for each variable X_i in $\mathbf{O}$, which is the set of recovered noises that are dependent of X_i. Under Assumption 2, the indicator set $\mathbf{I}_i$ corresponds to all the noises in $\mathbf{N_T}$ that are ancestors of X_i in $G_\mathbf{T}$, i.e., $An_{G_\mathbf{T}}(X_i) \cap \mathbf{N_T}$. Define $\mathcal{I}$ as the collection of sets $\{\mathbf{I}_i | i \in [n]\}$. In the following, we show how to identify the intervention targets by matching the recovered noises with the observed variables, based on the indicator sets and a limited number of extra CI tests.

3.3.2 Matching phase under causal sufficiency

For each variable X_i, define the the possible parent set $\mathbf{S}_i$ as the set of variables whose indicator set is a strict subset of $\mathbf{I}_i$, i.e., $\mathbf{S}_i := \{X_j | \mathbf{I}_j \subsetneq \mathbf{I}_i, 1 \leq j \leq n\}$. Define the *residual set* $\mathcal{N}_i$ as the set of noises in $\mathbf{I}_i$ that does not belong to any indicator set of the variables in $\mathbf{S}_i$, i.e. $\mathbf{I}_i \setminus \cup_{j: X_j \in \mathbf{S}_i} \mathbf{I}_j$. $\mathbf{S}_i$ includes a subset of the ancestors of X_i, and no descendants of X_i are included in $\mathbf{S}_i$. Further, $\mathcal{N}_i$ represent the noises in X_i that do not affect X_i through variables in $\mathbf{S}_i$. Under causal sufficiency assumption, $|\mathcal{N}_i|$ is either 0 or 1 (see Appendix C.3). The following proposition provides the conditions for checking whether a variable X_i is in the intervention target set or not given the indicator sets $\mathcal{I}$ and $\mathbf{S}_i$. Equipped with this proposition, we devise LIT Algorithm (see Algorithm 1) which recovers $\mathbf{T}$ under the causal sufficiency assumption.

Proposition 2. *Under causal sufficiency and Assumption 2, for each variable* X_i*, the following statements hold:*

(I) $X_i \notin \mathbf{T}$ *if the residual set is empty, i.e.,* $\mathcal{N}_i = \emptyset$.

(II) If $\mathcal{N}_i \neq \emptyset$ *and* $\mathbf{I}_i$ *is unique in* $\mathcal{I}$*, then* $X_i \in \mathbf{T}$.

(III) If $\mathcal{N}_i \neq \emptyset$ *and* $\mathbf{I}_i$ *is not unique, let* $\mathcal{X}_i = \{X_{i_1}, \cdots, X_{i_p}\}$*, for some* $p \geq 2$*, be the set of all variables with the same indicator sets as* X_i*, including* X_i *itself.*[6] *Suppose* $\mathcal{N}_i = \{\tilde{N}_l\}$ *for some* $\tilde{N}_l \in \tilde{\mathbf{N}}_\mathbf{T}$*. Then, the variable* X_{i_k} *satisfying the following condition is the only variable from* $\mathcal{X}_i$ *that is in* $\mathbf{T}$*, i.e.,* $X_{i_k} \in \mathbf{T}$ *if all other variables in* $\mathcal{X}_i$ *are independent of* $\tilde{N}_l$ *conditioned on* X_{i_k} *and* $\mathbf{S}_i$.

$$\tilde{N}_l \perp\!\!\!\perp X_{i_j} | \{X_{i_k}\} \cup \mathbf{S}_i, \textit{ for all } 1 \leq j \leq p, j \neq k. \tag{C1}$$

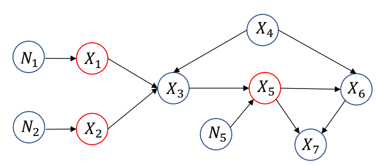

Figure 1: An example of SCM: Intervention targets $\mathbf{T} = \{X_1, X_2, X_5\}$ are shown by red circles.

Recall that one observed variable is in the intervention target set $\mathbf{T}$ if and only if its corresponding exogenous noise is recovered in $\tilde{\mathbf{N}}_\mathbf{T}$. The statement in (I) holds because if $X_i \in \mathbf{T}$, then $\tilde{N}_i$ cannot appear in $\mathbf{I}_j$ for any non-descendant X_j of X_i. When $\mathcal{N}_i \neq \emptyset$, this means that the only noise $\tilde{N}_l \in \mathcal{N}_i$ is either the exogenous noise of X_i, or the exogenous noise of some ancestor of X_i whose indicator set is the same as $\mathbf{I}_i$. We can then use conditions (II) and (III) to further distinguish between these two cases. If there are no other variables with the same indicator set (i.e., $\mathbf{I}_i$ is unique), then $X_i \in \mathbf{T}$. Otherwise, among the variables with the same indicator set, there is only one variable X_{i_k} that corresponds to $\tilde{N}_l$ and belongs to the intervention target set, and the rest are descendants of X_{i_k}. Herein, we use (C1) to find such X_{i_k}, as given X_{i_k} and $\mathbf{S}_i$, $\tilde{N}_l$ becomes independent from X_{i_j} for all $j \neq k$ under T-faithfulness assumption. The following example illustrates how the the conditions in Proposition 2 can be used to recover $\mathbf{T}$.

[6]As all the variables in $\mathcal{X}_i$ have the same indicator set, their corresponding possible parent sets $\mathbf{S}_{i_k}$ and residual sets $\mathcal{N}_{i_k}$ are also equal. In the statement of the proposition, we use $\mathbf{I}_i, \mathbf{S}_i, \mathcal{N}_i$ to denote the indicator set, possible parent set and residual set corresponding to any variable in $\{X_{i_1}, \cdots, X_{i_p}\}$.

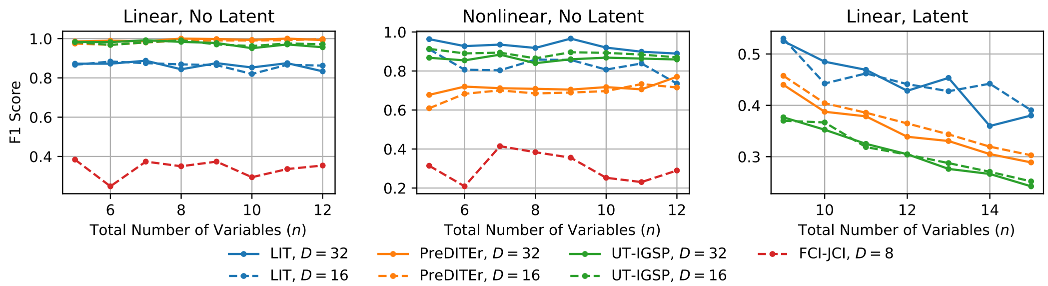

Figure 2: Comparison of LIT algorithm with previous work in locating intervention targets.

Example 1. *Figure 1 depicts the augmented graph of an SCM in which* $\mathbf{T} = \{X_1, X_2, X_5\}$ *(indicated by red circles). In the recovery phase, we recover three noises* $\tilde{N}_1, \tilde{N}_2, \tilde{N}_5$*, which are invertible translations of* N_1*,* N_2*,* N_5*, respectively. Note that we do not know the correspondence of the noises to the variables as there are permutation indeterminacy. The indicator sets for all the variables are:* $\mathbf{I}_1 = \{\tilde{N}_1\}$, $\mathbf{I}_2 = \{\tilde{N}_2\}$, $\mathbf{I}_3 = \{\tilde{N}_1, \tilde{N}_2\}$, $\mathbf{I}_4 = \emptyset$, $\mathbf{I}_5 = \mathbf{I}_6 = \mathbf{I}_7 = \{\tilde{N}_1, \tilde{N}_2, \tilde{N}_5\}$*. For* X_1 *and* X_2*, the condition in (II) is satisfied. Thus, they are in* $\mathbf{T}$*. As for* X_3 *and* X_4*, the condition in (I) holds and therefore they are not in* $\mathbf{T}$*. For the variables in* $\{X_5, X_6, X_7\}$*, the condition in (III) holds and the only variable satisfying the condition in* (C1) *is* X_5 *as* $\tilde{N}_5$ *is independent of* X_6 *and* X_7 *given* $\mathbf{S}_5 \cup \{X_5\}$ *where* $\mathbf{S}_5 = \{X_1, X_2, X_3, X_4\}$*.*

Based on Proposition 2, we propose LIT algorithm (see Algorithm 1) which returns a *candidate intervention set* $\mathbf{K}$. Specifically, we first check if a variable can be added to or excluded from $\mathbf{K}$ according to (I) and (II). We then partition the remaining variables in $\mathbf{U}$ into disjoint subsets (which correspond to the collection of all $\mathcal{X}_i$ in (III)), and find the candidate in each set (denoted by $\mathbf{K}_{\mathbf{U}_i}$) using condition (III). Note that LIT algorithm only requires quadratic number of CI tests: $O(n|\mathbf{T}|)$ for constructing the indicator set, and at most $O(n|\mathbf{T}|)$ for checking (C1). This is a significant reduction from the exponential number of independence/invariance tests with respect to n in the literature [JKSB20, MMC20].

Theorem 1. *Under causal sufficiency and T-faithfulness assumption (Assumption 2), Algorithm 1 uniquely identifies the intervention target set* $\mathbf{T}$*, i.e.,* $\mathbf{K} = \mathbf{T}$*.*

Algorithm 1: LIT algorithm

1 Obtain $\tilde{\mathbf{N}}_{\mathbf{T}}$ and $\mathcal{I}$; $\mathbf{U} \leftarrow \mathbf{X}$; $\mathbf{K} \leftarrow \emptyset$;
2 **for** $X_i \in \mathbf{X}$ **do**
3   **if** $\mathcal{N}_i = \emptyset$ **then** $\mathbf{U} \leftarrow \mathbf{U} \backslash \{X_i\}$;
4   **else if** (A) *holds* **then** $\mathbf{U} \leftarrow \mathbf{U} \backslash \{X_i\}$;
      // only with latent confounder
5   **else if** $\mathbf{I}_i$ *is unique* **then** $\mathbf{K} \leftarrow \mathbf{K} \cup \{X_i\}$, $\mathbf{U} \leftarrow \mathbf{U} \backslash \{X_i\}$;
6 Partition $\mathbf{U}$ to disjoint subsets $\mathbf{U}_1, \cdots, \mathbf{U}_r$ according to the indicator sets;
7 **for** $\mathbf{U}_i \in \{\mathbf{U}_1, \cdots, \mathbf{U}_r\}$ **do**
8     $\mathbf{K}_{\mathbf{U}_i} \leftarrow$ Find $X_{i_k} \in \mathbf{U}_i$ satisfying (C1) under causal sufficiency (resp. remove the subset of variables in $\mathbf{U}_i$ satisfying (C2) in the presence of latent confounders);
9     $\mathbf{K} \leftarrow \mathbf{K} \cup \mathbf{K}_{\mathbf{U}_i}$;
10 **return** $\mathbf{K}$

Lastly, we extend our results from the causally sufficient case to the case where latent confounders are present in Appendix B.2. Unlike the former case, the set of observed intervention targets, $\mathbf{T}_{\mathbf{O}}$, is not always uniquely identifiable. However, we provide conditions (Proposition 3) for finding variables that does not belong to $\mathbf{T}_{\mathbf{O}}$, and update the LIT algorithm based on the new conditions (lines 4 and 8). We then give graphical characterization of the recovered candidate intervention target set $\mathbf{K}$ through auxiliary graph in Definition 2. We show that $\mathbf{K}$ is always a superset of $\mathbf{T}_{\mathbf{O}}$, and it is always a subset of the recovery output of existing algorithms (e.g., [JKSB20]). See Appendix B.2 for more details.

4 Experiments

We evaluated the performance of LIT algorithm on randomly generated models. We considered the following three settings, with different numbers of environments $D = \{16, 32\}$ for data generation: (1) Linear Gaussian model under causal sufficiency assumption; (2) Nonlinear model under causal sufficiency assumption; (3) Linear Gaussian model in the presence of latent confounders. We considered the following methods in our empirical studies: 1- LIT (our proposed method); 2- PreDITEr algorithm [VSST22] which allows for latent confounders (that are not in $\mathbf{T}$) but assumes the model to be linear; 3- UT-IGSP algorithm [SWU20] which works for both linear and nonlinear SCMs under merely causal sufficiency; 4- FCI-JCI123 algorithm in [MMC20] which allows for both latent confounders and nonlinearity in the model.

We repeated each setting for 40 times, and reported the average F1-score in recovering $\mathbf{T_O}$ for each setting. The results are shown in Figure 2. Note that FCI-JCI is executable only under the first two settings with $D = 8$ due to huge run times. In the first setting, as we expected, PreDITEr has the best performance as it is designed specifically for linear Gaussian SCMs. UT-IGSP and LIT algorithm have decent performances, while FCI-JCI123 does not perform well. In the second setting, LIT and UT-IGSP can both recover the intervention targets with high accuracy. On the contrary, the performance of PreDITEr becomes worse as it is not designed for nonlinear models. Finally, in the third setting, in the presence of latent confounders, the performance of UT-IGSP becomes much worse because it cannot handle any latent confounders. Meanwhile, LIT outperforms PreDITEr and UT-IGSP for various numbers of variables in the system. Note that the best F1-score that can be achieved by any algorithm is strictly less than one as there are intervention targets that can never be recovered. Lastly, we note that LIT algorithm significantly reduces the number of CI tests performed: LIT algorithm takes at most 80 CI tests while PreDITER requires up to 30000 PDE estimates.

5 Conclusions

We addressed the problem of identifying unknown intervention targets in a multi-environment setting. Our two-phase algorithm recovers the exogenous noises and matches them with corresponding endogenous variables. Under the causal sufficiency assumption, our algorithm identifies uniquely the intervention targets. In the presence of latent confounders, we provided a candidate intervention target set which is more informative than previous work. Experiment results support the advantages of the proposed algorithm in identifying intervention targets. As a future work, in the recovery phase, it is an interesting direction to strengthen the identifiability result of non-linear SCM with latent variables which would broaden the applicability of our method to more complex systems.

Acknowledgments and Disclosure of Funding

This research was in part supported by the Swiss National Science Foundation under NCCR Automation, grant agreement 51NF40_180545 and Swiss SNF project 200021_204355 /1.

References

[AGM+21] Pooja Aggarwal, Ajay Gupta, Prateeti Mohapatra, Seema Nagar, Atri Mandal, Qing Wang, and Amit Paradkar. Localization of operational faults in cloud applications by mining causal dependencies in logs using golden signals. In *Service-Oriented Computing–ICSOC 2020 Workshops: AIOps, CFTIC, STRAPS, AI-PA, AI-IOTS, and Satellite Events, Dubai, United Arab Emirates, December 14–17, 2020, Proceedings*, pages 137–149. Springer, 2021.

[BMBJ22] Kailash Budhathoki, Lenon Minorics, Patrick Blöbaum, and Dominik Janzing. Causal structure-based root cause analysis of outliers. In *International Conference on Machine Learning*, pages 2357–2369. PMLR, 2022.

[Dar53] Georges Darmois. Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 21(1/2):2–8, 1953.

[EM07] Daniel Eaton and Kevin Murphy. Exact bayesian structure learning from uncertain interventions. In *Artificial intelligence and statistics*, pages 107–114. PMLR, 2007.

[GP90] Dan Geiger and Judea Pearl. On the logic of causal models. In *Machine Intelligence and Pattern Recognition*, volume 9, pages 3–14. Elsevier, 1990.

[HM16] Aapo Hyvarinen and Hiroshi Morioka. Unsupervised feature extraction by time-contrastive learning and nonlinear ica. *Advances in neural information processing systems*, 29, 2016.

[HST19] Aapo Hyvarinen, Hiroaki Sasaki, and Richard Turner. Nonlinear ica using auxiliary variables and generalized contrastive learning. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 859–868. PMLR, 2019.

[HSW89] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[JKSB20] Amin Jaber, Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Causal discovery from soft interventions with unknown targets: Characterization and learning. *Advances in neural information processing systems*, 33:9551–9561, 2020.

[KKMH20] Ilyes Khemakhem, Diederik Kingma, Ricardo Monti, and Aapo Hyvarinen. Variational autoencoders and nonlinear ica: A unifying framework. In *International Conference on Artificial Intelligence and Statistics*, pages 2207–2217. PMLR, 2020.

[Lju98] Lennart Ljung. *System identification*. Springer, 1998.

[MMC20] Joris M Mooij, Sara Magliacane, and Tom Claassen. Joint causal inference from multiple contexts. *The Journal of Machine Learning Research*, 21(1):3919–4026, 2020.

[NSMV17] Robert Osazuwa Ness, Karen Sachs, Parag Mallick, and Olga Vitek. A bayesian active learning experimental design for inferring signaling networks. In *Research in Computational Molecular Biology: 21st Annual International Conference, RECOMB 2017, Hong Kong, China, May 3-7, 2017, Proceedings 21*, pages 134–156. Springer, 2017.

[Pea88] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.

[Pea09] Judea Pearl. *Causality*. Cambridge university press, 2009.

[RS02] Thomas Richardson and Peter Spirtes. Ancestral graph markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

[SGSH00] Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000.

[Ski53] Viktor P Skitovitch. On a property of the normal distribution. *DAN SSSR*, 89:217–219, 1953.

[SPP$^+$05] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

[SRK20] Peter Sorrenson, Carsten Rother, and Ullrich Köthe. Disentanglement by nonlinear ica with general incompressible-flow networks (gin). In *International Conference on Learning Representations*, 2020.

[SWU20] Chandler Squires, Yuhao Wang, and Caroline Uhler. Permutation-based causal structure learning with unknown intervention targets. In *Conference on Uncertainty in Artificial Intelligence*, pages 1039–1048. PMLR, 2020.

[VSST22] Burak Varici, Karthikeyan Shanmugam, Prasanna Sattigeri, and Ali Tajer. Intervention target estimation in the presence of latent variables. In *Uncertainty in Artificial Intelligence*, pages 2013–2023. PMLR, 2022.

A Further discussion on the recovery phase

A.1 Detailed description about contrastive learning approach and nonlinear ICA

Nonlinear ICA refers to an instance of unsupervised learning, where the goal is to learn the independent components/features that generate multi-dimensional observed data. In particular, suppose that $\mathbf{X} = (X_1, \cdots, X_n)$, is a n-dimensional vector that is generated from n independent components $\mathbf{N} = (N_1, \cdots, N_n)$. Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be a smooth and invertible function transforming the latent components (aka sources) to the observed data, i.e., $\mathbf{X} = \mathbf{g}(\mathbf{N})$. Function $\mathbf{g}$ is called the "mixing function". The goal in nonlinear ICA is to recover the inverse function $\mathbf{g}^{-1}$ and also the latent components in $\mathbf{N}$.

We briefly describe the general approach in recent advances in nonlinear ICA [HST19, KKMH20, SRK20] for the case where no additional assumptions are made about the class of mixing functions. The main idea is to exploit non-stationarity in the data to recover the independent components. In particular, each component N_i depends on some auxiliary variable U and it is independent of other components given U, i.e., $\log p(\mathbf{N}|U) = \sum_i q_i(N_i|U)$, where q_is are some functions. The auxiliary variable U could be an index of a time segment or the index of some environments where we obtained samples from variables in $\mathbf{X}$. In this formulation, the distributions of components can change across the environments or time segments. It is often assumed that the distribution of each N_i given U is a member of the exponential family.

Definition 1. *A random variable N_i belongs to the exponential family of order one given a random variable U if its conditional probability distribution function (pdf) can be written as: $p(N_i|U) = \frac{Q_i(N_i)}{Z_i(U)} \exp\left(\lambda_i(U)\tilde{q}_i(N_i)\right)$ where Q_i, Z_i, λ_is, and $\tilde{q}_i$s are some scalar-valued functions.*

Example 2. *For $\tilde{q}_i(N_i) = -N_i^2/2$, and $Q_i(N_i) = 1$, the above conditional pdf reduces to Gaussian whose variance is changing across the environments.*

The general approach to exploit non-stationarity in data is to use contrastive learning to transform the unsupervised learning problem in nonlinear ICA to a supervised learning task. Specifically, a classifier is trained to discriminate samples of a real dataset from their randomized version, i.e., $\tilde{\mathbf{X}} = (\mathbf{X}, U)$ versus $\tilde{\mathbf{X}}' = (\mathbf{X}, U')$, where U' is drawn randomly from the distribution of U, which in practice can be obtained by randomized permutations of the samples of U.

In this approach, a nonlinear regression model is trained with the following form: $r(\mathbf{X}, U) = \mathbf{h}(\mathbf{X})^T \mathbf{v}(U) + a(\mathbf{X}) + b(U)$ where $\mathbf{h}(\mathbf{X}) : \mathbb{R}^n \to \mathbb{R}^n$, $\mathbf{v}(U) : \mathbb{R} \to \mathbb{R}^n$, and a, b are some scalar-valued functions. The model classifies a sample coming from the real data set with probability $1/(1 + \exp(-r(\mathbf{X}, U)))$. It has been shown in several work such as in [HST19, KKMH20] that if all the components are changed enough across the environments, then the independent components can be recovered from $\mathbf{h}(\mathbf{X})$ up to some permutation and component-wise nonlinear transformation.

How constrastive learning approach above is applied in our work As we have access to a collection of distributions in $\mathcal{P}$, we will exploit the heterogeneity in data to recover the exogenous noises in $\mathbf{N_T}$. Let auxiliary variable U denote the index of the environment and assume that the exogenous noises belong to the exponential family in Definition 1. Moreover, assume that λ_is corresponding to any $N_i \in \mathbf{N_T}$ are randomly generated across the environments and $\tilde{q}_i(N_i)$s are strictly monotonic functions of $|N_i|$. We utilize a contrastive learning approach similar to what we discussed above. We train the nonlinear regression model $r(\mathbf{O}, U) = \mathbf{h}(\mathbf{O})^T \mathbf{v}(U) + a(\mathbf{O}) + b(U)$ with universal approximation capability for the supervised learning task of discriminating $(\mathbf{O}, U)$ from $(\mathbf{O}, U')$ where $\mathbf{h}(\mathbf{O}) : \mathbb{R}^{|\mathbf{O}|} \to \mathbb{R}^{|\mathbf{O}|}$, $\mathbf{v}(U) : \mathbb{R} \to \mathbb{R}^{|\mathbf{O}|}$, and $a(\mathbf{O})$, $b(U)$ are some scalar-valued functions.

A.2 Additional discussion on Assumption 1

Assumption 1(a). We provide two examples on when Assumption 1(a) is satisfied. Note that Assumption 1(a) requires that $|\mathbf{O}| = n = |\mathbf{X}|$, which indicates that the model is causally sufficient.

First, for linear SCMs, the structural equations can be written in vector form as $\mathbf{X} = \mathbf{B}\mathbf{X} + \mathbf{N}$ where $\mathbf{B}$ is $n \times n$ matrix. Rewrite this equation as $\mathbf{X} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{N}$. Therefore, the corresponding mixing function is given by $(\mathbf{I} - \mathbf{B})^{-1}$ and the above assumption is satisfied if and only if $\mathbf{I} - \mathbf{B}$ is invertible, which is already satisfied when the causal graph is a DAG.

Second, for nonlinear SCMs, invertibility of the mixing function is satisfied when the model is an additive noise model, which can be written as $X_i = f_i(Pa_i) + N_i$. In this case, the inverse function that maps $\mathbf{N}$ to $\mathbf{O}$ can be constructed according to the following equation: $N_i = X_i - f_i(Pa_i)$. Note that the invertibility of the mixing function does not depend on f_i. This result can be generalized into the following remark:

Remark 1. *For a nonlinear SCM $X_i = f_i(Pa_i, N_i)$, Assumption 1(a) is satisfied if the model is acyclic, f_i is continuous, and the partial derivative $\partial f_i/\partial N_i$ is strictly negative or positive for all $X_i \in \mathbf{X}$ and any values of Pa_i.*

We note that the condition in Remark 3 can be satisfied when f_i in the data generating model is designed as a multi-layer perceptron with ReLU activation function, where all model weights are positive.

Assumption 1(b). For Assumption 1(b), we note that if the SCM is a linear non-Gaussian model, i.e., linear SCM with non-Gaussian exogenous noises, and latent confounders are present, then Assumption 1(b) implies that: (1) All intervention targets must be observed variables, i.e., $\mathbf{T} \subseteq \mathbf{O}$; and (2) Latent variables cannot have children in $\mathbf{T}$. In other words, $\mathbf{T}$ can only include observed variables that do not have latent parents. Please see Appendix C.2 for the proof. However, we observed experimentally that if the SCM is linear Gaussian within each environment, then we can recover the noises in $\mathbf{N_T}$ while allowing latent confounders to be in $\mathbf{T}$ using linear ICA methods.

B Further discussions on the matching phase

B.1 Markov property in the augmented graph $G_{\mathbf{T}}$

For the given SCM $\mathcal{M}$, we construct a modified version $\tilde{\mathcal{M}}_{\mathbf{T}}$ as follows. We add $\mathbf{N_T}$ to the set of endogenous variables in $\mathcal{M}$ and remove them from the set of exogenous noises. Moreover, for any $X_i \in \mathbf{T}$, we change its structural assignment as follows: $X_i := \tilde{f}_i(\tilde{PA}_i) = f_i(PA_i, N_i)$ where $\tilde{f}_i$ is the new causal mechanism of relating it to its new set of parents $\tilde{PA}_i = PA_i \cup \{N_i\}$. For each variable $N_i \in \mathbf{N_T}$, we add an exogenous noise to $\tilde{\mathcal{M}}_{\mathbf{T}}$ and set the value of N_i to its corresponding exogenous noise. Please note that with this construction, the joint distribution over $\mathbf{X}$ entailed by SCM $\tilde{\mathcal{M}}_{\mathbf{T}}$ is exactly the same as the one entailed by original SCM $\mathcal{M}$. With the exact same argument in the proof of Theorem 1.4.1 in [Pea09], it can be shown that the distribution $p(\mathbf{X}\backslash\mathbf{T_L}, \mathbf{N_T})$ entailed by SCM $\tilde{\mathcal{M}}_{\mathbf{T}}$ satisfies the local Markov property as the value of each observed variable is uniquely determined given the values of its parts and the corresponding exogenous noise. Moreover, in causal DAGs, the local Markov property implies the Global Markov property [GP90]. Hence, the joint distribution $p(\mathbf{X}\backslash\mathbf{T_L}, \mathbf{N_T})$ satisfies Markov property with respect to its corresponding causal graph, $G_{\mathbf{T}}$.

B.2 Matching phase in the presence of latent confounders

We extend our results from causally sufficient case to the case where latent confounders are present. Unlike previous work [JKSB20, VSST22], we allow latent confounders to be in $\mathbf{T}$. In this case, an exogenous noise in $\mathbf{N_T}$ may correspond to either an observed variable or a latent confounder, and the task is to recover the observed variables that are in the intervention target set, i.e., $\mathbf{T_O}$. Unlike the causally sufficient case, $\mathbf{T_O}$ is not always uniquely identifiable. However, by modifying the LIT algorithm according to the conditions in Proposition 3 below, the algorithm can recover a superset $\mathbf{K}$ of $\mathbf{T_O}$. Proposition 3 provides conditions for finding variables that do not belong to $\mathbf{T_O}$. In particular, compared with Proposition 2, the statement in (I) still holds, while condition (III) is replaced by condition (III-L). Further, the statement in (II) does not hold anymore, and we have one extra condition (IV) for excluding variables from $\mathbf{T_O}$.

Proposition 3. *In the presence of latent confounders, under Assumption 2, for each observed variable X_i, the following statements hold:*

(I) $X_i \notin \mathbf{T_O}$ if the residual set $\mathcal{N}_i = \emptyset$.

(IV) $X_i \notin \mathbf{T_O}$ *if* $\mathcal{N}_i \neq \emptyset$*, and every recovered noise in the residual set* $\mathcal{N}_i$ *belongs to at least one other indicator set* $\mathbf{I}_{j_l}$*, where* $\mathbf{I}_{j_l}$ *is not a strict superset of* $\mathbf{I}_i$*:*

$$\forall \tilde{N}_l \in \mathcal{N}_i,\ \exists j_l \text{ s.t. } \tilde{N}_l \in \mathbf{I}_{j_l},\ \text{and } \mathbf{I}_i \not\subsetneq \mathbf{I}_{j_l}. \tag{A}$$

(III-L) *If* $\mathcal{N}_i \neq \emptyset$ *and condition* (A) *does not hold, let* $\mathcal{X}_i = \{X_{i_1}, \cdots, X_{i_p}\}$ *be the set of all variables with the same indicator set as* X_i*, including* X_i *itself. Then for each* $j \in [p]$*,* $X_{i_j} \notin \mathbf{T_O}$ *if it is independent of some recovered noise* $\tilde{N}_l$ *in* $\mathcal{N}_i$*, conditioned on certain subsets of observed variables in* $\mathcal{X}_i$*,* $\mathbf{S}_i$ *and all other recovered noises in* $\mathbf{I}_i$*:*

$$\begin{aligned} &\exists K \subseteq [p] \setminus \{j\},\ \exists \mathbf{S} \subseteq \mathbf{S}_i,\ \exists \tilde{N}_l \in \mathcal{N}_i \text{ s.t.} \\ &\tilde{N}_l \perp\!\!\!\perp X_{i_j} | \left(\cup_{k' \in K} X_{i_{k'}}\right) \cup \mathbf{S} \cup \left(\mathbf{I}_i \setminus \{\tilde{N}_l\}\right). \end{aligned} \tag{C2}$$

The statement in (IV) holds because if $X_i \in \mathbf{T_O}$, then its corresponding exogenous noise $\tilde{N}_i$ must be in $\mathcal{N}_i$. Any variable (such as X_{j_l}) that is dependent on $\tilde{N}_l$ must be a descendant of X_i, and hence have $\mathbf{I}_i \subseteq \mathbf{I}_{j_l}$. For (III-L), similar to the argument for condition (III), there is at most one variable (say X_{i_k}) in $\mathcal{X}_i$ that belongs to $\mathbf{T_O}$. Moreover, if $X_{i_k} \in \mathbf{T_O}$, then all other variables in $\mathcal{X}_i$ are its descendants. The recovered noises in $\mathcal{N}_i$ cannot be conditionally independent of X_{i_k}, as they correspond to either X_{i_k} or some latent confounder that is a parent of X_{i_k}. Therefore, if an observed variable $X_{i_j} \in \mathcal{X}_i$ is conditionally independent of a recovered noise in $\mathcal{N}_i$ given some other variables in the system, then it cannot be an intervention target. Note that under the causal sufficiency assumption, X_{i_k} and $\mathbf{S}_i$ are sufficient for the conditioning set. Therefore condition (III-L) reduces to (III). When latent confounders are present, X_{i_k} and $\mathbf{S}_i$ may not be sufficient. However, condition (III-L) states that in order to perform such a CI test, it suffices to consider subsets of $\mathcal{X}_i \setminus \{X_{i_j}\}$ and $\mathbf{S}_i$ in the conditioning set as their union contains all the ancestors of X_{i_j} among the observed variables.

Based on Proposition 3, we update the LIT algorithm in the presence of latent confounders. We check condition (IV) in line 4, and replace condition (III) by (III-L) in line 8. We keep line 5 in the latent case. In fact, if X_i is not ruled out by conditions (I) and (IV), it is added to $\mathbf{K}$ if $\mathbf{I}_i$ is unique as it cannot be ruled out by condition (III-L) either. However, the uniqueness of the indicator set does not necessarily imply that the variable belongs to $\mathbf{T_O}$. Lastly, note that under causal sufficiency assumption, condition (IV) is automatically satisfied, and condition (III-L) reduces to condition (III). Hence the algorithm remains consistent with the causally sufficient case.

Example 3. *Consider an SCM whose corresponding causal graph is depicted in Figure 3(a). It includes three observed variables* X_1, X_2, X_3 *and a latent confounder* X_H*, where* $\mathbf{T} = \{X_1, X_2\}$ *(shown in red). Suppose we recovered two noises* $\tilde{N}_1$*,* $\tilde{N}_2$ *that correspond to* X_1*,* X_2*, respectively. We have* $\mathbf{I}_1 = \{\tilde{N}_1\}$ *and* $\mathbf{I}_2 = \mathbf{I}_3 = \{\tilde{N}_1, \tilde{N}_2\}$*. Following the LIT algorithm, we find that* $\mathbf{I}_1$ *is unique and conditions (I) and (IV) do not hold for* X_1*. Therefore* $X_1 \in \mathbf{K}$*. Further,* $\tilde{N}_2 \perp\!\!\!\perp X_3 | X_1, X_2$*, and* $\tilde{N}_2 \not\perp\!\!\!\perp X_2 | \mathbf{S}$ *for all* $\mathbf{S} \in \{\emptyset, \{X_1\}, \{X_3\}, \{X_1, X_3\}\}$*. Therefore* $X_2 \in \mathbf{K}$*,* $X_3 \notin \mathbf{K}$ *according to condition (III-L). In conclusion, we have* $\mathbf{K} = \{X_1, X_2\}$*. Please note that* $\mathbf{K}$ *could be a strict superset of* $\mathbf{T_O}$ *in some cases (see Example 4).*

In the following, we provide a theoretical analysis of the candidate intervention target set $\mathbf{K}$ returned by LIT algorithm in the presence of latent confounders. In particular, we show that $\mathbf{K}$ contains the true intervention targets in the observed variables (i.e., $\mathbf{T_O}$). Further, we provide a graphical characterization of what other types of variables are also included in the set $\mathbf{K}$, using the notion of *auxiliary graph* which is defined as follows.

Definition 2 (Auxiliary graph)**.** *For each variable* X_i*, denote* $\mathbf{I}_0(X_i)$ *as* $An_{G_\mathbf{T}}(X_i) \cap \mathbf{N_T}$*. Given a SCM and its corresponding augmented graph* $G_\mathbf{T}$*, the auxiliary graph* $Aux(G_\mathbf{T})$ *is constructed from* $G_\mathbf{T}$ *as follows:*

(a) For each $X_i \in \mathbf{T_O}$ *with its corresponding exogenous noise* N_i*, add the edge* $N_i \to X_j$ *if (i) there is an inducing path between them relative to* $\mathbf{L} \setminus \mathbf{T_L}$ *in* $G_\mathbf{T}$ *(i.e., there is an edge between* N_i *and* X_j *in the MAG corresponding to* $G_\mathbf{T}$*) , and (ii)* $\mathbf{I}_0(X_i) = \mathbf{I}_0(X_j)$*.*

(b) (i) For each $N_l \in \mathbf{N_T}$ *(noise that corresponds to a variable in* $\mathbf{T_L}$*) and each of its child* X_i*, keep the edge* $N_l \to X_i$ *if for any other child* X_j *of* N_l *in* $G_\mathbf{T}$*,* $\mathbf{I}_0(X_i) \subseteq \mathbf{I}_0(X_j)$*. Otherwise remove the edge* $N_l \to X_i$*. (ii) For each remaining edge* $N_l \to X_i$*, add (remove) the edge*

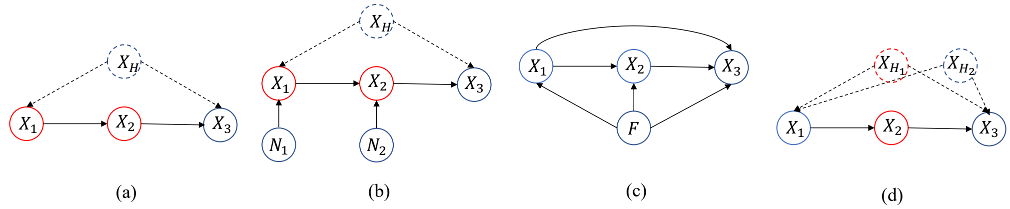

Figure 3: (a) The causal graph of the SCM considered in Example 4. (b) The corresponding auxiliary graph according to Definition 2. (c) The MAG of the augmented graph defined in [JKSB20], which indicates the output of their algorithm. (d) The causal graph of an alternative SCM that has the same auxiliary graph.

$N_l \to X_k$ if there is an (no) inducing path between X_k and all (some) parents of X_i in $\mathbf{N_T}$ relative to $\mathbf{L} \setminus \mathbf{T_L}$ in $G_\mathbf{T}$, and $\mathbf{I}_0(X_k) = \mathbf{I}_0(X_i)$.

Theorem 2. *In the presence of latent confounders and Assumption 2, the candidate intervention target set $\mathbf{K}$ returned by LIT algorithm is the set of observed variables that are children of $\mathbf{N_T}$ in $Aux(G_\mathbf{T})$, i.e., $\mathbf{K} = \cup_{N_i \in \mathbf{N_T}} Ch_{Aux(G_\mathbf{T})}(N_i)$.*

Theorem 2 gives a graphical characterization of the recovered candidate intervention set $\mathbf{K}$. In particular, according to Definition 2, $\mathbf{T_O}$ is a subset of $\mathbf{K}$. This is because the edge from $X_i \in \mathbf{T_O}$ to its corresponding exogenous noise N_i in $G_\mathbf{T}$ is not removed. This means that LIT algorithm returns a superset of $\mathbf{T_O}$. Further, two other types of variables are added to $\mathbf{K}$ according to the conditions in part (a) and part (b) of Definition 2, respectively.

Remark 2. *We can make the following observations from Theorem 2. First, under the causal sufficiency assumption, Theorem 2 implies that $\mathbf{T_O}$ can be uniquely identified. This is because no edges are added to $Aux(G_\mathbf{T})$ compared with $G_\mathbf{T}$ according to Definition 2. Second, if all latent variables are not intervention targets (i.e., $\mathbf{T} = \mathbf{T_O}$), then our identifiability result is stronger than the existing results in [JKSB20, VSST22]. In particular, they showed that in this case, $\mathbf{T_O}$ can only be identified up to the neighbors of $\mathbf{N_T}$ in the MAG corresponding to $G_\mathbf{T}$. We improve their results by adding part (a)(ii) in Definition 2, due to the recovery of $\tilde{\mathbf{N}}_\mathbf{T}$. See Example 4 below.*

Example 4. *Consider the same example as in Example 3. Following the results in [JKSB20, VSST22], the MAG of the augmented graph defined in [JKSB20] is shown in Figure 3(c). The recovery output of their algorithms is $\{X_1, X_2, X_3\}$ as all the observed variables are the neighbor of the F-node defined in their work. On the contrary, LIT algorithm has a more accurate recovery of the intervention targets. Specifically, the auxiliary graph is shown in Figure 3(b). The edge from N_1 to X_3 is not added since $\mathbf{I}_1 \neq \mathbf{I}_3$ (which violates part (a)(ii) in Definition 2), and the edge from N_2 to X_3 is not added because there is no inducing path (which violates part (a)(i)). Therefore $\mathbf{K} = \{X_1, X_2\}$, which is the same as the output in Example 3. Lastly, note that $\mathbf{K}$ is not always equal to $\mathbf{T_O}$. Consider the causal graph in Figure 3(d) where $\mathbf{T} = \{X_{H_1}, X_2\}$. Its corresponding auxiliary graph is exactly the one in Figure 3(b). However, $\mathbf{T_O} = \{X_2\} \subsetneq \{X_1, X_2\} = \mathbf{K}$, and we cannot distinguish whether X_1 or X_{H_1} is the intervention target.*

C Proofs

C.1 Proof of Proposition 1

We first prove the statement of proposition for the case of $\mathbf{L} = \emptyset$. With infinite samples and a model with universal approximation capability, after training, the regression model will equal the difference of the log-densities in the two classes:

$$\begin{aligned} r(\mathbf{X}, U) &= \log p(\mathbf{N}, U) + \log|\mathbf{Jg}(\mathbf{X})| - \log p(\mathbf{N}) - \log p(U) - \log|\mathbf{Jg}(\mathbf{X})| \\ &= \log p(\mathbf{N}|U) - \log p(\mathbf{N}) \\ &= \sum_i \log Q_i(N_i) - \log Z_i(U) + \tilde{q}_i(N_i)\lambda_i(U) - \log p(\mathbf{N}). \end{aligned} \tag{1}$$

If we consider the form of $r(\mathbf{X}, U) = \mathbf{h}(\mathbf{X})^T \mathbf{v}(U) + a(\mathbf{X}) + b(U)$ for $r(\mathbf{X}, U)$, we can set the functions $\mathbf{h}(\mathbf{X})$, $\mathbf{v}(U)$, $a(\mathbf{X})$, and $b(U)$ such that it is equal to the right hand side of above equation. In particular, we can have the following equality

$$\sum_i h_i(\mathbf{X})v_i(U)+a(\mathbf{X})+b(U) = \sum_i \log Q_i(N_i) - \log Z_i(U) + \sum_i \tilde{q}_i(N_i)\lambda_i(U) - \log p(\mathbf{N}), \quad (2)$$

with the following possible solution:

$$h_i(\mathbf{X}) = \tilde{q}_i(N_i), v_i(U) = \lambda_i(U), a(\mathbf{X}) = \sum_i \log Q_i(N_i) - \log p(\mathbf{N}), b(U) = -\sum_i \log Z_i(U). \quad (3)$$

The random variable U is equal to d, if the sample is drawn from the environment E_d where $1 \leq d \leq D$. Let $\mathbf{P}$ be $D \times n$ matrix where d-th row is equal to $[\lambda_1(d), \cdots, \lambda_n(d)]$. We collect $\tilde{q}_i$s in the vector $\tilde{\mathbf{q}}(\mathbf{N}) = [\tilde{q}_1(N_1), \cdots, \tilde{q}_n(N_n)]^T$. We also define the matrix $\mathbf{V}$ where d-th row is equal to $[v_1(d), \cdots, v_n(d)]$. Finally, we collect $-\sum_i \log Z_i(U) - b(U)$ for different values of U in a vector $\mathbf{Z}$. Based on these definitions, we have:

$$\mathbf{V}\mathbf{h}(\mathbf{X}) = \mathbf{P}\tilde{\mathbf{q}}(\mathbf{N}) + \mathbf{Z} + \mathbf{1}\left(\sum_i \log Q_i(N_i) - \log p(\mathbf{N}) - a(\mathbf{X})\right), \quad (4)$$

where $\mathbf{1}$ is a $D \times 1$ vector of all ones. If from both sides of the above equation, we subtract the first row from the others,

$$\mathbf{V}'\mathbf{h}(\mathbf{X}) = \mathbf{P}'\tilde{\mathbf{q}}(\mathbf{N}) + \mathbf{Z}', \quad (5)$$

where $\mathbf{V}', \mathbf{P}'$, and $\mathbf{Z}'$ denote the resulting matrices after subtraction corresponding to $\mathbf{V}, \mathbf{P}$, and $\mathbf{Z}$, respectively.

The columns corresponding to exogenous noises that are not changing across environments are zeros in $\mathbf{P}'$. Thus, we can remove these columns from $\mathbf{P}'$ and also the corresponding entries in $\tilde{\mathbf{q}}(\mathbf{N})$. We denote the resulting matrix and vector by $\mathbf{P}''$ and $\tilde{\mathbf{q}}(\mathbf{N_T})$, respectively. Hence, we can rewrite the above equation as follows:

$$\mathbf{V}'\mathbf{h}(\mathbf{X}) = \mathbf{P}''\tilde{\mathbf{q}}(\mathbf{N_T}) + \mathbf{Z}'. \quad (6)$$

As $\lambda_i(U)$s are generated randomly across the environments and $D \geq |\mathbf{T}| + 1$, $\mathbf{P}''$ is full column rank with measure one. Therefore, we have:

$$\tilde{\mathbf{q}}(\mathbf{N_T}) = (\mathbf{P}'')^\dagger \mathbf{V}'\mathbf{h}(\mathbf{X}) - \mathbf{Z}'', \quad (7)$$

where $(\mathbf{P}'')^\dagger$ is pseudo-inverse of matrix $\mathbf{P}''$ and $\mathbf{Z}'' = (\mathbf{P}'')^\dagger \mathbf{Z}'$. Since we know that the entries of $\tilde{\mathbf{q}}(\mathbf{N_T})$ are linearly independent, $(\mathbf{P}'')^\dagger \mathbf{V}'$ is full row rank. Moreover, $q_i(N_i)$s are non-Gaussian as they are bounded from above to ensure integrability. Thus, we can recover $\tilde{\mathbf{q}}(\mathbf{N_T})$ from $\mathbf{h}(\mathbf{X})$ by solving an under-complete linear ICA problem.

Now, let us assume that there are some latent variables in the system, i.e., $\mathbf{L} \neq \emptyset$. As we know that there exists an invertible function $\tilde{\mathbf{g}}$ such that $\tilde{\mathbf{g}}(\mathbf{O}) = (\mathbf{N_T}, \mathbf{V})$, we have:

$$\begin{aligned} r(\mathbf{O}, U) &= \log p(\mathbf{O}, U) - \log p(\mathbf{O}) - \log p(U) \\ &= \log p(\mathbf{O}|U) - \log p(\mathbf{O}) \\ &= \log p(\mathbf{N_T}, \mathbf{V}|U) + \log |\mathbf{J}\tilde{\mathbf{g}}(\mathbf{O})| - \log p(\mathbf{N_T}, \mathbf{V}) - \log |\mathbf{J}\tilde{\mathbf{g}}(\mathbf{O})| \\ &= \log p(\mathbf{N_T}|U) - \log p(\mathbf{N_T}), \end{aligned} \quad (8)$$

where the third equality is due to the existence of invertible function $\tilde{\mathbf{g}}$ and the last equality is according to the assumptions that $\mathbf{N_T} \perp\!\!\!\perp \mathbf{V}|U$, $\mathbf{V} \perp\!\!\!\perp U$. Please note that these two assumptions imply that $\mathbf{N_T} \perp\!\!\!\perp \mathbf{V}$. Similar to the causally sufficient case, based on the form of $r(\mathbf{O}, U)$, we can write the following equation:

$$\sum_j h_j(\mathbf{O})v_j(U)+a(\mathbf{O})+b(U) = \sum_{i:X_i \in \mathbf{T}} \log Q_i(N_i) - \log Z_i(U) + \tilde{q}_i(N_i)\lambda_i(U) - \log p(\mathbf{N_T}). \quad (9)$$

where $h(\mathbf{O}) : \mathbb{R}^{|\mathbf{O}|} \to \mathbb{R}^{|\mathbf{O}|}$. Let $\mathbf{M}$ be $D \times |\mathbf{T}|$ matrix where d-th row is equal to $[\lambda_1(d), \cdots, \lambda_{|\mathbf{T}|}(d)]$. We collect $\tilde{q}_i$s in the vector $\tilde{\mathbf{q}}(\mathbf{N_T})$. We also define the matrix $\mathbf{W}$ where

d-th row is equal to $[v_1(d), \cdots, v_{|\mathbf{O}|}(d)]$. Finally, we collect $-\sum_{i:X_i\in\mathbf{T}} \log Z_i(U) - b(U)$ for different values of U in a vector $\mathbf{Z}$. Based on these definitions, we have:

$$\mathbf{W}\mathbf{h}(\mathbf{O}) = \mathbf{M}\tilde{\mathbf{q}}(\mathbf{N}_\mathbf{T}) + \mathbf{Z} + \mathbf{1}\left(\sum_{i:X_i\in\mathbf{T}} \log Q_i(N_i) - \log p(\mathbf{N}_\mathbf{T}) - a(\mathbf{O})\right), \tag{10}$$

where $\mathbf{1}$ is a $D\times 1$ vector of all ones.

Now, if from both sides of the above equation, we subtract the first row from the others, we have

$$\mathbf{W}'\mathbf{h}(\mathbf{O}) = \mathbf{M}'\tilde{\mathbf{q}}(\mathbf{N}_\mathbf{T}) + \mathbf{Z}', \tag{11}$$

where $\mathbf{V}', \mathbf{L}'$, and $\mathbf{Z}'$ denote the resulting matrices after subtraction corresponding to $\mathbf{W}, \mathbf{M}$, and $\mathbf{Z}$, respectively.

As $\lambda_i(U)$s are generated randomly across the environments and $D \geq |\mathbf{T}|+1$, $\mathbf{M}'$ is full column rank with measure one. Therefore, we have:

$$\tilde{\mathbf{q}}(\mathbf{N}_\mathbf{T}) = (\mathbf{M}')^\dagger\mathbf{W}'\mathbf{h}(\mathbf{O}) - \mathbf{Z}'', \tag{12}$$

where $(\mathbf{M}')^\dagger$ is pseudo-inverse of matrix $\mathbf{M}'$ and $\mathbf{Z}'' = (\mathbf{M}')^\dagger\mathbf{Z}'$. Since we know that the entries of $\tilde{\mathbf{q}}(\mathbf{N}_\mathbf{T})$ are linearly independent, $(\mathbf{M}'')^\dagger\mathbf{W}'$ is full row rank. Moreover, $q_i(N_i)$s are non-Gaussian as they are bounded from above to ensure integrability. Thus, we can recover $\tilde{\mathbf{q}}(\mathbf{N}_\mathbf{T})$ from $\mathbf{h}(\mathbf{O})$ by solving an under-complete linear ICA problem.

So far, we showed that the recovery phase can be performed up to some component-wise nonlinear transformation (not necessarily an invertible one). However, similar to Corollary 2 in [HM16], it can be shown that from $\tilde{\mathbf{q}}(\mathbf{N}_\mathbf{T})$ and the observed vector $\mathbf{O}$, the exogenous noises in $\mathbf{N}_\mathbf{T}$ can be recovered up to some strictly monotonic transformation if each function $\tilde{q}_i(N_i)$ is a strictly monotonic function of $|N_i|$.

C.2 Regarding Assumption 1(b) on Linear non-Gaussian Models

In the following we prove that, in the linear non-Gaussian model (i.e., linear SCM with non-Gaussian exogenous noises) with the causal graph G, if $\mathbf{L} \neq \emptyset$, then the conditions in Proposition 1 imply that: (i) $\mathbf{T} \subseteq \mathbf{O}$; (ii) For each latent variable $H_i \in \mathbf{L}$, $Ch_G(H_i) \cap \mathbf{T} = \emptyset$, i.e., H_i cannot have children in $\mathbf{T}$ where $Ch_G(H_i)$ is the children of H_i.

The linear SCM has the following matrix form:

$$\mathbf{L} = \mathbf{N}_\mathbf{L}; \quad \mathbf{O} = \mathbf{A}\mathbf{O} + \mathbf{A}'\mathbf{L} + \mathbf{N}_\mathbf{O}, \tag{13}$$

where $\mathbf{N}_\mathbf{L}$ and $\mathbf{N}_\mathbf{O}$ represent the vector of exogenous noises associated with latent and observed variables, respectively. We also denote the exogenous noises whose distributions are not changing across the environments by $\mathbf{N}_{\mathbf{T}^c}$. $\mathbf{A}$ represents the direct causal relationships among observed variables and $\mathbf{A}'$ represents the direct causal relations from latent to observed variables. Note that if we permute the variables such that $\mathbf{X} = [\mathbf{L}, \mathbf{O}]$, then the adjacency matrix $\mathbf{B}$ after the same (row and column) permutation is $[\mathbf{0}, \mathbf{0}; \mathbf{A}', \mathbf{A}]$. Under the acyclicity assumption, $\mathbf{A}$ can be permuted into a strictly lower triangular matrix. Following (13), $\mathbf{O}$ can be written as a linear combination of the noise terms in $(\mathbf{N}_\mathbf{L}, \mathbf{N}_\mathbf{O})$:

$$\mathbf{O} = [\mathbf{D} \quad \mathbf{C}]\begin{bmatrix}\mathbf{N}_\mathbf{L}\\ \mathbf{N}_\mathbf{O}\end{bmatrix}, \tag{14}$$

where $\mathbf{D} = (\mathbf{I}-\mathbf{A})^{-1}\mathbf{A}'$ and $\mathbf{C} = (\mathbf{I}-\mathbf{A})^{-1}$. Denote $\mathbf{W} = [\mathbf{D} \quad \mathbf{C}]$, which represents the total causal effects (i.e., sum of product of path coefficients) among variables [SGSH00]. If all exogenous noises are non-Gaussian and no two columns of $\mathbf{W}$ are linearly dependent of each other, then $\mathbf{W}$ can be recovered up to permutation and scaling of the columns using overcomplete ICA. Given that some variables in $\mathbf{X}$ belong to $\mathbf{T}$, we can rewrite (14) as follows:

$$\mathbf{O} = [\mathbf{W}_\mathbf{T} \quad \mathbf{W}_{\mathbf{T}^c}]\begin{bmatrix}\mathbf{N}_\mathbf{T}\\ \mathbf{N}_{\mathbf{T}^c}\end{bmatrix}, \tag{15}$$

where $\mathbf{W}_\mathbf{T}$ and $\mathbf{W}_{\mathbf{T}^c}$ represent the submatrix of $\mathbf{W}$ that correspond to the exogenous noises in $\mathbf{N}_\mathbf{T}$ and $\mathbf{N}_{\mathbf{T}^c}$, respectively.

If the conditions in Proposition 1 hold, then there exists an invertible matrix $\mathbf{G}$, such that

$$\mathbf{GO} = [\mathbf{GW_T} \quad \mathbf{GW}_{\mathbf{T}^c}] \begin{bmatrix} \mathbf{N_T} \\ \mathbf{N}_{\mathbf{T}^c} \end{bmatrix} = \begin{bmatrix} \mathbf{N_T} \\ \mathbf{V} \end{bmatrix}. \tag{16}$$

Partition $\mathbf{G}$ into $[\mathbf{G}_1; \mathbf{G}_2]$, where $\mathbf{G}_1 \in \mathbb{R}^{|\mathbf{T}| \times n}$ represent the first $|\mathbf{T}|$ rows of $\mathbf{G}$, and $\mathbf{G}_2 \in \mathbb{R}^{(n-|\mathbf{T}|) \times n}$ represent the remaining rows. Therefore, according to (16), we have

$$\mathbf{G}_1 \mathbf{W_T} \mathbf{N_T} + \mathbf{G}_1 \mathbf{W}_{\mathbf{T}^c} \mathbf{N}_{\mathbf{T}^c} = \mathbf{N_T}. \tag{17}$$

We first show that if all noises in $\mathbf{N_T}$ and $\mathbf{N}_{\mathbf{T}^c}$ are mutually independent and non-Gaussian, then (17) implies that $\mathbf{G}_1 \mathbf{W_T} = \mathbf{I}$, and $\mathbf{G}_1 \mathbf{W}_{\mathbf{T}^c} = \mathbf{0}$. This is because for each noise $N_i \in \mathbf{N_T}$, $i \in [|\mathbf{T}|]$, according to (17), N_i can be written as a linear combination of exogenous noises in $\mathbf{N_T} \cup \mathbf{N}_{\mathbf{T}^c}$. Since all noises are mutually independent and non-Gaussian, according to Darmois-Skitovitch theorem [Dar53, Ski53], the coefficient of any N_j, $j \neq i$ on N_i must be zero.

For each exogenous noise $N \in \mathbf{N_T} \cup \mathbf{N}_{\mathbf{T}^c}$, denote its corresponding column vector in $\mathbf{W}$ in (15) as $\mathbf{w}_N$. Then $\mathbf{G}_1 \mathbf{W_T} = \mathbf{I}$, $\mathbf{G}_1 \mathbf{W}_{\mathbf{T}^c} = \mathbf{0}$ is equivalent to: For each exogenous noise N and its corresponding column vector $\mathbf{w}_N$, $\mathbf{G}_1 \mathbf{w}_N = \mathbf{0}$ if $N \in \mathbf{N}_{\mathbf{T}^c}$, and $\mathbf{G}_1 \mathbf{w}_N = \mathbf{e}_N$ if $N \in \mathbf{N_T}$, where $\mathbf{e}_N$ is the basis (one-hot) vector where the entry corresponding to N is one and the rest are zero. Further, for each latent variable $H_i \in \mathbf{L}$, we have

$$\mathbf{w}_{N_{H_i}} = \sum_{j: X_j \in Ch_G(H)} a'_{ji} \mathbf{w}_{N_{X_j}},$$

where a'_{ji} represent the (j, i)-th entry of matrix $\mathbf{A}'$ in (13). This is because for any observed variable $X \in \mathbf{O}$, the total causal effect of H_i on X can be written as summation of the total causal effect from each child of H_i to X multiplied by the direct causal effect from H_i to this child (i.e., a'_{ji}). Therefore, we have

$$\mathbf{G}_1 \mathbf{w}_{N_{H_i}} = \sum_{j: X_j \in Ch_G(H_i)} a'_{ji} \mathbf{G}_1 \mathbf{w}_{N_{X_j}}. \tag{18}$$

Since $\mathbf{G}_1 \mathbf{w}_N$ corresponds to either zero vector or basis vector, and different noises in $\mathbf{N_T}$ correspond to different basis vectors, (18) implies that $\mathbf{G}_1 \mathbf{w}_{N_{H_i}} = \mathbf{0}$, and $\mathbf{G}_1 \mathbf{w}_{N_{X_j}} = \mathbf{0}$ for all latent variable $H_i \in \mathbf{L}$, and all $X_j \in Ch_G(H_i)$. This means that $\mathbf{T}$ only includes observed variables that do not have latent parents.

C.3 Proof of Proposition 2

Consider any recovered noise $\tilde{N}_i \in \tilde{\mathbf{N}}_\mathbf{T}$. Without loss of generality, suppose that $\tilde{N}_i$ corresponds to X_i. We first show that $\tilde{N}_i$ only depends on the descendants of X_i. First, for any node $X_j \in Des(X_i)$, the path $N_i \to X_i \to \cdots \to X_j$ in graph $G_\mathbf{T}$ is not blocked without conditioning on any other variable and $\tilde{N}_i \not\perp\!\!\!\perp X_j$. Moreover, for any X_j which is a non-descendent of X_i, there is always a collider on any path between X_j and N_i and thus it is blocked. Hence, we have: $\tilde{N}_i \perp\!\!\!\perp X_j$.

Remark 3. *Based on what we proved above, we can imply that $\mathbf{I}_i$ contains only the noises in $\tilde{\mathbf{N}}_\mathbf{T}$ whose corresponding variables are ancestors of X_i.*

Remark 4. *For any two variables X_i, X_j, if we have $\mathbf{I}_j \subsetneq \mathbf{I}_i$, then X_j cannot be a descendent of X_i. Since if X_j is a descendent of X_i, then based on Remark 3, we can conclude that $\mathbf{I}_i \subseteq \mathbf{I}_j$ which violates our assumption.*

Now, we prove the three statements in the proposition based on the above two remarks:

(I) By contradiction, suppose that X_i is in $\mathbf{T}$. If for a variable X_j, we have $\mathbf{I}_j \subsetneq \mathbf{I}_i$, then based on Remark 4, it cannot be a descendent of X_i. Now if the condition in (I) satisfies, then there exists a set such as $\mathbf{I}_j$, where $j \in \mathbf{S}_i$, such that $i \in \mathbf{I}_j$. But according to Remark 3, this means that X_j is a descendant of X_i which is a contradiction.

(II) As the condition in (I) is not satisfied, there exists k such that $\tilde{N}_k \in \mathcal{N}_i = \mathbf{I}_i \setminus \cup_{j: X_j \in \mathbf{S}_i} \mathbf{I}_j$. Suppose that $\tilde{N}_k$ corresponds to X_k. Based on Remark 3, X_k should be an ancestor of X_i. Moreover, we have $\mathbf{I}_k \subseteq \mathbf{I}_i$. As $\mathbf{I}_i$ is unique, then $\mathbf{I}_k \subsetneq \mathbf{I}_i$ and $X_k \in \mathbf{S}_i$. This contradictsthe fact that $\tilde{N}_k \in \mathbf{I}_i \setminus \cup_{j: X_j \in \mathbf{S}_i} \mathbf{I}_j$ and the proof is complete.

(III) As the indicator sets do not satisfy the condition in (I), $\mathcal{N}_i$ is not empty. Suppose that $\tilde{N}_k \in \mathcal{N}_i$ and $\tilde{N}_k$ corresponds to X_k. We know that X_k should be in the set $\{X_{i_1}, \cdots, X_{i_p}\}$. Otherwise, based on Remark 3, $\mathbf{I}_k \subseteq \mathbf{I}_{i_1}$. Now, if $\mathbf{I}_k \subsetneq \mathbf{I}_{i_1}$, then $X_k \in \mathbf{S}_{i_1}$ which is a contradiction. Thus, $\mathbf{I}_k$ should be equal to $\mathbf{I}_i$ but in that case X_k is in the set $\{X_{i_1}, \cdots, X_{i_p}\}$. Thus, we can conclude that at least one of the variables in this set is in $\mathbf{T}$. Please note that at most one of the variable in $\{X_{i_1}, \cdots, X_{i_p}\}$ can be in $\mathbf{T}$. Otherwise, the variables in this set cannot have the same indicator set. Therefore, exactly one of the variables in $\{X_{i_1}, \cdots, X_{i_p}\}$ is in $\mathbf{T}$ and we can obtain the corresponding recovered noise by $\mathbf{I}_{i_1} \setminus \cup_{j:X_j \in \mathbf{S}_{i_1}} \mathbf{I}_j$. Without loss of generality, suppose that $X_{i_1} \in \mathbf{T}$. Then, based on Remark 3, other variables in $\{X_{i_1}, \cdots, X_{i_p}\}$ should be descendant of X_{i_1}. The set $\mathbf{S}_{i_1}$ contains the ancestors of X_{i_1}. Thus, it also includes parents of X_{i_1}. Now, for any path between N_{i_1} and X_{i_j}, $j \geq 2$, if it is outgoing from node X_{i_1}, then it is blocked by X_{i_1}. If it is in-going toward X_{i_1}, then it is blocked by a parent of X_{i_1} which is inside the set $\mathbf{S}_{i_1}$. Thus, under T-faithfulness assumption, it implies that $\tilde{N}_{i_1} \perp\!\!\!\perp X_{i_j} | \{X_{i_1}\} \cup \mathbf{S}_{i_1}$, for all $2 \leq j \leq p$. Please note that the variables in $\{X_{i_2}, \cdots, X_{i_p}\}$ cannot satisfy the condition in (III) as they cannot block the path $N_{i_1} \to X_{i_1}$ by any set $\mathbf{S}_{i_j}$ for $j \geq 2$ and the proof is complete.

C.4 Proof of Theorem 1

It can be easily seen that for any variable $X_i \in \mathbf{X}$, only one of the conditions in (I)-(III) in Proposition 2 is satisfied. Moreover, these conditions cover all possible cases regarding the relation of the set $\cup_{j:X_j \in \mathbf{S}_i} \mathbf{I}_j$ and $\mathbf{I}_i$ and also the uniqueness of $\mathbf{I}_i$ in $\mathcal{I}$. Furthermore, in either case, we know definitely whether X_i is in $\mathbf{T}$ or not. Thus, the intervention target $\mathbf{T}$ can be identified uniquely by checking these three conditions for any variable $X_i \in \mathbf{X}$ and the proof is complete.

C.5 Proof of Theorem 2

To better distinguish between intervention targets in $\mathbf{T_O}$ and in $\mathbf{T_L}$, we use $\tilde{N}_{h_i}$ to represent the exogenous noise corresponding to X_i for all $X_i \in \mathbf{T_L}$. For ease of notation, we denote the true noises N_j in the augmented graph as $\tilde{N}_j$ in the following proof, since there is a one-to-one correspondence between them.

Remark 5. *If a group of variables $\mathcal{X}_i = \{X_{i_1}, \cdots, X_{i_p}\}$ have the same indicator set and $\mathcal{N}_i \neq \emptyset$ (note that $\mathcal{N}_{i_k}$ is the same for all $k \in [p]$), then any root variable in $\mathcal{X}_i$ (i.e., variable with no parent in $\mathcal{X}_i$) must be a child of all recovered noises in $\mathcal{N}_{i_1}$.*

C.5.1 Proof of sufficiency

We show that if an observed variable X_i is a child of some noise $\tilde{N}_j \in \tilde{\mathbf{N}}_\mathbf{T}$ in $Aux(G_\mathbf{T})$, then it is included in the output $\mathbf{K}$ of the LIT algorithm. We consider each of the following four cases:

(i) $X_i \in \mathbf{T}$.

(ii) $X_i \notin \mathbf{T}$, and $\tilde{N}_j$ is the exogenous noise of some observed variable X_j.

(iii) $X_i \notin \mathbf{T}$, $\tilde{N}_j$ is the exogenous noise of some latent confounder H_j, and H_j is a parent of X_i in G.

(iv) $X_i \notin \mathbf{T}$, $\tilde{N}_j$ is the exogenous noise of some latent confounder H_j, and H_j is not a parent of X_i in G.

In the following we show that, under each of these four cases, X_i is included in the output $\mathbf{K}$. That is, X_i violates the condition in (I), violates the condition in (IV), and either satisfies the condition in (II) or violates (C2) in (III-L). Note that the conditions in (II) and in (III-L) only depends on the uniqueness of the indicator set, therefore we do not need to check (II).

Case (i). If $X_i \in \mathbf{T}$, then its corresponding exogenous noise $\tilde{N}_i$ satisfies $i \in \tilde{N}_i$. Therefore the condition in (I) Proposition 2 is not satisfied. Further, any observed variable X_{j_l} with $i \in \mathbf{I}_{j_l}$ must be a descendant of X_i, and hence satisfies $\mathbf{I}_i \subseteq \mathbf{I}_{j_l}$. Therefore the condition in (IV) is not satisfied. Lastly, if there are no other variables that have the same indicator set, then $X_i \in \mathbf{K}$ according to (II). Otherwise, all other variables that have the same indicator set as $\mathbf{I}_i$ must be descendants of X_i (because of $\tilde{N}_i$), hence X_i is a child of any recovered noise in $\mathcal{N}_i$ (according to Remark 5). Therefore there does not exist l such that (C2) holds, which means that $X_i \in \mathbf{K}$.

Case (ii). If $X_i \notin \mathbf{T}$ and $\tilde{N}_j$ is the exogenous noise of some observed variable X_j, then $\mathbf{I}_j = \mathbf{I}_i$, and there is at least one inducing path from $\tilde{N}_j$ to X_i. This means that any variable $X_k \in \mathbf{S}_i$ is not a descendant of $\tilde{N}_j$ in $G_{\mathbf{T}}$, which implies that $j \in \mathbf{I}_i$ but not in $\cup_{j':X_{j'} \in \mathbf{S}_i} \mathbf{I}_{j'}$. Therefore the condition in (I) is not satisfied. Next, note that $j \in \mathcal{N}_i = \mathcal{N}_j$, $\mathbf{I}_i = \mathbf{I}_j$, and (A) in (IV) does not hold for X_j (explained in Case (i)). Therefore (A) in (IV) does not hold for X_i either.

Lastly, note that there is at least one other variable (X_j) that has the same indicator set as X_i, and X_j is a root variable in $\mathcal{X}_i$ (following the same argument as in Case (i)). Therefore X_j is directly connected to all recovered noises in $\mathcal{N}_i$. Since $\tilde{N}_j$ is only directly connected to X_j, this means that if there is an inducing path from $\tilde{N}_j$ to X_i, then for any $l \in \mathcal{N}_i$, by changing the first edge on this path from $\tilde{N}_j \to X_j$ to $\tilde{N}_l \to X_j$, the new path is also an inducing path from $\tilde{N}_l$ to X_i. Therefore there does not exist l such that (C2) holds, which means that $X_i \in K$.

Case (iii). If $X_i \notin \mathbf{T}$, $\tilde{N}_j$ is the exogenous noise of latent confounder H_j, and H_j is a parent of X_i in G, then according to Definition 2, for all other children X_k of H_j, we have $\mathbf{I}_i \subseteq \mathbf{I}_k$. This immediately implies that the condition in (IV) is not satisfied. This also implies that for any variable $X_{k'} \in \mathbf{S}_i$, $j \notin \mathbf{I}_{k'}$. Therefore, we have $j \in \mathbf{I}_i$ but not in $\cup_{j':X_{j'} \in \mathbf{S}_i} \mathbf{I}_{j'}$, which means that the condition in (I) is not satisfied. Lastly, if $\mathbf{I}_i$ is unique, then $X_i \in \mathbf{K}$. Otherwise, since there is an inducing path from all recovered noises in $\mathcal{N}_i$ to X_i, there does not exist l such that (C2) holds. Therefore $X_i \in \mathbf{K}$.

Case (iv). If $X_i \notin \mathbf{T}$, $\tilde{N}_j$ is the exogenous noise of latent confounder H_j, and H_j is not a parent of X_i in G, then according to Definition 2, there exists some X_k such that H_j is directly connected to X_k, and $\mathbf{I}_k = \mathbf{I}_i$. Further, there is an inducing path from $\tilde{N}_j$ to X_i. Note that all children of $\tilde{N}_j$ after part (b) have the same indicator set, which is the same as $\mathbf{I}_i$. Therefore the condition in (I) does not hold. Further, since the condition in (A) does not hold for X_k (explained in Case (iii)) and $\mathbf{I}_k = \mathbf{I}_i$, the condition in (A) does not hold for X_i either. Lastly, since there is an inducing path from all recovered noises in $\mathcal{N}_i$ to X_i, there does not exist l such that (C2) holds. Therefore $X_i \in \mathbf{K}$.

Conclusion: We show that if an observed variable X_i belongs to $\cup_{\tilde{N}_i \in \tilde{\mathbf{N}}_{\mathbf{T}}} Ch_{Aux(G_{\mathbf{T}})}(\tilde{N}_i)$, i.e., it is a child of some noise $\tilde{N}_j \in \tilde{\mathbf{N}}_{\mathbf{T}}$ in $Aux(G_{\mathbf{T}})$, then it is included in the output $\mathbf{K}$ of the LIT algorithm.

C.5.2 Proof of necessity

We show that if an observed variable X_i is included in the output $\mathbf{K}$ of the LIT algorithm, then it must be a child of some noise $\tilde{N}_j \in \tilde{\mathbf{N}}_{\mathbf{T}}$. That is, if X_i violates the condition in (I) (i.e., $\mathcal{N}_i \neq \emptyset$), and the condition in (IV), and either:

(i) $\mathbf{I}_i$ is unique, or

(ii) $\mathbf{I}_i$ is not unique, and among all variables with the same indicator set, the condition in (C2) does not hold for X_i,

then it must be a child of some noise $\tilde{N}_j \in \tilde{\mathbf{N}}_{\mathbf{T}}$.

Case (i). If $\mathbf{I}_i$ is unique. This implies that any recovered noise $\tilde{N}_j \in \mathcal{N}_i$ is a parent of X_i in $G_{\mathbf{T}}$. This is because otherwise the child of $\tilde{N}_j$ is a ancestor of X_i but does not belong to $\mathbf{S}_i$. Hence it has the same indicator set as $\mathbf{I}_i$, which violates the uniqueness of $\mathbf{I}_i$.

Consider all these recovered noises $\tilde{N}_j \in \mathcal{N}_i$. If there exists $\tilde{N}_j$ such that X_i is the only child of $\tilde{N}_j$, then $\tilde{N}_j$ is the exogenous noise of X_i, i.e., $X_i \in \mathbf{T}_{\mathbf{O}}$, which is a subset of $\cup_{\tilde{N}_i \in \tilde{\mathbf{N}}_{\mathbf{T}}} Ch_{Aux(G_{\mathbf{T}})}(\tilde{N}_i)$. Otherwise, if all noises $\tilde{N}_j$ has at least two children, then all of them correspond to the exogenous noises of latent confounders. Since the condition in (IV) is violated, there exist $l \in \mathcal{N}_i$ such that for all j_l with $l \in \mathbf{I}_{j_l}$, $\mathbf{I}_i \subseteq \mathbf{I}_{j_l}$. This implies that the indicator set of any other child of $\tilde{N}_l$ in $G_{\mathbf{T}}$ must be a (strict) superset of X_i. Therefore the edge $\tilde{N}_l \to X_i$ is kept according to part (b)(i) of Definition 2, i.e., X_i is the child of $\tilde{N}_l$ in $Aux(G_{\mathbf{T}})$.

Case (ii). If $\mathbf{I}_i$ is not unique. This means that $\mathcal{X}_i$ is not a singleton. Consider the root variables in $\mathcal{X}_i$. Note that following the induced causal order induced on $\mathcal{X}_i$, there is at least one root variable. If X_i is a root variable, then according to Remark 5, all recovered noises in $\mathcal{N}_i$ is a parent of X_i in $G_{\mathbf{T}}$. In this case we can apply the same proof as in Case (i) to show that X_i is a child of some $\tilde{N}_j$ in $Aux(G_{\mathbf{T}})$. That is, if there exists $\tilde{N}_j$ such that X_i is the only child of $\tilde{N}_j$, then $X_i \in \mathbf{T}_{\mathbf{O}}$. Otherwise, if all noises $\tilde{N}_j$ has at least two children, then since the condition in (IV) is violated, part (b)(i) in Definition 2 is satisfied, and there is an inducing path from all recovered noises in $\mathcal{N}_i$ to X_i (because a direct connection is an inducing path). Therefore X_i is the child of all $\tilde{N}_j$ in $Aux(G_{\mathbf{T}})$.

Next, we consider the case if X_i is not a root variable. If there exists $\tilde{N}_j \in \mathcal{N}_i$ such that it has only one direct child X_k in $G_{\mathbf{T}}$, then X_k is a root variable in $\mathcal{X}_i$. Since (C2) does not hold for X_i, this means that $\tilde{N}_j$ is not independent of X_i, conditioned on all observed ancestors of X_i in $G_{\mathbf{T}}$ (i.e., $An_{G_{\mathbf{T}}}(X_i) \setminus (\{X_i\} \cup (\mathbf{L} \backslash \mathbf{T}))$). This is because all observed ancestors of X_i in $G_{\mathbf{T}}$ must belong to one of the three following cases: observed variable whose indicator set is a strict subset of $\mathbf{I}_i$ (i.e., belongs to $\mathbf{S}_i$), observed variable whose indicator set is the same as $\mathbf{I}_i$ (i.e., belongs to $\mathcal{X}_i$), and the recovered exogenous noise of a latent confounder (i.e., belongs to $\cup_{l' \in \mathbf{I}_{i_1} \setminus \{j\}} \tilde{N}_{l'}$).

In the following we show that the path that cannot be blocked by these observed ancestors is an inducing path. Suppose the path $\tilde{N}_j \rightarrow X_k - V_{k_1} - V_{k_2} \cdots X_i$ is not blocked, where $\{V_{k_m}\}$ are variables in $G_{\mathbf{T}}$, and $-$ represents either $\leftarrow$ or $\rightarrow$. Note that X_k is an ancestor of X_i thus it is in the condition set. Therefore X_k is a collider on this path, and we have the edge $X_k \leftarrow V_{k_1}$. This means that V_{k_1} is a non-collider on this path and an ancestor of X_i in $G_{\mathbf{T}}$. Since the path is active, V_{k_1} is a latent confounder that is not changing across environments (i.e., belongs to $\mathbf{L} \setminus \mathbf{T}$). Therefore there is an edge $V_{k_1} \rightarrow V_{k_2}$. Next consider V_{k_2}. If V_{k_2} is a non-collider and is not in the conditioning set (i.e., not an ancestor of X_i), then there is an edge $V_{k_2} \rightarrow V_{k_3}$. Since V_{k_2} is not an ancestor of X_i, neither is V_{k_3}. Therefore V_{k_3} is also a non-collider and we have $V_{k_3} \rightarrow V_{k_4}$. Repeat the same analysis on V_{k_m} for all $m = 4, 5, \cdots$, the path is $\tilde{N}_j \rightarrow X_k \leftarrow V_{k_1} \rightarrow V_{k_2} \rightarrow V_{k_3} \rightarrow \cdots \rightarrow X_i$. This violates the claim that V_{k_2} is not an ancestor of X_i. Therefore, V_{k_2} is a collider and is in the conditioning set (i.e., an ancestor of X_i). Then we have the edge $V_{k_2} \leftarrow V_{k_3}$, and following the same analysis in V_{k_1} on V_{k_3}, we have $V_{k_3} \in \mathbf{L} \setminus \mathbf{T}$ and is a confounder. Repeat the same analysis on V_{k_m} for all $m = 3, 4, 5, \cdots$, we have that variables V_{k_m} are either confounders in $\mathbf{L} \setminus \mathbf{T}$, or colliders that are ancestors of X_i. Therefore this path is an inducing path from $\tilde{N}_j$ to X_i. According to part (a) in Definition 2, X_i is the child of $\tilde{N}_j$ in $Aux(G_{\mathbf{T}})$.

Next, we consider the case when each recovered noise $\tilde{N}_j \in \mathcal{N}_i$ has at least two children in $G_{\mathbf{T}}$. Similar to the above argument, there is at least one root variable X_k in $\mathcal{X}_i$ that is an ancestor of X_i. This means that $\mathbf{I}_k = \mathbf{I}_i$, and the edge cannot be removed in part (b)(ii) in Definition 2. Further, for each recovered noise $\tilde{N}_j \in \mathcal{N}_i$, since (C2) does not hold for X_i, $\tilde{N}_j$ is not independent of X_i, conditioned on all observed ancestors of X_i in $G_{\mathbf{T}}$. This means that there is a path from $\tilde{N}_j$ to X_i that is not blocked by the observed ancestors. Similar to the argument above, suppose the path $\tilde{N}_j \rightarrow V_{k_1} - V_{k_2} \cdots X_i$ is not blocked. If V_{k_1} is a non-collider and is not an observed ancestor of X_i (i.e., not in the conditioning set), then there is an edge $V_{k_1} \rightarrow V_{k_2}$, and V_{k_2} is not an ancestor of X_i either. This means that V_{k_2} is also a non-collider and there is an edge $V_{k_2} \rightarrow V_{k_3}$. Repeat the same analysis on V_{k_m} for all $m = 3, 4, 5, \cdots$, the path is $\tilde{N}_j \rightarrow V_{k_1} \rightarrow V_{k_2} \rightarrow V_{k_3} \rightarrow \cdots \rightarrow X_i$. This violates the claim that V_{k_2} is not an ancestor of X_i. Therefore, V_{k_1} is a collider and is an ancestor of X_i. Then we can repeat the same analysis as above to conclude that this path is an inducing path. Therefore, there is an inducing path from each recovered noise $\tilde{N}_j \in \mathcal{N}_i$ to X_i, hence X_i is a child of $\tilde{N}_j$ in $Aux(G_{\mathbf{T}})$, according to part (b)(ii) in Definition 2.

Conclusion: We show that if an observed variable X_i is included in the output $\mathbf{K}$ of the LIT algorithm, then it must be a child of some recovered noise in $Aux(G_{\mathbf{T}})$, hence belongs to $\cup_{\tilde{N}_i \in \tilde{\mathbf{N}}_{\mathbf{T}}} Ch_{Aux(G_{\mathbf{T}})}(\tilde{N}_i)$.

C.6 Proof of Remark 2

Regarding the equivalency between the condition (C2) and the condition (C1) under causal sufficiency assumption, note that under causal sufficiency assumption, the condition in (C2) can be rewritten as follows:

(III-L) Let $\mathcal{X}_i = \{X_{i_1}, \cdots, X_{i_p}\}$, for some $p \geq 2$, be a set of all variables whose corresponding indicator sets are the same as X_i and $\mathcal{N}_i \neq \emptyset$. Then for each $j \in [p]$, $X_{i_j} \notin \mathbf{K}$ if and only if the following condition holds:

$$\exists K \subseteq [p] \setminus \{j\},\ \exists \mathbf{S} \subseteq \mathbf{S}_{i_1},\ \text{ s.t. } \tilde{N}_l \perp\!\!\!\perp X_{i_j} \,|\, \left(\cup_{k' \in K} X_{i_{k'}}\right) \cup \mathbf{S}. \tag{C2}$$

Please note that $\mathcal{N}_i$ only contains $\tilde{N}_l$ under causal sufficiency assumption (otherwise, at least two variables are descendants of each other which is impossible) and each recovered noise only has one child (i.e., their corresponding intervention target) and cannot block any path. Moreover, for any X_{i_j}, the set $K = \{i_k\}$ and $\mathbf{S} = \mathbf{S}_{i_1}$ is enough to guarantee that $\tilde{N}_l \perp\!\!\!\perp X_{i_j} | X_{i_k} \cup \mathbf{S}_{i_1}$ where $X_{i_k} \in \mathbf{T}$ and $\tilde{N}_l$ is its corresponding recovered noise. Thus, all X_{i_j} with $j \neq k$ are excluded from $\mathbf{K}$. Moreover, X_{i_k} is added to $\mathbf{K}$ as $\tilde{N}_l$ is the direct parent of X_{i_k} and cannot be d-separated by a subset of observed variables and recovered noises.

About comparing our result with [JKSB20] In the case that $\mathbf{T_O} = \mathbf{T}$, the augmented graph in [JKSB20] is constructed as follows (which we show it by $G'_\mathbf{T}$). For any pairs of environments such as E_i and E_j, a node denoted by F_{ij} is added to the original graph, and it is connected to all variables in $\mathbf{T_O}$. Please note that in our setting, all variables in $\mathbf{T}$ are changing across the environments and all F_{ij}s are directly connected to $\mathbf{T_O}$. The candidate set in [JKSB20] is the set of neighbors of F-nodes in the MAG of the augmented graph. Now, based on Theorem 2, it is just needed to show that if some $\tilde{N}_i$ is a parent of some X_j in $Aux(G_\mathbf{T})$, then some F-node is also connected to X_j in MAG of $G'_\mathbf{T}$. If X_j is $\mathbf{T_O}$, this is trivial as it should be connected to some recovered noise in $Aux(G_\mathbf{T})$ and also to some F-node in the MAG of $G'_\mathbf{T}$. Thus, suppose that the $X_i \notin \mathbf{T_O}$. In that case, based on part (a)(i) of Definition 2, there should be an inducing path between $\tilde{N}_i$ and X_j relative to $\mathbf{L}$ in $G_\mathbf{T}$. This path starts from $\tilde{N}_i$ and goes directly to its corresponding observed variable in $\mathbf{T_O}$ (without loss of generality, let us say X_i) and continues until getting to X_j in the augmented graph $G_\mathbf{T}$. Note that since there are no changing latent variables, this path only involves variables in $\mathbf{O}$ and $\mathbf{L}$. Therefore we can construct the same path on the original graph G. We denote the part of this path in the original graph by P. Now, in $G'_\mathbf{T}$, due to symmetry among F-nodes, consider any of them. This node is connected to X_i as $X_i \in \mathbf{T}$. Now, the path starting from the F-node and then going directly to X_i and continuing based on P until reaching X_j is also an inducing path in $G'_\mathbf{T}$. This shows that our identifiability result is stronger than the one in [JKSB20] when $\mathbf{T_O} = \mathbf{T}$, because of part (a)(ii) in Definition 2.