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

Forré, P.; Mooij, J.M.

Publication date
2019

Document Version
Author accepted manuscript

Published in
Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence

Citation for published version (APA):
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 additional 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 $\sigma$-separation criterion in the underlying graph, where $\sigma$-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 $\sigma$-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 definitions for ioSCMs to extend the ID algorithm for the identification of causal effects to the ioSCM setting, which reduces to the usual one in the acyclic case.

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 $v_1 \rightarrow \cdots \rightarrow v_2$ and $v_2 \rightarrow \cdots \rightarrow v_1$ 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 $v \rightarrow v \in E$ or not).
2. The set of loops of $G$ is written as $L(G)$.
3. The strongly connected component of $v$ in $G$ is defined to be: $Sc^G(v) := Anc^G(v) \cap Desc^G(v)$.
4. The set of strongly connected components is $S(G)$.

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

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

In the following all spaces are meant to be equipped with $\sigma$-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 $V^+ = V \cup U \cup J$, 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 $X$, for every $v \in V^+$, $X := \prod_{u \in V^+ \cup J} X_u$.
3. a product probability measure $\mathbb{P}_U = \bigotimes_{u \in U} \mathbb{P}_u$ on the latent space $X_U := \prod_{u \in U} X_u$.
4. a directed graph structure $G^+ = (V^+, E^+)$ with the properties:
   (a) $V = Ch^G^+ (U \cup J)$,
   (b) $Pa^G^+ (U \cup J) = \emptyset$, where $Ch^G^+$ and $Pa^G^+$ stand for children and parents in $G^+$, resp.
5. a system of causal mechanisms $g = \{gs\}_{S \subseteq L(G^+)}$: $S \subseteq V$

$$gs : \prod_{v \in Pa^{G^+}(S) \cup S} X_v \rightarrow \prod_{v \in S} X_v,$$

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

$$gs(x_{Pa^{G^+}(S) \setminus S'} \cup S) = x_S \implies gs(x_{Pa^{G^+}(S') \setminus S} \cup S') = x_{S'},$$

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

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

2Note that the index set runs over all “observable loops” $S \subseteq V$, $S \in L(G^+)$, not just the sets $\{v\}$ for $v \in V$. 
The ioSCM will be denoted by $M = (G^+, X, P_U, g)$.

**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 $g_{i_1} : X_{i_1} \to X_{i_2}$ for $i_1$ corresponding to $i_2$, e.g. $g_{i_2} = \text{id}$ if possible. We can now “glue” them together to get a new ioSCM $M_3$ given by $V_3 := V_1 \cup V_2 \cup J_2$, $P_{U} := U_1 \cup U_2$, $J_3 = J_1 \cup J_2 \setminus I_2$ and $G^+_3 := G^+_1 \cup G^+_2$, where we add the the edges $i_1 \to i_2$ and the mechanisms $g_{i_1}$ and $P_{U_i} := P_{U_1} \otimes P_{U_2}$.

**Example 2.6 (Constructing mSCMs from ioSCMs).** Given an ioSCM $M = (G^+, X, P_U, g)$ with graph $G^+ = (V \cup U \cup J, E^+)$ we can construct a well-defined mSCM by specifying a product distribution $P_J = \bigotimes_{j \in J} P_j$ on $X_J$ and the remaining functions then are clearly globally compatible and we get a well-defined ioSCM $M_{do}(W)$.

The actual joint distributions on the observed space $X_V$ and thus the random variables attached to any ioSCM will be defined in the following.

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

1. The latent variables are given by $(X_u)_{u \in U} \sim P_U$, i.e. by the canonical projections $X_u : X_U \to X_u$, which are jointly $P_U$-independent. We put $X_{do}(x_J) := X_u$, i.e., independent of $x_J$.
2. For $j \in J$ we put $X_{do}(x_J) := x_j$, the constant variable given by the $j$-component of $x_J$.
3. The observed variables $(X_{do}(x_J))_{x \in V}$ are inductively defined by:
   
   $$X_{do}(x_J) := g_{S,v}(\{(X_w)_{w \in Pa^{G^+}(S) \setminus S}\}),$$
   
   where $S := \text{Sc}^+ \{v\} := \text{Anc}^+ \{v\} \cap \text{Desc}^+ \{v\}$ 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:
   
   $$X_S := g_S(X_{do}(x_J)),$$
   
   where we put $X_{do}(x_J) := \prod_{v \in A, x \in A, u \in A}$ for subsets $A$.
5. We define the family of conditional distributions:
   
   $$P_U(X_A | X_B, X_J = x_J) := P_U(X_A | X_B, do(X_J = x_J)) := P_U(X_{do}(x_J) | X_B),$$

for $A, B \subseteq V$ and $x_J \in X_J$. 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 $P_J = \bigotimes_{j \in J} P_j$ on $X_J$, then we get a joint distribution $P$ on $X_{U \cup J}$ by setting:
   
   $$P(X_V, X_J) := P_U(X_{do}(x_J)) \otimes P_J(X_J).$$

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

$$X^{do}(x_J) := g_A(X^{do}(x_J)).$$

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

**Definition 2.9 (Marginalization of ioSCMs).** Let $M = (G^+, X, P_U, g)$ be an ioSCM with $G^+ = (V \cup U \cup J, E^+)$ and $W \subseteq V$ a subset. The marginalized ioSCM $M^{W \setminus W}$ 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 $w \in Pa^{G^+}(S) \setminus S$):

$$g_{S,v}(x_{Pa^{G^+}(S) \setminus S}^w); := g_{S,v}(x_{Pa^{G^+}(S) \setminus S}; g_{w}(x_{Pa^{G^+}(w) \setminus w}),$$

where $(G^+)^W$ is the marginalized graph of $G^+$ (see Supplementary Material B), $S' \subseteq V^W := V \setminus W$ is any loop of $(G^+)^W$ 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 $M = (G^+, X, P_U, g)$ be an ioSCM with $G^+ = (V \cup U \cup J, E^+)$. Let $W \subseteq V \cup J$ be a subset. We then define the post-interventional ioSCM $M_{do(W)}$ w.r.t. $W$:

1. Define the graph $G^+_{do(W)}$ by removing all the edges $v \to w$ for all nodes $w \in W$ and $v \in Pa^{G^+}(w)$.
2. Put $V_{do(W)} := V \setminus W$ and $J_{do(W)} := J \cup W$.
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 $M_{do(W)}$. 


3 CONDITIONAL INDEPENDENCE

Here we generalize conditional independence for structured families of distributions. The main application will be the distributions \( \{P(x_j | \mathbf{do}(x_j = x_{ij})) \}_{x_j \in X_j} \) coming from isoSCMs, but the following definition might be of more general importance.

**Definition 3.1** (Conditional independence). Let \( X_V := \prod_{v \in V} X_v \) and \( X_J := \prod_{j \in J} X_j \) be product spaces and \( \mathbb{P} := (\mathbb{P}_V (X_V | x_J))_{x_J \in X_j} \) a family of distributions on \( X_V \) (measurably\(^3\)) parametrized by \( X_J \). For subsets \( A, B, C \subseteq V \cup J \) we write:

\[
X_A \perp \perp X_B | X_C
\]

if and only if for every product distribution \( \mathbb{P}_J = \otimes_{j \in J} \mathbb{P}_j \) on \( X_J \) we have:

\[
\mathbb{P}_{V \cup J}(x_A | x_B, x_C) = \mathbb{P}_{V \cup J}(x_A | x_C) \quad \mathbb{P}_{V \cup J-a.s.}
\]

\[\text{where } \mathbb{P}_{V \cup J}(x_V | x_J) := \mathbb{P}_V (x_V | x_J) \otimes \mathbb{P}_J (x_J) \text{ is the distribution given by } x_J \sim \mathbb{P}_J \text{ and then } x_V \sim \mathbb{P}_V (\_ | x_J) .\]

**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. \( \perp \perp \) satisfies the separoid axioms (see [6, 7, 13, 25]) or see rules 1-5 in Lem. 4.5 for \( \perp \perp \) as these rules are preserved under conjunction.

4 \( \sigma \)-SEPARATION

In this section we will define \( \sigma \)-separation on directed mixed graphs (DMG) and present the generalized directed global Markov property stating that every isoSCM will entail the conditional independencies that come from \( \sigma \)-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 (\( \rightarrow \)) and bidirected edges (\( \leftrightarrow \)). In case \( G \) contains no directed cycles it is called an acyclic directed mixed graph (ADMG).

\[^1\text{We require that for every measurable } F \subseteq X_V \text{ the map } X_I \rightarrow [0,1] \text{ given by } x_I \mapsto \mathbb{P}_V (X_V \in F | x_J) \text{ is measurable. Such families of distributions are also called channels or (stochastic) Markov (transition) kernels (see [16]).}

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

\[
v_1 \leftrightarrow \cdots \leftrightarrow v_n\(^4\).\]

The walk will be called \( C\)-\( \sigma \)-open if:

1. the endnodes \( v_1, v_n \notin C \), and
2. every triple of adjacent nodes in \( \pi \) that is of the form:
   (a) collider: \( v_{i-1} \not\leftrightarrow v_i \not\leftrightarrow v_{i+1} \), satisfies \( v_i \in C \),
   (b) left chain: \( v_{i-1} \not\leftrightarrow v_i \leftrightarrow v_{i+1} \), satisfies \( v_i \notin C \) or \( v_i \in C \cap \text{Sc}_G(v_{i-1}) \),
   (c) right chain: \( v_{i-1} \leftrightarrow v_i \not\leftrightarrow v_{i+1} \), satisfies \( v_i \notin C \) or \( v_i \in C \cap \text{Sc}_G(v_{i+1}) \),
   (d) fork: \( v_{i-1} \leftrightarrow v_i \leftrightarrow v_{i+1} \), satisfies \( v_i \notin C \) or \( v_i \in C \cap \text{Sc}_G(v_{i-1}) \cap \text{Sc}_G(v_{i+1}) \).

Similar to d-separation we define \( \sigma \)-separation in a DMG.

**Definition 4.3** (\( \sigma \)-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 \( \sigma \)-connected by \( C \) or not \( \sigma \)-separated by \( C \) if there exists a walk \( \pi \) (with \( n \geq 1 \) nodes) in \( G \) with one endnode in \( A \) and one endnode in \( B \) that is \( C\)-\( \sigma \)-open. In symbols this statement will be written as follows:

\[
\overset{A \uparrow C}{\underset{G}{\sigma}} B.
\]

2. Otherwise, we will say that \( A \) and \( B \) are \( \sigma \)-separated by \( C \) and write:

\[
\overset{A \not\sigma C}{\underset{G}{\sigma}} B.
\]

**Remark 4.4.**

1. In any DMG we will always have that \( \sigma \)-separation implies d-separation, since every \( C\)-d-open walk is also \( C\)-\( \sigma \)-open because \( \{v\} \subseteq \text{Sc}_G(v) \).
2. If a DMG \( G \) is acyclic, i.e. an ADMG, then \( \sigma \)-separation coincides with d-separation (also called m- or m' -separation in this context).

It was shown in [10] that \( \sigma \)-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 \( \sigma \)-separation in \( G \) (with \( \perp \perp \) standing for \( \perp \perp \)):

\[^4\text{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: $A \perp B \mid A$ always holds.
2. Symmetry: $A \perp B \mid D \Rightarrow B \perp A \mid D$.
3. Decomposition: $A \perp B \cup C \mid D \Rightarrow A \perp B \mid C \cup D$.
4. Weak Union: $A \perp B \cup C \mid D \Rightarrow A \perp B \mid C \cup D$.
5. Contraction: $(A \perp B \mid C \cup D) \land (A \perp C \mid D) \Rightarrow A \perp B \mid C \cup D$.
6. Intersection: $(A \perp B \mid C \cup D) \land (A \perp C \mid B \cup D) \Rightarrow A \perp B \mid C \cup D$.
7. Composition: $(A \perp B \mid D) \land (A \perp C \mid D) \Rightarrow A \perp B \cup C \mid D$.

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

**Theorem 4.6** ($\sigma$-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:

$$A \overset{\sigma}{\perp} B \mid C \iff A \overset{G}{\perp} B \mid C,$$

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 $\sigma$-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 $M = (G^+, \mathcal{X}, \mathcal{P}_U, \mathcal{g})$ be an ioSCM with $G^+ = (V \cup \bar{U} \cup J, E^+)$. 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 $v,w \in \text{Ch}^G(u)$, 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** ($\sigma$-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:

$$A \overset{\sigma}{\perp} B \mid C \iff X_A \overset{\dag}{\perp} X_B \mid X_C.$$

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

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

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

**Remark 5.4** (Causal calculus for mechanism change). The $\sigma$-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 $A, B \subseteq V, \ G \subseteq J$ and $\tilde{J}$ the graphical separation $A \perp_{G,\tilde{J}} B \mid \tilde{J} \cup (J \setminus I)$ implies that the conditional probability $\mathcal{P}_U(X_A|X_B, \text{do}(X_J))$ 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 $M_{\text{do}(W)}$, 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 $\sigma$-separation criterion (Thm. 5.2).

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

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


2. Let \( \hat{G}^+ \) be the graph \( G^+ \) with the additional intervention nodes \( I_v \) and directed edges \( I_v \xrightarrow{\sigma} v \) for every \( v \in V \). For a uniform notation we sometimes write \( I_j \) instead of \( j \) for \( j \in J \). So we have:
\[
J := \cup \{ I_v | v \in V \} = \{ I_w | w \in V \cup J \}.
\]
3. For every \( A \subseteq V \) we will define the mechanism:
\[
\hat{g}_A : \hat{X}_{Pa^+(A) \setminus A} = \mathcal{I} \times \hat{X}_{Pa^+(A) \setminus A} \to \hat{X}_A = \hat{X}.
\]
First, for \( x_A \in \mathcal{I} \) we put \( I(x_A) := \{ v \in A | x_v \neq \sigma_v \} \). Consider the subgraph of \( G^+ \):
\[
H(x_A) := (Pa^+(A) \cup A)_{do((I(x_A)))}.
\]
Then define recursively for \( v \in A \):
\[
\hat{g}_{A,v}(x_A, Pa^+(A) \setminus A) := \begin{cases} 
  x_v & \text{if } v \in I(x_A), \\
  g_{sv}(x_{Pa^+(A) \setminus A}(S), x_A) & \text{if } v \notin I(x_A),
\end{cases}
\]
where \( S := S_{I(x_A)}(v) \) 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 \( (x_C, x_J) \in \mathcal{I} \times \mathcal{X}_J ; \)
\[
P_{U}(X_V|I_C = x_C, X_J = x_J) := P_{U}(X_V|do((I_C, I_J \setminus C, X_J = x_J)).
\]
6. The extended DMG \( \hat{G} \) of \( G^+ \) is then the induced DMG of \( \hat{G}^+ \), i.e. the induced DMG \( \hat{G} \) with the additional edges \( I_v \xrightarrow{\sigma} v \) for every \( v \in V \).

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

**Proposition 6.2.** Let \( M = (G^+, \mathcal{X}, \mathcal{P}, g) \) be an ioSCM with \( G^+ = (V \cup \{ U \cup J, E^+ \}) \) and \( \hat{M} \) the extended ioSCM. Let \( A, B, C \subseteq V \) be pairwise disjoint set of nodes and \( x_{C \cup J} \in \hat{X}_{C \cup J} \). Then we have the equations:
\[
P_{U}(X_A|X_B, do(X_{C \cup J} = x_{C \cup J})) = P_{U}(X_A|X_B, I_C = x_C, X_J = x_J).
\]
\[
P_{U}(X_A|X_B, I_C = x_C, X_C = x_C, X_J = x_J).
\]
Proof. This follows from \( I(x_C, \mathcal{G} \setminus C) = C \). 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 \( \mathcal{P}_U \). E.g. we write \( \mathcal{P}(Y|I_T, T, Z, do(W)) \) instead of
\[
\mathcal{P}_U(X_Y|I_T = x_T, X_T = x_T, X_Z = x_Z, do(W = x_W)),
\]
where the latter comes from the extended ioSCM of the intervened ioSCM \( \mathcal{M}_{do(W)} := \mathcal{M}_{do(W \setminus J)} \) of \( M \).
3. We abbreviate \( X_Y \perp \sigma_{T, do(X_W = x_W)} X_T | X_Z \) as \( Y \sigma_{T, do(W)} X_T | X_Z \), etc.
4. We write \( Y \sigma_{do(W)} X, Z, do(W) \) to mean \( Y \sigma_{G_{do(W)}} X | X, Z \), where \( G_{do(W)} \) is the extended DMG of the intervened graph \( G_{do(W)}^+ \).

**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:
\[
\sigma_{M} Y \xrightarrow{\mathcal{P}} X, Z, do(W) \quad \text{then:} \quad \mathcal{P}(Y|X, Z, do(W)) = \mathcal{P}(Y|Z, do(W)).
\]
2. Action/observation exchange:
\[
\sigma_{M} Y \xrightarrow{\mathcal{P}} X | X, Z, do(W) \quad \text{then:} \quad \mathcal{P}(Y|do(X), Z, do(W)) = \mathcal{P}(Y|X, Z, do(W)).
\]
3. Insertion/deletion of actions:
\[
\sigma_{M} Y \xrightarrow{\mathcal{P}} I_X | Z, do(W) \quad \text{then:} \quad \mathcal{P}(Y|do(X), Z, do(W)) = \mathcal{P}(Y|Z, do(W)).
\]
The proofs follow directly from the \( \sigma \)-separation criterion 5.2 and Prop. 6.2 applied to the extended ioSCM and can be found in Supplementary Material E.1.

8. **ADJUSTMENT CRITERIA**

**Notation 8.1.** Let \( M = (G^+, \mathcal{X}, \mathcal{P}, g) \) be an ioSCM with \( G^+ = (V \cup \{ U \cup J, E^+ \}) \). 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,
- \( Z_a \): additional adjustment variables,
- \( Z := Z_0 \cup Z_a \): 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 "do(X)-free" expression for the (conditional) causal effect \( P(Y|C, do(X), do(W)) \) only using data for \( C, X, Y, Z \) that was gathered under selection bias \( S = s \) and intervention \( do(W) \) and additional unbiased observational data for \( C, Z \) given \( do(W) \). 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, \( P(V|S = s, do(W)) \) (or under \( P(V|do(W)) \) and \( S = \emptyset \)), and there are unbiased samples from \( P(Z|C, do(W)) \). Further assume that the variables satisfy:

1. \((Z_0, L) \vdash_{G} I_X | C, do(W), \) and
2. \( Y \vdash_{G} (I_X, Z_+) | C, X, Z_0, L, do(W), \) and
3. \( Y \vdash_{G} S | C, X, Z, do(W), \) and
4. \( L \vdash_{G} X | C, Z, do(W). \)

Then one can estimate the conditional causal effect \( P(Y|C, do(X), do(W)) \) via the adjustment formula:

\[
P(Y|C, do(X), do(W)) = \int P(Y|X, Z, S = s, do(W)) dP(Z|C, do(W)).
\]

The proof again follows directly from the \( \sigma \)-separation criterion 5.2 and Prp. 6.2 applied to the extended IOSCM and can be found in the Supplementary Material F.1.

![Diagram](image)

**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 \( P(Y|C, do(X)) \) with \( L = \{L_1, I_L\} \) 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 \( v \rightarrow w \) by \( v \rightarrow \ell \rightarrow w \) 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]): \( C = L = Z_+ = \emptyset \):
   - (a) \( Z \vdash_{\sigma} I_X, \) and
   - (b) \( Y \vdash_{\sigma} I_X | X, Z, \) implies:
     \[
P(Y|do(X)) = \int P(Y|X, Z) dP(Z).
     \]

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 \( Y \perp \! \! \! \perp X|Z, W \) in the graph \( G_{do(W)} \) that is further mutilated on the set \( X(Z) \), the set of all \( X \)-nodes that are not ancestors of any \( Z \)-node in \( G_{do(W)} \).
2. For the backdoor criterion instead of \( L \perp \! \! \! \! \perp I_X \) we could have written that \( L \) does not contain any descendant of \( X \); and for \( Y \perp \! \! \! \! \perp I_X | X, Z \) that \( Z \) blocks all "backdoor paths" from \( X \) to \( Y \).

We presented the results in the formulaic terms of \( \sigma \)-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 \( P(Y|do(J)) \) in an IOSCM can be identified (i.e., expressed in terms of) the observational distributions \( P(V|do(J)) \) and the induced graph \( G \). Note that having more dependence structure (like latent founders, 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 $Cd(v)$ of $v$ 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 $v_i \in Sc(v_{i-1})$. For $B \subseteq V$ we write $Cd(B) := \bigcup_{v \in B} Cd(v)$. Let $CD(G)$ 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:
   
   $w \in Anc(v) \setminus Sc(v) \implies w < v$.

2. For every $v_1, v_2, w \in V$ we have:
   
   $v_2 \in Sc(v_1) \setminus (v_1 \leq w \leq v_2) \implies w \in Sc(v_1)$.

**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. $w \in PaG(v) \implies w < v$).

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 $<$ an apt-order on $G$. For elements $v \in V$ and subsets $B \subseteq V$ we put:

1. $PredG(v) := \{ w \in V | w < v \}$.

2. $PredG(v) := \{ w \in V | w = v \lor w < v \}$.

3. $PredG(B) := \bigcup_{v \in B} PredG(v)$.

4. $PredG(B) := PredG(B) \setminus B$.

**Remark 9.5.** If $B$ is strongly-connected, then $PredG(B)$ is ancestral in $G$, i.e., $AncG(PredG(B)) = PredG(B)$.

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

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

1. $G_{[C]}$ to be the subgraph of $G^+_{do(PaG(C) \setminus C)}$ induced by $C \cup PaG^+(C)$.

2. $V_{[C]} := C$, $J_{[C]} := PaG^+(C) \setminus (C \cup U)$, $U_{[C]} := U \cap PaG^+(C)$.

3. Keep all functions $g_B$ with $S \subseteq C$.

4. $P_{U_{[C]}} := \bigotimes_{v \in U_{[C]}} P_v$, i.e. the marginal of $P_U$ and we will use the notation $P_U$ (or just $P$) 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 $M = (G^+, X, P_U, g)$ be an ioSCM with $G^+ = (V \cup U \cup J, E^+)$ For $C \subseteq V$, we have (indices for emphasis):

$$P_{M_{[C]}(C \cap \{\text{do}(P_aG(C) \setminus C)\}) = P_{M}(C \cap \{\text{do}(J \cup W)\}),$$

for any $W \subseteq V \setminus C$ that contains $(PaG(C) \cap V) \setminus C$. As a special case: if $A \subseteq G$ is ancestral, i.e., $AncG(A) = A$,

$$P_{M_{[A]}(A \cap V \cup \{\text{do}(A \cap J)\}) = P_{M}(A \cap V \cup \{\text{do}(J \cup W)\})$$

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 $M = (G^+, X, P_U, g)$ be an ioSCM with $G^+ = (V \cup U \cup J, E^+)$ and $<$ an apt-order for $G^+$.

1. $P(V|\{\text{do}(J)\}) = \bigotimes_{S \subseteq V} P(S|\text{PredG}(S) \cap V, \{\text{do}(J)\})$.

2. For $S \subseteq V$ a strongly connected component of $G$, $D := CdG(S)$ its consolidated district in $G$ and $P := PaG^+(D) \setminus D$:

$$P_{M}(S|\text{PredG}(S) \cap V, \{\text{do}(J)\}) = P_{M_{[D]}(S|\text{PredG}(S) \cap V, \{\text{do}(P)\})}.$$

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

$$P_{D}(\{\text{do}(J \cup V \cup D)\}) = \bigotimes_{S \subseteq V} P(S|\text{PredG}(S) \cap V, \{\text{do}(J)\})$$

**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. □

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

$$P(V|\{\text{do}(J)\}) = \bigotimes_{D \subseteq V} P(D|\{\text{do}(J \cup V \setminus D)\}).$$

Note that the product might not be well-defined as the consolidated districts i.e. are not totally ordered by $<$.
This problem is usually not addressed in the literature. The problem disappears if every strongly connected component \( S \subseteq V \) comes with a measure \( \mu_S \) such that \( \mathbb{P}(V \mid \text{do}(J)) \) has a density w.r.t. the product measure \( \otimes_{S \in S(G)} \mu_S \). Then the densities \( p(D \mid \text{do}(J \cup V \setminus D)) \) can be multiplied in any order and the integration can be separately done via the \( \mu_S \) 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 \( M = (G^+, X, P, \text{id}, g) \) be an ioSCM with \( G^+ = (V \cup U \cup I, E^+) \) with set of observed nodes \( V \) and input nodes \( I \) and distributions \( \mathbb{P}(V \mid \text{do}(J)) \). Let \( < \) be an apt-order for \( G^+ \). Assume that for every strongly connected component \( S \subseteq V \) we have a measure \( \mu_S \) such that \( \mathbb{P}(V \mid \text{do}(J)) \) has a density w.r.t. the product measure \( \otimes_{S \in S(G)} \mu_S \). 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 \( \mathbb{P}(Y \mid \text{do}(W)) \) is identifiable, i.e. it can be computed from \( \mathbb{P}(V \mid \text{do}(J)) \) 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.

```plaintext
function Id(G, Y, W, \( \mathbb{P}(V \mid \text{do}(J)) \))
2:   require: \( Y \subseteq V \), \( W \subseteq V \), \( Y \cap W = \emptyset \)
3:   \( H \leftarrow \text{Anc}_{G^+ \setminus W}(Y) \)
4:   for \( C \in \text{CD}(H) \) do
5:       \( Q[C] \leftarrow \text{IdCD}(G, C, \text{Cd}_{G^+}(C), Q[\text{Cd}_{G^+}(C)]) \)
6:      if \( Q[C] \) = FAIL then
7:          return FAIL.
8:   end if
9: end for
10: \( Q[H] \leftarrow \left[ \bigotimes_{C \in \text{CD}(H)} \right] Q[C] \)
11: return \( \mathbb{P}(Y \mid \text{do}(J, W)) = \int Q[H] d\mu_{H \setminus Y} \)
12: end function

function IdCD(G, C, D, Q[D])
13:   require: \( C \subseteq D \subseteq V \), \( \text{CD}(G_D) = \{ D \} \)
14:   \( A \leftarrow \text{Anc}_{G^+ \setminus D}(C) \cap D \)
15:   \( Q[A] \leftarrow \int Q[D] d\mu_{D \setminus A} \)
16:   if \( A = C \) then
17:      return \( Q[A] \)
18: else if \( D = A \) then
19:      return FAIL.
20: end if
21: else if \( C \subseteq A \subseteq D \) then
22:      for \( S \in \text{S}(G[A]) \) s.t. \( S \subseteq \text{Cd}_{G^+}(A)(C) \) do
23:          \( R_A[S] \leftarrow \mathbb{P}(S \mid \text{Pred}_{G^+}(S) \cap A, \text{do}(J \cup V \setminus A)) \)
24:       end for
25: \( Q[\text{Cd}_{G^+}(A)(C)] \leftarrow \mathbb{P}[S \mid \text{Pred}_{G^+}(S) \cap A, \text{do}(J \cup V \setminus A)) \)
26: \( \otimes_{S \in \text{S}(G[A])} R_A[S] \)
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]).

**Acknowledgements**

This work was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement 639466).
References


