On Positivity Condition for Causal Inference

Inwoo Hwang [*1] **Yesong Choe** [*2] **Yeahoon Kwon** [2] **Sanghack Lee** [1 2]

Abstract

Identifying and estimating a causal effect is a fundamental task when researchers want to infer a causal effect using an observational study without experiments. A conventional assumption is the strict positivity of the given distribution, or so called positivity (or overlap) under the unconfounded assumption that the probabilities of treatments are positive. However, there exist many environments where neither observational data exhibits strict positivity nor unconfounded assumption holds. Against this background, we examine the graphical counterpart of the conventional positivity condition so as to license the use of identification formula without strict positivity. In particular, we explore various approaches, including analysis in a post-hoc manner, do-calculus, Q-decomposition, and algorithmic, to yielding a positivity condition for an identification formula, where we relate them, providing a comprehensive view. We further discuss the design of a positivity-aware identification algorithm based on the theoretical characterization of identification formulas.

1. Introduction

Knowing only associations among variables is insufficient to establish the direction of causation, thereby limiting its utility in making informed decisions. On the other hand, understanding cause-effect relationships can help elucidate the effect of intervention so that precise control of the underlying environment can lead to the desired outcome. Causal effect identification is the task of answering whether eliciting a causal quantity under given information (typically a set of causal assumptions encoded as a causal diagram) is probable and, if so, providing a mapping from an observational distribution to the causal quantity.

[*]Equal contribution [1]Artificial Intelligence Institute, Seoul National University, Seoul, South Korea [2]Graduate School of Data Science, Seoul National University, Seoul, South Korea. Correspondence to: Sanghack Lee <sanghack@snu.ac.kr>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

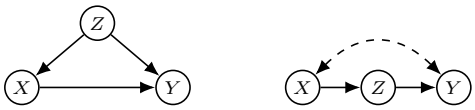


Figure 1: Causal diagrams where $P_x(y)$ is identified via a (a) backdoor and (b) front-door criterion, respectively.

Formally, the causal effect of a set of treatment variables $\mathbf{X}$ on a disjoint set of outcome variables $\mathbf{Y}$ is said to be identifiable from a causal graph $\mathcal{G}$ if the quantity $P_{\mathbf{x}}(\mathbf{y}) = P(\mathbf{y} \mid do(\mathbf{x}))$ can be uniquely computed from any positive distribution over the observed variables (Tian & Pearl, 2002; Shpitser & Pearl, 2006a). If an effect is non-identifiable, there exist two models eliciting the same causal structure and agreeing on the provided observational distribution while disagreeing on the effect of the intervention of interest.

In the absence of unmeasured confounders, all causal effects are trivially identifiable (Robins, 1986; Pearl, 1993; Spirtes et al., 1993; Galles & Pearl, 1998). However, if there are unmeasured confounders (i.e., a semi-Markovian model), the problem of causal effect identifiability is getting sophisticated. Tian & Pearl (2003); Huang & Valtorta (2006); Shpitser & Pearl (2006a) developed *complete* identification algorithms taking an arbitrary causal graph. When not identifiable, one might rely on partial identification, which provides bounds for the causal effect, or alike (Tian & Pearl, 2000; Balke, 1995; Zhang et al., 2022; Li & Pearl, 2022).

Regardless of the existence of complete graphical identification algorithms, one simple, widely adopted identification condition is the *adjustment criterion* (i.e., backdoor criterion (Pearl, 1993); g-computation (Robins, 1986)). It yields the following form of the formula for $P_x(y)$:

$$P_x(y) = \textstyle\sum_z P(y \mid x, \mathbf{z}) P(\mathbf{z}), \tag{1}$$

where $\mathbf{Z}$ is an admissible set. Here, *positivity* $P(x \mid \mathbf{z}) > 0$ is assumed to license the use of the above formula for the causal effect. In the context of estimating average treatment effect $\mathbb{E}[Y \mid do(X=1)] - \mathbb{E}[Y \mid do(X=0)]$, Hernán & Robins (2006) stated that, for each value of the covariate in the population, there are some subjects that received the treatment—i.e., $P(X \mid \mathbf{z}) > 0$ for all $\mathbf{z}$ with $P(\mathbf{z}) \neq 0$. With all individuals receiving the same treatment, it would

be impossible to estimate the causal effect from observed data (Hernán & Robins, 2006; 2020).

The same positivity condition *might* be derived directly from the given formula, although its validity is less clear at this moment. Given the summation in (1), if there exists $\mathbf{z}$ such that $P(y \mid x, \mathbf{z})P(\mathbf{z})$ is undefined, then the estimate will be undefined. Although $P(y \mid x, \mathbf{z})$ is *undefined* if $P(x, \mathbf{z}) = 0$, let us ignore the case when $P(\mathbf{z}) = 0$, for the sake of explanation. Therefore, we are concerned about $P(x, \mathbf{z}) > 0$ only when $P(\mathbf{z}) > 0$. That is, we want to ensure $P(x \mid \mathbf{z}) > 0$ whenever $P(\mathbf{z}) \neq 0$. This leads to the original positivity condition for the adjustment:

$$\forall \mathbf{z}(P(\mathbf{z}) = 0 \vee P(x \mid \mathbf{z}) > 0), \tag{2}$$

which we will denote by $\mathrm{adj}(x; \mathbf{Z})$.

There is limited literature on the topic of positivity conditions in causal identification. Shpitser & Pearl (2006a) mentioned that most literature only considers $P(\mathbf{V})$ as a positive distribution, where $\mathbf{V}$ is a set of endogenous variables. However, they postulated that $P(\mathbf{A})$ must be positive for every distribution $P(\mathbf{B} \mid \mathbf{A})$ involved during the derivation of a formula. Other than this, positivity assumptions are mentioned only for specific identification tasks. Petersen et al. (2012) examined a positivity condition for assessing model and parametric-specific identifiability of causal effects within a marginal structural working model, where the result was identical to the condition for the adjustment criterion. Dahabreh et al. (2019) investigated the positivity conditions to generalize causal inferences under single world intervention graphs of a randomized trial in the target population of trial-eligible individuals.

Nonetheless, previous studies have not yet explicated positivity conditions when an identifiable causal formula exhibits an intricate form, such as fractions, given a general, arbitrary causal structure. Against this background, this paper seeks a principled way of eliciting a positivity condition for identifying a causal effect given a causal graph.

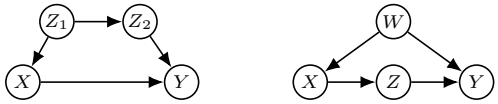
1.1. Motivating examples

We motivate our work by illustrating an example in which the backdoor adjustment is inapplicable. We then describe how different identification formulas involve different positivity conditions, illustrating practical implications.

Positivity on a front-door criterion Another well-known identification criterion is front-door adjustment (Pearl, 1995), exemplified by the causal diagram in Fig. 1(b). Its identification formula for $P_x(y)$ is as follows:

$$P_x(y) = \textstyle\sum_z P(z \mid x) \sum_{x'} P(y \mid x', z)P(x'),$$

which can be derived by applying rules of do-calculus and probability operations. Without assuming general positivity



(a) (b)

Figure 2: Causal diagrams where $P_x(y)$ is identified via (a) backdoor adjustment either using $\{Z_1\}$, $\{Z_2\}$, or both (b) both backdoor and front-door criteria.

$P(\mathbf{V}) > 0$, one must ensure which positivity conditions are sufficient for each application and combine those conditions to derive a condition licensing the use of given $P(\mathbf{V})$ for $P_x(y)$. It turns out that a sufficient condition is

$$\big(P(x) > 0 \vee \forall z(P_z(y) = 0)\big) \wedge \big(\forall z(P(z \mid x) = 0 \vee \mathrm{adj}(z; X))\big), \tag{3}$$

where the condition over an interventional distribution $P_z(y) = 0$ is further expanded to $\forall x'(P(y|x', z) = 0 \vee P(x') = 0)$ (see Appendix C.2 for the derivation). Furthermore, the condition derived directly from the resulting formula that avoids its undefinedness is indeed equivalent to the above condition (3). In this paper, we discuss the relationships between the two conditions derived differently.

We note that the term 'positivity' might be a misnomer since the detailed condition (3) for front-door criterion is described not only by inequality > 0 but also by equality, logical connectives, and quantifiers (see, for comparison, $P(x \mid z) > 0$ for adjustment or $P(\mathbf{V}) > 0$ for a graphical criterion). Yet, we will stick to the conventional term 'positivity' for the condition of an observational distribution under which an identification formula is validly evaluated.

Multiplicity of identification formulas and conditions An identification algorithm (e.g., ID (Shpitser & Pearl, 2006a) or IDENTIFY (Tian & Pearl, 2003)) typically returns a single identification formula for a causal query if it is identifiable from a causal graph. However, there may exist multiple formulas (Tikka & Karvanen, 2017; 2018), which may be obtained by applying the identification algorithm over different latent projections onto $\mathbf{V}' \subset \mathbf{V}$ (Pearl & Verma, 1991) (i.e., a causal graph over a subset of variables while maintaining causal relationships) of a full causal graph over $\mathbf{V}$. Here, we show by examples that different formulas imply different positivity conditions.

Consider Fig. 2(a) where there exist three different backdoor admissible sets $\{Z_1\}$, $\{Z_2\}$, and $\{Z_1, Z_2\}$. In this case, it is possible for $\mathrm{adj}(x; Z_2)$ holds true but not $\mathrm{adj}(x; Z_1)$ (see Appendix D.1). This implies that one can estimate the causal effect with a formula but not with the other, which was not the case under strict positivity. Now consider an example in Fig. 2(b) where both backdoor and front-door criteria can be applied to identify $P_x(y)$. Similarly, it might

be the case that the backdoor is not applicable, but the front-door criterion is, and vice versa (see Appendix D.2 for details). These examples manifest the importance of knowing the characteristics of the given data and tailoring the formula accordingly, rather than simply calculating the formula assuming $P(\mathbf{V}) > 0$.

Practical implications Under the violation of strict positivity, the conventional identification formula for a causal quantity may not be properly computed, e.g., due to undefined terms. Furthermore, the main tools for eliciting the formula (i.e., do-calculus and Q-decomposition) are established under the strict positivity assumption. Thus, their validity and mathematical correctness are unclear under relaxed positivity. For practitioners, it is crucial to understand which identification formula is valid for the causal quantity, as the observational distribution is rarely strictly positive in many real-world scenarios.

1.2. Contributions

Our contributions are as follows. (i) We provide a comprehensive view of eliciting a positivity condition over an observational distribution for an identifiable causal query given an arbitrary causal graph. (ii) In particular, we offer positivity conditions for do-calculus and generalized Q-decomposition, which are the main drivers of sound and complete identification algorithms, establishing rigorous foundations for obtaining sufficient positivity for causal effect identification. (iii) We devise an algorithmic approach to eliciting positivity by incorporating a relaxed version of generalized Q-decomposition into an existing identification algorithm. We establish a connection to post-hoc analysis of positivity. Further, this principled approach can be modified to design a positivity-aware identification algorithm.

1.3. Organization

We illustrate the structure of our paper in Fig. 3. In Sec. 3, we first explore, without theoretical guarantee, the possibility of eliciting a positivity condition by directly examining a given identification formula, which is typically the output of an identification algorithm such as IDENTIFY (Tian & Pearl, 2003) or ID (Shpitser & Pearl, 2006a).

In Sec. 4, we formally re-establish the rules of do-calculus (Lem. 4.1) under relaxed positivity, which is the building block of identification algorithms. Consequently, we show that the conjunction of the positivity conditions, invoked during the derivation of an identification formula through sequential applications of do-calculus, corresponds to the positivity condition for the formula (Prop. 4.2 and Corollary E.1 in Appendix).

Next, we provide a more systematic, algorithmic approach to acquiring positivity in Secs. 5 and 6. Specifically, we

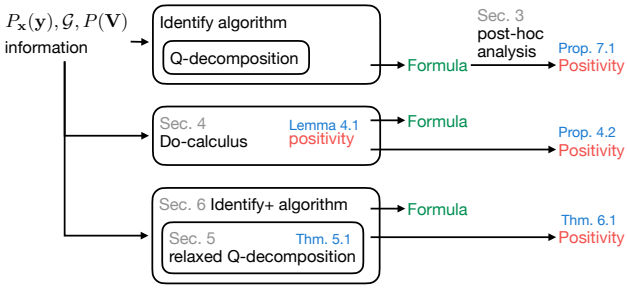


Figure 3: A schematic diagram representing the overall structure of this paper.

first relax generalized Q-decomposition where an additional sufficient condition for identification is provided (Lem. 5.1 and Thm. 5.1). We then develop IDENTIFY+ (Algs. 1 and 2), which incorporates the relaxed decomposition into IDENTIFY (Tian & Pearl, 2003) and an additional component that simultaneously constructs a positivity condition while eliciting an identification formula (Thm. 6.1).

We note that one may skip the post-hoc analysis (Sec. 3), read through to the algorithmic approach, and then come back to the post-hoc analysis. This sequence would be beneficial to some readers who want to avoid some unresolved uncertainties regarding the validity of post-hoc analysis. All the omitted proofs are deferred to Appendix E.

2. Preliminaries

Each variable is represented with a capital letter, e.g., X, and its realization with the corresponding lowercase x. We use bold letters, e.g., $\mathbf{X}$, to denote sets of variables (or values). We follow the conventional definition of probability based on a set of axioms by Kolmogorov (1933). Throughout this paper, we focus only on the *discrete* case. We discuss possible challenges to the continuous case in Appendix F.

Causal framework We adopt a structural causal model (SCM (Pearl, 2009)) as our causal framework.

Definition 2.1 (SCM (Pearl, 2009))**.** A structural causal model (SCM) $\mathcal{M}$ is a 4-tuple $\langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$, where $\mathbf{V} = \{V_1, \ldots, V_n\}$ is a set of endogenous variables; $\mathbf{U}$ is a set of exogenous variables; $\mathbf{F} = \{f_1, \ldots, f_n\}$ is a set of functions determining $\mathbf{V}$: $v_i \leftarrow f_i(\mathbf{pa}_i, \mathbf{u}_i)$ where $\mathbf{Pa}_i \subseteq \mathbf{V} \backslash \{V_i\}$ and $\mathbf{U}_i \subseteq \mathbf{U}$; $P(\mathbf{U})$ is a distribution over $\mathbf{U}$. ■

Each SCM $\mathcal{M}$ induces a distribution $P(\mathbf{V})$ and a causal graph $\mathcal{G}$, where directed edges encode functional relationships between observed variables, and bidirected edges encode unobserved confounders. Conducting an intervention is represented through the *do*-operator, $do(\mathbf{X} = \mathbf{x})$, which corresponds to replacing the original equations of $\mathbf{X}$ with constant $\mathbf{x}$. Such an intervention induces a submodel $\mathcal{M}_{\mathbf{x}}$

and an interventional distribution $P_{\mathbf{x}}(\mathbf{V})$. We adopt graphical kinship terminology $Pa(\cdot)$, $Ch(\cdot)$, $An(\cdot)$, and $De(\cdot)$ to represent its parents, children, ancestors, and descendants in $\mathcal{G}$ where ancestors and descendants include its argument. For a set of arguments, $Pa(\mathbf{X}) = \bigcup_{X \in \mathbf{X}} Pa(X)$, and others are defined similarly. We denote by $\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ the manipulated graph of $\mathcal{G}$ by deleting the edges onto $\mathbf{X}$ and edges outgoing from $\mathbf{Z}$. We denote by $\perp\!\!\!\perp_p$ for conditional independence and $\perp\!\!\!\perp_d$ for d-separation. We may use $\perp\!\!\!\perp$ for $\perp\!\!\!\perp_d$ if no ambiguity arises. Given $\mathbf{W}$ a subset of vertices in $\mathcal{G}$, we use $\mathcal{G}[\mathbf{W}]$ to denote a subgraph induced by $\mathbf{W} \subseteq \mathbf{V}$. See Appendix A for formal definitions of d-separation and latent projection.

Causal identification problem The causal effect of an action $do(\mathbf{x})$ on a set of variables $\mathbf{Y}$ such that $\mathbf{Y} \cap \mathbf{X} = \emptyset$ is said to be identifiable from P in $\mathcal{G}$ if $P_{\mathbf{x}}(\mathbf{y})$ is uniquely computable from $P(\mathbf{V})$ in any causal model which induces $\mathcal{G}$. That is, if $P_{\mathbf{x}}^{\mathcal{M}_1}(\mathbf{y}) = P_{\mathbf{x}}^{\mathcal{M}_2}(\mathbf{y})$ for every pair of causal models $\mathcal{M}_1$ and $\mathcal{M}_2$ with $P^{\mathcal{M}_1}(\mathbf{V}) = P^{\mathcal{M}_2}(\mathbf{V}) > 0$ and $\mathcal{G}(\mathcal{M}_1) = \mathcal{G}(\mathcal{M}_2) = \mathcal{G}$, then $P_{\mathbf{x}}(\mathbf{y})$ is identifiable in $\mathcal{G}$. On the other hand, if one can construct two models such that $P_{\mathbf{x}}^{\mathcal{M}_1}(\mathbf{y}) \neq P_{\mathbf{x}}^{\mathcal{M}_2}(\mathbf{y})$, then $P_{\mathbf{x}}(\mathbf{y})$ is not identifiable in $\mathcal{G}$.

However, simply sticking to strict positivity for the definition of identifiability will lose the opportunity to utilize an available observational distribution for identifying causal effects since one can still obtain the effect once the distribution satisfies a sufficient positivity condition, e.g., Eq. (2). Given that our objective is *not* devising a complete graphical criterion for identification, we will be negligent to the strict positivity from the definition and focus on the positivity condition under which graphical causal effect identification is sound. Note that the causal effect $P_{\mathbf{x}}(\mathbf{y})$ itself is *always* well-defined regardless of the strict positivity assumption (see Appendix A).

3. Informal Examination of Post-hoc Approach to Positivity

In this section, we informally[1] examine a positivity condition under which the identification formula *itself* is well-defined[2]. The complete algorithms (Tian & Pearl, 2003; Shpitser & Pearl, 2006a) return formulas which are composed of fractions, summation (i.e., marginalization), or multiplication of the given joint probability or expressions thereof. The basic idea to derive whether the formula is well-defined consists of the following: (a) the addition of any undefined term yields an undefined value; (b) the multiplication of an undefined term yields an undefined value unless zero is multiplied to; and (c) the zero denominator of a fraction makes the fraction undefined (conditional probability is its special case). We will later illustrate the sufficiency of these rules to determine positivity conditions by connecting an identification algorithm and the pattern of the formula.

[1]We wanted this section to be inspirational to later, more principled sections. Further, a formal treatment to post-hoc analysis requires making sense of $0/0$ in formulas, investigated in Thm. 5.1

[2]Here, we say a formula is *well-defined* if it is computable. It should not be conflated with the well-definedness of mathematical terms. A formula can still be computable even if it contains undefined terms. For example, strictly speaking, $\sum_{\{z|P(x|z)>0\}} P(y \mid x, z)P(z)$ if $\forall z(P(z) = 0 \vee P(x \mid z) > 0)$. We will henceforth denote it as $\sum_z P(y \mid x, z)P(z) \geq 0 \Leftarrow \text{adj}(x; Z)$, and say the formula $\sum_z P(y \mid x, z)P(z)$ is *well-defined* under the positivity condition $\text{adj}(x; Z)$.

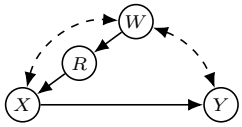


Figure 4: A causal diagram called Napkin.

We introduce the concept of well-definedness of a formula through a causal diagram called Napkin (Fig. 4) (Pearl & Mackenzie, 2018) where its formula is as follows, $P_x(y) = \frac{\sum_w P(y,x|r,w)P(w)}{\sum_w P(x|r,w)P(w)}$. Based on the formula, since the causal quantity is a fraction, we consider the (1) numerator and (2) denominator separately. If the denominator is zero, the formula is undefined. We derive the positivity condition as follows:

$$\exists r \left(\frac{\sum_w P(y,x|r,w)P(w)}{\sum_w P(x|r,w)P(w)} \geq 0 \right) \Leftarrow \exists r(① \geq 0 \wedge ② > 0),$$

where ① and ② is a numerator and denominator, respectively. Each condition can be expressed as

$$\begin{aligned} ① \geq 0 &\Leftarrow \text{adj}(r; W), \\ ② > 0 &\Leftarrow \text{adj}(r; W) \wedge \exists w(P(x \mid r, w)P(w) > 0) \\ &\Leftarrow \text{adj}(r; W) \wedge \exists w(P(x \mid r, w) > 0 \wedge P(w) > 0) \\ &\Leftarrow \text{adj}(r; W) \wedge \exists w(P(x, r, w) > 0) \\ &\Leftarrow \text{adj}(r; W) \wedge P(x, r) > 0. \end{aligned}$$

Hence, the sufficient condition for the well-definedness of the formula is $\exists r\big(\text{adj}(r; W) \wedge P(x, r) > 0\big)$.

While it is true that the positivity condition derived directly from a formula ensures that the formula is well-defined, yet its validity is unclear for now since the formula is derived under strict positivity, and there *might* be some conditions that cannot be read off from the formula. We provide a formal treatment to the validity of post-hoc analysis in Sec. 7.1.

4. Positivity for Do-calculus

We now consider developing a general and principled approach for deriving a positivity condition by examining the conditions for the application of *do-calculus* (Pearl, 1995), which is a building block for identification algorithms and

shown to be complete for the task of causal effect identification (Shpitser & Pearl, 2006a). Do-calculus, capable of reducing an expression for $P_{\mathbf{x}}(\mathbf{y})$ to a subscript-free expression, consists of the following three rules:

Rule 1 (addition/deletion of observation):
$P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}}}$

Rule 2 (exchange of action and observation):
$P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$

Rule 3 (addition/deletion of action):
$P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ if $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}},\overline{\mathbf{Z}(\mathbf{W})}}}$, where $\mathbf{Z}(\mathbf{W}) = \mathbf{Z} \setminus An(\mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}}}$.

Pearl (1995) implicitly stated that a causal effect is identifiable from any *positive* distribution of the observed variables (i.e., $P(\mathbf{v}) > 0$) in a model characterized by a graph $\mathcal{G}$ if there exists a finite sequence of transformations, each conforming to one of the three rules. We can relax the positivity conditions for do-calculus taking advantage of the results from Lauritzen et al. (1990); Verma & Pearl (1990).

Proposition 4.1 (Prop. 2 in (Lauritzen et al., 1990))**.** *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are d-separated by* $\mathbf{S}$*, then* $\mathbf{X} \perp\!\!\!\perp_p \mathbf{Y} \mid \mathbf{S}$. ■

While the result looks familiar, it is *without* strict positivity: whenever d-separation ($\mathbf{X} \perp\!\!\!\perp_d \mathbf{Y} \mid \mathbf{S}$) holds true, then $P(\mathbf{x}, \mathbf{y} \mid \mathbf{s}) = P(\mathbf{x} \mid \mathbf{s})P(\mathbf{y} \mid \mathbf{s})$ for $P(\mathbf{s}) > 0$ and, similarly, $P(\mathbf{x} \mid \mathbf{y}, \mathbf{s}) = P(\mathbf{x} \mid \mathbf{s})$ for $P(\mathbf{s}, \mathbf{y}) > 0$ without strict positivity over all the variables (nodes) in the DAG.

We generalize the proposition to the space of observational and interventional distributions and figure out a sufficient positivity condition for each rule. Before we proceed, we slightly rewrite do-calculus replacing $\mathcal{G}_{\overline{\mathbf{X}}}$ with $\mathcal{G} \setminus \mathbf{X}$ and dropping the condition on $\mathbf{X}$ (Lee & Bareinboim, 2020).

Lemma 4.1 (Do-calculus with relaxed positivity)**.** *Let* $\mathcal{G}$ *be the directed acyclic graph (DAG) associated with a causal model, and let* $P(\cdot)$ *be the probability distribution induced by the model. Then,*

Rule 1: $P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ *if* $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G}\setminus\mathbf{X})}$ *and* $P_{\mathbf{x}}(\mathbf{z}, \mathbf{w}) > 0$

Rule 2: $P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w})$ *if* $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G}\setminus\mathbf{X})_{\underline{\mathbf{Z}}}}$ *and* $P_{\mathbf{x}}(\mathbf{z}, \mathbf{w}) > 0$

Rule 3: $P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ *if* $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G}\setminus\mathbf{X})_{\overline{\mathbf{Z}(\mathbf{W})}}}$ *and* $P_{\mathbf{x}}(\mathbf{w}) > 0$. ■

Note that, for example, if we are to claim $P_{\mathbf{x}}(\mathbf{y}, \mathbf{z} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})P_{\mathbf{x}}(\mathbf{z} \mid \mathbf{w})$ as an alternative for Rule 1, the condition can be further relaxed to $P_{\mathbf{x}}(\mathbf{w}) > 0$. This lemma asserts sufficient positivity conditions *not exclusively* for an observational distribution. However, this does not hinder us from inducing a positivity condition over an observational distribution since all the constrained interventional probabilities appeared in intermediate expressions are later expressed using only observational probabilities. In the following, we illustrate examples (i.e., front-door and Napkin) where positivity for do-calculus is applied while taking multiplied-by-zero into consideration.[3] Due to the space constraint, the backdoor example is provided in Appendix C.1.

Example 4.1 (Front-door)**.** We demonstrate the lemma by deriving the formula based on probability axioms and do-calculus (Pearl, 1995) on the front-door graph.

$$\begin{aligned}
P_x(y) &= \textstyle\sum_z P_x(z)P_x(y \mid z) \\
&= \textstyle\sum_z P_x(z)P_{xz}(y) && (4)\\
&= \textstyle\sum_z P_x(z)P_z(y) \\
&= \textstyle\sum_z P(z \mid x)P_z(y) && (5)\\
&= \textstyle\sum_z P(z \mid x) \sum_{x'} P_z(y \mid x')P_z(x') \\
&= \textstyle\sum_z P(z \mid x) \sum_{x'} P_z(y \mid x')P(x') \\
&= \textstyle\sum_z P(z \mid x) \sum_{x'} P(y \mid x', z)P(x'), && (6)
\end{aligned}$$

where (5) and (6), respectively, is granted if

$$P(x){>}0 \vee \forall z(P_z(y){=}0) \text{ and } \forall z(P(z|x){=}0 \vee \mathrm{adj}(z; X)).$$

The conjunction of both conditions grants the identification of $P_x(y)$ without strict positivity. During the derivation, other lines do not require any condition. For example, Eq. (4) does not invoke $P_x(z) > 0$ since $P_x(z) = 0$ makes $P_x(y \mid z)$ ignorable. We present the full condition in terms of the observational distribution in Appendix C.2 together with a post-hoc analysis, which yields the same positivity condition. ■

Example 4.2 (Napkin)**.** We derive the condition for Napkin:

$$\begin{aligned}
P_x(y) &= P_{w,r,x}(y) \\
&= P_{w,r}(y \mid x) && \text{if } P_{w,r}(x) > 0 \\
&= P_{w,r}(y, x)/P_{w,r}(x) && \text{if } P_{w,r}(x) > 0 \\
&= P_r(y, x)/P_r(x) \\
&= \tfrac{\sum_{w'} P(y,x|r,w')P(w')}{\sum_{w'} P(x|r,w')P(w')}. && \text{if } \mathrm{adj}(r; W)
\end{aligned}$$

This derivation requires two positivity conditions. In reverse, there exists r such that (1) $\mathrm{adj}(r; W)$ and; (2) $P_{w,r}(x) > 0$, which is equivalent to $P_r(x) > 0$ by Rule 3. Condition (2) is implied by $\sum_{w'} P(x \mid r, w')P(w') > 0$ under (1). This leads to $\mathrm{adj}(r; W) \wedge \exists w(P(x \mid r, w)P(w) > 0)$. Since $P(x \mid r, w')P(w') > 0 \Leftarrow P(x, r, w) > 0$ and $\exists w(P(x, r, w) > 0) \Leftrightarrow P(x, r) > 0$, we finally obtain

$$\exists r(\mathrm{adj}(r; W) \wedge P(x, r) > 0). \tag{7}$$

This turns out to be equivalent to the condition derived by post-hoc analysis (see Appendix C.3). ■

[3] To make sense of an undefined term multiplied by zero, consider the following derivation $P(y) = \sum_z P(y, z) = \sum_{\{z|P(z)>0\}} P(y, z) = \sum_{\{z|P(z)>0\}} P(y|z)P(z)$. Thus, $P(y) = \sum_z P(y|z)P(z)$ is valid ignoring z such that $P(z) = 0$.

Proposition 4.2. *The conjunction of positivity conditions required in deriving an identification formula for a causal effect applying do-calculus and probability operations sequentially is a sufficient positivity condition for the identification of the causal effect.* ■

5. Generalized Q-decomposition and Positivity

Given the theoretical underpinnings of the sufficient positivity conditions over do-calculus, we now investigate the feasibility of creating an identification algorithm capable of simultaneously taking a non-strictly positive observational distribution into account. IDENTIFY is a causal effect identification algorithm (Tian & Pearl, 2003), which is sound and complete, that identifies a causal effect in a semi-Markovian model. In this section, we characterize the key components of the algorithm, called Q-decomposition (Tian & Pearl, 2003), which decomposes a probability into c-factors, a special type of interventional probability. The c-factors derived from the given observational distribution are used to answer the c-factors derived from the query (Tian & Pearl, 2003; Shpitser & Pearl, 2006a). We review Q-decomposition and establish the correctness of the decomposition without relying on the strict positivity.

We can rewrite a causal effect

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{y}^+ \setminus \mathbf{y}} P_{\mathbf{x}^+}(\mathbf{y}^+), \tag{8}$$

where $\mathbf{Y}^+ = An(\mathbf{Y})_{\mathcal{G}_{\underline{\mathbf{X}}}}$ and $\mathbf{X}^+ = An(\mathbf{X} \cap An(\mathbf{Y}))_{\mathcal{G}_{\underline{\mathbf{Y}}}}$. It is proved that the identifiability of $P_{\mathbf{x}^+}(\mathbf{y}^+)$ is equivalent that of $P_{\mathbf{x}}(\mathbf{y})$ (Shpitser & Pearl, 2006a). Since marginalization and Rule 3 do not invoke positivity, we do not impose any positivity condition over the observation at this stage.

The modified query $P_{\mathbf{x}^+}(\mathbf{y}^+)$ is then decomposed into c-factors typically (and implicitly) under $P(\mathbf{V}) > 0$ based on the c-components (connected components via bidirected edges) over $\mathcal{G}[\mathbf{Y}^+]$. Next, each c-factor is identified individually and combined to ultimately solve $P_{\mathbf{x}}(\mathbf{y})$:

$$P_{\mathbf{x}^+}(\mathbf{y}^+) = \prod_{\mathbf{C} \in \mathrm{cc}(\mathcal{G}[\mathbf{Y}^+])} Q[\mathbf{C}], \tag{9}$$

where $\mathrm{cc}(\cdot)$ is c-components and

$$Q[\mathbf{W}](\mathbf{w}, \mathbf{pa_w}) = \sum_{\mathbf{u}(\mathbf{W})} P(\mathbf{u}(\mathbf{W})) \prod_{V_i \in \mathbf{W}} 1_{v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)}, \tag{10}$$

with $\mathbf{U}(\mathbf{W}) = \bigcup_{V_i \in \mathbf{W}} \mathbf{U}_i$. This c-factor is simply written $Q[\mathbf{W}]$ for readability. C-factor $Q[\mathbf{W}]$ is equivalent to $P_{\mathbf{V} \setminus \mathbf{W}}(\mathbf{W})$, an interventional distribution over $\mathbf{W}$ where all the other variables are intervened (i.e., $\mathbf{W}$'s parents). We show the validity of the decomposition without positivity condition:

Proposition 5.1. *Under structural causal model and without $P(\mathbf{V})$ being positive, the following holds:*

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{y}^+ \setminus \mathbf{y}} \prod_{\mathbf{C} \in \mathrm{cc}(\mathcal{G}[\mathbf{Y}^+])} Q[\mathbf{C}]. \quad ■$$

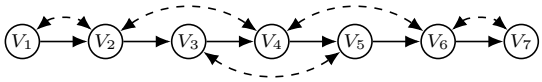


Figure 5: An exemplar causal diagram.

Now we will relate the following characterization of c-factor with the positivity condition. We revisit 'generalized Q-factorization' (Tian & Pearl, 2003, Lemma 4) which consists of two parts, (i) and (ii). We present it as two separate lemmas with assumptions on the positivity *explicit* where the first part (i) can be claimed without positivity.

Lemma 5.1. *Given $\mathbf{H} \subseteq \mathbf{V}$, let $\mathbf{H}_1, \ldots, \mathbf{H}_k$ be the c-components of $\mathcal{G}[\mathbf{H}]$ where $\mathbf{H} \subseteq \mathbf{V}$. Then, without $Q[\mathbf{H}]$ being positive, $Q[\mathbf{H}] = \prod_j Q[\mathbf{H}_j]$.* ■

The second part (ii) involves positivity, which we will later generalize that can take the condition of $P(\mathbf{V})$ into account.

Lemma 5.2 (reproduced from Tian & Pearl 2003 with explicit positivity)**.** *Given $\mathbf{H} \subseteq \mathbf{V}$, let $\mathbf{H}_1, \ldots, \mathbf{H}_k$ be the c-components of $\mathcal{G}[\mathbf{H}]$. Let $\prec$ be a topological order over the variables in $\mathbf{H}$ according to $\mathcal{G}[\mathbf{H}]$ such that $V^{(1)} \prec V^{(2)} \cdots \prec V^{(|\mathbf{H}|)}$. Let $\mathbf{H}^{\preceq i}$ be the variables in $\mathbf{H}$ that come before $V^{(i)}$ including $V^{(i)}$. Let $\mathbf{H}^{\succ i}$ be the variables in $\mathbf{H}$ that come after $V^{(i)}$. Given $Q[\mathbf{H}] > 0$,*

$$Q[\mathbf{H}_j] = \prod_{V^{(i)} \in \mathbf{H}_j} \frac{Q[\mathbf{H}^{\preceq i}]}{Q[\mathbf{H}^{\preceq i-1}]}, \tag{11}$$

where $Q[\mathbf{H}^{\preceq i}] = \sum_{\mathbf{h}^{\succ i}} Q[\mathbf{H}]$. ■

One may think that the equality, which plays a key role in IDENTIFY, is well-defined if the numerators are well-defined (i.e., ≥ 0) and the denominators are positive. However, this approach does not fully take into account an interplay between positivity and decomposition since Lem. 5.2 is derived under strict positivity, $Q[\mathbf{H}] > 0$.

5.1. Relaxed Generalized Q-decomposition

This motivates us to generalize Lem. 5.2 for algorithmic identification under relaxed positivity. The intuition behind the generalization is that the product of fractions often can be shortened by canceling out depending on the topological order. We illustrate an example in Fig. 5. Here, $Q[\mathbf{H}]$ is factorized as $Q[\mathbf{H}] = Q[\mathbf{H}_1] \cdot Q[\mathbf{H}_2]$ where $\mathbf{H}_1 = \{V_1, V_2, V_4, V_6, V_7\}$ and $\mathbf{H}_2 = \{V_3, V_5\}$. Denoting $Q[\mathbf{H}^{\preceq i}]$ as Q_i for brevity, if $Q[\mathbf{H}] = Q_7 > 0$, then $Q[\mathbf{H}_1] = \frac{Q_7}{Q_6} \cdot \frac{Q_6}{Q_5} \cdot \frac{Q_4}{Q_3} \cdot \frac{Q_2}{Q_1} \cdot \frac{Q_1}{Q_0}$ and $Q[\mathbf{H}_2] = \frac{Q_5}{Q_4} \cdot \frac{Q_3}{Q_2}$ by Lem. 5.2. Since Q_6 and Q_1 can be canceled out, we can write $Q[\mathbf{H}_1] = \frac{Q_7}{Q_5} \cdot \frac{Q_4}{Q_3} \cdot \frac{Q_2}{Q_0}$.

We will show that this expression is valid if $Q_5 > 0$, and further show that it is still possible to identify $Q[\mathbf{H}_1]$ when some of the denominators are 0, i.e., $Q_5 = 0$ or $Q_3 =$

0, relaxing the strict positivity condition of $Q[\mathbf{H}] > 0$ in Lem. 5.2. To begin with, we define some indices, called anchors, useful to characterize such cancellation.

Definition 5.1 (Anchors)**.** Given $\mathbf{H} \subseteq \mathbf{V}$, let $\prec$ be a topological order over the variables in $\mathbf{H}$ according to $\mathcal{G}[\mathbf{H}]$. Let $\mathbf{C} \in \text{cc}(\mathcal{G}[\mathbf{H}])$ where $\mathbf{C}$ can be partitioned into variables (maximally) consecutive in a topological order $\prec$ such that $1 \leq l_1 \leq r_1 < l_2 \leq r_2 < \cdots < l_T \leq r_T \leq |\mathbf{H}|$ and $\mathbf{C} = \{V^{(i)} \mid i \in \cup_{d=1}^{T}[l_d, r_d]\}$. We say $\{(l_d, r_d)\}_{d=1}^{T}$ as *anchors* for $\mathbf{C}$ and denoted as $I_{\mathcal{G}[\mathbf{H}],\prec}(\mathbf{C})$. ■

In Fig. 5, $\{(1,2),(4,4),(6,7)\}$ and $\{(3,3),(5,5)\}$ is anchors for $\mathbf{H}_1$ and $\mathbf{H}_2$, respectively, with maximally consecutive segments. Note that $\{(1,1),(2,2),(4,4),(6,6),(7,7)\}$ is also valid anchors for $\mathbf{H}_1$ employed in Lem. 5.2.

Theorem 5.1. *Given $\mathbf{H} \subseteq \mathbf{V}$, let $\mathbf{H}' \in \text{cc}(\mathcal{G}[\mathbf{H}])$ where $I_{\mathcal{G}[\mathbf{H}],\prec}(\mathbf{H}') = \{(l_d, r_d)\}_{d=1}^{T}$. Then, the following holds:*

(i) If $Q[\mathbf{H}^{\preceq l_T - 1}] > 0$,

$$Q[\mathbf{H}'] = \prod_{d=1}^{T} \frac{Q[\mathbf{H}^{\preceq r_d}]}{Q[\mathbf{H}^{\preceq l_d - 1}]}. \tag{12}$$

(ii) If $Q[\mathbf{H}^{\preceq r_m}] = 0$ and $Q[\mathbf{H}^{\preceq l_m - 1}] > 0$ for some m,

$$Q[\mathbf{H}'] = 0.$$ ■

This generalizes Lem. 5.2 where irrelevant c-factors can be canceled out taking the positivity of $Q[\mathbf{H}^{\preceq l_T - 1}]$. Further, even when such positivity assumption is violated, still it provides a condition where the c-factor is identified as zero. In Fig. 5, Thm. 5.1 states that $Q[\mathbf{H}_1] = \frac{Q_7}{Q_5} \cdot \frac{Q_4}{Q_3} \cdot \frac{Q_2}{Q_0}$ if $Q_5 > 0$, and $Q[\mathbf{H}_2] = \frac{Q_5}{Q_4} \cdot \frac{Q_3}{Q_2}$ if $Q_4 > 0$. If $Q_4 = 0$ and $Q_3 > 0$, $Q[\mathbf{H}_1] = 0$. Similarly, if $Q_2 = 0$ then $Q[\mathbf{H}_1] = 0$. On the other hand, if $Q_4 > 0$ and $Q_5 = 0$, we cannot make a conclusion on $Q[\mathbf{H}_1]$. We demonstrate an example where a c-factor is non-identifiable when Thm. 5.1 is not applicable.

Example 5.1 (Front-door)**.** Consider two SCMs $\mathcal{M}_1, \mathcal{M}_2$ that share the same $\mathbf{F}$, i.e., $X \leftarrow U_1, Z \leftarrow X \wedge \neg U_2, Y \leftarrow \neg Z \vee (\neg U_1 \wedge U_3)$, but differ in the distribution over $\mathbf{U}$. For $\mathcal{M}_1$, $U_1, U_2 \sim \text{Ber}(0.5), U_3 \sim \text{Ber}(0.2)$ and, for $\mathcal{M}_2$, $U_3 \sim \text{Ber}(0.8)$. Here, $P^{\mathcal{M}_1}(\mathbf{V}) = P^{\mathcal{M}_2}(\mathbf{V})$. Thm. 5.1-(i) states that if $P(x,z,y) > 0$, then

$$Q[X,Y](x,z,y) = \frac{Q[X,Z,Y]}{Q[X,Z]} \frac{Q[X]}{Q[\emptyset]} = P(y \mid x,z)P(x) > 0.$$

Thm. 5.1-(ii) states that if ($P(x,z,y) = 0$ and $P(x,z) > 0$), or $P(x) = 0$, then $Q[X,Y](x,z,y) = 0$. Thm. 5.1 is inapplicable to identify $Q[X,Y]$ from $Q[X,Z,Y] = P(\mathbf{V})$ if $P(x,z) = 0$ and $P(x) > 0$. Consider $Q[X,Y](x = 0, z = 1, y = 1)$ where $P(x = 0, z = 1) = 0$ and $P(x = 0) = 0.5$. Here, the two models disagree: $Q[X,Y]^{\mathcal{M}_1}(x = 0, z = 1, y = 1) = 0.1 \neq Q[X,Y]^{\mathcal{M}_2}(x = 0, z = 1, y = 1) = 0.4$, thus, $Q[X,Y](x = 0, z = 1, y = 1)$ is non-identifiable. ■

Unlike Lem. 5.2 where any topological order can be applied *safely*, Thm. 5.1 depends on the choice of topological order and arguments of the c-factors since they determine which c-factor that can be zero to appear as a denominator.

Example 5.2 (Order dependency without strict positivity)**.** Consider the following functions: $W \leftarrow U_1$, $Z \leftarrow \neg U_1 \wedge U_2$, $X \leftarrow \neg W \wedge U_3$, $Y \leftarrow \neg Z \wedge U_3$, where $U_1, U_2, U_3 \sim \text{Ber}(0.5)$. Here, we consider identifying $Q[X,Y](w = 1, z = 1, x = 1, y = 1)$. Under a topological order $W \prec Z \prec X \prec Y$, the expression of $Q[X,Y]$ following Eq. (12), if valid, is $Q[X,Y] = \frac{Q[W,Z,X,Y]}{Q[W,Z]}$. Here, we cannot reach a conclusion under the topological order $\prec$ since $Q[W,Z](w = 1, z = 1) = 0$. However, with a different topological order $W \prec X \prec Z \prec Y$, the expression of $Q[X,Y]$ from Eq. (12) becomes $Q[X,Y] = \frac{Q[W,X,Z,Y]}{Q[W,X,Z]} \cdot \frac{Q[W,X]}{Q[W]}$. Since $Q[W,X](w = 1, x = 1) = 0$ and $Q[W](w = 1) = 0.5 > 0$, we have $Q[X,Y](w = 1, z = 1, x = 1, y = 1) = 0$ by Thm. 5.1-(ii). ■

As demonstrated in Examples 5.1 and 5.2, we were not able to make a conclusion when Thm. 5.1 is inapplicable. Thus, a few research questions remain such as an efficient way of finding a topological order that admits the theorem, or the necessity of the theorem, i.e., the non-identifiability of a causal quantity when the theorem is not applicable in any topological order. Answering those questions is beyond the scope of this paper, and we leave them for future work.

6. Algorithmic Positivity

Based on the characterization of c-factors with respect to positivity (Prop. 5.1, Lem. 5.1, and Thm. 5.1), we devise an identification algorithm named IDENTIFY+, which simultaneously returns a positivity condition implied by the resulting identification formula. We limit the algorithm to be *data-agnostic*, where the formula is driven without the knowledge of the positivity of an underlying observational distribution (we discuss the design of positivity-aware identification algorithm in Sec. 7.2.) This condition is constructed on-the-fly by repeatedly updating the conditions for a c-factor of interest to other c-factors that induce the c-factor.

Propagating a positivity condition We present rules that translate a constraint over a c-factor Q to the constraints over the c-factors in the formula for Q. This makes use of the fact that every identifiable c-factor can be expressed using operators such as summation, product, and fraction of expressions or, simply, an observational distribution $P(\mathbf{V})$.

Proposition 6.1. *Consider a c-factor and its expression induced by marginalization, Lem. 5.1, or Thm. 5.1 for the c-factor. Then, a constraint over the expression of the c-factor is implied by a statement of first-order logic as follows:*

$$\textstyle\sum_{\mathbf{z}} g(\mathbf{z}) \geq 0 \impliedby \forall \mathbf{z}(g(\mathbf{z}) \geq 0),$$

Algorithm 1 IDENTIFY+ (outer)

1: **input**: $\mathbf{x}, \mathbf{y} \subseteq \mathfrak{X}_{\{X,Y\}}$, $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$, $\mathcal{G}$.
2: **output**: an expression and positivity condition for $P_{\mathbf{x}}(\mathbf{y})$.
3: Initialize constraint E with $\exists \mathbf{x}^+ \backslash \mathbf{x}\, (P_{\mathbf{x}^+}(\mathbf{y}) \geq 0)$.
4: EXPAND$(E, P_{\mathbf{x}^+}(\mathbf{y}), \sum_{\mathbf{y}^+ \backslash \mathbf{Y}} \prod_{\mathbf{C}_j \in \mathrm{cc}(\mathcal{G}[\mathbf{Y}^+])} Q[\mathbf{C}_j])$
5: **for** $\mathbf{C}_j \in \mathrm{cc}(\mathcal{G}[\mathbf{Y}^+])$ **do**
6: Let $\mathbf{S}_k \in \mathrm{cc}(\mathcal{G}[An(\mathbf{Y})_{\mathcal{G}}])$ such that $\mathbf{C}_j \subseteq \mathbf{S}_k$.
7: $f_{Q[\mathbf{C}_j]|Q[\mathbf{S}_k]} =$ IDENTIFY+$(\mathbf{C}_j, \mathcal{G}[\mathbf{S}_k], Q[\mathbf{S}_k], E)$ **if** $\mathbf{C}_j \neq \mathbf{S}_k$
8: EXPAND$(E, Q[\mathbf{S}_k], f_{Q[\mathbf{S}_k]|P})$
9: record $f_{Q[\mathbf{C}_j]|P} = f_{Q[\mathbf{C}_j]|Q[\mathbf{S}_k]} \circ f_{Q[\mathbf{S}_k]|P}$
10: **end for**
11: **return** $\sum_{\mathbf{y}^+ \backslash \mathbf{Y}} \prod_{\mathbf{C}_j \in \mathrm{cc}(\mathcal{G}[\mathbf{Y}^+])} f_{Q[\mathbf{C}_j]|P}, E$

Algorithm 2 IDENTIFY+ (inner)

1: **input**: $\mathbf{C}, \mathcal{G} = \mathcal{G}[\mathbf{T}], Q = Q[\mathbf{T}]$, constraint E.
2: **output**: an expression for $Q[\mathbf{C}]$ made with $Q[\mathbf{T}]$.
3: **if** $(\mathbf{A} := An(\mathbf{C})_{\mathcal{G}[\mathbf{T}]}) = \mathbf{C}$ **then**
4: EXPAND$(E, Q[\mathbf{C}], f_{Q[\mathbf{C}]|Q[\mathbf{T}]} := \sum_{\mathbf{t} \backslash \mathbf{c}} Q[\mathbf{T}])$
5: **return** $f_{Q[\mathbf{C}]|Q[\mathbf{T}]}$
6: **else if** $\mathbf{A} = \mathbf{T}$ **then**
7: **raise** FAIL.
8: **else**
9: Let $\mathbf{T}'$ be a c-component in $\mathcal{G}[\mathbf{A}]$ s.t. $\mathbf{C} \subseteq \mathbf{T}'$.
10: $f_{Q[\mathbf{C}]|Q[\mathbf{T}']} =$ IDENTIFY+$(\mathbf{C}, \mathcal{G}[\mathbf{T}'], Q[\mathbf{T}'], E)$ **if** $\mathbf{C} \neq \mathbf{T}'$
11: EXPAND$(E, Q[\mathbf{T}'], f_{Q[\mathbf{T}']|Q[\mathbf{A}]} \circ f_{Q[\mathbf{A}]|Q[\mathbf{T}]})$
12: **return** $f_{Q[\mathbf{C}]|Q[\mathbf{T}']} \circ f_{Q[\mathbf{T}']|Q[\mathbf{A}]} \circ f_{Q[\mathbf{A}]|Q[\mathbf{T}]}$
13: **end if**

$$\sum_{\mathbf{z}} g(\mathbf{z}) > 0 \impliedby (\forall \mathbf{z}(g(\mathbf{z}) \geq 0)) \wedge (\exists \mathbf{z}(g(\mathbf{z}) > 0)),$$
$$\sum_{\mathbf{z}} g(\mathbf{z}) = 0 \impliedby \forall \mathbf{z}(g(\mathbf{z}) = 0),$$
$$\prod_i g_i \geq 0 \impliedby (\exists i(g_i = 0)) \vee (\forall i(g_i \geq 0)),$$
$$\prod_i g_i > 0 \impliedby \forall i(g_i > 0),$$
$$\prod_i g_i = 0 \impliedby \exists i(g_i = 0),$$
$$g_1/g_2 \geq 0 \impliedby g_1 \geq 0 \wedge g_2 > 0,$$
$$g_1/g_2 > 0 \impliedby g_1 > 0 \wedge g_2 > 0,$$
$$g_1/g_2 = 0 \impliedby g_1 = 0 \wedge g_2 > 0.$$

■

We denote by EXPAND an algorithm that rewrites a constraint following Prop. 6.1. It takes a constraint expression E, a c-factor Q, and its expression using other c-factors. For example, if $Q_1 = \sum_a Q_2$, then EXPAND$(Q_1 = 0, Q_1, \sum_a Q_2)$ would return $\forall a(Q_2 = 0)$. Further, if we have $Q_2 = \frac{Q_6}{Q_5}\frac{Q_4}{Q_3}$ using Thm. 5.1-(i), EXPAND$(\forall a(Q_2 = 0), Q_2, \frac{Q_6}{Q_5} \cdot \frac{Q_4}{Q_3})$ returns $\forall a\big((Q_6{=}0 \wedge Q_5{>}0) \vee (Q_4{=}0 \wedge Q_3{>}0)\big)$. We would like to emphasize that (ii) of Thm. 5.1 plays a pivotal role in inducing such positivity conditions in conjunction with Prop. 6.1, while deriving the formula itself follows (i) of Thm. 5.1 and other machineries.

IDENTIFY+ algorithm We incorporate EXPAND into algorithmic identification[4]. Algorithm IDENTIFY+ transforms and factorizes a given query based on Q-decomposition (Alg. 1) and delegates the identification of each c-factor to its inner part (Alg. 2). To begin with, we introduce a new notation $f_{g|h}$ which is a formula for g with respect to h derived based on marginalization, Lem. 5.1, and Thm. 5.1. We denote symbolic substitution by $\circ$ so that $f_{g|h} \circ f_{h|i}$ is $f_{g|h}$ with h replaced with $f_{h|i}$ yielding $f_{g|i}$. Further, abusing notation, we denote by $P_{\mathbf{x}}(\mathbf{y}) \geq 0$ to mean a constraint on an expression made of $P(\mathbf{V})$ such that $P_{\mathbf{x}}(\mathbf{y}) = f(P(\mathbf{V})) \geq 0$, where the expression f is the output of the algorithm.

Alg. 1 begins with the initialization of $\mathbf{X}^+$, $\mathbf{Y}^+$, and constraint $E = \exists \mathbf{x}^+ \backslash \mathbf{x}\, (P_{\mathbf{x}^+}(\mathbf{y}) \geq 0)$ (Line 3), which is then expanded with respect to the formula (Eqs. (8) and (9), Line 4). Each c-factor is then identified, and its condition is examined (Lines 5–10). Line 7 induces $f_{Q[\mathbf{C}_j]|Q[\mathbf{S}_k]}$ an expression for $Q[\mathbf{C}_j]$ with $Q[\mathbf{S}_k]$ following Alg. 2 (or identity if $\mathbf{C}_j = \mathbf{S}_k$), and Line 8 translates constraint E written in $Q[\mathbf{S}_k]$ to P regarding $f_{Q[\mathbf{S}_k]|P}$, which is made by Thm. 5.1. Finally, the identification formula and condition are returned (Line 11) with formulas for c-factors recorded in Line 9.

The inner algorithm (Alg. 2) attempts to identify $Q[\mathbf{C}]$ with $Q[\mathbf{T}]$ (i.e., $Q[\mathbf{C}_j]$ and $Q[\mathbf{S}_k]$, respectively, in the outer, Line 7) and returns $f_{Q[\mathbf{C}]|Q[\mathbf{T}]}$ updating constraint E with $Q[\mathbf{C}]$ replaced to $Q[\mathbf{T}]$. In a simple case, marginalization is sufficient. Otherwise, a recursion occurs to identify $Q[\mathbf{C}]$ with $Q[\mathbf{T}']$, based on Thm. 5.1 (i), which is acquired as a c-factor from $Q[\mathbf{A}]$, a marginalization of $Q[\mathbf{T}]$. Once the recursion is terminated, we propagate the constraint over $Q[\mathbf{C}]$ to that over $Q[\mathbf{T}']$ (Line 11). Once terminated, the outer algorithm will take care of further expressing the constraint using $Q[\mathbf{T}]$ (Line 8 in Alg. 1). We provide running examples in Appendix C.

Theorem 6.1. *Whenever* IDENTIFY+ *returns a positivity condition and identification formula, they are correct.* ■

We focused on deriving a positivity condition *on-the-fly* while deriving a formula relaxed in part by Thm. 5.1. However, this does not take the given observational distribution's positivity into account. We present in Sec. 7.2 modifications that need to be augmented for IDENTIFY+ to identify a query adapting to the condition of given data.

7. Discussion

We discuss on (i) the validity of post-hoc analysis, (ii) the design of a positivity-aware identification algorithm, and (iii) generalization to conditional causal effects. An extended discussion on other topics is provided in Appendix F.

[4]We focus on returning an identification formula rather than evaluating the value (i.e., estimation).

7.1. Validity of Post-hoc Analysis

Now we formally examine the validity of post-hoc analysis. In post-hoc analysis, the sufficient conditions we considered in deriving fractions, summation, and multiplication (Sec. 6) are identical to those derived from the form of the identification formula (Prop. 6.1). With the formula *intact* (without further reduction from the resulting algorithm), the recursive application of marginalization and Q-decomposition will leave imprints on the formula that can be back-traced. Hence, the positivity condition can be reconstructed equally.

Proposition 7.1. *Consider an identification formula derived through* IDENTIFY *or* IDENTIFY+. *Post-hoc analysis driven by Prop. 6.1 yields a sufficient positivity condition.* ■

We conjecture that the post-hoc analysis is sound for any identification expression made with product, marginalization, and fraction, irrespect to the choice of methods, e.g., do-calculus, IDENTIFY, ID, identification criteria, etc. However, it is beyond the scope of this paper due to its generality.

7.2. Considerations for Positivity-Aware Identification

For the design of a positivity-aware identification algorithm, we present three key components that need to be considered, namely, topological order, fixing values, and latent projection. With positivity checked for the given distribution, the new algorithm can further take advantage of (ii) of Thm. 5.1.

To utilize Thm. 5.1, IDENTIFY+ algorithm needs to take topological orders into account. In particular, within a loop over topological orders, the inner part of IDENTIFY+ needs to be called taking the order as an additional argument.

The choice of value for rule 3 (e.g., r in Napkin) can be crucial under non-strict positivity. Thus, the outer algorithm, where the value $\mathbf{x}^+ \setminus \mathbf{x}$ is fixed, needs to enumerate different instantiations for a subset of $\mathbf{X}^+ \setminus \mathbf{X}$, e.g., $\{R\} \subset \{R, W\}$ for Napkin. However, we do not take into account of the value $\mathbf{y}^+ \setminus \mathbf{y}$ since it is already associated with marginalization, which iterates over all possible values.

Under strict positivity, the absence of a hedge for $P_{\mathbf{x}}(\mathbf{y})$ is necessary and sufficient for its identification. If a latent projection does not induce a hedge, IDENTIFY or ID yields a simpler expression with fewer variables, but it does not turn a non-identifiable effect into an identifiable one. However, under non-strict positivity, Q-decomposition under a projection (without inducing a hedge) may avoid the case of $\frac{0}{0}$ in Thm. 5.1 as demonstrated in Appendix D. In particular, the violation of positivity in a specific c-factor can be avoided by projecting out the variables involved, resulting in the consolidation of multiple c-components. This leads to different c-factors for $P_{\mathbf{x}}(\mathbf{y})$. Subsequently, the algorithm can be rerun with the projection to identify a causal effect. Hence, positivity-aware projection is a crucial component.

7.3. Generalization to Conditional Causal Effects

A sound and complete identification criterion for a conditional causal effect under the strict positivity is shown in Tian (2004); Shpitser & Pearl (2006b). Given a query $P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$, one can apply Rule 2 of do-calculus to maximally change conditions into interventions. For example, $P_{\mathbf{x},\mathbf{w}'}(\mathbf{y} \mid \mathbf{w}'')$ with $\mathbf{W} = \mathbf{W}' \dot{\cup} \mathbf{W}''$. Then, identifying the unconditional causal effect $P_{\mathbf{x},\mathbf{w}'}(\mathbf{y}, \mathbf{w}'')$ leads to the identification of $P_{\mathbf{x},\mathbf{w}'}(\mathbf{y} \mid \mathbf{w}'')$ (if $P_{\mathbf{x},\mathbf{w}'}(\mathbf{w}'') > 0$), which is equivalent to the original query $P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$. Since we have established sufficient conditions for both do-calculus and identification of marginal effects, our results indeed generalize to conditional causal effects as well.

8. Conclusion

Strict positivity is a long-standing critical assumption for causal inference, which is often unrealistic in many practical scenarios. We provided a comprehensive treatment to constructing a positivity condition required in graphical causal effect identification. In particular, we first informally explored ways of eliciting positivity innate in an identification formula in a post-hoc manner. We then provided positivity conditions for the rules of do-calculus so that one can induce a positivity condition together with an identification formula. Given the lack of a principled way to derive a formula using do-calculus, we delved into an algorithmic approach based on Q-decomposition, where we provided a relaxed positivity condition for Q-decomposition. Equipped with such characterization, we devised an identification algorithm that can simultaneously elicit a positivity condition. We discussed the design of a positivity-aware identification algorithm, which should account for topological orders and latent projections altogether, and generalization of our results to conditional causal effects. We hope this research sparks further investigation into the development of an identification algorithm that adapts to the positivity of given data.

Impact Statement

This paper presents work whose goal is to advance the field of causal inference and machine learning. In particular, relaxing some of the assumptions underlying the task of causal effect identification will help understand the characteristics between the available observational distribution and identification, further allowing the applicability of causal effect identification for previously considered impossible. There are some potential societal consequences of our work, none of which we feel must be specifically highlighted here.

Acknowledgements

We would like to thank Jin Tian for providing helpful and extensive feedback. We also thank anonymous reviewers for constructive comments to improve the manuscript. This work was partly supported by the IITP (2022-0-00953-PICA/20%), MFDS (23212MFDS202/20%), and NRF (RS-2023-00222663/20%, RS-2023-00211904/40%) grant funded by the Korean government.

References

Balke, A. *Probabilistic Counterfactuals: Semantics, Computation, and Applications*. PhD thesis, Computer Science Department, University of California, Los Angeles, CA, 11 1995.

Bareinboim, E., Correa, J. D., Ibeling, D., and Icard, T. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and causal inference: the works of judea pearl*, pp. 507–556. Association for Computing Machinery, 2022.

Dahabreh, I. J., Robins, J. M., Haneuse, S. J., and Hernán, M. A. Generalizing causal inferences from randomized trials: counterfactual and graphical identification. *arXiv preprint arXiv:1906.10792*, 2019.

Galles, D. and Pearl, J. An axiomatic characterization of causal counterfactuals. *Foundation of Science*, 3(1):151–182, 1998.

Geiger, D. and Pearl, J. On the logic of causal models. In *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, pp. 136–147, St Paul, MN, 1988.

Hernán, M. A. and Robins, J. M. Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health*, 60(7):578–586, 2006.

Hernán, M. A. and Robins, J. M. *Causal Inference: What if.* Chapman & Hall/CRC, Boca Raton, 1st edition, 2020.

Huang, Y. and Valtorta, M. Pearl's Calculus of Intervention Is Complete. In T.S. Richardson, R. D. a. (ed.), *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pp. 217–224, Corvallis, OR, 2006. AUAI Press. ISBN 0974903922.

Kolmogorov, A. Grundbegriffe der wahrscheinlichkeitrechnung, ergebnisse der mathematik. published in english in 1950 as foundations of the theory of probability, bronx, ny: Chelsea. 1941a, stationary sequences in hilbert space,(russian) bull. *Math. Univ. Moscow*, 2:3–14, 1933.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H. G.-G. Independence Properties of Directed Markov Fields. *Networks*, 20(5):491–505, 1990. ISSN 10970037. doi: 10.1002/net.3230200503.

Lee, S. and Bareinboim, E. Causal Effect Identifiability under Partial-Observability. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119. PMLR, 2020.

Li, A. and Pearl, J. Bounds on causal effects and application to high dimensional data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 5773–5780, 2022.

Pearl, J. [bayesian analysis in expert systems]: comment: graphical models, causality and intervention. *Statistical Science*, 8(3):266–269, 1993.

Pearl, J. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

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

Pearl, J. and Mackenzie, D. *The Book of Why*. Basic Books, New York, 2018. ISBN 978-0-465-09760-9.

Pearl, J. and Verma, T. S. A Theory of Inferred Causation. In Allen, J. A., Fikes, R., and Sandewall, E. (eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pp. 441–452, San Mateo, CA, 1991. Morgan Kaufmann.

Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., and Van Der Laan, M. J. Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research*, 21(1):31–54, 2012.

Robins, J. M. A new approach to causal inference in mortality studies with a sustained exposure period – applications to control of the healthy workers survivor effect. *Mathematical Modeling*, 7:1393–1512, 1986.

Shpitser, I. and Pearl, J. Identification of Joint Interventional Distributions in Recursive semi-Markovian Causal Models. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence*, volume 2, pp. 1219–1226, 2006a. ISBN 978-1-57735-281-5.

Shpitser, I. and Pearl, J. Identification of Conditional Interventional Distributions. In R. Dechter and T.S. Richardson (eds.), *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pp. 437–444, Corvallis, OR, 2006b. AUAI Press.

Spirtes, P., Glymour, C. N., and Scheines, R. *Causation, Prediction, and Search*. Springer-Verlag, New York, 1993.

Tian, J. Identifying Conditional Causal Effects. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, pp. 561–568, Arlington, Virginia, United States, 2004. AUAI Press. ISBN 0-9749039-0-6. URL http://dl.acm.org/citation.cfm?id=1036843.1036911.

Tian, J. and Pearl, J. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1-4):287–313, 2000.

Tian, J. and Pearl, J. A General Identification Condition for Causal Effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002)*, pp. 567–573, Menlo Park, CA, 2002. AAAI Press/The MIT Press.

Tian, J. and Pearl, J. On the identification of causal effects. Technical Report R-290-L, Department of Computer Science, University of California, Los Angeles, CA, 2003.

Tikka, S. and Karvanen, J. Simplifying probabilistic expressions in causal inference. *The Journal of Machine Learning Research*, 18(1):1203–1232, 2017.

Tikka, S. and Karvanen, J. Enhancing identification of causal effects by pruning. *Journal of Machine Learning Research*, 18(1):7072–7094, 2018. ISSN 15337928. doi: 10.1016/j.cellbi.2006.10.004.

Verma, T. and Pearl, J. Causal networks: Semantics and expressiveness. In *Machine intelligence and pattern recognition*, volume 9, pp. 69–76. Elsevier, 1990.

Verma, T. S. Graphical Aspects of Causal Models. Technical Report R-191, Cognitive Systems Laboratory, Computer Science Department, University of California, Los Angeles, Los Angeles, CA, 1993.

Zhang, J., Tian, J., and Bareinboim, E. Partial counterfactual identification from observational and experimental data. In *International Conference on Machine Learning*, pp. 26548–26558. PMLR, 2022.

A. Preliminaries

Here we present definitions of d-separation, latent projection, and causal effect $P_{\mathbf{x}}(\mathbf{y})$.

d-separation Excerpt from (Geiger & Pearl, 1988). For any three disjoint subsets $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ of nodes in a DAG $\mathcal{G}$, $\mathbf{Z}$ is said to d-separate $\mathbf{X}$ from $\mathbf{Y}$ if there is no path from a node in $\mathbf{X}$ to a node in $\mathbf{Y}$ along which (i) every node that delivers an arrow is outside $\mathbf{Z}$, and (ii) every node with converging arrows either is in $\mathbf{Z}$ or has a descendant in $\mathbf{Z}$.

Latent projection Excerpt from (Tian & Pearl, 2003). The latent projection of a graph $\mathcal{G}$ over $\mathbf{V}$ on $\mathbf{C} \subseteq \mathbf{V}$, denoted $\mathcal{G}_{\langle \mathbf{C} \rangle}$, is a graph over $\mathbf{C}$ such that, in addition to edges in $\mathcal{G}_{\langle \mathbf{C} \rangle}$, for every pair of vertices $(V_i, V_j) \in \mathbf{C}$, (i) add a directed edge $V_i \to V_j$ in $\mathcal{G}_{\langle \mathbf{C} \rangle}$ if there exists a directed path from V_i to V_j in $\mathcal{G}$ such that every vertex on the path is not in $\mathbf{C}$, and (ii) add a bidirected edge $V_i \leftrightarrow V_j$ in $\mathcal{G}_{\langle \mathbf{C} \rangle}$ if there exists a divergent path between V_i and V_j in $\mathcal{G}$ such that every vertex on the path is not in $\mathbf{C}$.

Causal effect $P_{\mathbf{x}}(\mathbf{y})$ Under Pearl's causal hierarchy (Bareinboim et al., 2022, Def. 5), $P(\mathbf{y}_{\mathbf{x}}) := \sum_{\{\mathbf{u} | \mathbf{Y}_{\mathbf{x}}(\mathbf{u}) = \mathbf{y}\}} P(\mathbf{u})$ where the potential response $\mathbf{Y}_{\mathbf{x}}(\mathbf{u})$ is defined as the solution for $\mathbf{Y}$ of the set of equations $\mathbf{F}_{\mathbf{x}}$ with respect to SCM $\mathcal{M}$. Importantly, this definition is ***irrelevant*** to strict positivity. Similarly, the subsequent definition of conditional causal effect, i.e., $P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y}, \mathbf{w}) / P_{\mathbf{x}}(\mathbf{w})$ if $P_{\mathbf{x}}(\mathbf{w}) > 0$, is also well-established ***regardless*** of the strict positivity assumption.

B. Properties of C-factors

Proposition B.1. *Let* $\mathbf{C} \subseteq \mathbf{D}$*, if* $Q[\mathbf{D}](\mathbf{d}, \mathbf{pa}_{\mathbf{D}}) > 0$*, then* $Q[\mathbf{C}](\mathbf{c}, \mathbf{pa}_{\mathbf{C}}) > 0$ *with* $\mathbf{c} \cup \mathbf{pa}_{\mathbf{C}}$ *consistent with* $\mathbf{d} \cup \mathbf{pa}_{\mathbf{D}}$. ■

Proof. First, $\mathbf{C} \cup Pa(\mathbf{C}) \subseteq \mathbf{D} \cup Pa(\mathbf{D})$. By definition of c-factor, there exists value $\mathbf{u}(\mathbf{D})$ with $P(\mathbf{u}(\mathbf{D})) > 0$ that yields $\mathbf{d}$ given $Pa(\mathbf{D}) = \mathbf{pa}_{\mathbf{D}}$. Since $\mathbf{u}(\mathbf{C}) \subseteq \mathbf{u}(\mathbf{D})$, $P(\mathbf{u}(\mathbf{C})) > 0$, and $\mathbf{C}$ yields $\mathbf{c}$ consistent to $\mathbf{d}$ with $Pa(\mathbf{C})$ fixed to the values consistent with $\mathbf{d} \cup \mathbf{pa}_{\mathbf{D}}$. □

Corollary B.1. *Let* $\mathbf{C} \subseteq \mathbf{D}$*, if* $Q[\mathbf{C}](\mathbf{c}, \mathbf{pa}_{\mathbf{C}}) = 0$*, then* $Q[\mathbf{D}](\mathbf{d}, \mathbf{pa}_{\mathbf{D}}) = 0$ *with* $\mathbf{c} \cup \mathbf{pa}_{\mathbf{C}}$ *consistent with* $\mathbf{d} \cup \mathbf{pa}_{\mathbf{D}}$. ■

Corollary B.2. *Given* $\mathbf{H} \subseteq \mathbf{V}$*, let* $\prec$ *be a topological order over the variables in* $\mathbf{H}$ *according to* $\mathcal{G}[\mathbf{H}]$*. If* $i \preceq j$ *and* $Q[\mathbf{H}^{\preceq i}] = 0$*, then* $Q[\mathbf{H}^{\preceq j}] = 0$. ■

Corollary B.3. *Given* $\mathbf{H} \subseteq \mathbf{V}$*, let* $\prec$ *be a topological order over the variables in* $\mathbf{H}$ *according to* $\mathcal{G}[\mathbf{H}]$*. If* $i \preceq j$ *and* $Q[\mathbf{H}^{\preceq j}] > 0$*, then* $Q[\mathbf{H}^{\preceq i}] > 0$. ■

Proposition B.2. *Given* $\mathbf{H} \subseteq \mathbf{V}$*, let* $\prec$ *be a topological order over the variables in* $\mathbf{H}$ *according to* $\mathcal{G}[\mathbf{H}]$*. Let* $i \preceq j$*. Then, the following statements are equivalent.*

(a) $Q[\mathbf{H}^{\preceq i-1}] > 0$ *and* $Q[\mathbf{H}^{\preceq j}] = 0$.

(b) There exists $t \in [i, j]$ *such that* $Q[\mathbf{H}^{\preceq t-1}] > 0$ *and* $Q[\mathbf{H}^{\preceq t}] = 0$.

■

Proof. ((a) ⇒ (b)) We give a proof by contradiction. Suppose there does not exist $t \in [i, j]$ such that $Q[\mathbf{H}^{\preceq t-1}] > 0$ and $Q[\mathbf{H}^{\preceq t}] = 0$. Therefore, for all $t \in [i, j]$, $Q[\mathbf{H}^{\preceq t-1}] = 0$ if $Q[\mathbf{H}^{\preceq t}] = 0$. Since $Q[\mathbf{H}^{\preceq j}] = 0$, we have $Q[\mathbf{H}^{\preceq j}] = Q[\mathbf{H}^{\preceq j-1}] = \cdots Q[\mathbf{H}^{\preceq i-1}] = 0$, which contradicts the hypothesis $Q[\mathbf{H}^{\preceq t-1}] > 0$.

((b) ⇒ (a)) By Corollary B.2, $Q[\mathbf{H}^{\preceq j}] = 0$ since $t \leq j$ and $Q[\mathbf{H}^{\preceq t}] = 0$. Also, by Corollary B.3, $Q[\mathbf{H}^{\preceq i-1}] > 0$ since $i - 1 \leq t - 1$ and $Q[\mathbf{H}^{\preceq t-1}] > 0$. □

C. Derivation of Positivity for Backdoor, Front-door, and Napkin graphs

We highlight a few sentences in order for the readers not to be confused when deriving some of the positivity conditions in this section. We present positivity conditions for backdoor, front-door, and Napkin graphs derived based on post-hoc analysis, do-calculus, generalized Q-decomposition (manually), and algorithmic approach (IDENTIFY+).

C.1. Back-door graph

To begin with note that the term $P(x, z) > 0$ is equivalent to $P(x \mid z) > 0$.

Post-hoc Analysis

$$
\begin{aligned}
P_x(y) \geq 0 &\Leftarrow \sum_z P(y \mid x, z) P(z) \geq 0 \\
&\Leftarrow \forall z \big(\overbrace{P(y \mid x, z)}^{①} \overbrace{P(z)}^{②} \geq 0 \big) \\
&\Leftarrow \forall z \big((① = 0) \vee (② = 0) \vee (① \geq 0 \wedge ② \geq 0) \big) \\
&\Leftarrow \forall z (((P(y, x, z) = 0 \wedge P(x, z) > 0) \vee P(z) = 0) \vee (P(x, z) > 0 \wedge P(z) \geq 0) \\
&\Leftrightarrow \forall z (((P(y, x, z) = 0 \wedge P(x, z) > 0) \vee P(z) = 0) \vee (P(x, z) > 0) \\
&\Leftrightarrow \forall z (P(x, z) > 0 \vee P(z) = 0).
\end{aligned}
$$

This is equivalent to $\forall z (P(x \mid z) > 0 \vee P(z) = 0)$.

Do-Calculus We show do-calculus-based derivation and check the positivity condition for the backdoor graph:

$$
\begin{aligned}
P_x(y) &= \sum_z P_x(y \mid z) P_x(z) \\
&= \sum_z P(y \mid x, z) P_x(z) && \text{if } \forall z (P(x, z) > 0 \vee P_x(z) = 0) \\
&= \sum_z P(y \mid x, z) P(z).
\end{aligned}
$$

Plugging in $P_x(z) = P(z)$, this derivation yields $\mathrm{adj}(x; Z)$, which is the same for the post-hoc analysis.

Q-decomposition (manually)

$$
P_x(y) = \sum_z Q[\{Y, Z\}] = \sum_z Q[\{Y\}](y, x, z) Q[\{Z\}](z),
$$

where

$$
\begin{aligned}
P_x(y) \geq 0 &\Leftarrow \forall z Q[\{Y, Z\}] \geq 0 \\
&\Leftarrow \forall z (Q[\{Y\}] = 0 \vee Q[\{Z\}] = 0 \vee (Q[\{Y\}] \geq 0 \wedge Q[\{Z\}] \geq 0)).
\end{aligned}
$$

By Thm. 5.1, for all z, each of $Q[\{Y\}](y, x, z)$ and $Q[\{Z\}](z)$ is further identified as

$$
\begin{aligned}
&Q[\{Y\}](y, x, z) = \frac{Q[\{X, Y, Z\}]}{Q[\{X, Z\}]} = \frac{P(x, y, z)}{P(x, z)} \quad \text{if } P(x, z) > 0, \\
&Q[\{Y\}](y, x, z) = 0 \quad \text{if } P(y \mid x, z) = 0,
\end{aligned}
$$

and

$$
Q[\{Z\}](z) = \sum_{x, y} Q[\{X, Y, Z\}] = P(z).
$$

Combining those, we arrive at

$$
\begin{aligned}
P_x(y) \geq 0 &\Leftarrow \forall z (Q[\{Y\}] = 0 \vee Q[\{Z\}] = 0 \vee (Q[\{Y\}] \geq 0 \wedge Q[\{Z\}] \geq 0)) \\
&\Leftarrow \forall z (P(y \mid x, z) = 0 \vee P(z) = 0 \vee (P(x, z) > 0 \wedge P(z) \geq 0)) \\
&\Leftrightarrow \forall z (P(y \mid x, z) = 0 \vee P(z) = 0 \vee P(x, z) > 0) \\
&\Leftrightarrow \forall z (P(z) = 0 \vee P(x, z) > 0) \\
&\Leftrightarrow \forall z (P(z) = 0 \vee P(x \mid z) > 0).
\end{aligned}
$$

We end up with exactly the same positivity conditions as post-hoc analysis and do-calculus aforementioned.

Algorithmic Approach

- (outer, line 3) $E = (P_x(y) \geq 0)$
- (outer, line 4) with expression $\sum_z Q[\{Y\}]Q[\{Z\}]$, E is expanded as
$$\forall z(Q[\{Y\}] = 0 \vee Q[\{Z\}] = 0 \vee (Q[\{Y\}] \geq 0 \wedge Q[\{Z\}] \geq 0)).$$
- (outer, line 5) case $Q[\{Y\}]$
  - (outer, line 6) $\mathbf{S}_k = \{Y\}$,
  - (outer, line 7) $f_{Q[\{Y\}]|Q[\{Y\}]}$ is an identity function, and calling the inner part is skipped.
  - (outer, line 8) computes $Q[\{Y\}] = \frac{P(y,x,z)}{P(x,z)}$ and expands E as
$$\forall z((P(y,x,z) = 0 \wedge P(x,z) > 0) \vee Q[\{Z\}] = 0 \vee (P(x,z) > 0 \wedge Q[\{Z\}] \geq 0)).$$
  - (outer, line 9) records $Q[\{Y\}] = \frac{P(y,x,z)}{P(x,z)}$.
- (outer, line 5) case $Q[\{Z\}]$
  - (outer, line 6) $\mathbf{S}_k = \{Z\}$,
  - (outer, line 7) is skipped.
  - (outer, line 8) computes $Q[\{Z\}] = P(z)$ and expands E as
$$\forall z((P(y,x,z) = 0 \wedge P(x,z) > 0) \vee P(z) = 0 \vee (P(x,z) > 0 \wedge P(z) \geq 0)).$$
  - (outer, line 9) records $Q[\{Z\}] = P(z)$.
- (outer, line 11) returns $\sum_z \frac{P(y,x,z)}{P(x,z)} P(z)$ and up-to-date E.

Further reduction leads to the same conclusion as other approaches:

$$\begin{aligned}
&\forall z((P(y,x,z) = 0 \wedge P(x,z) > 0) \vee P(z) = 0 \vee (P(x,z) > 0 \wedge P(z) \geq 0)) \\
&\Leftrightarrow \forall z((P(y,x,z) = 0 \wedge P(x,z) > 0) \vee P(z) = 0 \vee P(x,z) > 0) \\
&\Leftrightarrow \forall z(P(z) = 0 \vee P(x,z) > 0) \\
&\Leftrightarrow \forall z(P(z) = 0 \vee P(x \mid z) > 0).
\end{aligned}$$

C.2. Front-door graph

<u>We would like to note that $P(z \mid x) \geq 0$ is equivalent to $P(x) > 0$.</u>

$P(y \mid x', z)P(x') \geq 0$ is implied by <u>$P(x') = 0 \vee P(z \mid x') > 0$ (i.e., $\mathrm{adj}(z; X)$ except the universal quantifier)</u>.

Post-hoc Analysis

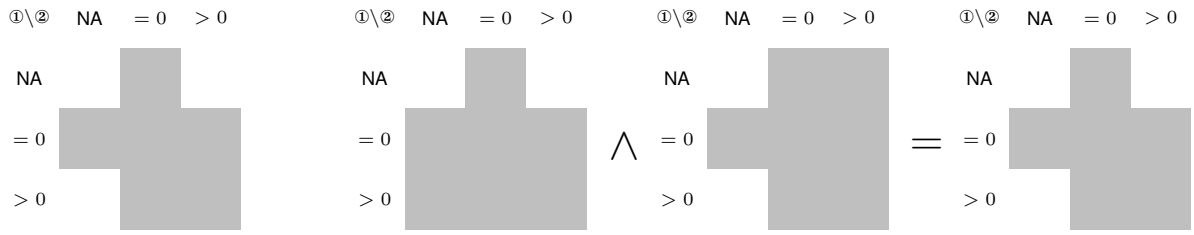
$$\begin{aligned}
P_x(y) \geq 0 &\Leftarrow \sum_z P(z \mid x) \sum_{x'} P(y \mid x', z)P(x') \geq 0 \\
&\Leftarrow \forall z\Big(\overbrace{P(z \mid x)}^{①} \overbrace{\sum_{x'} P(y \mid x', z)P(x')}^{②} \geq 0\Big) \\
&\Leftarrow \forall z\big((① = 0) \vee (② = 0) \vee (① \geq 0 \wedge ② \geq 0)\big).
\end{aligned}$$

The conditions involving ② are

$$\begin{aligned}
&② = 0 \Leftarrow \forall x'(P(y \mid x', z) = 0 \vee P(x') = 0) \\
&② \geq 0 \Leftarrow \mathrm{adj}(z; X).
\end{aligned}$$

Finally,

$$P_x(y) \geq 0 \Leftarrow \forall z\Big(P(z \mid x) = 0 \vee (\forall x'(P(y \mid x', z) = 0 \vee P(x') = 0)) \vee (P(x) > 0 \wedge \mathrm{adj}(z; X)\Big). \tag{13}$$



(a) Post-hoc analysis (13) (b) Do-calculus (14)

Figure 6: Equivalence between post-hoc analysis and do-calculus driven positivity for the front-door graph. Rows represent ① being undefined (NA), equal to zero, and greater than zero and columns are similarly defined for ②. Shaded cells are the condition satisfied by the equations.

Do-Calculus We show do-calculus-based derivation and check the positivity condition for the front-door graph:

$$
\begin{aligned}
P_x(y) &= \sum_z P_x(y \mid z) P_x(z) \\
&= \sum_z P_{xz}(y) P_x(z) \\
&= \sum_z P_z(y) P_x(z) \\
&= \sum_z P_z(y) P(z \mid x) && \text{if } \forall z (P(x) > 0 \vee P_z(y) = 0) \\
&= \sum_z P(z \mid x) \sum_{x'} P_z(y \mid x') P_z(x') \\
&= \sum_z P(z \mid x) \sum_{x'} P_z(y \mid x') P(x') \\
&= \sum_z P(z \mid x) \sum_{x'} P(y \mid x', z) P(x') && \text{if } \forall z (P(z \mid x) = 0 \vee \text{adj}(z; X))
\end{aligned}
$$

This derivation requires only two positivity conditions (with numbers from post-hoc analysis):

- $\forall z (P(x) > 0 \vee P_z(y) = 0)$ (i.e., $\forall z (① \geq 0 \vee ② = 0)$)
- $\forall z (P(z \mid x) = 0 \vee \text{adj}(z; X))$ (i.e., $\forall z (① = 0 \vee ② \geq 0)$

Combining those with replacing $P_z(y)$ with $\sum_{x'} P(y \mid x', z) P(x')$ (which is granted since we are taking the conjunction over the conditions including those allowing $P_z(y) = \sum_{x'} P(y \mid x', z) P(x')$ other than the assumptions encoded in the causal graph $\mathcal{G}$),

$$
\forall z (① \geq 0 \vee ② = 0) \wedge (① = 0 \vee ② \geq 0). \tag{14}
$$

Its equivalence to the post-hoc analysis is illustrated in Fig. 6.

Q-decomposition (manually)

$$
P_x(y) = \sum_z Q[\{Y, Z\}] = \sum_z Q[\{Y\}] Q[\{Z\}] = \sum_z Q[\{Z\}](x, z) \sum_{x'} Q[\{X, Y\}](x', y, z).
$$

Therefore,

$$
P_x(y) \geq 0 \Leftarrow \forall z Q[\{Y, Z\}] \geq 0
$$

$$\Leftarrow \forall z(Q[\{Z\}] = 0 \vee Q[\{Y\}] = 0 \vee (Q[\{Z\}] \geq 0 \wedge Q[\{Y\}] \geq 0))$$
$$\Leftarrow \forall z\Big[\underbrace{Q[\{Z\}] = 0}_{①=0} \vee \underbrace{(\forall x' Q[\{X,Y\}] = 0)}_{②=0} \vee \big(\underbrace{Q[\{Z\}] \geq 0}_{①\geq 0} \wedge \underbrace{(\forall x' Q[\{X,Y\}] \geq 0)}_{②\geq 0}\big)\Big]. \tag{14}$$

One can see the equivalence of Eq. (14) to the condition (13) since

$$Q[\{X,Y\}](x',y,z) = \frac{Q[\{X,Z,Y\}]}{Q[\{X,Z\}]}\frac{Q[\{X\}]}{Q[\emptyset]} = P(y \mid x',z)P(x') \geq 0, \quad \text{if} \quad P(x',z) > 0 \vee P(x') = 0,$$
$$Q[\{X,Y\}](x',y,z) = 0 \quad \text{if} \quad P(x') = 0 \vee P(y \mid x',z) = 0,$$

and

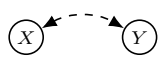
$$Q[\{Z\}](x,z) = \frac{Q[\{X,Z\}]}{Q[\{X\}]} = P(z \mid x) \quad \text{if} \quad P(x) > 0,$$
$$Q[\{Z\}](x,z) = 0 \quad \text{if} \quad P(z \mid x) = 0,$$

by Thm. 5.1, for all z and x'.

Algorithmic Approach Some of the expressions are simplified (e.g., a conditional probability for a fraction) for readability.

- (outer, line 3) initializes E with $P_x(y) \geq 0$.
- (outer, line 4) expands E with $\sum_z Q[\{Z\}]Q[\{Y\}]$ and yields

$$\forall z(Q[\{Z\}] = 0 \vee Q[\{Y\}] = 0 \vee (Q[\{Z\}] \geq 0 \wedge Q[\{Y\}] \geq 0)).$$

- (outer, line 5) case $Q[\{Y\}]$
  - (outer, line 6) $\mathbf{S}_k = \{X, Y\}$
  - (outer, line 7) calls the inner with $\mathbf{C}_j = \{Y\}$, $\mathcal{G}[\mathbf{S}_k]$ 
    - (inner, line 3) the condition is satisfied, $\mathbf{A} = \mathbf{C} = \{Y\}$
    - (inner, line 4) expands with $Q[\{Y\}] = \sum_{x'} Q[X,Y]$, yielding

$$\forall z\Big(Q[\{Z\}] = 0 \vee (\forall x' Q[X,Y] = 0) \vee \big(Q[\{Z\}] \geq 0 \wedge (\forall x' Q[X,Y] \geq 0)\big)\Big).$$

    - (inner, line 5) returns $Q[\{Y\}] = \sum_{x'} Q[X,Y]$.
  - (outer, line 8) expands with $Q[X,Y] = \frac{P(y,z,x')}{P(z,x')}P(x')$ (shortened for readability)

$$\forall z\Big(Q[\{Z\}] = 0 \vee (\forall x'(P(y \mid z,x') = 0) \vee P(x') = 0) \vee \big(Q[\{Z\}] \geq 0 \wedge (\forall x'(P(x') = 0 \vee P(z \mid x') > 0))\big)\Big).$$

  - (outer, line 9) records $Q[\{Y\}] = \sum_{x'} P(y \mid z,x')P(x')$.
- (outer, line 5) case $Q[\{Z\}]$
  - (outer, line 6) $\mathbf{S}_k = \{Z\}$
  - (outer, line 7) skips due to $\mathbf{C}_j = \mathbf{S}_k = \{Z\}$
  - (outer, line 8) expands E with $Q[\{Z\}] = \frac{P(x,z)}{P(x)}$,

$$\forall z\Big(P(z \mid x) = 0 \vee (\forall x'(P(y \mid z,x') = 0) \vee P(x') = 0) \vee \big(P(z \mid x) > 0 \wedge (\forall x'(P(x') = 0 \vee P(z \mid x') > 0))\big)\Big)$$

  - (outer, line 9) records $Q[\{Z\}] = \frac{P(x,z)}{P(z)}$
- (outer, line 11) returns $\sum_z P(x \mid z) \sum_{x'} P(y \mid z,x')P(x')$ and up-to-date E.

C.3. Napkin graph

Post-hoc Analysis The causal formula for $P_x(y)$ is as follows,

$$P_x(y) = \frac{\sum_w P(y, x \mid r, w)P(w)}{\sum_w P(x \mid r, w)P(w)}.$$

Based on the formula, since the causal quantity is a fraction, we should consider the (1) numerator and (2) denominator separately. Especially when the denominator is zero, the causal quantity is undefined. We derive the positive condition as follows:

$$\begin{aligned} P_x(y) \geq 0 &\Leftarrow \exists r \frac{\overbrace{\sum_w P(y, x \mid r, w)P(w)}^{①}}{\underbrace{\sum_w P(x \mid r, w)P(w)}_{②}} \geq 0 \\ &\Leftarrow \exists r (① \geq 0 \wedge ② > 0), \end{aligned}$$

where each condition can be expressed as

$$\begin{aligned} ① \geq 0 &\Leftarrow \mathrm{adj}(r; W), \\ ② > 0 &\Leftarrow \mathrm{adj}(r; W) \wedge \exists w (P(x \mid r, w)P(w) > 0) \\ &\Leftrightarrow \mathrm{adj}(r; W) \wedge \exists w (P(x \mid r, w) > 0 \wedge P(w) > 0) \\ &\Leftrightarrow \mathrm{adj}(r; W) \wedge \exists w (P(x, r, w) > 0) \\ &\Leftrightarrow \mathrm{adj}(r; W) \wedge P(x, r) > 0. \end{aligned}$$

Hence,

$$P_x(y) \geq 0 \Leftarrow \exists r \big(\mathrm{adj}(r; W) \wedge P(x, r) > 0\big). \tag{15}$$

Do-Calculus We start by $P_x(y) = P_{w,r,x}(y)$.

$$\begin{aligned} P_x(y) &= P_{w,r,x}(y) \\ &= P_{w,r}(y \mid x) && \text{if } P_{w,r}(x) > 0 \\ &= \frac{P_{w,r}(y, x)}{P_{w,r}(x)} \\ &= \frac{P_r(y, x)}{P_r(x)} \\ &= \frac{\sum_{w'} P(y, x \mid r, w')P(w')}{\sum_{w'} P(x \mid r, w')P(w')} && \text{if } \mathrm{adj}(r; W). \end{aligned}$$

By combining both conditions with $P_{w,r}(x)$ substituted to the denominator and applying the derivation shown in post-hoc analysis for ② > 0, we obtain

$$\exists r \big(\mathrm{adj}(r; W) \wedge P(x, r) > 0\big). \tag{16}$$

This corresponds to the condition (15) illustrated by only examining the resulting identification formula.

Q-decomposition (manually) We start by $P_x(y) = Q[\{Y\}]$. By Thm. 5.1, for any r,

$$Q[\{Y\}](y, x) = \frac{Q[\{X, Y\}](x, y, r)}{Q[\{X\}](x, r)} \quad \text{if} \qquad Q[\{X\}](x, r) > 0.$$

Therefore,

$$
\begin{aligned}
P_x(y) &= \frac{Q[\{X,Y\}](x,y,r)}{Q[\{X\}](x,r)} \geq 0 \\
&\Leftarrow \exists r(Q[\{X,Y\}](x,y,r) \geq 0 \wedge Q[\{X\}](x,r) > 0).
\end{aligned}
$$

Since $Q[\{X\}](x,r) = \sum_{y'} Q[\{X,Y\}](x,y',r)$, we have

$$
\begin{aligned}
P_x(y) &= \frac{Q[\{X,Y\}](x,y,r)}{\sum_{y'} Q[\{X,Y\}](x,y',r)} \geq 0 \\
&\Leftarrow \exists r(Q[\{X,Y\}](x,y,r) \geq 0 \wedge \forall y' Q[\{X,Y\}](x,y',r) \geq 0 \wedge \exists y' Q[\{X,Y\}](x,y',r) > 0) \\
&\Leftrightarrow \exists r(\forall y' Q[\{X,Y\}] \geq 0 \wedge \exists y' Q[\{X,Y\}] > 0).
\end{aligned}
$$

Since $Q[\{X,Y\}] = \sum_w Q[\{W,X,Y\}]$, we have

$$
\begin{aligned}
P_x(y) &= \frac{\sum_w Q[\{W,X,Y\}](x,y,w,r)}{\sum_{y',w} Q[\{W,X,Y\}](x,y',w,r)} \geq 0 \\
&\Leftarrow \exists r(\forall (y',w)(Q[\{W,X,Y\}] \geq 0) \wedge \exists y'(\forall w Q[\{W,X,Y\}] \geq 0 \wedge \exists w Q[\{W,X,Y\}] > 0)) \\
&\Leftarrow \exists r(\forall (y',w) Q[\{W,X,Y\}] \geq 0 \wedge \exists (y',w) Q[\{W,X,Y\}] > 0).
\end{aligned}
$$

By Thm. 5.1 and the definition of conditional probability,

$$
\begin{aligned}
Q[\{W,X,Y\}] &= \frac{Q[\{W,R,X,Y\}]}{Q[\{W,R\}]} \cdot \frac{Q[\{W\}]}{Q[\emptyset]} = P(y,x \mid w,r)P(w) > 0 \quad \text{if} \quad P(y,x,w,r) > 0, \\
Q[\{W,X,Y\}] &= P(y,x \mid w,r)P(w) \geq 0 \quad \text{if} \quad P(r \mid w) > 0 \vee P(w) = 0.
\end{aligned}
$$

Therefore,

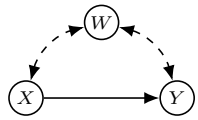
$$
\begin{aligned}
P_x(y) &= \frac{\sum_w P(y,x \mid w,r)P(w)}{\sum_{y',w} P(y',x \mid w,r)P(w)} \geq 0 \\
&\Leftarrow \exists r(\underbrace{\forall (y',w)(P(r \mid w) > 0 \vee P(w) = 0)}_{①} \wedge \underbrace{\exists (y',w)(P(y',x,w,r) > 0)}_{②}).
\end{aligned}
$$

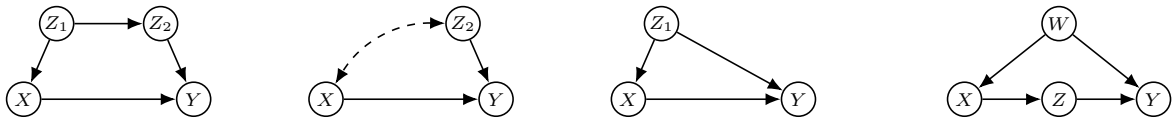
Here, y' can be removed from ①. Also, ② is equivalent to $P(x,r) > 0$. Therefore,

$$
\begin{aligned}
P_x(y) &= \frac{\sum_w P(y,x \mid w,r)P(w)}{\sum_w P(x \mid w,r)P(w)} \geq 0 \\
&\Leftarrow \exists r(\forall w(P(r \mid w) > 0 \vee P(w) = 0) \wedge P(x,r) > 0) \\
&\Leftrightarrow \exists r(\text{adj}(r;W) \wedge P(x,r) > 0).
\end{aligned} \tag{16}
$$

Consequently, the resulting condition is the same as both post-hoc analysis and do-calculus.

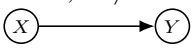
Algorithmic Approach

- (outer, line 3) initializes E with $\exists w, r(P_{w,r,x}(y) \geq 0)$.
- (outer, line 4) expands E simply with $Q[\{Y\}]$, resulting in $Q[\{Y\}] \geq 0$.
- (outer, line 5) case $Q[\{Y\}]$
  - (outer, line 6) $\mathbf{S}_k = \{W, X, Y\}$
  - (outer, line 7) calls with $\mathcal{G}[\mathbf{S}_k]$ 



(a) (b) (c) (d)

Figure 7: (a) a causal diagram where $P_x(y)$ is identified via backdoor adjustment either using Z_1 or Z_2, (b) with Z_1 projected out, (c) with Z_2 projected out, and (d) a causal diagram with an observed mediator and confounder where both back- and front-door criteria are applicable.

  - ∗ (inner, line 3) the condition is not satisfied, $\mathbf{A} = \{X, Y\} \neq \mathbf{C}$.
  - ∗ (inner, line 6) the condition is not satisfied, $\mathbf{A} \neq \mathbf{T}$.
  - ∗ (inner, line 9) $\mathbf{T}' = \{Y\}$ with $\mathcal{G}[\mathbf{A}]$ 
  - ∗ (inner, line 10) is skipped.
  - ∗ (inner, line 11) $f_{Q[\mathbf{T}']|Q[\mathbf{A}]} = \frac{Q[\{Y,X\}]}{\sum_{y'} Q[\{Y,X\}]}$ and $f_{Q[\mathbf{A}]|Q[\mathbf{T}]} = \sum_{w'} Q[\{W, X, Y\}]$. Therefore, $f_{Q[\mathbf{T}']|Q[\mathbf{A}]} \circ f_{Q[\mathbf{A}]|Q[\mathbf{T}]} = \frac{\sum_{w} Q[\{W,X,Y\}]}{\sum_{y',w'} Q[\{W,X,Y\}]}$. Expanded E becomes

$$\exists r, w \frac{\sum_{w'} Q[\{W, X, Y\}]}{\sum_{y',w'} Q[\{W, X, Y\}]} \geq 0.$$

    We can ignore $\exists w$ since w is not involved any term. Then,

$$\exists_r \Big(\sum_{w'} Q[\{W, X, Y\}] \geq 0 \wedge \sum_{y',w'} Q[\{W, X, Y\}] > 0 \Big).$$

  - ∗ (inner, line 12) returns $\frac{\sum_{w'} Q[\{W,X,Y\}]}{\sum_{y',w'} Q[\{W,X,Y\}]}$.
- (outer, line 8) expands $Q[\{W, X, Y\}]$ with $\frac{P(y,x,r,w')}{P(r,w')} P(w') = P(y, x \mid r, w')P(w')$

$$\exists r \Big(\Big(\forall w' (P(y, x|r, w') = 0 \vee P(w') = 0 \vee P(y, x, r, w') > 0) \Big) \wedge \Big(\exists y', w' (P(y', x, r, w') > 0) \Big) \Big),$$

  which can be further reduced to

$$\exists r \Big(\Big(\forall w' (P(r, w') > 0 \vee P(w') = 0) \Big) \wedge \Big(P(x, r) > 0 \Big) \Big).$$

- (outer, line 9) records $Q[\{Y\}] = \frac{\sum_{w'} P(y,x|r,w')P(w')}{\sum_{y',w'} P(y',x|r,w')P(w')}$.

• (outer, line 11) returns $\frac{\sum_{w'} P(y,x|r,w')P(w')}{\sum_{w'} P(x|r,w')P(w')}$ with up-to-date E, that is, $\exists r \, \mathrm{adj}(r; W) \wedge P(x, r) > 0$.

D. On Different Positivity Conditions for Different Identification Formulae

Causal graphs can be simplified by latent projection (Pearl & Verma, 1991; Verma, 1993), which may lead to the simplification of identification formula (Tikka & Karvanen, 2017; 2018). Latent projection induces a causal graph over a subset of variables while maintaining causal relationships among the subset of variables.

D.1. Backdoor Graph with $\{Z_1, Z_2\}$

Through the example illustrated in Fig. 2(a), we would like to argue that there can be multiple sufficient positivity conditions. Consider two identification formulae for $P_x(y)$ in $\mathcal{G}$ illustrated in Fig. 2(a):

$$\sum_{z_1} P(y \mid x, z_1)P(z_1) \quad \text{and} \quad \sum_{z_2} P(y \mid x, z_2)P(z_2),$$

where two different positivity conditions are implied, i.e., $\mathrm{adj}(x; Z_1)$ and $\mathrm{adj}(x; Z_2)$.

$\{Z_2\}$ works as adjustment set but not $\{Z_1\}$. We claim that there exists a distribution where $\text{adj}(x; Z_2)$ holds true but not $\text{adj}(x; Z_1)$ with the following setting. Without loss of generality, we present the parametrization with marginal and conditional distributions. (We denote by $\mathfrak{X}_Z$ represent the state space of Z.)

- Let $P(Z_1, Z_2) = P(Z_2 \mid Z_1)P(Z_1) > 0$ with $\mathfrak{X}_{Z_2} = \{0, 1\}$ and $\mathfrak{X}_{Z_1} = \{0, 1, 2\}$.
- $P(x \mid Z_1 = 2) = 0$ and $P(X \mid Z_1 = z_1) > 0$ for $z_1 \in \{0, 1\}$.

First, this satisfies $\text{adj}(x; Z_2)$ since $P(x \mid z_2) = \sum_{z_1} P(x|z_1)P(z_1|z_2) > 0$ because, for every z_2 value, z_1 value exists (i.e., $\{0, 1\}$ but not 2) such that $P(x \mid z_1) > 0$ and $P(z_1 \mid z_2) > 0$. However, $\text{adj}(x; Z_1)$ does not hold since $P(x \mid Z_1 = 2) = 0$ while $P(Z_1 = 2) > 0$.

$\{Z_1\}$ as an adjustment set implies $\{Z_2\}$ as an adjustment set. However, the other way around is not possible. Assume for contradiction that $\text{adj}(x; Z_2)$ is false, that is, $P(x \mid z_2') = 0 \wedge P(z_2') > 0$ for some z_2', or simply, $P(x \mid z_2') = 0$. Since $P(Z_2) = \sum_{z_1} P(Z_2 \mid z_1)P(z_1)$, for every $P(z_2) > 0$ including z_2', there exists a value of z_1 such that $P(z_2 \mid z_1) > 0$ and $P(z_1) > 0$. Thus, there exists z_1' such that $P(z_1' \mid z_2') > 0$ and $P(z_1') > 0$. Now, $P(x \mid z_2') = \sum_{z_1} P(x \mid z_1)P(z_1 \mid z_2')$ being 0 implies that $P(x \mid z_1) = 0$ for every $P(z_1 \mid z_2') > 0$. Therefore, $P(x \mid z_1') = 0$ and $P(z_1') > 0$, which violates $\text{adj}(x; Z_1)$.

D.2. Backdoor/Front-door Graph

We now consider a causal graph illustrated in Fig. 2(b) where both backdoor and front-door criteria are applicable. Further, IDENTIFY algorithm returns a formula different than those criteria:

$$P_x(y) = \sum_w P(y \mid x, w)p(w) \qquad \text{backdoor} \tag{17}$$

$$P_x(y) = \sum_z P(z \mid x) \sum_{x'} P(y \mid x', z)p(x') \qquad \text{front-door} \tag{18}$$

$$P_x(y) = \sum_z P(z \mid x)P(y \mid z, w) \qquad \text{IDENTIFY} \tag{19}$$

Even though the formulae are different from each other, they produce the same causal quantity under positivity. Then, one might consider the backdoor formula more applicable when specific information about Z is unknown, or vice versa for W.

Front-door but not backdoor We claim that there exists a distribution where the front-door criterion holds true, but the backdoor criterion does not. Consider the following setting:

- Let X and Z be binary and W be ternary.
- Let W be uniformly distributed with $\mathfrak{X}_W = \{0, 1, 2\}$.
- Let $P(x \mid W = 0) = 0$ and $P(x \mid W = w) > 0$ for $w \in \{1, 2\}$.
- We set $P(Z \mid X)$ to be positive.

This satisfies the positivity condition for the front-door criterion since $P(X, Z) > 0$ leads $\forall z(P(z \mid x) > 0 \wedge \text{adj}(z; X))$ to be true (the last term in Eq. (13)). On the other hand, the backdoor criterion does not hold because $P(x \mid W = 0) = 0$ and $P(W = 0) > 0$ violate $\text{adj}(x; W)$.

In terms of Q-decomposition,

$$P_x(y) = \sum_{w,z} Q[\{Y, W, Z\}] = \sum_{w,z} Q[\{Y\}]Q[\{W\}]Q[\{Z\}], \tag{20}$$

where, if valid,

$$Q[\{Y\}] = \frac{Q[\{Y, Z, X, W\}]}{Q[\{Z, X, W\}]}, \tag{21}$$

but both the denominator and numerator are zero when $X = x$ and $W = 0$. With W projected out, X and Y are now confounded. As we have seen before,

$$Q[\{X, Y\}] = \frac{Q[\{Y, Z, X\}]}{Q[\{Z, X\}]} \frac{Q[\{X\}]}{Q[\emptyset]},$$

that is, W is removed from both numerator and denominator with a fraction for X added.

Backdoor but not Front-door We similarly design the model,

- Let $P(X, W) > 0$.
- For $P(Z \mid x) > 0$ but there exists $x' \neq x$ such that $P(z' \mid x') = 0$.

Since $P(x \mid W) > 0$ satisfies $\mathrm{adj}(x; W)$, the backdoor criterion works. However, for the front-door criterion, $P(z' \mid x) > 0$ (that is, ① > 0) and $P(y \mid x', z')$ is undefined due to $P(x', z') = 0$ where $P(x') > 0$ (i.e., ② $\ngeq 0$). Thus, the front-door criterion does not hold.

Again, let's consider Q-decomposition where we focus again on $Q[\{Y\}]$ Eq. (21) in Eq. (20). Now, due to X and Z, both parts of the fraction are zero. With Z projected out,

$$Q[\{Y\}] = \frac{Q[\{Y, X, W\}]}{Q[\{X, W\}]},$$

the denominator becomes positive.

E. Omitted Proofs

Lemma 4.1 (Do-calculus with relaxed positivity)**.** *Let $\mathcal{G}$ be the directed acyclic graph (DAG) associated with a causal model, and let $P(\cdot)$ be the probability distribution induced by the model. Then,*

Rule 1: $P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ *if*
$(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})}$ *and* $P_{\mathbf{x}}(\mathbf{z}, \mathbf{w}) > 0$

Rule 2: $P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{z}, \mathbf{w})$ *if*
$(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})_{\underline{\mathbf{Z}}}}$ *and* $P_{\mathbf{x}}(\mathbf{z}, \mathbf{w}) > 0$

Rule 3: $P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ *if*
$(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})_{\overline{\mathbf{Z}(\mathbf{W})}}}$ *and* $P_{\mathbf{x}}(\mathbf{w}) > 0$. ■

Proof. (Rule 1) By definition, $P_{\mathbf{x}}$ is Markov to $\mathcal{G} \setminus \mathbf{X}$.

(Rule 2) We employ a lifted model and then relate the results in the lifted model to the underlying model. To cover $P_{\mathbf{x},\mathbf{z}}$ and $P_{\mathbf{x}}$ simultaneously, we augment the underlying model with indicators $\mathbf{I} = \{I_Z \mid Z \in \mathbf{Z}\}$ with $I_Z \to Z$ for each $Z \in \mathbf{Z}$. Each indicator can be understood as an independent switch to change the mechanism of Z to implement both original f_Z when $I_Z = 0$ and interventional mechanism z when $I_Z = 1$. We denote by Q the distribution from the model. Then,

$$P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w}, \mathbf{z}) = Q(\mathbf{y} \mid \mathbf{w}, \mathbf{z}, \mathbf{I} = 0),$$
$$P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}, \mathbf{z}) = Q(\mathbf{y} \mid \mathbf{w}, \mathbf{z}, \mathbf{I} = 1).$$

Let us denote by $\mathcal{H}$, the graph $\mathcal{G}$ augmented by $\mathbf{I}$, and we simply set $P(\mathbf{I}) = 0.5^{|\mathbf{Z}|}$. Rule 2 $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})_{\underline{\mathbf{Z}}}}$ is equivalent to $(\mathbf{I} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}, \mathbf{W})_{\mathcal{H}}$ (Pearl, 1995), which licenses $Q(\mathbf{Y} \mid \mathbf{z}, \mathbf{w}) = Q(\mathbf{Y} \mid \mathbf{z}, \mathbf{w}, \mathbf{I})$ under $Q(\mathbf{z}, \mathbf{w}, \mathbf{I}) > 0$, that is $Q(\mathbf{z}, \mathbf{w}) > 0$ since $\mathbf{I}$ is marginally independent to $\{\mathbf{Z}, \mathbf{W}\}$. Thus, $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})_{\underline{\mathbf{Z}}}}$ and $P_{\mathbf{x}}(\mathbf{z}, \mathbf{w}) > 0$ are sufficient.

(Rule 3) We similarly define indicator variables over $\mathbf{Z}$ and an augmented graph $\mathcal{H}$ where $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})_{\overline{\mathbf{Z}(\mathbf{W})}}}$ is equivalent to $(\mathbf{I} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{W})_{\mathcal{H}}$. Thus, $Q(\mathbf{Y} \mid \mathbf{I}, \mathbf{W}) = Q(\mathbf{Y} \mid \mathbf{W})$ under $Q(\mathbf{w}, \mathbf{I}) > 0$. Hence, a sufficient condition to imply $P_{\mathbf{x},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$ is the conjunction of $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W})_{(\mathcal{G} \setminus \mathbf{X})_{\overline{\mathbf{Z}(\mathbf{W})}}}$ and $P_{\mathbf{x}}(\mathbf{w}) > 0$. □

Proposition 4.2. *The conjunction of positivity conditions required in deriving an identification formula for a causal effect applying do-calculus and probability operations sequentially is a sufficient positivity condition for the identification of the causal effect.* ■

Proof. Each sufficient condition is satisfied if the conjunction of sufficient conditions holds true. Hence, we can arrive at the last expression. □

Additionally, if the derivation does not involve any reductions (e.g., merging multiple terms into one by multiplication), then any positivity condition implied in the interventional probability will also be expressed as the given observational distribution (i.e., recursively identified) at the end. Hence, we can obtain a sufficient positivity condition expressed only over the given observational distribution $P(\mathbf{V})$ without a need to separate the identification of the intermediate interventional probabilities.

Corollary E.1. *The conjunction of positivity conditions required in deriving an identification formula for a causal effect applying do-calculus, marginalization, and the definition of conditional probability to factorize sequentially is a sufficient positivity condition over an observational distribution for the identification of the causal effect.* ■

Proposition 5.1. *Under structural causal model and without $P(\mathbf{V})$ being positive, the following holds:*

$$P_{\mathbf{x}}(\mathbf{y}) = \textstyle\sum_{\mathbf{y}^+ \setminus \mathbf{y}} \prod_{\mathbf{C} \in \mathrm{cc}(\mathcal{G}[\mathbf{Y}^+])} Q[\mathbf{C}].$$ ■

Proof. With Rule 3 and the definition of c-factor,

$$\begin{aligned} P_{\mathbf{x}^+}(\mathbf{y}^+) &= P_{\mathbf{x}^+, \mathbf{v} \setminus (\mathbf{X}^+ \cup \mathbf{Y}^+)}(\mathbf{y}^+) = Q[\mathbf{Y}] \\ &= \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}^+} 1_{v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)}. \end{aligned}$$

Based on algebra, we can rearrange the terms as

$$\begin{aligned} Q[\mathbf{Y}] &= \prod_{\mathbf{C} \in \mathrm{cc}(\mathcal{G}[\mathbf{Y}^+])} \sum_{\mathbf{u}(\mathbf{C})} P(\mathbf{u}(\mathbf{C})) \prod_{V_i \in \mathbf{C}} 1_{v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)} \\ &= \prod_{\mathbf{C} \in \mathrm{cc}(\mathcal{G}[\mathbf{Y}^+])} Q[\mathbf{C}] \end{aligned}$$

so that terms related to each c-component are grouped. □

Lemma 5.1. *Given $\mathbf{H} \subseteq \mathbf{V}$, let $\mathbf{H}_1, \ldots, \mathbf{H}_k$ be the c-components of $\mathcal{G}[\mathbf{H}]$ where $\mathbf{H} \subseteq \mathbf{V}$. Then, without $Q[\mathbf{H}]$ being positive, $Q[\mathbf{H}] = \prod_j Q[\mathbf{H}_j]$.* ■

Proof. We start by definition:

$$\begin{aligned} Q[\mathbf{H}](\mathbf{h}, \mathbf{pa}_{\mathbf{h}}) &= \sum_{\mathbf{u}(\mathbf{H})} P(\mathbf{u}(\mathbf{H})) \prod_{V_i \in \mathbf{H}} 1_{v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)} \\ &= \sum_{\mathbf{u}(\mathbf{H})} P(\mathbf{u}(\mathbf{H})) \prod_j \prod_{V_i \in \mathbf{H}_j} 1_{v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)}. \end{aligned}$$

Since, $\mathbf{U}(\mathbf{H}_m) \cap \mathbf{U}(\mathbf{H}_n) = \emptyset$ for any $m \neq n$,

$$\begin{aligned} Q[\mathbf{H}](\mathbf{h}, \mathbf{pa}_{\mathbf{h}}) &= \sum_{\mathbf{u}(\mathbf{H})} P(\mathbf{u}(\mathbf{H})) \prod_j \prod_{V_i \in \mathbf{H}_j} 1_{v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)} \\ &= \sum_{\mathbf{u}(\mathbf{H})} \prod_j P(\mathbf{u}(\mathbf{H}_j)) \prod_{V_i \in \mathbf{H}_j} 1_{v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)} \\ &= \prod_j \sum_{\mathbf{u}(\mathbf{H}_j)} P(\mathbf{u}(\mathbf{H}_j)) \prod_{V_i \in \mathbf{H}_j} 1_{v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)} \end{aligned}$$

$$= \prod_j Q[\mathbf{H}_j](\mathbf{h}_j, \mathbf{pa}_{\mathbf{h}_j}).$$

□

Theorem 5.1. *Given* $\mathbf{H} \subseteq \mathbf{V}$*, let* $\mathbf{H}' \in \mathrm{cc}(\mathcal{G}[\mathbf{H}])$ *where* $I_{\mathcal{G}[\mathbf{H}],\prec}(\mathbf{H}') = \{(l_d, r_d)\}_{d=1}^T$*. Then, the following holds:*

(i) If $Q[\mathbf{H}^{\preceq l_T - 1}] > 0$*,*

$$Q[\mathbf{H}'] = \prod_{d=1}^{T} \frac{Q[\mathbf{H}^{\preceq r_d}]}{Q[\mathbf{H}^{\preceq l_d - 1}]}. \tag{12}$$

(ii) If $Q[\mathbf{H}^{\preceq r_m}] = 0$ *and* $Q[\mathbf{H}^{\preceq l_m - 1}] > 0$ *for some* m*,*

$$Q[\mathbf{H}'] = 0.$$

■

Proof. We first note that Eq. (12) is well-defined under the assumption $Q[\mathbf{H}^{\preceq l_T - 1}] > 0$ since it implies $Q[\mathbf{H}^{\preceq i-1}] > 0$ for all $i \leq l_T - 1$ and thus the denominators are positive.

(i) Suppose $Q[\mathbf{H}^{\preceq l_T - 1}] > 0$. Since (ii) covers the case of $Q[\mathbf{H}^{\preceq r_T}] = 0$, we now suppose $Q[\mathbf{H}^{\preceq r_T}] > 0$. We show $\mathbf{H}' \in \mathrm{cc}(\mathcal{G}[\mathbf{H}^{\preceq r_T}])$ since $\mathbf{H}' \subseteq \mathbf{H}^{\preceq r_T} \subseteq \mathbf{H}$ and $\mathbf{H}' \in \mathrm{cc}(\mathcal{G}[\mathbf{H}])$. Invoking Lem. 5.2,

$$Q[\mathbf{H}'] = \prod_{V^{(i)} \in \mathbf{H}'} \frac{Q[\mathbf{H}^{\preceq i}]}{Q[\mathbf{H}^{\preceq i-1}]},$$

where, for each $1 \leq d \leq T$ and $i \in [l_d + 1, r_d]$, we cancel out the terms $\frac{\cdot}{Q[\mathbf{H}^{\preceq i-1}]} \frac{Q[\mathbf{H}^{\preceq i-1}]}{\cdot}$ in the telescoping product

$$= \prod_{d=1}^{T} \frac{Q[\mathbf{H}^{\preceq r_d}]}{Q[\mathbf{H}^{\preceq l_d - 1}]} > 0,$$

since $V^{(i)} \in \mathbf{H}'$ for $i \in [l_d, r_d]$.

(ii) Suppose $Q[\mathbf{H}^{\preceq r_m}] = 0$ and $Q[\mathbf{H}^{\preceq l_m - 1}] > 0$ for some m. Then, there exists $t \in [l_m, r_m]$ such that $Q[\mathbf{H}^{\preceq t}] = 0$ and $Q[\mathbf{H}^{\preceq t-1}] > 0$ (Prop. B.2). Note that $V^{(t)} \in \mathbf{H}'$. Now, we have $Q[\mathbf{H}_j] > 0$ for all $\mathbf{H}_j \in \mathrm{cc}(\mathcal{G}[\mathbf{H}^{\preceq t-1}])$ since $\mathbf{H}_j \subseteq \mathbf{H}^{\preceq t-1}$ (Prop. B.1). Now, let $\mathbf{H}^* \in \mathrm{cc}(\mathcal{G}[\mathbf{H}^{\preceq t}])$ where $V^{(t)} \in \mathbf{H}^*$. Applying Lem. 5.1 for $Q[\mathbf{H}^{\preceq t}]$,

$$0 = Q[\mathbf{H}^{\preceq t}] = Q[\mathbf{H}^*] \times \prod_{\mathbf{H}'' \in \mathrm{cc}(\mathcal{G}[\mathbf{H}^{\preceq t}]) \setminus \{\mathbf{H}^*\}} Q[\mathbf{H}''] \tag{22}$$

For all $\mathbf{H}'' \in \mathrm{cc}(\mathcal{G}[\mathbf{H}^{\preceq \mathrm{t}}]) \setminus \{\mathbf{H}^*\}$, $\mathbf{H}''$ is a c-component of $\mathcal{G}[\mathbf{H}^{\preceq t-1}]$ and thus $Q[\mathbf{H}''] > 0$. Therefore, by Eq. (22), we have $Q[\mathbf{H}^*] = 0$.

Given $Q[\mathbf{H}^*] = 0$, we finally show $Q[\mathbf{H}'] = 0$ by invoking Corollary B.1 with $\mathbf{H}^* \subseteq \mathbf{H}'$, which is due to $V^{(t)} \in \mathbf{H}^* \in \mathrm{cc}(\mathcal{G}[\mathbf{H}^{\preceq t}])$ and $V^{(t)} \in \mathbf{H}' \in \mathrm{cc}(\mathcal{G}[\mathbf{H}])$. □

Proposition 6.1. *Consider a c-factor and its expression induced by marginalization, Lem. 5.1, or Thm. 5.1 for the c-factor. Then, a constraint over the expression of the c-factor is implied by a statement of first-order logic as follows:*

$$\begin{aligned}
\textstyle\sum_{\mathbf{z}} g(\mathbf{z}) \geq 0 &\Longleftarrow \forall \mathbf{z}(g(\mathbf{z}) \geq 0),\\
\textstyle\sum_{\mathbf{z}} g(\mathbf{z}) > 0 &\Longleftarrow (\forall \mathbf{z}(g(\mathbf{z}) \geq 0)) \wedge (\exists \mathbf{z}(g(\mathbf{z}) > 0)),\\
\textstyle\sum_{\mathbf{z}} g(\mathbf{z}) = 0 &\Longleftarrow \forall \mathbf{z}(g(\mathbf{z}) = 0),\\
\textstyle\prod_i g_i \geq 0 &\Longleftarrow (\exists i(g_i = 0)) \vee (\forall i(g_i \geq 0)),\\
\textstyle\prod_i g_i > 0 &\Longleftarrow \forall i(g_i > 0),\\
\textstyle\prod_i g_i = 0 &\Longleftarrow \exists i(g_i = 0),\\
g_1/g_2 \geq 0 &\Longleftarrow g_1 \geq 0 \wedge g_2 > 0,
\end{aligned}$$

$$g_1/g_2 > 0 \impliedby g_1 > 0 \wedge g_2 > 0,$$
$$g_1/g_2 = 0 \impliedby g_1 = 0 \wedge g_2 > 0.$$

■

Proof. The result follows from basic algebra if there exist no undefined terms. In case of possible undefined terms, we only consider the two cases where we treat multiplication by zero as zero, namely, $\prod_i g_i \geq 0$ and $\prod_i g_i = 0$. By relaxed generalized Q-decomposition, whenever we have an undefined term (divide by zero), it is covered by Thm. 5.1-(ii). Since Thm. 5.1 generalizes Lem. 5.2 (with anchors minimally consecutive, i.e., all separated), this also covers the equation, output by IDENTIFY. □

Note that $g_1/g_2 > 0$ can be simply implied by $g_1 > 0$ under generalized Q-decomposition since the denominator is the marginalization of the numerator. Further, for each i, $g_i > 0$ and $g_i \geq 0$ are interchangeable in $(\exists i\,(g_i = 0)) \vee (\forall i\,(g_i \geq 0))$.

Theorem 6.1. *Whenever* IDENTIFY+ *returns a positivity condition and identification formula, they are correct.* ■

Proof. The soundness of the identification formula comes from the correctness of Thm. 5.1, which generalizes Lem. 5.2, and IDENTIFY algorithm, which provides the basis for IDENTIFY+. The soundness of the positivity condition is due to Props. 5.1 and 6.1. □

F. Further Discussion

We discuss the generalization to continuous variables. The use of summation $\sum$ in identification formulas is for the convention, and such formulas are valid for the continuous case with integral $\int$ (Shpitser & Pearl, 2006a; Tian & Pearl, 2003). However, challenges remain for the positivity condition since it involves probability density functions and measure-zero sets, which complicates its analysis, especially when fractions are involved. For the case of Napkin (Fig. 4), ensuring the denominator positive with $\exists w(P(x|r,w)P(w) > 0)$ may be translated to $\exists \mathcal{W}' \int_{\mathcal{W}'} f(x|r,w)\mathrm{d}F(w) > 0$ where $\mathcal{W}'$ is a subset of the domain of W. While we believe such translation is sufficient for imposing constraints over point estimates for the continuous case, we defer its formal treatments and deeper discussion to future work.

Here, we compare our result to the relaxed positivity mentioned by Shpitser & Pearl (2006a). The authors relaxed $P(\mathbf{V}) > 0$ and, without a concrete proof, relied on $P(\mathbf{Z}) > 0$ whenever $P(\mathbf{W} \mid \mathbf{Z})$ is considered in the identification. However, this yields a stricter positivity condition, e.g., $P(x, Z) > 0$ versus $\mathrm{adj}(x; Z)$ for adjustment criterion. In this work, we explicitly considered canceling out terms, avoiding checking positivity for unnecessary c-factors. Further, we have incorporated the case where such positivity is violated yet the identification is feasible (Thm. 5.1, ii). Hence, we further advance the understanding of the formal connection between positivity conditions and causal effect identification.