Causal Calculus in the Presence of Cycles, Latent Confounders and Selection Bias

Patrick Forré
Informatics Institute
University of Amsterdam
The Netherlands
p.d.forre@uva.nl

Joris M. Mooij
Informatics Institute
University of Amsterdam
The Netherlands
j.m.mooij@uva.nl

Abstract

We prove the main rules of *causal calculus* (also called *do-calculus*) for *i/o structural causal models (ioSCMs)*, a generalization of a recently proposed general class of non-/linear structural causal models that allow for cycles, latent confounders and arbitrary probability distributions. We also generalize *adjustment criteria and formulas* from the acyclic setting to the general one (i.e. ioSCMs). Such criteria then allow to estimate (conditional) causal effects from observational data that was (partially) gathered under selection bias and cycles. This generalizes the *backdoor criterion*, the *selection-backdoor criterion* and extensions of these to arbitrary ioSCMs. Together, our results thus enable causal reasoning in the presence of cycles, latent confounders and selection bias. Finally, we extend the ID algorithm for the identification of causal effects to ioSCMs.

1 INTRODUCTION

Statistical models are governed by the rules of probability (e.g. sum and product rule), which link joint distributions with the corresponding (conditional) marginal ones. *Causal models* follow additonal rules, which relate the observational distributions with the interventional ones. In contrast to the rules of probability theory, which directly follow from their axioms, the rules of *causal calculus* need to be proven, when based on the definition of *structural causal models* (SCMs). As SCMs will among other things depend on the underlying graphical structure (e.g. with or without cycles or bidirected edges, etc.), the used function classes (e.g. linear or non-linear, etc.) and the allowed probability distributions (e.g. discrete, continuous, singular or mixtures, etc.) the respective endeavour is not immediate.

Such a framework of causal calculus contains rules about when one can 1.) insert/delete observations, 2.) exchange action/observation, 3.) insert/delete actions; and about when and how to recover from interventions and/or selection bias (backdoor and selection-backdoor criterion), etc. (see [1, 4, 5, 14, 21–24, 26, 27, 32–35]). While these rules have been extensively studied for *acyclic* causal models, e.g. (semi-)Markovian models, which are attached to directed acyclic graphs (DAGs) or acyclic directed mixed graphs (ADMGs) (see [1, 4, 5, 14, 21–24, 26, 27, 32–35]), the case of causal models with *cycles* stayed in the dark.

To deal with cycles and latent confounders at the same time in this paper we will introduce the class of *input/output structural causal models (ioSCMs)*, a "conditional" version of the recently proposed class of *modular structural causal models (mSCMs)* (see [10, 11]) to also include "input" nodes that can play the role of parameter/context/action/intervention nodes. ioSCMs have several desirable properties: They allow for arbitrary probability distributions, non-/linear functional relations, latent confounders and cycles. They can also model non-/probabilistic external and probabilistic internal nodes in one framework. The cycles are modelled in a least restrictive way such that the class of ioSCMs still becomes closed under arbitrary marginalizations and interventions. All causal models that are based on acyclic graphs like DAGs, ADMGs or mDAGs (see [9, 28]) can be interpreted as special acyclic ioSCMs. Besides feedback over time ioSCMs can also express instantaneous and equilibrated feedback under the made model assumptions (e.g. the ODEs in [2, 18]). All models where the non-trivial cycles are "contractive" (negative feedback loops, see [11]) are ioSCMs without further assumptions. Thus ioSCMs generalize all these classes of causal models in one framework, which goes beyond the acyclic setting and also allows for conditional

versions of those (e.g. CADMGs), expressed via external non-/probabilistic "input" nodes. Also the *generalized directed global Markov property* for mSCMs (see [10, 11]) generalizes to ioSCMs, i.e. ioSCMs entail the conditional independence relations that follow from the σ-*separation criterion* in the underlying graph, where σ-*separation* generalizes the usual d-separation (also called m- or m*-separation, see [9, 20, 24, 28, 38]) from acyclic graphs to directed mixed graphs (DMGs) (and even HEDGes [10] and σ-CGs [11]) with or without cycles in a non-naive way.

This paper now aims at proving the mentioned main rules of causal calculus for ioSCMs and derive adjustment criteria with corresponding adjustment formulas like generalized (selection-)backdoor adjustments. We also provide an extension of the ID algorithm for the identification of causal effects to the ioSCM setting, which reduces to the usual one in the acyclic case.

The paper is structured as follows: We will first give the precise definition of ioSCMs closely mirroring mSCMs from [10, 11]. We will then review σ-separation and generalize its criterion from mSCMs (see [10, 11]) to ioSCMs. As a preparation for the causal calculus, which relates observational and interventional distributions, we will then show how one can extend a given ioSCM to one that also incorporates additional interventional variables indicating the regime of interventions on the observed nodes. We will then show how the rules of causal calculus directly follow from applying the σ-separation criterion to such an extended ioSCM. We then derive the mentioned general adjustment criteria with corresponding adjustment formulas. Finally, we introduce the right definitions for ioSCMs to extend the ID algorithm for the identification of causal effects to the general setting.

2 INPUT/OUTPUT STRUCTURAL CAUSAL MODELS

In this section we will define *input/output structural causal models* (ioSCMs), which can be seen as a "conditional" version of modular structural causal models (mSCMs) defined in [10, 11]. We will then construct marginalized ioSCMs and intervened ioSCMs. To allow for cycles we first need to introduce the notion of loop of a graph and its strongly connected components.

Definition 2.1 (Loops). *Let $G = (V, E)$ be a directed graph (with or without cycles).*

1. *A set of nodes $S \subseteq V$ is called a* loop *of G if for every two nodes $v_1, v_2 \in S$ there are two directed walks $v_1 \longrightarrow \cdots \longrightarrow v_2$ and $v_2 \longrightarrow \cdots \longrightarrow v_1$ in G such that all the intermediate nodes are also in S (if any). The sets $S = \{v\}$ are also considered as*

loops (independent of $v \longrightarrow v \in E$ or not).

2. *The set of loops of G is written as $\mathcal{L}(G)$.*
3. *The* strongly connected component *of v in G is defined to be:* $\mathrm{Sc}^G(v) := \mathrm{Anc}^G(v) \cap \mathrm{Desc}^G(v)$.
4. *The set of strongly connected components is $\mathcal{S}(G)$.*

Remark 2.2. *Let $G = (V, E)$ be a directed graph.*

1. *We always have $v \in \mathrm{Sc}^G(v)$ and $\mathrm{Sc}^G(v) \in \mathcal{L}(G)$.*
2. *If G is acyclic then: $\mathcal{L}(G) = \{\{v\} \,|\, v \in V\}$.*

In the following all spaces are meant to be equipped with σ-algebras and all maps to be measurable. Whenever (regular) conditional distributions occur we implicitly assume standard measurable spaces (to ensure existence).

Definition 2.3 (Input/Output Structural Causal Model). *An* input/output (i/o) structural causal model (ioSCM) *by definition consists of:*

1. *a set of nodes $V^+ = V \dot\cup U \dot\cup J$, where elements of V correspond to output/observed variables, elements of U to probabilistic latent variables and elements of J to input/intervention variables.*
2. *an observation/latent/action space $\mathcal{X}_v$ for every $v \in V^+$, $\mathcal{X} := \prod_{v \in V^+} \mathcal{X}_v$,*
3. *a product probability measure $\mathbb{P}_U = \bigotimes_{u \in U} \mathbb{P}_u$ on the latent space $\mathcal{X}_U := \prod_{u \in U} \mathcal{X}_u$,*
4. *a directed graph structure $G^+ = (V^+, E^+)$ with the properties:*
   (a) $V = \mathrm{Ch}^{G^+}(U \cup J)$,
   (b) $\mathrm{Pa}^{G^+}(U \cup J) = \emptyset$,
   where Ch^{G^+} and Pa^{G^+} stand for children and parents in G^+, resp.,[1]
5. *a system of causal mechanisms $g = (g_S)_{\substack{S \in \mathcal{L}(G^+) \\ S \subseteq V}}$:*

$$g_S : \prod_{v \in \mathrm{Pa}^{G^+}(S) \setminus S} \mathcal{X}_v \to \prod_{v \in S} \mathcal{X}_v, [2]$$

that satisfy the following global compatibility *conditions: For every nested pair of loops $S' \subseteq S \subseteq V$ of G^+ and every element $x_{\mathrm{Pa}^{G^+}(S) \cup S} \in \prod_{v \in \mathrm{Pa}^{G^+}(S) \cup S} \mathcal{X}_v$ we have the implication:*

$$g_S(x_{\mathrm{Pa}^{G^+}(S) \setminus S}) = x_S$$
$$\implies g_{S'}(x_{\mathrm{Pa}^{G^+}(S') \setminus S'}) = x_{S'},$$

where $x_{\mathrm{Pa}^{G^+}(S') \setminus S'}$ and $x_{S'}$ denote the corresponding components of $x_{\mathrm{Pa}^{G^+}(S) \cup S}$.

---

[1] To have a "reduced" form of the latent space one can in addition impose the condition: $\mathrm{Ch}^{G^+}(u_1) \nsubseteq \mathrm{Ch}^{G^+}(u_2)$ for every two distinct $u_1, u_2 \in U$. This can always be achieved by gathering latent nodes together if $\mathrm{Ch}^{G^+}(u_1) \subseteq \mathrm{Ch}^{G^+}(u_2)$.

[2] Note that the index set runs over all "observable loops" $S \subseteq V$, $S \in \mathcal{L}(G^+)$, not just the sets $\{v\}$ for $v \in V$.

The ioSCM will be denoted by $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$.

Definition 2.4 (Modular structural causal model, see [10, 11]). *A modular structural causal model (mSCM) is an ioSCM without input nodes, i.e. $J = \emptyset$.*

Remark 2.5 (Composition of ioSCMs). *Consider two ioSCMs M_1, M_2 and an identification of subsets $I_1 \subseteq V_1^+$ with $I_2 \subseteq J_2$ and maps $g_{i_2} : \mathcal{X}_{i_1} \to \mathcal{X}_{i_2}$, for i_1 corresponding to i_2, e.g. $g_{i_2} = \mathrm{id}$ if possible. We can now "glue" them together to get a new ioSCM M_3 given by $V_3 := V_1 \dot\cup V_2 \dot\cup I_2$, $U_3 := U_1 \dot\cup U_2$, $J_3 := J_1 \dot\cup J_2 \setminus I_2$ and $G_3^+ := G_1^+ \cup G_2^+$, where we add the the edges $i_1 \longrightarrow i_2$, and the mechanisms g_{i_2} and $\mathbb{P}_{U_3} := \mathbb{P}_{U_1} \otimes \mathbb{P}_{U_2}$.*

Example 2.6 (Constructing mSCMs from ioSCMs). *Given an ioSCM $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ with graph $G^+ = (V \dot\cup U \dot\cup J, E^+)$ we can construct a well-defined mSCM by specifying a product distribution $\mathbb{P}_J := \bigotimes_{j \in J} \mathbb{P}_j$ on $\mathcal{X}_J := \prod_{j \in J} \mathcal{X}_j$ and following 2.5 with M_1 with only $U_1 := J_2$ without any edges and gluing maps $g_i := \mathrm{id}$.*

The actual joint distributions on the observed space $\mathcal{X}_V$ and thus the random variables attached to any ioSCM will be defined in the following.

Definition 2.7. *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. The following constructions will depend on the choice of a fixed value $x_J \in \mathcal{X}_J$.*

1. *The latent variables are given by $(X_u)_{u \in U} \sim \mathbb{P}_U$, i.e. by the canonical projections $X_u : \mathcal{X}_U \to \mathcal{X}_u$, which are jointly $\mathbb{P}_U$-independent. We put $X_u^{\mathrm{do}(x_J)} := X_u$, i.e., independent of x_J.*

2. *For $j \in J$ we put $X_j^{\mathrm{do}(x_J)} := x_j$, the constant variable given by the j-component of x_J.*

3. *The observed variables $(X_v^{\mathrm{do}(x_J)})_{v \in V}$ are inductively defined by:*
$$X_v^{\mathrm{do}(x_J)} := g_{S,v}\big((X_w^{\mathrm{do}(x_J)})_{w \in \mathrm{Pa}^{G^+}(S) \setminus S}\big),$$
where $S := \mathrm{Sc}^{G^+}(v) := \mathrm{Anc}^{G^+}(v) \cap \mathrm{Desc}^{G^+}(v)$ and where the second index v refers to the v-component of g_S. The induction is taken over any topological order of the strongly connected components of G^+, which always exists (see [10]).

4. *By the compatibility condition for g we then have that for every $S \in \mathcal{L}(G^+)$ with $S \subseteq V$ the following equality holds:*
$$X_S^{\mathrm{do}(x_J)} = g_S(X_{\mathrm{Pa}^{G^+}(S) \setminus S}^{\mathrm{do}(x_J)}),$$
where we put $\mathcal{X}_A := \prod_{v \in A} \mathcal{X}_v$ and $X_A := (X_v)_{v \in A}$ for subsets A.

5. *We define the family of conditional distributions:*
$$
\begin{aligned}
&\mathbb{P}_U(X_A | X_B, X_J = x_J) \\
:=\ &\mathbb{P}_U(X_A | X_B, \mathrm{do}(X_J = x_J)) \\
:=\ &\mathbb{P}_U(X_A^{\mathrm{do}(x_J)} | X_B^{\mathrm{do}(x_J)}),
\end{aligned}
$$

for $A, B \subseteq V$ and $x_J \in \mathcal{X}_J$. Note that in the following we will use the do *and the* do-*free notation (only) for the J-variables interchangeably.*

6. *If we, furthermore, specify a product distribution $\mathbb{P}_J = \bigotimes_{j \in J} \mathbb{P}_j$ on $\mathcal{X}_J$, then we get a joint distribution $\mathbb{P}$ on $\mathcal{X}_{V \cup J}$ by setting:*
$$\mathbb{P}(X_V, X_J) := \mathbb{P}_U(X_V | \mathrm{do}(X_J)) \otimes \mathbb{P}_J(X_J).$$

Remark 2.8. *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. For every subset $A \subseteq V$ we get a well-defined map $g_A : \mathcal{X}_{\mathrm{Pa}^{G^+}(A) \setminus A} \to \mathcal{X}_A$, by recursively plugging in the g_S into each other for the biggest occuring loops $S \subseteq A$ by the same arguments as before. These then are all globally compatible by construction and satisfy:*
$$X_A^{\mathrm{do}(x_J)} = g_A(X_{\mathrm{Pa}^{G^+}(A) \setminus A}^{\mathrm{do}(x_J)}).$$

Similar to mSCMs (see [10, 11]) we can define the marginalization of an ioSCM.

Definition 2.9 (Marginalization of ioSCMs). *Let $M = (G^+, \mathcal{X}, \mathbb{P}, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$ and $W \subseteq V$ a subset. The marginalized ioSCM $M^{\backslash W}$ w.r.t. W can be defined by plugging the functions g_S related to W into each other. For example, when marginalizing out $W = \{w\}$ we can define (for the non-trivial case $w \in \mathrm{Pa}^{G^+}(S) \setminus S$):*
$$
\begin{aligned}
&g_{S',v}(x_{\mathrm{Pa}(G^+)^{\backslash W}(S') \setminus S'}) := \\
&g_{S,v}\big(x_{\mathrm{Pa}^{G^+}(S) \setminus (S \cup \{w\})}, g_{\{w\}}(x_{\mathrm{Pa}^{G^+}(w) \setminus \{w\}})\big),
\end{aligned}
$$
where $(G^+)^{\backslash W}$ is the marginalized graph of G^+ (see Supplementary Material B), $S' \subseteq V^{\backslash W} := V \setminus W$ is any loop of $(G^+)^{\backslash W}$ and S the corresponding induced loop in G^+.

Similar to mSCMs (see [10, 11]) we now define what it means to intervene on *observed* nodes in an ioSCM.

Definition 2.10 (Perfect interventions on ioSCMs). *Let $M = (G^+, \mathcal{X}, \mathbb{P}, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. Let $W \subseteq V \cup J$ be a subset. We then define the post-interventional ioSCM $M_{\mathrm{do}(W)}$ w.r.t. W:*

1. *Define the graph $G_{\mathrm{do}(W)}^+$ by removing all the edges $v \longrightarrow w$ for all nodes $w \in W$ and $v \in \mathrm{Pa}^{G^+}(w)$.*
2. *Put $V_{\mathrm{do}(W)} := V \setminus W$ and $J_{\mathrm{do}(W)} := J \cup W$.*
3. *Remove the functions g_S for loops S with $S \cap W \neq \emptyset$.*

The remaining functions then are clearly globally compatible and we get a well-defined ioSCM $M_{\mathrm{do}(W)}$.

3 CONDITIONAL INDEPENDENCE

Here we generalize conditional independence for structured families of distributions. The main application will be the distributions $(\mathbb{P}_U(X_V|\operatorname{do}(X_J = x_J)))_{x_J \in \mathcal{X}_J}$ coming from ioSCMs, but the following definition might be of more general importance.

Definition 3.1 (Conditional independence). *Let $\mathcal{X}_V := \prod_{v \in V} \mathcal{X}_v$ and $\mathcal{X}_J := \prod_{j \in J} \mathcal{X}_j$ be product spaces and*

$$\mathbb{P} := (\mathbb{P}_V(X_V|x_J))_{x_J \in \mathcal{X}_J}$$

a family of distributions on $\mathcal{X}_V$ (measurably[3]) parametrized by $\mathcal{X}_J$. For subsets $A, B, C \subseteq V \dot\cup J$ we write:

$$X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \,|\, X_C$$

if and only if for every product distribution $\mathbb{P}_J = \bigotimes_{j \in J} \mathbb{P}_j$ on $\mathcal{X}_J$ we have: $X_A \perp\!\!\!\perp_{\mathbb{P}_{V \cup J}} X_B \,|\, X_C$, i.e.:

$$\mathbb{P}_{V \cup J}(X_A|X_B, X_C) = \mathbb{P}_{V \cup J}(X_A|X_C) \quad \mathbb{P}_{V \cup J}\text{-}a.s.,$$

where $\mathbb{P}_{V \cup J}(X_{V \cup J}) := \mathbb{P}_V(X_V|X_J) \otimes \mathbb{P}_J(X_J)$ is the distribution given by $X_J \sim \mathbb{P}_J$ and then $X_V \sim \mathbb{P}_V(\_|X_J)$.

Remark 3.2. *1. The definition 3.1 assumes that the input variables J are considered independent, in contrast to [3, 29], where all J are implicitly assumed to be jointly confounded. We discuss this further in Supplementary Material C.*

2. In contrast with [3, 6, 29] definition 3.1 can accommodate any variable from V or J at any spot of the conditional independence statement.

3. $\perp\!\!\!\perp_{\mathbb{P}}$ satisfies the separoid axioms (see [6, 7, 13, 25] or see rules 1-5 in Lem. 4.5 for $\perp\!\!\!\perp_{\mathbb{P}}$) as these rules are preserved under conjunction.

4 σ-SEPARATION

In this section we will define σ-separation on directed mixed graphs (DMG) and present the generalized directed global Markov property stating that every ioSCM will entail the conditional independencies that come from σ-separation in its induced DMG. We will again closely follow the work in [11].

Definition 4.1 (Directed mixed graph (DMG)). *A directed mixed graph (DMG) G consists of a set of nodes V together with a set of directed edges ($\longrightarrow$) and bidirected edges ($\leftrightarrow$). In case G contains no directed cycles it is called an* acyclic directed mixed graph (ADMG).

---

[3]We require that for every measurable $F \subseteq \mathcal{X}_V$ the map $\mathcal{X}_J \to [0, 1]$ given by $x_J \mapsto \mathbb{P}_V(X_V \in F|x_J)$ is measurable. Such families of distributions are also called *channels* or *(stochastic) Markov (transition) kernels* (see [16]).

Definition 4.2 (σ-Open walk in a DMG). *Let G be a DMG with set of nodes V and $C \subseteq V$ a subset. Consider a walk π in G with $n \geq 1$ nodes:*

$$v_1 \;\leftrightarrows\; \cdots \;\leftrightarrows\; v_n.^{[4]}$$

The walk will be called C-σ-open if:

1. the endnodes $v_1, v_n \notin C$, and
2. every triple of adjacent nodes in π that is of the form:
  (a) collider: $\quad v_{i-1} \leftrightarrow v_i \leftrightarrow v_{i+1}$, satisfies $v_i \in C$,
  (b) left chain: $\quad v_{i-1} \longleftarrow v_i \leftrightarrows v_{i+1}$, satisfies $v_i \notin C$ or $v_i \in C \cap \operatorname{Sc}^G(v_{i-1})$,
  (c) right chain: $\quad v_{i-1} \leftrightarrows v_i \longrightarrow v_{i+1}$, satisfies $v_i \notin C$ or $v_i \in C \cap \operatorname{Sc}^G(v_{i+1})$,
  (d) fork: $\quad v_{i-1} \longleftarrow v_i \longrightarrow v_{i+1}$, satisfies $v_i \notin C$ or $v_i \in C \cap \operatorname{Sc}^G(v_{i-1}) \cap \operatorname{Sc}^G(v_{i+1})$.

Similar to d-separation we define σ-separation in a DMG.

Definition 4.3 (σ-Separation in a DMG). *Let G be a DMG with set of nodes V. Let $A, B, C \subseteq V$ be subsets.*

1. We say that A and B are σ-connected by C or not σ-separated by C if there exists a walk π (with $n \geq 1$ nodes) in G with one endnode in A and one endnode in B that is C-σ-open. In symbols this statement will be written as follows:

$$A \overset{\sigma}{\underset{G}{\not\perp\!\!\!\perp}} B \,|\, C.$$

2. Otherwise, we will say that A and B are σ-separated by C and write:

$$A \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} B \,|\, C.$$

Remark 4.4. *1. In any DMG we will always have that σ-separation implies d-separation, since every C-d-open walk is also C-σ-open because $\{v\} \subseteq \operatorname{Sc}^G(v)$.*

2. If a DMG G is acyclic, i.e. an ADMG, then σ-separation coincides with d-separation (also called m- or m^*-separation in this context).

It was shown in [10] that σ-separation satisfies the *graphoid/separoid axioms* (see [6, 7, 13, 25]):

Lemma 4.5 (Graphoid and separoid axioms). *Let G be a DMG with set of nodes V and $A, B, C, D \subseteq V$ subsets. Then we have the following rules for σ-separation in G (with $\perp\!\!\!\perp$ standing for $\perp\!\!\!\perp_G^\sigma$):*

---

[4]The stacked edges are meant to be read as an "OR" at each place independently. We also allow for repeated nodes in the walks. Some authors also use the term "path" instead, which other authors use to refer to walks without repeated nodes.

1. *Redundancy:* $A \perp\!\!\!\perp B \,|\, A$ *always holds.*
2. *Symmetry:* $A \perp\!\!\!\perp B \,|\, D \implies B \perp\!\!\!\perp A \,|\, D.$
3. *Decomposition:* $A \perp\!\!\!\perp B \cup C \,|\, D \implies A \perp\!\!\!\perp B \,|\, D.$
4. *Weak Union:* $A \perp\!\!\!\perp B \cup C \,|\, D \implies A \perp\!\!\!\perp B \,|\, C \cup D.$
5. *Contraction:* $(A \perp\!\!\!\perp B \,|\, C \cup D) \wedge (A \perp\!\!\!\perp C \,|\, D)$
   $\implies A \perp\!\!\!\perp B \cup C \,|\, D.$
6. *Intersection:* $(A \perp\!\!\!\perp B \,|\, C \cup D) \wedge (A \perp\!\!\!\perp C \,|\, B \cup D)$
   $\implies A \perp\!\!\!\perp B \cup C \,|\, D,$
   whenever A, B, C, D are pairwise disjoint.
7. *Composition:* $(A \perp\!\!\!\perp B \,|\, D) \wedge (A \perp\!\!\!\perp C \,|\, D)$
   $\implies A \perp\!\!\!\perp B \cup C \,|\, D.$

It was also shown that σ-separation is stable under *marginalization* (see [10, 11]):

Theorem 4.6 (σ-Separation under marginalization, see [10,11]). *Let G be a DMG with set of nodes V. Then for any sets $A, B, C \subseteq V$ and $L \subseteq V \setminus (A \cup B \cup C)$ we have the equivalence:*

$$A \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} B \,|\, C \iff A \underset{G^{\setminus L}}{\overset{\sigma}{\perp\!\!\!\perp}} B \,|\, C,$$

where $G^{\setminus L}$ is the DMG that arises from G by marginalizing out the variables from L.

5 A GLOBAL MARKOV PROPERTY

The most important ingredient for our results is a *generalized directed global Markov property* that relates the graphical structure of any ioSCM M to the conditional independencies of the observed random variables via a σ-separation criterion. Since we have no access to the latent nodes $u \in U$ of an ioSCM with graph G^+ we need to marginalize them out (see Supplementary Material B). This will give us an induced directed mixed graph (DMG) G.

Definition 5.1 (Induced DMG of an ioSCM). *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. The induced directed mixed graph (DMG) G of M is defined as follows:*

1. *G contains all nodes from $V \cup J$.*
2. *G contains all the directed edges of G^+ whose endnodes are both in $V \cup J$.*
3. *G contains the bidirected edge $v \leftrightarrow w$ with $v, w \in V$ if and only if $v \neq w$ and there exists a $u \in U$ with $v, w \in \mathrm{Ch}^{G^+}(u)$, i.e. v and w have a common latent confounder.*

The following *generalized directed global Markov property* directly generalizes from mSCMs (see [10, 11]) to ioSCMs. An alternative version with confounded input is given in C.5.

Theorem 5.2 (σ-Separation criterion). *Let M be an ioSCM with induced DMG G. Then for all subsets*

$A, B, C \subseteq V \cup J$ *we have the implication:*

$$A \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} B \,|\, C \implies X_A \underset{\mathbb{P}}{\perp\!\!\!\perp} X_B \,|\, X_C.$$

In words, if A and B are σ-separated by C in G then the corresponding variables X_A and X_B are conditionally independent given X_C under $\mathbb{P}$, i.e. under the joint distribution $\mathbb{P}_U(X_V \,|\, \mathrm{do}(X_J)) \otimes \mathbb{P}_J(X_J)$ for any product distribution $\mathbb{P}_J = \bigotimes_{j \in J} \mathbb{P}_j$.

Proof. As mentioned, after specifying the product distribution $\mathbb{P}_J$ the ioSCM M constitutes a well-defined mSCM with the same induced DMG G. So the σ-separation criterion for ioSCMs directly follows from the mSCM-version proven in [10, 11]. $\square$

Remark 5.3. *Note that, since σ-separation is stable under marginalization (see [10, 11]), also the σ-separation criterion is stable under marginalization.*

Remark 5.4 (Causal calculus for mechanism change). *The σ-separation criterion 5.2 can be viewed as the causal calculus for mechanism change (also sometimes called "soft" interventions, see [8, 17, 19, 24]). As an example consider $A, B \subseteq V$, $I \subseteq J$. Then the graphical separation $A \perp\!\!\!\perp_G^\sigma I \,|\, B \cup (J \setminus I)$ implies that the conditional probability $\mathbb{P}_U(X_A | X_B, \mathrm{do}(X_J))$ is independent of the actual input variables in I.*

6 THE EXTENDED IOSCM

In this section we want to consider (perfect) interventions onto the observed nodes and improve upon the general rules mentioned in 5.4. For an elegant treatment of this we need to gather for a given ioSCM M all interventional ioSCMs $M_{\mathrm{do}(W)}$, where W runs through all subsets of observed variables, and glue them all together into one big *extended ioSCM* $\hat{M}$. To consider all interventions at once we will need to introduce additional intervention variables I_v to the graph G^+, $v \in V$, which indicate which interventional mechanisms to use. Such techniques were already used in the acyclic case in [21, 22, 24]. The definition will be made in such a way that $\hat{M}$ will still be a well-defined ioSCM. So all the results for ioSCMs will apply to $\hat{M}$, most importantly the σ-separation criterion (Thm. 5.2).

Definition 6.1. *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. The extended ioSCM $\hat{M} = (\hat{G}^+, \hat{\mathcal{X}}, \mathbb{P}_U, \hat{g})$ will be defined as follows:*

1. *For every $v \in V$ define the interventional domain $\mathcal{I}_v := \mathcal{X}_v \dot\cup \{\varnothing_v\}$, where $\varnothing_v$ is a new symbol corresponding to the observational (non-interventional) regime. For a set $A \subseteq V$ we put $\mathcal{I}_A := \prod_{v \in A} \mathcal{I}_v$ and $\varnothing_A := (\varnothing_v)_{v \in A}$.*

2. Let $\hat{G}^+$ be the graph G^+ with the additional intervention nodes I_v and directed edges $I_v \longrightarrow v$ for every $v \in V$. For a uniform notation we sometimes write I_j instead of j for $j \in J$. So we have:
$$\hat{J} := J \cup \{I_v \mid v \in V\} = \{I_w \mid w \in V \dot\cup J\}.$$

3. For every $A \subseteq V$ we will define the mechanism:
$$\hat{g}_A : \; \hat{\mathcal{X}}_{\mathrm{Pa}^{\hat{G}^+}(A) \setminus A} = \mathcal{I}_A \times \mathcal{X}_{\mathrm{Pa}^{G^+}(A) \setminus A} \to \mathcal{X}_A = \hat{\mathcal{X}}_A.$$

First, for $x_A \in \mathcal{I}_A$ we put $I(x_A) := \{v \in A | x_v \neq \varnothing_v\}$. Consider the subgraph of G^+:
$$H(x_A) := (\mathrm{Pa}^{G^+}(A) \cup A)_{\mathrm{do}(I(x_A))}.$$

Then define recursively for $v \in A$:
$$\hat{g}_{A,v}(x_A, x_{\mathrm{Pa}^{G^+}(A) \setminus A})$$
$$:= \begin{cases} x_v & if \quad v \in I(x_A), \\ g_{S,v}(x_{\mathrm{Pa}^{H(x_A)}(S) \setminus S}) & if \quad v \notin I(x_A), \end{cases}$$

where $S := \mathrm{Sc}^{H(x_A)}(v)$ is also a loop in G^+.

4. These functions then are again globally compatible and $\hat{M}$ constitutes a well-defined ioSCM.

5. All the distributions in $\hat{M}$ then are given by the general procedure of ioSCMs (see Def. 2.7). We introduce the notation for $C \subseteq V$ and $(x_C, x_J) \in \mathcal{I}_C \times \mathcal{X}_J$:
$$\mathbb{P}_U(X_V | I_C = x_C, X_J = x_J) :=$$
$$\mathbb{P}_U(X_V | \mathrm{do}((I_C, I_{V \setminus C}, X_J) = (x_C, \varnothing_{V \setminus C}, x_J))).$$

6. The extended DMG $\hat{G}$ of G^+ is then the induced DMG of $\hat{G}^+$, i.e. the induced DMG G with the additional edges $I_v \longrightarrow v$ for every $v \in V$.

The following result now relates the interventional distributions of the ioSCM M with the ones from the extended ioSCM $\hat{M}$. These relations will be used in the following.

Proposition 6.2. Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$ and $\hat{M}$ the extended ioSCM. Let $A, B, C \subseteq V$ be pairwise disjoint set of nodes and $x_{C \cup J} \in \mathcal{X}_{C \cup J}$. Then we have the equations:
$$\mathbb{P}_U(X_A | X_B, \mathrm{do}(X_{C \cup J} = x_{C \cup J}))$$
$$= \mathbb{P}_U(X_A | X_B, I_C = x_C, X_J = x_J)$$
$$= \mathbb{P}_U(X_A | X_B, I_C = x_C, X_C = x_C, X_J = x_J).$$

Proof. This follows from $I(x_C, \varnothing_{V \setminus C}) = C$. See Supplementary Material D.1. □

7 THE THREE MAIN RULES OF CAUSAL CALCULUS

Notation 7.1. *Since everything has been defined in detail in the last section we now want to make use of a simplified and more suggestive notation for better readability.*

1. We identify variables X_A with the set of nodes A.
2. We omit values x_V and the subscript in $\mathbb{P}_U$. E.g. we write $\mathbb{P}(Y | I_T, T, Z, \mathrm{do}(W))$ instead of
$$\mathbb{P}_U(X_Y \mid I_T = x_T, X_T = x_T, X_Z = x_Z,$$
$$\mathrm{do}(X_W = x_W)),$$

where the latter comes from the extended ioSCM of the intervened ioSCM $M_{\mathrm{do}(W)} := M_{\mathrm{do}(W \setminus J)}$ of M.

3. We abbreviate $X_Y \perp\!\!\!\perp_{\mathbb{P}_U(\_| \mathrm{do}(X_W = x_W))} X_T | X_Z$ as $Y \perp\!\!\!\perp_{\mathbb{P}} T | Z, \mathrm{do}(W)$, etc..

4. We write $Y \perp\!\!\!\perp_G^\sigma I_X | X, Z, \mathrm{do}(W)$ to mean $Y \perp\!\!\!\perp_{\hat{G}_{\mathrm{do}(W)}}^\sigma I_X | X, Z$, where $\hat{G}_{\mathrm{do}(W)}$ is the extended DMG of the intervened graph $G^+_{\mathrm{do}(W)}$.

Theorem 7.2 (The three main rules of causal calculus). *Let M be an ioSCM with set of observed nodes V and input nodes J and induced DMG G. Let $X, Y, Z \subseteq V$ and $J \subseteq W \subseteq V \cup J$ be subsets.*

1. *Insertion/deletion of observation:*
   If $\quad Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} X | Z, \mathrm{do}(W) \quad$ then:
   $$\mathbb{P}(Y | X, Z, \mathrm{do}(W)) = \mathbb{P}(Y | Z, \mathrm{do}(W)).$$

2. *Action/observation exchange:*
   If $\quad Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X | X, Z, \mathrm{do}(W) \quad$ then:
   $$\mathbb{P}(Y | \mathrm{do}(X), Z, \mathrm{do}(W)) = \mathbb{P}(Y | X, Z, \mathrm{do}(W)).$$

3. *Insertion/deletion of actions:*
   If $\quad Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X | Z, \mathrm{do}(W) \quad$ then:
   $$\mathbb{P}(Y | \mathrm{do}(X), Z, \mathrm{do}(W)) = \mathbb{P}(Y | Z, \mathrm{do}(W)).$$

The proofs follow directly from the σ-separation criterion 5.2 and Prp. 6.2 applied to the extended ioSCM and can be found in Supplementary Material E.1.

8 ADJUSTMENT CRITERIA

Notation 8.1. *Let $M = (G^+, \mathcal{X}, \mathbb{P}, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. The following set of nodes/variables will play the described roles:*

Y: *the outcome variables,*
X: *the treatment or intervention variables,*
Z_0: *the core set of adjustment variables,*
Z_+: *additional adjustment variables,*
$Z := Z_0 \cup Z_+$: *all actual adjustment variables,*
L: *"marginalizable" adjustment variables,*
C: *context variables,*
W: *default intervention variables containing J,*
S: *variables inducing selection bias given $S = s$.*

We are interested in finding a "$\mathrm{do}(X)$-free" expression for the (conditional) causal effect $\mathbb{P}(Y|C, \mathrm{do}(X), \mathrm{do}(W))$ only using data for C, X, Y, Z that was gathered under selection bias $S = s$ and intervention $\mathrm{do}(W)$ and additional unbiased observational data for C, Z given $\mathrm{do}(W)$. The task can be achieved via the following criterion, which is a generalization of the acyclic case of the selection-backdoor criterion (see [1]), the backdoor criterion (see [21, 22, 24]) and its extensions (also see [4, 26, 27, 32]) to general ioSCMs.

Theorem 8.2 (General adjustment criterion and formula)**.** *Let the setting be like in 8.1. Assume that data was collected under selection bias, $\mathbb{P}(V|S = s, \mathrm{do}(W))$ (or under $\mathbb{P}(V|\mathrm{do}(W))$ and $S = \emptyset$), and there are unbiased samples from $\mathbb{P}(Z|C, \mathrm{do}(W))$. Further assume that the variables satisfy:*

1. *$(Z_0, L) \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X \mid C, \mathrm{do}(W)$, and*
2. *$Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} (I_X, Z_+) \mid C, X, Z_0, L, \mathrm{do}(W)$, and*
3. *$Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} S \mid C, X, Z, \mathrm{do}(W)$, and*
4. *$L \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} X \mid C, Z, \mathrm{do}(W)$.*

Then one can estimate the conditional causal effect $\mathbb{P}(Y|C, \mathrm{do}(X), \mathrm{do}(W))$ via the adjustment formula*:*

$$\mathbb{P}(Y|C, \mathrm{do}(X), \mathrm{do}(W))$$
$$= \int \mathbb{P}(Y|X, Z, C, S = s, \mathrm{do}(W))\, d\mathbb{P}(Z|C, \mathrm{do}(W)).$$

The proof again follows directly from the σ-separation criterion 5.2 and Prp. 6.2 applied to the extended ioSCM and can be found in the Supplementary Material F.1.
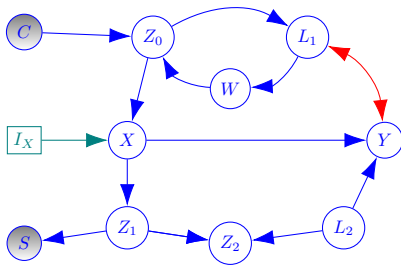

Figure 1: An induced DMG G with input node I_X (the others are left out for readability). The variables satisfy the general adjustment criterion for $\mathbb{P}(Y|C, \mathrm{do}(X))$ with $L = \{L_1, L_2\}$ and $Z_+ = \{Z_1, Z_2\}$. Note that L_2 could also have been a latent variable. Different colours for different node and/or edge types.

Remark 8.3. *Note that the adjustment* formula *in theorem 8.2 does not depend on L. This thus allows us to even choose variables for L that come from an ioSCM*

M' that marginalizes to M, e.g. $L \subseteq U$ or by extending directed edges $v \longrightarrow w$ by $v \longrightarrow \ell \longrightarrow w$ with $\ell \in L$. This technique was used in [32] to find all adjustment sets in the acyclic case with $C = S = \emptyset$.

Corollary 8.4. *Let the notations be like in 8.1 and 8.2 and $W = J = \emptyset$. We have the following special cases, which in the* acyclic *case will reduce to the ones given by the indicated references:*

1. *General selection-backdoor (see [4]): $C = \emptyset$.*
2. *Selection-backdoor (see [1]): $C = L = \emptyset$.*
3. *Extended backdoor (see [26, 32]): $C = S = \emptyset$.*
4. *Backdoor (see [21, 22, 24]): $C = S = L = Z_+ = \emptyset$:*
   (a) *$Z \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X$, and*
   (b) *$Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X \mid X, Z$, implies:*
   $$\mathbb{P}(Y|\mathrm{do}(X)) = \int \mathbb{P}(Y|X, Z)\, d\mathbb{P}(Z).$$

More details can be found in the Supplementary Material F.2. Also a generalization of the criterion for selection without/partial external data of [4, 5] is given there.

Remark 8.5. *The conditions in theorems 7.2, 8.2 and corollary 8.4 are in the acyclic setting usually phrased in terms of sub-structures of the graph G (see [21, 22, 24]):*

1. *For rule 3 in Thm. 7.2 one usually requires $Y \perp\!\!\!\perp^d X|Z, W$ in the graph $G_{\mathrm{do}(W)}$ that is further mutilated on the set $X(Z)$, the set of all X-nodes that are not ancestors of any Z-node in $G_{\mathrm{do}(W)}$.*
2. *For the backdoor criterion instead of $L \perp\!\!\!\perp^d_G I_X$ we could have written that L does not contain any descendent of X; and for $Y \perp\!\!\!\perp^d_G I_X \mid X, Z$ that Z blocks all "backdoor paths" from X to Y.*

We presented the results in the formulaic terms of σ-separation because the relations to their use is directly indicated (e.g. in the proofs), it makes the generalization to ioSCMs possible and when reduced to the acyclic case it will be equivalent to the usual description.

9 IDENTIFYING CAUSAL EFFECTS

Here we extend the *ID algorithm* for the identification of causal effects to ioSCMs. The main references are [12, 14, 15, 24, 29, 34–37]. The task is to decide if a causal effect $\mathbb{P}(Y|\mathrm{do}(W))$ in an ioSCM can be *identified* from (i.e., expressed in terms of) the observational distributions $\mathbb{P}(V|\mathrm{do}(J))$ and the induced graph G. Note that having more dependence structure (like latent confounders, feedback cycles, etc.) will leave us with less identifiable causal effects in general. Due to space limitations, we can only provide here the bare necessities to state the generalized ID algorithm. We assume that the

reader is already familiar with the ID algorithm formulated for ADMGs (for example, the treatment in [36]).

We generalize the notion of districts / C-components:

Definition 9.1 (Consolidated districts). *Let G be a directed mixed graph (DMG) with set of nodes V. Let $v \in V$. The* consolidated district $\mathrm{Cd}^G(v)$ *of v in G is given by all nodes $w \in V$ for which there exist $k \geq 1$ nodes $(v_1, \ldots, v_k)$ in G such that $v_1 = v$, $v_k = w$ and for $i = 2, \ldots, k$ we have that the bidirected edge $v_{i-1} \leftrightarrow v_i$ is in G or that $v_i \in \mathrm{Sc}^G(v_{i-1})$. For $B \subseteq V$ we write $\mathrm{Cd}^G(B) := \bigcup_{v \in B} \mathrm{Cd}^G(v)$. Let $\mathcal{CD}(G)$ be the set of consolidated districts of G.*

We also generalize the notion of topological order:

Definition 9.2 (Apt-order, see [10]). *Let G be a DMG with set of nodes V. An* assembling pseudo-topological order (apt-order) *of G is a total order $<$ on V with the following two properties:*

1. *For every $v, w \in V$ we have:*
$$w \in \mathrm{Anc}^G(v) \setminus \mathrm{Sc}^G(v) \implies w < v.$$

2. *For every $v_1, v_2, w \in V$ we have:*
$$v_2 \in \mathrm{Sc}^G(v_1) \wedge (v_1 \leq w \leq v_2) \implies w \in \mathrm{Sc}^G(v_1).$$

Remark 9.3. *Let G be a DMG.*

1. *If G is acyclic then an apt-order $<$ is the same as a topological order (i.e. $w \in \mathrm{Pa}^G(v) \implies w < v$).*
2. *If G has a topological order then G is acyclic.*
3. *For any DMG G there always exists an apt-order $<$ (in contrast to topological orders).*

Notation 9.4. *Let G be a DMG with set of nodes V and $<$ a apt-order on G. For elements $v \in V$ and subsets $B \subseteq V$ we put:*

1. $\mathrm{Pred}^G_<(v) := \{w \in V \mid w < v\}$,
2. $\mathrm{Pred}^G_\leq(v) := \{w \in V \mid w = v \text{ or } w < v\}$,
3. $\mathrm{Pred}^G_\leq(B) := \bigcup_{v \in B} \mathrm{Pred}^G_\leq(v)$,
4. $\mathrm{Pred}^G_<(B) := \mathrm{Pred}^G_\leq(B) \setminus B$.

Remark 9.5. *If B is strongly-connected, then $\mathrm{Pred}^G_\leq(B)$ is ancestral in G, i.e., $\mathrm{Anc}^G(\mathrm{Pred}^G_\leq(B)) = \mathrm{Pred}^G_\leq(B)$.*

The notion of input variables enables the following convenient and intuitive construction:

Definition 9.6 (Sub-ioSCMs). *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. For $C \subseteq V$ non-empty define the ioSCM $M_{[C]}$ as follows:*

1. *Put $G^+_{[C]}$ to be the subgraph of $G^+_{\mathrm{do}(\mathrm{Pa}^G(C) \setminus C)}$ induced by $C \cup \mathrm{Pa}^{G^+}(C)$.*
2. $V_{[C]} := C$, $J_{[C]} := \mathrm{Pa}^{G^+}(C) \setminus (C \cup U)$, $U_{[C]} := U \cap \mathrm{Pa}^{G^+}(C)$.

3. *Keep all functions g_S with $S \subseteq C$.*
4. $\mathbb{P}_{U_{[C]}} := \bigotimes_{u \in U_{[C]}} \mathbb{P}_u$, *i.e. the marginal of $\mathbb{P}_U$ and we will use the notation $\mathbb{P}_U$ (or just $\mathbb{P}$) for both.*

For $C \subseteq V \cup J$ with $C \cap V \neq \emptyset$ put $M_{[C]} := M_{[C \cap V]}$.

By the definition of the random variables induced by an ioSCM we immediately get the following basic result:

Lemma 9.7. *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. For $C \subseteq V$, we have (indices for emphasis):*

$$\mathbb{P}_{M_{[C]}}(C \mid \mathrm{do}(\mathrm{Pa}^G(C) \setminus C)) = \mathbb{P}_M(C \mid \mathrm{do}(J \cup W)),$$

for any $W \subseteq V \setminus C$ that contains $(\mathrm{Pa}^G(C) \cap V) \setminus C$. As a special case: if $A \subseteq G$ is ancestral, i.e., $\mathrm{Anc}^G(A) = A$,

$$\mathbb{P}_{M_{[A]}}(A \cap V \mid \mathrm{do}(A \cap J)) = \mathbb{P}_M(A \cap V \mid \mathrm{do}(J \cup W))$$

for any $W \subseteq V \setminus A \cap V$.

The ID algorithm works by repeatedly applying the previous lemma and the following rules:

Proposition 9.8. *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$ and $<$ an apt-order for G^+.*

1.
$$\mathbb{P}(V \mid \mathrm{do}(J)) = \bigotimes_{\substack{S \in \mathcal{S}(G) \\ S \subseteq V}} \mathbb{P}(S \mid \mathrm{Pred}^G_<(S) \cap V, \mathrm{do}(J)).$$

2. *For $S \subseteq V$ a strongly connected component of G, $D := \mathrm{Cd}^G(S)$ its consolidated district in G and $P := \mathrm{Pa}^G(D) \setminus D$:*

$$\mathbb{P}_M(S \mid \mathrm{Pred}^G_<(S) \cap V, \mathrm{do}(J))$$
$$= \mathbb{P}_{M_{[D]}}(S \mid \mathrm{Pred}^{G_{[D]}}_<(S) \cap D, \mathrm{do}(P)).$$

3. *For $D \subseteq V$ a consolidated district of G:*

$$\mathbb{P}(D \mid \mathrm{do}(J \cup V \setminus D))$$
$$= \bigotimes_{\substack{S \in \mathcal{S}(G) \\ S \subseteq D}} \mathbb{P}(S \mid \mathrm{Pred}^G_<(S) \cap V, \mathrm{do}(J)).$$

Proof. 1. uses the chain rule; 2. is proved in Supplementary Material G.2; 3. is shown by applying 1. and Remark 9.7 to $G_{[D]}$ and then making use of 2.. $\qquad\square$

Remark 9.9. *Naively putting the equations of Prp. 9.8 into each other would give us the equation:*

$$\mathbb{P}(V \mid \mathrm{do}(J)) = \left[\bigotimes_{\substack{D \in \mathcal{CD}(G) \\ D \subseteq V}} \right] \mathbb{P}(D \mid \mathrm{do}(J \cup V \setminus D)).$$

Note that the product might not be well-defined as the consolidated districts i.g. are not totally ordered by $<$

(in contrast to strongly connected components), even in the acyclic case. For example, consider the graph:
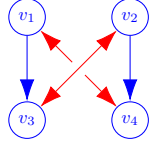
This problem is usually not addressed in the literature. The problem disappears if every strongly connected component $S \subseteq V$ comes with a measure μ_S such that $\mathbb{P}(V \,|\, \mathrm{do}(J))$ has a density w.r.t. the product measure $\bigotimes_{\substack{S \in \mathcal{S}(G) \\ S \subseteq V}} \mu_S$. Then the densities $p(D \,|\, \mathrm{do}(J \cup V \setminus D))$ can be multiplied in any order and the integration can be separately done via the μ_S in reverse order of $<$.

We now have all the prerequisites to state the generalized ID algorithm (Algorithm 1) and prove its correctness:

Theorem 9.10 (Consequence of 9.8, 9.9). *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$ with set of observed nodes V and input nodes J and distributions $\mathbb{P}(V \,|\, \mathrm{do}(J))$. Let $<$ be an apt-order for G^+. Assume that for every strongly connected component $S \subseteq V$ we have a measure μ_S such that $\mathbb{P}(V \,|\, \mathrm{do}(J))$ has a density w.r.t. the product measure $\bigotimes_{\substack{S \in \mathcal{S}(G) \\ S \subseteq V}} \mu_S$. Let $Y \subseteq V$ and $W \subseteq J \cup V$ be subsets. If the extended ID algorithm (see Algorithm 1) does not "FAIL" then the causal effect $\mathbb{P}(Y \,|\, \mathrm{do}(W))$ is identifiable, i.e. it can be computed from $\mathbb{P}(V \,|\, \mathrm{do}(J))$ alone, and the expression is obtained by postprocessing the output of the algorithm.*

Remark 9.11. *1. We make no claim about the completeness of the algorithm here.*

2. The algorithm reduces to the usual version in the acyclic case (see [29, 35–37]).

3. The main idea of the generalized ID algorithm is to exploit that the causal effects onto ancestral subsets and consolidated districts are identifiable. The algorithm then alternates these constructions to shrink towards the queried set C until convergence, i.e. until a set A is reached that is both the ancestral closure of C and a consolidated district in itself. If $C = A$ then the causal effect onto C is identifiable, otherwise it outputs "FAIL" as no shrinking can be done with these techniques anymore. Also see Supplementary Material G.1.

10 CONCLUSION

We proved the three main rules of causal calculus and general adjustment criteria with corresponding formulas to recover from interventions and selection bias

---

Algorithm 1 ID: Generalized ID algorithm for the identification of causal effects in general ioSCMs.

1: **function** ID$(G, Y, W, \mathbb{P}(V \,|\, \mathrm{do}(J)))$
2:    **require:** $Y \subseteq V$, $W \subseteq V$, $Y \cap W = \emptyset$
3:    $H \leftarrow \mathrm{Anc}^{G_{V \setminus W}}(Y)$
4:    **for** $C \in \mathcal{CD}(H)$ **do**
5:       $Q[C] \leftarrow \mathrm{IDCD}(G, C, \mathrm{Cd}^G(C), Q[\mathrm{Cd}^G(C)])$
6:       **if** $Q[C] = \mathrm{FAIL}$ **then**
7:          **return** FAIL
8:       **end if**
9:    **end for**
10:    $Q[H] \leftarrow \left[\bigotimes_{C \in \mathcal{CD}(H)}\right] Q[C]$
11:    **return** $\mathbb{P}(Y \,|\, \mathrm{do}(J, W)) = \int Q[H] dx_{H \setminus Y}$
12: **end function**

13: **function** IDCD$(G, C, D, Q[D])$
14:    **require:** $C \subseteq D \subseteq V$, $\mathcal{CD}(G_D) = \{D\}$
15:    $A \leftarrow \mathrm{Anc}^{G_{[D]}}(C) \cap D$
16:    $Q[A] \leftarrow \int Q[D] d(x_{D \setminus A})$
17:    **if** $A = C$ **then**
18:       **return** $Q[A]$
19:    **else if** $A = D$ **then**
20:       **return** FAIL
21:    **else if** $C \subsetneq A \subsetneq D$ **then**
22:       **for** $S \in \mathcal{S}(G_{[A]})$ s.t. $S \subseteq \mathrm{Cd}^{G_{[A]}}(C)$ **do**
23:          $R_A[S] \leftarrow \mathbb{P}(S | \mathrm{Pred}^G_<(S) \cap A, \mathrm{do}(J \cup V \setminus A))$
24:       **end for**
25:       $Q[\mathrm{Cd}^{G_{[A]}}(C)] \leftarrow \bigotimes_{\substack{S \in \mathcal{S}(G_{[A]}) \\ S \subseteq \mathrm{Cd}^{G_{[A]}}(C)}} R_A[S]$
26:       **return** IDCD$(G, C, \mathrm{Cd}^{G_{[A]}}(C), Q[\mathrm{Cd}^{G_{[A]}}(C)])$
27:    **end if**
28: **end function**

---

for general ioSCMs, which allow for arbitrary probability distributions, non-/linear functional relations, latent confounders, external non-/probabilistic parameter/action/intervention/context/input nodes and *cycles*. This generalizes all the corresponding results of *acyclic* causal models (see [1, 4, 21, 22, 24, 26, 27, 32]) to general ioSCMs. We also showed how to extend the ID algorithm for the identification of causal effects from the acyclic setting to general ioSCMs. In supplementary material A we also show how to do counterfactual reasoning in ioSCMs. Future work might address completeness questions of the ID algorithm (see [14, 24, 33, 34]).

References

[1] E. Bareinboim, J. Tian, and J. Pearl. Recovering from Selection Bias in Causal and Statistical Inference. In *AAAI*, 2014.

[2] S. Bongers, T. Blom, and J. M. Mooij. From Random Differential Equations to Structural Causal Models: the stochastic case. *https://arxiv.org/abs/1803.08784*, 2018.

[3] P. Constantinou and A. P. Dawid. Extended conditional independence and applications in causal inference. *Ann. Statist.*, 45(6):2618–2653, 2017.

[4] J. D. Correa and E. Bareinboim. Causal Effect Identification by Adjustment under Confounding and Selection Biases. In *AAAI*, 2017.

[5] J. D. Correa, J. Tian, and E. Bareinboim. Generalized Adjustment Under Confounding and Selection Biases. In *AAAI*, 2018.

[6] A.P. Dawid. Conditional independence in statistical theory. *J. R. Stat. Soc., Ser. B*, 41:1–31, 1979.

[7] A.P. Dawid. Separoids: a mathematical framework for conditional independence and irrelevance. *Ann. Math. Artif. Intell.*, 32(1-4):335–372, 2001.

[8] D. Eaton and K. P. Murphy. Bayesian structure learning using dynamic programming and MCMC. In *UAI*, 2007.

[9] R. J. Evans. Graphs for margins of Bayesian networks. *Scand. J. Stat.*, 43(3):625–648, 2016.

[10] P. Forré and J. M. Mooij. Markov Properties for Graphical Models with Cycles and Latent Variables. *https://arxiv.org/abs/1710.08775*, 2017.

[11] P. Forré and J. M. Mooij. Constraint-based Causal Discovery for Non-Linear Structural Causal Models with Cycles and Latent Confounders. In *UAI*, 2018.

[12] D. Galles and J. Pearl. Testing Identifiability of Causal Effects. In *UAI*, 1995.

[13] D. Geiger, T.S. Verma, and J. Pearl. Identifying Independence in Bayesian Networks. *Networks*, 20(5):507–534, 1990.

[14] Y. Huang and M. Valtorta. Pearl's calculus of intervention is complete. In *UAI*, 2006.

[15] Y. Huang and M. Valtorta. On the completeness of an identifiability algorithm for semi-Markovian models. *Annals of Mathematics and Artificial Intelligence*, 54(4):363–408, 2008.

[16] Achim Klenke. *Probability Theory - A Comprehensive Course*. Universitext. Springer, 2nd edition, 2014.

[17] F. Markowetz, S. Grossmann, and R. Spang. Probabilistic Soft Interventions in Conditional Gaussian Networks. In *AISTATS-05*, 2005.

[18] J. M. Mooij, D. Janzing, and B. Schölkopf. From Ordinary Differential Equations to Structural Causal Models: the deterministic case. In *UAI*, 2013.

[19] J. M. Mooij, S. Magliacane, and T. Claassen. Joint Causal Inference from Multiple Contexts. *https://arxiv.org/abs/1611.10351v3*, 2018.

[20] J. Pearl. Fusion, propagation and structuring in belief networks. Technical Report 3, UCLA Computer Science Department, 1986. Technical Report 850022 (R-42).

[21] J. Pearl. Aspects of graphical models connected with causality. In *In Proceedings of the 49th Session of the International Statistical Institute*, pages 391–401, 1993.

[22] J. Pearl. Comment: Graphical Models, Causality, And Intervention. *Statistical Science*, 8(3):266–269, 1993.

[23] J. Pearl. Causal diagrams for empirical research (with discussion). *Biometrika*, 82(4):669–710, 1995.

[24] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

[25] J. Pearl and A. Paz. Graphoids: a graph-based logic for reasoning about relevance relations. *Advances in Artificial Intelligence-II*, pages 357–363, 1987.

[26] J. Pearl and A. Paz. Confounding equivalence in causal inference. In *UAI*, 2010.

[27] E. Perkovic, J. Textor, M. Kalisch, and M. Maathuis. A complete generalized adjustment criterion. In *UAI*, 2015.

[28] T. Richardson. Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.*, 30(1):145–157, 2003.

[29] T. S. Richardson, R. J. Evans, J. M. Robins, and I. Shpitser. Nested Markov Properties for Acyclic Directed Mixed Graphs. *https://arxiv.org/abs/1701.06686*, 2017.

[30] T. S. Richardson and J. M. Robins. Single world intervention graphs: a primer. In *UAI workshop*, 2013.

[31] P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55, 1983.

[32] I. Shpitser, T. J. Van der Weele, and J. M. Robins. On the validity of covariate adjustment for estimating causal effects. In *UAI*, 2010.

[33] I. Shpitser and J. Pearl. Identification of conditional interventional distributions. In *UAI*, 2006.

[34] I. Shpitser and J. Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *AAAI*, 2006.

[35] J. Tian. *Studies in Causal Reasoning and Learning*. PhD thesis, University of California, Los Angeles, 2002.

[36] J. Tian. Identifying Conditional Causal Effects. In *UAI*, 2004.

[37] J. Tian and J. Pearl. A General Identification Condition for Causal Effects. In *AAAI*, 2002.

[38] T. S. Verma and J. Pearl. Causal Networks: Semantics and Expressiveness. In *UAI*, 1990.