TECHNICAL REPORT
R-70
Feb 2021

Causal Identification with Matrix Equations

Sanghack Lee [1] **Elias Bareinboim** [1]

Abstract

Causal effect identification is concerned with determining whether a causal effect is computable from a combination of qualitative assumptions about the underlying system (e.g., a causal graph) and distributions collected from this system. Many identification algorithms typically rely on graphical criteria made of a non-trivial combination of probability axioms, do-calculus, and refined c-component factorization (e.g., Lee & Bareinboim, 2020). In a sequence of increasingly sophisticated results, it has been shown how *proxy* variables can be used to identify certain effects that would not be otherwise recoverable in challenging scenarios through solving matrix equations (e.g., Kuroki & Pearl, 2014; Miao et al., 2018). In this paper, we establish a connection between graphical criteria and matrix equations from first principles, and develop new identification conditions and algorithms. Specifically, we characterize the relationships between certain identifiable formulae and matrix multiplications. Second, we devise a general identification condition for proxy variables, which subsumes existing methods. Further, we propose a novel *intermediary* criteria based on the idea of pseudoinverse of a matrix. Finally, we provide an in-depth discussion on how to incorporate matrix-based methods into factorization-based identification approaches.

1. Introduction

One of the important tasks in data-intensive disciplines (artificial intelligence) and empirical sciences is to uncover cause and effect relationships. Once causal relationships are well-understood and a collection of distributions are available, one may be able to estimate causal effects without performing experiments (Pearl, 2000; Spirtes et al., 2000).

[1]Causal Artificial Intelligence Laboratory, Columbia University, New York, NY, USA. Correspondence to: Sanghack Lee <sanghacklee@cs.columbia.edu>.

A causal effect of a set of variables $\mathbf{X}$ being held fixed to $\mathbf{x}$ on another set of variables $\mathbf{Y}$ is denoted by $P(\mathbf{y}|do(\mathbf{x}))$ or using a subscript $P_{\mathbf{x}}(\mathbf{y})$. Earlier results focused on controlling for a confounding bias (Fig. 1a) by adjusting a back-door admissible set,

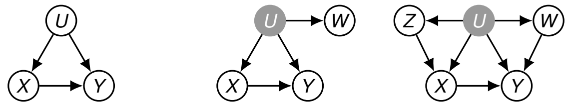
$$P_x(y) = \sum_u P(y|x, u)P(u), \quad (1)$$

where U blocks paths that leave behind X to Y (for a detailed discussion please refer (Pearl, 2000, Sec. 3.3.1)). Necessary and sufficient conditions have been developed for identifying a causal effect given a causal graph $\mathcal{G}$ over endogenous variables $\mathbf{V}$ and observational data $P(\mathbf{V})$ (Pearl, 1995; Tian & Pearl, 2002; Shpitser & Pearl, 2006; Huang & Valtorta, 2006). With growing interest in a data fusion framework (Bareinboim & Pearl, 2016), Lee et al. (2019) devised a sound and complete algorithm for the *general identifiability* problem (GID) that identifies a causal effect given an arbitrary collection of distributions under different experimental conditions. The key ideas of the algorithm are (i) the decomposition of a given causal query into the *factors* (*c-factorization*, (Tian & Pearl, 2002)) relative to a causal graph and the query, and (ii) the identification of each factor individually by one of the available distributions.

The aforementioned problems (ID and GID) relied on an assumption that each dataset contains measurements over *all* the variables modeled in a causal graph. Relaxing the assumption, researchers have investigated the problem of identification under *partial-observability*, which makes use of *marginal* distributions (e.g., $P_{\mathbf{Z}}(\mathbf{W})$ where $\mathbf{Z} \cup \mathbf{W} \subsetneq \mathbf{V}$).[1] A recent characterization (Lee & Bareinboim, 2020) is that the c-factorization needs to be applied at a certain abstraction of a causal graph (i.e., latent projection). On another note, distributions may be available in a *conditional* form, e.g., $P_{\mathbf{Z}}(\mathbf{W}|\mathbf{T})$. When dealing with marginal or conditional distributions, traditional factorization-based approaches are *insufficient*. We will describe in the sequel methods capable of dealing with such challenges.

Identification with Proxies Greenland & Lash (2008); Kuroki & Pearl (2014); Miao et al. (2018); Wang & Blei (2019) developed identification methods when the back-

[1]Strictly speaking, $P_{\mathbf{Z}}(\mathbf{W})$ is a collection of interventional distributions $\{P_{\mathbf{z}}(\mathbf{W})\}_{\mathbf{z}}$ for all $\mathbf{z}$ values. Without loss of generality, we refer $P_{\mathbf{Z}}(\mathbf{W})$ a distribution.



(a) (b) (c)

Figure 1: Causal graphs: (a) with a back-door condition; (b) with W as a proxy for U; (c) with proxies W and Z for U.

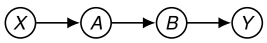


Figure 2: A causal diagram demonstrating the identification of $P_x(y) = P(y|x)$ given $P(B|x)$, $P(B|A)$, and $P(y|A)$.

door admissible set is unmeasured, thus, cannot be directly controlled for. Given that U is unobserved, the query $P_x(y)$ is typically not identifiable. Greenland & Lash (2008) considered a case (Fig. 1b), where W is a *proxy* for the unmeasured confounder, and both distributions $P(W, X, Y)$ and $P(W|U)$ are available. Let $\mathsf{P}(W|U)$ be a matrix representation of $P(W|U)$ where $\mathsf{P}(W|U)_{i,j} = P(w_i|u_j)$ (note the difference in font family). When $\mathsf{P}(W|U)$ is invertible, $\mathsf{P}(y|U, x)$ becomes recoverable together with $\mathsf{P}(U)$, which we will elaborate later. Hence, the effect is identifiable by plugging in these quantities into $P_x(y) = \mathsf{P}(y|U, x)\mathsf{P}(U)$ (rewritten Eq. (1)).

Further, Kuroki & Pearl (2014) and Miao et al. (2018) considered the case of a pair of *proxy variables* depicted in Fig. 1c. Given an observational distribution $P(W, X, Y, Z)$, the causal effect is identified as

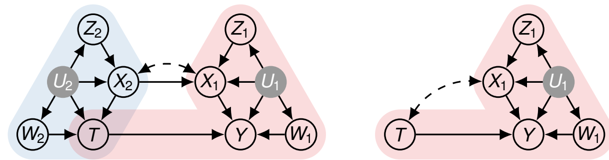
$$P_x(y) = \mathsf{P}(y|Z, x)\mathsf{P}(W|Z, x)^{-1}\mathsf{P}(W), \tag{2}$$

where $\mathsf{P}(W|Z, x)$ is *invertible*.[2] We refer this identification condition *MGT criterion* where the acronym MGT comes from the surnames of the authors in (Miao et al., 2018). Although typically not framed in terms of identification with multiple datasets, this setting corresponds to identifying $P_x(y)$ with, e.g., $\{P(X, Y, Z), P(W, X, Z)\}$ or, simply $\{P(y|Z, x), P(W|Z, x), P(W)\}$ as appeared in Eq. (2).

Identification with Intermediaries In addition to the proxy methods, we will consider novel identification opportunities using *intermediary* distributions, which will serve as a link between two other distributions. For concreteness, consider the causal graph in Fig. 2. A causal effect $P_x(y)$ can be expressed as

$$\begin{aligned} P_x(y) = P(y|x) &= \sum_{a,b} P(y|b)P(b|a)P(a|x) \\ &= \mathsf{P}(y|B)\mathsf{P}(B|A)\mathsf{P}(A|x). \end{aligned} \tag{3}$$

[2]In this paper, we focus on discrete variables. Detailed discussion on a continuous case can be found in Miao et al. (2018).



(a) $\mathcal{G}$ (b) projected

Figure 3: A causal graph $\mathcal{G}$ where a causal effect $P_{\mathbf{x}}(y)$ can be computed by estimating each of $P_{x_2}(t)$ (light blue, left) and $P_{t,x_1}(t)$ (light red, right) using MGT criterion.

Given three distributions $P(y|A)$, $P(B|A)$, and $P(B|x)$, $P_x(y)$ is identifiable under the invertibility of the *intermediary* distribution $\mathsf{P}(B|A)$ (since $P(A|x)$ and $P(y|B)$ are recovered). Surprisingly, such invertibility is *not* necessary. Let $\mathsf{P}(B|A)^\dagger$ be the *pseudoinverse* of $\mathsf{P}(B|A)$, which always exists and is unique. It satisfies $\mathsf{P}(B|A) = \mathsf{P}(B|A)\mathsf{P}(B|A)^\dagger\mathsf{P}(B|A)$. By plugging in the expression into Eq. (3),

$$\begin{aligned} P_x(y) &= \mathsf{P}(y|B)\mathsf{P}(B|A)\mathsf{P}(B|A)^\dagger\mathsf{P}(B|A)\mathsf{P}(A|x) \\ &= \mathsf{P}(y|A)\mathsf{P}(B|A)^\dagger\mathsf{P}(B|x). \end{aligned}$$

To the best of our knowledge, this is the first time the pseudoinverse of a matrix is used within the context of causal effect identification.[3]

A Motivating Example Let us demonstrate the opportunities of incorporating identification methods based on *generalized inverse* of matrix (i.e., both typical inverse and pseudoinverse) inside a matrix equation for a probability or distribution into factorization-based identifiability approaches. Observe the causal graph in Fig. 3. Neither factorization nor proxy-based approaches are helpful here in answering the query $P_{\mathbf{x}}(y)$. It is only once we combine the two ideas by expressing the equation as two factors that a solution becomes apparent:

$$P_{\mathbf{x}}(y) = \sum_t P_{\mathbf{x}}(t, y) = \sum_t P_{t,x_1}(y)P_{x_2}(t),$$

Each factor can then be identified based on MGT criterion treating T and X_1 in the first factor as a single variable.

Contributions Identification problems have evolved to accommodate increasingly general collections of distributions. This paper offers a unifying framework for the problem, with the following main contributions: (i) We characterize matrix equations of probability distributions driven by graphical constraints in a causal graph to improve the understanding

[3]In case of continuous variables, $\mathsf{P}(B|A)$ can be understood as a linear operator on Hilbert spaces. See (Nashed, 1971) for more detailed discussion on the existence and uniqueness of generalized inverse (pseudoinverse) for a bounded linear operator.

of the identification through solving a system of equations. (ii) Building on this new characterization, we generalize proxy-based criteria and devise intermediary criteria so as to identify a causal effect by utilizing the (pseudo)inverse of a matrix within a matrix equation. (iii) We provide an in-depth discussion on how to exploit matrix equations and (pseudo)inverses so as to combine existing identification algorithms to output an identification formula for a causal query given a causal diagram and a set of marginal, experimental, and conditional distributions.

2. Preliminaries

We follow notational conventions from literature on causal inference. We denote a variable by an upper case letter Y, and its value is denoted by its corresponding lower case letter y in the domain $\mathfrak{X}_Y$. A set of variables will be denoted by a bold capital letter $\mathbf{Y}$ with its value $\mathbf{y}$. We may use $\dot\cup$, instead of $\cup$, to emphasize the union of two *disjoint* sets. Given $\mathbf{Z} \subseteq \mathbf{W}$, $\mathbf{w}\backslash\mathbf{Z}$ denotes the value of $\mathbf{W}\backslash\mathbf{Z}$ consistent with $\mathbf{w}$. Without loss of generality, we refer $P_{\mathbf{Z}}(\mathbf{V}'|\mathbf{W})$ a distribution. We may employ *conditional* to emphasize $\mathbf{W} \supsetneq \emptyset$, *experimental* or *interventional* $\mathbf{Z} \supsetneq \emptyset$ compared to *observational* $\mathbf{Z} = \emptyset$, and *marginal* if $\mathbf{Z} \cup \mathbf{W} \cup \mathbf{V}' \subsetneq \mathbf{V}$.

Structural Causal Model We employ structural causal models (SCMs) (Pearl, 2000, Ch. 7) as the semantical framework to represent a domain of interest. An SCM $\mathcal{M}$ is a quadruple $\langle \mathbf{U}, \mathbf{V}, P(\mathbf{U}), \mathbf{F}\rangle$. A set of exogenous variables $\mathbf{U}$ is determined by factors outside the model. It follows a joint distribution $P(\mathbf{U})$. $\mathbf{V}$ is a set of endogenous variables whose values are determined by functions $\mathbf{F} = \{f_i\}_{V_i \in \mathbf{V}}$ such that $V_i \leftarrow f_i(\mathbf{pa}_i, \mathbf{u}_i)$ where $\mathbf{PA}_i \subseteq \mathbf{V}\backslash\{V_i\}$ and $\mathbf{U}_i \subseteq \mathbf{U}$. Further, $do(\mathbf{X} = \mathbf{x}) = do(\mathbf{x})$ represents the operation of holding a set $\mathbf{X}$ to a constant $\mathbf{x}$ regardless of their original mechanisms. Such intervention induces a submodel $\mathcal{M}_{\mathbf{x}}$, which is $\mathcal{M}$ with f_X replaced to x for $X \in \mathbf{X}$. The distribution over $\mathbf{V}$ induced by the submodel is denoted by $P(\mathbf{V}|do(\mathbf{x})) = P_{\mathbf{x}}(\mathbf{V})$. We may employ letter Q to denote an interventional distribution, e.g., $Q = P_{\mathbf{r}}$.

Each SCM (model, for short) induces a causal diagram (or causal graph) $\mathcal{G} = \langle \mathbf{V}, \mathbf{E}\rangle$, where each type of edge represents a different causal relationship among the variables: (i) $X \rightarrow Y$ if X is used as an argument of f_Y; and (ii) $X \leftrightarrow Y$ if $\mathbf{U}_X$ and $\mathbf{U}_Y$ are correlated.[4] Given a causal diagram $\mathcal{G}$, familial relationships among its vertices are denoted by pa and an for parents and ancestors, respectively. Further, An is a set of ancestors including its argument as well. We denote by $\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ an edge subgraph of $\mathcal{G}$ which removes edges incoming to $\mathbf{X}$ and outgoing from $\mathbf{Z}$. A submodel $\mathcal{M}_{\mathbf{x}}$ can be presented as $\mathcal{G}_{\overline{\mathbf{X}}}$ with $\mathbf{X}$ fixed to $\mathbf{x}$.

[4] More precisely, any set of variables $\mathbf{W}$ lacking bidirected edges among them satisfies that their exogenous parents should be jointly independent.

Causal relationships among other variables are captured in $\mathcal{G}\backslash\mathbf{X}$, which is the subgraph of $\mathcal{G}$ over $\mathbf{V}\backslash\mathbf{X}$. A vertex induced subgraph is denoted by $\mathcal{G}[\mathbf{V}']$ where $\mathbf{V}' \subseteq \mathbf{V}$. Causal effect identification relies heavily on standard graphical constraints imposed by a causal diagram such as d-separation (reading off conditional independence from the graph, Verma & Pearl, 1988; Geiger et al., 1990) and do-calculus (equivalence among interventional probabilities, Pearl, 1995). For completeness, we include d-separation and the rules of do-calculus in the Appendix.

The *latent projection* (or projection, for short) of a causal diagram is a causal diagram retaining the causal relationships among a subset of variables. We denote by $\mathcal{G}\langle\mathbf{V}'\rangle$ the latent projection of $\mathcal{G}$ onto $\mathbf{V}' \subseteq \mathbf{V}$, the causal graph over $\mathbf{V}'$ (Verma & Pearl, 1990). Conditional independence (CI) statements and do-calculus (Pearl, 1995) on a projection are valid in $\mathcal{G}$, vice versa. We formally define a latent projection in the supplementary material. The omitted proofs and derivations are also provided in the Appendix.

3. Characterization of Matrix Equations of Graphical Criteria

In this section, we present characterizations of graphical constraints underlying a given causal diagram $\mathcal{G}$, leading to equations expressed as the multiplication of matrices. The characterizations will further advance our understanding on the constraints imposed over the distributions generated by the underlying system compared to simple equivalence relationships such as conditional independence and do-calculus.

To begin with, we denote by $\mathsf{P}(A|B)$ a $|\mathfrak{X}_A| \times |\mathfrak{X}_B|$ matrix whose element at (i, j) corresponds to $P(a_i|b_j)$ with $\mathfrak{X}_A = \{a_1, \ldots, a_n\}$ and $\mathfrak{X}_B = \{b_1, \ldots, b_m\}$. Similarly, $\mathsf{P}_{\mathbf{C}',\mathbf{c}''}(\mathbf{A}', \mathbf{a}''|\mathbf{B}', \mathbf{b}'')$ is a $|\mathfrak{X}_{\mathbf{A}'}| \times |\mathfrak{X}_{\mathbf{B}' \times \mathbf{C}'}|$ matrix, which is a submatrix of $\mathsf{P}_{\mathbf{C}}(\mathbf{A}|\mathbf{B})$ by selecting the rows and columns corresponding to the constants $\mathbf{a}'', \mathbf{b}'', \mathbf{c}''$ where $\mathbf{W} = \mathbf{W}' \dot\cup \mathbf{W}''$ ($\mathbf{W} \in \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$).

Chain Rule and Conditional Independence The definitions of conditional and marginal distributions naturally lead to a sum of a product of probabilities. Let $Q = P_{\mathbf{r}}$ be an arbitrary interventional distribution. Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{R}$ be disjoint. A marginal probability over chain rule induced multiplication is expressed as

$$\begin{aligned} Q(\mathbf{a}, \mathbf{b}'|\mathbf{c}) &= \textstyle\sum_{\mathbf{b}''} Q(\mathbf{a}|\mathbf{b}, \mathbf{c}) Q(\mathbf{b}|\mathbf{c}) \\ &= \mathsf{Q}(\mathbf{a}|\mathbf{b}', \mathbf{B}'', \mathbf{c}) \mathsf{Q}(\mathbf{B}'', \mathbf{b}'|\mathbf{c}). \end{aligned}$$

Hence, it is $\mathsf{Q}(\mathbf{A}, \mathbf{b}'|\mathbf{c}) = \mathsf{Q}(\mathbf{A}|\mathbf{b}', \mathbf{B}'', \mathbf{c})\mathsf{Q}(\mathbf{B}'', \mathbf{b}'|\mathbf{c})$. However, one cannot simply change $\mathbf{c}$ to $\mathbf{C}$ in this expression. Considering conditional independence, we can further enrich such a characterization.

Lemma 1. *Given a causal diagram $\mathcal{G}$, let $Q = P_{\mathbf{r}}$ for some $\mathbf{r} \in \mathfrak{X}_{\mathbf{R}}$ where $\mathbf{R} \subsetneq \mathbf{V}$. Let $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$ be disjoint subsets of $\mathbf{V} \backslash \mathbf{R}$. With $(\mathbf{D} \perp\!\!\!\perp \mathbf{A} \mid \mathbf{B}, \mathbf{C}, \mathbf{E})$ and $(\mathbf{E} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C}, \mathbf{D})$ in $\mathcal{G} \backslash \mathbf{R}$, the following equality holds:*

$$\mathsf{Q}(\mathbf{A}, \mathbf{b}' | \mathbf{c}, \mathbf{D}, \mathbf{e}) = \mathsf{Q}(\mathbf{A} | \mathbf{b}', \mathbf{B}'', \mathbf{c}, \mathbf{e}) \mathsf{Q}(\mathbf{B}'', \mathbf{b}' | \mathbf{c}, \mathbf{D}).$$

The lemma implies that the result *is* a matrix *not* a row or column. This makes MGT criterion possible through inverting the resulting matrix so as to cancel out an unknown distribution $\mathsf{P}(W|U)$ (we will formally revisit this in Sec. 5.2).

Adjustment Criterion Given a graph $\mathcal{G}$ and a causal effect of interest $P_{\mathbf{x}}(\mathbf{y})$, the adjustment criterion (Shpitser et al., 2010) seeks a set of covariates $\mathbf{Z} \subseteq \mathbf{V} \backslash \mathbf{X} \backslash \mathbf{Y}$, called an *adjustment set* for a causal effect $P_{\mathbf{x}}(\mathbf{y})$, which grants the following expression, $P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{z}} P(\mathbf{y}|\mathbf{x}, \mathbf{z}) P(\mathbf{z})$. Adjustment criterion generalizes back-door criterion (Pearl, 2000). Its matricized expression with employing $Q = P_{\mathbf{r}}$ is

$$Q_{\mathbf{x}}(\mathbf{y}) = \textstyle\sum_{\mathbf{z}} Q(\mathbf{y}|\mathbf{x}, \mathbf{z}) Q(\mathbf{z}) = \mathsf{Q}(\mathbf{y}|\mathbf{x}, \mathbf{Z}) \mathsf{Q}(\mathbf{Z}). \quad (4)$$

This simple expression plays a central role in the identification with proxy variables.

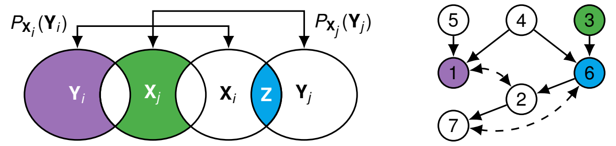
In many settings, the left hand side (LHS) is the query of interest and two terms in the RHS are usually available or to be inferred using other available quantities. However, substituting value $\mathbf{y}$ with $\mathbf{Y}$, we can further yield (under invertibility assumption) $\mathsf{Q}(\mathbf{Z}) = \mathsf{Q}(\mathbf{Y}|\mathbf{x}, \mathbf{Z})^{-1} \mathsf{Q}_{\mathbf{x}}(\mathbf{Y})$, which restores the covariate distribution of interest given a causal effect and conditional distribution.

In the case of $\mathbf{X} = \emptyset$, any $\mathbf{Z}$ disjoint to $\mathbf{Y}$ becomes an adjustment set, bridging the adjustment criterion and chain rule, i.e., $\mathsf{Q}(\mathbf{Y}) = \mathsf{Q}(\mathbf{Y}|\mathbf{Z}) \mathsf{Q}(\mathbf{Z})$.

C-Factorization C-factorization (Tian, 2002) decomposes a causal effect $P_{\mathbf{x}}(\mathbf{y})$ into a sum-product of c-factors (simply, factors) with respect to the given causal diagram $\mathcal{G}$. Without loss of generality, let $\mathbf{X}$ be minimal such that no $\mathbf{X}' \subsetneq \mathbf{X}$ satisfies $P_{\mathbf{x}'}(\mathbf{y}) \neq P_{\mathbf{x}}(\mathbf{y})$ (i.e., overriding $\mathbf{X}$ by $\mathrm{an}^{\mathcal{G}_{\overline{\mathbf{X}}}}(\mathbf{Y}) \cap \mathbf{X}$). For any projection $\mathcal{H}$ of $\mathcal{G}$ that preserves $\mathbf{X} \cup \mathbf{Y}$, the following holds.

$$\begin{aligned} P_{\mathbf{x}}(\mathbf{y}) &= \textstyle\sum_{\mathbf{y}^+ \backslash \mathbf{y}} P_{\mathbf{x}}(\mathbf{y}^+) \\ &= \textstyle\sum_{\mathbf{y}^+ \backslash \mathbf{y}} \prod_{\mathbf{Y}_i \in \mathcal{C}(\mathcal{H}[\mathbf{Y}^+])} P_{\mathrm{pa}^{\mathcal{H}}(\mathbf{y}_i) \backslash \mathbf{y}_i}(\mathbf{y}_i) \end{aligned} \quad (5)$$

where $\mathbf{Y}^+ = \mathrm{An}^{\mathcal{H}_{\underline{\mathbf{X}}}}(\mathbf{Y})$ and $\mathcal{C}(\cdot)$ is the c-component decomposition (partitioning the variables in the graph based on their connectivity through bidirected edges). Given a query $P_{\mathbf{x}}(\mathbf{y})$, we let c-factors $\mathbb{F}_{\mathcal{H}} = \{\langle \mathbf{X}_i, \mathbf{Y}_i \rangle\}_i$ where $\{\mathbf{Y}_i\}_i$ form the c-component decomposition of $\mathcal{H}[\mathbf{Y}^+]$ and $\mathbf{X}_i = \mathrm{pa}^{\mathcal{H}}(\mathbf{Y}_i) \backslash \mathbf{Y}_i$. The identification method (Lee & Bareinboim, 2020) can be summarily described as finding $\mathcal{H}$ such



(a) Euler diagram (b) $\mathcal{G}$

Figure 4: (a) set relationships among $\mathbf{X}$ and $\mathbf{Y}$ in two factors $\mathbf{Y}_i \cap \mathbf{Y}_j = \emptyset$, $\mathbf{X}_i \cap \mathbf{Y}_i = \emptyset$, and $\mathbf{X}_j \cap \mathbf{Y}_j = \emptyset$. (b) a causal graph with $P_{3,4,5}(1,2,7) = \sum_6 P_{4,5,6}(1,2) P_{2,3,4}(6,7)$ where each number corresponds to the section of (a) from the left (i.e., $\mathbf{Y}_i \backslash \mathbf{X}_j$) to right (i.e., $\mathbf{Y}_j \backslash \mathbf{Z}$).

that each factor $P_{\mathbf{x}_i}(\mathbf{y}_i)$ is identified by one of the available distributions. For instance, Fig. 2 yields $\mathbf{Y}^+ = \{Y, A, B\}$ with three c-components $\{\{Y\}, \{A\}, \{B\}\}$ so that

$$P_x(y) = \textstyle\sum_{a,b} P_x(y, a, b) = \sum_{a,b} P_b(y) P_a(b) P_x(a)$$

where the interventional part of each factor can be exchanged with conditionals (Rule 2 of do-calculus).

We focus here on the sum-product of *two* factors. A pair of factors at some latent projection $\mathcal{G}'$ satisfies the following, while abusing notation $\mathbf{X}_{ij} = \mathbf{X}_i \cup \mathbf{X}_j$,

$$P_{\mathbf{x}_i}(\mathbf{y}_i) P_{\mathbf{x}_j}(\mathbf{y}_j) = P_{\mathbf{x}_{ij} \backslash \mathbf{y}_{ij}}(\mathbf{y}_{ij}), \quad (6)$$

assuming that values are consistent between $\mathbf{x}_{ij}$ and $\mathbf{y}_{ij}$ (Lee & Bareinboim, 2020). Now consider the summation over $\mathbf{Z} \subseteq \mathbf{X}_i \cap \mathbf{Y}_j$. Then,

$$\textstyle\sum_{\mathbf{z}} P_{\mathbf{x}_i \backslash \mathbf{z}, \mathbf{z}}(\mathbf{y}_i) P_{\mathbf{x}_j}(\mathbf{y}_j \backslash \mathbf{z}, \mathbf{z}) = P_{\mathbf{x}_{ij} \backslash \mathbf{y}_{ij}}(\mathbf{y}_{ij} \backslash \mathbf{Z}).$$

To properly represent this as a matrix multiplication, the two matrices should share only $\mathbf{Z}$, and other shared variables or non-matching variables[5] need to be set to constants. To help understand, we illustrate in Fig. 4a the set relationships among $\mathbf{X}_i$, $\mathbf{Y}_i$, $\mathbf{X}_j$, and $\mathbf{Y}_j$ in the case of $\mathbf{Z} = \mathbf{X}_i \cap \mathbf{Y}_j$.

Therefore, $\mathbf{Y}_i \cap \mathbf{X}_j$ (the shared variables other than $\mathbf{Z}$) and $(\mathbf{X}_i \cup \mathbf{Y}_j) \backslash \mathbf{Z}$ (the non-matching variables) are fixed. These fixed variables are not colored in Fig. 4a. As a result, we can obtain a submatrix of $\mathsf{P}_{\mathbf{X}_{ij} \backslash \mathbf{Y}_{ij}}(\mathbf{Y}_{ij} \backslash \mathbf{Z})$ as the multiplication of the submatrices of $\mathsf{P}_{\mathbf{X}_i}(\mathbf{Y}_i)$ and $\mathsf{P}_{\mathbf{X}_j}(\mathbf{Y}_j)$ obtained via fixing variables.

Let $\mathbf{a}/\mathbf{B} = (\mathbf{A} \cap \mathbf{B}, \mathbf{a} \backslash \mathbf{B})$ which retains $\mathbf{B}$ as a set of variables and values of $\mathbf{a}$ excluding $\mathbf{B}$.

Lemma 2 (Matrix Equation of C-Factorization with Two Factors)**.** *Given a causal diagram $\mathcal{G}$ and an experimental distribution $Q = P_{\mathbf{r}}$ where a causal effect is c-factorized as $Q_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{z}} Q_{\mathbf{x}_i}(\mathbf{y}_i) Q_{\mathbf{x}_j}(\mathbf{y}_j)$ in $\mathcal{G} \backslash \mathbf{R}$, the effect can be represented as a matrix multiplication, if $\mathbf{Z} \subseteq \mathbf{X}_i \cap \mathbf{Y}_j$.*

[5] Those are the variables corresponding to the columns in the left matrix and the rows in the right matrix other than $\mathbf{Z}$.

Further, the corresponding matrix equation is

$$\begin{aligned}&\mathsf{Q}_{(\mathbf{x}_{ij}\setminus\mathbf{y}_{ij})/(\mathbf{X}_j\setminus\mathbf{X}_i\setminus\mathbf{Y}_i)}((\mathbf{y}_{ij}\setminus\mathbf{Z})/(\mathbf{Y}_i\setminus\mathbf{X}_j))\\&=\mathsf{Q}_{\mathbf{x}_i/\mathbf{Z}}(\mathbf{y}_i/(\mathbf{Y}_i\setminus\mathbf{X}_j))\mathsf{Q}_{\mathbf{x}_j/(\mathbf{X}_j\setminus\mathbf{X}_i\setminus\mathbf{Y}_i)}(\mathbf{y}_j/\mathbf{Z}),\end{aligned}$$

where the values are consistent.

We illustrate a causal graph in Fig. 4b where a variable matches to a partition in Fig. 4a—the number corresponds to the position in the Euler diagram. With $\mathbf{Z}=\{6\}$ and for an arbitrary instantiation of variables 2, 4, 5, and 7,

$$\mathsf{P}_{V_3,v_4,v_5}(V_1,v_2,v_7)=\mathsf{P}_{v_4,v_5,V_6}(V_1,v_2)\mathsf{P}_{v_2,V_3,v_4}(V_6,v_7),$$

that is, the resulting matrix is $|\mathfrak{X}_{V_1}|\times|\mathfrak{X}_{V_3}|$.

Similar to CI incorporated into chain-rule for a matrix equation, one might surmise that we may combine Rule 3 of do-calculus, which removes redundant interventions, into c-factorization. However, no Rule 3 is applicable to those terms on the RHS because those interventions (e.g., $\mathbf{X}_i$) are directly connected to the measurement (e.g., $\mathbf{Y}_i$).

In this section, we connected three different graphical criteria (chain-rule with conditional independence, adjustment criterion, and c-factorization) induced identification formulae to matrix equations. Results presented in this section are by no means complete. Nevertheless, this suite of characterizations will provide a fundamental understanding of the mathematical structures involved in the identification methods for proxy variables.

4. Generalized Proxy-based Criteria

Equipped with the characterization from the previous section, we revisit single- and double-proxy settings more formally which identify a causal effect through the combination of chain-rule, adjustment criteria, and inverses of matrices, and investigate its extension to deal with more restrictive conditions at the expense of other available distributions.

4.1. Generalizing a Single-Proxy Setting

We illustrate in Fig. 5 the available distributions and unknown distributions, considered in Fig. 1b, that can lead to a causal effect $P_x(y)$. First, note that $(X,Y \perp\!\!\!\perp W \mid U)$ in a causal graph $\mathcal{G}$ is central, which grants $P(W|U,x,y)=P(W|U,x)=P(W|U)$ so that

$$\mathsf{P}(y,W|x)=\mathsf{P}(W|U)\mathsf{P}(y,U|x) \tag{7}$$
$$\mathsf{P}(W|x)=\mathsf{P}(W|U)\mathsf{P}(U|x) \tag{8}$$
$$\mathsf{P}(W)=\mathsf{P}(W|U)\mathsf{P}(U). \tag{9}$$

If $\mathsf{P}(W|U)$ is invertible, the three distributions $\mathsf{P}(y,U|x)$, $\mathsf{P}(U|x)$, and $\mathsf{P}(U)$ are obtained where $\mathsf{P}(y|U,x)$ is computed by a chain rule. Then, $P_x(y)$ is identified as,

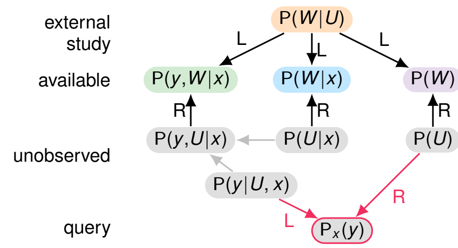


Figure 5: A schematic for identifiability with an external study $P(W|U)$ where a matrix multiplication $A=B\cdot C$ is represented as $B\overset{\text{L}}{\rightarrow}A\overset{\text{R}}{\leftarrow}C$ with positions annotated. Red lines into $P_x(y)$ highlight adjustment as matrix multiplication. Gray lines among $\mathsf{P}(y|U,x)$, $\mathsf{P}(y,U|x)$, and $\mathsf{P}(U|x)$ represent a chain rule relationship.

(a) (b)

Figure 6: Exemplary causal diagrams admitting a single proxy setting with a surrogate experiment (a) $P_x(W)$ without $W \perp\!\!\!\perp X \mid U,Y$ and (b) $P_{x,y}(W)$ without the condition.

$$\left(\frac{\mathsf{P}(W|U)^{-1}\mathsf{P}(y,W|x)}{\mathsf{P}(W|U)^{-1}\mathsf{P}(W|x)}\right)^{\top}\mathsf{P}(W|U)^{-1}\mathsf{P}(W) \tag{10}$$

where the fraction is an elementwise division.

However, this scheme would not work if X and W are e.g., directly connected (Fig. 6a). This challenging scenario can be handled if a different external study is available altogether with a surrogate experiment.

Case $(Y \perp\!\!\!\perp W \mid U,X)$**:** Consider a case $(Y \perp\!\!\!\perp W \mid U,X)$ as illustrated in Fig. 6a. Further, unlike the original setting, suppose an invertible matrix $\mathsf{P}(W|U,x)$ is available instead of $\mathsf{P}(W|U)$. Then, the identification cannot be completed since Eq. (9) becomes invalid. Nevertheless, if U is an admissible set for a surrogate experiment $P_x(w)$ and an experiment $P_x(W)$ is available, then $P_x(y)$ is identified as

$$\left(\frac{\mathsf{P}(W|U)^{-1}\mathsf{P}(y,W|x)}{\mathsf{P}(W|U)^{-1}\mathsf{P}(W|x)}\right)^{\top}\mathsf{P}(W|U,x)^{-1}\mathsf{P}_x(W).$$

Case without $(X,Y \perp\!\!\!\perp W \mid U)$**:** Now, suppose that neither X nor Y is conditionally independent to W given U. $\mathsf{P}(W|U,x,y)$ is required to obtain $\mathsf{P}(y,U|x)$ through $\mathsf{P}(y,W|x)=\mathsf{P}(W|U,x,y)\mathsf{P}(y,U|x)$. Since $P(W|U,x,y)$ is no longer equal to $P(W|U,x)$, one may additionally require $\mathsf{P}(W|U,x)$ to identify $\mathsf{P}(U|x)$. By the way, having $P(W|U,x,Y)$ is sufficient to obtain

$\mathsf{P}(U|x)$ through marginalizing out Y from $P(Y, U|x) = [P(y', U|x)]_{y'}$ so as to compute $P(y|U, x)$. Similar to the previous setting, the existence of a surrogate experiment $P_{x,y}(W)$ where U is an admissible set for $P_{x,y}(W)$ in $\mathcal{G}$ allows the identification of $\mathsf{P}(U)$, thus eliciting $P_x(y)$,

$$\left(\frac{\mathsf{P}(W|U)^{-1}\mathsf{P}(y, W|x)}{\sum_{y'}\mathsf{P}(W|U)^{-1}\mathsf{P}(y', W|x)}\right)^{\top}\mathsf{P}(W|U, x, y)^{-1}\mathsf{P}_{x,y}(W).$$

Extending these examples, we present Thm. 3 in Appendix which generalizes a single-proxy variable setting to take advantage of surrogate experiments and external studies, providing a broader spectrum of identification results.

4.2. Generalizing a Double-Proxy Setting

In this section, we generalize MGT criterion (Eq. (2)) to utilize distributions other than the originally considered observational study. MGT criterion for a double-proxy setting relies on the following CI statements in $\mathcal{G}$ to identify $P_x(y)$ with $P(X, Y, Z, W)$:

(C1) U is an adjustment set for $P_x(y)$ in $\mathcal{G}$, (11)

(C2) $Y \perp\!\!\!\perp Z \mid U, X$, (12)

(C3) $Z, X \perp\!\!\!\perp W \mid U$, (13)

(C4) $\mathsf{P}(W|Z, x)$ is invertible. (14)

Under these conditions, algebraic relationships between a causal effect $P_x(y)$ and other distributions can be illustrated as in Fig. 7 where the distributions form a closed loop alternating between (i) distributions with an unmeasured confounder and (ii) given distributions and a query. Among the four multiplications, $P_x(y)$ corresponds to an adjustment criterion (C1) so as to $P_x(y) = \mathsf{P}(y|U, x)\mathsf{P}(U)$. Others are due to the chain-rule combined with CI, e.g., (C2) $\mathsf{P}(y|U, x) = \mathsf{P}(y|U, x, z)$ and (C3) $\mathsf{P}(W|U) = \mathsf{P}(W|U, x, z)$. With (C4), which implies that both $\mathsf{P}(W|U)$ and $\mathsf{P}(U|Z, x)$ are invertible (Miao et al., 2018; Banerjee & Roy, 2014), the causal effect $P_x(y)$ can be expressed as Eq. (2) by subsequently rewriting $\mathsf{P}(U|Z, x)$ and $\mathsf{P}(y|U, x)$, and contracting $\mathsf{P}(W) = \mathsf{P}(W|U)\mathsf{P}(U)$ (see Fig. 15 in Appendix).

We are interested in relaxing assumption C3, where the CI grants the use of matrix multiplication leading to a chain-rule $\mathsf{P}(W) = \mathsf{P}(W|U)\mathsf{P}(U)$. It allows us to transform the multiplication into adjustment criterion as follows.

Case $(Z \perp\!\!\!\perp W \mid U, X)$ The assumption grants $\mathsf{P}(W|U, Z, x)$=$\mathsf{P}(W|U, x)$. Given that a surrogate experiment $P_x(W)$ is accessible and it can be decomposed as $\mathsf{P}_x(W) = \mathsf{P}(W|U, x)\mathsf{P}(U)$ where U is also an admissible set for $P_x(w)$, then, $P_x(y)$ is identified. An example is illustrated in Fig. 8a where an additional directed edge from X to W is allowed.

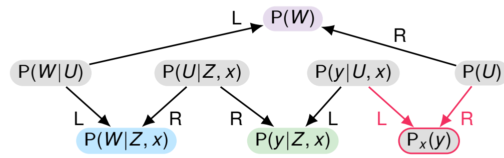
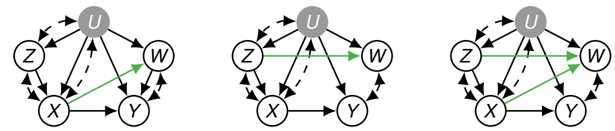


Figure 7: Schematic of MGT criterion.



(a) $P_x(W)$ (b) $P_z(W)$ (c) $P_{x,z}(W)$

Figure 8: Causal diagrams admitting a generalized MGT criterion under the availability of (a,b,c) surrogate experiments (captioned) and (b,c) external study.

Case $(X \perp\!\!\!\perp W \mid U, Z)$ One might surmise that we can similarly assume the existence of a surrogate experiment and its factorization based on an adjustment criterion,

$$\mathsf{P}_z(W) = \mathsf{P}(W|U, z)\mathsf{P}(U)$$

However, this dependence $Z \not\perp\!\!\!\perp W \mid U$ prohibits a matrix multiplication involving $\mathsf{P}(W|Z, x)$, i.e., Z needs to be instantiated as $\mathsf{P}(U|Z, x) = [\mathsf{P}(U|z', x)]_{z'}$ where $\mathsf{P}(U|z, x) = \mathsf{P}(W|U, z)^{-1}\mathsf{P}(W|z, x)$. Ultimately, we are unable to invoke the *double-inversion* trick as depicted in Fig. 15 to contract $\mathsf{P}(W|U, z)$ and $\mathsf{P}(U)$ into a U-free term.

This challenging situation can be addressed by assuming the existence of an external study $P(W|U, Z)$. Then,

$$P_x(y) = \mathsf{P}(y|Z, x)\mathsf{P}(U|Z, x)^{-1}\mathsf{P}(U) \quad (15)$$

where $\mathsf{P}(U|Z, x)$ is plugged-in and $\mathsf{P}(U) = \mathsf{P}(W|U, z')^{-1}\mathsf{P}_{z'}(W)$ for some $z' \in \mathfrak{X}_Z$. An example is shown in Fig. 8b. Note the absence of a directed edge from W to Y to preserve (C2).

Further dropping the condition $(X \perp\!\!\!\perp W \mid U, Z)$ yields a similar result where $\mathsf{P}(W|U, z, x)$ and $\mathsf{P}_{x,z}(W)$ replace $\mathsf{P}(W|U, z)$ and $\mathsf{P}_z(W)$, respectively, since x, unlike z, is *fixed* throughout the derivation (see Fig. 8c for an instance). We present a theorem (Thm. 4) that extends MGT criterion with varying degrees of the assumption in Appendix.

We have presented generalized identification results using proxy variables. Our results demonstrate the trade-offs between the relaxed CI assumptions and the requirements for an external study and surrogate experiment.

5. Intermediary Criteria

We present a novel condition where the probability of interest can be expressed as the multiplication of three matrices

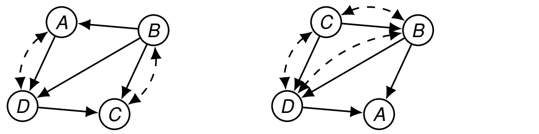


Figure 9: Causal diagrams where chain-rule intermediary criterion is applicable to identify $P(a|d)$ given $\mathsf{P}(a|C,d)$, $\mathsf{P}(B|C,d)$, and $\mathsf{P}(B|d)$.

and the probability is identifiable with available distributions relevant to the three terms.

Theorem 1 (Base Intermediary Criterion)**.** *Given a causal diagram $\mathcal{G}$, let $\{P_1,P_2,P_3,P_4\}$ be distributions compatible with $\mathcal{G}$. Let $\{\mathsf{P}_i\}_{i=1}^4$ be their matrix representations. If submatrices $\{\mathsf{P}'_i\}_{i=1}^4$ of $\{\mathsf{P}_i\}_{i=1}^4$ satisfy $\mathsf{P}'_1=\mathsf{P}'_2\mathsf{P}'_3\mathsf{P}'_4$ and $\mathsf{P}'_2\mathsf{P}'_3$, $\mathsf{P}'_3\mathsf{P}'_4$, and P'_3 are given, then, every probability in P'_1 is identified as $(\mathsf{P}'_2\mathsf{P}'_3)\mathsf{P}'^{\dagger}_3(\mathsf{P}'_3\mathsf{P}'_4)$.*

Although the theorem itself is rather general, we concretely characterize probability distributions satisfying $\mathsf{P}'_1=\mathsf{P}'_2\mathsf{P}'_3\mathsf{P}'_4$ with respect to two interpretations of multiplication, namely, chain-rule (Sec. 5.1) and c-factorization (Sec. 5.2).

5.1. Chain Rule based Intermediary Criterion

We start by characterizing an intermediary criterion with a chain-rule using a simple illustrative example. Let Q be an arbitrary interventional distribution. One way to decompose $Q(a,b,c|d)$ into three probabilities is

$$Q(a,b,c|d)=Q(a|b,c,d)Q(b|c,d)Q(c|d).$$

Let a probability of interest be $Q(a|d)=\sum_{b,c}Q(a,b,c|d)$ where the following distributions are available: $Q(B|C,D)$, $Q(A|C,D)$, and $Q(B|D)$. Given the first term $Q(a|b,c,d)$ being equal to $Q(a|b,d)$, the term can be multiplied by $Q(b|d)=\sum_c Q(b|c,d)Q(c|d)$. Hence, the matricized expression becomes,

$$\mathsf{Q}(A|d)=\overbrace{\mathsf{Q}(A|B,d)\,\mathsf{Q}(B|C,d)}^{\mathsf{Q}(A|C,d)}\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\underbrace{\phantom{\mathsf{Q}(B|C,d)}\mathsf{Q}(C|d)}_{\mathsf{Q}(B|d)},$$

for any $d\in\mathfrak{X}_D$. Two illustrative examples where this expression is applicable (i.e., $(C\perp\!\!\!\perp A\mid B,D)$ in $\mathcal{G}$) are shown in Fig. 9. Now, we propose a chain-rule intermediary criterion.

Lemma 3 (Chain-Rule Intermediary Criterion)**.** *Given a causal diagram $\mathcal{G}$, let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{R}$ be disjoint subsets of $\mathbf{V}$ with $\mathbf{D}$ and $\mathbf{R}$ can be empty. Let $\mathbf{B}=\mathbf{B}'\,\dot{\cup}\,\mathbf{B}''$ and $\mathbf{C}=\mathbf{C}'\,\dot{\cup}\,\mathbf{C}''$ where $\mathbf{B}'$ and $\mathbf{C}'$ are not empty. Given an interventional distribution $Q=P_{\mathbf{r}}$, if*

$$Q(\mathbf{a},\mathbf{b}'',\mathbf{c}''|\mathbf{d})=\textstyle\sum_{\mathbf{b}',\mathbf{c}'}Q(\mathbf{a}|\mathbf{b},\mathbf{c},\mathbf{d})Q(\mathbf{b}|\mathbf{c},\mathbf{d})Q(\mathbf{c}|\mathbf{d})$$

and $(\mathbf{C}'\perp\!\!\!\perp\mathbf{A}\mid\mathbf{B},\mathbf{C}'',\mathbf{D})$ in $\mathcal{G}\backslash\mathbf{R}$, then, $\mathsf{Q}(\mathbf{A},\mathbf{b}'',\mathbf{c}''|\mathbf{d})$ is

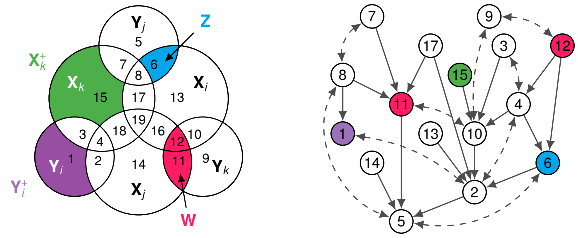
$$\mathsf{Q}(\mathbf{A},\mathbf{b}''|\mathbf{C}',\mathbf{c}'',\mathbf{d})\cdot\mathsf{Q}(\mathbf{B}',\mathbf{b}''|\mathbf{C}',\mathbf{c}'',\mathbf{d})^{\dagger}\cdot\mathsf{Q}(\mathbf{B}',\mathbf{b}'',\mathbf{c}''|\mathbf{d}).$$

Additional CI allows the criterion to return a matrix, which can then be nested into another matrix equation.

Corollary 1. *Given Lemma 3, if $(\mathbf{D}\perp\!\!\!\perp\mathbf{A},\mathbf{B}''\mid\mathbf{C})$ and $(\mathbf{D}\perp\!\!\!\perp\mathbf{B}\mid\mathbf{C})$ in $\mathcal{G}\backslash\mathbf{R}$, then $\mathsf{Q}(\mathbf{A},\mathbf{b}'',\mathbf{c}''|\mathbf{D})$ is*

$$\mathsf{Q}(\mathbf{A},\mathbf{b}''|\mathbf{C}',\mathbf{c}'')\cdot\mathsf{Q}(\mathbf{B}',\mathbf{b}''|\mathbf{C}',\mathbf{c}'')^{\dagger}\cdot\mathsf{Q}(\mathbf{B}',\mathbf{b}'',\mathbf{c}''|\mathbf{D}).$$

These results advance the current understanding of the causal or non-causal identification with fragments of information presented as conditional distributions.



(a) Euler diagram (b) a causal graph

Figure 10: (a) an Euler diagram depicting set relationships. (b) a causal graph where each number in (a) corresponds to a variable excluding 16, 18, and 19 for simplicity.

5.2. C-Factorization based Intermediary Criterion

Now, we proceed to characterize an intermediary criterion where a matrix equation corresponds to a c-factorization resulting in three factors, extending the case of a pair of factors (Lemma 2). Concretely speaking, we are interested in the matrix form of

$$P_{\mathbf{x}_\ell}(\mathbf{y}_\ell)=\textstyle\sum_{\mathbf{z}}P_{\mathbf{x}_i}(\mathbf{y}_i)\sum_{\mathbf{w}}P_{\mathbf{x}_j}(\mathbf{y}_j)P_{\mathbf{x}_k}(\mathbf{y}_k) \qquad (16)$$

where $\mathbf{Z}\subseteq\mathbf{X}_i\cap\mathbf{Y}_j$ and $\mathbf{W}\subseteq\mathbf{X}_j\cap\mathbf{Y}_k$ are disjoint sets of variables to be marginalized. In addition, available distributions are of the form $P_{\mathbf{X}_j}(\mathbf{Y}_j)$, $P_{\mathbf{X}_{ij}\backslash\mathbf{Y}_{ij}}(\mathbf{Y}_{ij}\backslash\mathbf{Z})$, and $P_{\mathbf{X}_{jk}\backslash\mathbf{Y}_{jk}}(\mathbf{Y}_{jk}\backslash\mathbf{W})$.

We illustrate in Fig. 10a the set relationships among the six terms ($\mathbf{X}_\cdot$, $\mathbf{Y}_\cdot$) together with $\mathbf{Z}$ and $\mathbf{W}$, and a matching causal graph in Fig. 10b. By definition, three $\mathbf{Y}$ terms are disjoint to each other and each pair of $\mathbf{X}$ and $\mathbf{Y}$ of the same index are disjoint. A noticeable feature is that $\mathbf{Z}$ is disjoint to $\mathbf{X}_k$. We will present a theorem first, then briefly explain the reasons behind the exclusion.

Theorem 2 (C-Factorization Intermediary Criterion)**.** *Let $\mathcal{G}$ be a causal diagram and $Q=P_{\mathbf{r}}$. Let $Q_{\mathbf{x}_\ell}(\mathbf{y}_\ell)$ be c-factorized as Eq. (16). Let $\mathbf{X}_k^+$ be a subset of $\mathbf{X}_k$ excluding the rest five sets, $\{\mathbf{Y}_i,\mathbf{Y}_j,\mathbf{Y}_k,\mathbf{X}_i,\mathbf{X}_j\}$. $\mathbf{Y}_i^+$ is similarly* *defined. If* $\mathbf{Z} \subseteq (\mathbf{X}_i \cap \mathbf{Y}_j)\backslash\mathbf{X}_k$ *and* $\mathbf{W} \subseteq \mathbf{X}_j \cap \mathbf{Y}_k$*, then* $\mathsf{Q}_{\mathbf{x}_\ell/\mathbf{X}_k^+}(\mathbf{y}_\ell/\mathbf{Y}_i^+)$*, a submatrix of* $\mathsf{Q}_{\mathbf{X}_\ell}(\mathbf{Y}_\ell)$*, equals to*

$$\mathsf{Q}_{(\mathbf{x}_{ij}\backslash\mathbf{y}_{ij})/\mathbf{W}}((\mathbf{y}_{ij}\backslash\mathbf{Z})/\mathbf{Y}_i^+) \cdot \mathsf{Q}_{\mathbf{x}_j/\mathbf{W}}(\mathbf{y}_j/\mathbf{Z})^\dagger \cdot \mathsf{Q}_{(\mathbf{x}_{jk}\backslash\mathbf{y}_{jk})/\mathbf{X}_k^+}((\mathbf{y}_{jk}\backslash\mathbf{W})/\mathbf{Z}).$$

The theorem imposes an additional constraint that $\mathbf{Z}$ should be disjoint to $\mathbf{X}_k$ compared to naively interpreting the three-matrix multiplication as two individual matrix multiplications as seen in Lemma 2. Briefly speaking, given that the summation over $\mathbf{W}$ is nested (Eq. (16)), $(\mathbf{X}_j \cup \mathbf{Y}_j)\backslash\mathbf{W}$ is fixed along with $\mathbf{Y}_j \cap \mathbf{X}_k$ (7 and 8 in the figure). Thus, $\mathbf{X}_i \cap \mathbf{Y}_j \cap \mathbf{X}_k$ can't be part of $\mathbf{Z}$. In other words, the constraints imposed in the original expression asymmetrically affect what $\mathbf{Z}$ can be but not what $\mathbf{W}$ can be.

In this section, we introduced novel intermediary criteria by characterizing both chain-rule and c-factorization with respect to matrix multiplications of three matrices exploiting the pseudoinverse, which has never been employed in the context of causal identification to the best of our knowledge.

6. A Causal Identification Framework

We provide a framework for the problem of causal identification given a collection of distributions and a causal graph where the framework can unify different approaches such as proxy variable based criteria (Greenland & Lash, 2008; Kuroki & Pearl, 2014; Miao et al., 2018), their generalizations (Sec. 4), factorization approaches (Shpitser & Pearl, 2006; Lee et al., 2019; Lee & Bareinboim, 2020), and intermediary criteria (Sec. 5).

The basic structure of the framework is built on (Lee & Bareinboim, 2020; Lee & Shpitser, 2020) (illustrated in black in Fig. 11), where the given query $P_{\mathbf{x}}(\mathbf{y})$ is c-factorized with respect to some latent projection of the given causal graph $\mathcal{G}$ so as to identify every resulting c-factor. Considering that the given distributions $\mathbb{D}$ can contain conditional ones, one can expand the distributions by repeatedly applying do-calculus and chain rule where such an expanded data set $\mathbb{D}'$ is formally called 'chain rule closed' (Lee & Shpitser, 2020). Once a larger data set is obtained, ID algorithm adapted to conditional distribution (Bareinboim & Tian, 2015; Lee & Shpitser, 2020) (shown as ID-RC) is used to identify each factor with a (conditional) distribution.

However, this existing structure does not exploit the insights demonstrated in Sec. 3 to 5, and is incapable of identifying cases like the example in Fig. 3. We now show how to expand this existing framework, through a series of simple augmentations, and enable it to take advantage of the newly developed matrix-based approach. First, *chain rule closure* alone is insufficient to expand $\mathbb{D}$. We can take advantage of the characterizations in Sec. 3 by examining distributions

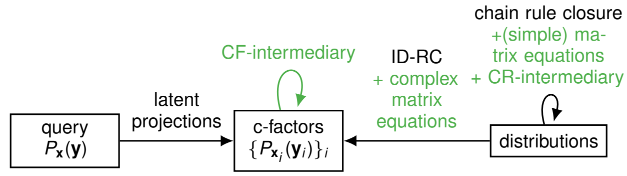


Figure 11: A framework for causal identification

that can be recovered through a simple matrix inversion. Further, the chain rule intermediary criterion is adopted.

With expanded distributions $\mathbb{D}'$, each factor can be identified not only through ID-RC using one of the available distributions (one-to-one) but also through complex matrix equations involving multiple distributions (one-to-many) such as proxy based methods (Thms 3 and 4 in Appendix) as we have seen in the motivational example (Fig. 3a). A complex matrix equation can be understood as an abstract syntax tree made of matrix multiplications and other machinery where the c-factor to identify is its root. By collapsing the tree through subsequently replacing each term with its expression, one may identify the factor (e.g., MGT).

Finally, c-factorization intermediary criterion (Thm. 2) identifies a c-factor, that is not identified by any other methods, with other three identified c-factors. Lee & Bareinboim (2020, Prop. 6) investigated how c-factors based on different latent projections form a hierarchy. Given an unidentified c-factor, such hierarchy provides a guidance on which identified c-factors to look for. Once no more c-factor is identifiable, the algorithm can finally examines whether the given query is composed of the identified c-factors, and returns an identification formula if true. The components newly added to the basic framework is shown in green (Fig. 11).

7. Conclusion

In this paper, we studied the use of matrix equations in causal identification given general distributions. In particular, we characterized matrix equations made of distributions driven by graphical constraints, deepening our understanding on algebraic constraints imposed by the graph. We generalized proxy-based identification methods to cover a broad spectrum of problem instances beyond the identification problems with an observation study and optional external study. We developed novel intermediary criteria that identify a query by utilizing the pseudoinverse of the center matrix within the multiplication of three matrices. Finally, we provide an in-depth discussion on how to integrate many existing identification results to produce an identification formula for a causal query given a causal graph and a set of distributions that can be marginal, experimental, and conditional taking advantage of matrix equations and (pseudo)inverses.

References

Banerjee, S. and Roy, A. *Linear Algebra and Matrix Analysis for Statistics*. Boca Raton: Taylor & Francis, 2014.

Bareinboim, E. and Pearl, J. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113:7345–7352, 2016.

Bareinboim, E. and Tian, J. Recovering causal effects from selection bias. In Koenig, S. and Bonet, B. (eds.), *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pp. 3475–3481, Palo Alto, CA, 2015. AAAI Press.

Geiger, D., Verma, T., and Pearl, J. Identifying independence in Bayesian networks. In *Networks*, volume 20, pp. 507–534. John Wiley, Sussex, England, 1990.

Greenland, S. and Lash, T. Bias analysis. In Rothman, K., Greenland, S., and Lash, T. (eds.), *Modern Epidemiology*, pp. 345–380. Lippincott Williams & Wilkins, Philadelphia, PA, 3rd edition, 2008.

Huang, Y. and Valtorta, M. Pearl's calculus of intervention is complete. In Dechter, R. and Richardson, T. (eds.), *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI 2006)*, pp. 217–224. AUAI Press, Corvallis, OR, 2006.

Kuroki, M. and Pearl, J. Measurement bias and effect restoration in causal inference. *Biometrika*, 101:423–437, 2014.

Lee, J. J. R. and Shpitser, I. Identification methods with arbitrary interventional distributions as inputs, 2020. arXiv:2004.01157.

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.

Lee, S., Correa, J. D., and Bareinboim, E. General identifiability with arbitrary surrogate experiments. In *Proceedings of the Thirty-Fifth Conference Annual Conference on Uncertainty in Artificial Intelligence*, Corvallis, OR, 2019. AUAI Press, in press.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 08 2018. ISSN 0006-3444.

Nashed, M. Z. Generalized inverses, normal solvability, and iteration for singular operator equations. In Rall, L. B. (ed.), *Nonlinear Functional Analysis and Applications*, pp. 311 – 359. Academic Press, 1971.

Pearl, J. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1988.

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

Pearl, J. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000.

Shpitser, I. and Pearl, J. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI 2006)*, pp. 1219–1226. AAAI Press, Menlo Park, CA, 2006.

Shpitser, I., VanderWeele, T., and Robins, J. On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pp. 527–536. AUAI, Corvallis, OR, 2010.

Spirtes, P., Glymour, C. N., and Scheines, R. *Causation, prediction, and search*. MIT press, 2000.

Tian, J. *Studies in Causal Reasoning and Learning*. PhD thesis, Computer Science Department, University of California, Los Angeles, CA, November 2002.

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.

Verma, T. and Pearl, J. Causal networks: Semantics and expressiveness. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*, pp. 352–359, Mountain View, CA, 1988.

Verma, T. and Pearl, J. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence (UAI 1990)*, pp. 220–227, Cambridge, MA, July 1990.

Wang, Y. and Blei, D. M. The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528): 1574–1596, 2019.