# 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 *interventional structural causal models (iSCMs)*, 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. iSCMs). 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 iSCMs. Together, our results thus enable causal reasoning in the presence of cycles, latent confounders and selection bias.

## 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, 3, 4, 12, 18–21, 23, 24, 28–30]). 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, 3, 4, 12, 18–21, 23, 24, 28–30]), 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 *interventional structural causal models (iSCMs)*, a "conditional" version of the recently proposed class of *modular structural causal models (mSCMs)* (see [8, 9]) to also include external nodes that can play the role of parameter/action/intervention nodes. They have several desirable properties: iSCMs 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. Furthermore, the class of iSCMs is closed under arbitrary marginalisations and interventions. All causal models that are based on acyclic graphs like DAGs, ADMGs or mDAGs (see [7, 25]) can be interpreted as special acyclic iSCMs. Thus iSCMs generalize all these classes of causal models in one framework, but also allow for cycles and external non-/probabilistic nodes. Also the *generalized directed global Markov property* for mSCMs (see [8, 9]) generalizes to iSCMs, i.e. iSCMs 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 [7, 17, 21, 25, 33]) from acyclic graphs to directed mixed graphs (DMGs) (and even HEDGes and $\sigma$-CGs) with or without cycles in a non-naive way.

This paper now aims at proving the mentioned main rules of causal calculus for iSCMs and derive adjustment criteria with corresponding adjustment formulas like generalized (selection-)backdoor adjustments.

The paper is structured as follows: We will first give the precise definition and main constructions of interventional structural causal models (iSCMs) closely mirroring mSCMs from [8, 9]. We then will review the definition of $\sigma$-separation and generalize its criterion from mSCMs (see [8,9]) to iSCMs. As a preparation for the causal calculus, which relates observational and interventional distributions, we will then show how one can extend a given iSCM to one that also incorporates additional interventional variables indicating the regime of interventions onto the observed nodes. We will then basically show how the rules of causal calculus directly follow from the existence of such an extended iSCM and the $\sigma$-separation criterion applied to it. Finally, we will derive the mentioned general adjustment criteria with corresponding adjustment formulas.

## 2 INTERVENTIONAL STRUCTURAL CAUSAL MODELS

In this section we will define *interventional structural causal models* (iSCMs), which could be seen as a "conditional" version of modular structural causal models (mSCMs) defined in [8, 9]. We will then construct marginalized iSCMs and intervened iSCMs. 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 paths $v_1 \longrightarrow \cdots \longrightarrow v_2$ 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.*
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:* $\mathrm{Sc}^G(v) := \mathrm{Anc}^G(v) \cap \mathrm{Desc}^G(v)$.

**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: $\mathcal{L}(G) = \{\{v\} \mid v \in V\}$.*

In the following, all spaces are meant to be equipped with $\sigma$-algebras, forming standard measurable spaces, and all maps to be measurable.

**Definition 2.3** (Interventional Structural Causal Model). *An* interventional structural causal model (iSCM) *by definition consists of:*

1. *a set of nodes $V^+ = V \dot\cup U \dot\cup J$, where elements of $V$ correspond to observed variables, elements of $U$ to latent variables and elements of $J$ to 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) $V = \mathrm{Ch}^{G^+}(U)$,*
   *(b) $\mathrm{Pa}^{G^+}(U \cup J) = \emptyset$,*
   *where $\mathrm{Ch}^{G^+}$ and $\mathrm{Pa}^{G^+}$ stand for children and parents in $G^+$, resp.,*[1]
5. *a system of structural equations $g = (g_S)_{\substack{S \in \mathcal{L}(G^+) \\ S \subseteq V}}$:*

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

   *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:*

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

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

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

**Definition 2.4** (Modular structural causal model, see [8,9])**.** *A* modular structural causal model (mSCM) *is an iSCM without intervention nodes, i.e. $J = \emptyset$.*

**Remark 2.5** (Relation between iSCMs and mSCMs)**.** *Given an iSCM $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ 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 $\mathcal{X}_J := \prod_{j \in J} \mathcal{X}_j$. For every node $j \in J$ we can decide to change $j$ either to a latent node ($U$) or to an observed node ($V$). In the latter case we then formally need to add a latent node $u_j$ to $U$ and an edge $u_j \longrightarrow j$ to $G^+$, put $\mathcal{X}_{u_j} := \mathcal{X}_j$ and $g_{\{j\}} := \mathrm{id}_{\mathcal{X}_j}$ and consider $\mathbb{P}_j$ to live on the latent space $\mathcal{X}_{u_j}$ (corresponding to $u_j$ rather then to $j$ directly).*

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

**Definition 2.6.** *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an* iSCM *with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. We fix a value $x_J \in \mathcal{X}_J$. The following constructions will depend on the choice of $x_J$.*

[1] To have a "reduced" form of the latent space one can in addition impose the condition: $\mathrm{Ch}^{G^+}(u_1) \not\subseteq \mathrm{Ch}^{G^+}(u_2)$ 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 $v \in V$.

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. These are still independent of $x_J$, but we put $X_u^{\mathrm{do}(x_J)} := X_u$.*
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 [8]).*
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:*

$$\mathbb{P}(X_V, X_J) := \mathbb{P}_U(X_V|\,\mathrm{do}(X_J))\,\mathbb{P}_J(X_J).$$

**Remark 2.7.** *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an iSCM 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 [8, 9]) we can define the marginalisation of an iSCM.

**Definition 2.8** (Marginalisation of iSCMs)**.** *Let $M = (G^+, \mathcal{X}, \mathbb{P}, g)$ be an iSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$ and $W \subseteq V$ a subset. The* marginalised *iSCM $M^{\setminus W}$ 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 $w \in \mathrm{Pa}^{G^+}(S) \setminus S$):*

$$\begin{aligned} & g_{S',v}(x_{\mathrm{Pa}^{(G^+)^{\setminus W}}(S')\setminus S'}) := \\ & g_{S,v}(x_{\mathrm{Pa}^{G^+}(S)\setminus(S\cup\{w\})}, g_{\{w\}}(x_{\mathrm{Pa}^{G^+}(w)\setminus\{w\}})), \end{aligned}$$

*where $(G^+)^{\setminus W}$ is the marginalised graph of $G^+$, $S' \subseteq V^{\setminus W} := V \setminus W$ is any loop of $(G^+)^{\setminus W}$ and $S$ the corresponding induced loop in $G^+$.*

Similar to mSCMs (see [8, 9]) we now define what it means to intervene on *observed* nodes in an iSCM.

**Definition 2.9** (Perfect interventions on iSCMs)**.** *Let $M = (G^+, \mathcal{X}, \mathbb{P}, g)$ be an iSCM 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 *iSCM $M_{\mathrm{do}(W)}$ 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 iSCM $M_{\mathrm{do}(W)}$.*

## 3 CONDITIONAL INDEPENDENCE

Here we shortly generalize conditional independence for structured families of distributions. The main application will be the distributions $(\mathbb{P}_U(X_V|\,\mathrm{do}(X_J = x_J)))_{x_J\in\mathcal{X}_J}$ coming from iSCMs, 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: $\quad X_A \perp\!\!\!\perp_{\mathbb{P}_{V\cup J}} X_B \,|\, X_C, \quad$ 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 $\mathbb{P}_{V\cup J}(X_{V\cup J}) := \mathbb{P}_V(X_V|X_J)\mathbb{P}_J(X_J)$, the distribution given by $X_J \sim \mathbb{P}_J$ and then $X_V \sim \mathbb{P}_V(\_|X_J)$.*

[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 [14]).

**Lemma 3.2.** *Let the situation be like in 3.1. If $A, B, C$ are pairwise disjoint, $A \cap J = \emptyset$ and $J \subseteq B \cup C$ then we have the equivalence:*

1. $X_A \perp\!\!\!\perp_{\mathbb{P}} X_B \,|\, X_C$*, if and only if*
2. $\mathbb{P}_V(X_A | X_B = x_B, X_C = x_C)$ *is only a function of* $x_C$ *(for every setting of* $X_A$*).*

*Proof.* Since the latter does not depend on the choice of $\mathbb{P}_J$ it clearly implies the former. Now assume the former. For every two values $x_B \neq x'_B \in \mathcal{X}_B$ and every $x_{V \cup J \setminus B} \in \mathcal{X}_{V \cup J \setminus B}$ put $\mathbb{P}_J := \bigotimes_{j \in B \cap J} \left( \frac{1}{2} \delta_{x_j} + \frac{1}{2} \delta_{x'_j} \right) \otimes \bigotimes_{j \in J \setminus B} \delta_{x_j}$. Since the former holds for every $\mathbb{P}_J$ of product form we get:

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

with $\mathbb{P}_{V \cup J}$ given by $\mathbb{P}_V(X_V | X_J) \mathbb{P}_J(X_J)$ using $\mathbb{P}_J$ from above. The claim follows:

$$\begin{aligned} &\mathbb{P}_V(X_A | X_B = x_B, X_C = x_C) \\ &= \mathbb{P}_V(X_A | X_B = x'_B, X_C = x_C). \end{aligned}$$ □

**Remark 3.3.**
1. *Lem. 3.2 shows that definition 3.1 generalizes the one from [26].*
2. *Thm. 4.4 in [2] shows that definition 3.1 also generalizes the one from [2].*
3. *In contrast with [2, 5, 26] definition 3.1 can accommodate any variable from $V$ or $J$ at any spot of the conditional independence statement.*
4. $\perp\!\!\!\perp_{\mathbb{P}}$ *satisfies the* semi-graphoid/separoid axioms *(see [5, 11, 22] or see rules 1-5 in Lem. 4.5 for* $\perp\!\!\!\perp_{\mathbb{P}}$*) as these rules hold for any distribution and 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 iSCM will entail the conditional independencies that come from $\sigma$-separation in its induced DMG. We again will closely follow the work in [9].

**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 ($\longleftrightarrow$). In case $G$ contains no directed cycles it is called an* acyclic directed mixed graph (ADMG).

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

$$v_1 \; \overset{\longleftrightarrow}{\underset{\longleftarrow}{\longrightarrow}} \; \cdots \; \overset{\longleftrightarrow}{\underset{\longleftarrow}{\longrightarrow}} \; v_n.^4$$

*The path 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*:* $v_{i-1} \overset{\longrightarrow}{\longleftrightarrow} v_i \overset{\longleftarrow}{\longleftrightarrow} v_{i+1}$,
   *satisfies* $v_i \in C$,
   (b) left chain*:* $v_{i-1} \longleftarrow v_i \overset{\longleftarrow}{\longleftrightarrow} v_{i+1}$,
   *satisfies* $v_i \notin C$ *or* $v_i \in C \cap \mathrm{Sc}^G(v_{i-1})$,
   (c) right chain*:* $v_{i-1} \overset{\longrightarrow}{\longleftrightarrow} v_i \longrightarrow v_{i+1}$,
   *satisfies* $v_i \notin C$ *or* $v_i \in C \cap \mathrm{Sc}^G(v_{i+1})$,
   (d) fork*:* $v_{i-1} \longleftarrow v_i \longrightarrow v_{i+1}$,
   *satisfies* $v_i \notin C$ *or*
   $v_i \in C \cap \mathrm{Sc}^G(v_{i-1}) \cap \mathrm{Sc}^G(v_{i+1})$.

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 path $\pi$ (with some $n \geq 1$ 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:*

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

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

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

**Remark 4.4.**
1. *In any DMG we will always have that $\sigma$-separation implies d-separation, since every $C$-d-open path is also $C$-$\sigma$-open because $\{v\} \subseteq \mathrm{Sc}^G(v)$.*
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 [8] that $\sigma$-separation satisfies the *graphoid/separoid axioms* (see [5, 11, 22]):

**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 $\perp\!\!\!\perp^{\sigma}_{G}$):*

1. *Redundancy:* $A \perp\!\!\!\perp B \,|\, B$ *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:* $(A \perp\!\!\!\perp B \,|\, D) \wedge (A \perp\!\!\!\perp C \,|\, D)$
   $\implies A \perp\!\!\!\perp B \cup C \,|\, D$.

[4] The stacked edges are meant to be read as an "OR" at each place independently. We also allow for repeated nodes in the paths.

It was also shown that $\sigma$-separation is stable under *marginalisation* (see [8, 9]):

**Theorem 4.6** ($\sigma$-Separation under marginalisation, see [8, 9])**.** *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:*

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

*where $G^{\setminus L}$ is the DMG that arises from $G$ by marginalising 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 iSCM $M$ 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 iSCM with graph $G^+$ we need to marginalize them out. This will give us an induced directed mixed graph (DMG) $G$.

**Definition 5.1** (Induced DMG of an iSCM)**.** *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an iSCM 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 an $u \in U$ 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 [8, 9]) to iSCMs.

**Theorem 5.2** ($\sigma$-Separation criterion)**.** *Let $M$ be an iSCM with induced DMG $G$. Then for all subsets $A, B, C \subseteq V \cup J$ we have the implication:*

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

*In words, if $A$ and $B$ are $\sigma$-separated by $C$ in $G$ then the corresponding variables $X_A$ and $X_B$ are conditional independent given $X_C$ under $\mathbb{P}$, i.e. under the joint distribution $\mathbb{P}_U(X_V \,|\, \mathrm{do}(X_J))\mathbb{P}_J(X_J)$ 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 iSCM $M$ constitutes a well-defined mSCM with the same induced DMG $G$. So the $\sigma$-separation criterion for iSCMs directly follows from the mSCM-version proven in [8, 9]. $\square$

**Remark 5.3.** *Note that, since $\sigma$-separation is stable under marginalisation (see [8, 9]) also the $\sigma$-separation criterion is stable under marginalisation.*

**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 [6, 15, 16, 21]). As an example consider $A, B \subseteq V$, $I \subseteq J$ and $x'_I \in \mathcal{X}_I$, $x_J \in \mathcal{X}_J$. Then we have:*

$$\begin{aligned} & A \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I \,|\, B \cup (J \setminus I) \\ \implies\ & \mathbb{P}_U(X_A | X_B, \mathrm{do}(X_J = x_J)) \\ =\ & \mathbb{P}_U(X_A | X_B, \mathrm{do}(X_I = x'_I), \mathrm{do}(X_{J \setminus I} = x_{J \setminus I})). \end{aligned}$$

*So the graphical independence of the intervention nodes implies that the conditional probability is independent of the actual intervention on $I$.*

# 6 CAUSAL CALCULUS FOR PERFECT INTERVENTIONS

## 6.1 The extended iSCM

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 iSCM $M$ *all* interventional iSCMs $M_{\mathrm{do}(W)}$, where $W$ runs through all subsets of observed variables, and glue them all together into one big *extended iSCM* $\hat{M}$. 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 [18, 19, 21]. The definition will be made in such a way that $\hat{M}$ will still be a well-defined iSCM. So all the results for iSCMs will apply to $\hat{M}$, most importantly the $\sigma$-separation criterion (Thm. 5.2).

**Definition 6.1.** *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an iSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. The* extended iSCM *$\hat{M} = (\hat{G}^+, \hat{\mathcal{X}}, \mathbb{P}_U, \hat{g})$ 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 $I_v \to v$ for every $v \in V$. For a uniform notation we sometimes write $I_j$ instead of $j$ for $j \in J$. So we have:* $\hat{J} := J \cup \{I_v \,|\, v \in V\} = \{I_w \,|\, w \in V \dot\cup J\}$.
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 $v \in A$:*

$$\hat{g}_{A,v}(x_A, x_{\mathrm{Pa}^{G^+}(A)\setminus A})$$

$$:= \begin{cases} x_v & \text{if} \quad v \in I(x_A), \\ g_{S,v}(x_{\mathrm{Pa}^{H(x_A)}(S)\setminus S}) & \text{if} \quad v \notin I(x_A), \end{cases}$$

*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 iSCM.*
5. *All the distributions in $\hat{M}$ then are given by the general procedure of iSCMs (see Def. 2.6). 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 iSCM $M$ with the ones from the extended iSCM $\hat{M}$. These relations will be used in the following.

**Proposition 6.2.** *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an iSCM with $G^+ = (V\dot{\cup}U\dot{\cup}J, E^+)$ and $\hat{M}$ the extended iSCM. 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:*

$$\begin{aligned} &\mathbb{P}_U(X_A|X_B, \mathrm{do}(X_{C\cup J} = x_{C\cup J})) \\ =\, &\mathbb{P}_U(X_A|X_B, I_C = x_C, X_J = x_J) \\ =\, &\mathbb{P}_U(X_A|X_B, I_C = x_C, X_C = x_C, X_J = x_J). \end{aligned}$$

*Proof.* Consider the first equality. For any subset $D \subseteq V$ the variable $X_D^{\mathrm{do}(X_{C\cup J}=x_{C\cup J})}$ was recursively defined in $M_{\mathrm{do}(C)}$ via $g$ using $G^+_{\mathrm{do}(C)}$, whereas the variable $X_D^{\mathrm{do}((I_C, I_{V\setminus C}, X_J)=(x_C, \varnothing_{V\setminus C}, x_J))}$ was recursively defined in $\hat{M}$ via the same $g$ but using $I(x_C, \varnothing_{V\setminus C})$ and $G^+_{\mathrm{do}(I(x_C, \varnothing_{V\setminus C}))}$. Since $x_C \in \mathcal{X}_C$ we have that $I(x_C, \varnothing_{V\setminus C}) = C$ and thus $G^+_{\mathrm{do}(I(x_C,\varnothing_{V\setminus C}))} = G^+_{\mathrm{do}(C)}$. It directly follows that:

$$X_D^{\mathrm{do}(X_{C\cup J}=x_{C\cup J})} = X_D^{\mathrm{do}((I_C, I_{V\setminus C}, X_J)=(x_C, \varnothing_{V\setminus C}, x_J))}.$$

This shows the equality of top and middle line. For the equality between the middle and bottom line note that:

$$I_C = x_C \overset{x_C \in \mathcal{X}_C}{\Longrightarrow} X_C = x_C. \qquad \square$$

## 6.2 The three main rules of causal calculus

**Notation 6.3.** *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 a 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*

$$\mathbb{P}_U(X_Y|I_T = x_T, X_T = x_T, X_Z, \mathrm{do}(X_W = x_W)),$$

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

3. *We write $Y \perp\!\!\!\perp_{\mathbb{P}} T|Z, \mathrm{do}(W)$ for $X_Y \perp\!\!\!\perp_{\mathbb{P}_U(\_|\,\mathrm{do}(X_W=x_W))} X_T|X_Z$, etc..*
4. *We write $Y \perp\!\!\!\perp^{\sigma}_{G} I_X \,|\, X, Z, \mathrm{do}(W)$ to mean $Y \perp\!\!\!\perp^{\sigma}_{\hat{G}_{\mathrm{do}(W)}} I_X \,|\, X, Z$, where $\hat{G}_{\mathrm{do}(W)}$ is the extended DMG of the intervened graph $G^+_{\mathrm{do}(W)}$.*

**Theorem 6.4** (The three main rules of causal calculus)**.** *Let $M$ be an iSCM with set of observed nodes $V$ and intervention 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* $\quad Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} X|Z, \mathrm{do}(W) \quad$ *then:*

$$\mathbb{P}(Y|X, Z, \mathrm{do}(W)) = \mathbb{P}(Y|Z, \mathrm{do}(W)).$$

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

$$\mathbb{P}(Y|\,\mathrm{do}(X), Z, \mathrm{do}(W)) = \mathbb{P}(Y|Z, \mathrm{do}(W)).$$

*Proof.* 1. Thm. 5.2 applied to $G_{\mathrm{do}(W)}$ gives:

$$Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} X|Z, \mathrm{do}(W) \overset{5.2}{\Longrightarrow} Y \underset{\mathbb{P}}{\perp\!\!\!\perp} X|Z, \mathrm{do}(W).$$

The latter directly gives the claim:

$$\mathbb{P}(Y|X, Z, \mathrm{do}(W)) = \mathbb{P}(Y|Z, \mathrm{do}(W)).$$

2. The $\sigma$-separation criterion 5.2 w.r.t. to $\hat{G}_{\mathrm{do}(W)}$ gives:

$$Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} I_X|X, Z, \mathrm{do}(W) \overset{5.2}{\Longrightarrow} Y \underset{\mathbb{P}}{\perp\!\!\!\perp} I_X|X, Z, \mathrm{do}(W).$$

Together with Prp. 6.2 (applied to $M_{\mathrm{do}(W)}$) we have:

$$\begin{aligned} & \mathbb{P}(Y|\,\mathrm{do}(X), Z, \mathrm{do}(W)) \\ \overset{6.2}{=}\; & \mathbb{P}(Y|I_X, X, Z, \mathrm{do}(W)) \\ \overset{Y \perp\!\!\!\perp I_X|X,Z,\mathrm{do}(W)}{=}\; & \mathbb{P}(Y|X, Z, \mathrm{do}(W)). \end{aligned}$$

3. As before we have:

$$Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X | Z, \mathrm{do}(W) \overset{5.2}{\implies} Y \underset{\mathbb{P}}{\perp\!\!\!\perp} I_X | Z, \mathrm{do}(W).$$

And again: $\quad \mathbb{P}(Y | \mathrm{do}(X), Z, \mathrm{do}(W))$

$$\overset{6.2}{=} \mathbb{P}(Y | I_X, Z, \mathrm{do}(W))$$
$$\overset{Y \perp\!\!\!\perp I_X | Z, \mathrm{do}(W)}{=} \mathbb{P}(Y | Z, \mathrm{do}(W)). \qquad \square$$

**Remark 6.5.** *The conditions for aboves rules are usually phrased in terms of further graph transformations. E.g. for rule 3 in the acyclic setting one 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 $G_{\mathrm{do}(W)}$ (see [18,19,21]). When reduced to those acyclic settings they will be equivalent to the our formulation via $\sigma$-separation including the intervention nodes.*

*We presented the above only in terms of $\sigma$-separation in the extended iSCM because in this form it was clear how to generalize to arbitrary iSCMs. We also believe that the formulaic expressions in terms of $\sigma$-separation will be easier to parse and to remember as their relations to the claims and implications are directly visible.*

### 6.3 The general adjustment criterion

**Notation 6.6.** *Let $M = (G^+, \mathcal{X}, \mathbb{P}, g)$ be an iSCM with $G^+ = (V \dot\cup U \dot\cup J, E^+)$. The following set of nodes/variables will play the described roles:*

1. *$Y$: the outcome variables,*
2. *$X$: the treatment or intervention variables,*
3. *$Z_0$: the core set of adjustment variables,*
4. *$Z_+$: additional adjustment variables,*
5. *$L$: "latent adjustment variables",*
6. *$Z := Z_0 \cup Z_+$: all actual adjustment variables,*
7. *$C$: context variables,*
8. *$W$: default intervention variables containing $J$,*
9. *$S$: variables inducing selection bias when $S = s$.*

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 [18, 19, 21]) and its extensions (also see [3,23,24,28]) to general iSCMs.

**Theorem 6.7** (General adjustment criterion and formula)**.** *Let the setting be like in 6.6. 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 $S = \emptyset$), but there are unbiased samples from $\mathbb{P}(Z | C, \mathrm{do}(W))$. Further assume that the variables satisfy:*

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

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

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

*Proof.* Since $C$, $\mathrm{do}(W)$ occur everywhere as a conditioning set, we will suppress $C$, $\mathrm{do}(W)$ in the following everywhere. Then note that the $\sigma$-separation criterion 5.2 implies the corresponding conditional independencies in following when indicated. The adjustment formula then derives from the following computations:

$$\begin{aligned}
& \mathbb{P}(Y | \mathrm{do}(X)) \\
= & \int \mathbb{P}(Y | Z_0, L, \mathrm{do}(X)) \, d\mathbb{P}(Z_0, L | \mathrm{do}(X)) \\
\overset{6.2}{=} & \int \mathbb{P}(Y | I_X, X, Z_0, L) \, d\mathbb{P}(Z_0, L | I_X) \\
\overset{Y \perp\!\!\!\perp I_X | X, Z_0, L;\ (Z_0, L) \perp\!\!\!\perp I_X}{=} & \int \mathbb{P}(Y | X, Z_0, L) \, d\mathbb{P}(Z_0, L) \\
\overset{\int d\mathbb{P}(Z_+ | Z_0, L) = 1}{=} & \int \int \mathbb{P}(Y | X, Z_0, L) \, d\mathbb{P}(Z_+ | Z_0, L) \, d\mathbb{P}(Z_0, L) \\
\overset{Y \perp\!\!\!\perp Z_+ | X, Z_0, L}{=} & \int \mathbb{P}(Y | X, Z_0, Z_+, L) \, d\mathbb{P}(Z_+, Z_0, L) \\
\overset{Z = Z_+ \cup Z_0}{=} & \int \mathbb{P}(Y | X, Z, L) \, d\mathbb{P}(Z, L) \\
= & \int \int \mathbb{P}(Y | X, Z, L) \, d\mathbb{P}(L | Z) \, d\mathbb{P}(Z) \\
\overset{L \perp\!\!\!\perp X | Z}{=} & \int \int \mathbb{P}(Y | L, X, Z) \, d\mathbb{P}(L | X, Z) \, d\mathbb{P}(Z) \\
= & \int \mathbb{P}(Y | X, Z) \, d\mathbb{P}(Z) \\
\overset{Y \perp\!\!\!\perp S | X, Z}{=} & \int \mathbb{P}(Y | X, Z, S) \, d\mathbb{P}(Z). \qquad \square
\end{aligned}$$

**Remark 6.8.** *Note that the adjustment* formula *in theorem 6.7 does not depend on $L$. This thus allows us to even choose variables for $L$ that come from an iSCM $M'$ 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 [28] to find* all *adjustment sets in the* acyclic *case with $S = C = \emptyset$, see below.*

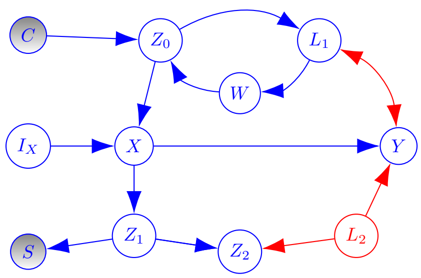


Figure 1: An induced DMG $G$ with intervention node $I_X$ (the others are left out for readability). The variables satisfy the general adjustment criterion for $\mathbb{P}(Y|C, \mathrm{do}(X))$ with $L = \{L_1, L_2\}$ and $Z_+ = \{Z_1, Z_2\}$. Note that $L_2$ could also have been a latent variable.

**Corollary 6.9.** *Let the notations be like in 6.6 and 6.7 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 [3]): $C = \emptyset$, and*
   (a) $(Z_0, L) \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X$, *and*
   (b) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} (I_X, Z_+) \,|\, X, Z_0, L$, *and*
   (c) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} S \,|\, X, Z$, *and*
   (d) $L \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} X \,|\, Z$, *implies:*
   $$\mathbb{P}(Y|\,\mathrm{do}(X)) = \int \mathbb{P}(Y|X, Z, S = s)\, d\mathbb{P}(Z).$$
2. *Selection-backdoor (see [1]): $C = L = \emptyset$, and*
   (a) $Z_0 \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X$, *and*
   (b) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} (I_X, Z_+, S) \,|\, X, Z_0$ *implies:*
   $$\mathbb{P}(Y|\,\mathrm{do}(X)) = \int \mathbb{P}(Y|X, Z, S = s)\, d\mathbb{P}(Z).$$
3. *Extended backdoor[5] (see [23, 28]): $C = S = \emptyset$,*
   (a) $(Z_0, L) \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X$, *and*
   (b) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} (I_X, Z_+) \,|\, X, Z_0, L$, *and*
   (c) $L \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} X \,|\, Z$, *implies:*
   $$\mathbb{P}(Y|\,\mathrm{do}(X)) = \int \mathbb{P}(Y|X, Z)\, d\mathbb{P}(Z).$$
4. *Backdoor (see [18, 19, 21]): $C = S = L = Z_+ = \emptyset$,*
   (a) $Z \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X$,
   (b) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X \,|\, X, Z$ *implies:*
   $$\mathbb{P}(Y|\,\mathrm{do}(X)) = \int \mathbb{P}(Y|X, Z)\, d\mathbb{P}(Z).$$

A generalization of the crition for selection without (or only partial) external data of [3, 4] is given in the appendix.

**Remark 6.10.** *The conditions in theorem 6.7 and corollary 6.9 are in the acyclic setting usually phrased in terms of sub-structures of the graph $G$ (see [18, 19, 21]). E.g. for the backdoor criterion for a DAG $G$ instead of $L \perp\!\!\!\perp^d_G I_X$ we could have written that $L$ does not contain any descendent of $X$; and for $Y \perp\!\!\!\perp^d_G I_X \,|\, X, Z$ that $Z$ blocks all "backdoor paths" from $X$ to $Y$. As before, we presented the results only in terms of $\sigma$-separation because the relations to their use is directly indicated (e.g. in the proofs). Furthermore, the formulaic framework of $\sigma$-separation makes the generalization to arbitrary iSCMs possible.*

## 7 TWIN NETWORKS AND COUNTERFACTUALS

We shortly want to mention that with iSCMs one can do counterfactual reasoning as well. Given an iSCM $M$ with graph $G^+ = (V \dot\cup U \dot\cup J, E^+)$, a set $W \subseteq V \cup J$ and $x_W \in \mathcal{X}_W$ and the corresponding intervened iSCM $M_{\mathrm{do}(W)}$ with graph $G^+_{\mathrm{do}(W)}$ and distribution $\delta_{x_W}$ on $W$ one can construct a (merged) *twin iSCM* $M_{\mathrm{twin}}$ similarly to the acyclic case (see [21]). This is done by identifying/merging the nodes, mechanisms and variables from the non-descendents of $W$, $\mathrm{NonDesc}^{G^+}(W)$ and $\mathrm{NonDesc}^{G^+_{\mathrm{do}(W)}}(W)$, which are unchanged by the action $\mathrm{do}(x_W)$. Then one has the two different branches $\mathrm{Desc}^{G^+}(W)$ and $\mathrm{Desc}^{G^+_{\mathrm{do}(W)}}(W)$ in the network. This construction then allows one to formulate counterfactual statements like in the acyclic case (see [21]), but now for general iSCMs. E.g. one could state the assumption of *strong ignorability* (see [21, 27]) as:

$$\left(Y^{\mathrm{do}(X=\varnothing)}, Y^{\mathrm{do}(X=x)}\right) \overset{\sigma}{\underset{G_{\mathrm{twin}}}{\perp\!\!\!\perp}} X \,|\, Z.$$

All the causal reasoning rules derived so far can thus also be applied to reason about counterfactuals.

## 8 CONCLUSION

We proved the three main rules of causal calculus and general adjustment criteria with corresponding formulas to recover from interventions and selection bias for general iSCMs, which allow for arbitrary probability distributions, non-/linear functional relations, latent confounders, external non-/probabilistic parameter/action/intervention nodes and *cycles*. This generalizes all the corresponding results of *acyclic* causal models (see [1, 3, 18, 19, 21, 23, 24, 28]) to general iSCMs. Future work might address completeness

[5] In the acyclic case it was shown in [28] that when $L$ is allowed to represent latent variables in a graph $G'$ that marginalizes to $G$ then this criterion actually characterizes all adjustment sets for $G$ and $\mathbb{P}(Y|\,\mathrm{do}(X))$.

questions (see [12, 21, 29]) and extensions of Tian's algorithm for identifiability (see [10, 12, 13, 21, 26, 30–32]) to general iSCMs.

### 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

[1] E. Bareinboim, J. Tian, and J. Pearl. Recovering from Selection Bias in Causal and Statistical Inference. In *AAAI-14*, 2014.

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

[3] J. D. Correa and E. Bareinboim. Causal Effect Identification by Adjustment under Confounding and Selection Biases. In *AAAI-17*, 2017.

[4] J. D. Correa, J. Tian, and E. Bareinboim. Generalized Adjustment Under Confounding and Selection Biases. In *AAAI-18*, 2018.

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

[6] D. Eaton and K. P. Murphy. Bayesian structure learning using dynamic programming and MCMC. In *UAI-07*, 2007.

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

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

[9] P. Forré and J. M. Mooij. Constraint-based Causal Discovery for Non-Linear Structural Causal Models with Cycles and Latent Confounders. In *UAI-18*, 2018.

[10] D. Galles and J. Pearl. Testing Identifiability of Causal Effects. In *UAI-95*, 1995.

[11] D. Geiger, T.S. Verma, and J. Pearl. Identifying Independence in Bayesian Networks. *Networks*, 20(5):507–534, 1990. UCLA Cognitive Systems Laboratory, Technical Report CSD-890028 (R-116).

[12] Y. Huang and M. Valtorta. Pearl's calculus of intervention is complete. In *UAI-06*, 2006.

[13] 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.

[14] Achim Klenke. *Probability Theory - A Comprehensive Course*. Universitext. Springer, London, second edition, 2014.

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

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

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

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

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

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

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

[22] J. Pearl and A. Paz. Graphoids: a graph-based logic for reasoning about relevance relations. *Advances in artificial intelligence-II*, pages 357–363, 1987. UCLA Computer Science Department Technical Report R-53, CSD-850038.

[23] J. Pearl and A. Paz. Confounding equivalence in causal inference. In *UAI-10*, 2010.

[24] E. Perkovic, J. Textor, M. Kalisch, and M. Maathuis. A complete generalized adjustment criterion. In *UAI-15*, 2015.

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

[26] 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.

[27] 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.

[28] I. Shpitser, T. J. Van der Weele, and J. M. Robins. On the validity of covariate adjustment for estimating causal effects. In *UAI-10*, 2010.

[29] I. Shpitser and J. Pearl. Identification of conditional interventional distributions. In *UAI-06*, 2006.

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

[31] J. Tian. Identifying Conditional Causal Effects. In *UAI-04*, 2004.

[32] J. Tian and J. Pearl. A General Identification Condition for Causal Effects. In *AAAI-02*, 2002.

[33] T. S. Verma and J. Pearl. Causal Networks: Semantics and Expressiveness. In *UAI-90*, 1990.

# SUPPLEMENTARY MATERIAL

## A MORE ON ADJUSTMENT CRITERIA

The following generalizes the adjustment criterion of type I in [3].

**Theorem A.1** (General adjustment without external data)**.** *Let the setting be like in 6.6. Assume that data was collected under selection bias, $\mathbb{P}(V|S=s)$. Further assume that the variables satisfy:*

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

*Then one can estimate the causal effect $\mathbb{P}(Y|\,\mathrm{do}(X))$ via the following* adjustment formula *from the biased data:*

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

*Proof.* First note that the $\sigma$-separation criterion 5.2 implies the corresponding conditional independencies in following when indicated. We implicitly make use of proposition 6.2 when needed. The adjustment formula then derives from the following computations:

$$\begin{aligned}
&\mathbb{P}(Y|\,\mathrm{do}(X)) \\
\overset{Y \perp\!\!\!\perp S\,|\,\mathrm{do}(X)}{=}\;& \mathbb{P}(Y|S,\mathrm{do}(X)) \\
\overset{\text{chain rule}}{=}\;& \int \mathbb{P}(Y|Z_0,S,\mathrm{do}(X))\, d\mathbb{P}(Z_0|S,\mathrm{do}(X)) \\
\underset{6.2}{\overset{Z_0 \perp\!\!\!\perp I_X\,|\,S}{=}}\;& \int \mathbb{P}(Y|Z_0,S,\mathrm{do}(X))\, d\mathbb{P}(Z_0|S) \\
\overset{\int d\mathbb{P}(Z_+|Z_0,S)=1}{=}\;& \int \mathbb{P}(Y|Z_0,S,\mathrm{do}(X))\, d\mathbb{P}(Z_+,Z_0|S) \\
\overset{Y \perp\!\!\!\perp Z_+\,|\,Z_0,S,\mathrm{do}(X)}{=}\;& \int \mathbb{P}(Y|Z_+,Z_0,S,\mathrm{do}(X))\, d\mathbb{P}(Z_+,Z_0|S) \\
\overset{Z=Z_+\cup Z_0}{=}\;& \int \mathbb{P}(Y|Z,S,\mathrm{do}(X))\, d\mathbb{P}(Z|S) \\
\underset{6.2}{\overset{Y \perp\!\!\!\perp I_X\,|\,X,Z,S}{=}}\;& \int \mathbb{P}(Y|Z,S,X)\, d\mathbb{P}(Z|S).
\end{aligned}$$

□

The following theorem generalizes the adjustment criterion of type III in [4]. For this we have to introduce even more adjustment sets: $Z_0^A, Z_0^B, Z_1^A, Z_1^B, Z_2, Z_3$ and $L_0, L_1$.

**Theorem A.2** (General adjustment with partial external data)**.** *Assume that data was collected under selection bias, $\mathbb{P}(V|S=s)$, but we have unbiased data from $\mathbb{P}(Z_{\leq 1}^B)$. Further assume the conditions that are indicated on the equality signs in the proof. Then we have the adjustment formula: $\mathbb{P}(Y|\,\mathrm{do}(X)) =$*

$$\int\!\!\int \mathbb{P}(Y|S=s,Z,X)\, d\mathbb{P}(Z\setminus Z_{\leq 1}^B|S=s,Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B).$$

*Note that this formula does not depend on $L_1$ and $L_2$. So $L_1$ and $L_2$ can be chosen in a graph $G'$ that marginalizes to $G$.*

*Proof.*

$$\begin{aligned}
&\mathbb{P}(Y|\,\mathrm{do}(X)) \\
\overset{\text{chain rule}}{=}\;& \int \mathbb{P}(Y|L_0,Z_0,\mathrm{do}(X))\, d\mathbb{P}(L_0,Z_0|\,\mathrm{do}(X)) \\
\underset{6.2}{\overset{(L_0,Z_0)\perp\!\!\!\perp I_X}{=}}\;& \int \mathbb{P}(Y|L_0,Z_0,\mathrm{do}(X))\, d\mathbb{P}(L_0,Z_0) \\
\underset{Z_{\leq 1}=Z_0\cup Z_1}{\overset{\int d\mathbb{P}(Z_1|L_0,Z_0)=1}{=}}\;& \int \mathbb{P}(Y|L_0,Z_0,\mathrm{do}(X))\, d\mathbb{P}(L_0,Z_{\leq 1}) \\
\overset{Y\perp\!\!\!\perp Z_1\,|\,L_0,Z_0,\mathrm{do}(X)}{=}\;& \int \mathbb{P}(Y|L_0,Z_{\leq 1},\mathrm{do}(X))\, d\mathbb{P}(L_0,Z_{\leq 1}) \\
\underset{Z_{\leq 1}=Z_{\leq 1}^A\cup Z_{\leq 1}^B}{\overset{\text{chain rule}}{=}}\;& \int \mathbb{P}(Y|L_0,Z_{\leq 1},\mathrm{do}(X))\, d\mathbb{P}(L_0|Z_{\leq 1})\, d\mathbb{P}(Z_{\leq 1}^A|Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B) \\
\overset{Z_{\leq 1}^A\perp\!\!\!\perp S\,|\,Z_{\leq 1}^B}{=}\;& \int \mathbb{P}(Y|L_0,Z_{\leq 1},\mathrm{do}(X))\, d\mathbb{P}(L_0|Z_{\leq 1})\, d\mathbb{P}(Z_{\leq 1}^A|S,Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B) \\
\underset{6.2}{\overset{L_0\perp\!\!\!\perp I_X\,|\,Z_{\leq 1}}{=}}\;& \int \mathbb{P}(Y|L_0,Z_{\leq 1},\mathrm{do}(X))\, d\mathbb{P}(L_0|Z_{\leq 1},\mathrm{do}(X))\, d\mathbb{P}(Z_{\leq 1}^A|S,Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B) \\
\overset{\text{chain rule}}{=}\;& \int \mathbb{P}(Y|Z_{\leq 1},\mathrm{do}(X))\, d\mathbb{P}(Z_{\leq 1}^A|S,Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B) \\
\overset{Y\perp\!\!\!\perp S\,|\,Z_{\leq 1},\mathrm{do}(X)}{=}\;& \int \mathbb{P}(Y|S,Z_{\leq 1},\mathrm{do}(X))\, d\mathbb{P}(Z_{\leq 1}^A|S,Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B) \\
\overset{\text{chain rule}}{=}\;& \int \mathbb{P}(Y|L_1,Z_2,S,Z_{\leq 1},\mathrm{do}(X))\, d\mathbb{P}(L_1,Z_2|S,Z_{\leq 1},\mathrm{do}(X))\, d\mathbb{P}(Z_{\leq 1}^A|S,Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B)
\end{aligned}$$

$$
\begin{aligned}
\overset{Z_{\leq 2}=Z_{\leq 1}\cup Z_2}{=} \quad & \int \mathbb{P}(Y|L_1, S, Z_{\leq 2}, \mathrm{do}(X)) \\
& d\mathbb{P}(L_1, Z_2|S, Z_{\leq 1}, \mathrm{do}(X)) \\
& d\mathbb{P}(Z^A_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
\underset{6.2}{\overset{(L_1,Z_2) \perp\!\!\!\perp I_X \mid S,Z_{\leq 1}}{=}} \quad & \int \mathbb{P}(Y|L_1, S, Z_{\leq 2}, \mathrm{do}(X)) \\
& d\mathbb{P}(L_1, Z_2|S, Z_{\leq 1}) \\
& d\mathbb{P}(Z^A_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
\underset{\int \mathbb{P}(Z_3|L_1,S,Z_{\leq 2})=1}{\overset{Y \perp\!\!\!\perp Z_3 \mid L_1,S,Z_{\leq 2},\mathrm{do}(X)}{=}} \quad & \int \mathbb{P}(Y|L_1, S, Z_{\leq 2}, Z_3, \mathrm{do}(X)) \\
& d\mathbb{P}(L_1, Z_2, Z_3|S, Z_{\leq 1}) \\
& d\mathbb{P}(Z^A_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
\underset{Z=Z_{\leq 2}\cup Z_3}{\overset{\text{chain rule}}{=}} \quad & \int \mathbb{P}(Y|L_1, S, Z, \mathrm{do}(X)) \\
& d\mathbb{P}(L_1|S, Z) \\
& d\mathbb{P}(Z \setminus Z^B_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
\underset{6.2}{\overset{L_1 \perp\!\!\!\perp I_X \mid S,Z}{=}} \quad & \int \mathbb{P}(Y|L_1, S, Z, \mathrm{do}(X)) \\
& d\mathbb{P}(L_1|S, Z, \mathrm{do}(X)) \\
& d\mathbb{P}(Z \setminus Z^B_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
\overset{\text{chain rule}}{=} \quad & \int \mathbb{P}(Y|S, Z, \mathrm{do}(X)) \\
& d\mathbb{P}(Z \setminus Z^B_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
\overset{Y \perp\!\!\!\perp I_X \mid X,S,Z}{=} \quad & \int \mathbb{P}(Y|S, Z, X) \\
& d\mathbb{P}(Z \setminus Z^B_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}).
\end{aligned}
$$

□