A Stronger Calculus of Intervention for Max-Linear Bayesian Networks

Leon Sierau[1] **Francesco Nowell**[2] **Nihat Ay**[1,3,4,5]

[1]Technische Universität Hamburg, Institute for Data Science Foundations, 21079 Hamburg, Germany
[2]Technische Universität Berlin, Institute for Mathematics, 10623 Berlin, Germany
[3]Santa Fe Institute, Santa Fe, NM 87501, USA
[4]Complexity Science Hub, 1030 Vienna, Austria
[5]Leipzig University, 04109 Leipzig, Germany

Abstract

Max-linear Bayesian networks (MLBNs) are a novel class of directed acyclic graphical models which are of interest to statistics and data science due to their relevance to causality and probabilistic inference, particularly of extreme events. Interestingly, MLBNs encode strictly more conditional independence than general structural causal models. Because of this, MLBNs are Markov to a strengthening of the d-separation criterion called $*$-separation. As many tools for causal inference (most notably, Judea Pearl's do-calculus) are based on d-separation, we address the natural question of whether assuming Markovianity to $*$-separation can lead to additional causal effect identifiability under intervention. We answer this question in the affirmative and develop a stronger version of the do-calculus rules for models which are Markov to $*$-separation. However, we note that restrictions inherent to the max-linear setting (in particular, that they give rise to non strictly positive joint probability distributions) mean that our calculus is only applicable to specific kinds of policy intervention.

1 INTRODUCTION

Causal relationships between the components of a random vector $\mathbf{X} = (X_1, \dots, X_n)$ are often represented by *directed acyclic graphs* (DAGs) in the form of *structural causal models* (SCMs). In a SCM, the component X_i of $\mathbf{X}$ is a function of its parents and an independent noise term: $X_i = f_i(\mathbf{X}_{\mathrm{pa}(i)}, \epsilon_i)$.

Max-linear Bayesian networks (MLBNs), sometimes also called *recursive max-linear models* (RMLMs), are a class of structural causal models first introduced by Gissibl and Klüppelberg [2018], in which the structural equations are of the following form:

$$X_i = \bigvee_{j \in \mathrm{pa}(i)} c_{ji} X_j \vee \epsilon_i, \qquad c_{ji}, \epsilon_i \geq 0, \tag{1.1}$$

where $\vee = \max$, the ϵ_i are independent, atom-free, continuous random variables, and the quantities c_{ji} may be seen as non-negative weights on the edges of the DAG. They are the simplest class of model that exhibit *cascading failure*, where disproportionately high measurements at many points X_i in the system can be attributed to a few common sources ϵ_j. For this reason, MLBNs have found applications in the modeling of financial risk [Einmahl et al., 2018] and extreme natural events [Asadi et al., 2015, Tran et al., 2024], and are considered an important tool in the graphical modeling of *multivariate extremes* [Engelke and Hitz, 2020].

A central problem of *causal inference* is the study of the behavior of a SCM under *intervention*, i.e. the replacement of the structural equations that determine certain components of the random vector with another function. In particular, one is interested in determining whether the effect of an intervention can be computed from the distribution of the observed variables. The *do-calculus* rules of Pearl [1995] and the corresponding *ID-algorithm* [Shpitser and Pearl, 2006a] provide a complete framework for the identification of causal effects of atomic interventions, under the assumption that the observed joint distribution is strictly positive.

While some work on interventions has been done in the parallel approach to extremal statistics developed by Engelke et al. [2025], the authors restrict their attention to atomic interventions in fully observed SCMs. In this work, we study the identifiability of causal effects in MLBNs with latent confounding, expanding on recent work of Krali et al. [2025] on MLBNs with hidden variables. [1] We show that, while restrictions inherent to the max-linear setting imply that identifying expressions for the effects of atomic interventions in MLBNs are often ill-defined, the additional

[1]At the time of writing, this preprint is the only other work on MLBNs with unobserved variables of which we are aware.

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:6342–6365.

conditional independence which they encode allows for the identification of strictly more causal effects under *policy interventions* (to be introduced in Section 2.2) compatible with the MLBN structure. This is illustrated in the following example.

Example 1.1. Consider a SCM on the following graph $\mathcal{G}$:

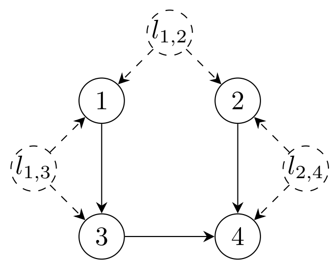


Figure 1: A DAG with observed node set $\{1, 2, 3, 4\}$ and unobserved node set $\{l_{1,2}, l_{1,3}, l_{2,4}\}$.

Upon deleting the edge $3 \to 4$, 3 and 4 are $*-$separated but $d-$connected, conditioned on $\{1, 2\}$. We will see that, due to this fact, the causal effect of the policy intervention σ'_3, $P(X_4\ ;\ \sigma'_3 = P^*(X_3 \mid X_1))$ (see Section 2.2), is identifiable if the underlying distribution is Markov to $*-$separation (as is the case for any MLBN on $\mathcal{G}$), but not if it is merely Markov to $d-$separation.

Examples such as the one above give us our first main result.

Theorem 3.6. *In a SCM that is Markov with respect to $*-$separation, there are strictly more identifiable causal effects compared to a general SCM.*

Motivated by this result, we develop a stronger version of the $\sigma-$*calculus* rules of Correa and Bareinboim [2020a], which is a generalization of $do-$calculus for stochastic interventions. We call our calculus σ^*-*calculus*, prove its correctness, and our second main result:

Theorem 5.3. *The rules of σ^*-calculus allow for the derivation of strictly more identifying expressions, compared to $d-$separation based $\sigma-$calculus.*

The remainder of this paper is structured as follows: in Section 2 we introduce the $*-$separation criterion, causal effect identifiability, and $\sigma-$calculus. In Section 3 we investigate causal effect identifiability in MLBNs, providing a first example of additional identifiability. In Section 4 we develop and prove the correctness of σ^*-calculus. In Section 5 we revisit Example 1.1, proving the claimed identifiability using σ^*-calculus. In Section 6 we outline practical limitations of this work and potential future directions.

2 PRELIMINARIES

Let $\mathcal{G} = (O \dot{\cup} U, E)$ be a DAG. The vertices in O, corresponding to the random vector $\mathbf{X}_O$, represent the observable part of the system. Analogously, U represents the unobservable part of the system, $\mathbf{X}_U$. In our examples, we denote the observed nodes by natural numbers $\{1, \ldots, n\}$ and the hidden nodes in U by $l_{i,j}$, where the subscript i, j refers to the observed children of $l_{i,j}$. We adopt the graphical nomenclature used in [Lauritzen, 1996], the essentials of which are recalled in Appendix A.

An important goal in the theory of causal inference is to identify causal effects of some components of $\mathbf{X}_O$ on the distribution on other components of $\mathbf{X}_O$, independently of the unobserved part $\mathbf{X}_U$ (we refer to Section 2.6 for a formal definition of the identification of causal effects).

2.1 SEPARATION CRITERIA

The $d-$*separation* criterion for general SCMs is recalled in Appendix A.1. Améndola et al. [2022] showed that MLBNs encode strictly more conditional independence statements than those implied by $d-$separation, and developed the stronger $*-$separation criterion.

Definition 2.1 ($*-$Separation)**.**
Let $\mathcal{G} = (O \dot{\cup} U, E)$ be a DAG. A path between two nodes $i, j \in O \dot{\cup} U$ is $*-$*connecting* given (or conditioned on) $K \subset O \dot{\cup} U$ if the path is $d-$connecting given K and contains at most one collider. If no such path exists, we say that i and j are $*-$*separated given* K, and denote this by $(i \perp_* j \mid K)_{\mathcal{G}}$.

Analogously, we say that two sets of vertices $A, B \subset O \dot{\cup} U$ are $*-$separated given K if no $*-$connecting path exists between any $a \in A$ and $b \in B$, and denote this by $(A \perp_* B \mid K)_{\mathcal{G}}$.

Theorem 2.2 (Améndola et al. [2022],Theorem 5.15)**.** *Let $\mathbf{X}_{O \dot{\cup} U}$ be a random vector distributed according to a MLBN on a DAG $\mathcal{G}$. For arbitrary subsets of vertices $A, B, C \subseteq O \dot{\cup} U$ the following holds:*

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C \iff (A \perp_* B \mid C)_{\mathcal{G}}.$$

Example 2.3. Let $\mathcal{G}$ be the DAG in Figure 2.

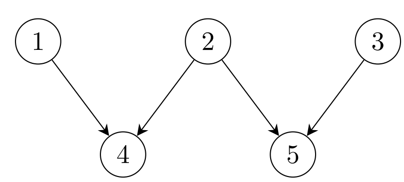


Figure 2: The "Cassiopeia" graph.

Here it holds that $(1 \perp_* 3 \mid 4, 5)$, implying the statement $X_1 \perp\!\!\!\perp X_3 \mid \mathbf{X}_{4,5}$ in any MLBN on $\mathcal{G}$. The same conditional independence statement does not hold in a general SCM on $\mathcal{G}$, as 1 and 3 are $d-$connected given 4 and 5.

We will show in Sections 3.3 and 5 that this stronger separation criterion allows us to identify the causal effects of interventions that are impossible to identify in general SCMs. The remainder of this chapter will introduce the necessary notions of interventions, causal effects of interventions and causal effect identifiability.

2.2 INTERVENTIONS AND THE IDENTIFIABILITY OF CAUSAL EFFECTS

A canonical way of defining interventions in SCMs is given by the notion of "atomic" or "hard" interventions [Pearl, 2009]. This refers to manually setting the values of a treatment set $T \subseteq O$ to constants and thereby replacing the structural mechanisms that generate $\mathbf{X}_T$ by the constant function $X_t = c_t$ for all $t \in T$. This operation is denoted by $do(\mathbf{X}_T = \mathbf{c}_T)$. As we will see in Section 3.2, the effect identification of atomic interventions often leads to problems due to inherent support-restrictions in probability distributions generated by MLBNs. Thus, we will work with the more general notion of *policy interventions* [Correa and Bareinboim, 2020a], of which hard interventions are a special case.

In this generalized setting, the conditional distribution $P(\mathbf{X}_T \mid \mathbf{X}_{\mathrm{pa}(T)})$, induced by the natural mechanisms of $\mathbf{X}_T$, is replaced by a new conditional distribution. Correa and Bareinboim [2020a] denote policy interventions by $\sigma_T = P^*(\mathbf{X}_T \mid \mathbf{X}_{\mathrm{pa}^{\sigma_T}(T)})$, where $\mathrm{pa}^{\sigma_T}(T) = \bigcup_{t \in T} \mathrm{pa}^{\sigma_T}(t) \setminus T$ denotes the parent set of T in the graph corresponding to the model induced by σ_T. Technically, the intervention takes place at the generating mechanisms, such that the structural functions $f_t(\mathbf{X}_{\mathrm{pa}(t)}, \epsilon_t)$ are replaced by a new function $f_t^{\sigma_T}(\mathbf{X}_{\mathrm{pa}^{\sigma_T}(t)})$ for all $t \in T$. One can intervene in the system by defining the new functions and compute the resulting conditional distribution $P^*(\mathbf{X}_T \mid \mathbf{X}_{\mathrm{pa}^{\sigma_T}(T)})$. Alternatively, one can directly compute the effect of an intervention on the level of the induced $P^*(\mathbf{X}_T \mid \mathbf{X}_{\mathrm{pa}^{\sigma_T}(T)})$. Strictly speaking, in this case one does not identify the causal effect of a specific intervention on the structural level but rather the effect of any structural intervention that would induce $P^*(\mathbf{X}_T \mid \mathbf{X}_{\mathrm{pa}^{\sigma_T}(T)})$, which could generally be achieved by infinitely many different interventions.

Remark 2.4. An intervention induces a new structural causal model, which we denote by $\mathcal{M}_{\sigma_T}$, on a DAG $\mathcal{G}_{\sigma_T} = (O \dot\cup U^{\sigma_T}, E^{\sigma_T})$, which we refer to as the *post-intervention DAG*. Note that policy interventions allow for changing the parent sets of the intervened variables not only within the un-intervened vertex set $O \dot\cup U$, but also allow for the introduction of new unobserved parents such that $U \subseteq U^{\sigma_T}$.

Definition 2.5 (Causal Effect of an intervention)**.**
For subsets $A, B \subseteq O$, the *causal effect of an intervention* σ_A on $\mathbf{X}_B$ is the distribution of $\mathbf{X}_B$ under the intervened model $\mathcal{M}_{\sigma_A}$, denoted by $P(\mathbf{X}_B\,;\,\sigma_A)$ or equivalently as $P^{\sigma_A}(\mathbf{X}_B)$.

In the setting where both $\mathbf{X}_O$ and $\mathbf{X}_{U^{\sigma_T}}$ are observable, the causal effect of any intervention can be computed by

$$
\begin{aligned}
&P(\mathbf{X}_O\,;\,\sigma_T) = \\
&\int_{\mathcal{X}_{U^{\sigma_T}}} \prod_{v \in (O \cup U) \setminus T} P(X_v \mid \mathbf{X}_{\mathrm{pa}(v)}) P(\mathrm{d}\mathbf{x}_U) \\
&\qquad \times \prod_{t \in T} P(X_t \mid \mathbf{X}_{\mathrm{pa}^{\sigma_T}(t)}) P(\mathrm{d}\mathbf{x}_{U^{\sigma_T} \setminus U}\,;\,\sigma_T),
\end{aligned}
$$

where $\mathcal{X}_{U^{\sigma_T}}$ denotes the state space of $\mathbf{X}_{U^{\sigma_T}}$.

However, we assume $\mathbf{X}_{U^{\sigma_T}}$ as unobserved. The goal of causal inference is to find a unique function that maps the joint distribution of the observed variables $P(\mathbf{X}_O)$ to the causal effect in question [Pearl, 2009, Correa and Bareinboim, 2020a]. This is what is meant when we speak of the *identification* of causal effects, which we define analogously to Correa and Bareinboim [2020a], Definition 2.

Definition 2.6 (Causal Effect Identifiability)**.**
Let $A, B, C \subseteq O$ with $B \cap C = \emptyset$. The (conditional) effect of an intervention specified by $\sigma_A = \{\sigma_{a_1}, \ldots, \sigma_{a_{|A|}}\}$ on a set of outcome variables B, conditional on C, $P(\mathbf{X}_B \mid \mathbf{X}_C\,;\,\sigma_A)$, is said to be *identifiable* in $\mathcal{G}$ if it is uniquely computable from the joint distribution $P(\mathbf{X}_O)$, for every assignment $(\mathbf{X}_B = \mathbf{x}_B, \mathbf{X}_C = \mathbf{x}_C)$, in every model on $\mathcal{G}$ that induces $P(\mathbf{X}_O)$.

2.3 NOTE ON CAUSAL DIAGRAMS AND LATENT PROJECTION OF MLBNS

In the following, we work in the full DAG $\mathcal{G} = (O \dot\cup U, E)$ corresponding to the underlying the SCM $\mathcal{M}$. In the literature on causal inference, projections of the true DAG onto O, sometimes referred to as *causal diagrams*, are commonly used. In Appendix E we show that all our results can be transferred to this setting.

2.4 RULES OF σ-CALCULUS

Correa and Bareinboim [2020a] provide a generalized version of Pearl's $do-$calculus [Pearl, 1995] for the derivation of identifying expressions, which they refer to as $\sigma-$*calculus*. The rules of $\sigma-$calculus are formulated analogously to the rules of $do-$calculus and are presented in List 1.

While the first rule is essentially a restatement of the global Markov property for $\mathcal{G}_{\sigma_Z}$, rules 2 and 3 provide graphical criteria for the equality of expressions under *change of mechanism* from σ_Z to σ'_Z, which are based on $d-$separation statements in modified versions of the two post-intervention

DAGs. The modifications are sometimes referred to as *graph surgeries*, and are defined in Definition 2.7. Unlike the framework of Pearl [2009] for atomic interventions, two distinct policy interventions will generally lead to post-intervention graphs which are not subgraphs of one another. Hence the requirement for explicitly checking that the $d-$separation statements hold in both graphs.

***σ*-Rule 1**
$P(\mathbf{X}_A \mid \mathbf{X}_B, \mathbf{X}_C; \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_C; \sigma_Z)$
if $(A \perp_d B|C)$ holds in $\mathcal{G}_{\sigma_Z}$
***σ*-Rule 2**
$P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B; \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B; \sigma_W, \sigma'_Z)$
if $(A \perp_d Z \mid B)$ holds in $\mathcal{G}_{\sigma_W \sigma_Z \underline{Z}}$ and $\mathcal{G}_{\sigma_W \sigma'_Z \underline{Z}}$
***σ*-Rule 3**
$P(\mathbf{X}_A \mid \mathbf{X}_B; \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_B; \sigma_W, \sigma'_z)$
if $(A \perp_d Z \mid B)$ holds in $\mathcal{G}_{\sigma_W \sigma_Z \overline{Z(B)}}$ and $\mathcal{G}_{\sigma_W \sigma'_Z \overline{Z(B)}}$

List 1: The rules of $\sigma-$calculus.
$\mathcal{G}_{\sigma_W \sigma_Z}$ is the graph resulting from applying σ_Z to $\mathcal{G}_{\sigma_W}$, whereas $Z(B) := Z \setminus \mathrm{an}_{\mathcal{G}_{\sigma_W}}(B)$.

Definition 2.7. Let $\mathcal{G} = (O \dot{\cup} U, E)$ be a DAG and $Z \subset O$. The *child-cut* of $\mathcal{G}$ w.r.t Z, denoted $\mathcal{G}_{\underline{Z}}$, is the graph obtained by removing every outgoing edge from Z from E.
The *parent-cut* of $\mathcal{G}$ w.r.t Z, denoted $\mathcal{G}_{\overline{Z}}$, is the graph obtained by removing every incoming edge to Z from E.

Remark 2.8. Like the rules of $do-$calculus [Pearl, 1995], that have been shown to be complete for the identification of causal effects of atomic interventions in recursive SCMs [Huang and Valtorta, 2006], the rules of $\sigma-$calculus have been shown to be complete for the more general problem of *soft transportability* [Correa and Bareinboim, 2020b]. This means that it is possible to derive an identifying expression for the causal effect of a policy intervention, based on the joint distribution of the observed variables, if and only if there exists a finite sequence of applications of the $\sigma-$calculus rules with which this expression can be derived. Correa and Bareinboim [2020b] also introduce a complete graphical criterion for the identifiability of a policy intervention. However, presenting and using this criterion would require additional definitions and examples beyond the scope of this paper. Thus, we make use of more common graphical criteria for the identifiability of atomic interventions or provide direct proofs of non-identifiability, when necessary.

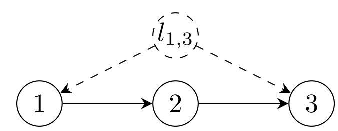
3 CAUSAL IDENTIFICATION IN MLBNS

Many classical results in the field of causal inference assume that the SCMs generate strictly positive joint probability distributions [Pearl, 2009, Shpitser and Pearl, 2006b, Tian, 2004, Huang and Valtorta, 2006]. This assumption guarantees the well-definedness of every symbolically derived identifying expression. While strict positivity is a valid assumption for important classes of SCMs, most prominently for linear Gaussian models, this can lead to problems when working in SCMs that generate non strictly positive joint distributions. Despite its importance, non-positivity is an often overlooked problem in the field of causal inference (see Hwang et al. [2024] for a notable exception).

3.1 CAUSAL EFFECT IDENTIFICATION IN NON-POSITIVE SCM

Before we explain why non-positivity is an issue for the identification of causal effects in (continuous) MLBNs, we illustrate the more general problem in the setting of discrete random variables.

Example 3.1. Consider a discrete SCM that factorizes according to the classic backdoor graph:



Assume we are given the joint distribution $P(X_1, X_2, X_3)$ and want to identify the causal effect of setting X_2 to a constant value x_2 on the distribution of X_3. That is we want to compute $P(X_3\ ;\ \sigma_2 = do(X_2 = x_2))$, which is equivalent to the causal effect $P(X_3 \mid do(X_2 = x_2))$ in the notation used in Pearl [2009].

An identifying expression for any assignment $X_3 = x_3$ can be derived by the well known backdoor adjustment formula or by the rules of $do-$calculus and is given by

$$
\begin{aligned}
&P(X_3 = x_3 \mid do(X_2 = x_2)) \\
&= \sum_{x_1 \in \mathcal{X}_1} P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) P(X_1 = x_1).
\end{aligned}
$$

Assume further that the marginal distribution $P(X_1)$ is strictly positive on some finite, discrete domain and that X_2 is lower bounded by some function of X_1. For concreteness, assume that the underlying structural mechanism is $X_2 = X_1 + Z_2$, where $Z_2 \sim \mathrm{Ber}(\lambda)$. This implies that all events of the form $(X_1 > x_2,\ X_2 = x_2)$ have probability 0 under the natural mechanisms.

In order to to evaluate the identifying expression above, we have to sum over *all* values x_1 in X_1's state space, while keeping $X_2 = x_2$ fixed. Now consider for some value $x_1 > x_2$ the summand $P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) P(X_1 = x_1)$.

If the first factor could be uniquely computed from $P(X_1, X_2, X_3)$, it would be given by the following fraction:

$$P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) \\ = \frac{P(X_1 = x_1, X_2 = x_2, X_3 = x_3)}{P(X_1 = x_1, X_2 = x_2)}.$$

But this expression is not well defined, as both the numerator and the denominator evaluate to 0.

Note that this does not imply that the causal effect $P(X_3 = x_3 \mid do(X_2 = x_2))$ is not defined. It does however imply that there exist arbitrarily many different *versions* of $P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2)$ that satisfy $P(X_3 = x_3, X_1 = x_1, X_2 = x_2) = P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2)P(X_1 = x_1, X_2 = x_2)$ and hence are all compatible with the observed joint distribution. Thereby, despite the existence of an identifying expression, the causal effect is not uniquely computable from $P(X_3 = x_3, X_1 = x_1, X_2 = x_2)$. This problem arises necessarily when we consider a (continuous) MLBN on the same example DAG.

3.2 CHALLENGES DUE TO NON-POSITIVITY IN MLBNS

The distributions generated by MLBNs are inherently not strictly positive (except for the trivial case of an MLBN that corresponds to a DAG of only isolated vertices). Recall the form of the structural equations that define a MLBN, given in (1.1). The max-linear structure leads to every variable being bounded from below by the values of its parents. To make this more explicit, consider the following example.

Example 3.2. Let $c_{12} > 0$, ϵ_1, ϵ_2 be continuous non-negative atom-free random variables, and

$$X_1 = \epsilon_1 \\ X_2 = c_{12}X_1 \vee \epsilon_2.$$

Observe that X_2 is bounded from below by $c_{12}X_1$, which implies that $P(X_2 < c_{12}x_1, X_1 \geq x_1) = 0$ for all $x_1 \in \mathbb{R}$.

The presence of regions on which the joint distribution of a MLBN is non-positive renders the identification of causal effects impossible in many settings that require the integration over the whole state space of an adjustment set. The next example illustrates this issue.

Example 3.3. Consider a MLBN that factors according to the backdoor graph in Example 3.1:

If we want to identify the causal effect $P(X_3 \,;\, \sigma_2 = P^*(X_2))$, we can derive an identifying expression by the rules of $\sigma-$calculus:

$$P(X_3 \,;\, \sigma_2) = \\ \int_{\mathcal{X}_1} \left[\int_{\mathcal{X}_2} P(X_3 \mid X_1, X_2) P^*(\mathrm{d}x_2) \right] P(\mathrm{d}x_1).$$

We refer to Appendix D.1 for the derivation of this expression.

Since $P^*(X_2)$ is fixed and known, this expression may appear to be computable from $P(X_1, X_2, X_3)$ at first glance. However, note that the conditional probability $P(X_3 \mid X_1, X_2)$ is only unique up to (X_1, X_2)-Nullsets. But because $P^*(X_2)$ is not necessarily absolutely continuous with respect to $P(X_1, X_2)$, the inner integral $\int_{\mathcal{X}_2} P(X_3 \mid X_1, X_2)P^*(dx_2)$ depends on the values attained by $P(X_3 \mid X_1, X_2)$ on (X_1, X_2)-Nullsets. In other words, the expression is dependent on the version of $P(X_3 \mid X_1, X_2)$, which can be chosen arbitrarily. Hence, the integral is not well-defined and the causal effect is not uniquely determined by the joint distribution on the observable random variables.

Examples 3.1 and 3.3 shed light on the importance of considering the well-definedness of identifying expressions in SCMs that do not induce strictly positive joint probability distributions. This justifies passing to the more general notion of a policy intervention, which allows us to define interventions whose causal effects are uniquely determined by the joint distribution of the observed variables. For example, in the backdoor graph from Example 3.1, the causal effect of an intervention $\sigma_2 = P^*(X_2 \mid X_1)$ on X_3 can be identified with

$$P(X_3 \,;\, \sigma_2) = \\ \int_{\mathcal{X}_1} \left[\int_{\mathcal{X}_2} P(X_3 \mid X_1, X_2)\, P^*(\mathrm{d}x_2 \mid X_1) \right] P(\mathrm{d}x_1).$$

Again, we refer to Appendix D.1 for the derivation.

This integral *is* well defined if we intervene such that $P^*(X_2 \mid X_1) \ll P(X_2 \mid X_1)$, where $\ll$ denotes absolute continuity.

Remark 3.4. Arguably, the most relevant type of intervention in extreme value theory is one in which certain components of $\mathbf{X}_O$ are set to disproportionately high values. While, with respect to the graph in Example 3.3, we cannot identify the causal effect of a constant "shock" event of the form $do(X_2 = x_2)$, we *can* identify the causal effect of something qualitatively similar with a policy intervention; the effect of changing the mechanism by which X_2 is generated from X_1 such that $P^*(X_2 \mid X_1)$ is heavy-tailed. Thus, causal effects of interest in the modeling of extreme events would still be identifiable.

This subsection outlined some challenges and caveats in the application of the theory of the identification of causal effects in the max-linear setting. Luckily, the additional structure inherent to MLBNs does not only lead to challenges. In the next subsection we provide a motivating example which shows that the additional conditional independence statements implied by $*-$separation (see Definition 2.1) make a substantial difference for the identifiability of causal effects.

3.3 ADDITIONAL IDENTIFIABILITY IN MLBNS

Example 3.5. Let $\mathcal{G}$ be the DAG depicted in Figure 3 and let $\mathbf{X}$ be an MLBN on $\mathcal{G}$.

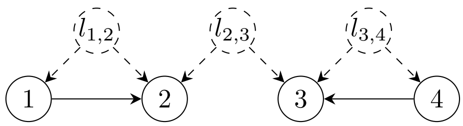


Figure 3: A DAG on four observable and three unobservable nodes.

In this graph, it holds that $(1 \perp_* 4 \mid 2, 3)$. Due to Theorem 2.2, this implies $(X_4 \perp\!\!\!\perp X_1 \mid X_2, X_3)$. It is thereby immediately clear that changing the distribution of X_1 does not impact the conditional distribution of X_4, given X_2, X_3.

Let σ_1 be the intervention which sets $X_1 = \epsilon^* \sim \exp(\lambda)$, such that $P^*(X_1) = \exp(\lambda)$, where $\exp(\lambda)$ is the exponential distribution with parameter $\lambda > 0$. Then, the intervened model $\mathcal{M}_{\sigma_1}$ is a MLBN in which the conditional independence $(X_4 \perp\!\!\!\perp X_1 \mid X_2, X_3)$ still holds. Thus,

$$P(X_4 \mid X_2, X_3 \ ; \ \sigma_1 = P^*(X_1)) \qquad (3.1)$$

is identifiable with

$$\begin{aligned} &P(X_4 \mid X_2, X_3 \ ; \ \sigma_1 = P^*(X_1)) \\ &= P(X_4 \mid X_2, X_3 \ ; \ \sigma_1' = idle) \\ &= P(X_4 \mid X_2, X_3), \end{aligned}$$

where $\sigma_1' = idle$ indicates the natural mechanism, i.e. is equivalent to performing no intervention. This implies that the causal effect $P(X_4 \mid X_2, X_3 \ ; \ \sigma_1)$ is identifiable in any MLBN on $\mathcal{G}$.

This minimal example already leads to an important result, presented in the following theorem.

Theorem 3.6. *There exists a DAG $\mathcal{G}$ and an intervention σ, for which the following two statements hold:*

(i) In any MLBN $\mathcal{M}^*$ on $\mathcal{G}$, the causal effect of σ is identifiable.

(ii) There exists a model $\mathcal{M}$ on $\mathcal{G}$ which is Markov to $d-$separation, in which the causal effect of σ is not identifiable.

In particular, there are generally more identifiable causal effects in MLBNs, compared to general SCMs.

Proof. Let $\mathcal{G}$ and σ_1 be the DAG and the intervention from Example 3.5. The first part of Theorem 3.6 is already implied by Example 3.5, which shows that the effect (3.1) is identifiable in any MLBN on $\mathcal{G}$.

For the second part, note that the conditional independence $(X_4 \perp\!\!\!\perp X_1 \mid X_2, X_3)$ is *not* implied by $d-$separation. Hence, it is intuitively plausible that the causal effect (3.1) *can* depend on $P^*(X_1)$ in a general SCM.

In order to prove the statement rigorously, we explicitly construct two SCMs on $\mathcal{G}$, that are Markov with respect to $d-$separation and induce the same joint distribution on the observable variables, $P(X_1, X_2, X_3, X_4)$, while they disagree on the causal effect (3.1). That is, we construct two SCMs $\mathcal{M}_1, \mathcal{M}_2$ on $\mathcal{G}$ such that

$$P^{\mathcal{M}_1}(X_4 \mid X_2, X_3 \ ; \ \sigma_1) \neq P^{\mathcal{M}_2}(X_4 \mid X_2, X_3 \ ; \ \sigma_1).$$

Due to space restrictions we refer to Appendix D.2 for the complete construction. $\square$

Theorem 3.6 naturally raises the question of whether we can strengthen the rules of σ-calculus in MLBNs in order to provide us with a tool for the symbolic derivation of identifying expressions that cannot be derived by $d-$separation based $\sigma-$calculus. It turns out that this is indeed possible, as we show in the next section.

4 A STRONGER σ-CALCULUS FOR MLBNS

The following theorem states a stronger version of the $d-$separation based $\sigma-$calculus in SCMs that are Markov to $*-$separation. We refer to it as σ^*-*calculus*.

Theorem 4.1 (Rules of σ^*-calculus)**.** *Assume that the joint distributions $P^{\sigma_W}, P^{\sigma_Z}, P^{\sigma_Z'}$, generated by $\mathcal{M}_{\sigma_W}$, $\mathcal{M}_{\sigma_Z}$ and $\mathcal{M}_{\sigma_Z'}$, are all Markov to $\perp_*$.*

Then the following stronger versions of the rules of $\sigma-$calculus (List 1) apply:

σ^*-Rule 1
$P(\mathbf{X}_A \mid \mathbf{X}_B, \mathbf{X}_C; \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_C; \sigma_Z)$
if $(A \perp_* B|C)$ in $\mathcal{G}_{\sigma_Z}$

σ^*-Rule 2
$P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B; \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B; \sigma_W, \sigma_Z')$
if $(A \perp_* Z \mid B)$ holds in $\mathcal{G}_{\sigma_W \sigma_Z \underline{Z}}$ and $\mathcal{G}_{\sigma_W \sigma_Z' \underline{Z}}$

σ^*-Rule 3
$P(\mathbf{X}_A \mid \mathbf{X}_B; \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_B; \sigma_W, \sigma_Z')$
if $(A \perp_* Z \mid B)$ holds in $\mathcal{G}_{\sigma_W \sigma_Z \overline{Z(B)}}$ and $\mathcal{G}_{\sigma_W \sigma_Z' \overline{Z(B)}}$,

where the equality in Rule 2 holds outside the union of P^{σ_Z}- and $P^{\sigma_Z'}$-Nullsets in $\sigma(\mathbf{X}_Z, \mathbf{X}_B)$ and the equality in Rule 3 holds outside the union of P^{σ_Z}- and $P^{\sigma_Z'}$-Nullsets in $\sigma(\mathbf{X}_B)$, where $\sigma(\cdot)$ denotes the smallest $\sigma-$algebra generated by the respective random vector.

Proof. We refer to Appendix C for the proof of Theorem 4.1, since it is quite involved and requires multiple auxiliary technical results. That is because the proof strategy of both Pearl [1995] and Correa and Bareinboim [2020a] which uses graphical arguments in a graph augmented with regime indicators cannot be used. The corresponding augmented models in our setting would be mixtures of MLBNs and, to the best of our knowledge, there are no results on the $*-$Markovianity of such mixtures as of today. $\square$

Remark 4.2. The assumption of Theorem 4.1 restricts the interventions to which the rules of σ^*-calculus apply. Note however, that these restrictions do not apply to the $d-$separation based rules, which of course still hold in MLBNs. This is discussed in more detail in Section 6.1.

In the next section we show that the rules of σ^*-calculus can be used to derive identifying expressions that cannot be derived by regular $d-$separation based $\sigma-$calculus.

5 σ^*-CALCULUS IS STRICTLY STRONGER THAN ITS $d-$SEPARATION COUNTERPART

We apply the rules presented in Theorem 4.1 to a MLBN on the graph shown in the introductory example Example 1.1.

Example 5.1 (Example 1.1 revisited)**.** Our goal is to derive an identifying expression for the causal effect

$$P(X_4 \; ; \; \sigma_3' = P^*(X_3 \mid X_1)). \tag{5.1}$$

We do this by checking the conditions of the rules in parallel with respect to the graph induced by σ_3' and the natural graph induced by $\sigma_3 = idle$, displayed in Figure 4. We write $P(\mathbf{X})$ instead of $P(\mathbf{X} \; ; \; \sigma_3 = idle)$ as shorthand notation for the probability distribution $P(\mathbf{X})$ generated by the natural mechanisms.

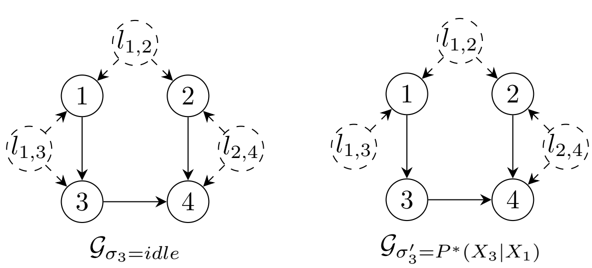


Figure 4: The graphs $\mathcal{G}_{\sigma_3}$ (left) and $\mathcal{G}_{\sigma_3'}$ (right).

We start by writing

$$\begin{aligned} &P(X_4 \; ; \; \sigma_3') = \\ &\int_{\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3} P(X_4 \mid \mathbf{X}_{1,2,3} \; ; \; \sigma_3') \\ &\quad \times P(\mathrm{d}x_3 \mid \mathbf{X}_{1,2} \; ; \; \sigma_3') P(\mathrm{d}\mathbf{x}_{1,2} \; ; \; \sigma_3'). \end{aligned}$$

First, we handle the term $P(X_3 \mid \mathbf{X}_{1,2} \; ; \; \sigma_3')$. We can check that

$$(3 \perp_d 2 \mid 1) \text{ in } \mathcal{G}_{\sigma_3'}.$$

Now σ-Rule 1 implies

$$P(X_3 \mid \mathbf{X}_{1,2} \; ; \; \sigma_3') = P(X_3 \mid X_1 \; ; \; \sigma_3')$$

By definition, we have that

$$P(X_3 \mid X_1 \; ; \; \sigma_3') = P^*(X_3 \mid X_1).$$

Now it only remains to handle the terms

$$P(X_4 \mid \mathbf{X}_{1,2,3} \; ; \; \sigma_3') \text{ and } P(\mathbf{X}_{1,2} \; ; \; \sigma_3').$$

For the first term we can check that

$$(4 \perp_* 3 \mid \{1,2\}) \text{ in } \mathcal{G}_{\sigma_3' \underline{3}} \text{ and } \mathcal{G}_{idle \underline{3}} \; .$$

Hence, we can use σ^*-Rule 2 to write

$$P(X_4 \mid \mathbf{X}_{1,2,3} \; ; \; \sigma_3') = P(X_4 \mid \mathbf{X}_{1,2,3})$$

outside of P^{σ_Z}- and $P^{\sigma_Z'}$-Nullsets in $\sigma(\mathbf{X}_{1,2,3})$.

For the second term we can check that

$$(\{1,2\} \perp_d 3 \mid \emptyset) \text{ in } \mathcal{G}_{\sigma_3' \overline{3(\emptyset)}} \text{ and } \mathcal{G}_{idle \overline{3(\emptyset)}} \; .$$

Hence, we can use σ-Rule 3 to write

$$P(\mathbf{X}_{1,2} \; ; \; \sigma_3') = P(\mathbf{X}_{1,2}).$$

Taking the previous steps together, this finally yields the identifying expression

$$\begin{aligned} &P(X_4 \; ; \; \sigma_3') = \\ &\int_{\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3} P(X_4 \mid \mathbf{X}_{1,2,3}) P^*(\mathrm{d}x_3 \mid X_1) P(\mathrm{d}\mathbf{x}_{1,2}). \end{aligned} \tag{5.2}$$

outside of P^{σ_3}- and $P^{\sigma_3'}$-Nullsets in $\sigma(\mathbf{X}_{1,2,3})$.

Numerical experiments validate the correctness of the identifying integral 5.2 in a MLBN[2]. We set up MLBNs and linear Gaussian SCMs corresponding to the left graph in Figure 4 and then perform an intervention by exchanging the innovation Z_3 for a more heavily tailed one, which we refer to as Z_3^*. This pushes X_4 towards higher values with larger probability than under the natural mechanisms.

Figure 5 shows that the cumulative distribution function (CDF) of X_4 under the policy intervention can be computed correctly by the identifying integral 5.2 in the MLBN case (left), while the evaluation of the same integral leads to an incorrectly identified CDF of X_4 in the Gaussian case (right).

[2]Code is available under: `https://github.com/LeonSierau/MLBN_Causal_Effects_UAI`.

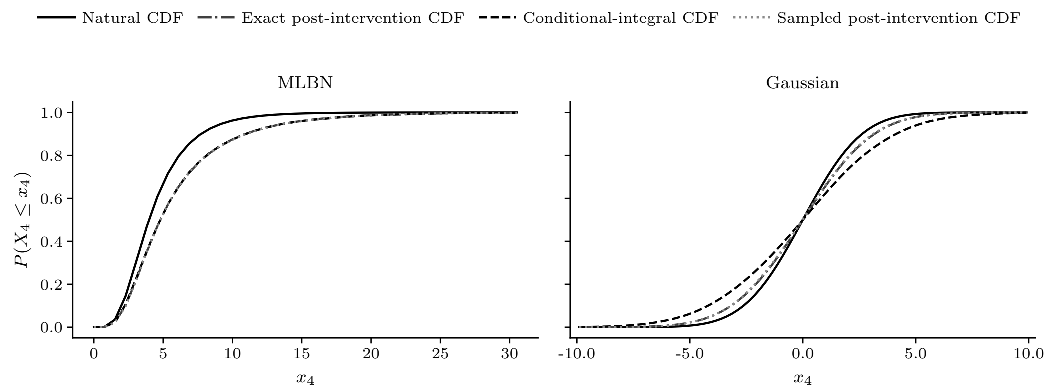


Figure 5: The cumulative distribution function (CDF) of X_4 under the natural mechanism and under a policy intervention, replacing the innovation Z_3 with a more heavily tailed Z^*, in SCMs corresponding to the graph in Figure 4 (left). The exact post-intervention CDF is computed based on the true max-linear coefficient matrix and by analytically derived Gaussian distributions, respectively. An approximated post-intervention CDF, based on sampling X_4 from the true intervened model, is included as a sanity check. The "conditional-integral CDF" refers to the curve computed from evaluating the identifying integral 5.2 for small intervals with a moving window. Left: In the MLBN case the CDF based on the identifying integral matches the exact and sampled post-intervention CDF. Right: In the Gaussian case the curves do not match, as expected.

Remark 5.2. Note that, because of the equality outside a union of $P^{\sigma_3}-$ and $P^{\sigma'_3}-$Nullsets, in order for this integral to be equal to the causal effect in (5.1), we must assert that P^{σ_3} and $P^{\sigma'_3}$ have exactly the same Nullsets in $\sigma(\mathbf{X}_{1,2,3})$, i.e. that they are *equivalent* with respect to $\sigma(\mathbf{X}_{1,2,3})$. Since $P(\mathbf{X}_{1,2}\ ;\ \sigma'_3) = P(\mathbf{X}_{1,2})$ holds both $P^{\sigma'_3}$- and P^{σ_3}-almost everywhere, it remains to choose $P^*(X_3 \mid X_1)$ such that it is equivalent to $P(X_3 \mid \mathbf{X}_{1,2})$ with respect to $\sigma(\mathbf{X}_{1,2,3})$. That is, we must make sure that both

$$P^*(X_3 \mid X_1) \ll P(X_3 \mid \mathbf{X}_{1,2}) \text{ and}$$
$$P(X_3 \mid \mathbf{X}_{1,2}) \ll P^*(X_3 \mid X_1)$$

hold. We will discuss this issue further in Section 6.1.

The derivation above directly leads to the following result:

Theorem 5.3. *The Rules of σ^*-calculus allow for the derivation of strictly more identifying expressions, compared to $d-$separation based $\sigma-$calculus.*

For the proof of Theorem 5.3 we require the notions of a $\mathcal{C}$-component, a $\mathcal{C}$-tree and a result that implies non-identifiability of certain causal effects if a graph forms a $\mathcal{C}$-tree (see Appendix F).

Proof. We have already shown that σ^*-calculus can be used to derive an identifying expression for the causal effect

$$P(X_4\ ;\ \sigma'_3 = P^*(X_3 \mid X_1)) \tag{5.3}$$

with respect to the graph in Example 1.1. Hence, it suffices to show that there exists no finite sequence of applications of $d-$separation based $\sigma-$calculus, by which an identifying expression for this effect can be derived.

Observe that the identifiability of (5.3) by a finite sequence of rules of $\sigma-$calculus would imply the identifiability of the special case where $\sigma'_3 = do(X_3 = x_3)$ by the same finite sequence of rules. This would imply the identifiability of

$$\begin{aligned} &P(X_4\ ;\ \sigma'_3 = P^*(X_3 \mid X_1)) \\ &= P(X_4\ ;\ \sigma'_3 = do(X_3 = x_3)) \\ &= P(X_4 \mid do(X_3 = x_3)) \end{aligned}$$

by the rules of $\sigma-$calculus.

However, it is easy to see that the causal diagram corresponding to the DAG in Example 1.1, obtained by replacing every latent variable with exactly one bidirected edge connected to both of its children, forms a 4-rooted $\mathcal{C}$-tree in the sense of Definition F.2.

Hence, Theorem F.3 immediately implies that $P(X_4 \mid do(X_3 = x_3))$ cannot be identifiable in a general SCM and thereby no finite sequence of σ-calculus rules that identify $P(X_4\ ;\ \sigma'_3 = P^*(X_3 \mid X_1))$ can exist. $\square$

6 CONCLUDING REMARKS

We have seen that not only is it possible to identify strictly more effects in MLBNs, compared to general SCMs, but that we can even formulate rules that allow for symbolic manipulations to derive identifying expressions in a systematic fashion. In the remainder of this work we revisit some of the limitations that we observed along the way and give an outlook on future research.

6.1 PRACTICAL LIMITATIONS

As mentioned in Remarks 4.2 and 5.2 there are restrictions on the design of interventions that lead to well-defined identifying expressions for causal effects.

The first restriction is that one must define interventions such that the intervened system is still Markov with respect to $*-$separation. This can be achieved by defining interventions on the functional level. As an example, one can define each treatment variable σ_Z to be a max-linear function of its observed parent set under σ'_Z and one or multiple additional unobserved parents. From there, one can compute the resulting conditional probability distribution $P^*(\mathbf{X}_Z \mid \mathbf{X}_{\mathrm{pa}^{\sigma_Z}(Z)\cap O})$. This ensures that the resulting system is still a MLBN and hence Markov with respect to $*-$separation.

For the second restriction recall Remark 5.2. In the example given there, we observed that the intervention σ'_3 must induce a conditional distribution $P^*(X_3 \mid X_1)$ which is equivalent to $P(X_3 \mid \mathbf{X}_{1,2})$. While this condition is trivially fulfilled in classes of SCM that induce strictly positive and continuous joint distributions, the joint distributions induced by MLBNs are inherently not strictly positive, as clarified by Example 3.2. Specifically, conditional distributions $P(\mathbf{X}_A \mid \mathbf{X}_B)$ induced by MLBNs naturally contain regions of zero measure for $A < c_{BA}B$ whenever B contains parents of A in $\mathcal{G}$.

Moreover, as discussed in [Klüppelberg and Lauritzen, 2019], every conditional distribution $P(\mathbf{X}_A \mid \mathbf{X}_B)$ generated by an MLBN features atoms if B is among A's ancestors in $\mathcal{G}$. In the case of singleton A and B, for instance, $P(X_A \mid X_B)$ has an atom exactly at the point $w_{BA}B$, where w_{BA} is equal to weight of a *max-weighted path* from B to A (see for example section 2 of [Gissibl et al., 2021] for a rigorous definition of this notion).

In the concrete example discussed in Remark 5.2, the presence of atoms requires consideration when defining an intervention $P^*(X_3 \mid X_1)$ that is equivalent to $P(X_3 \mid \mathbf{X}_{1,2})$. A valid construction for an intervention σ_3 that achieves this would be to exchange the natural innovation at X_3 with a new innovation ϵ^*, leaving the rest of the max-linear equation for X_3 unchanged. It should be noted, however, that computing the induced $P^*(X_3 \mid X_1)$ in this manner assumes knowledge of both the distribution of X_3's unobserved parents and the respective weight of a max-weighted path from them to X_3.

Developing an intervention procedure that is consistent with the restrictions of the strengthened rules 2 and 3, yet less reliant on local knowledge about the mechanisms generating the variables in the treatment set, is an open problem. Most notably, finding alternative ways to intervene in the system, beyond defining max-linear mechanisms, while preserving the $*-$Markovianity of at least the relevant remaining parts of the intervened system will be an important next step.

Finally, an alternative approach for strengthening our results would be to derive versions of our σ^*-calculus with less restrictive assumptions. This could potentially be achieved by approaching the problem of deriving the rules directly at the level of the structural equations, rather than relying on probability theoretic means.

6.2 OUTLOOK

To our knowledge, our work is only the second treatment of MLBNs with hidden variables, alongside the recent work of Krali et al. [2025]. In that work, the authors note that MLBNs with hidden nodes generally contain edges which are redundant for the max-weighted path structure of $\mathcal{G}$; they then propose an algorithm for reducing a generic MLBN $\mathcal{G}$ to an equivalent representation $\mathcal{G}^O$, in which the number of hidden nodes is minimal. A first step towards minimizing the amount of knowledge about the true DAG $\mathcal{G}$ needed to construct compatible interventions could therefore be to work in the reduced graph $\mathcal{G}^O$ whenever possible; the feasibility of this is left for future work.

Another avenue for future work is the development of a σ^*-calculus equivalent to the ID-Algorithm [Shpitser and Pearl, 2006a, Huang and Valtorta, 2006] for determining identifiability of causal expressions via the successive application of σ^*-calculus rules. As there is no $*-$separation analogue of the notion of a $\mathcal{C}-$component, we expect the design, proof of soundness, and proof of completeness of such a procedure to be a very challenging problem.

Our considerations surrounding the non-positivity of joint and conditional distributions of MLBNs already hold true in the fully-observed setting. While this makes causal identifiability a highly non-trivial task even in MLBNs with no hidden nodes, there may be interesting work to be done in studying, for a fixed graph and edge weights, the regions of values attainable by $\mathbf{X}$ (and its conditional distributions) with positive probability. As is the case for the characterization of edge weight matrices on a fixed graph [Améndola and Ferry, 2026, Boege et al., 2026], it may be fruitful to do this through the lens of tropical or polyhedral geometry.

Author Contributions

Francesco Nowell and Leon Sierau contributed equally to the development of the main results and the writing of this manuscript. Nihat Ay contributed to finalising technical parts of the proofs. All authors have read and approved the final manuscript.

Acknowledgements

Leon Sierau was supported by the project “KiMeKo: Aufbau einer norddeutschen KI-Med-Kollaborationsplattform", funded by the Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR, German Federal Ministry for Research, Technology and Aeronautics) - project number 01 | S24056F.

Francesco Nowell was supported by the SPP 2458 "Combinatorial Synergies", funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 539875257).

The authors thank Carlos Améndola and Benjamin Hollering for their useful insights.

References

Carlos Améndola, Claudia Klüppelberg, Steffen Lauritzen, and Ngoc M. Tran. Conditional independence in max-linear Bayesian networks. *The Annals of Applied Probability*, 32(1):1 – 45, 2022. doi: 10.1214/21-AAP1670.

Carlos Améndola and Kamillo Ferry. Tropical combinatorics of max-linear bayesian networks. *Journal of Symbolic Computation*, 134:102518, 2026. ISSN 0747-7171. doi: 10.1016/j.jsc.2025.102518.

Peiman Asadi, Anthony C. Davison, and Sebastian Engelke. Extremes on river networks. *The Annals of Applied Statistics*, 9(4):2023 – 2050, 2015. doi: 10.1214/15-AOAS863.

Tobias Boege, Kamillo Ferry, Benjamin Hollering, and Francesco Nowell. Polyhedral aspects of maxoids. *arXiv preprint*, 2026. doi: 10.48550/arXiv.2504.21068.

Juan Correa and Elias Bareinboim. A Calculus for Stochastic Interventions:Causal Effect Identification and Surrogate Experiments. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(06):10093–10100, April 2020a. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v34i06.6567.

Juan D. Correa and Elias Bareinboim. General transportability of soft interventions: completeness results. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS ’20, Red Hook, NY, USA, 2020b. Curran Associates Inc. ISBN 9781713829546.

John H. J. Einmahl, Anna Kiriliouk, and Johan Segers. A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes*, 21(2): 205–233, June 2018. ISSN 1572-915X. doi: 10.1007/s10687-017-0303-7.

Sebastian Engelke and Adrien S. Hitz. Graphical models for extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4):871–932, 06 2020. ISSN 1369-7412. doi: 10.1111/rssb.12355.

Sebastian Engelke, Nicola Gnecco, and Frank Röttger. Extremes of structural causal models. *arXiv preprint*, 2025. doi: 10.48550/arXiv.2503.06536.

Patrick Forré and Joris M. Mooij. A mathematical introduction to causality: Lecture notes, July 2025. URL `https://staff.fnwi.uva.nl/j.m.mooij/articles/causality_lecture_notes_2025.pdf`. Lecture notes, version dated July 18, 2025.

Nadine Gissibl and Claudia Klüppelberg. Max-linear models on directed acyclic graphs. *Bernoulli*, 24(4A):2693 – 2720, 2018. doi: 10.3150/17-BEJ941.

Nadine Gissibl, Claudia Klüppelberg, and Steffen Lauritzen. Identifiability and estimation of recursive max-linear models. *Scandinavian Journal of Statistics*, 48(1):188–211, March 2021. ISSN 0303-6898, 1467-9469. doi: 10.1111/sjos.12446.

Yimin Huang and Marco Valtorta. Pearl’s calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI’06, page 217–224, Arlington, Virginia, USA, 2006. AUAI Press. ISBN 0974903922.

Inwoo Hwang, Yesong Choe, Yeahoon Kwon, and Sanghack Lee. On positivity condition for causal inference. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 20818–20841. PMLR, July 2024.

Claudia Klüppelberg and Steffen L. Lauritzen. Bayesian networks for max-linear models. In *Network Science: An Aerial View*, pages 79–97. Springer International Publishing, 2019. ISBN 978-3-030-26813-8. doi: 10.1007/978-3-030-26814-5_6.

Mario Krali, Anthony C. Davison, and Claudia Klüppelberg. Heavy-tailed max-linear structural equation models in networks with hidden nodes. *arXiv preprint*, 2025. doi: 10.48550/arXiv.2306.15356.

Steffen L Lauritzen. *Graphical Models*. Oxford University Press, 05 1996. ISBN 9780198522195. doi: 10.1093/oso/9780198522195.001.0001.

Steffen L. Lauritzen. *Lectures on Graphical Models, 3rd edition*. Department of Mathematical Sciences, Faculty of Science, University of Copenhagen, 2019. ISBN 978-87-70787-53-6. URL `https://web.math.ku.dk/~lauritzen/papers/gmnotes.pdf`.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995. doi: 10.1093/biomet/82.4.669.

Judea Pearl. *Causality*. Cambridge University Press, 2 edition, 2009. ISBN 9780521895606.

Thomas Richardson and Peter Spirtes. Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962 – 1030, 2002. doi: 10.1214/aos/1031689015.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI '06)*, pages 1219–1226, Boston, MA, USA, 2006a. AAAI Press. ISBN 0974903922.

Ilya Shpitser and Judea Pearl. Identification of conditional interventional distributions. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence (UAI)*, UAI'06, page 437–444, Arlington, Virginia, USA, 2006b. AUAI Press. ISBN 0974903922.

Jin Tian. Identifying conditional causal effects. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 561–568, Arlington, Virginia, USA, 2004. AUAI Press. ISBN 9780974903903.

Jin Tian and Judea Pearl. On the identification of causal effects. Technical Report R-290-L, Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA, 2003. URL `https://ftp.cs.ucla.edu/pub/stat_ser/R290-L.pdf`.

Ngoc Mai Tran, Johannes Buck, and Claudia Klüppelberg. Estimating a directed tree for extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(3):771–792, 07 2024. ISSN 1369-7412. doi: 10.1093/jrsssb/qkad165.

A Stronger Calculus of Intervention for Max-Linear Bayesian Networks (Supplementary Material)

Leon Sierau[1] **Francesco Nowell**[2] **Nihat Ay**[1,3,4,5]

[1]Technische Universität Hamburg, Institute for Data Science Foundations, 21079 Hamburg, Germany
[2]Technische Universität Berlin, Institute for Mathematics, 10623 Berlin, Germany
[3]Santa Fe Institute, Santa Fe, NM 87501, USA
[4]Complexity Science Hub, 1030 Vienna, Austria
[5]Leipzig University, 04109 Leipzig, Germany

A GRAPHICAL TERMINOLOGY

A *simple directed graph* $\mathcal{G} = (V, E)$ consists of nodes V and directed edges E, i.e., edges of the form $i \to j$ such that there is at most one edge between any two nodes. We say that two nodes i and j are adjacent if they are connected by an edge. The set of all nodes adjacent to i are the *neighbors* of i, denoted $\mathrm{ne}(i)$. A *path* between two nodes i and j (normally denoted π) is a sequence of distinct nodes $(i, \dots, j)$ such that all consecutive nodes in π are adjacent. We say $\pi = (k_0 = i, \dots, j = k_s)$ is a *directed path* from i to j if all edges on π are of the form $k_i \to k_{i+1}$, i.e, point towards j. A path π between i and j is called a *trek* if there exists a $k \in \pi$ such that the subpaths $(i, \dots k)$ and $(k, \dots, j)$ are directed paths from k to i and k to j respectively. A directed path from i to j and the edge $j \to i$ form a *directed cycle*. We call a simple directed graph without directed cycles a *directed acyclic graph* (DAG). Given an edge $i \to j$ we say that i is a *parent* of j and denote the set of parents of j with $\mathrm{pa}(j)$. The cardinality of $\mathrm{pa}(j)$ is the *in-degree* of j, denoted $\mathrm{indeg}(j)$. Given a directed path $i \to \cdots \to j$ we say that i is an *ancestor* of j, respectively j is a *descendant* of i. We denote the set of ancestors with $\mathrm{an}(j)$ and the set of descendants with $\mathrm{de}(i)$. For a subset of nodes $K \subset V$ we define $\mathrm{an}(K) := \cup_{k \in K} \mathrm{an}(k) \setminus K$ and $\mathrm{An}(K) := \mathrm{an}(K) \cup K$. The sets $\mathrm{de}(K)$ and $\mathrm{De}(K)$, and $\mathrm{pa}(K)$ are defined analogously. We say that a set $S \subset V$ is *ancestral* if $\mathrm{An}(S) = S$.

A.1 d-SEPARATION

Graphical models identify conditional independence relations through a separation condition $\perp$. The best-known instance of this in directed acyclic graphical models is *d-separation* for linear structural equation models. A triple of nodes $\{i, k, j\} \subset V$ form a *collider* in a DAG $\mathcal{G} = (V, E)$ if $i \to k \in E$ and $j \to k \in E$. An $i - j$ path π in a DAG is *d-connecting* given $K \subset V \setminus ij$ if all colliders along π lie in $\mathrm{An}(K)$ and no non-collider along π lies in K. If a d-connecting path exists, we say that i and j are *d-connected* given K. Otherwise, we say that they are *d-separated* and denote this with $(i \perp_d j \mid K)$. Whenever we consider multiple graphs at once and there is potential ambiguity in regards to which DAG a given separation statement holds in, we write $(i \perp_d j \mid K)_{\mathcal{G}}$ as shorthand for the statement "i and j are d-separated in $\mathcal{G}$, given K". The notion of $d-$separation of sets of nodes is defined analogously: we say that two sets $A, B \subset V$ are $d-$connected given $K \subset V$ if there exists a $d-$connecting path in $\mathcal{G}$ conditioned on K for some $a \in A$ and $b \in B$. If no such path exists, we say that A and B are d-separated in $\mathcal{G}$ given K and denote this by $(A \perp_{\mathrm{d}} B \mid K)$, adding the subscript $\mathcal{G}$ whenever there is ambiguity in regards to which graph the statement holds in.

B AUXILIARY RESULTS FOR THE PROOF OF THEOREM 4.1

B.1 PROBABILITY THEORY

Proposition B.1 ([Lauritzen, 2019], Proposition 2.48)**.** *Let $\mathcal{G} = (O \dot\cup U, E)$ be a DAG and let P be a distribution that factorizes according to $\mathcal{G}$. If $A \subseteq (O \cup U)$ is an ancestral set in $\mathcal{G}$, then the marginal distribution P_A factorizes according to the induced subgraph $\mathcal{G}_A$.*

Corollary B.2. *Let A be an ancestral set in $\mathcal{G}$. Proposition B.1 implies that intervening on variables outside of $\mathbf{X}_A$, i.e. changing mechanisms on $\mathbf{X}_{(O\cup U\setminus A)}$ has no effect on the marginal distribution $P_A(\mathbf{X}_A)$.*

The next Lemma formalizes an intuitively plausible result: For disjoint $A, B, Z \subseteq V$ (note that $B = \emptyset$ is an admissible choice) in a DAG $\mathcal{G} = (V, E)$ corresponding to a SCM, the conditional distribution $P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_{\mathrm{pa}(Z)}, \mathbf{X}_B)$ does not depend on the mechanisms that generate Z from its parents. Informally speaking: Once we fix a configuration of $(x_Z, x_{\mathrm{pa}(Z)})$ it is irrelevant for the remaining system how this configuration was generated and how likely it was.

Lemma B.3. *Let $\mathcal{G} = (V, E)$ be a DAG. For each $v \in V$, let $(\mathcal{X}_v, \mathcal{B}_v)$ be measurable spaces such that $\mathcal{B}_v$ is countably generated and let $\mathcal{X}_V := \times_{v\in V}\mathcal{X}_v$ and $\mathcal{B}_V := \otimes_{v\in V}\mathcal{B}_v$ be the product sigma-algebra on $\mathcal{X}_V$.*

Let σ_Z and σ'_Z be two policy interventions acting on the same set $Z \subseteq V$. Write $\mathrm{pa}^{\sigma_Z}(v)$ and $\mathrm{pa}^{\sigma'_Z}(v)$ for the parent sets of v in the graphs $\mathcal{G}_{\sigma_Z}$ and $\mathcal{G}_{\sigma'_Z}$, respectively. Fix an order $v_1, \ldots, v_n$ of V such that for every i,

$$\mathrm{pa}^{\sigma_Z}(v_i) \cup \mathrm{pa}^{\sigma'_Z}(v_i) \subseteq \{v_1, \ldots, v_{i-1}\}.$$

For each $v \in V$, let

$$\kappa_v^{\sigma_Z}(\,\cdot \mid x_{\mathrm{pa}^{\sigma_Z}(v)}) : \mathcal{X}_{\mathrm{pa}^{\sigma_Z}(v)} \times \mathcal{B}_v \to [0,1], \qquad \kappa_v^{\sigma'_Z}(\,\cdot \mid x_{\mathrm{pa}^{\sigma'_Z}(v)}) : \mathcal{X}_{\mathrm{pa}^{\sigma'_Z}(v)} \times \mathcal{B}_v \to [0,1]$$

be Markov kernels, and let P^{σ_Z} and $P^{\sigma'_Z}$ be the joint distributions on $(\mathcal{X}_V, \mathcal{B}_V)$ obtained by composing these kernels along $v_1, \ldots, v_n$ [Forré and Mooij, 2025, Cor. A.10.3].

Assume

$$\mathrm{pa}^{\sigma'_Z}(v) = \mathrm{pa}^{\sigma_Z}(v) \ \text{ and } \ \kappa_v^{\sigma'_Z} = \kappa_v^{\sigma_Z} \quad \text{for all } v \notin Z.$$

Let $A, B \subseteq V$ be disjoint from Z, write

$$\mathrm{pa}^{\sigma_Z}(Z) := \bigcup_{v\in Z} \mathrm{pa}^{\sigma_Z}(v), \qquad \mathrm{pa}^{\sigma'_Z}(Z) := \bigcup_{v\in Z} \mathrm{pa}^{\sigma'_Z}(v).$$

Let $C_Z = \mathbf{X}_{Z\cup \mathrm{pa}^{\sigma_Z}(Z)\cup \mathrm{pa}^{\sigma'_Z}(Z)\cup B}$ and let

$$\mathcal{C} := \sigma\big(C_Z\big)$$

be the smallest sigma-algebra generated by C_Z.

Then for every bounded measurable $\varphi : \mathcal{X}_A \to \mathbb{R}$ there exist versions of $\mathbb{E}_{P^{\sigma_Z}}[\varphi(\mathbf{X}_A) \mid \mathcal{C}]$ and $\mathbb{E}_{P^{\sigma'_Z}}[\varphi(\mathbf{X}_A) \mid \mathcal{C}]$ and a set $N \in \mathcal{B}_V$ with $P^{\sigma_Z}(N) = P^{\sigma'_Z}(N) = 0$ such that

$$\mathbb{E}_{P^{\sigma'_Z}}\big[\varphi(\mathbf{X}_A) \mid \mathcal{C}\big](x_V) = \mathbb{E}_{P^{\sigma_Z}}\big[\varphi(\mathbf{X}_A) \mid \mathcal{C}\big](x_V) \quad \text{for all } x_V \in N^c.$$

Proof. For $v \in Z$ let $S_v := \mathrm{pa}^{\sigma_Z}(v) \cup \mathrm{pa}^{\sigma'_Z}(v)$. For $v \in V$ define

$$P_v^{\sigma_Z}(D \mid x_{\mathrm{pa}^{\sigma_Z}(v)}) := \kappa_v^{\sigma_Z}(D \mid x_{\mathrm{pa}^{\sigma_Z}(v)}), \qquad P_v^{\sigma'_Z}(D \mid x_{\mathrm{pa}^{\sigma'_Z}(v)}) := \kappa_v^{\sigma'_Z}(D \mid x_{\mathrm{pa}^{\sigma'_Z}(v)}), \qquad D \in \mathcal{B}_v.$$

For $v \in Z$ we also view these kernels as functions on x_{S_v} by ignoring irrelevant coordinates, i.e.,

$$P_v^{\sigma_Z}(D \mid x_{S_v}) := \kappa_v^{\sigma_Z}\big(D \mid x_{\mathrm{pa}^{\sigma_Z}(v)}\big), \qquad P_v^{\sigma'_Z}(D \mid x_{S_v}) := \kappa_v^{\sigma'_Z}\big(D \mid x_{\mathrm{pa}^{\sigma'_Z}(v)}\big), \qquad D \in \mathcal{B}_v.$$

Step 1. Define Markov kernels P_v^0 by setting $P_v^0 := P_v^{\sigma_Z} = P_v^{\sigma'_Z}$ for $v \notin Z$, and for $v \in Z$,

$$P_v^0(D \mid x_{S_v}) := \frac{1}{2} P_v^{\sigma_Z}(D \mid x_{S_v}) + \frac{1}{2} P_v^{\sigma'_Z}(D \mid x_{S_v}), \qquad D \in \mathcal{B}_v.$$

Then for every $v \in Z$ and every x_{S_v} we have

$$P_v^{\sigma_Z}(\cdot \mid x_{S_v}) \ll P_v^0(\cdot \mid x_{S_v}), \qquad P_v^{\sigma'_Z}(\cdot \mid x_{S_v}) \ll P_v^0(\cdot \mid x_{S_v}),$$

by construction. Since $\mathcal{B}_v$ is countably generated, the Doob-Radon-Nikodym theorem for kernels [Forré and Mooij, 2025, Thm. 2.3.17] yields jointly measurable

$$\ell_v^{\sigma Z} : \mathcal{X}_v \times \mathcal{X}_{S_v} \to [0, \infty), \qquad \ell_v^{\sigma' Z} : \mathcal{X}_v \times \mathcal{X}_{S_v} \to [0, \infty),$$

such that for all $D \in \mathcal{B}_v$ and all $x_{S_v} \in \mathcal{X}_{S_v}$,

$$P_v^{\sigma Z}(D \mid x_{S_v}) = \int_D \ell_v^{\sigma Z}(x_v, x_{S_v})\, P_v^0(dx_v \mid x_{S_v}), \qquad P_v^{\sigma' Z}(D \mid x_{S_v}) = \int_D \ell_v^{\sigma' Z}(x_v, x_{S_v})\, P_v^0(dx_v \mid x_{S_v}).$$

Define

$$L^{\sigma Z}(x_V) := \prod_{v \in Z} \ell_v^{\sigma Z}(x_v, x_{S_v}), \qquad L^{\sigma' Z}(x_V) := \prod_{v \in Z} \ell_v^{\sigma' Z}(x_v, x_{S_v}).$$

Then $L^{\sigma Z}$ and $L^{\sigma' Z}$ are $\mathcal{C}$-measurable.

Step 2. Let P^0 be the joint distribution on $(\mathcal{X}_V, \mathcal{B}_V)$ obtained by composing the kernels $\{P_{v_i}^0\}_{i=1}^n$ along $v_1, \ldots, v_n$ [Forré and Mooij, 2025, Cor. A.10.3]. For bounded measurable $f : \mathcal{X}_V \to \mathbb{R}$ we have

$$\int f \, dP^0 = \int \cdots \int f(x_{v_1}, \ldots, x_{v_n}) \prod_{i=1}^n P_{v_i}^0\big(dx_{v_i} \mid x_{\mathrm{pa}^0(v_i)}\big),$$

Substituting the kernel Radon-Nikodym representations from Step 1 into the iterated integrals for $P^{\sigma Z}$ and $P^{\sigma' Z}$ yields

$$\int f \, dP^{\sigma Z} = \int f \, L^{\sigma Z} \, dP^0, \qquad \int f \, dP^{\sigma' Z} = \int f \, L^{\sigma' Z} \, dP^0. \tag{B.1}$$

Step 3. Fix bounded measurable $\varphi : \mathcal{X}_A \to \mathbb{R}$ and set

$$U := \mathbb{E}_{P^0}[\varphi(\mathbf{X}_A) \mid \mathcal{C}].$$

Then for every bounded $\mathcal{C}$-measurable random variable T,

$$\mathbb{E}_{P^0}[\varphi(\mathbf{X}_A)\, T] = \mathbb{E}_{P^0}[U\, T]. \tag{B.2}$$

Now fix $H \in \mathcal{C}$. Since $L^{\sigma Z}$ is nonnegative and $\mathcal{C}$-measurable, for $m \in \mathbb{N}$ the random variable $T_m := \mathbf{1}_H\,(L^{\sigma Z} \wedge m) = \mathbf{1}_H \min\{L^{\sigma Z}, m\}$ is bounded and $\mathcal{C}$-measurable. Thus by (B.2),

$$\mathbb{E}_{P^0}\big[\varphi(\mathbf{X}_A)\, \mathbf{1}_H\,(L^{\sigma Z} \wedge m)\big] = \mathbb{E}_{P^0}\big[U\, \mathbf{1}_H\,(L^{\sigma Z} \wedge m)\big].$$

Since φ is bounded, letting $m \to \infty$ and applying monotone convergence gives

$$\mathbb{E}_{P^0}[\varphi(\mathbf{X}_A)\, \mathbf{1}_H\, L^{\sigma Z}] = \mathbb{E}_{P^0}[U\, \mathbf{1}_H\, L^{\sigma Z}].$$

Using (B.1) on both sides yields

$$\mathbb{E}_{P^{\sigma Z}}[\varphi(\mathbf{X}_A)\, \mathbf{1}_H] = \mathbb{E}_{P^{\sigma Z}}[U\, \mathbf{1}_H].$$

Since this holds for all $H \in \mathcal{C}$, almost sure uniqueness of conditional expectation implies

$$\mathbb{E}_{P^{\sigma Z}}[\varphi(\mathbf{X}_A) \mid \mathcal{C}] = U \quad P^{\sigma Z}\text{-a.s..}$$

The same argument with $L^{\sigma' Z}$ instead of $L^{\sigma Z}$ shows that

$$\mathbb{E}_{P^{\sigma'_Z}}[\varphi(\mathbf{X}_A) \mid \mathcal{C}] = U \quad P^{\sigma' Z}\text{-a.s..}$$

Combining the two equalities, we have that

$$\mathbb{E}_{P^{\sigma'_Z}}[\varphi(\mathbf{X}_A) \mid \mathcal{C}] = \mathbb{E}_{P^{\sigma Z}}[\varphi(\mathbf{X}_A) \mid \mathcal{C}]$$

outside of a union of $P^{\sigma Z}$- and $P^{\sigma' Z}$-Nullsets in $\mathcal{C}$. $\square$

Corollary B.4. *Let $C_{Z\setminus B} := \mathbf{X}_{Z\cup \mathrm{pa}^{\sigma Z}(Z)\cup \mathrm{pa}^{\sigma' Z}(Z)\cup B\setminus B}$ and let C_Z be defined as in Lemma B.3. Let $A, B \subseteq V$ be disjoint from Z, and let*

$$\mathcal{B} := \sigma(\mathbf{X}_B), \qquad \mathcal{C} := \sigma(C_Z).$$

Then Lemma B.3 immediately implies that

$$P^{\sigma Z}\big(\mathbf{X}_A \in \cdot \mid \mathcal{C}\big) = P^{\sigma' Z}\big(\mathbf{X}_A \in \cdot \mid \mathcal{C}\big).$$

outside of $P^{\sigma Z}$- and $P^{\sigma' Z}$-Nullsets in $\mathcal{C}$.

If moreover it holds that $(\mathbf{X}_A \perp\!\!\!\perp C_{Z\setminus B} \mid \mathbf{X}_B)$, we even have that

$$P^{\sigma Z}\big(\mathbf{X}_A \in \cdot \mid \mathcal{B}\big) = P^{\sigma' Z}\big(\mathbf{X}_A \in \cdot \mid \mathcal{B}\big)$$

outside of $P^{\sigma Z}$- and $P^{\sigma' Z}$-Nullsets in $\mathcal{B}$.

B.2 GRAPH THEORY

In this section, we prove a series of auxiliary results concerning the comparison of $*-$separation across DAGs which differ only in the parent set of a subset of nodes Z. We start by concretizing the latter notion.

Definition B.5. Let $\mathcal{G} = (V, E)$ and $\mathcal{G}' = (V, E')$ be DAGs on the same node set V, and fix $Z \subset V$. We say that $\mathcal{G}$ and $\mathcal{G}'$ are *locally equivalent outide of* Z (denoted $\mathcal{G} \equiv_Z \mathcal{G}'$) if the following condition holds:

$$\mathrm{pa}_{\mathcal{G}}(x) = \mathrm{pa}_{\mathcal{G}'}(x) \text{ for all } x \in V \setminus Z.$$

We collect some observations about $\equiv_Z$.

Remark B.6. Let $\mathcal{G} = (V, E)$ and $\mathcal{G} = (V, E')$ be graphs on the same node set, and define their *union* $\mathcal{G} \cup \mathcal{G}' := (V, E \cup E')$. For fixed $Z \subset V$, the following hold:

(i) The relation $\equiv_Z$ is an equivalence relation on the set of DAGs on V.

(ii) If $\mathcal{G} \equiv_Z \mathcal{G}'$, then $\mathcal{G} \cup \mathcal{G}' \equiv_Z \mathcal{G} \equiv_Z \mathcal{G}'$.

(iii) Removing a subset of edges common to both $\mathcal{G}$ and $\mathcal{G}'$ preserves $\equiv_Z$. Expressed rigorously: for $F \subset E \cap E'$ and $\mathcal{G}^{\setminus F} := (V, E \setminus F)$, $\mathcal{G}'^{\setminus F} := (V, E' \setminus F)$, it holds that

$$\mathcal{G} \equiv_Z \mathcal{G} \implies \mathcal{G}^{\setminus F} \equiv_Z \mathcal{G}^{\setminus F}.$$

In particular, local equivalence outside of Z is preserved under either of the graph surgeries defined in Definition 2.7.

We also state some observations about the $\perp_*$ separation condition that will be used several times.

Remark B.7. Let $\mathcal{G} = (V, E)$ be a DAG.

(i) The relation $\perp_*$ fulfills the *compositional graphoid* axioms (see Proposition 4.6 of [Boege et al., 2026]). In particular, *decomposition* holds, i.e:

$$(A \perp_* B_1 \cup B_2 \mid C)_{\mathcal{G}} \implies (A \perp_* B_1 \mid C)_{\mathcal{G}} \text{ and } (A \perp_* B_2 \mid C)_{\mathcal{G}} \tag{B.3}$$

for all disjoint $A, B_1, B_2, C \subset V$.

(ii) If $\mathcal{G}' = (V, E')$ is a *subgraph* of $\mathcal{G}$ (i.e. $E' \subset E$), then

$$(A \perp_* B \mid C)_{\mathcal{G}} \implies (A \perp_* B \mid C)_{\mathcal{G}'} \tag{B.4}$$

for any disjoint $A, B, C \subset V$. This becomes self-evident when one considers the contrapositive implication; clearly any $*-$connecting path in $\mathcal{G}'$ is also $*-$connecting in $\mathcal{G}$ due to $E' \subset E$.

(iii) Any subpath of a $*-$connecting path is itself $*-$connecting.

The first Lemma a general result concerning the translation of $\perp_*$ results to the union of two DAGs.

Lemma B.8. *Let $\mathcal{G} = (V, E)$ and $\mathcal{G}' = (V, E')$ be locally equivalent outside of Z and $\mathcal{G} \cup \mathcal{G}'$ defined as in Remark B.6. For all $A, B \subset V \setminus Z$.*

$$(Z \perp_* A|B) \text{ holds in } \mathcal{G} \text{ and } \mathcal{G}' \implies (Z \perp_* A|B) \text{ holds in } \mathcal{G} \cup \mathcal{G}'.$$

Proof. We prove the statement by contraposition. Assume that there exists a $*-$connecting path π between Z and A in $\mathcal{G} \cup \mathcal{G}'$ given B, and assume without restriction that the first node z is the only node in Z contained in it (if this is not the case, then one may consider instead a subpath). That is, π is of the form

$$\pi : z - x - \cdots - a$$

where $z \in Z$, $x \notin Z$, $a \in A$ and every $-$ stands for either $\leftarrow$ or $\rightarrow$. Denote the portion of the path from x to a as $\tilde{\pi}$. Because it is a subpath of π , $\tilde{\pi}$ is a $*-$connecting from x to a in $\mathcal{G} \cup \mathcal{G}'$ given B. Furthermore, as $\mathcal{G}$ and $\mathcal{G}'$ have edge sets which are identical up to those incoming to Z, and $\tilde{\pi}$ contains no edges of this form, $\tilde{\pi}$ is in fact a $*-$connecting path given B in both $\mathcal{G}$ and $\mathcal{G}'$. Finally, the edge connecting z and x in π is, by construction of $\mathcal{G} \cup \mathcal{G}'$, present in $\mathcal{G}$ or $\mathcal{G}'$. This implies that the whole path π is contained in $\mathcal{G}$ or $\mathcal{G}'$ and $*-$connecting given B. $\square$

We now prove results specific to the settings of Rules 2 and 3 of $\sigma*$-calculus. Recall from Definition 2.7 that $\mathcal{G}_{\underline{Z}}$ is the subgraph of $\mathcal{G}$ obtained by deleting all outgoing edges from $Z \subset O$

Lemma B.9. *Let $\mathcal{G}$ be a DAG and A, B, Z be disjoint subsets of vertices. We define $\mathrm{pa}_{\mathcal{G}\setminus B}(Z) := \mathrm{pa}_{\mathcal{G}}(Z) \setminus (Z \cup B)$. Then*

$$(Z \perp_* A \mid B) \text{ in } \mathcal{G}_{\underline{Z}} \implies (\mathrm{pa}_{\mathcal{G}\setminus B}(Z) \perp_* A \mid Z \cup B) \text{ in } \mathcal{G}.$$

Proof. We prove the statement by contraposition. Let $\pi : x - \cdots - a$ be a $*-$connecting path between $x \in \mathrm{pa}_{\mathcal{G}\setminus B}(Z)$ and $a \in A$ in $\mathcal{G}$, given $Z \cup B$. If no intermediate nodes along π are contained in Z, then π may be extended via the edge $z \leftarrow x$ for some $z \in Z$. This edge exists due to the assumption that $x \in \mathrm{pa}_{\mathcal{G}\setminus B}(Z)$, and the resulting path $\pi' : z \leftarrow x - \cdots - a$ is $*-$connecting in both $\mathcal{G}$ and $\mathcal{G}_{\underline{Z}}$, as $z \leftarrow x$ exists in $\mathcal{G}_{\underline{Z}}$, and appending it to the path does not give rise to additional colliders. If, on the other hand, π contains an intermediate node $z \in Z$, then z must be the unique collider along π in order for it to be $*-$connecting in $\mathcal{G}$ given $Z \cup B$. However in this case, the portion of π between z and a is $*-$connecting in $\mathcal{G}$ given $Z \cup B$, has no colliders, and is in particular not blocked by B. As the first edge of this portion is an incoming edge to z, it is also present in $\mathcal{G}_Z$ and $*-$connecting given B due to the observations above. In either of these two cases $Z \not\perp_* A|B$ holds, proving the lemma. $\square$

Lemma B.10. *Let $\mathcal{G}$ and $\mathcal{G}'$ be locally equivalent outside of Z and let $A, B \subset V \setminus Z$ be disjoint. The following implication holds:*

$$(Z \perp_* A \mid B) \text{ in } \mathcal{G}_{\underline{Z}} \text{ and } \mathcal{G}'_{\underline{Z}} \implies (\mathrm{pa}_{(\mathcal{G}\cup\mathcal{G}')\setminus B}(Z) \perp_* A \mid Z \cup B)_{\mathcal{G}} \text{ in } \mathcal{G} \text{ and } \mathcal{G}'.$$

Proof. Let $(Z \perp_* A \mid B)_{\mathcal{G}_{\underline{Z}}}$ and $(Z \perp_* A \mid B)_{\mathcal{G}'_{\underline{Z}}}$ hold. By point (iii) of Remark B.6 it holds that $\mathcal{G}_{\underline{Z}} \equiv_Z \mathcal{G}'_{\underline{Z}}$, and applying Lemma B.8 gives $(Z \perp_* A \mid B)_{(\mathcal{G}\cup\mathcal{G}')_{\underline{Z}}}$. Applying Lemma B.9 yields $(\mathrm{pa}_{(\mathcal{G}\cup\mathcal{G}')\setminus B}(Z) \perp_* A \mid Z \cup B)_{\mathcal{G}\cup\mathcal{G}'}$. From this one can conclude both $(\mathrm{pa}_{(\mathcal{G}\cup\mathcal{G}')\setminus B}(Z) \perp_* A \mid Z \cup B)_{\mathcal{G}}$ and $(\mathrm{pa}_{(\mathcal{G}\cup\mathcal{G}')\setminus B}(Z) \perp_* A \mid Z \cup B)_{\mathcal{G}'}$ by applying the implication in (B.4) twice .This concludes the proof. $\square$

Lemma B.11. *Let A, B, Z be distinct node sets in a DAG $\mathcal{G}$ and consider the partition $Z = Z_0 \dot{\cup} Z_1$, where $Z_0 := Z \setminus \mathrm{an}_{\mathcal{G}}(B)$ and $Z_1 := Z \cap \mathrm{an}_{\mathcal{G}}(B)$. We further assume $Z_0 \cap \mathrm{an}_{\mathcal{G}}(A) = \emptyset$. The following implication holds:*

$$(A \perp_* Z \mid B)_{\mathcal{G}_{\overline{Z}_0}} \implies (A \perp_* Z_1 \cup \mathrm{pa}_{\mathcal{G}}(Z_1) \mid B)_{\mathcal{G}}$$

Proof. We first show

$$(A \perp_* Z \mid B)_{\mathcal{G}_{\overline{Z}_0}} \implies (A \perp_* Z_1 \cup \mathrm{pa}_{\mathcal{G}}(Z_1) \mid B)_{\mathcal{G}_{\overline{Z}_0}} \tag{B.5}$$

by showing

$$(A \perp_* Z \mid B)_{\mathcal{G}_{\overline{Z}_0}} \implies (A \perp_* Z_1 \mid B)_{\mathcal{G}_{\overline{Z}_0}} \text{ and} \tag{B.6}$$

$$(A \perp_* Z \mid B)_{\mathcal{G}_{\overline{Z}_0}} \implies (A \perp_* \mathrm{pa}_{\mathcal{G}}(Z_1) \mid B)_{\mathcal{G}_{\overline{Z}_0}} \tag{B.7}$$

separately. The implication (B.6) immediately follows from $Z_1 \subset Z$ the decomposition property of $\perp_*$ (B.3). We prove (B.7) by contraposition. Let $\pi : p - \cdots - a$ be a $*-$connecting path given B in $\mathcal{G}_{\overline{Z}_0}$, where $p \in \mathrm{pa}_{\mathcal{G}}(Z_1)$ and $a \in A$. If π contains an intermediate node $z \in Z$, then the portion of π from z to a is $*-$connecting, immediately implying $(A \not\perp_* Z \mid B)_{\mathcal{G}_{\overline{Z}_0}}$ as desired. If π is disjoint from Z, then it may be extended to a path $\pi' : z \leftarrow p - \cdots - a$, where $z \in Z_1 \subset Z$ is the child of p which exists due to the assumption $p \in \mathrm{pa}_{\mathcal{G}}(Z_1)$. The path π' clearly has the same colliders as π, and is thus also a $*-$connecting path, proving $(A \not\perp_* Z \mid B)_{\mathcal{G}_{\overline{Z}_0}}$.

To prove the lemma, it remains to show that the right hand side of (B.5) implies $(A \perp_* Z_1 \cup \mathrm{pa}_{\mathcal{G}}(Z_1) \mid B)_{\mathcal{G}}$. We prove this by showing that no $*-$connecting path between $Z_1 \cup \mathrm{pa}_{\mathcal{G}}(Z_1)$ and A given B can contain an edge of the form $x \to z_o$ for some $z_0 \in Z_0$. Indeed, assume (towards a contradiction) that such a path $\pi : z - \ldots x \to z_0 - \cdots - a$ exists. Denote the portion of π between z_0 and a by π' and observe that π' is a $*$-connecting path given B. The first edge of π' must be of the form $z_0 \to y$, as otherwise z_0 would be the unique collider in the original path π which is $*-$connecting given B, implying $z_0 \in \mathrm{an}_{\mathcal{G}}(B)$ and thus a contradiction. However π' cannot be a directed path from z_0 to a, as this would contradict the assumption that $Z_0 \cap \mathrm{an}_{\mathcal{G}}(A) = \emptyset$. The only option remaining is that π' is of the form $z_0 \to \cdots \to k \leftarrow \cdots \leftarrow a$, containing a collider $k \in \mathrm{an}_{\mathcal{G}}(B)$ which is a descendant of z_0, but this would once again imply $z_0 \in \mathrm{an}_{\mathcal{G}}(B)$ and thus a contradiction. Thus, any $*-$connecting path between $Z_1 \cup \mathrm{pa}_{\mathcal{G}}(Z_1)$ and A given B in $\mathcal{G}$ is also $*-$connecting given B in $\mathcal{G}_{\overline{Z}_0}$. This implies the remaining implication by contraposition. □

Lemma B.12. *Let A, B, Z be node sets in a DAG $\mathcal{G}$ fulfilling the conditions of Lemma B.11, with Z partitioned once again into $Z = Z_0 \dot\cup Z_1$, where $Z_0 := Z \setminus \mathrm{an}_{\mathcal{G}}(B)$ and $Z_1 := Z \cap \mathrm{an}_{\mathcal{G}}(B)$. Let $\mathcal{G}'$ be locally equivalent to $\mathcal{G}$ outside of Z. The following implication holds:*

$$(A \perp_* Z \mid B) \text{ in } \mathcal{G}_{\overline{Z}_0} \text{ and } \mathcal{G}'_{\overline{Z}_0} \implies (A \perp_* Z_1 \cup \mathrm{pa}_{\mathcal{G}\setminus B}(Z_1) \cup \mathrm{pa}_{\mathcal{G}'\setminus B}(Z_1) \mid B) \text{ in } \mathcal{G} \text{ and } \mathcal{G}'.$$

Proof. The proof is analogous to that of Lemma B.10. The left hand side of the implication implies $(A \perp_* Z \mid B)_{(\mathcal{G}\cup\mathcal{G}')_{\overline{Z}_0}}$ by Lemma B.8, which in turn implies $(A \perp_* Z_1 \cup \mathrm{pa}_{\mathcal{G}\cup\mathcal{G}'}(Z_1) \mid B)_{\mathcal{G}\cup\mathcal{G}'}$ by Lemma B.9 (observe again that $\mathcal{G}_{\overline{Z}_0} \equiv_Z \mathcal{G}_{\overline{Z}_0}$ follows again from Remark B.6(iii)). The statement remains true upon passing to subgraphs of $\mathcal{G} \cup \mathcal{G}'$, therefore $(A \perp_* Z_1 \cup \mathrm{pa}_{\mathcal{G}\cup\mathcal{G}'}(Z_1) \mid B)_{\mathcal{G}}$ and $(A \perp_* Z_1 \cup \mathrm{pa}_{\mathcal{G}\cup\mathcal{G}'}(Z_1) \mid B)_{\mathcal{G}'}$ hold, and observing that $\mathrm{pa}_{\mathcal{G}\cup\mathcal{G}'}(Z_1) = \mathrm{pa}_{\mathcal{G}}(Z_1) \cup \mathrm{pa}_{\mathcal{G}'}(Z_1)$ completes the proof. □

Observation B.13. Let Z_0, Z_1 be defined as in Lemma B.11 (assuming again that $Z_0 \cap \mathrm{an}_{\mathcal{G}}(A) = \emptyset$) and let $S := (A \cup B \cup Z_1 \cup ((\mathrm{pa}^{\sigma_Z}(Z_1) \cup \mathrm{pa}^{\sigma'_Z}(Z_1)) \setminus B))$.

By definition, Z_0 has no descendants in B and can also not have descendants in $S \setminus A$, since these all have descendants in B by definition.

By the assumption $(A \perp_* Z \mid B)_{\mathcal{G}_{\overline{Z}_0}}$, Z_0 can also not have any descendants in A. In order to see this, assume that there existed a directed path $\pi : Z_0 \rightsquigarrow A$. Since Z_0 has no ancestors in B, π cannot pass through B. But this would imply $*$-connectedness of Z and A, given B in $\mathcal{G}_{\overline{Z_0}}$, contradicting the assumption.

This together with Corollary B.2 implies that the distribution $P(\mathbf{X}_S)$ is independent of the mechanisms generating Z_0 from its parents. Formally that is

$$P(\mathbf{X}_S; \sigma_{Z_0}) = P(\mathbf{X}_S; \sigma'_{Z_0})$$

and in particular

$$P(\mathbf{X}_A \mid \mathbf{X}_{S\setminus A}; \sigma_{Z_0}) = P(\mathbf{X}_S \mid \mathbf{X}_{S\setminus A}; \sigma'_{Z_0}).$$

C PROOF OF THEOREM 4.1

Remark C.1. In the proofs of rules 2 and 3, we compare two SCMs $\mathcal{M}_{\sigma_Z}$ and $\mathcal{M}_{\sigma'_Z}$ by comparing their graphs $\mathcal{G}_{\sigma_Z}$ and $\mathcal{G}_{\sigma'_Z}$. Graphically, performing the intervention σ_Z on a DAG $\mathcal{G} = (O \cup U, E)$ equates to modifying the parent set of some $Z \subset O$, potentially introducing unobserved ancestors of Z in the process (see also Remark 2.4). This results in a post-intervention graph $\mathcal{G}_{\sigma_Z} = (O \cup U_{\sigma_Z}, E_{\sigma_Z})$, where $U \subset U_{\sigma_Z}$. While clearly performing two distinct interventions σ_Z and σ'_Z generally leads to $U_{\sigma_Z} \neq U_{\sigma'_Z}$, we implicitly identify $\mathcal{G}_{\sigma_Z}$ and $\mathcal{G}_{\sigma'_Z}$ with their respective *embeddings* in $V_{\sigma_Z \sigma'_Z} := O \cup U_{\sigma_Z} \cup U_{\sigma'_Z}$, i.e. the DAGs $\overline{\mathcal{G}}_{\sigma_Z} := (V_{\sigma_Z \sigma'_Z}, E_{\sigma_Z})$ and $\overline{\mathcal{G}}_{\sigma'_Z} := (V_{\sigma_Z \sigma'_Z}, E_{\sigma'_Z})$, noting that $\mathcal{M}_{\sigma_Z}$ and $\mathcal{M}_{\sigma'_Z}$ are SCMs on $\overline{\mathcal{G}}_{\sigma_Z}$ and $\overline{\mathcal{G}}_{\sigma'_Z}$ respectively; this holds because nodes in $U_{\sigma_Z} \setminus U_{\sigma'_Z}$ are isolated in $\overline{\mathcal{G}}_{\sigma'_Z}$ and vice versa. As the two embeddings are defined on a common set, this identification allows us to apply the results from Appendix B.2 to $\mathcal{G}_{\sigma_Z}$ and $\mathcal{G}_{\sigma'_Z}$, as well as their post-surgery counterparts $\mathcal{G}_{\sigma_Z \underline{Z}}$, $\mathcal{G}_{\sigma'_Z \underline{Z}}$ and $\mathcal{G}_{\sigma_Z \overline{Z_0}}$ and $\mathcal{G}_{\sigma'_Z \overline{Z_0}}$.

C.1 PROOF OF THE STRENGTHENED RULE 1

We want to show that, granted that the intervention σ_Z leads to a SCM which is Markov with respect to $*$-separation, the following implication holds:

$$(A \perp_* B|C)_{\mathcal{G}_{\sigma_Z}} \implies P(\mathbf{X}_A \mid \mathbf{X}_B, \mathbf{X}_C; \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_C; \sigma_Z).$$

Proof. The proof is immediate, but included here for completeness.

$(A \perp_* B|C)_{\mathcal{G}_{\sigma_Z}} \implies (\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C)_{\mathcal{M}^{\sigma_Z}}$, where M^{σ_Z} is the SCM induced by the intervention σ_Z. This is equivalent to the statement we wanted to show. □

C.2 PROOF OF THE STRENGTHENED RULE 2

We want to show that the following implication

$$(A \perp_* Z \mid B)_{\mathcal{G}_{\sigma_W \sigma_Z \underline{Z}}} \text{ and } (A \perp_* Z \mid B)_{\mathcal{G}_{\sigma_W \sigma'_Z \underline{Z}}} \implies P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B; \sigma_W, \sigma_Z) \;=\; P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B; \sigma_W, \sigma'_z)$$

holds outside of P^{σ_Z}- and $P^{\sigma'_Z}$-Nullsets in $\sigma(\mathbf{X}_Z, \mathbf{X}_B)$, under the assumption that $\sigma_W, \sigma_Z, \sigma'_Z$ induce SCM that are Markov with respect to $*$-separation.

Proof. For disjoint $A, B, W, Z \subseteq O$, assume

$$(A \perp_* Z \mid B)_{\mathcal{G}_{\sigma_W \sigma_Z \underline{Z}}} \text{ and } (A \perp_* Z \mid B)_{\mathcal{G}_{\sigma_W \sigma'_Z \underline{Z}}}. \tag{C.1}$$

By Lemma B.8, (C.1) implies that

$$(A \perp_* Z \mid B)_{\mathcal{G}_{\sigma_W \sigma_Z \underline{Z}} \cup \mathcal{G}_{\sigma_W \sigma'_Z \underline{Z}}}. \tag{C.2}$$

And by Lemma B.10, (C.2) gives us

$$(A \perp_* (\mathrm{pa}^{\sigma_Z}(Z) \cup \mathrm{pa}^{\sigma'_Z}(Z)) \setminus B \mid B \cup Z)_{\mathcal{G}_{\sigma_W \sigma_Z \underline{Z}}} \tag{C.3}$$

and

$$(A \perp_* (\mathrm{pa}^{\sigma_Z}(Z) \cup \mathrm{pa}^{\sigma'_Z}(Z)) \setminus B \mid B \cup Z)_{\mathcal{G}_{\sigma_W \sigma'_Z \underline{Z}}}. \tag{C.4}$$

Now we can use Lemma B.3 in order to see that

$$P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B, \mathbf{X}_{(\mathrm{pa}^{\sigma_Z}(Z) \cup \mathrm{pa}^{\sigma'_Z}(Z)) \setminus B}\,;\, \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B, \mathbf{X}_{(\mathrm{pa}^{\sigma_Z}(Z) \cup \mathrm{pa}^{\sigma'_Z}(Z)) \setminus B}\,;\, \sigma_W, \sigma'_Z) \tag{C.5}$$

outside of $P^{\sigma_Z}(N)$- and $P^{\sigma'_Z}$-Nullsets in $\sigma(\mathbf{X}_Z, \mathbf{X}_B, \mathbf{X}_{(\mathrm{pa}^{\sigma_Z}(Z) \cup \mathrm{pa}^{\sigma'_Z}(Z)) \setminus B})$.

Finally, (C.3) and (C.4), together with the fact that the models are Markov with respect to $*$-separation, allow us to use the implied conditional independence of $(\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{(\mathrm{pa}^{\sigma_Z}(Z) \cup \mathrm{pa}^{\sigma'_Z}(Z)) \setminus B} \mid \mathbf{X}_Z, \mathbf{X}_B)$.

Now corollary B.4 yields

$$P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B \ ; \ \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_Z, \mathbf{X}_B \ ; \ \sigma_W, \sigma'_Z) \tag{C.6}$$

outside of P^{σ_Z}- and $P^{\sigma'_Z}$-Nullsets in $\sigma(\mathbf{X}_Z, \mathbf{X}_B)$, which is exactly what we wanted to show.

□

C.3 PROOF OF THE STRENGTHENED RULE 3

We want to show that the following implication

$$(A \perp_* Z \mid B)_{\mathcal{G}_{\sigma_W \sigma_Z \overline{Z(B)}}} \text{ and } (A \perp_* Z \mid B)_{\mathcal{G}_{\sigma_W \sigma'_Z \overline{Z(B)}}} \implies P(\mathbf{X}_A \mid \mathbf{X}_B; \sigma_W, \sigma_Z) \ = \ P(\mathbf{X}_A \mid \mathbf{X}_B; \sigma_W, \sigma'_z)$$

holds outside of P^{σ_Z}- and $P^{\sigma'_Z}$-Nullsets in $\sigma(\mathbf{X}_B)$, under the assumption that $\sigma_W, \sigma_Z, \sigma'_Z$ induce SCM that are Markov with respect to $*$-separation.

Proof. For disjoint $A, B, W, Z \subseteq O$, assume

$$(A \perp_* Z \mid B)_{\mathcal{G}_{\sigma_W \sigma_Z \overline{Z(B)}}} \text{ and } (A \perp_* Z \mid B)_{\mathcal{G}_{\sigma_W \sigma'_Z \overline{Z(B)}}}, \tag{C.7}$$

where $Z(B) = Z \setminus \mathrm{an}_{\mathcal{G}}(B)$. For brevity, let $Z_0 := Z(B)$ and $Z_1 = Z \setminus Z_0$, as in Lemma B.11.
Also let $S := (A \cup B \cup Z_1 \cup ((\mathrm{pa}^{\sigma_Z}(Z_1) \cup \mathrm{pa}^{\sigma'_Z}(Z_1)) \setminus B))$, as in Observation B.13.

Observation B.13 implies that

$$P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A}; \sigma_W, \sigma_{Z_1}, \sigma_{Z_0}) = P(\mathbf{X}_S \mid \mathbf{X}_{S \setminus A}; \sigma_W, \sigma_{Z_1} \sigma'_{Z_0}). \tag{C.8}$$

By Lemma B.12, (C.7) implies that

$$(A \perp_* S \setminus (B \cup A) \mid B)_{\mathcal{G}_{\sigma_W \sigma_Z}} \text{ and } (A \perp_* S \setminus (B \cup A) \mid B)_{\mathcal{G}_{\sigma_W \sigma'_Z}}. \tag{C.9}$$

This, together with the fact that the Models are Markov with respect to $*$-separation, yields

$$P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A} \ ; \ \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_B \ ; \ \sigma_Z) \text{ and } P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A} \ ; \ \sigma_W, \sigma'_Z) = P(\mathbf{X}_A \mid \mathbf{X}_B \ ; \ \sigma'_Z). \tag{C.10}$$

Now we write

$$P(\mathbf{X}_A \mid \mathbf{X}_B \ ; \ \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A} \ ; \ \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A} \ ; \ \sigma_W, \sigma_{Z_1}, \sigma_{Z_0}). \tag{C.11}$$

Together with (C.8) we get that

$$P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A} \ ; \ \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A} \ ; \ \sigma_W, \sigma_{Z_1}, \sigma_{Z_0}) = P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A} \ ; \ \sigma_W, \sigma_{Z_1}, \sigma'_{Z_0}). \tag{C.12}$$

Because $S \setminus A$ contains Z_1, $\mathrm{pa}^{\sigma_Z}(Z_1)$ and $\mathrm{pa}^{\sigma'_Z}(Z_1)$, we can again make use of Corollary B.4 in order to see that

$$P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A} \ ; \ \sigma_W, \sigma_{Z_1}, \sigma'_{Z_0}) = P(\mathbf{X}_A \mid \mathbf{X}_{S \setminus A} \ ; \ \sigma_W, \sigma'_{Z_1}, \sigma'_{Z_0}) \tag{C.13}$$

outside of P^{σ_Z}- and $P^{\sigma'_Z}$-Nullsets in $\sigma(\mathbf{X}_B)$.

We can then use (C.10) for

$$P(\mathbf{X}_A \mid \mathbf{X}_{S\setminus A} \,;\, \sigma_W, \sigma'_{Z_1}, \sigma'_{Z_0}) = P(\mathbf{X}_A \mid \mathbf{X}_B \,;\, \sigma_W, \sigma'_{Z_1}, \sigma'_{Z_0}) = P(\mathbf{X}_A \mid \mathbf{X}_B \,;\, \sigma_W, \sigma'_Z). \tag{C.14}$$

Piecing together (C.11) - (C.14) finally yields

$$P(\mathbf{X}_A \mid \mathbf{X}_B \,;\, \sigma_W, \sigma_Z) = P(\mathbf{X}_A \mid \mathbf{X}_B \,;\, \sigma_W, \sigma'_Z) \tag{C.15}$$

outside of P^{σ_Z}- and $P^{\sigma'_Z}$-Nullsets in $\sigma(\mathbf{X}_B)$, which is exactly what we wanted to show.

□

D ADDITIONAL DERIVATIONS AND PROOFS

D.1 DERIVATION OF THE IDENTIFIYING EXPRESSIONS IN SECTION 3.2

First we show that $P(X_3 \,;\, \sigma_2) = \int_{\mathcal{X}_1} \left[\int_{\mathcal{X}_2} P(X_3 \mid X_1, X_2) P^*(\mathrm{d}x_2) \right] P(\mathrm{d}x_1)$.

This can be derived by

$$\begin{aligned}
P(X_3 \,;\, \sigma_2) &= \int_{\mathcal{X}_1} P(X_3, \mathrm{d}x_1 \,;\, \sigma_2) \\
&= \int_{\mathcal{X}_1} P(X_3 \mid X_1 \,;\, \sigma_2) P(\mathrm{d}x_1 \,;\, \sigma_2) \\
&= \int_{\mathcal{X}_1} \Bigg[\int_{\mathcal{X}_2} P(X_3 \mid X_1, X_2 \,;\, \sigma_2) \\
&\qquad\qquad \times P(\mathrm{d}x_2 \mid X_1 \,;\, \sigma_2) \Bigg] P(\mathrm{d}x_1) \\
&= \int_{\mathcal{X}_1} \left[\int_{\mathcal{X}_2} P(X_3 \mid X_1, X_2) P^*(\mathrm{d}x_2) \right] P(\mathrm{d}x_1),
\end{aligned}$$

where we used

Rule 3 to conclude that that
$P(X_1 \,;\, \sigma_2)$
$= P(X_1)$,
Rule 2 to conclude that
$P(X_3 \mid X_1, X_2 \,;\, \sigma_2)$
$= P(X_3 \mid X_1, X_2)$,
and Rule 1, combined with the defintion of P^*, to conclude that
$P(X_2 \mid X_1 \,;\, \sigma_2) = P(X_2 \,;\, \sigma_2)$
$= P^*(X_2)$.

The derivation for $P(X_3 \,;\, \sigma_2 = P^*(X_2 \mid X_1))$ is completely analogous, except that in the third step we do not need Rule 1 and can directly use the definition of $P^*(X_2 \mid X_1)$.

D.2 PROOF OF THEOREM 3.6

Let $\mathcal{G}$ be the DAG from Example 3.5. We have already established that the causal effect $P(X_4 \mid X_2, X_3 \; ; \; \sigma_1 = P^*(X_1))$ is identifiable in any MLBN on $\mathcal{G}$ for $P^*(X_1) = \text{Exp}(\lambda), \; \lambda \geq 0$.

To prove that this effect is unidentifiable in general SCM, we construct two explicit examples of SCM on $\mathcal{G} = (O \dot{\cup} U^{\sigma_1}, E), \; O = \{1, 2, 3, 4\}, \; U^{\sigma_1} = \{\epsilon^*, l_{1,2}, l_{2,3}, l_{3,4}, \varepsilon_4\}$, that are Markov with respect to $d-$separation and generate the same joint distribution $P(X_1, X_2, X_3, X_4)$, but differ on $P(X_4 \mid X_2, X_3 \; ; \; \sigma_1 = P^*(X_1))$.

Let $l_{1,2}, l_{2,3}, l_{3,4}, \varepsilon_4$ be independent Bernoulli random variables with $P(l_{i,j} = 1) = P(\varepsilon_4 = 1) = \frac{1}{2}$. Let ϵ^* be independent of $(l_{1,2}, l_{2,3}, l_{3,4}, \varepsilon_4)$. All X_i are binary under the natural mechanisms, and we use the operations $\oplus$ (XOR), $\vee$ (OR) and $\wedge$ (AND).

Fix the map $\phi : \mathbb{R} \to \{0, 1\}$ given by

$$\phi(x) := \mathbb{1}\{x > 1/2\}.$$

Under the natural mechanisms we have $X_1 \in \{0, 1\}$, hence $\phi(X_1) = X_1$.

We define two structural causal models $\mathcal{M}_1$ and $\mathcal{M}_2$ on this DAG that agree on the structural equations for X_1, X_3 and X_4 but differ on X_2.

Both models share natural structural equations

$$\begin{aligned} X_1 &= l_{1,2}, \\ X_4 &= l_{3,4} \oplus \varepsilon_4, \\ X_3 &= (X_4 \oplus l_{2,3}) \vee l_{3,4}. \end{aligned}$$

Under model $\mathcal{M}_1$

$$X_2 \;=\; (\phi(X_1) \oplus l_{2,3}) \oplus (l_{1,2} \wedge \phi(X_1)),$$

whereas under model model $\mathcal{M}_2$

$$X_2 \;=\; (l_{1,2} \oplus l_{2,3}) \oplus (l_{1,2} \wedge \phi(X_1)).$$

Since under the natural mechanisms $X_1 = l_{1,2} \in \{0, 1\}$ almost surely, we have $\phi(X_1) = X_1 = l_{1,2}$ almost surely. Consequently, in both models we obtain

$$X_2 = (l_{1,2} \oplus l_{2,3}) \oplus (l_{1,2} \wedge l_{1,2}) = l_{2,3} \qquad \text{almost surely},$$

and thus the induced joint distribution $P(X_1, X_2, X_3, X_4)$ is the same under $\mathcal{M}_1$ and $\mathcal{M}_2$.

Now consider the soft intervention σ_1 that replaces the natural mechanism for X_1 by $P^*(X_1) = P(\epsilon^*)$, where ϵ^* is a newly introduced latent variable. Concretely, under σ_1 we set

$$X_1 := \epsilon^*, \qquad \epsilon^* \sim \text{Exp}(\lambda).$$

Fix, for example, $\lambda := 2 \log(5/4)$, so that

$$P^{\sigma_1}(X_1 > 1/2) = e^{-\lambda/2} = \tfrac{4}{5}.$$

Under both $\mathcal{M}_1$ and $\mathcal{M}_2$, the event $\{(X_2, X_3) = (0, 1)\}$ has strictly positive probability under the natural mechanisms and also under σ_1. But under σ_1 we can compute that

$$\begin{aligned} P^{\mathcal{M}_1}\big(X_4 = 1 \mid X_2 = 0, X_3 = 1; \sigma_1\big) &= \frac{8}{15}, \\ P^{\mathcal{M}_2}\big(X_4 = 1 \mid X_2 = 0, X_3 = 1; \sigma_1\big) &= \frac{19}{30}. \end{aligned}$$

Hence, the causal effect $P(X_4 \mid X_2, X_3; \sigma_1)$ is generally not identifiable in SCMs that fulfill the Markov property with respect to d-separation.

E LATENT PROJECTION

All of our results and examples hold in the full DAG $\mathcal{G} = (O \dot\cup U, E)$ underlying the SCM $\mathcal{M}$, which we refer to as the *true DAG*. In the literature on causal inference, one often chooses instead to work in a projection of $\mathcal{G}$ onto the observed nodes O, with bidirected edges representing the presence of latent confounding. The result is an *acyclic directed mixed graph* (ADMG) which is sometimes referred to as a *causal diagram* (Definition 2 in Tian and Pearl [2003]). We opt against defining ADMGs for the max-linear setting. However, for completeness we provide a method for projecting the full DAG onto a simplified DAG G which we call the *latent projection* of $\mathcal{G}$ and show in Theorem E.5 that G and $\mathcal{G}$ encode the same $*-$separation statements on the observed part. As our notion of the latent projection is precisely the *canonical expansion* (see Definition E.4) of the ADMG corresponding to $\mathcal{G}$, Theorem E.5 allows for the reframing of all the results in this paper in the terms of causal diagrams, provided one defines the $*-$separation analogue of $m-$*separation* (see [Richardson and Spirtes, 2002]) in the intuitive way: i.e., that one considers two sets to be $*$-separated in an ADMG if and only if they are $*$-separated in their canonical expansion DAG. Ultimately, an identifying expression involving only the observable part of the system can be derived with respect to the projected DAG G if and only if it can be derived with respect to the true graph $\mathcal{G}$, therefore we stress that our approach is equivalent to one which uses causal diagrams.

Definition E.1 (Latent Projection)**.** Let $\mathcal{G} = (O \dot\cup U, E)$ be a DAG with observed node set O and hidden node set U. The *latent projection* of $\mathcal{G}$ is the graph $G = (O \dot\cup U', E')$ on the same observed node set O, and with hidden node and edge sets U' and E' determined by applying the following procedure:

Start with $U' = \emptyset$ and $E' = \emptyset$. For $a, b \in O$ with $a \neq b$, do the following:

(P1) Add $a \to b$ to E' $\iff$ there exists a directed path π from a to b such that every intermediate node in π lies in U.

(P2) Add $u_{a,b}$ to U' and $u_{a,b} \to a$, $u_{a,b} \to b$ to E' $\iff$ there exists a trek between a and b such that all intermediate nodes in the trek lie in U.

Remark E.2. The latent projection G of $\mathcal{G}$ has the following properties:

(i) Due to (P1) , every edge between observed nodes in $\mathcal{G}$ is an edge in G.

(ii) Any two $a, b \in O$ have at most one common parent in U'

(iii) Each node in U' has no parents and exactly two observed children.

(iv) G and $\mathcal{G}$ have the same ancestral relations on the observed nodes. That is, $\mathrm{an}_G(o) = \mathrm{an}_{\mathcal{G}}(o)$ for all $o \in O$.

Remark E.3. The projected DAG G resulting from the projection provided above is exactly the *canonical expansion* of the causal diagram resulting from the projection defined in [Tian and Pearl, 2003].

Next we provide a definition of the canonical expansion.

Definition E.4 (Canonical Expansion)**.** Let $\tilde{G} = (O, \tilde{E})$ be an ADMG (a causal diagram).

Start with $U' = \emptyset,\ E' = \tilde{E}$.

For each bidirected edge $e_{a,b}$ in E', add exactly one parentless vertex $u_{a,b}$ to U' and add directed edges $u_{a,b} \to a$, $u_{a,b} \to b$ to E'. Remove the bidirected edge $e_{a,b}$ from E'

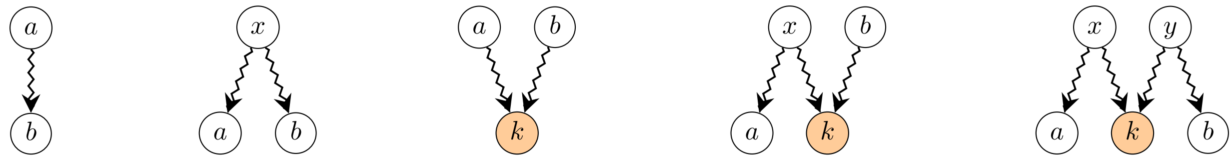
We call the resulting DAG $G = (O \dot\cup U', E')$ the *canonical expansion* of $\tilde{G}$.

Theorem E.5 (Latent projection preserves $\perp_*$)**.** *Let $\mathcal{G} = (O \dot\cup U, E)$ be a DAG and $G = (O \dot\cup U', E')$ its latent projection. For disjoint subsets $A, B, C \subset O$ the following holds:*

$$(A \perp_* B \mid C)_{\mathcal{G}} \iff (A \perp_* B \mid C)_G.$$

Proof. We show both implications by contraposition, starting with $(A \not\perp_* B \mid C)_G \implies (A \not\perp_* B \mid C)_{\mathcal{G}}$.
Let π be a $*-$connecting path between $a \in A$ and $b \in B$ given C in G. From the definition of $*$-connectedness, π must be one of the following five forms, where each zig-zag arrow $\rightsquigarrow$ denotes a directed path not blocked by C, and the highlighted nodes are in $\mathrm{An}_G(C)$.

We now argue by case distinction that any π in G of one of the forms (I)-(V) must give rise to a $*-$connecting path in $\mathcal{G}$.



(I) (II) (III) (IV) (V)

Figure 6: The types of $*$-connecting paths between a and b in a DAG, given C. The highlighted nodes must lie in $\mathrm{An}(C)$, whereas the non-highlighted nodes must not lie in C.

π **is of type (I):** First note that by Remark E.2(iii), π contains no intermediate node in U'. Thus, each edge of π is of the form $o_1 \to o_2$ for some $o_1, o_2 \in O$, which by corresponds to a directed path from o_1 to o_2 in $\mathcal{G}$. Concatenating these paths in $\mathcal{G}$ creates a directed path from a to b in $\mathcal{G}$.

π **is of type (II):** We perform a further case distinction:

$x \in O$**:** By (I), the paths $x \rightsquigarrow a$ and $x \rightsquigarrow b$ give rise to directed paths from x to a and b respectively in $\mathcal{G}$. Concatenating the two paths gives a path of type (II) in $\mathcal{G}$.

$x \in U'$**:** By point Remark E.2(iii), x is the parent of exactly two observed nodes $o_1, o_2 \in O$, which must lie along π (assume w.l.o.g. that π is of the form $\pi : a \leftarrow \dots o_1 \leftarrow x \to o_2 \cdots \to b$). The subpaths $a \leftsquigarrow o_1$ and $o_2 \rightsquigarrow b$ give rise to $*-$connecting paths in $\mathcal{G}$ of type (I), whereas by (P2), the triple $o_1 \leftarrow x \to o_2$ in G corresponds to a trek $o_1 \leftsquigarrow\rightsquigarrow o_2$ in $\mathcal{G}$ with intermediate nodes fully contained in U (and in particular, not blocked by $C \subset O$). Concatenating these three paths in $\mathcal{G}$ gives rise to a path between a and b of type (II).

π **is of type (III)-(V):** We consider in detail only the case that π is of type (IV); the arguments for types (III) and (V) are completely analogous. Again, via a case distinction:

$k \in O$**:** Then the arguments for (I) and (II) applied to the portions $a \leftsquigarrow x \rightsquigarrow k$ and $b \leftsquigarrow k$ of π give rise to $*-$connecting paths in $\mathcal{G}$, and as $k \in \mathrm{An}_{\mathcal{G}}(C)$ holds due to Remark E.2(iv), their concatenation is $*-$connecting of type (IV) in $\mathcal{G}$.

$k \in U'$**:** This case cannot occur, as it would imply that a node in U' has a parent, contradicting Remark E.2(iii).

This proves the first implication of the equivalence.
The converse $(A \not\perp_* B \mid C)_{\mathcal{G}} \implies (A \not\perp_* B \mid C)_G$ is similarly shown via case distinction. Let π now be a $*-$connecting path between $a \in A$ and $b \in B$ in $\mathcal{G}$ given C. Again, π is of one of the five general forms depicted in Figure 6. We show that a path of any form corresponds to a $*-$connecting path in the projection G.

π **is of type (I):** Any π of type (I) decomposes into a concatenation of directed subpaths $x \rightsquigarrow y$ in which every intermediate node is contained in U. By (P1) such a path gives rise to an edge $x \to y$ in G. Concatenating these edges gives a path of type (I) from a to b in G, which is $*-$connecting given C as π contains no nodes in C.

π **is of type (II):** We perform a case distinction for x.

$x \in O$**:** The argument from (I) applies to the subpaths $a \leftsquigarrow x$ and $x \rightsquigarrow b$, giving a path of type (II) in G.

$x \in U$**:** There exists a subpath $\pi' : o_1 \leftsquigarrow x \rightsquigarrow o_2$ of π such that $o_1, o_2 \in O$ are the only observable nodes in π' (Intuitively, o_1 and o_2 are the "first" observable nodes along the directed paths $x \rightsquigarrow a$ and $x \rightsquigarrow b$ respectively. In the projection G, π' corresponds to a trek on three nodes $o_1 \leftarrow u' \to o_2$ for some $u' \in U'$. Concatenating this trek with the paths of type (I) induced by $o_1 \rightsquigarrow a$ and $o_2 \rightsquigarrow b$ gives a path of type (II) in G.

π **is of type (III):** We perform a case distinction for the collider k

$k \in O$**:** Then $k \in \mathrm{An}_{\mathcal{G}}(C)$ implies $k \in \mathrm{An}_G(C)$ by Remark E.2(iv). Thus, the concatenations of the paths in G corresponding to the portions $a \rightsquigarrow k$ and $b \rightsquigarrow k$ of π give rise to a $*-$connecting path of type (III).

$k \in U$**:** Then, as $k \in \mathrm{An}_{\mathcal{G}}(C)$, there exists a directed path $k \rightsquigarrow c$ for some $c \in C$. Let $\tilde{o} \in O$ be the first observed node along this path (at least one exists, as $c \in O$). Further let o_1 and o_2 be the last observed nodes along the paths $a \rightsquigarrow k$ and $b \rightsquigarrow k$ respectively. By construction, the paths $o_1 \rightsquigarrow \tilde{o}$ and $o_2 \rightsquigarrow \tilde{o}$ contain no intermediate observed nodes. Thus, by (P1), G contains the edges $o_1 \to \tilde{o}$ and $o_2 \to \tilde{o}$. Concatenating the directed paths in G arising from the subpaths $a \rightsquigarrow o_1$ and $b \rightsquigarrow o_2$ with the triple $o_1 \to \tilde{o} \leftarrow o_2$ gives a $*-$connecting path of type (III) in G as $o \in \mathrm{An}_G(C)$.

π **is of type (IV)-(V):** By a combination of the arguments used in the previous three cases, π gives rise to a path of the same type in G.

$\square$

F DEFINITIONS AND RESULTS ON IDENTIFIABILITY USED THROUGHOUT THIS WORK

Definition F.1 ($\mathcal{C}$-component)**.** Let G be a causal diagram on vertex set $V(G)$. If for every pair $v, w \in V(G)$ there exists a path from v to w consisting only of bidirected arcs, then G is called a *$\mathcal{C}$-component* (confounded component).

Definition F.2 ($\mathcal{C}$-tree)**.** Let G be a causal diagram and let $Y \in V(G)$. Assume that (i) G is a $\mathcal{C}$-component, (ii) every observable node has at most one child in G, and (iii) $\mathrm{An}(Y)_G = V(G)$. Then G is called a *Y-rooted $\mathcal{C}$-tree* (confounded tree).

Theorem F.3 (Shpitser and Pearl [2006a], Theorem 3)**.** *Let G be a Y-rooted $\mathcal{C}$-tree. Then the effect of any set of nodes X in G on Y is not identifiable if $Y \notin X$.*