Weakly-Supervised Abstractions for Linear Additive Models

Riccardo Massidda[1] **Davide Bacciu**[1] **Sara Magliacane**[2]

[1]Department of Computer Science, Università di Pisa, Pisa, IT
[2]Informatics Institute, University of Amsterdam, Amsterdam, NL

Abstract

Causal Abstraction provides a way to summarize complex low-level models into smaller and more interpretable causal models, on which we can perform causal inference more efficiently. Despite pioneering work in learning causal abstractions, most approaches still require significant knowledge of the abstract model, e.g., the abstract graph, joint observational samples, interventional samples, or a map from low-level to abstract interventions. In this paper, we instead focus on the setting with a *weak supervision signal*: we require that the low-level model is a known linear Additive Noise Model and that we have an initial set of relevant variables, i.e., groups of low-level variables that correspond to an abstract variable. Given these relevant sets, we show that, in general, a *consistent* abstract model might not be causally sufficient even when the low-level model is causally sufficient. We then study how to extend these initial relevant sets by defining new abstract variables in an unsupervised way to preserve the causal sufficiency of the abstract model. In particular, we focus on identifying the smallest set of variables to add to a user-defined set of relevant variables to guarantee abstract sufficiency. We propose the Relevant Sufficiency Enforcer (RSE) algorithm, a weakly supervised method that, based on an initial set of relevant variables, determines the set of minimal extensions to induce abstract models that preserve causal sufficiency.

1 INTRODUCTION

In many settings, e.g., climatology [Chalupka et al., 2016], brain imaging [Dubois et al., 2020], macroeconomics [Stoker, 2010], and more, data is gathered at a lower-level of *abstraction* than the actual phenomenon of interest. In these cases, the causal effects in a large number of low-level causal variables are potentially not interesting or difficult to interpret for the domain experts, who are instead interested in estimating causal effects between a small number of interpretable high-level causal variables. An additional hurdle in these settings is that these high-level variables might not be defined a priori or evident to domain experts, so they would instead need to be learned.

Causal Abstraction [Rubenstein et al., 2017, Beckers and Halpern, 2019] is a theoretical framework to describe such scenarios by defining the necessary properties to aggregate variables while preserving interventional distributions. Intuitively, whenever two models are in an abstractive relation, intervening on the low-level and abstracting the realizations is equivalent to performing an adequate intervention directly on the abstract model. Due to this property, known as *interventional consistency* [Rubenstein et al., 2017, Zennaro et al., 2023], abstraction offers a clear advantage in estimating causal effects of multivariate treatments on a multivariate outcome. In fact, whenever these aggregations induce an abstract causal model, causal inference techniques [Pearl, 2009] can directly operate on smaller abstract models, with clear computational and interpretability advantages.

Causal sufficiency [Spirtes et al., 2000], i.e., the absence of latent confounders or selection bias in the model, simplifies causal effect estimation, so it is a desirable property that, if present in the low-level causal model, we might want to preserve also in the abstract model. This is in general not the case for any possible abstract model. For example, an abstract model might lack an abstract variable that represents the low-level variables that act as a confounder between two other abstract variables, introducing latent confounding.

In this paper, we formalize and address the problem of learning a causally sufficient abstract graph from a given causally sufficient low-level linear Additive Noise Model with a *weak supervision signal*. In particular, we focus on the scenario where we are provided an initial set of abstract

Accepted for the 1st Workshop on Causal Abstractions and Representations (CAR) at UAI 2025 (CAR 2025).

variables and, for each of abstract variable, an initial set of relevant variables, i.e., a set of the low-level variables on which each abstract variable depends directly, *without* necessarily specifying the functional form of the aggregation. We show under which conditions a set of relevant variables induces a causally sufficient abstract model. Moreover, if the initially provided set of relevant variables does not induce a sufficient abstract model, we provide a method that determines in an unsupervised way possible additions to the set of abstract variables to ensure causal sufficiency.

Example 1 (Motivating Example)**.** *Consider a neuroscientific study, which measures impulses from electrodes in distinct areas of a brain. Having a low-level causal model of the interactions between neighboring areas, practitioners might be interested in studying the causal relation between two cognitive capacities, say "attention" and "long-term memory usage". In this scenario, the supervision would consist of the areas of the brain known to be associated with these two capacities. Notably, other abstract variables confounding the provided two might need to be defined from the low-level model, without additional supervision.*

Overall, we summarize our contributions as follows:

1. First, we study the problem of extending a set of abstract variables to guarantee causal sufficiency on the abstract model. We show that the problem is underdetermined and that multiple abstract models are in general compatible with the provided weak supervision, which contains only partial information on the aggregation of the low-level variables. Therefore, we define a class of abstractions that guarantee causal sufficiency whenever the corresponding low-level model is causally sufficient, which we name "minimal sufficient extension", or MiSE (Section 4.2).
2. Then, we introduce the **RSE** algorithm, for "Relevant Sufficiency Enforcer", which correctly returns possible solutions from the minimal sufficient extension of a low-level causal graph given an initial set of relevant variables. In this way, **RSE** produces a set of candidate abstract graphs without requiring additional information on the abstract model in the form of data or domain knowledge (Section 4.3).

2 RELATED WORKS

Our proposal is significantly related to causal abstraction learning approaches, which aim to learn an abstract representation of a low-level model from data [Dyer et al., 2024, Felekis et al., 2024, Kekić et al., 2023, Massidda et al., 2024, Xia and Bareinboim, 2024, Zennaro et al., 2023]. The focus of most methods in causal abstraction learning lies in the recovery of the function mapping low-level variables to abstract variables. To this end, they require additional information such as paired low-level and abstract samples, an explicit map from low-level to high-level interventions, the abstract graph, complete knowledge over the map of low-level variables into high-level variables, or a combination of these elements, as summarized in the Table 21 in the Appendix by Felekis et al. [2024].

On the other hand, our work shares similar motivation with D'Acunto et al. [2025], who also require only partial knowledge on the map between low-level and high-level variables. While their work also relates to *causal discovery* [Spirtes et al., 2000] and tries to recover an abstraction from low-level data, we instead operate over known graphical properties of causal abstraction for linear models [Massidda et al., 2024], and define a class of possible abstract graphs compatible with the partial knowledge.

Due to our focus on graphical properties, instead of learning abstractions from data, our proposal also relates to existing works on Cluster DAGs [Anand et al., 2023] and Partial Cluster DAGs [Schooltink and Zennaro, 2024]. Both works define the existence of edges in an abstract graph according to the causal relations between low-level variables. Notably, linear causal abstraction [Massidda et al., 2024] entails different graphical conditions, as we will briefly discuss in our background (Section 3). Furthermore, in this work we study the problem of extending a partial specification of the abstract variables to ensure causal sufficiency of the abstract model, and provide a practical algorithm to this end.

3 BACKGROUND

In this section we summarize our notation, assumptions, and the definitions and results on learning linear causal abstractions for linear Additive Noise Models (ANMs) proposed by Massidda et al. [2024]. Linear ANMs are of particular interest as they are identifiable from observational data under the assumption of non-Gaussianity of the noise terms, as in the LiNGAM model [Shimizu et al., 2006].

We denote individual variables in plain upper-case, e.g., V, and sets of variables in bold, e.g., $\boldsymbol{V}$. We denote the domain of V as $\mathcal{D}(V)$ and a possible realization as $v \in \mathcal{D}(V)$. Similarly, we denote the domain of $\boldsymbol{V}$ as $\mathcal{D}(\boldsymbol{V})$ and the realization of the set as $\boldsymbol{v} \in \mathcal{D}(\boldsymbol{V})$. We focus on Structural Causal Models [Pearl, 2009], defined in our notation as:

Definition 2. *A Structural Causal Model (SCM) is a tuple* $\mathcal{M} = (\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{E}})$*, where*

- $\boldsymbol{X}$ *is a set of* d *distinct endogenous variables,*
- $\boldsymbol{E}$ *is a set of* d *distinct exogenous variables,*
- $f_X : \mathcal{D}(\mathrm{Pa}(X) \cup \{E_X\}) \to \mathcal{D}(X)$ *is a* causal mechanism*, i.e. a function that determines the value of the variable* $X \in \boldsymbol{X}$ *given its parents* $\mathrm{Pa}(X)$ *and the exogenous noise term* $E_X \in \boldsymbol{E}$*,*

- $\mathbb{P}_{\boldsymbol{E}}$ *is the joint distribution over* $\boldsymbol{E}$.

We denote as $\mathcal{M}: \mathcal{D}(\boldsymbol{E}) \to \mathcal{D}(\boldsymbol{X})$ also the push-forward function of a SCM $\mathcal{M}$, i.e., the map from the exogenous to the endogenous variables. The parental relations in the SCM then induce a directed graph, which we denote as $\mathcal{G}_{\mathcal{M}}$, that represents causal relations between variables. We assume the causal graph to be an Acyclic Directed Mixed Graph (ADMG), and denote with bidirected edges the presence of a hidden confounder between two variables, as in $X_1 \leftrightarrow X_2$.

Interventions are a fundamental concept underlying the causality literature, representing external manipulations on the variables. While there are many possible types of interventions, here we focus on *do* interventions [Pearl, 2009], which replace a structural mechanism with a constant value.

Definition 3 (Do Intervention)**.** *Let* $\mathcal{M} = (\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{E}})$ *be an SCM. A do intervention* i *is a tuple composed of a subset of variables* $\boldsymbol{V} \subseteq \boldsymbol{X}$ *and a subset of values* $\boldsymbol{v} \in \mathcal{D}(\boldsymbol{V})$, *which we denote as* $i = (\boldsymbol{V} \leftarrow \boldsymbol{v})$. *Each do intervention* i *induces an intervened SCM* $\mathcal{M}^i = (\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{f}^i, \mathbb{P}_{\boldsymbol{E}})$ *s.t.*

$$f_X^i(\cdot) = \begin{cases} f_X(\cdot) & \text{if } X \notin \boldsymbol{V} \\ v_X & \text{otherwise,} \end{cases} \tag{1}$$

for any endogenous variable $X \in \boldsymbol{X}$.

To leverage graphical properties of causal abstraction on linear ANMs [Massidda et al., 2024], we also restrict our setting to linear ANMs, i.e., SCMs in which the functions are linear and the exogenous variables are additive.

Definition 4. *A linear Additive Noise Model (ANM) is an SCM* $\mathcal{M} = (\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{E}})$ *such that for each variable* $X_j \in \boldsymbol{X}$ *it holds*

$$f_{X_j}(\boldsymbol{e}, \boldsymbol{x}) = \sum_{X_i \in \boldsymbol{X}} w_{ij} \cdot x_i + e_j, \tag{2}$$

where w_{ij} *is the element in the* i*-th row and* j*-th column of the matrix* $\mathbf{W} \in \mathbb{R}^{d \times d}$.

Given two linear ANMs, $\mathbf{T}$-abstraction [Massidda et al., 2024] specializes causal abstraction [Beckers and Halpern, 2019] by defining abstract variables as a linear transformation of low-level variables.

Definition 5 ($\mathbf{T}$-abstraction)**.** *Let* $\mathcal{L} = (\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{E}})$ *and* $\mathcal{H} = (\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{g}, \mathbb{P}_{\boldsymbol{U}})$ *be two linear ANMs with intervention sets* $\boldsymbol{I}$ *and* $\boldsymbol{J}$. *Then, given a surjective linear transformation* $\mathbf{T} \in \mathbb{R}^{d \times b}$, *where* $|\boldsymbol{X}| = d$ *and* $|\boldsymbol{Y}| = b$, $\mathcal{H}$ *is a* $\mathbf{T}$*-abstraction of* $\mathcal{L}$ *if and only if there exists a linear transformation* $\mathbf{S} \in \mathbb{R}^{d \times b}$ *such that, for any abstract intervention* $j \in \boldsymbol{J}$, *there exists a low-level intervention* $i \in \boldsymbol{I}$ *such that, for any exogenous configuration* $\boldsymbol{e} \in \mathcal{D}(\boldsymbol{E})$,

$$\tau(\mathcal{L}^i(\boldsymbol{e})) = \mathcal{H}^j(\gamma(\boldsymbol{e})) \tag{3}$$

where $\mathcal{L}^i$ *and* $\mathcal{H}^j$ *represent the push-forward functions of the intervened models,* $\tau(\boldsymbol{x}) = \mathbf{T}^\top \boldsymbol{x}$ *and* $\gamma(\boldsymbol{e}) = \mathbf{S}^\top \boldsymbol{e}$.

Similarly to Massidda et al. [2024], we assume that both low and high-level causal models are faithful, i.e., in each model a conditional indepedence in the distribution implies a corresponding d-separation in the causal graph, and that the abstraction function does not yield cancelling paths from the low-level to the abstract variables.

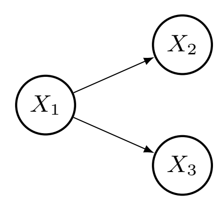
Assumption 6. *Let* $\tau: \mathcal{D}(\boldsymbol{X}) \to \mathcal{D}(\boldsymbol{Y})$ *be a surjective function and* $\mathcal{L} = (\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{E}})$ *and* $\mathcal{H} = (\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{g}, \mathbb{P}_{\boldsymbol{U}})$ *be two linear ANMs. Then, we assume that* $\mathcal{L}$ *and* $\mathcal{H}$ *are faithful and that the joint SCM* $\mathcal{M}$ *defined as*

$$\tau \circ \mathcal{L} = (\boldsymbol{X} \cup \boldsymbol{Y}, \boldsymbol{E}, \boldsymbol{f} \cup \{\tau_j\}_{Y_j \in \boldsymbol{Y}}, \mathbb{P}_{\boldsymbol{E}})$$

does not have cancelling paths.

While the requirement on the faithfulness of the causal models is standard, assuming that the joint SCM does not yield cancelling paths requires more careful consideration, as we show in the following example. The example shows the pathological case in which the influence of a low-level variable is cancelled on an abstract variable, even when both the low-level and the high-level models are faithful.

Example 7 (Cancelling Paths Assumption Violation)**.** *Let* $\mathcal{L} = (\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{E}})$ *be a linear ANM with the causal graph*



where we assume all causal relations to have unitary weight. Then, let $\mathcal{H} = (\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{g}, \mathbb{P}_{\boldsymbol{U}})$ *be a linear ANM with two disconnected variables* Y_1 *and* Y_2, *as in the causal graph*

Then, we can prove that given the following transformation

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix}^\top, \tag{4}$$

$\mathcal{H}$ *is a valid* $\mathbf{T}$*-abstraction of* $\mathcal{L}$ *under all possible abstract interventions. Intuitively, this abstract model is not a faithful or interpretable representation of the concrete model. The issue arises as there is a canceling path from* X_1 *to* Y_2 *through* X_2 *and* X_3 *in the joint SCM* $\tau \circ \mathcal{L}$, *even if both* $\mathcal{L}$ *and* $\mathcal{H}$ *follow the Causal Faithfulness assumption.*

A $\mathbf{T}$-abstraction considers different types of low-level variables: *relevant*, *block*, and *ignorable* variables. Intuitively, *relevant variables* are the subset of low-level variables from which we can reconstruct high-level realizations. Similarly, *block variables* are the subset of low-level variables whose exogenous terms are necessary to reconstruct realizations of the high-level exogenous term. Finally, any variable that is neither relevant or block is defined as *ignorable*.

Definition 8 (Relevant Variables). *Let $\mathbf{T} \in \mathbb{R}^{d \times b}$ be a linear abstraction function from a set of low-level variables $\boldsymbol{X}$ to a set of high-level variables $\boldsymbol{Y}$. For each abstract variable $Y_j \in \boldsymbol{Y}$ we define its relevant variables as $\boldsymbol{R}_{Y_j} = \{X_i \mid t_{ij} \neq 0\}$.*

Definition 9 (Block Variables). *Let $\mathbf{T} \in \mathbb{R}^{d \times b}$ be a linear abstraction function from a set of d low-level variables $\boldsymbol{X}$ to a set of high-level variables $\boldsymbol{Y}$. For each abstract variable $Y_j \in \boldsymbol{Y}$, we define its block variables as $\boldsymbol{B}_{Y_j} = \{X_i \mid s_{ij} \neq 0\}$.*

Example 10 (Relevant, Block, and Ignorable Variables). *Let $\mathcal{H}$ and $\mathcal{L}$ be two linear ANMs respectively on variables $\boldsymbol{Y}$ and $\boldsymbol{X}$. We consider an endogenous abstraction linear transformation $\mathbf{T} \in \mathbb{R}^{d \times b}$ from $\boldsymbol{X}$ to $\boldsymbol{Y}$, and an exogenous abstraction linear transformation $\mathbf{S} \in \mathbb{R}^{d \times b}$ from $\boldsymbol{E}$ to $\boldsymbol{U}$. Then, given the following coefficients,*

$$\mathbf{T} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}^\top \tag{5}$$

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}^\top . \tag{6}$$

we can define the relevant variables, block variables, and abstraction vectors as

$$\boldsymbol{R}_{Y_1} = \{X_1, X_2\}, \tag{7}$$

$$\boldsymbol{R}_{Y_2} = \{X_5, X_6\}, \tag{8}$$

$$\boldsymbol{B}_{Y_1} = \{X_1, X_2\}, \tag{9}$$

$$\boldsymbol{B}_{Y_2} = \{X_3, X_4, X_5, X_6\} \tag{10}$$

$$\boldsymbol{t}_1 = \begin{bmatrix} 1 & 1 \end{bmatrix} \tag{11}$$

$$\boldsymbol{t}_2 = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix} \tag{12}$$

$$\boldsymbol{s}_1 = \begin{bmatrix} 1 & 1 \end{bmatrix} \tag{13}$$

$$\boldsymbol{s}_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}. \tag{14}$$

Consequently, the only ignorable variable is X_7.

Massidda et al. [2024] define the notion of $\mathbf{T}$-direct path as a path in the low-level graph that does not cross any relevant variable apart from the source and the target of the path. These paths will be central to the definition of causal sufficient abstractions that we will develop in this paper. To highlight the dependence of these paths on the set of relevant variables and not on the particular coefficients in $\mathbf{T}$, we equivalently refer to them as $\boldsymbol{R}$-direct paths.

Definition 11 ($\boldsymbol{R}$-direct Path). *Let $\mathcal{H}$ and $\mathcal{L}$ be two linear ANMs respectively on variables $\boldsymbol{Y}$ and $\boldsymbol{X}$. Then, given the low-level causal graph $\mathcal{G}_{\mathcal{L}}$ and a set of relevant variables $\boldsymbol{R}$, we say that a directed path $\boldsymbol{p}$ from a node X_{p_1} to a node X_{p_n} is $\boldsymbol{R}$-direct if and only if any variable $X_i \in \boldsymbol{p}_{2:n-1}$ in the middle of the path is not relevant to any abstract variable $Y_j \in \boldsymbol{Y}$. We denote the existence of a $\boldsymbol{R}$-direct path from X_i to X_j in the low-level graph as $X_i \xrightarrow{\boldsymbol{R}} X_j \in \mathcal{G}_{\mathcal{L}}$.*

Similarly to Anand et al. [2023], Massidda et al. [2024], Schooltink and Zennaro [2024], we define the induced graph by studying $\boldsymbol{R}$-direct paths between relevant variables of different abstract variables. Then, we define the relevant sets to be *graphically consistent* if they satisfy the necessary graphical conditions of linear $\mathbf{T}$-abstraction. This requirement is stronger than similar conditions for Cluster DAGs [Anand et al., 2023] and Partial Cluster DAGs [Schooltink and Zennaro, 2024], as it requires multiple $\boldsymbol{R}$-direct paths.

Definition 12 (Induced Abstract Graph). *Given a low-level ANM $\mathcal{L}$ on variables $\boldsymbol{X}$, a set of abstract variables $\boldsymbol{Y}$, and a set of relevant variables $\boldsymbol{R} = \{\boldsymbol{R}_Y\}_{Y \in \boldsymbol{Y}}$, we define the induced graph on $\boldsymbol{Y}$ as the graph $\mathcal{G}_{\boldsymbol{R}}$ such that*

$$\begin{aligned} & Y_1 \to Y_2 \in \mathcal{G}_{\boldsymbol{R}} \\ \iff & \exists X_1 \in \boldsymbol{R}_{Y_1}, X_2 \in \boldsymbol{R}_{Y_2} \text{ s.t. } X_1 \xrightarrow{\boldsymbol{R}} X_2 \in \mathcal{G}_{\mathcal{L}}. \end{aligned} \tag{15}$$

Definition 13 (Graphical Consistency of Relevant Sets). *Given a set of abstract variables $\boldsymbol{Y}$ and a low-level causal graph $\mathcal{G}_{\mathcal{L}}$, a set of relevant variables $\boldsymbol{R} = \{\boldsymbol{R}_Y\}_{Y \in \boldsymbol{Y}}$ is graphically consistent whenever*

$$\begin{aligned} & Y_1 \to Y_2 \in \mathcal{G}_{\boldsymbol{R}} \\ \implies & \forall X_1 \in \boldsymbol{R}_{Y_1}, \exists X_2 \in \boldsymbol{R}_{Y_2} \text{ s.t. } X_1 \xrightarrow{\boldsymbol{R}} X_2 \in \mathcal{G}_{\mathcal{L}}. \end{aligned} \tag{16}$$

Finally, we report a graphical characterization of the block variables, which we will leverage in this paper to determine them from the low-level graph $\mathcal{G}_{\mathcal{L}}$ and the relevant variables $\boldsymbol{R}$ only, regardless of the abstraction coefficients $\mathbf{T}, \mathbf{S}$.

Lemma 14 (Lemma 3 in [Massidda et al., 2024]). *Let $\mathcal{H}$ be a $\mathbf{T}$-abstraction of $\mathcal{L}$, where $\mathcal{H}$ and $\mathcal{L}$ are two linear ANMs respectively on variables $\boldsymbol{Y}$ and $\boldsymbol{X}$. Then, for any abstract variable $Y \in \boldsymbol{Y}$, it holds $X \in \boldsymbol{B}_Y$ if and only if*

1. *$X \in \boldsymbol{R}_Y$, or*
2. *X is not relevant for any $Y' \in \boldsymbol{Y}$ and exists $X' \in \boldsymbol{R}_Y$ such that $X \xrightarrow{\boldsymbol{R}} X'$.*

4 CAUSALLY SUFFICIENT AGGREGATION OF LOW-LEVEL CAUSAL VARIABLES

Under the assumption that the low-level model is known, we propose an approach to determine a graphically consistent and causally sufficient aggregation of the low-level variables that respects the necessary graphical conditions of causal abstraction. This scenario is particularly relevant in the context of linear Additive Noise Models, where the low-level model can be recovered from data whenever it has additive non-Gaussian noise [Shimizu et al., 2006] or Gaussian and homoskedastic noise [Loh and Bühlmann, 2014]. Further, this approach holds whenever the model is known by domain-knowledge, sufficient interventional data is available for causal discovery, or a simulator for the model exists [Kekić et al., 2023].

Therefore, our problem consists of determining the abstract causal graph $\mathcal{G}_{\mathcal{H}}$ from the low-level graph $\mathcal{G}_{\mathcal{L}}$. In practical applications, we can expect a practitioner to define the relevant sets for at least two variables, i.e., the treatment and the outcome of interest. However, we define our procedure for a more general scenario, where the user defines at least $b \geq 2$ abstract variables and their corresponding low-level relevant variables. In this way, the treatment-to-outcome estimation immediately results as a special case, where $b = 2$. We do *not* require the user to define the functional form of the abstraction function, but only the two groups of variables on which two abstract variables should depend.

4.1 SUFFICIENCY OF RELEVANT VARIABLES

Causal effects are always identifiable in *causal sufficient* settings, i.e., when there are no latent confounders or selection bias [Spirtes et al., 2000, Pearl, 2009]. Abstracting causal models could potentially introduce violations of causal sufficiency, since two abstract variables might be confounded by latent factors, if we fail to abstract from the low-level model also their common causes. As we prove in the following proposition, whenever the low-level model is causally sufficient, a corresponding abstract model is *not* causally sufficient if and only if its relevant variables lead to overlapping blocks. For this reason, we refer to this graphical condition as the *sufficiency* of the relevant sets. To determine blocks from the low-level graph and the relevant variables, we recall the result reported in Lemma 14.

Definition 15 (Sufficient Relevant Sets)**.** *Given a set of abstract variables $\boldsymbol{Y}$ and a low-level causal graph $\mathcal{G}_{\mathcal{L}}$, a set of relevant variables $\boldsymbol{R} = \{\boldsymbol{R}_Y\}_{Y \in \boldsymbol{Y}}$ is sufficient whenever*

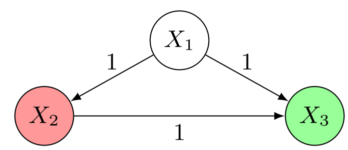
$$\forall Y_1, Y_2 \in \boldsymbol{Y} \text{ s.t. } \boldsymbol{B}_{Y_1} \cap \boldsymbol{B}_{Y_2} = \emptyset, \tag{17}$$

where $\boldsymbol{B} = \{\boldsymbol{B}_Y\}_{Y \in \boldsymbol{Y}}$ is the set of block variables induced by $\boldsymbol{R}$ on $\mathcal{G}_{\mathcal{L}}$.

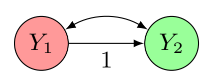
With this new definition, we can now rephrase Lemma 4 from Massidda et al. [2024] to prove the relation between the sufficiency of the relevant variables and causal sufficiency on the abstract model. We show an example of violation of abstract sufficiency, due to the failure to abstract a concrete variable that becomes a latent confounder.

Proposition 16 (Abstract Causal Sufficiency)**.** *Let $\mathcal{H}$ be a $\mathbf{T}$-abstraction of $\mathcal{L}$. Then, if $\mathcal{L}$ is causally sufficient, the abstract ANM $\mathcal{H}$ is causally sufficient only if the relevant sets $\boldsymbol{R}$ are sufficient relevant sets.*

Example 17 (Violation of Abstract Sufficiency)**.** *Let $\mathcal{L}$ be a low-level model with the following graph*



Then, we can abstract the model with two abstract variables Y_1, Y_2 such that $\boldsymbol{R}_{Y_1} = \{X_2\}$ and $\boldsymbol{R}_{Y_2} = \{X_3\}$, leading to the following overlapping *blocks $\boldsymbol{B}_{Y_1} = \{X_1, X_2\}$ and $\boldsymbol{B}_{Y_2} = \{X_2, X_3\}$, and the following abstract graph*



Notably, the abstract model is provably a $\mathbf{T}$-abstraction of the low-level model for the transformation

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^\top. \tag{18}$$

The causal effect $\mathbb{P}_{Y_2|\mathrm{Do}(Y_1)}$ is not identifiable from the abstract model. However, by using the corresponding low-level relevant variables, the causal effect $\mathbb{P}_{\boldsymbol{R}_{Y_2}|\mathrm{Do}(\boldsymbol{R}_{Y_1})} = \mathbb{P}_{X_3|\mathrm{Do}(X_2)}$ is identifiable, since $\{X_1\}$ is a valid adjustment set. The issue arises due to the fact that the adjustment set, namely $\{X_1\}$, is not abstracted.

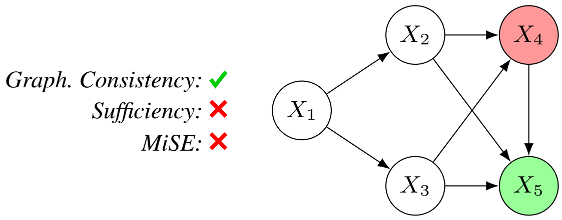
4.2 MINIMAL SUFFICIENT EXTENSION

As the causal insufficiency on the abstract model depends on the overlapping of the blocks, one possible strategy is to add more abstract variables containing the variables in the intersections. This problem is under-determined, as by partitioning intersected variables differently, we could introduce novel abstract variables in several ways. To enable the computational advantages and follow the intuition that an abstract model should somehow "compress" the low-level one, we also want to avoid inserting superfluous abstract variables. Therefore, we propose a notion of *minimality* to extend relevant variables and guarantee causal sufficiency on the abstract model. As we show in Example 19, multiple relevant sets can induce sufficient abstract models. Therefore, we define a set containing all "minimal" additions of the relevant variables, which we name *minimal sufficient extension (MiSE)*.

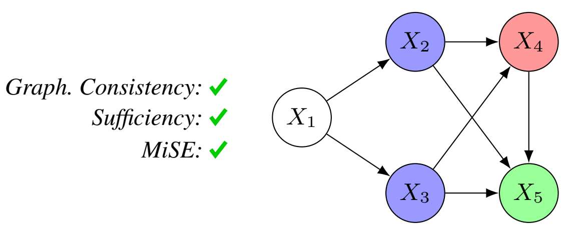
Definition 18 (Minimal Sufficient Extension)**.** *Given a set of relevant variables $\boldsymbol{R}$ and a causal graph $\mathcal{G}_{\mathcal{L}}$, the minimal sufficient extension* $\mathrm{MiSE}\left(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}}\right)$ *is a set such that, for each relevant set $\boldsymbol{R}' \in \mathrm{MiSE}\left(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}}\right)$, it holds that*

1. $\boldsymbol{R}' \supseteq \boldsymbol{R}$,
2. $\boldsymbol{R}'$ *is sufficient and graphically consistent on* $\mathcal{G}_{\mathcal{L}}$, *and*
3. *any* $\boldsymbol{R} \subset \boldsymbol{R}'' \subset \boldsymbol{R}'$ *is not sufficient on* $\mathcal{G}_{\mathcal{L}}$.

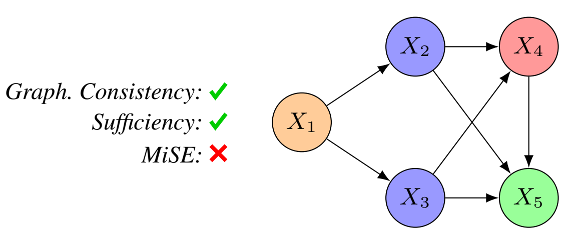
Example 19 (MiSE)**.** *Consider the following low-level graph with relevant sets* $\boldsymbol{R}_{Y_1} = \{X_4\}$ *and* $\boldsymbol{R}_{Y_2} = \{X_5\}$. *The resulting abstract graph is not causally sufficient as* X_1, X_2, X_3 *are in both blocks* $\boldsymbol{B}_{Y_1}$ and $\boldsymbol{B}_{Y_2}$.



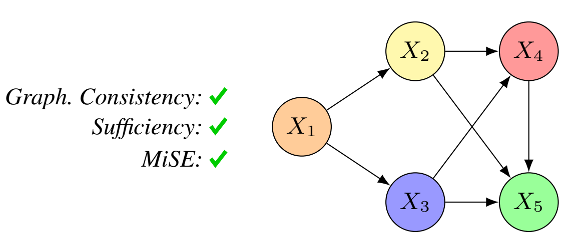
Introducing a new abstract variable Y_3 *with relevant set* $\{X_3, X_4\}$ *produces a causally sufficient abstract model, as* $X_1 \in \boldsymbol{B}_{Y_3}$ *and not in other blocks. Given the conditions in Definition 18,* $\boldsymbol{R}' = \{\{4\}, \{5\}, \{3, 4\}\} \in \text{MiSE}(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}})$.



On the other hand, a further abstract variable Y_4 *with relevant set* $\boldsymbol{R}_{Y_4} = \{X_1\}$ *would still lead to a graphically consistent and causally sufficient relevant set* $\boldsymbol{R}''$. *However, since* $\boldsymbol{R}'' \supset \boldsymbol{R}'$, *it is not minimal. Intuitively, we do not need* Y_4 *to adjust the causal effect among the abstract variables introduced by the original relevant set* $\boldsymbol{R}$.

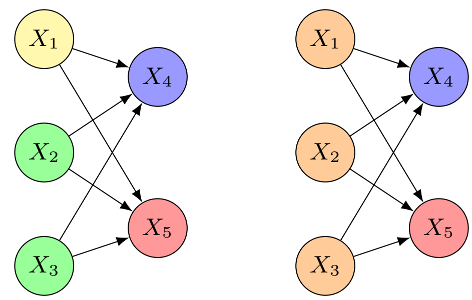


Notably, if X_2 *and* X_3 *are assigned to two distinct abstract variables,* X_1 *would also need to become relevant to avoid being in the intersection of the new abstract variables.*



As all the relevant sets in the *minimal sufficient extension* are graphically consistent and sufficient, we might be tempted to only select the one with the fewest number of abstract variables and discard the remaining. However, while graphically correct, this might not lead to a consistent abstraction as we will provide some evidence in the following example.

Example 20 (MiSE Parametrical Inconsistency)**.** *Consider the following two clustering of the low-level variables* X_1, X_2, X_3*: (i.) two abstract variables* Y_3 *and* Y_4 *with* $\boldsymbol{R}_{Y_3} = \{X_1\}$ *and* $\boldsymbol{R}_{Y_4} = \{X_2, X_3\}$, *(ii.) one abstract variable* Y_5 *with* $\boldsymbol{R}_{Y_5} = \{X_1, X_2, X_3\}$. *We assume that (i.) is the "ground-truth" abstraction. We represent the two sets of relevant variables graphically with different colors as follows:*



Intuitively, the abstractions of X_4 *and* X_5 *are a function of the abstraction of* $\tau_3(X_1)$ *and of* $\tau_4(X_2, X_3)$, *i.e.,*

$$\tau_1(X_4) = f_1(\tau_3(X_1), \tau_4(X_2, X_3)),$$
$$\tau_2(X_5) = f_2(\tau_3(X_1), \tau_4(X_2, X_3)).$$

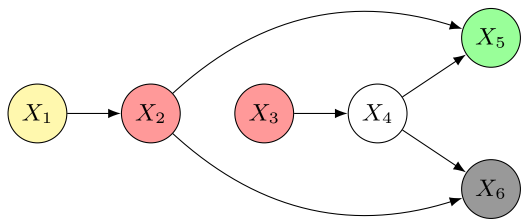
In general, without further assumptions, there might not exist two functions g_1, g_2 *such that*

$$\tau_1(X_4) = g_1(\tau_5(X_1, X_2, X_3)),$$
$$\tau_2(X_5) = g_2(\tau_5(X_1, X_2, X_3)).$$

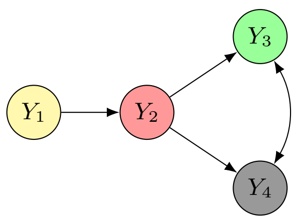
Therefore, while both are in the minimal sufficient extension *of* $\boldsymbol{R} = \{\{X_4\}, \{X_5\}\}$ *on the low-level graph, only for one of them we will be able to retrieve abstract parameters.*

Finally, adding new variables is not always possible, as there are scenarios where no further additions can render the abstract model causally sufficient without breaking its graphical consistency, as in the next example. In other words, we show that some low-level models cannot be further abstracted in a sufficient way, hence the MiSE is empty.

Example 21 (Empty MiSE)**.** *Let* $\mathcal{H}$ *be an abstract SCM with four variables and* $\mathcal{L}$ *be a causally sufficient low-level SCM with six variables. We represent the low-level causal graph as follows, where we employ colours to denote relevant variables of distinct abstract variables, as in* $\boldsymbol{R}_{Y_1} = \{X_1\}$, $\boldsymbol{R}_{Y_2} = \{X_2, X_3\}$, $\boldsymbol{R}_{Y_3} = \{X_5\}$, *and* $\boldsymbol{R}_{Y_4} = \{X_6\}$.



This choice of relevant variables is graphically consistent, as it respects the conditions from Definition 13. However, it is not sufficient, as the low-level variable X_4 is shared among the block of Y_3 and Y_4, as shown in the following abstract graph.



To ensure causal sufficiency of the abstract model, we have to assign X_4 to either one of the existing relevant sets or to the one of a new abstract variable Y_5. However, while ensuring sufficiency, any of these operations would break the graphical consistency of the relevant sets (Appendix A). Therefore this low-level model cannot be abstracted in any non-trivial way, while ensuring abstract causal sufficiency.

4.3 RELEVANT SUFFICIENCY ENFORCEMENT

Building on the previous examples, we address the problem of ensuring sufficiency of a resulting abstract model starting from a graphically consistent set of relevant variables. To this end, we introduce the **`RSE`** algorithm, for Relevant Sufficiency Enforcer (Algorithm 1). Given a set of relevant sets and a causally sufficient low-level graph, the algorithm provably returns only extensions of the relevant sets that are graphically consistent, sufficient, and *minimal*, as defined in Definition 18. Formally, from a relevant set $\boldsymbol{R}$ defining a set of abstract variables $\boldsymbol{Y}$ on a graph $\mathcal{G}_{\mathcal{L}}$, **`RSE`** returns a set of solutions $\boldsymbol{S}$, where each solution $\boldsymbol{R}' \in \boldsymbol{S}$ is in the *minimal sufficient extension* of $\boldsymbol{R}$ on the low-level graph $\mathcal{G}_{\mathcal{L}}$. While we postulate also the completeness of the **`RSE`** algorithm, in this work we only prove its correctness. Therefore, we leave the proof of $\texttt{RSE}(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}}) = \text{MiSE}(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}})$ to future work.

Theorem 22 (RSE Correctness)**.** *Let $\boldsymbol{R}$ be a set of graphically consistent relevant variables on a low-level graph $\mathcal{G}_{\mathcal{L}}$, and $\boldsymbol{S} \leftarrow \texttt{RSE}(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}})$ be the output of the RSE algorithm. Then, it holds that $\boldsymbol{S} \subseteq \text{MiSE}(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}})$.*

Proof. We provide the proof in Appendix B.2. □

Algorithm 1 Relevant Sufficiency Enforcer (RSE)

Input: Relevant set $\boldsymbol{R}$ and low-level graph $\mathcal{G}_{\mathcal{L}}$ on $\boldsymbol{X}$.
Output: $\boldsymbol{S} \subseteq$ Minimal Sufficient Extension of $\boldsymbol{R}$ on $\mathcal{G}_{\mathcal{L}}$.

1: **if** not $\text{GraphicalConsistency}(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}})$ **then**
2:     **return** $\emptyset$
3: **end if**
4: $\boldsymbol{Y} \leftarrow \{1, \ldots, |\boldsymbol{R}|\}$
5: $\boldsymbol{B} \leftarrow \text{Blocks}(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}})$
6: $\boldsymbol{I} \leftarrow \left\{\boldsymbol{V} \in \mathcal{P}(\boldsymbol{Y}) \mid \cap_{Y_j \in \boldsymbol{V}} \boldsymbol{B}_{Y_j} \neq \emptyset\right\}$
7: **if** $\boldsymbol{I} = \emptyset$ **then**
8:     **return** $\boldsymbol{R}$
9: **end if**
10: $\boldsymbol{Z} \leftarrow \boldsymbol{X} \setminus \cup_{\boldsymbol{V} \in \boldsymbol{I}} \cap_{Y_j \in \boldsymbol{V}} \boldsymbol{B}_{Y_j}$
11: $\boldsymbol{I}_0 \leftarrow \text{Any}(\boldsymbol{I})$
12: $\boldsymbol{W} \leftarrow \text{NewRelevants}(\boldsymbol{I}_0, \boldsymbol{Z}, \boldsymbol{R}, \mathcal{G}_{\mathcal{L}})$ (Eq. 19)
13: $\boldsymbol{Q}, \boldsymbol{L} \leftarrow \text{TargetVariables}(\boldsymbol{I}_0, \boldsymbol{Z}, \boldsymbol{R}, \mathcal{G}_{\mathcal{L}})$ (Eq. 20-21)
14: $\hat{X} \leftarrow \text{Any}(\{X \in \boldsymbol{W} \mid \boldsymbol{L}_X = \emptyset\})$
15: $\boldsymbol{M} \leftarrow \text{MatchingTargets}(\hat{X}, \boldsymbol{W}, \boldsymbol{Q}, \boldsymbol{L})$ (Eq. 22)
16: $\boldsymbol{S} \leftarrow \emptyset$
17: **for** $\boldsymbol{P} \in \text{Partitions}(\boldsymbol{M})$ **do**
18:     $\boldsymbol{S} \leftarrow \boldsymbol{S} \cup \text{RSE}(\boldsymbol{R} \cup \boldsymbol{P}, \mathcal{G}_{\mathcal{L}})$
19: **end for**
20: **return** $\boldsymbol{S}$

At each call, **`RSE`** computes the block variables (Line 5) and the set $\boldsymbol{I}$ containing all subsets of abstract variables whose low-level blocks have a non empty intersection (Line 6). Then, it arbitrarily selects one of the intersections $\boldsymbol{I}_0 \in \boldsymbol{I}$, and identifies which of the low-level variables $\boldsymbol{W} \subset \boldsymbol{X}$ in the intersection must become relevant to ensure abstract sufficiency (Line 11). A low-level variable $X \in \boldsymbol{X}$ is flagged as new relevant variable, if it is in the intersection $\boldsymbol{I}_0$ but has at least an outgoing edge $X \to X'$ to a variable $X' \in \boldsymbol{Z}$, where $\boldsymbol{Z}$ is the set of low-level variables *not* in any intersection (Line 6). Formally, we define the set $\boldsymbol{W} \subseteq \boldsymbol{X}$ of new relevant variables as

$$\boldsymbol{W} = \left\{X \in \cap_{Y_j \in \boldsymbol{I}_0} \boldsymbol{B}_{Y_j} \mid \exists X' \in \boldsymbol{Z}. X \to X' \in \mathcal{G}_{\mathcal{L}}\right\}. \tag{19}$$

As we prove in the following lemma, which will is central for the proof of the correctness of **`RSE`** (Theorem 22), all variables in $\boldsymbol{W}$ must be assigned to ensure sufficiency.

Lemma 23. *Let $\boldsymbol{R}$ be a set of graphically consistent relevant variables on a low-level causal graph $\mathcal{G}_{\mathcal{L}}$ and $\boldsymbol{W}$ be the set of new relevant variables identified by the **`RSE`** algorithm. Then, for all possible extensions $\boldsymbol{R}' \supseteq \boldsymbol{R}$, if it exists a low-level variable $X \in \boldsymbol{W}$ not in any relevant set $\boldsymbol{R}'_{Y_j} \in \boldsymbol{R}'$, then $\boldsymbol{R}'$ is either not sufficient or not graphically consistent on $\mathcal{G}_{\mathcal{L}}$.*

Proof. We provide the proof in Appendix B.1. □

Therefore, once identified the new low-level relevants $\boldsymbol{W}$, we have to assign them to new abstract variables to solve the intersections of the blocks. One trivial strategy would be to enumerate all possible partitions of $\boldsymbol{W}$ into up to $|\boldsymbol{W}|$ new abstract variables and propagate the solution, letting **RSE** check graphical consistency and sufficiency in the recursive calls. In general, it is not possible to cluster all of them for a unique new abstract variable, as graphical consistency requires that, if a variable $X_1 \in \boldsymbol{W}$ has a $\boldsymbol{R}$-direct path to a relevant $X_2 \in \boldsymbol{R}_Y$ of an abstract variable Y, then also any other $X'_1 \in \boldsymbol{W}$ must have a $\boldsymbol{R}$-direct path to at least one relevant $X'_2 \in \boldsymbol{R}_Y$. Therefore, to exploit this condition, we study the outgoing paths from each new relevant $X \in \boldsymbol{W}$ and return the set of targeted abstract variables $\boldsymbol{Q}_X \subset \boldsymbol{Y}$ and the set of targeted new relevant variables $\boldsymbol{L}_X \subset \boldsymbol{W}$ (Line 11). To define these two sets, we slightly extend the notation of $\boldsymbol{R}$-direct paths to $(\boldsymbol{R}, \boldsymbol{W})$-direct paths, where we denote as $X \xrightarrow{\boldsymbol{R},\boldsymbol{W}} X'$ the existence of a path between two variables $X, X' \in \boldsymbol{X}$ that is not crossed by (i.) any relevant variable in $\boldsymbol{R}$ or (ii.) any candidate relevant in $\boldsymbol{W}$. Therefore, we can define the target sets of the new relevant $X \in \boldsymbol{W}$ as:

$$\boldsymbol{Q}_X = \{Y \in \boldsymbol{Y} \mid \exists X' \in \boldsymbol{B}_Y \cap \boldsymbol{Z}. X \to X'\}, \tag{20}$$

$$\boldsymbol{L}_X = \left\{X' \in \boldsymbol{W} \mid X \xrightarrow{\boldsymbol{R},\boldsymbol{W}} X'\right\}. \tag{21}$$

At this point, for each new relevant variable $X \in \boldsymbol{W}$, we know which abstract variables it surely targets ($\boldsymbol{Q}_X \subset \boldsymbol{Y}$) and which low-level other new relevants it has a $(\boldsymbol{R}, \boldsymbol{W})$-direct path to ($\boldsymbol{L}_X \subset \boldsymbol{W}$). By using these sets, we can determine whether two variables in set of new relevants $\boldsymbol{W}$ can be clustered together or not. In practice, we select one of the variables $\hat{X} \in \boldsymbol{W}$ (Line 13) and find all the variables satisfying the following recursive definition:

$$\boldsymbol{M}_{\hat{X}} = \left\{X \in \boldsymbol{W} \mid \boldsymbol{Q}_X = \boldsymbol{Q}_{\hat{X}} \wedge \boldsymbol{L}_X \subseteq \boldsymbol{M}_{\hat{X}}\right\}. \tag{22}$$

5 CONCLUSION

In this work, we tackled the problem of inducing a causally sufficient abstract model from a known and causally sufficient low-level linear Additive Noise Model (ANM). In particular, we focused on the weakly supervised scenario where, for a subset of abstract variables, we know which low-level variables they are aggregating, without further information on their functional form. To this end, we introduced the set of *minimal sufficient extension* (MiSE), to formalize how to extend the provided partial knowledge to ensure causal sufficiency on the abstract model, without introducing superfluous variables. Then, we defined the Relevant Sufficiency Enforcer (**RSE**) algorithm to compute these extensions in practice. Finally, we proved the correctness of **RSE**, i.e, that it returns only solutions in the minimal sufficient extension of the provided partial information on the abstract variables.

This paper highlights how the problem of inducing abstract models is essentially undetermined when only partial knowledge on the abstract variables is provided. However, it also remarks how graphical conditions guaranteeing causal sufficiency can reduce the problem to a well-defined class of solutions, the MiSE, which we can explore algorithmically, through the **RSE** algorithm.

The paper opens up to several future directions. First, as we hinted in Example 20, while respecting the graphical conditions, solutions contained in the MiSE might be incompatible with the actual parameters of the low-level model. A promising direction then lies in combining the exploration of the MiSE with existing causal abstraction learning approaches to further restrict the class according to low-level samples. Furthermore, we remark that this work directly deals with graphical properties known to characterize linear abstraction over linear ANMs. It is still an open question whether similar conditions also hold without the linearity assumption; hence, if the concept of MiSE and the **RSE** algorithm could also be applied to non-linear models in a causal abstraction relation. Finally, while we have proven the correctness of the **RSE** algorithm, proving its completeness, i.e., whether **RSE** returns the whole MiSE class, constitutes an interesting theoretical result open for future works.

Acknowledgements

This research was partially supported by TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No 952215.

References

Tara V Anand, Adele H Ribeiro, Jin Tian, and Elias Bareinboim. Causal effect identification in cluster dags. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12172–12179, 2023.

Sander Beckers and Joseph Y Halpern. Abstracting causal models. In *Proceedings of the aaai conference on artificial intelligence*, volume 33 of *AAAI'19/IAAI'19/EAAI'19*, pages 2678–2685. AAAI Press, 2019. ISBN 978-1-57735-809-1. doi: 10.1609/aaai.v33i01.33012678. URL https://doi.org/10.1609/aaai.v33i01.33012678.

Krzysztof Chalupka, Tobias Bischoff, Pietro Perona, and Frederick Eberhardt. Unsupervised discovery of el nino using causal feature learning on microlevel climate data. In Alexander T. Ihler and Dominik Janzing, editors, *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, pages 72–81. AUAI Press, 2016. URL http://auai.org/uai2016/proceedings/papers/11.pdf.

Gabriele D'Acunto, Fabio Massimo Zennaro, Yorgos Felekis, and Paolo Di Lorenzo. Causal abstraction learning based on the semantic embedding principle. *arXiv preprint arXiv:2502.00407*, 2025.

Julien Dubois, Hiroyuki Oya, Julian Michael Tyszka, Matthew A. Howard, Frederick Eberhardt, and Ralph Adolphs. Causal mapping of emotion networks in the human brain: Framework and initial findings. *Neuropsychologia*, 145, 2020.

Joel Dyer, Nicholas George Bishop, Yorgos Felekis, Fabio Massimo Zennaro, Ani Calinescu, Theodoros Damoulas, and Michael J Wooldridge. Interventionally consistent surrogates for complex simulation models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Yorgos Felekis, Fabio Massimo Zennaro, Nicola Branchini, and Theodoros Damoulas. Causal optimal transport of abstractions. In *Causal Learning and Reasoning*, pages 462–498. PMLR, 2024.

Armin Kekić, Bernhard Schölkopf, and Michel Besserve. Targeted reduction of causal models. *arXiv preprint arXiv:2311.18639*, 2023.

Po-Ling Loh and Peter Bühlmann. High-dimensional learning of linear causal networks via inverse covariance estimation. *The Journal of Machine Learning Research*, 15 (1):3065–3105, 2014.

Riccardo Massidda, Sara Magliacane, and Davide Bacciu. Learning causal abstractions of linear structural causal models. In *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024. URL `https://openreview.net/forum?id=XlFqI9TMhf`.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Paul K. Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M. Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. In Gal Elidan and Kristian Kersting, editors, *Proceedings of the 33rd Conference on Uncertainty in A rtificial Intelligence (UAI-17)*, volume abs/1707.00819. AUAI Press, August 2017. URL `http://auai.org/uai2017/proceedings/papers/11.pdf`.

Wilem Schooltink and Fabio Massimo Zennaro. Aligning graphical and functional causal abstractions. *arXiv preprint arXiv:2412.17080*, 2024.

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

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. Adaptive computation and machine learning. MIT press, 2000.

Thomas M. Stoker. Aggregation (econometrics). In Steven N. Durlauf and Lawrence E. Blume, editors, *Macroeconometrics and Time Series Analysis*, pages 1–14. Palgrave Macmillan UK, 2010. ISBN 978-0-230-28083-0. doi: 10.1057/9780230280830_1. URL `https://doi.org/10.1057/9780230280830_1`.

Kevin Xia and Elias Bareinboim. Neural causal abstractions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20585–20595, 2024.

Fabio Massimo Zennaro, Máté Drávucz, Geanina Apachitei, W Dhammika Widanage, and Theodoros Damoulas. Jointly learning consistent causal abstractions over multiple interventional distributions. In *2nd Conference on Causal Learning and Reasoning*, 2023.

Weakly-Supervised Abstractions for Linear Additive Models Supplementary Material

Riccardo Massidda[1] **Davide Bacciu**[1] **Sara Magliacane**[2]

[1]Department of Computer Science, Università di Pisa, Pisa, IT
[2]Informatics Institute, University of Amsterdam, Amsterdam, NL

A EMPTY MINIMAL SUFFICIENT EXTENSION

We show why in Example 21 all solutions to ensure abstract causal sufficiency would break graphical consistency. We consider all possible additions of X_4 to an existing or a novel relevant set. In all scenarios, the problem is the creation of a novel $\boldsymbol{R}$-direct path $X_3 \xrightarrow{\boldsymbol{R}} X_4$.

$\boldsymbol{R}_{Y_1}$. Would require a $\boldsymbol{R}$-direct path from X_2 to either X_1 or X_4, which does not exist.

$\boldsymbol{R}_{Y_2}$. Breaks the $\boldsymbol{R}$-direct paths from X_3 to X_5 and X_6.

$\boldsymbol{R}_{Y_3}$. Would require a $\boldsymbol{R}$-direct path from X_5 to X_6.

$\boldsymbol{R}_{Y_4}$. Would require a $\boldsymbol{R}$-direct path from X_6 to X_5.

$\boldsymbol{R}_{Y_5}$. Would require a $\boldsymbol{R}$-direct path from X_2 to X_4.

B PROOFS

B.1 PROOF OF LEMMA 23

Lemma 23 Let $\boldsymbol{R}$ be a set of graphically consistent relevant variables on a low-level causal graph $\mathcal{G}_\mathcal{L}$ and $\boldsymbol{W}$ be the set of new relevant variables identified by the `RSE` algorithm. Then, for any set of relevant variables $\boldsymbol{R}' \supseteq \boldsymbol{R}$, if it exists a low-level variable $X \in \boldsymbol{W}$ not in any relevant set $\boldsymbol{R}'_{Y_j} \in \boldsymbol{R}'$, then $\boldsymbol{R}'$ is either not sufficient or not graphically consistent on $\mathcal{G}_\mathcal{L}$.

Let $\boldsymbol{Y} = \{1, \dots, |\boldsymbol{R}|\}$ be the corresponding set of abstract variables. By construction (Equation 19), a variable $X \in \boldsymbol{X}$ is a new relevant variable, only if it is in the intersection of the blocks of a subset of abstract variables $\boldsymbol{V} \subseteq \boldsymbol{Y}$ and has at least one child that is not in any intersection. Let $X \in \boldsymbol{W}$ be a new relevant variable, it is immediate that X must have at least two children. If it had only one child out of any intersection, it would have not been in an intersection itself. Otherwise, if it had only one child in an intersection, it would not have a child satisfying the condition for becoming a new relevant variable. Let $\boldsymbol{A} \subseteq \mathrm{Ch}(X)$ be the children of X that are outside any intersection and $\boldsymbol{B} \subset \mathrm{Ch}(X)$ the children that are within an intersection, where $\boldsymbol{A} \uplus \boldsymbol{B} = \mathrm{Ch}(X)$. Consequently, the only possible scenarios are the following: (i.) $|\boldsymbol{B}| \geq 0, |\boldsymbol{A}| \geq 2$, or (ii.) $|\boldsymbol{B}| > 0, |\boldsymbol{A}| \geq 1$.

We prove that all possible extensions $\boldsymbol{R}' \supset \boldsymbol{R}$ not containing $X \in \boldsymbol{W}$ are not sufficient or graphically consistent.

Case $\boldsymbol{A} \geq 2$. If $X \in \boldsymbol{W}$ has two or more children outside an intersection, it means that through these edges (i.) it has $\boldsymbol{R}$-direct paths to relevant variables of multiple abstract variables $\{Y_1, \dots Y_m\} \subseteq \boldsymbol{Y}$, and (ii.) these paths only contain other variables not in any intersection. Without directly adding $X \in \boldsymbol{W}$ to a novel abstract variable, the only way to avoid being in an intersection is to block all these paths with a novel *and* unique abstract variable. This new variable has to be unique, otherwise $X \in \boldsymbol{W}$ would still be in an intersection, not of the original variables, but of the newly introduced. Suppose that we cluster together a variable X_1 in the $\boldsymbol{R}$-direct path towards $\boldsymbol{R}_{Y_1}$, a variable X_2 towards $\boldsymbol{R}_{Y_2}$, and so on. Now, X has only $\boldsymbol{R}$-direct paths to the new introduced variable, and it is not any more in an intersection. However, this operation breaks graphical consistency. In fact, X_1 has now a $\boldsymbol{R}$-direct path to $\boldsymbol{R}_{Y_1}$ only, X_2 has a $\boldsymbol{R}$-direct path to $\boldsymbol{R}_{Y_2}$ only, and so on. However, since they are now in the same relevant set, X_1 should have a $\boldsymbol{R}$-direct path towards $\boldsymbol{R}_{Y_2}, \dots, \boldsymbol{R}_{Y_m}$ as well, similarly for the other variables. Consequently, if we do not add $W \in \boldsymbol{W}$ to any relevant variable, any model $\boldsymbol{R}' \supset \boldsymbol{R}$ is not graphically consistent (if we block all the outgoing $\boldsymbol{R}$-direct paths towards distinct abstract variables) or not causally sufficient (if we do not block these paths).

Case $\boldsymbol{A} \geq 1, \boldsymbol{B} > 0$. If $X \in \boldsymbol{W}$ only has one child outside any intersection, we consider two scenarios. In the first, there exist some extension that solves the intersections of the children in $\boldsymbol{B} \subset \mathrm{Ch}(X)$. In this case, the proof for the case $\boldsymbol{A} \geq 2$ applies, hence any $\boldsymbol{R}' \supset \boldsymbol{R}$ is not sufficient or graphically consistent. Otherwise, if the intersection is not

Accepted for the 1st Workshop on Causal Abstractions and Representations (CAR) at UAI 2025 (CAR 2025).

solved, being a parent of a variable in an intersection also $X \in \boldsymbol{W}$ will be in an intersection. Therefore, any $\boldsymbol{R}' \supset \boldsymbol{R}$ without containing $X \in \boldsymbol{W}$ is not sufficient.

B.2 PROOF OF THEOREM 22

We prove the correctness of the algorithm, i.e., we prove that

$$\boldsymbol{R}' \in \texttt{RSE}(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}}) \implies \boldsymbol{R}' \in \mathrm{MiSE}\,(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}})\,. \qquad (23)$$

First, we notice that if $\boldsymbol{R}$ is not graphically consistent of $\mathcal{G}_{\mathcal{L}}$ the property holds, since the minimal sufficient extension is empty and **RSE** correctly returns the empty set. Furthermore, if $\boldsymbol{R}$ is both graphically consistent and sufficient on $\mathcal{G}_{\mathcal{L}}$, **RSE** returns a set containing only $\boldsymbol{R}$, which again coincides with the *MiSE* of $\boldsymbol{R}$ on $\mathcal{G}_{\mathcal{L}}$. Therefore, we focus on the scenario where $\boldsymbol{R}$ is graphically consistent but not sufficient on $\mathcal{G}_{\mathcal{L}}$.

At each call, **RSE** identifies an intersection between abstract variables and creates one or more novel abstract variables by selecting new relevant variables. Formally, we say that for each solution $\boldsymbol{R}' \in \boldsymbol{S}$, the **RSE** induces a sequence

$$\boldsymbol{R}^{(1)} \subset \boldsymbol{R}^{(2)} \subset \ldots \subset \boldsymbol{R}^{(n)}, \qquad (24)$$

where $\boldsymbol{R}^{(1)} = \boldsymbol{R}$ and $\boldsymbol{R}^{(n)} = \boldsymbol{R}'$.

Suppose that we remove a variable Y from the solution $\boldsymbol{R}'$, as in $\boldsymbol{R}'' = \boldsymbol{R}' \setminus \boldsymbol{R}_Y$. By construction of the **RSE** algorithm, all relevants for a new abstract variables are identified in the same step. Therefore, there must exist a set $\boldsymbol{W}^{(i)}$ at the i-th step, such that

$$\boldsymbol{W}^{(i)} \supseteq \boldsymbol{R}_Y. \qquad (25)$$

Furthermore, it holds that $\boldsymbol{R}^{(i)} \subset \boldsymbol{R}''$, as $\boldsymbol{R}''$ contains all new relevant variables except for $\boldsymbol{R}_Y$ and $\boldsymbol{R}^{(i)}$ might lack also other variables, that will be introduced in later recursive calls of the algorithm.

Due to Lemma 23, all extension of $\boldsymbol{R}^{(i)}$ lacking variables from $\boldsymbol{W}^{(i)}$ are either not causally sufficient or not graphically consistent. Therefore, since $\boldsymbol{R}'' \supseteq \boldsymbol{R}^{(i)}$, it holds that $\boldsymbol{R}''$ is not causally sufficient or graphically consistent. Notably, removing further variables would not reintroduce sufficiency *and* graphically consistency. Therefore, since any $\boldsymbol{R} \subset \boldsymbol{R}'' \subset \boldsymbol{R}'$ is not sufficient or graphically consistent, and $\boldsymbol{R}'$ is sufficient and graphically consistent, it holds that $\boldsymbol{R}' \in \mathrm{MiSE}\,(\boldsymbol{R}, \mathcal{G}_{\mathcal{L}})$.