Causality and independence in systems of equations

Blom, T.

Publication date
2021
Document Version
Final published version

Tineke Blom Causality and independence in systems of equations

Causality and independence in systems of equations

Tineke Blom

Causality and independence in systems of equations

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 23 september 2021, te 10.00 uur

door Tineke Blom
geboren te Dongeradeel

Promotiecommissie

Promotor:	prof. dr. J.M. Mooij	Universiteit van Amsterdam
Copromotor:	prof. dr. M. Welling	Universiteit van Amsterdam
Overige leden:	prof. dr. A.P. Dawid	University of Cambridge
	prof. dr. T.S. Richardson	University of Washington
	dr. J.C. Textor	Radboud Universiteit Nijmegen
	prof. dr. S.J.L. Smets	Universiteit van Amsterdam
	prof. dr. P.M.A. Sloot	Universiteit van Amsterdam
	dr. P.D. Forré	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Het hier beschreven onderzoek werd mede mogelijk gemaakt door steun van: European Research Council onder het Horizon 2020 research and innovation programma van de Europese Unie (grant agreement 639466).

Contents

Chapter 1

Introduction

Learning about causal relations is a central objective in science. Scientific explanations of observed phenomena usually start with identifying the *causes* that produced a certain *effect* (i.e. outcome). In agriculture, for instance, a failed harvest may be explained by determining whether it was caused by heavy rainfall, drought, insects, fungus, disease, or a lack of certain nutrients in the soil. A causal relation can be established by varying the type of crops or controlling the soil conditions and recording the changes in crop yield. This way, knowledge about the causes of a failed harvest can be leveraged to adjust agricultural practices and reduce the chances of disappointing crop yields in future years. Moreover, understanding the causal mechanisms that are involved with crop yield might even pave the way for the development of sustainable farming methods that maximize crop yields. It is imperative to be aware that not all statistical associations are causal relations. For instance, a statistical dependence between average wages and wheat yields in EU countries would not imply that a farmer can increase crop yields by raising his own salary. Instead, a plausible causal explanation of such a correlation would be that certain economic factors affect both the wages and agricultural development. The task of learning causal relations thus goes beyond making predictions based on mere statistical associations because they allow us to make predictions about the effects of interactions with the world around us. Additionally, awareness of the underlying causal mechanisms provides valuable insights into the workings of complex real-world systems.

In many scientific disciplines, the gold standard for causal discovery is randomized controlled experimentation. Consider, for example, a clothing store that wants to determine the effect of advertisement on profit. To estimate this, they could implement A/B testing on their website, i.e. they can try out various marketing strategies and observe how this affects profits from customers that are randomly assigned to receive different types of advertisement compared to customers who received no marketing at all. Similarly, during the trial period of newly developed medicine, patients are randomly assigned to either the treatment or the control group so that treatment

effects and potential side-effects can be estimated. On the other hand, scientists cannot, for the sake of curiosity, vary fiscal policies to determine the effect of stimulus on economic growth. Neither would it be ethical to assign newborn babies into treatment and control groups to assess the effect of bonding on child development. In many practical applications, randomized controlled trials are too expensive, time-consuming, unethical, or otherwise infeasible. To remedy this, researchers have studied ways in which causal relations can be learned from purely observational data, possibly in combination with background knowledge and experimental data. This has led to applications of causal inference methods in a variety of domains, e.g. sociology, economics, and biology.

The novel methods and ideas for causal inference that we develop in this thesis were inspired by a wish to understand causal pathways in protein signalling networks. In living organisms, cellular responses to external inputs are governed by biochemical reaction networks that facilitate the flow of information within cells. These signalling networks play a central role in cell activities such as immunity and tissue repair. Since errors in the way that cells process information may cause autoimmune diseases or cancer, the hope is that a better understanding of protein signalling pathways may help treat these diseases more effectively in the future. Therefore, the design of methods that facilitate or automate causal discovery from experimental protein expression data is a meaningful and far-reaching direction of research. Although previous work along these lines is already impressive, the impact of many causal inference methods in these and other areas remains limited due to remaining issues regarding e.g. measurement error, feedback mechanisms, latent confounders, and selection bias. In this thesis, we work towards bridging this gap between theory and practice by considering how and when the presence of measurement error or feedback mechanisms may affect the outcome of causal discovery algorithms.

The behaviour of complex systems that evolve over time and that consist of many interacting particles, e.g. a protein signalling network, is often modelled with a (stochastic) dynamical model. The variables in such a model may influence one another through the presence of feedback mechanisms. The properties of cyclic causal models that arise from describing the stationary characteristics of a dynamical system form a fascinating research topic. In this work, we take a closer look at a peculiar phenomenon called *perfect adaptation*, which refers to the ability of variables in a system to have a transient response to a persistent external input and then revert back to their initial value. Systems that are able to achieve perfect adaptation possess a particular type of feedback mechanism which occurs quite frequently in chemical reaction networks, protein signalling networks, and control systems. To study their behaviour, we extended the causal ordering framework of Simon (1953), allowing us to deduce causality and independence from sets of (equilibrium) equations. This way, we connect properties of the equilibrium distribution of stochastic dynamical models to interventionist causal modelling frameworks and to the inputs of causal discovery algorithms.

Building on this theoretical framework, we obtained new insights into the behaviour of perfectly adapted dynamical systems and we discovered new approaches to causal discovery. First, we propose the presence of perfect adaptation as a compelling explanation for the apparent differences between the output of causal discovery algorithms applied to protein expression data and the biological consensus networks for the corresponding signalling pathways that have been reported in literature. Second, we formulate sufficient graphical conditions to identify the presence of perfect adaptation in a model, and we demonstrate how it can be recognized in data using a combination of experimental data and background knowledge. Third, we show that perfect adaptation explains non-robustness of qualitative model predictions that is sometimes observed (e.g. in epidemiology or ecology) and we characterize a large class of model extensions that preserve these predictions. Finally, we propose a novel holistic approach to causal discovery and model selection where data for submodels is used to reason about unobserved feedback loops in a larger model. It is our hope that the results in this thesis bring the world of causal discovery and causal modelling one step closer to the world of dynamical models and their application domains.

In Sections 1.1 to 1.3, we provide additional historical, philosophical, and mathematical background information to place the topics of this thesis into a broader scientific context. These sections are optional reading and one can safely skip ahead to Section 1.4 where we present a detailed overview of our main contributions, as well as an outline of this thesis. In Section 1.1 we discuss some historical background on certain metaphysical and epistemological questions that philosophers have raised regarding the notion of causation. In Section 1.2 we introduce recent advances in mathematical frameworks for the representation of causal semantics and the relationship to dynamical models. In Section 1.3 we provide a concise history of methodological and practical scientific methods to establish causal relations.

1.1 Philosophical background

The notion of causation remains an elusive concept. Even though causal claims are ubiquitous in everyday life and the discovery of causal relations is one of the primary goals in countless scientific endeavours, finding an explicit definition of causation or a universal method to establish causal relations has been a controversial topic for many centuries. There are many thought-provoking metaphysical questions (e.g. “How do causes bring about affects?”) and epistemological issues (e.g. “How can causal knowledge be justified”) to think about. The contemporary debate on topics of causation is ongoing in diverse scientific disciplines such as philosophy, physics, statistics, economics, and artificial intelligence. In this section, we will briefly discuss different philosophical perspectives.

1.1.1 Regularity accounts

Human beings are often able to tell the difference between causal and non-causal relations through memories, reasoning, and by observing regularities in the world. Consider, for instance, the following sequence of events: a player in a game of snooker strikes a white ball with his cue, the white ball starts moving, the white ball hits the black ball, the black ball starts moving. Intuitively, we know that the white ball *causes* the movement of the black ball and that the process would play out in the exact same way if it didn't take place on a snooker table but on a billiard table. Can this type of causal inference be justified? We will follow Reiss (2007) in an examination of Hume (1928), who attempted to construct a naturalistic and reductive explanation of how people are able to empirically infer causal relations from regularities that they experience. We then briefly discuss criticisms of regularity accounts of causation.

According to David Hume (1711-1776), causal knowledge includes neither mathematical or logical truths that can be ascertained through reasoning alone (e.g. the Pythagorean theorem $a^2 + b^2 = c^2$) nor facts that can be established through observations and memories alone (e.g. "There are two balls on the billiard table"). Instead, he proposed three criteria that compel human beings to infer causal knowledge from observing a black ball, formerly at rest, acquiring motion after it is hit by a white ball on a billiard table. The first was *contiguity in time and space*, meaning that there is no interval of space or time between the white ball striking the black ball and the black ball acquiring motion. The second is *temporal priority of causes*, meaning that the cause (i.e. the movement of the white ball) precedes the effect (i.e. the black ball acquiring motion). The third and last is *constant conjunction of cause and effect*, meaning that the impulse of one billiard ball will result in movement of the other billiard ball in similar circumstances, so that like causes produce like effects. Together, these criteria form a naturalistic explanation of how humans infer a causal relation from observations.

Hume worried that although causal inference may arise from observing regularities, there is actually nothing about the cause (a moving billiard ball) that necessitates the effect (a second ball acquiring motion). Indeed, for all we know, the second billiard ball could be glued to the table or an earth quake may knock over the table preventing the collision from ever occurring. The idea that notions of causation cannot be derived from observing regularities alone is known as *Hume's critique*. Immanuel Kant (1724-1804) famously wrote that this interrupted him from his "dogmatic slumber" and inspired him to conclude that the notion of causation is *a priori* (Kant, 1783; Langsam, 1994). Roughly speaking, because we assume beforehand that the black billiard ball acquiring motion must have a cause, we are able to infer from observations that it must have been the collision with the white ball that caused the black ball to move. From this viewpoint, we are able to infer causal relations from observing regularities because we already presume that every event has a cause.

Regularity accounts that define causation either in non-causal terms or as fundamental building blocks of nature have been developed much further over the years. Dowe (1992), Mackie (1974), and Salmon (1984) described causation via *causal mechanisms* and used principles of invariance and the notion of conserved quantities to recognize *causal processes*. However, regularity accounts of causation have been widely criticized for not capturing our ordinary notion of causation (Cartwright, 2007; Langsam, 1994).

First of all, Cartwright (2007) and Hoover (1988) pointed out that the notion of contiguity in time and space is not always sensible for causal relations between macro-economic or sociological variables. Consider, for instance, a macro-economic model predicting that central banks printing money causes inflation. Obviously, it is hard to imagine how abstract concepts like money and inflation are exactly connected in space and time. Secondly, the requirement that causes precede effects in time is problematic in macro-economic models that allow for simultaneous causation (F. M. Fisher, 1970). Finally, the idea that cause and effect are related to one another by necessity is contested in the work of Max Weber (1864-1920) (Ringer, 2002) and more recently by Cartwright (2007), who regard causes as enablers or objective possibilities that do not necessarily produce the effect. To see this, consider the following example. Suppose your bicycle gets stolen after parking it in the city centre without locking it. According to ordinary notions of causation, one might say that parking your bicycle in the city centre caused it to get stolen because not parking it indoors enabled the thief to steal it. Likewise, notions of probabilistic causation in Pearl (2009), Reichenbach (1956), and Suppes (1970) are also at odds with the requirement of necessity. For instance, it is common to say that forgetting to lock your bicycle is a cause of it getting stolen because it increases the probability of this event happening.

In pursuit of capturing the intuitive human notion of causation, philosophers like Lewis (1973) and Mackie (1974) have dedicated their work to counterfactual accounts of causation. From this viewpoint, a causal relation from A (e.g. parking your bicycle outdoors) to B (e.g. your bicycle getting stolen) relates to counterfactual claims of the form "If A had not occurred, then B would not have followed" (i.e. if you hadn't parked your bicycle outside then it would not have been stolen). Counter-examples to the counterfactual regularity account of causation has led Hitchcock (2003) to "abandon the attempt to characterize *the* causal relation" and Woodward (2003) to decide against *any* reductive theory of causation.

1.1.2 Causal explanations

Perhaps the idea of a universal account of causation as a monolithic and fundamental concept that intuitively explains phenomena in the world around us should be abandoned altogether, as Cartwright (2007), N. Hall (2004), and Hitchcock (2003) contend. In practice, humans seek out causal explanations for particular phenomena of interest by looking for answers to specific 'why?'-questions. In this light, we can treat

causation as an umbrella term for more explicit causal concepts and shift our focus towards the goal of obtaining causal explanations.

> Perhaps we should be seeking not theories of causality but rather causal theories of the world. Cartwright, 2007

From the viewpoint that causal relations are effective instruments to explain phenomena that only share "family resemblance" (Reiss, 2007), formalizing different valid causal concepts is more fruitful than the quest for a universal account of causation (Cartwright, 2007). According to Cartwright (2007), the use of the term "cause" is warranted through an understanding of *thick causal concepts*. Consider, for instance, a bimetallic thermostat consisting of a bimetallic strip (i.e. two strips of different metals bolted together) that *regulates* the temperature. As temperature *increases*, one type of metal *expands* more than the other so that the strip *bends*. If the strip *bends too far* then an electrical circuit connected to a heater is *broken* so that the heating *switches off*, after which the temperature *decreases*. With the help of very specific causal concepts and knowledge of the way in which they operate we are able to give a minute description of a thermostat controlling the temperature.

Adequate causal explanations were already contemplated by Aristotle (384 BC - 322 BC), who posited four distinct types of causal explanations, all of which he deemed necessary to wholly explain real-world phenomena (Moravcsik, 1974). For the bimetallic thermostat, the metal of the strip is a substance that undergoes change and is called a *material cause*. The design of the thermostat provides an account of what it means to be a thermostat and is called the *formal cause*. Sources of change, referred to as *efficient causes*, are the engineer who builds the thermostat and the person who sets the temperature on the thermostat. The reason that the bimetallic thermostat undergoes changes (to keep the temperature constant) is called the *final cause*. Although Aristotle's final cause (i.e. teleological causation) can be reduced to the operations of efficient causes (Evans, 1959), discarding them may result in a reduction of explanatory power when phenomena that are shaped by selection must be treated at a manageable level of description (e.g. efficient causes do not have the explanatory power of the teleological statement "Birds have wings in order to fly") (Cartwright, 1983). Note that Aristotle's four causes are more closely related to our contemporary notion of *explanation* (Evans, 1959), while only the efficient cause resembles contemporary notions of causation (Licata, 2019).

1.1.3 Interventionist perspective

We know that children not only learn about causal relations by asking for explanations, but they also obtain causal knowledge by interacting with the world around them and observing what happens (Piaget, 1930). This is very similar to the way that *interventionists* think about causality:

> I suggest that the distinguishing feature of causal explanations, so conceived, is that they are explanations that furnish information that is potentially relevant to manipulation and control: they tell us how, if we were able to change the value of one or more variables, we could change the value of other variables. Woodward, 2003

Interventionists say that "A causes B" when B changes in a predictable manner if the value of A is altered through an *intervention*. Although defining causes in terms of interventions is circular (because the definition of an intervention relies on the notion of causation) and anthropocentric (Woodward, 2003), the systematic development of the interventionist approach has turned out to be particularly useful both theoretically and practically (Pearl, 2009). This practical approach, popular among statisticians, computer scientists, economists and experimentalists, has many successes in medicine, economic, and the social sciences. In this thesis, we focus on the interventionist perspective on causation.

1.2 Causal Models

In many cases, causal relations are easy to infer and simple to explain. For example, we intuitively understand that heat causes ice to melt and that a moving billiard ball causes another ball to acquire motion after impact. However, in many cases it is much more complicated to get a good grasp of causal relations that contribute to an outcome. For example, we do not (yet) have a full understanding of the underlying causes of serious phenomena such as financial crises, climate, societal unrest, or the development of disease in the human body. It turns out that dynamical models or structural causal models have great explanatory power when it comes to making predictions about time-dependent or static behaviour of certain complex systems. This way, (causal) modelling frameworks can help us to better understand the properties of these systems.

Note that although explanatory power does not imply that a (causal) model truly represents reality (Cartwright, 1983), this would be irrelevant to the question of whether the concept of causality is admissible when all defined concepts refer to the model and not to the real-world (Simon, 1952). From a purely mathematical perspective causation is simply a concept that is defined in terms of a causal model and a mathematical operation called an *intervention*. In this section, we introduce dynamical models together with an intervention operation and explain how structural causal models may arise through *equilibration* of a dynamical model.

1.2.1 Structural causal models

Structural causal models are a popular causal modelling framework that form the basis of many statistical methods for causal inference (Pearl, 2009). Their origins can

be traced back to early work in genetics (S. Wright, 1921), econometrics (Haavelmo, 1943; P. G. Wright, 1928), electrical engineering (Mason, 1953; Mason, 1956), and the social sciences (Duncan, 1975; Goldberger et al., 1973).

To give a concrete example, consider a simple model for the grade that a student receives on an exam. Suppose that the model takes into account the effect of preparation X_{v_P} and the ability to concentrate X_{v_C} on his final test score X_{v_T}. How well the student is able to concentrate depends on the amount of time that he spent sleeping X_{v_S} during the previous night. A simple structural causal model would include a set of independent exogenous random variables $\mathbf{U} = \{U_{w_P}, U_{w_S}, U_{w_C}, U_{w_T}\}$ that determine the values of endogenous variables $\mathbf{X} = \{X_{v_P}, X_{v_S}, X_{v_C}, X_{v_T}\}$ via the *structural equations*:

$$X_{v_P} = f_P(U_{w_P}), \tag{1.1}$$
$$X_{v_S} = f_S(U_{w_S}), \tag{1.2}$$
$$X_{v_C} = f_C(X_{v_S}, U_{w_C}), \tag{1.3}$$
$$X_{v_T} = f_T(X_{v_S}, X_{v_C}, U_{w_T}), \tag{1.4}$$

where f_P, f_S, f_C, and f_T are measurable functions that represent *causal mechanisms*. The functional dependences among the endogenous variables are represented by the directed acyclic graph in Figure 1.1.

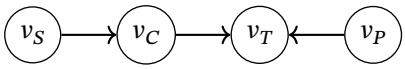


Figure 1.1: The causal graph associated with the structural equations (1.1) to (1.4).

The properties of acyclic SCMs (i.e. recursive SEMs) have been widely studied and are well-understood, see for example Lauritzen, Dawid, et al. (1990), Pearl (2009), and Spirtes, Glymour, et al. (2000). Notably, SCMs with acyclic causal graphs have an intuitive causal interpretation (e.g. the presence (absence) of directed paths in the causal graph in Figure 1.1 correspond to the presence (absence) of causal relations.). Furthermore, they induce a unique distribution over endogenous variables and they obey a Markov property. Cyclic SCMs have been proposed to model systems that contain causal cycles (Mooij, Janzing, and Schölkopf, 2013; Spirtes and Richardson, 1995). Although the convenient properties of acyclic SCMs also apply in linear and discrete cyclic SCMs, they do not hold in more general settings (Bongers, Forré, et al., 2020; Forré et al., 2017; Spirtes and Richardson, 1995). Recently, Bongers, Forré, et al. (2020) and Forré et al. (2017) showed that *modular* and *simple* SCMs retain many of the attractive properties of acyclic SCMs. Here, we will closely follow Bongers, Forré, et al. (2020) for a succinct but formal treatment of cyclic SCMs and their properties. In Section 1.2.3, we also discuss how cyclic SCMs may arise from equilibrating dynamical models.

The definition of an SCM in Bongers, Forré, et al. (2020) slightly deviates from previous notions of (acyclic) SCMs because it separates the model from the (endogenous) random variables that solve it. Due to this change, interventions on SCMs are always well-defined, even if the resulting *intervened* SCM does not have a (unique) solution. In Definition 1.1 below, we explicitly include exogenous random variables, which may be observed or unobserved, and the graph of the SCM. The endogenous random variables that solve an SCM are defined in Definition 1.2.

Definition 1.1. A structural causal model (SCM) is a tuple $\langle \mathfrak{X}, \mathbb{P}_W, \mathbf{f}, \mathcal{G} \rangle$ where

(i) $\mathfrak{X} = \bigotimes_{v \in V} \mathcal{X}_v$, with each $\mathcal{X}_v$ a standard measurable space and the domain of a variable X_v,

(ii) $\mathbb{P}_W = \prod_{w \in W} \mathbb{P}_w$ specifies the *exogenous distribution*, a product probability measure on $\bigotimes_{w \in W} \mathcal{X}_w$, where each $\mathbb{P}_w$ is a probability measure on $\mathcal{X}_w$, with $W \subseteq V$ a set of indices corresponding to exogenous variables,[1]

(iii) $\mathbf{f} : \mathfrak{X}_V \to \mathfrak{X}_{V \setminus W}$ is a measurable function that specifies *causal mechanisms*.[2]

(iv) $\mathcal{G} = \langle V, E \rangle$ is a directed graph with:

  (i) a set of nodes V corresponding to variables,

  (ii) a set of edges $E = \{(v_i \to v_j) : v_i \text{ is a parent of } v_j\}$.[3]

Definition 1.2. We say that a random variable $\mathbf{X}$ taking value in $\mathfrak{X}$ is a *solution to an SCM* $\langle \mathfrak{X}, \mathbb{P}_W, \mathbf{f}, \mathcal{G} \rangle$ if $\mathbb{P}^{\mathbf{X}_W} = \mathbb{P}_W$ (i.e., if the marginal distribution of $\mathbf{X}$ on $\mathfrak{X}_W$ equals the exogenous distribution specified by the SCM), and

$$\mathbf{X}_{V \setminus W} = \mathbf{f}(\mathbf{X}) \quad \text{a.s.} \tag{1.5}$$

The notion of *unique solvability w.r.t. a subset* is given in Definition 1.3 below.

Definition 1.3. An SCM $\langle \mathfrak{X}, \mathbb{P}_W, \mathbf{f}, \mathcal{G} \rangle$ is uniquely solvable w.r.t. $S \subseteq V \setminus W$ if there exists a measurable function $\mathbf{g}_S : \mathfrak{X}_{\mathrm{pa}_{\mathcal{G}}(S) \setminus S} \to \mathfrak{X}_S$ such that for $\mathbb{P}_W$-almost every $\mathbf{x}_W \in \mathfrak{X}_W$ and for all $\mathbf{x}_{V \setminus W} \in \mathfrak{X}_{V \setminus W}$

$$\mathbf{x}_S = \mathbf{g}_S(\mathbf{x}_{\mathrm{pa}_{\mathcal{G}}(S) \setminus S}) \iff \mathbf{x}_S = \mathbf{f}_S(\mathbf{x}). \tag{1.6}$$

SCMs that are uniquely solvable w.r.t. every subset of variables are called *simple SCMs* (Bongers, Forré, et al., 2020). It can be shown that SCMs with acyclic graphs are simple SCMs (Proposition 3.6 in Bongers, Forré, et al. (2020)). Furthermore, SCMs are uniquely solvable w.r.t. a single variable if and only if there is no self-cycle at that variable (Proposition 3.9 in Bongers, Forré, et al. (2020)). The notion of (perfect) interventions on an SCM is given in Definition 1.4.

[1] This means that the nodes $V \setminus W$ correspond to endogenous variables.

[2] The structural equations of the model are given by $x_v = f_v(\mathbf{x}), \mathbf{x} \in \mathfrak{X}$ for $v \in V \setminus W$.

[3] We say that v_i is a parent of v_j if and only if $v_j \in V \setminus W$ and there does not exist a measurable function $\tilde{f}_j : \mathfrak{X}_{V \setminus \{v_i\}} \to \mathcal{X}_j$ such that for $\mathbb{P}_W$-almost every $\mathbf{x}_W \in \mathfrak{X}_W$ and for all $\mathbf{x}_{V \setminus W} \in \mathfrak{X}_{V \setminus W}$ we have $x_j = f_j(\mathbf{x}) \iff x_j = \tilde{f}_j(\mathbf{x}_{V \setminus \{v_i\}})$, see Definition 2.7 in Bongers, Forré, et al. (2020).

Definition 1.4. Let $\mathcal{M} = \langle \mathfrak{X}, \mathbb{P}_W, \mathbf{f}, \mathcal{G} \rangle$ be an SCM, $I \subseteq V$ an *intervention target* and $\boldsymbol{\xi}_I \in \mathfrak{X}_I$ the *intervention value*. A perfect intervention $\mathrm{do}(I, \boldsymbol{\xi}_I)$ on the SCM maps it to an intervened SCM $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)} = \langle \mathfrak{X}, \mathbb{P}_W, \tilde{\mathbf{f}}, \mathcal{G}_{\mathrm{do}(I)} \rangle$ with

$$\widetilde{f}_v(\mathbf{x}) := \begin{cases} \xi_v & v \in I \\ f_v(\mathbf{x}) & v \in V \setminus I. \end{cases} \tag{1.7}$$

Cyclic SCMs may have no solution, solutions with different distributions, or all solutions may have the same distribution. This may even change as a result of a perfect intervention. Because changes in the solution after an intervention may not be compatible with the structure of the functional relations between variables, the causal interpretation of cyclic SCMs may not be intuitive (Bongers, Forré, et al., 2020). It can be shown that the graph of a simple SCMs, whose unique solvability is preserved under intervention (Proposition 8.2 in Bongers, Forré, et al. (2020)), has an intuitive causal interpretation; direct and indirect causal effects can be read off from the graph of the SCM by checking for the presence of directed edges and directed paths between variables (Bongers, Forré, et al., 2020). For general cyclic models, Bongers, Forré, et al. (2020) give a sufficient condition for detecting direct and indirect causes in an SCM with cycles. Roughly speaking, an indirect cause v_i of v_j can be detected if by controlling v_i we can bring about a change in the distribution of v_j and a direct cause v_d of v_j can be detected if by controlling v_d and keeping all other variables constant we can bring about a change in the distribution of v_j. For the exact formulation we refer to Proposition 7.1 in Bongers, Forré, et al. (2020).

1.2.2 Dynamical models

In application domains (e.g. mechanics, chemical kinetics, or economics) dynamical models describe the time-dependent behaviour of a system. However, they are often interpreted causally as well. Bertrand Russell (1872-1970), who was a vocal critic of the notion of causation, supported the idea that dynamical models are not purely descriptive because he believed that inferences from what occurs at some times to what occurs at other times has to involve a relation that is causal in nature (Russell, 1903).[4] This is in line with the view of Albert Einstein (1879-1955) on causality in physics:

> The differential law is the only form which completely satisfies the modern physicist's demand for causality. Einstein, 1927

[4] At the start of the 20th century the notion of causation had been seriously discredited in the field of physics. Russell (1903) contended that causation was an obsolete and archaic notion in a scathing critique of causation as a fundamental building block in nature: "[...] the notion of 'cause' is so inextricably bound up with misleading associations so as to make its complete extrusion from the philosophical vocabulary desirable. [...] The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm." Intriguingly, he did support the idea that dynamical models may represent causal relations.

Here, we will discuss how the causal effect on a dynamical system of an interaction with the system can be modelled by an operation called *intervention*. To do so, we will first introduce the notion of first-order random differential equations. For a thorough analysis of causality in dynamical models we refer to Bongers and Mooij (2018).

Differential equations were invented by Isaac Newton (1643-1727) and Gottfried Leibniz (1646-1716) after the invention of calculus. Throughout this thesis, we consider systems of *first-order ordinary differential equations* that relate the time-derivative of a function to other functions. The *canonical form* of a differential equation is given by:

$$\frac{dX_v(t)}{dt} = f_v(t, \mathbf{X}(t)),$$

where $\mathbf{X}(t) = (X_v(t))_{v \in V}$ is a set of functions indexed by V and taking value in $\mathbb{R}^{|V|}$. A set of ordinary differential equations (ODEs) for $v \in V$, together with an *initial condition* $(t_0, \mathbf{x}_0) \in \mathbb{R} \times \mathbb{R}^{|V|}$ is called an *initial value problem*. For Lipschitz continuous functions $\mathbf{f} = (f_v)_{v \in V} : \mathcal{D} \subset \mathbb{R} \times \mathbb{R}^{|V|} \to \mathbb{R}^{|V|}$, the Picard-Lindelöf theorem guarantees that $\forall t \geq 0$ the initial value problem has a unique solution of the form

$$\mathbf{X}(t) = \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{f}(\mathbf{X}(s))ds.$$

Uncertainty is a prominent feature in many dynamical systems.[5] The solution to a set of *random* first-order differential equations is a stochastic process. Formally, a stochastic process is a function $(X_v)_{v \in V} : T \times \Omega \to \mathbb{R}^{|V|}$ where T is an index set for time so that for all $t \in T$ we have that $X_v(t, \cdot)$ is a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *endogenous stochastic process* $(X_v)_{v \in V} : T \times \Omega \to \mathbb{R}^{|V|}$ is a solution to first-order random differential equations

$$\frac{dX_v(t, \cdot)}{dt} = f_v(\mathbf{X}(t, \cdot), \mathbf{U}(t, \cdot)), \qquad \forall v \in V,$$

with $(f_v)_{v \in V} : \mathbb{R}^{|V|} \times \mathbb{R}^{|W|} \to \mathbb{R}^{|V|}$ a measurable function and $(U_w)_{w \in W} : T \times \Omega \to \mathbb{R}^{|W|}$ an *exogenous stochastic process* if for almost all $\omega \in \Omega$

$$\frac{dX_v(t, \omega)}{dt} = f(\mathbf{X}(t, \omega), \mathbf{U}(t, \omega)), \qquad \forall v \in V.$$

The random fluctuations in many financial, biological, and physical models are naturally modelled through random differential equations with initial conditions $\mathbf{X}(t_0)$ that are determined by independent exogenous random variables $\mathbf{U} = (U_w)_{w \in W}$ taking value in $\mathbb{R}^{|W|}$. These initial conditions can be modelled by an exogenous

[5] The inclusion of noise terms can be traced back to Albert Einstein and Paul Langevin (1872-1946), who used it to explain Brownian motion in physical systems (Han et al., 2017).

stochastic process that is constant in time. Under certain regularity conditions (see (Kloeden et al., 1992) for more details) the solution for $t \geq 0$ and $\mathbf{u}(\omega) \in \mathbb{R}^{|W|}$ for $\omega \in \Omega$ to the initial value problem is then given by

$$\mathbf{X}(t, \mathbf{u}(\omega)) = \mathbf{X}(t_0, \mathbf{u}(\omega)) + \int_{t_0}^{t} \mathbf{f}\big(\mathbf{X}(s, \mathbf{u}(\omega))\big) ds.$$

We consider causal relations in a dynamical model that are defined through the effects of *interventions*, see for example Bongers and Mooij (2018) and Mooij, Janzing, and Schölkopf (2013). The options include fixing the value of targeted values at one time-point by changing the initial conditions $\mathbf{X}_I(t_0, \cdot)$ and *soft interventions* which alter parameter values in the model. A stochastic version of perfect interventions, also known as ideal, hard, structural, or independent interventions, replace an endogenous process $\mathbf{X}_I(t, \cdot)$, with $I \subseteq V$, that is targeted by an intervention $\mathrm{do}(I, \boldsymbol{\xi}_I)$ by an independent process $\boldsymbol{\xi}_I$ (Eberhardt et al., 2007). In this work, we are interested in dynamical models that converge to an equilibrium, and therefore we mostly focus on perfect interventions that fix the value of targeted variables $\mathbf{X}_I(t, \cdot)$ to a constant intervention value $\boldsymbol{\xi}_I$ for all time. This is accomplished by setting the time-derivatives of targeted variables equal to zero and replacing the initial conditions with the intervention value (Mooij, Janzing, and Schölkopf, 2013). More formally, we consider an operation $\mathrm{do}(I, \boldsymbol{\xi}_I)$ that set $\frac{dX_i(t,\omega)}{dt} = 0$ for $i \in I$, for all $t \geq 0$, and for all $\omega \in \Omega$, while putting $\mathbf{X}_I(t_0, \omega) = \boldsymbol{\xi}_I$. The *causal effects* of an intervention are the solution components of the endogenous process that solves the (intervened) system which are affected by that intervention.[6]

1.2.3 Equilibrating dynamical models

Dynamical models capture the time-dependent behaviour of systems that evolve over time. By specifying the operation of *equilibration*, we can study the stationary behaviour of the dynamical model under certain stability assumptions. Additionally, by specifying operations called interventions, we can model how the stationary and dynamic solutions to a system react to certain actions. The *causal semantics* of an equilibrated dynamical model are obtained by comparing the equilibrium state before and after an intervention.[7]

The stochastic process that solves a dynamical model may converge to a random

[6] Although mathematical analysis of the properties of causal models are of interest in and of themselves, the hope is that these models resemble certain aspects of reality so that they can be used to make causal predictions for physical systems. In other words, we would like to have that interventions relate to the ways in which a variable can be controlled in an experiment in the real world that the dynamical model purports to describe.

[7] In economics, comparative statics is a type of statistical analysis that compares equilibria before and after adjustments to an exogenous parameter in a model are made. In this thesis, we consider other types of interventions as well.

variable that solves an associated structural causal model that preserves the causal semantics. Under stability assumptions, an SCM can be obtained by *equilibrating* a dynamical model (Bongers and Mooij, 2018; Dash, 2005; Mooij, Janzing, and Schölkopf, 2013). In the deterministic setting, Mooij, Janzing, and Schölkopf (2013) showed that a set of first-order differential equations in a globally asymptotically stable system can be mapped to a set of *labelled equilibrium equations* by setting the time derivatives of variables equal to zero and labelling them as belonging to the time derivative of particular variables. If each labelled equilibrium equation can be solved for the corresponding variable then the labelled equilibrium equations can be mapped to a cyclic SCM without self-cycles. The idea that a dynamical model can be *equilibrated to an SCM* was formalized in a general stochastic setting with zeroth and higher order differential equations by Bongers and Mooij (2018), who showed how to equilibrate a *causal dynamics model* to an SCM that may contain self-cycles.

The approach is easily illustrated by a simple econometric model. Consider the following dynamical model for the price $X_P(t)$, supply $X_S(t)$, and demand $X_D(t)$ of a product:

$$\frac{dX_P(t)}{dt} = \alpha_P(X_D - X_S) \tag{1.8}$$

$$\frac{dX_S(t)}{dt} = \alpha_S(c_S X_P + U_S - X_S) \tag{1.9}$$

$$\frac{dX_D(t)}{dt} = \alpha_D(c_D X_P + U_D - X_D), \tag{1.10}$$

where U_S, U_D are independent exogenous random variables and $\alpha_P, \alpha_S, \alpha_D, c_S$, and c_D are constant parameters. From this, labelled equilibrium equations are obtained by setting the time derivatives in (1.8)-(1.10) equal to zero. The equilibrium equations obtained from (1.9) and (1.10) can easily be solved for supply and demand, resulting in the structural equations in (1.12) and (1.13) below, respectively. If we set (1.8) equal to zero then we are not able to solve for price. Bongers and Mooij (2018) show that the equilibration of the dynamical model for price, supply, and demand is given by the structural equations

$$X_P = X_P + X_D - X_S, \tag{1.11}$$

$$X_S = c_S X_P + U_S, \tag{1.12}$$

$$X_D = c_D X_P + U_D, \tag{1.13}$$

with a self-cycle at X_P. This equilibration preserves the causal semantics of the dynamical model (Bongers and Mooij, 2018). That is, the outcome of a perfect intervention targeting an endogenous variable in the SCM of the equilibrated system would be the same if these operations were applied in the reverse order (i.e. if the intervention was first applied to the dynamical model which was then equilibrated).

Several (causal) models have been proposed that may arise from equilibration in

a dynamical system. In economics, simultaneous equation models are used to model equilibrium mechanisms induced by dynamical models (F. M. Fisher, 1970). By distinguishing between *exogenous* en *endogenous* variables in a simultaneous equations model, a *causal ordering* of endogenous variables can be deduced from the symmetric equations of the model (Berndsen, 1995; De Kleer et al., 1986; Iwasaki et al., 1994; Simon, 1953). Another approach is given by Lauritzen and Richardson (2002), who show that the dependence structure of the equilibrium distribution generated by a dynamical model with feedback can be represented by (causal) chain graphs. Many authors have noted how cyclic SCMs may represent the equilibrium distribution of a dynamic process (Bongers and Mooij, 2018; Dash, 2005; Hyttinen et al., 2012; Lacerda et al., 2008; Mooij, Janzing, Heskes, et al., 2011; Mooij, Janzing, and Schölkopf, 2013; Richardson, 1996; Spirtes and Richardson, 1995).

1.3 Causal discovery and the scientific method

The progress of scientific knowledge over the past centuries was achieved by means of the *scientific method*. It is an empirical approach that usually starts out with a hypothesis, which can be either stumbled upon by pure chance or which could be formulated through several steps of abductive and inductive reasoning. Subsequently, deductive reasoning is applied to generate predictions from the hypothesis so that an experiment can be designed to either falsify or confirm the theory. When an experiment confirms observations or predicts something that was previously not known it is usually assigned greater credibility.

The strength of the scientific method is exemplified by the story of Ignaz Semmelweis (1818-1865) who discovered that bacteria are a cause of a deadly disease called childbed fever but who was ridiculed by the medical community in Vienna at the time (Semmelweis et al., 1983). Semmelweis noticed a large discrepancy between the mortality rates after childbirth in the clinic that was run by midwives and the one in which doctors were in charge. Moreover, women who gave birth to their babies in the streets or at home, had an even smaller chance of contracting and dying of childbed fever. Semmelweis considered many hypothesis for causes leading to the disease but discarded almost all of them because they were inconsistent with the data that he had managed to gather. When a professor of forensic medicine died with symptoms of the disease after sustaining a cut with a knife that was used on a cadaver, Semmelweis conjectured that cadaverous particles adhering to hands cause childbed fever in maternity patients. To confirm his suspicions he mandated hand washing to destroy the cadaverous material before examinations, after which he observed a significant drop in mortality rate. At the time, a causal relation between hand-washing and childbed fever did not fit in with the dominant narrative about the nature of diseases and Semmelweis' thoughtful analysis was met with scorn and derision. After suffering from depression he died in a psychiatric ward before his theory was confirmed

by Louis Pasteur (1822-1895). Due to this careful deliberation of facts and diligent application of the scientific method Semmelweis was able to save many lives.

1.3.1 Empirical discovery

The inductive-deductive method of Aristotle, who repeatedly applied induction to derive general principles from observations and then deduction to test the resulting theory against further observations, has influenced many philosophers. Both the physicist Alhazen (965-1040), who studied light, and mathematical scientist Al-Biruni (973-1050), who studied mechanics and mineralogy, emphasized the role of empiricism and experimentation in deriving universal truths from observations. The philosopher Avicenna (980-1037), who was one of the first to describe some of the crucial methods that are used in the scientific method, believed that experimentation follows from theories, instead of the other way around. In practice, scientific knowledge is often acquired after many iterations of induction, deduction, and experimentation.

These days, Francis Bacon (1561-1626) is often regarded as pioneer of the modern scientific method and sometimes he is even referred to as the 'father of empiricism'. Later, John Stuart Mill (1806-1873) developed four rules for experimental inquiry, thereby significantly refining the Baconian method for the identification of causal relations from observed regularities and experiments. It is easiest to explain Mill's rules with a practical example. Suppose you suffer from food poisoning after eating at a buffet with a group of friends and you wish to identify which dish caused it. To do so, you make some calls to verify who ate from which dish in order to construct Table 1.1. From this, you observe that everyone who did not eat fish remained healthy while those that did eat fish got sick. Mill's method of differences states that: "If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring in the former; the circumstances in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause of the phenomenon." From this early version of controlled experimentation, we can conclude that it was the fish that caused the sickness. The method of agreement says that if everybody who got sick ate fish while there is no other dish from which the people who got sick all ate then fish can be identified as the cause.

If it turns out that the more fish people ate the more they suffered from symptoms like nausea and stomach-ache then by the rule of concomitant variation the fish is identified as a possible cause of the food poisoning. This method can be seen as an early version of Reichenbach's principle of common cause, which states that a correlation between two variables points to either a causal relation between them or that they share a common cause. Finally, the method of residues says that when the influence of factors that do not fully explain a certain effect are subtracted from it then the residual can be attributed to an unexamined cause. Although Mill's methods

Table 1.1: Overview of the dishes that a group of friends ate indicating who got sick and how did not. Using the method of differences or the method of agreement the fish can be identified as a possible cause of the sickness.

	Alice	Bob	Charlie	Daniel
Fish	✓	×	×	✓
Chicken	×	✓	×	✓
Vegetables	✓	✓	×	×
Soup	×	×	✓	✓
Sick?	Yes	No	No	Yes

play an important role in causal reasoning and evaluating hypothesis, they are neither suitable for causal discovery nor for proving the existence of certain causal relations because they can only help to find single likely causes from a set of predetermined possibilities.

The most influential method for establishing causal relations is the *randomized controlled trial* (R. A. Fisher, 1935). The method applies to systems with a context variable C and an outcome variable X, under the condition that the outcome is not a cause of the context and the outcome and context do not share a common cause (Mooij, Magliacane, et al., 2020). In drug trials, for example, the context C indicates whether a patient is in the treatment group ($C = 1$) or in the control group ($C = 0$). When patients are randomly assigned to these groups, we can be sure that the outcome X, e.g. blood pressure, does not cause the context nor does it share a common cause with it. A statistical test can then confirm whether there is a significant dependence between the context and the outcome (whether blood pressure of patients in the treatment group is significantly lower than that of patients in the control group). Using Reichenbach's principle of common cause there must be a causal effect of the context (i.e. treatment) on the outcome.

A variant of randomized controlled trials are *natural experiments* that do not require randomization, but instead rely on background knowledge to ensure that the context variable is not caused by the outcome and does not share a common cause with it. For instance, Angrist et al. (1998) aimed to estimated the effect of family size on the career of the mother. For this they could not simply look at the correlation between the two because they expected that lifestyle choices would be a common cause of both family size and career prospects and that their would be a reverse causal effect of career success on family size. It turns out that families with two children of the same sex are more likely to have a third child. This allowed them to interpret data as an experiment where nature had randomly assigned some families to have two or three children. Although these type of experiments are sometimes successfully applied in economics, epidemiology, and sociology, the method is in many cases difficult or impossible to implement.

1.3.2 Causal discovery from observational data

In the early 90s, researchers started to develop structure learning methods for the purpose of learning a graphical representation of the dependences and independences in purely observational data. To do so, one typically assumes that the data was generated by a graphical model, which is a pair consisting of a probability distribution and a graph that encodes exactly the conditional independences in the probability distribution. Structure learning algorithm are able to recover an equivalence class of that graph from data that was sampled from that probability distribution. Graphical models originated in the work of (Gibbs, 1902), who studied the total energy of large systems of locally interacting particles and the work of (S. Wright, 1921), who studied genetics using path-analysis. Nowadays, graphical models are commonly used in Bayesian statistics and machine learning (e.g. for causal inference, automatic speech recognition, modelling gene regulatory networks, computer vision).

Structure learning algorithms can be used for causal discovery if one assumes that the directed edges in the learned graph correspond to stable and autonomous mechanisms in the real-world that can be individually manipulated without disturbing the others, see e.g. Pearl (2009) and Spirtes, Glymour, et al. (2000) for more details. In other words, under additional assumptions on the data-generating process (e.g. the underlying model is an acyclic structural causal model), it can be shown that the presence and absence of directed paths in the learned graph corresponds to the absence or presence of causal relations. Although the phrase 'correlation does not imply causation' is notorious in statistics, it can be shown that under certain assumptions on the underlying model, *causal discovery algorithms* can construct graphs representing causal relations from observational data alone. For more background on causal discovery algorithms in the literature we refer to Section 5.2.1. If the causal graph is known in advance, than do-calculus is a powerful technique to estimate the strength of causal effects from observational data, see (Forré et al., 2019; Pearl, 2009).

1.4 Thesis outline

This thesis was inspired by a desire to understand the output of causal discovery algorithms when they are applied to protein expression data. We believe that the ideas presented here are promising steps towards bridging the gap between theory and practice for causal discovery in the context of dynamical systems at equilibrium. In the remainder of this section, we will discuss our contributions in more detail. We also provide an outline of the thesis.

First, in Section 1.4.1, we discuss novel theoretical contributions to the technique of causal ordering. In Section 1.4.2, we present ways in which these techniques may be used to study and identify models with feedback from both data and model equations. In Section 1.4.3, we discuss how the *Causal Constraints Models* that we propose overcome certain limitations of SCMs for systems with feedback that have

reached equilibrium. Finally, in Section 1.4.4, we consider how we have contributed to bridging the gap between theory and practice in causal discovery from protein expression data.

1.4.1 Extending the causal ordering technique

The causal ordering algorithm of Simon (1953) takes as input a *self-contained* set of equations and returns an ordering of the endogenous variables that appear in these equations. Our contributions are as follows:

- We generalize the notion of *perfect interventions* to models consisting of a set of symmetric (i.e. non-structural) equations.
- We extend the causal ordering algorithm so that it can be applied to sets of equations that are not self-contained. These arise, for example, from the equations and constants of motion in certain dynamical systems.
- We formalize the causal ordering algorithm by defining its output as a *directed cluster graph* that we call the *causal ordering graph*. We prove that it represents the effects of soft interventions and certain perfect interventions under mild assumptions.
- We show how to construct the *Markov ordering graph* from the output of the causal ordering algorithm. We prove that it implies conditional independences between variables. We demonstrate that the distinction between Markov ordering and causal ordering prevents ambiguous causal interpretations of the model for a bathtub at equilibrium in Iwasaki et al. (1994).

These theoretical results will be presented and elaborated upon in Chapter 4. Additionally, in Chapter 5 we consider applications of the causal ordering technique to sets of equilibrium equations implied by first-order differential equations with initial conditions. We demonstrate how, under certain solvability assumptions, causal relations and conditional independences are represented by the *equilibrium causal ordering graph* and the Markov ordering graph respectively. We also use the causal ordering algorithm to construct a *dynamic causal ordering graph* from dynamic equations that represents transient effects of interventions. This sheds new light on commutation results regarding equilibration and intervention and a phenomenon called *perfect adaptation* that frequently appears in biological models.

1.4.2 Causal discovery for perfectly adapted systems

Perfect adaptation in a dynamical system is the phenomenon that one or more variables have an initial transient response to a persistent change in one of the external inputs but revert to their original value as the system converges to equilibrium. Feedback mechanisms ensure that the values of certain variables are robust against many types of external disturbances. In Chapter 5, we study the causal and probabilistic aspects of perfectly adapted systems with the help of the causal ordering algorithm:

- We provide a sufficient graphical criterion and mild assumptions under which the presence of perfect adaptation can be identified from the dynamic and equilibrium causal ordering graphs associated with the model.
- We give a sufficient condition and assumptions under which the presence of perfect adaptation can be tested for in experimental equilibrium data with background knowledge.
- We apply these ideas to several simple examples of perfectly adapted dynamical systems. We demonstrate that the presence of perfect adaptation explains the reversal of edges when such a system reaches equilibrium.

In Chapter 6 we consider qualitative predictions for the equilibrium distribution. This includes the implied presence or absence of causal relations and the presence or absence of conditional independences. We show that these predictions strongly depend on whether the model is extended with a feedback mechanism for perfect adaptation or not, and we demonstrate how the technique of causal ordering is a convenient tool to establish robustness of qualitative model predictions. We then characterize a large class of model extensions that preserve qualitative model predictions. We propose a holistic approach to causal discovery where model extensions are selected or rejected based only on data for the submodel (i.e. the model without extension). We discuss how this idea may also be used to reason about the presence or absence of feedback loops that result in perfect adaptation.

1.4.3 Beyond structural causal models

In this thesis we explore the relationship between static causal models and dynamical models at equilibrium. In Chapter 3, we show that (cyclic) SCMs have limitations when it comes to modelling the causal semantics and equilibrium distribution of perfectly adapted dynamical systems. Here, perfect interventions are modelled as experiments where the targeted variables are held at a fixed value by intervening on their causal mechanisms (i.e. by setting their time-derivatives equal to zero) and then waiting until the system has reached equilibrium again. In a nutshell, we propose a more general framework, that we call causal constraints models, to overcome the limitations of SCMs in this context. We consider the technique of causal ordering as a means to construct both a causal ordering graph that represents the effects of a possibly different set of interventions and a Markov ordering graph that encodes conditional independences from a set of active causal constraints. Our contributions are as follows:

- We provide a concrete example, corresponding to a simple biochemical reaction, for which there exists no SCM that fully captures the causal semantics at equilibrium. That is, the equilibrium distribution is not uniquely specified and the presence or absence of certain causal effects are not implied by the structural equations of the SCM. We show that, under some stability assumptions,

CCMs are capable of fully capturing the causal semantics and equilibrium distributions under intervention of dynamical models at equilibrium.
- We also provide an example, corresponding to a filling bathtub, which shows that the notion of perfect interventions on SCMs is too narrow to capture all relevant causal concepts of the system. In other words, the more general notion of interventions on symmetric equilibrium equations may reveal additional causal structure. By construction, the CCM framework models all of the interventions that are defined for SCMs. In Chapter 5 we show that the technique of causal ordering may reveal additional causal structure if it is applied to the active constraints in a CCM.
- We give examples where the equilibrium distribution of perfectly adapted dynamical systems is not faithful to the graph of its associated SCM in Chapter 5. A perfectly adapted system contains a particular type of feedback loop that ensures such a faithfulness violation in a way that does not depend on specific tuning of model parameters. We demonstrate that the Markov ordering graph associated with active causal constraints may also reveal additional conditional independences.

1.4.4 Bridging the gap between theory and practice

This PhD project started out with attempts to apply constraint-based causal discovery algorithms to a dataset of protein expression levels in cells that were treated with various stimuli. One issue that we encountered was that without knowing the true underlying causal model, we were unable to verify the correctness of our graphical outputs. Comparisons with biological consensus networks revealed that some of the pathways that we detected were oriented in the opposite direction. From this point, we set out to better understand how our observations in the data were related to the biochemical reaction networks and dynamical models that are used in cell biology. This led us to investigate causality and independence in dynamical models and how they help us to understand the output of causal discovery algorithms. Summarizing, the ideas that are presented in this thesis form an important step towards bringing the world of causal discovery closer to dynamical modelling:

- Initially, we identified the presence of unknown measurement error as a potential cause for erroneous interpretations of the output of causal discovery algorithms. More details are given in Chapter 2, where we also present a partial solution, although its application is limited.
- To fully capture certain causal semantics of dynamical systems at equilibrium, we proposed a generalization of the SCM framework that we call Causal Constraints Models (CCMs). The details of this modelling class are given in Chapter 3.
- We extended the causal ordering algorithm so that we could analyse causality and independence in sets of (equilibrium) equations and in CCMs. In particu-

lar, this technique allowed us to investigate subtle differences between graphs encoding conditional independences and those depicting the presence and absence of causal relations. These ideas are presented in Chapter 4.

- We showed that the presence of perfect adaptation is a plausible explanation for seeming differences between the direction of causal relations implied by causal discovery algorithms and biological consensus networks for protein pathways. Additionally, we outlined how this can be used to identify the presence of feedback loops that achieve perfect adaptation from model equations and from certain experiments. These results are given in Chapter 5.
- Finally, we applied the technique of causal ordering to assess the robustness of qualitative model predictions under extensions of a dynamical model at equilibrium. We demonstrated how this idea can be used to reason about a larger system, given only observations and a model for a subsystem. The details of this novel approach to model selection can be found in Chapter 6.

Chapter 2

Causal discovery in the presence of measurement error

Adaptation based on:
An upper bound for random measurement error in causal discovery,
T. Blom, A. Klimovskaia, S. Magliacane, and J.M. Mooij,
Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18).

Causal discovery algorithms aim to infer causal relations from data. To do so, they rely on several assumptions, including the absence of measurement error. This assumption is most likely violated in practical applications, which may result in erroneous results that cannot be reproduced in independent experiments. In this chapter, we propose a novel method to correct for measurement error, which relies on the strong faithfulness assumption and makes use of an upper bound for the variance of measurement error. We also show how to obtain such an upper bound for the variance of random measurement error from the covariance matrix of measured variables in a linear Gaussian model. We illustrate our method on both simulated data and real-world protein expression data.

2.1 Introduction

The discovery of causal relations is a fundamental objective in science and the interest in causal discovery algorithms has increased rapidly since they were established in the 1990s (Pearl, 2009; Spirtes, Glymour, et al., 2000). In practice, it may happen that their predictions are not reproducible in independent experiments. This chapter focuses on the ramifications of the presence of measurement error for the reliability of causal discovery. In particular, we show how measurement error is a possible explanation for incorrect or inconsistent output of causal discovery algorithms

on different datasets obtained from a single system. To mitigate adverse effects of measurement error, we introduce a novel method to estimate an upper bound for the variance of measurement error in linear Gaussian models and show how such an upper bound can be used as a partial solution to issues that are due to the presence of measurement error.

Intuitively it is easy to understand why measurement error complicates causal discovery. For example, the room temperature may cause ice-cream to melt while the reading on a thermometer is not a cause of melting ice-cream at all. Clearly, measured quantities themselves are not causes of one another even when the variables that they represent are. Consider an experiment in a gym where the type and amount of exercise X_{v_E} of participants is controlled in an experiment and where weight loss X_{v_W} at a later time is measured with great accuracy. Additionally, we have access to a noisy measurement $\widetilde{X}_{\tilde{v}_C} = X_{v_C} + M_{v_C}$ of their amount of burned calories X_{v_C} where M_{v_C} is an independent exogenous random variable representing measurement error. Suppose that the directed acyclic graph in Figure 2.1 is the causal graph associated with this system. It can be seen that even though v_E and v_W are d-separated[1] by v_C, they are not d-separated by $\tilde{v}_C$, which implies that even though exercise and weight loss are independent conditional on burned calories, they are not when we condition on the measurement $\tilde{v}_C$. In this chapter we will demonstrate that, for very large measurement error, one might even find that the *measurements* of the calories are dependent on both exercise and weight loss, but conditionally independent of exercise given weight loss. Since constraint-based causal discovery algorithms rely on conditional independences, they are not equipped to learn the underlying causal structure of latent variables that are not corrupted by measurement error. A researcher who is unaware of the measurement error could then draw incorrect conclusions (e.g. weight loss causes the burning of calories).

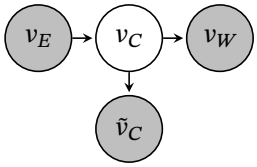


Figure 2.1: Causal graph corresponding to a simple model for amount of exercise v_E, burned calories v_C, and weight loss v_W, in the presence of measurement error on v_C. Notice that the vertices v_E and v_W are d-separated by v_C but not by $\tilde{v}_C$. Gray shaded nodes are observed variables, white nodes are latent variables.

[1] D-separation is a relation between vertices in a graph, the formal definition will be given in Definition 4.14 in Section 4.A.2 of Chapter 4.

2.1.1 Related work

A measurement process can be modelled in many different ways. Here, we follow (Kuroki et al., 2014; Pearl, 2010; Scheines and Ramsey, 2016; K. Zhang et al., 2017) and focus on *random* measurement error in structural causal models. In this case, an independent exogenous random variable representing measurement noise is added to the structural equation of a variable to obtain a structural equation for its measurement. We propose a novel method that identifies an upper bound for the variance of random measurement error in a linear Gaussian model using vanishing tetrad constraints. These results build upon previous work regarding the identification of sets of variables that are d-separated by a common latent variable, see (Bollen, 1989; Pearl, 2010; Silva et al., 2006; Sullivant et al., 2010) for more details. We show that uncertainty regarding the variance of measurement error can be propagated to an uncertainty in the partial correlations of the latent variables that are unperturbed by measurement error, see also (Harris et al., 2013). This uncertainty can then be taken into account when performing statistical tests so that we have outputs: dependent, independent, or unknown. Although these types of outputs for independence tests have been already used in previous work, e.g. (Triantafillou et al., 2017), in that case the thresholds for the different decisions were hyper-parameters of the algorithm, while we provide an adaptive and more principled way to set them. Similarly to previous work, our approach relies on strong faithfulness (Kalisch et al., 2007; Maathuis et al., 2010; Spirtes, Glymour, et al., 2000) but crucially it does not require causal sufficiency, i.e. the absence of unmeasured confounders, as K. Zhang et al. (2018) do. Recently, Saeed et al. (2020) used the method-of-moments to estimate the causal structure of the unmeasured latent variables, although they require more detailed knowledge of the measurement process.

In this work, we propose a practical correction method for measurement error in the context of constraint-based causal discovery. We demonstrate the effectiveness of our approach in identifying causal structures using Local Causal Discovery (LCD) (Cooper, 1997) on simulated data. We discuss how measurement error possibly provides an explanation for apparent discrepancies between the output of causal discovery algorithm applied to protein expression data and biological consensus networks. Furthermore, we apply the ideas that we develop in this chapter to real-world protein expression data. Although we focus on one particular causal discovery algorithm, our ideas can in principle be applied to other constraint-based causal discovery algorithms as well.

2.1.2 Preliminaries

For the remainder of this chapter, variables will be denoted by capital letters and sets of variables by bold capital letters. We will assume that the data-generating processes described here can be modelled by an acyclic Structural Causal Model (SCM)

with associated causal graph $\mathcal{G} = \langle V, E, B \rangle$ where vertices V represent latent or observed variables and E, B are sets of directed and bi-directed edges respectively. For graphs without directed cycles we say that X_v is a direct cause of X_w (where $v, w \in V$) whenever there is a directed edge from v to w in E. When there is a sequence of directed edges from v to w in $\mathcal{G}$ with all arrowheads pointing towards w (i.e. a directed path) then we say that X_v is an ancestor or indirect cause of X_w. Bi-directed edges in B between two vertices v and w are used to represent hidden confounders between variables X_v and X_w.

If $C \subseteq V$ d-separates $A \subseteq V$ from $B \subseteq V$ then we write $A \perp B \,|\, C$ (see Definition 4.14 in Section 4.A.2 of Chapter 4 for a formal definition). In the absence of measurement error, there are additional assumptions that allow us to relate conditional (in)dependences between sets of variables $\mathbf{X}_A$ and $\mathbf{X}_B$ while controlling for variables in $\mathbf{X}_C$ to d-separation in the underlying causal graph $\mathcal{G}$ (Pearl, 2009; Spirtes, Glymour, et al., 2000). Throughout the remainder of this paper we will assume that the following commonly made assumptions hold.

(i) There are no directed cycles in the causal graph.
(ii) Markov Property: For all disjoint sets of vertices $A, B, C \subseteq V$ it holds that $A \perp B \,|\, C \implies \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \,|\, \mathbf{X}_C$.
(iii) Faithfulness: For all disjoint sets of variables $\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C \subseteq \mathbf{X}$ it holds that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \,|\, \mathbf{X}_C \implies A \perp B \,|\, C$.
(iv) No selection bias is present.

The Local causal discovery (LCD) algorithm is a straightforward and efficient search method to detect one specific causal structure from experimental data using dependence and independence relations between variables combined with background knowledge (Cooper, 1997). The algorithm looks in an (experimental) dataset for triples of variables $(X_{v_i}, X_{v_j}, X_{v_k})$ for which (a) it is known that X_{v_i} is not caused by any observed variable and (b) the following (in)dependences hold: $X_{v_i} \not\perp\!\!\!\perp X_{v_j}$, $X_{v_j} \not\perp\!\!\!\perp X_{v_k}$, and $X_{v_i} \perp\!\!\!\perp X_{v_k} \mid X_{v_j}$. We will henceforth call such triples *LCD triples*. Under the common assumptions, the causal graphs that correspond to this independence pattern are shown in Figure 2.2. A conservative variant of LCD is given by Triantafillou et al. (2017) who also require that $X_{v_i} \not\perp\!\!\!\perp X_{v_k}$, $X_{v_i} \not\perp\!\!\!\perp X_{v_j} \,|\, X_{v_k}$, and $X_{v_j} \not\perp\!\!\!\perp X_{v_k} \,|\, X_{v_i}$ and use it in a real-world application to protein expression data.

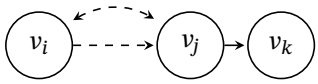


Figure 2.2: An LCD triple has the above causal structure, with at least one of the dashed arrows present.

In practice, constraint-based causal discovery algorithms rely on statistical tests

to assess the (in)dependence relationships between variables and exploit this information to construct a causal graph. For data that has a multivariate Gaussian distribution, a (conditional) independence corresponds to a vanishing (partial) correlation coefficient. For random variables $(X_1, \ldots, X_D) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the Pearson partial correlation can be calculated from the inverse covariance matrix, which we will henceforth denote by $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$. Conventionally, one calculates a p-value p_T for the (conditional) dependence between variables X and Y conditional on a set of variables $\mathbf{Z}$, so that dependence relations can be determined by

$$\begin{cases} X \not\perp\!\!\!\perp Y \mid \mathbf{Z} & \text{if } p_T < \alpha \\ X \perp\!\!\!\perp Y \mid \mathbf{Z} & \text{if } p_T > \beta, \end{cases} . \tag{2.1}$$

where α and β are thresholds for dependence and independence respectively. The nature of the relation is undecided when $\alpha \leq p_T \leq \beta$. Usually only a single threshold $\alpha = \beta = 0.01$ or $\alpha = \beta = 0.05$ is used.

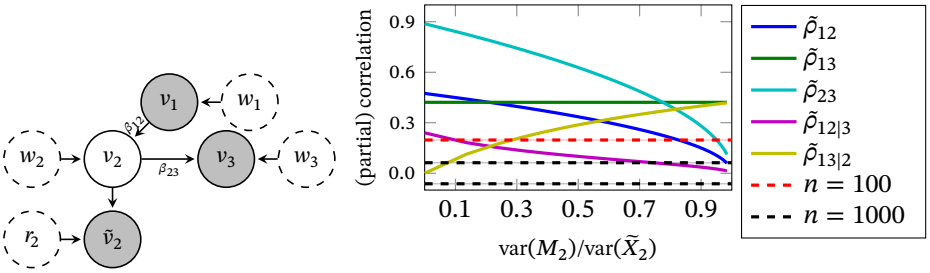
Especially in cases where many latent variables are present, e.g. when there is measurement error, there may be very few vanishing partial correlations between observed variables. Additional information about the model structure may be provided by vanishing *tetrad differences* (Scheines, Spirtes, et al., 1998). For linear SCMs these are constraints on the covariance matrix of the variables in the model that can be read off from the causal graphs associated with the SCM. For linear SCMs these constraints are graphically characterized by the Tetrad Representation Theorem (Spirtes, 1989). The advantages of analysing tetrad constraints in under-identified SCMs is discussed by Bollen and Ting (1993).

2.2 Effects of measurement error on local causal discovery

The aim of this section is to illustrate potential adverse effects on the output of constraint-based causal discovery algorithms caused by the presence of (unknown) random measurement error. To that end, we provide a detailed analysis of the behaviour of partial correlations in response to increasing measurement error in a particular SCM. We assume that observed variables are subject to *random* measurement error, which is a vector of independent noise variables $\mathbf{M} = (M_{r_1}, \ldots, M_{r_n})$. Measurements of a random vector $\mathbf{X} = (X_{v_1}, \ldots, X_{v_n})$ are then given by $\widetilde{\mathbf{X}} = (\widetilde{X}_{\tilde{v}_1}, \ldots \widetilde{X}_{\tilde{v}_2}) = \mathbf{X} + \mathbf{M}$, so that a measurement vertex $\widetilde{X}_{\tilde{v}_i}$ is always childless and has precisely two parents: X_{v_i} and the source of its measurement error M_{r_i}.

In many practical applications, it is reasonable to assume that the measurement noise follows a Gaussian distribution. For instance, when the measurement noise is the sum of many small independent sources of error, it approximates a normal distribution.[2] From here on, we will assume that measurement errors are independ-

[2]The central limit theorem in probability theory establishes that the sum of independent random vari-



(a) Causal graph. (b) Partial correlations.

Figure 2.3: The causal graph and partial correlations of endogenous variables in the SCM with structural equations (2.2) to (2.5). There is random measurement error on X_{v_2}. Gray shaded vertices are observed variables (the others are latent) and dashed vertices represent independent exogenous random variables (the others are endogenous). The coefficients alongside the arrows represent the coefficients in the model. The partial correlations in the random measurement error model are given in Figure 2.3(b). The dotted lines represent the critical values for the correlations at a significance level of $\alpha = 5\%$ for sample sizes $n = 100$ and $n = 1000$. The partial correlations were calculated for noise variables following a standard normal distribution and parameters $\beta_{12} = 0.6$, $\beta_{23} = 1.2$.

ent and follow a Gaussian distribution, i.e. the measurement noise variables satisfy $\mathbf{M} = (M_{r_1}, \dots, M_{r_n}) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{M}})$, where $\boldsymbol{\Sigma}_M$ is a diagonal covariance matrix.

2.2.1 Partial correlations

We illustrate the effects of measurement error on a structural causal model with observed endogenous variables $X_{v_1}, \widetilde{X}_{\tilde{v}_2}, X_{v_3}$, latent variable X_{v_2}, and independent exogenous random variables $U_{w_1}, U_{w_2}, U_{w_3}, M_{r_2}$ following a Gaussian distribution, with the following structural equations:

$$X_{v_1} = U_{w_2} \tag{2.2}$$
$$X_{v_2} = \beta_{12} X_{v_1} + U_{w_2} \tag{2.3}$$
$$X_{v_3} = \beta_{23} X_{v_2} + U_{w_3} \tag{2.4}$$
$$\widetilde{X}_{\tilde{v}_2} = X_{v_2} + M_{r_2}, \tag{2.5}$$

This model contains a latent variable X_{v_2} which is corrupted by independent random measurement error M_{r_2}, while both X_{v_1} and X_{v_3} are not affected by measurement error. To be able to apply the LCD algorithm, we will consider the fact that X_{v_1} is not caused by any of the endogenous variables in the SCM to be background knowledge that can be used by LCD. The corresponding causal graph is displayed in Figure 2.3(a).

From this, we see that the triple $(X_{v_1}, X_{v_2}, X_{v_3})$ has the structure and properties of

ables tends to a normal distribution, regardless of the distribution of the original random variables.

an LCD triple. Therefore we have that $X_{v_1} \perp\!\!\!\perp X_{v_3} \,|\, X_{v_2}$ and the partial correlation for the latent unmeasured variables satisfies $\rho_{13|2} = 0$. Let $\widetilde{\boldsymbol{\Sigma}}$ be the covariance matrix of $(X_{v_1}, \widetilde{X}_{\tilde{v}_2}, X_{v_3})$ and $\widetilde{\boldsymbol{\Lambda}}$ its inverse. By definition of partial correlation:

$$\tilde{\rho}_{13|2} = -\frac{\widetilde{\Lambda}_{13}}{\sqrt{\widetilde{\Lambda}_{11}\widetilde{\Lambda}_{33}}} \tag{2.6}$$

$$= \frac{-1}{\sqrt{\widetilde{\Lambda}_{11}\widetilde{\Lambda}_{33}}} \cdot \frac{\widetilde{\Sigma}_{12}\widetilde{\Sigma}_{23} - \widetilde{\Sigma}_{13}\widetilde{\Sigma}_{22}}{|\widetilde{\boldsymbol{\Sigma}}|} \tag{2.7}$$

$$= \frac{-1}{\sqrt{\widetilde{\Lambda}_{11}\widetilde{\Lambda}_{33}}} \cdot \frac{\Sigma_{12}\Sigma_{23} - \Sigma_{13}\left(\Sigma_{22} + \mathrm{var}\left(M_{r_2}\right)\right)}{|\widetilde{\boldsymbol{\Sigma}}|} \tag{2.8}$$

$$= \frac{\beta_{12}\beta_{23}\widetilde{\Sigma}_{11}\mathrm{var}\left(M_{r_2}\right)}{|\widetilde{\Sigma}|} \cdot \frac{-1}{\sqrt{\widetilde{\Lambda}_{11}\widetilde{\Lambda}_{33}}} \neq 0, \tag{2.9}$$

for non-zero parameters, so that $X_{v_1} \not\perp\!\!\!\perp X_{v_3} \,|\, \widetilde{X}_{\tilde{v}_2}$. Therefore $(X_{v_1}, \widetilde{X}_{\tilde{v}_2}, X_{v_3})$ is not an LCD triple.

Remark 2.1. A statistical test with conventional thresholds would conclude that X_{v_1} and X_{v_3} are conditionally dependent conditional on the measurement $\widetilde{X}_{\tilde{v}_2}$, if the measurement error would be large enough. If we would incorrectly assume that there is no measurement error, so that $X_{v_2} = \widetilde{X}_{\tilde{v}_2}$, then the Markov assumption would appear to be violated. △

2.2.2 Empirical study

To better understand the way in which measurement error could impact the output of causal discovery algorithms in real-world applications, we consider the effect of varying the measurement error variance var(M_2) relative to the total variance of the measurement $\widetilde{X}_{\tilde{v}_2}$ on the partial correlations in the SCM in the previous section.

The effects of increasing this relative random measurement error can be seen in Figure 2.3(b). The dotted lines represent the $\alpha = 0.05$ threshold at different sample sizes and (partial) correlations above this threshold are often considered to be dependent. We see that for zero measurement error (so that $\widetilde{X}_{\tilde{v}_2} = X_{v_2}$), only the yellow line is below the red and black dotted lines. In that case a conventional statistical test would indicate that all variables are marginally dependent while $X_{v_1} \perp\!\!\!\perp X_{v_3} \,|\, \widetilde{X}_{\tilde{v}_2}$. From this the LCD algorithm would conclude that $(X_{v_1}, \widetilde{X}_{\tilde{v}_2}, X_{v_3})$ is an LCD triple, and hence we would be able to detect the directed edge from $\widetilde{X}_{\tilde{v}_2}$ to X_{v_3}. For relative measurement errors larger than approximately 0.25 this conditional independence would no longer be detected by testing for vanishing partial correlations at the 5% level (because the yellow line is above the black-dotted line).

From Figure 2.3(b) we also observe that for a sample size of 100 and a relative measurement error larger than approximately 0.3, a conventional statistical test would indicate that $X_{v_1} \perp\!\!\!\perp \widetilde{X}_{\tilde{v}_2} \,|\, X_{v_3}$ since the partial correlation $\tilde{\rho}_{12|3} \approx 0$ (i.e. below the red dotted line) and all other (partial) correlations indicate a dependence (i.e. above the red dotted line). From these constraints, the LCD algorithm would conclude that $(X_{v_1}, X_{v_3}, \widetilde{X}_{\tilde{v}_2})$ is an LCD triple, suggesting the presence of a directed edge from X_{v_3} to $\widetilde{X}_{\tilde{v}_2}$ in the *reverse* direction.

Remark 2.2. The results of constraint-based causal discovery algorithms like LCD may depend on the sample size. This can be better understood by observing that the dependences that are identified by a statistical test, depend both on the size of the measurement error and the sample size. This way, we may obtain inconsistent causal discoveries that cannot be reproduced on new datasets. △

This example shows how measurement error interferes with detecting the correct causal structures, which may lead to edge deletions, insertions or reversals. Note that although we focused on the LCD algorithm here, the conclusions that we draw are more generally applicable to constraint-based causal discovery algorithms.

Remark 2.3. For relative measurement error of approximately 0.25 a conflicting set of (in)dependences arises for $n = 100$. Since both the yellow and purple line are below the red dotted line, a statistical test would indicate that $X_{v_1} \perp\!\!\!\perp X_{v_3} \,|\, \widetilde{X}_{\tilde{v}_2}$ and $X_{v_1} \perp\!\!\!\perp \widetilde{X}_{\tilde{v}_2} \,|\, X_{v_3}$, while all variables are marginally dependent. There is no model that satisfies both the common assumptions and this set of (in)dependences, resulting in a conflict of conditional independences. △

2.3 Upper bound for measurement error variance

In this section we show how, under certain conditions, an upper bound for the variance of random measurement error can be obtained from observational data with random measurement error.

2.3.1 Theoretical results

First, consider the definition of random measurement error, which implies that the true covariance matrix $\boldsymbol{\Sigma}$ of random variables $X_{v_1}, \ldots, X_{v_n}$, the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{M}}$ of random measurement errors $M_{r_1}, \ldots, M_{r_n}$, and the covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ of measurements $\widetilde{X}_{\tilde{v}_1}, \ldots, \widetilde{X}_{\tilde{v}_n}$ are related as follows:

$$\Sigma_{\mathbf{M}} = \widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} = \operatorname{diag}(\sigma^2_{r_1}, \ldots, \sigma^2_{r_n}),$$

where $\sigma^2_{r_1}, \ldots, \sigma^2_{r_n} > 0$ are the variances of random measurement errors associated with each variable.

Remark 2.4. Given an (unbiased) estimate of $\boldsymbol{\Sigma}_{\mathbf{M}}$, we can simply adjust the covariance matrix $\widetilde{\Sigma}$ as suggested by Pearl (2010). In practical applications such an estimate of the covariance matrix of measurement error might not be available. △

To detect an upper-bound for random measurement error, we consider a minimum of four latent random variables $X_{v_1}, \dots, X_{v_4}$ and their corresponding measurements $\widetilde{X}_{\tilde{v}_1}, \dots, \widetilde{X}_{\tilde{v}_4} \in V$ with covariance matrices $\boldsymbol{\Sigma}$ and $\widetilde{\boldsymbol{\Sigma}}$ respectively. Our detection method relies on Lemma 2.1 which is due to Silva et al. (2006) and gives conditions under which there exists a latent variable that d-separates the measured variables $\widetilde{X}_{\tilde{v}_1}, \dots, \widetilde{X}_{\tilde{v}_4}$.[3]

Lemma 2.1. *Let $X_{v_1}, \dots, X_{v_4}$ be variables in a linear Gaussian model and let $\widetilde{X}_{\tilde{v}_1}, \dots, \widetilde{X}_{\tilde{v}_4}$ be their measurements with random measurement error. If the correlations satisfy $\tilde{\rho}_{ij} \neq 0$ for all $i, j \in \{1, \dots, 4\}$ and $\widetilde{\Sigma}_{12}\widetilde{\Sigma}_{34} = \widetilde{\Sigma}_{13}\widetilde{\Sigma}_{24} = \widetilde{\Sigma}_{14}\widetilde{\Sigma}_{23}$, then there exists a vertex v_l in the true underlying DAG such that $\widetilde{X}_{\tilde{v}_i} \perp \widetilde{X}_{\tilde{v}_j} \mid v_l$ for all $i \neq j \in \{1, \dots, 4\}$.*

Proof. The proof can be found in Silva et al. (2006). □

When there exists a vertex v_l that d-separates $\widetilde{X}_{\tilde{v}_1}, \dots, \widetilde{X}_{\tilde{v}_4}$, then the causal graph and latent structure are represented by the causal graph in Figure 2.4. This follows from the fact that the variables with measurement error $\widetilde{X}_{\tilde{v}_i}$ can only have incoming arrows from X_{v_i} and M_{r_i} and hence they never have any outgoing arrows. Because v_l d-separates all $\widetilde{X}_{\tilde{v}_i}$ it follows that there is no collider at v_l with arrow coming from these vertices.

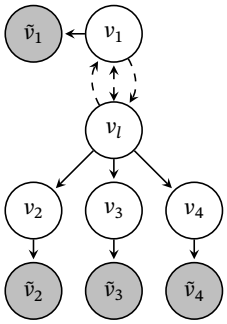


Figure 2.4: Causal graph of upper bound pattern for model with random measurement error, where at least one of the dashed edges is present. Gray shaded vertices are observed, while the other vertices are latent. The indices 1, ... 4 can be permuted. Vertices representing exogenous random variables $U_{w_1}, \dots, U_{w_4}$ that serve as noise variables are not drawn, neither are the vertices representing random measurement errors $M_{r_1}, \dots, M_{r_4}$.

[3] These conditions are known as tetrad conditions in the literature, see Bollen (1989), Drton et al. (2008), and Sullivant et al. (2010) for more details.

In Theorem 2.1 we will use Lemma 2.1 to obtain a necessary and sufficient condition to detect the pattern in Figure 2.4 from the covariance matrix of the observed variables. To do so, we first introduce the notion of an *adjusted covariance matrix*:

$$\widetilde{\boldsymbol{\Sigma}}(a, j) = \widetilde{\boldsymbol{\Sigma}} - a\,\mathrm{diag}(e_j),$$

where $j \in \{1, 2, 3, 4\}$ and e_j is a standard basis vector. For all a such that $\widetilde{\boldsymbol{\Sigma}}(a, j)$ is a valid covariance matrix, the associated *adjusted partial correlations* $\tilde{\rho}^a_{ik|j}$ may be calculated from $\widetilde{\boldsymbol{\Lambda}}(a, j) = (\widetilde{\boldsymbol{\Sigma}}(a, j))^{-1}$ as follows:

$$\tilde{\rho}^a_{ik|j} = -\frac{(\widetilde{\boldsymbol{\Lambda}}(a, j))_{ik}}{\sqrt{(\widetilde{\boldsymbol{\Lambda}}(a, j))_{ii}(\widetilde{\boldsymbol{\Lambda}}(a, j))_{kk}}}. \tag{2.10}$$

Theorem 2.1 shows how the adjusted partial correlation is related to the underlying causal graph in Figure 2.4. Corollary 2.1 shows how we can use adjusted partial correlations to find an upper bound for the measurement error on one variable.

Theorem 2.1. *Let $X_{v_1}, \dots, X_{v_4}, \widetilde{X}_{\tilde{v}_1}, \dots, \widetilde{X}_{\tilde{v}_4}$ and $\tilde{\rho}_{ij}$ be as in Lemma 2.1. The true underlying DAG is as in Figure 2.4 if and only if there exists $a > 0$ such that $\tilde{\rho}^a_{13|2} = \tilde{\rho}^a_{14|2} = \tilde{\rho}^a_{34|2} = 0$.*

Proof. Throughout the proof, we denote the covariance between variables $\widetilde{X}_{\tilde{v}_i}$ and $\widetilde{X}_{\tilde{v}_j}$ as $\widetilde{\Sigma}_{ij}$ for $i, j \in \{1, 2, 3, 4\}$. We start with proving the direction ' $\Longleftarrow$ ':

$$\begin{aligned}
&\begin{cases}
\tilde{\rho}^a_{13|2} = 0 \iff \widetilde{\Sigma}_{12}\widetilde{\Sigma}_{23} - \widetilde{\Sigma}_{13}(\widetilde{\Sigma}_{22} - a) = 0 \\
\tilde{\rho}^a_{14|2} = 0 \iff \widetilde{\Sigma}_{12}\widetilde{\Sigma}_{24} - \widetilde{\Sigma}_{14}(\widetilde{\Sigma}_{22} - a) = 0 \\
\tilde{\rho}^a_{34|2} = 0 \iff \widetilde{\Sigma}_{32}\widetilde{\Sigma}_{24} - \widetilde{\Sigma}_{34}(\widetilde{\Sigma}_{22} - a) = 0
\end{cases} \\
&\iff (\widetilde{\Sigma}_{22} - a) = \frac{\widetilde{\Sigma}_{12}\widetilde{\Sigma}_{23}}{\widetilde{\Sigma}_{13}} = \frac{\widetilde{\Sigma}_{12}\widetilde{\Sigma}_{24}}{\widetilde{\Sigma}_{14}} = \frac{\widetilde{\Sigma}_{23}\widetilde{\Sigma}_{24}}{\widetilde{\Sigma}_{34}} \\
&\implies \frac{\widetilde{\Sigma}_{23}}{\widetilde{\Sigma}_{13}} = \frac{\widetilde{\Sigma}_{24}}{\widetilde{\Sigma}_{14}}, \qquad \frac{\widetilde{\Sigma}_{12}}{\widetilde{\Sigma}_{14}} = \frac{\widetilde{\Sigma}_{23}}{\widetilde{\Sigma}_{34}} \\
&\iff \widetilde{\Sigma}_{12}\widetilde{\Sigma}_{34} = \widetilde{\Sigma}_{13}\widetilde{\Sigma}_{24} = \widetilde{\Sigma}_{14}\widetilde{\Sigma}_{23}.
\end{aligned}$$

The result follows by applying Lemma 2.1 and observing that these are the only structures for a random measurement model where all d-separations hold.

Next we prove the direction ' $\Longrightarrow$ '. If the true underlying causal graph is as in Figure 2.4, we have that there exists $\alpha \neq 0$ such that $\widetilde{X}_{\tilde{v}_2} = \alpha X_{v_l} + U_{w_2} + M_{r_2}$, where U_{w_2} is an independent noise variable with variance τ for X_{v_2} and M_{r_2} is an independent random measurement error for $\widetilde{X}_{\tilde{v}_2}$. The covariance between X_{v_l} and

variables $(X_{v_1}, X_{v_2}, X_{v_3}, X_{v_4})$ are denoted as Σ_{li} and Σ_{il} for $i \in \{1,2,3,4\}$. Hence

$$\mathrm{Cov}(\widetilde{X}_{\tilde{v}_1}, \widetilde{X}_{\tilde{v}_2}, \widetilde{X}_{\tilde{v}_3}) = \begin{pmatrix} \Sigma_{11} + m_1 & \alpha\sigma_{l1} & \Sigma_{13} \\ \alpha\Sigma_{l1} & \alpha^2\Sigma_{ll} + \tau + m_2 & \alpha\Sigma_{l3} \\ \Sigma_{13} & \alpha\Sigma_{l3} & \Sigma_{33} + m_3, \end{pmatrix}.$$

where m_1, m_2, m_3 are the variances of $M_{r_1}, M_{r_2}, M_{r_3}$ respectively and Σ_{ll} denotes the variance of the latent variable X_{v_l}. By the definition of the adjusted partial correlation we obtain:

$$\tilde{\rho}^a_{13|2} = 0 \qquad \Longleftrightarrow \qquad \alpha^2(\Sigma_{l1}\Sigma_{l3} - \Sigma_{13}\Sigma_{ll}) - \Sigma_{13}(\tau + m_2 - a) = 0.$$

Since v_l d-separates v_1 and v_3, the Markov assumption implies that $\tilde{\rho}_{13|l} = 0$ and hence we have that

$$\Sigma_{l1}\Sigma_{l3} - \Sigma_{13}\Sigma_{ll} = 0.$$

Because $\Sigma_{13} \neq 0$ by assumption, we find that $\tilde{\rho}^a_{13|2} = 0$ if and only if $a = \tau + m_2$. Via a similar argument we can show that for this a we also have that $\tilde{\rho}^a_{14|2} = 0$ and $\tilde{\rho}^a_{34|2} = 0$. □

Corollary 2.1. *Let $X_{v_1}, \ldots, X_{v_4}$, $\widetilde{X}_{\tilde{v}_1}, \ldots, \widetilde{X}_{\tilde{v}_4}$ and $\tilde{\rho}_{ij}$ be as in Lemma 2.1, and m_2 the variance of the measurement error on X_{v_2}. If $\tilde{\rho}^{u^*}_{13|2} = \tilde{\rho}^{u^*}_{14|2} = \tilde{\rho}^{u^*}_{34|2} = 0$ for some $u^* > 0$ then $m_2 \leq u^*$.*

Proof. This follows from the proof of Theorem 2.1, which shows that $u^* = \tau + m_2$, where τ is the variance of the latent variable X_{v_l}. □

Corollary 2.1 shows that an adjustment on the covariance matrix that results in vanishing tetrad constraints (i.e. the equalities in Lemma 2.1) is an upper bound for the measurement error on the adjusted variable. In practical settings, we can test for the constraints in Lemma 2.1 (see (Bollen, 1989; Silva et al., 2006; Thoemmes et al., 2018)). When all variables are measured in a similar manner, it may be reasonable to assume that the variance of the measurement error is the same for all variables. Under this assumption, the upper bound for the measurement error can be extended to an upper bound for the measurement error variance on all variables.

2.3.2 Data simulations

To empirically test the performance of the upper bound, we simulated 10000 data points for 10000 random models with causal structures as in Figure 2.4 with parameters chosen uniformly from the interval $[-1, 1]$ and error variances chosen uniformly from the interval $[0.5, 1]$. We added random measurement error to each variable with the same variance for all variables and minimized the sum of adjusted partial correlations in Corollary 2.1 to obtain an upper bound. Figure 2.5 shows that this leads to a correct upper bound on the variance of the measurement error.

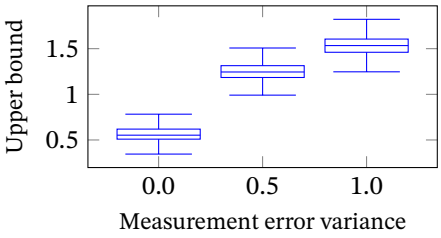


Figure 2.5: Detected upper bounds for measurement error in linear Gaussian models with a structure as in Figure 2.4 where either no measurement error or measurement error with a variance of 0.5 or 1.0 was applied.

2.4 Measurement error correction for independence testing

The following proposition shows that conditional independences between variables cannot be reliably tested given only noisy observations that have been corrupted with measurement error.

Proposition 2.1. *Let* $\mathbf{X}, \mathbf{Y}$ *and* $\widetilde{\mathbf{Z}}$ *be three sets of (disjoint) variables. If* $\widetilde{\mathbf{Z}}$ *has measurement error with non-zero variance, then the (in)dependences*

$$\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} \qquad \mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \widetilde{\mathbf{Z}},$$

must be due to a violation of the faithfulness assumption.

Proof. A faithfulness violation occurs when $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \widetilde{\mathbf{Z}}$ but $\widetilde{\mathbf{Z}}$ does not d-separate $\mathbf{X}$ and $\mathbf{Y}$. Since $\mathbf{X}$ and $\mathbf{Y}$ are dependent in the data there must be an open path between them by the Markov assumption. By definition of random measurement error the variables in $\widetilde{\mathbf{Z}}$ are leaf nodes. Therefore $\widetilde{\mathbf{Z}}$ cannot block the path between $\mathbf{X}$ and $\mathbf{Y}$, and hence $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} | \widetilde{\mathbf{Z}}$. □

In this section we will discuss how, under the strong faithfulness assumption, an upper-bound for the variance of measurement error can be propagated to refine conventional independence testing for cases where only noisy measurements are available.

2.4.1 Strong faithfulness assumption

Proposition 2.2 shows that, under the assumption that all variables in the model have the same measurement error variance (e.g. because they are subject to the same source of measurement error), the variance of the measurement error must be zero whenever a marginal dependence and a conditional independence is detected.

Proposition 2.2. *Let $\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ be three sets of (disjoint) variables with measurement errors that have equal (possibly zero) variances. Under the faithfulness assumption, if $\widetilde{\mathbf{X}} \not\perp\!\!\!\perp \widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{X}} \perp\!\!\!\perp \widetilde{\mathbf{Y}} | \widetilde{\mathbf{Z}}$, then the measurement error on all variables has zero variance.*

Proof. Follows directly from Proposition 2.1. □

Constraint-based causal discovery algorithms rely both on the faithfulness assumption and on the results of conditional independence tests and therefore poor performance is to be expected when variables are measured with error. To remedy this, we need to make additional assumptions and in this chapter we consider how issues related to measurement error can be alleviated under the *strong faithfulness* assumption (see e.g. Spirtes, Glymour, et al. (2000) for more details on strong faithfulness).

Definition 2.1. (*Strong faithfulness*) We assume that the data of the unobserved measurement-error-free variables is λ-strong faithful to the true underlying causal graph that generated it. That is, for all disjoint sets of variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$:

$$|\rho_{\mathbf{X},\mathbf{Y}|\mathbf{Z}}| < \lambda \implies \mathbf{X} \perp \mathbf{Y} \,|\, \mathbf{Z}.$$

The example in Section 2.2.2 and the correlations in Figure 2.3(b) illustrate how the strong faithfulness assumption can be leveraged to adjust for the presence of measurement error, although this could aggravate the risk of false negative test results. To see this, first note that λ-faithful data is also μ-faithful for all $0 < \mu \leq \lambda$, so that μ can be treated as a tuning parameter. Figure 2.3(b) shows that for zero measurement error all non-zero correlations are larger than approximately 0.25. We conclude that the data is μ-faithful to the causal graph, as long as μ does not exceed 0.25. If we would use 0.25 as a threshold for (partial) correlations to detect (conditional) independences, then we would find that $X_{\nu_1} \perp\!\!\!\perp X_{\nu_3} \,|\, \widetilde{X}_{\tilde{\nu}_2}$ up to a relative measurement error of approximately 0.3. But for even larger relative measurement error, we would wrongly conclude that $X_{\nu_1} \perp\!\!\!\perp \widetilde{X}_{\tilde{\nu}_2} \,|\, X_{\nu_3}$.[4] The tuning parameter μ thus represents a trade-off between detecting as many as possible of the true conditional independences and wrongfully detecting conditional independences.

2.4.2 Error propagation

In this section we consider propagation of an error bound on random measurement error to partial correlations in the LCD setting. As we saw in the previous section, the presence of measurement error may result in the failure to detect LCD triples from noisy observations while relying on the strong faithfulness assumption may result in

[4] Small enough correlations correspond to d-separations in the underlying graph by the strong faithfulness assumption and by the causal Markov assumption d-separations correspond to conditional independences. Therefore, a conditional independence between X_{ν_1} and $\widetilde{X}_{\tilde{\nu}_2}$ given X_{ν_3} would conflict with the causal graph in Figure 2.3(a).

wrongfully detecting LCD triples because of some false negatives. We aim to alleviate the adverse effect of wrongfully detecting conditional independences by including the possibility to adaptively assign 'unknown' to a statistical test result. In that case we aim to get the best of both worlds by detecting both the true conditional independences and by assigning either 'unknown' or 'dependent' to the true conditional dependences.

To accomplish this, we start by defining an adjusted covariance matrix for three variables. Let $\mathbf{m} = (m_1, m_2, m_3)$ be the variances of the random measurement errors $(M_{r_1}, M_{r_2}, M_{r_3})$ on the latent (unmeasured) variables $(X_{v_1}, X_{v_2}, X_{v_3})$, and suppose that $\mathbf{u}^* = (u_1^*, u_2^*, u_3^*)$ is an upper bound such that $\mathbf{m} \preceq \mathbf{u}^*$.[5] Suppose that $\widetilde{\boldsymbol{\Sigma}}$ is the true covariance matrix of the measured variables $\widetilde{X}_{\tilde{v}_1}, \widetilde{X}_{\tilde{v}_2}, \widetilde{X}_{\tilde{v}_3}$. The adjusted covariance matrix is given by

$$\widetilde{\boldsymbol{\Sigma}}(\mathbf{u}) = \widetilde{\boldsymbol{\Sigma}} - \mathbf{u}^T \mathbf{I}, \tag{2.11}$$

where $\mathbf{I}$ denotes the identity matrix. For $\mathbf{0} \preceq \mathbf{u} \preceq \mathbf{u}^*$ we can find minimal and maximal absolute values of partial correlations based on $\widetilde{\boldsymbol{\Lambda}}(\mathbf{u}) = (\widetilde{\boldsymbol{\Sigma}}(\mathbf{u}))^{-1}$, assuming that $\widetilde{\boldsymbol{\Sigma}}(\mathbf{u})$ has an inverse. We define

$$\tilde{\rho}^{\min}_{12|3} = \underset{\mathbf{0} \preceq \mathbf{u} \preceq \mathbf{u}^*}{\arg\min} \left| \frac{(\widetilde{\Lambda}(\mathbf{u}))_{12}}{\sqrt{(\widetilde{\Lambda}(\mathbf{u}))_{11}(\widetilde{\Lambda}(\mathbf{u}))_{22}}} \right|, \tag{2.12}$$

$$\tilde{\rho}^{\max}_{12|3} = \underset{\mathbf{0} \preceq \mathbf{u} \preceq \mathbf{u}^*}{\arg\max} \left| \frac{(\widetilde{\Lambda}(\mathbf{u}))_{12}}{\sqrt{(\widetilde{\Lambda}(\mathbf{u}))_{11}(\widetilde{\Lambda}(\mathbf{u}))_{22}}} \right|. \tag{2.13}$$

Under the λ-strong faithfulness assumption, the conditional (in)dependence relations can be determined as follows:

$$\begin{cases} X_{v_1} \not\perp\!\!\!\perp X_{v_2} \,|\, X_{v_3} & \text{if } \tilde{\rho}^{\min}_{12|3} > \lambda \\ X_{v_1} \perp\!\!\!\perp X_{v_2} \,|\, X_{v_3} & \text{if } \tilde{\rho}^{\max}_{12|3} < \lambda, \end{cases} . \tag{2.14}$$

The nature of the relation is undecided when $\tilde{\rho}^{\min}_{12|3} < \lambda$ and $\tilde{\rho}^{\max}_{12|3} > \lambda$.[6]

Although we consider a measurement error correction in cases where only one variable is conditioned upon, our ideas can be trivially extended to accommodate larger conditioning sets when an upper bound on the measurement error is known for all variables involved.[7]

[5] $\preceq$ is the component-wise inequality between two vectors.

[6] In practical applications the covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ is estimated from data. The added uncertainty can be taken into account by using bootstrapping to obtain confidence intervals for $\tilde{\rho}^{\min}_{12|3}$ and $\tilde{\rho}^{\max}_{12|3}$.

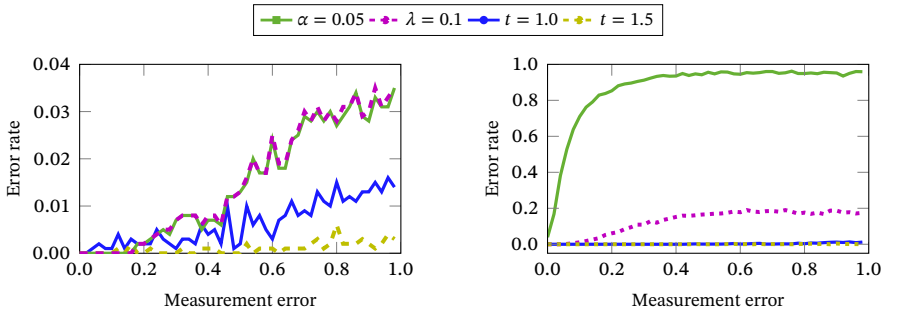
[7] In that case one considers a larger adjusted covariance matrix, and since the partial correlations are calculated from the covariance matrix one can use the same scheme to find minimal and maximal values for the absolute partial correlation.

2.4.3 Data simulations

To illustrate the effectiveness of measurement error correction for the identification of conditional dependence relations between unobserved variables for which we only have noisy measurements, we generated data for three variables $(X_{v_1}, X_{v_2}, X_{v_3})$ from linear Gaussian acyclic causal structures, possibly with latent confounders. We generated random DAGs for 6 variables with a connection probability of 0.7, parameters chosen uniformly at random from the interval $[-1.0, 1.0]$, and error variances chosen uniformly from the interval $[0.5, 1.0]$. We let three out of the six variables be observed variables, and the other three were considered to be latent. We used rejection sampling to select models for which the observed variables $(X_{v_1}, X_{v_2}, X_{v_3})$ satisfied: the λ-strong faithfulness assumption for $\lambda = 0.1$, $X_{v_1} \not\perp\!\!\!\perp X_{v_2}$, $X_{v_2} \not\perp\!\!\!\perp X_{v_3}$. We selected 2000 models with the conditional dependence $X_{v_1} \not\perp\!\!\!\perp X_{v_3} \,|\, X_{v_2}$ and 2000 models with the conditional independence $X_{v_1} \perp\!\!\!\perp X_{v_3} \,|\, X_{v_2}$. For each model, we generated 10000 data points and added normally distributed measurement error with mean zero and varying variances. The conditional (in)dependence between $\widetilde{X}_{\tilde{v}_1}$ and $\widetilde{X}_{\tilde{v}_3}$ given $\widetilde{X}_{\tilde{v}_2}$ was tested in various ways: using a threshold on the p-value $\alpha = 0.05$, using a threshold $\lambda = 0.1$ on the partial correlation, and using the same threshold with a measurement error correction with an upper bound on the measurement error of t times the true variance. We then calculated the error rate as the number of incorrect classifications relative to the total number of tests. Figure 2.6(a) shows that the measurement error correction slightly reduces the error rate for conditional dependences, and 2.6(b) shows that the error rate of detecting incorrect conditional independences is greatly reduced. Figures 2.6(c) and 2.6(d) show that the rate of 'unknowns' combined with false positives and false negatives respectively, increases with the size of the measurement error and the looseness of the upper bound that is used. This increase is mostly due to an increase in the number of assigned 'unknowns'.

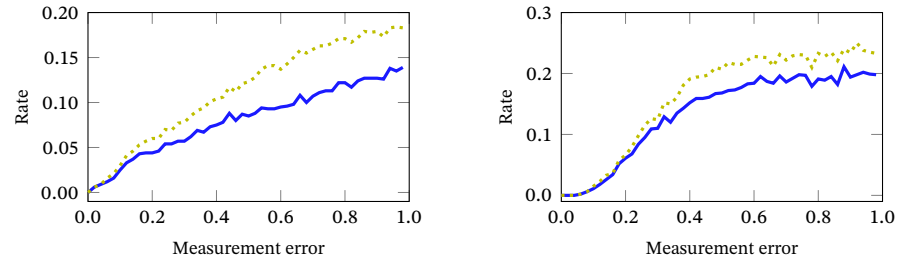
2.5 Experiments with local causal discovery

Both in theory and in practical simulations a measurement error correction on the threshold for a conditional independence test may help to improve the accuracy of testing for dependence and independence of variables from noisy measurements. This comes at the price of introducing both a tuning parameter λ (which was derived from the notion of the λ-strong faithfulness condition) and having to deal with 'unknown' output of a conditional independence test in causal discovery. In this section we will show how the measurement error correction may improve the accuracy of local causal discovery in the presence of random measurement error in simulated data and in real-world protein signalling data.



(a) Fraction of detected conditional independences given noisy measurements of conditionally dependent variables, i.e. false negatives.

(b) Fraction of detected conditional dependences given noisy measurements of conditionally independent variables, i.e. false positives.



(c) Rate of combined unknowns and false negatives for noisy measurements of conditionally dependent variables.

(d) Rate of combined unknowns and false positives for noisy measurements of conditionally independent variables.

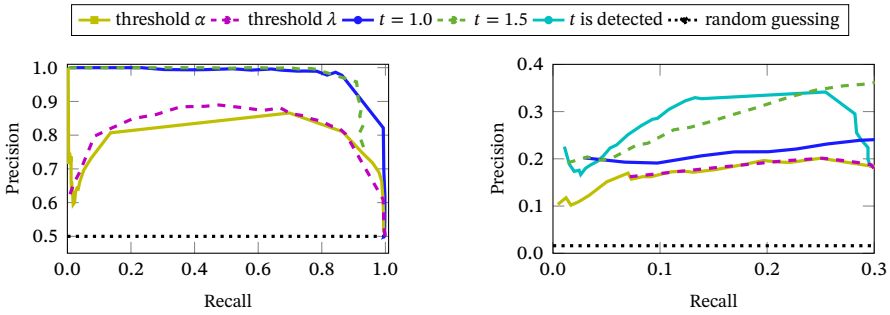
Figure 2.6: Error rates for different (adjusted) conditional independence tests applied to simulated data with added measurement noise. Figures (a) and (b) show the error rate for detecting conditional dependences and independences, respectively, in the presence of measurement error for λ-strong faithful data (with $\lambda = 0.1$). Figure (c) shows the rate of conditional dependences not detected from the noisy measurements by the adjusted partial correlation test (i.e. the false negatives and unknowns). Similarly, Figure (d) shows the rate of conditional independences that cannot be identified from noisy measurements.

2.5.1 Simulations with LCD triples

First, we generated data sets by sampling from random causal models for triples of variables $(X_{v_1}, X_{v_2}, X_{v_3})$. We made sure that half of the models had the graphical structure of an LCD triple. We generated parameter values as described in Section 2.4.3. The other half of the random causal models was obtained by selecting models from the ones that we constructed in Section 2.4.3, for which X_{v_1} can be treated as an intervention variable. To do so, we selected those models in the previous section, where X_{v_1} was caused neither by X_{v_2} nor by X_{v_3}. For each model, we generated 10000 data points and added random measurement error with a fixed variance of 0.8. We then applied the LCD algorithm to these data sets using different ways to evaluate a con-

ditional independence test based on partial correlations: using a threshold α on the p-value, using a threshold of λ on the partial correlation, and using a measurement error correction of t times the true measurement error variance and the parameter λ to evaluate adjusted partial correlations. We treat both α and λ as tuning parameters to obtain the precision recall curve in Figure 2.7(a). From this we see that LCD based on measurement error corrected partial correlations outperforms more standard testing procedures. In particular, due to the possibility of letting a conditional independence test output the value 'unknown', the method is able to achieve high precision at low recall.

In practice, an upper bound for the variance of measurement error is rarely known in advance and the data may not be λ-strong faithful. To simulate a somewhat more realistic scenario we generated 200 random DAGs with 15 vertices and a connection probability of 0.15. We made sure that there was one special intervention vertex that was not caused by any of the other vertices. We sampled 10000 data points from linear Gaussian models with these causal structures, where edge weights were drawn uniformly at random from the interval $[0.8, 1.2]$ and error variances were chosen uniformly at random from the interval $[0.5, 1.0]$. The noisy measurements of the 15 variables were obtained by adding random measurement error with a fixed variance of 0.8. From these 200 data sets we tried to obtain upper bounds for the variance of measurement error. To do so, we applied Wishart's test (Wishart, 1928) to test for vanishing tetrad constraints and then used Corollary 2.1 to estimate this upper bound using the adjusted partial correlations amongst the four variables with vanishing tetrad constraints. In 79% of case the detected upper bound was correct (i.e. it exceeded the true measurement error variance of 0.8), while we were not able to detect any upper bound in 18.5% of the cases, and in 1.5% of cases the estimated upper bound was too small. When multiple upper bounds could be detected we chose the median as an estimate of the upper bound of the measurement error variance. Using the detected upper bound we then applied adjusted partial correlations to test for conditional (in)dependences as in Section 2.4.2. If no upper bound was detected then we assigned 'unknown' to every conditional independence test. These conditional independences were then used in the LCD algorithm, where we treated the variable that was not caused by any of the other endogenous variables in the model as an intervention variable, and where marginal dependences were tested using Pearson's correlation test with an $\alpha = 0.05$ threshold on the p-value. Figure 2.7(b) shows the precision recall curves on these data sets for different methods to evaluate conditional (in)dependence from partial correlations. From this, we see that causal discovery based on one of the three adjusted partial correlation tests to test for conditional independences outperforms conditional independence tests based on a threshold on the p-value of a partial correlation test. Furthermore, the LCD algorithm performs significantly better than a random guessing baseline. Adjustment with the detected upper bound or the upper bound of 1.5 times the true measurement error variance outperforms adjustments with the true variance of the measurement



(a) LCD application.

(b) Upper bound detection and LCD application.

Figure 2.7: In (a) the precision-recall curve for detecting LCD triples from λ-strong faithful data subject to random measurement error with fixed variance and fixed upper bound is given. The precision-recall curves are shown in cases where we evaluate conditional independence by comparing the p-value of a partial correlation to a fixed threshold α, and by comparing adjusted partial correlation (based on an upper bound for measurement error of either 1.0 or 1.5 times the true measurement error variance) to a threshold λ. Both α and λ are treated as tuning parameters. Figure (b) shows the precision-recall curve for simulations of 15 variables. Here the lightblue curve shows the precision-recall for an experiment where we first apply the upper bound detection and then the measurement error corrections. The random baseline for this scenario is drawn at 0.016.

error (i.e. the curve for $t = 1.0$ lies below that of $t = 1.5$ and that of the detected upper bound). This suggests that it may be prudent to assign the outcome 'unknown' to the outcome of a partial correlation test.

2.5.2 Protein expression data

Our ideas can be applied to a real-world protein expression data-set in which the measurements of protein abundances might have been corrupted by measurement error. We used a dataset that was gathered by Lun et al. (2017) to study the influence of protein abundances on signalling cascades in a protein network in human kidney cells. Using Corollary 2.1 we were able to obtain an estimate for an upper bound of random measurement error variance from this data set. Furthermore, we applied the LCD algorithm in combination with the adjusted partial correlation test that we proposed in Section 2.4.2. In absence of a reliable ground truth for this experiment, we validated the results of a measurement error correction applied to the LCD algorithm by comparing it to a baseline derived from interventions in the data. LCD with adjusted partial correlations seemed to work well on this data set, although LCD with standard conditional independence tests already gives good results on this data.

Data description The abundance of different proteins labelled $(\text{GFP})_j$ with $j = 1, \dots, 20$ were over-expressed in 20 different experiments and their abundance was

measured using mass cytometry (Lun et al., 2017). In most experimental conditions there were around 10000 single cells in which this abundance was measured. Proteins may send signals through chemical alterations of the protein called *phosphorylation*. For each condition $j = 1, \ldots, 20$, the abundances of an additional 34 phosphorylated proteins P_i^j (with $i = 1, \ldots, 20$) were measured after stimulation of the network. We labelled the experimental conditions j in such a way that the over-expression of a protein $(\text{GFP})_j$ corresponds to measurements of the abundance of the corresponding phosphorylated protein P_j^j (if it was measured). The abundance of an over-expressed protein typically differed between cells and not every cell was affected (Lun et al., 2017). Because of the experimental design, the abundance of an over-expressed protein $(\text{GFP})_j$ is not caused by the abundance of any of the measured phosphorylated proteins.

Data pre-processing To ensure that the data is more similar to data generated by a linear Gaussian model, we follow (Lun et al., 2017) and use the following data transformation. Each raw data point x is transformed by

$$\hat{x} = \text{arcsinh}(x/5). \tag{2.15}$$

As a further preprocessing step we filtered out cells that are in the M cell cycle phase according to the gating procedure described in (Behbehani et al., 2012). This filtering step can be motivated as follows. Cells that are in the M phase form a distinct cluster and therefore strongly violate assumptions of linearity or Gaussianity, although they can easily be filtered out. Secondly, cells that are in the M phase already have doubled nuclei and other organelles so we cannot safely assume that the causal mechanisms of cell signalling behave in the same way. The removal of these cells from the data should therefore be seen as the removal of a contaminating population. Practically, this came down to selecting only those single cell measurements for which the abundance of the phosphorylated protein pHH3 was smaller than 3.0 (after application of equation (2.15)). We took into account the detection limit of mass cytometry by only including single cell measurements for which all relevant protein abundances exceeded a lower threshold of 0.5 (after application of the data transformation).

For the purpose of validation, we only considered proteins that were over-expressed in one of the experiments and for which the abundance of the corresponding phosphorylated protein was measured as well. Furthermore, because we assume strong faithfulness, we only considered the measurements that were obtained 5 minutes after stimulation, because at this time the signalling responses were generally strong according to Figure 3 in Lun et al. (2017). We used data from the first replica of the experiment which had the most measurements under each condition. We analysed a subset of the available proteins, based on the recommendations in Lun et al. (2017), and excluded proteins from the cause variables when spill-over effects were reported under the condition that they were over-expressed, see also Table

2.1. We further excluded the S6 protein because over-expressing it induced no strong signalling responses, and we also discarded the SHP2 protein because the measurements of the abundance of the phosphorylated protein pSHP2 was affected by spill-over effects in multiple experiments where other proteins were over-expressed.

Table 2.1: Proteins that are both over-expressed in one of the conditions and whose phosphorylated abundance is measured under all conditions, with an indication whether spillover effects are present.

Over-expressed protein	Phosphorylated protein	Spill-over effect present?
JNK1	pJNK	no
MKK6	pMKK3/6	no
PDPK1	pPDPK1	yes
P38	pP38	no
AKT1	pAKT	no
ERK2	pERK	no
SHP2	pSHP2	no
GSK3B	pGSK3B	yes
S6	pS6	no
P90RSK	pP90RSK	yes
MEK1	pMEK1/2	no
P70S6K	pS6K	no

Upper bound detection We used the available data to detect an upper bound for the variance of measurement error, using the Corollary 2.1. To that end, we considered all measurements under the condition that the SRC protein was over-expressed and which resulted in the presence of strong signalling relations, see also Lun et al. (2017). We looped over all triples (P_i, P_j, P_k) and selected those that were marginally dependent amongst each other and also on $(\text{GFP})_{\text{SRC}}$ according to a t-test on the 1%-level. We then tested whether all three tetrads vanished using a Wishart test at the 5% level. We found that these constraints were satisfied for the triple (pS6K, pMAPKAPK2, pMAP2K3). The upper bounds for the variance of measurement error that we found were 0.10 for the adjustment on pS6K, 0.15 for an adjustment on pMAPKAPK2, and 0.14 for adjustments on pMAP2K3. Other triples that satisfied the constraints gave similar or (much) higher upper bounds for the measurement error. Since all proteins were measured with the same device, we assumed that the variance of the measurement error is the same for each variable, so that 0.14 was a suitable upper bound for the measurement error on any variable. Although the detected upper bound was large for weak signals, the proteins with stronger signals typically had variances > 1, so that the relative amount of measurement error for proteins with strong signalling relations amounted to less than 10%.

Validation To validate the results of LCD, we created a 'ground truth' from the interventions (corresponding to over-expression of certain proteins) in the dataset. This baseline starts from the assumption that the abundance $(\text{GFP})_j$ of the over-expressed protein j is a direct cause of the abundance P_j^* of phosphorylated protein j. The reasoning behind this assumption is that a higher abundance of a protein facilitates phosphorylation of that protein, while the mechanism of phosphorylation would not directly lead to an increase in the abundance of the protein. Lun et al. (2017) show that over-expression of a protein P_j^* does not alter the network structure. Therefore, under the additional assumption that $(\text{GFP})_j$ does not directly cause any of the other proteins P_i^* (with $i \neq j$), we have that P_j^* is a cause of P_k^*, whenever $(\text{GFP})_j$ and P_k^* are dependent.

We constructed the 'ground truth' for cause-effect pairs (P_j^*, P_k^*). We considered the 7 phosphorylated proteins P_j^* that were over-expressed in one of the conditions as potential causes and effects, while we did not consider the effects of any on the other measured proteins. We considered a pair (P_j^*, P_k^*) a causal pair if a t-test indicated that $(\text{GFP})_j$ and P_k^* were dependent at a conservative level of 10^{-4}. Out of 231 possible cause-effect pairs, we identified 164 causal relations using this method, corresponding to a 71% rate of identified cause-effect pairs.

Methods and results We used the detected upper bound for random measurement to adjust partial correlations used in the independence test for the LCD algorithm. To apply LCD, we treated $(\text{GFP})_i$ as an intervention variable for conditions $i \in \{1, \ldots, 20\}$ and looked for LCD triples $((\text{GFP})_i, P_j^*, P_k^*)$ with $i \neq j$ and $i \neq k$. To make our our predictions of LCD triples more conservative and more robust we only identified the triple if we were able to detect it under a minimum of 2 conditions.[8] We applied three methods of (in)dependence testing in combination with the LCD algorithm: a threshold α on the p-value of t-tests, a threshold λ on the absolute value of partial correlations, and a threshold λ on partial correlations with a measurement error correction using the upper bound $u = 0.14$. We treated λ and α as tuning parameters and compared the output to the 'ground truth' that we obtained for validation. Figure 2.8 shows that precision-recall curves for each method of conditional (in)dependence testing significantly outperform a baseline of random guessing at lower recall. These curves appear somewhat unusual, because the recall initially increases when the threshold for dependence increases (i.e. we find more LCD triples) but starts to decrease as the threshold is raised further because we are then no longer able to identify the marginal dependencies of LCD triples. Figure 2.8 shows that all methods are able to significantly outperform a random baseline, although the adjustment for measurement error does not significantly improve upon the results that can be obtained with LCD using the usual (partial) correlation testing procedures.

[8] Since central proteins in the network were over-expressed, true causal pairs were expected to appear under multiple conditions.

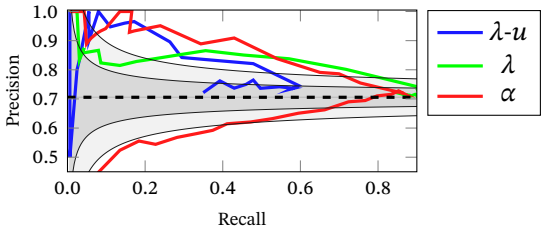


Figure 2.8: LCD applied to protein signaling data with α or λ as tuning parameter and a measurement-error correction. The results are compared with the random baseline, the gray-shaded areas represent one and two standard deviations from the random baseline. The precision-recall curve circles back for high values of the tuning parameter, because the same threshold is used to test for the dependences and conditional independences that are required for the detection of an LCD triple.

2.6 Conclusion

In this chapter we have demonstrated that measurement error, when not taken into account, may fool causal discovery methods into wrongfully inserting, deleting or reversing edges in the predicted causal graph. We showed that regular statistical tests with conventional thresholds would fail to detect conditional independences between the uncorrupted variables from the data when measurement error is present. The key result that we presented in this work is that, under certain conditions, we can find an upper bound for the variance of random measurement error from noisy measurements. We show how to propagate this uncertainty to adjust partial correlations for the presence of measurement error, so that under the strong faithfulness condition the test indicates a dependence for large correlations, an independence for small correlations, and 'unknown' for correlations that cannot be decided either way. We demonstrated in simulations that this may improve the accuracy of the LCD algorithm. We also applied our ideas to a real-world protein signalling dataset, and we found an upper bound for the variance of the measurement error in this dataset. Although LCD with a measurement error correction yielded significant results compared to a random baseline, it did not outperform conventional methods. Nevertheless, it is our belief that taking measurement error into account is a promising step towards successful real-world applications of (constraint-based) causal discovery.

Chapter 3

Causal constraints models

Adaptation based on:
Beyond structural causal models: causal constraints models,
T. Blom, S. Bongers, and J.M. Mooij,
Proceedings of the 35th Annual Conference on Uncertainty in Artificial Intelligence (UAI-19).

Structural Causal Models (SCMs) provide a popular causal modelling framework. In this work, we show that SCMs are not flexible enough to give a complete causal representation of dynamical systems at equilibrium. Instead, we propose a generalization of the notion of an SCM, that we call Causal Constraints Models (CCMs), and prove that they do capture the causal semantics of such systems. We show how CCMs can be constructed from differential equations and initial conditions. We then illustrate our ideas further on a simple but ubiquitous (bio)chemical reaction and demonstrate that our framework also allows to model functional laws, such as the ideal gas law, in a sensible and intuitive way.

3.1 Introduction

Real-world processes are often complex and time-evolving. The dynamics of such systems can be modelled by (random) differential equations, which offer a fine-grained description of how the variables in the system change over time. A coarser but more tractable approach is to model the system with a Structural Causal Model (SCM), a modelling class that provides a framework that is used in many fields such as biology, the social sciences, and economy (Pearl, 2009). Although SCMs have been successfully applied to certain static systems, a pressing concern is whether SCMs are able to completely model the causal semantics of the stationary behaviour of a dynamical system. In this chapter, we prove that generally SCMs are not flexible enough to completely model the equilibrium distribution of dynamical systems

under certain interventions.

We generalize the notion of SCMs and introduce a novel type of causal model, that we call Causal Constraints Models (CCMs). We prove that they give a complete description of the causal semantics of dynamical systems at equilibrium and show how a CCM can be derived from differential equations and initial conditions. We further motivate our approach by pointing out that CCMs, contrary to SCMs, fully capture the causal semantics of functional laws (e.g. the ideal gas law), which describe relations between variables that are invariant under all interventions. We illustrate the benefits of CCMs on a simple but ubiquitous (bio)chemical reaction.

Causal models that arise from studying the behaviour of dynamical systems have received much attention over the years. F. M. Fisher (1970), Mogensen et al. (2018), Rubenstein et al. (2018), Sokol et al. (2014), and Voortman et al. (2010) consider causal relations in systems that can be modelled by (stochastic) differential equations that have not (yet) reached equilibrium. In contrast, we consider the stationary behaviour of dynamical systems, which does not require us to model the system's dependence on time. Bongers and Mooij (2018), Hyttinen et al. (2012), Lacerda et al. (2008), Mooij, Janzing, Heskes, et al. (2011), and Mooij, Janzing, and Schölkopf (2013) show how cyclic SCMs may arise from studying the stationary behaviour of certain dynamical time-series or differential equations, and how in some cases cyclic SCMs can be learned from equilibrium data. SCMs are well-understood and have recently been extended to also include the cyclic case (Bongers, Forré, et al., 2020; Forré et al., 2017). The drawback of the extension in Forré et al. (2017), with respect to modelling equilibria of dynamical systems, is that it requires the model to have a globally compatible solution under any intervention, which dynamical systems do not, in general, possess. Another modelling approach for dynamical systems at equilibrium is to construct a, possibly cyclic, SCM from the differential equations as Mooij, Janzing, and Schölkopf (2013) and Bongers and Mooij (2018) do. In this work, we show that these modelling approaches for the causal semantics of the stationary behaviour in dynamical systems cannot accommodate the dependence of equilibria on initial conditions of the system.

In previous work, researchers have come across subtleties regarding the relation between the causal semantics and conditional independence properties of dynamical systems at equilibrium (Dash, 2005; Iwasaki et al., 1994; Lacerda et al., 2008). Previously, researchers have made additional assumptions about the underlying dynamical system to circumvent these. Although Rubenstein* et al. (2017) and Bongers and Mooij (2018) do not make such restrictions, the price that one pays is that either one must limit the interventions that can be modelled or the equilibrium is no longer uniquely specified and one is limited to modelling the fixed points of the system. To the best of our knowledge, Causal Constraints Models are the first models that can completely capture the causal semantics of the stationary behaviour of dynamical

systems in general.[1]

3.1.1 Structural causal models

A statistical model over random variables, taking value in a measurable space $\mathfrak{X}$, usually is a pair $(\mathfrak{X}, \mathbb{P}^{\mathfrak{X}})$ where $\mathbb{P}^{\mathfrak{X}}$ is a (parametrized) family of probability distributions on $\mathfrak{X}$. A causal model on the other hand, can be thought of as a family of statistical models, one for each (perfect) intervention,

$$\bar{\mathbb{P}}^{\mathfrak{X}} = \left(\mathbb{P}^{\mathfrak{X}}_{\mathrm{do}(I,\xi_I)} : \quad I \in \mathcal{P}(\mathcal{I}), \quad \xi_I \in \mathfrak{X}_I\right), \tag{3.1}$$

where $\mathcal{I}$ is an index set and $\mathcal{P}(\mathcal{I})$ denotes the power set of $\mathcal{I}$ (i.e. the set of all subsets of $\mathcal{I}$). I represents the intervention target and ξ_I a tuple of intervention values. The null intervention $\mathrm{do}(\emptyset)$ for $I = \emptyset$ corresponds to the observed system.

SCMs are a special type of causal models that are specified by structural equations. Our formal treatment of SCMs mostly follows Bongers, Forré, et al. (2020) and Pearl (2009). We deviate from the usual definition of SCMs by not requiring acyclicity (i.e. recursiveness).[2]

Definition 3.1. Let $\mathcal{I}$ and $\mathcal{J}$ be index sets. A *Structural Causal Model (SCM)* $\mathcal{M}$ is a triple $(\mathfrak{X}, F, \mathbf{E})$, with:

- a product of standard measurable spaces $\mathfrak{X} = \prod_{i\in\mathcal{I}} \mathcal{X}_i$ representing the domains of endogenous variables,
- a tuple of independent exogenous random variables $\mathbf{E} = (E_j)_{j\in\mathcal{J}}$ taking value in a product of standard measurable spaces $\mathfrak{E} = \prod_{j\in\mathcal{J}} \mathcal{E}_j$,
- a family F of measurable functions:[3]

$$f_i : \mathfrak{X}_{\mathrm{pa}(i)\cap\mathcal{I}} \times \mathfrak{E}_{\mathrm{pa}(i)\cap\mathcal{J}} \to \mathcal{X}_i, \quad \forall i \in \mathcal{I}.$$

Note that a cyclic structural causal model does not need to imply a unique joint distribution $\mathbb{P}^{\mathfrak{X}}_{\mathrm{do}(\emptyset)}$ on the space of endogenous variables in the observed system, although acyclic and simple SCMs do (Bongers, Forré, et al., 2020). When there exists a unique solution $\mathbf{x} \in \mathfrak{X}$ to the *structural equations*

$$x_i = f_i(\mathbf{x}_{\mathrm{pa}(i)\cap\mathcal{I}}, \mathbf{e}_{\mathrm{pa}(i)\cap\mathcal{J}}), \quad \forall i \in \mathcal{I}$$

for almost all $\mathbf{e} \in \mathfrak{E}$, we say that the model is uniquely solvable.

[1] It has been shown that simple SCMs have a graphical representation with an intuitive causal interpretation (Bongers, Forré, et al., 2020). We will discuss graphical representations of the causal relations and independence structure for CCMs in Chapter 4.

[2] In the original version of this article we did not assume independence of exogenous variables because it is not needed within this chapter. For the sake of consistency with the rest of the thesis, the independence of exogenous variables is included in the definition.

[3] Here $\mathrm{pa}(i) \subseteq \mathcal{I} \cup \mathcal{J}$ denotes a subset of indices so that the values of $(\mathbf{X}_{\mathrm{pa}(i)}, \mathbf{E}_{\mathrm{pa}(i)})$ are sufficient to determine the values of f_i.

Definition 3.2. We say that a random variable $\mathbf{X} = (X_i)_{i\in\mathcal{I}}$ is a *solution* to an SCM $\mathcal{M} = (\mathfrak{X}, F, \mathbf{E})$ if

$$X_i = f_i(\mathbf{X}_{\mathrm{pa}(i)\cap\mathcal{I}}, \mathbf{E}_{\mathrm{pa}(i)\cap\mathcal{J}}) \quad \text{a.s.}, \quad \forall i \in \mathcal{I}.$$

An SCM may have a unique (up to zero sets) solution, multiple solutions, or there may not exist any solution at all. There are many types of interventions, corresponding to different experimental procedures, that can be modelled in an SCM. For the remainder of this chapter, we only consider *perfect* (also known as "surgical" or "atomic") interventions that force variables to take on a specific value through some external force acting on the system.

Definition 3.3. A *perfect intervention* $\mathrm{do}(I, \boldsymbol{\xi}_I)$ with target $I \subseteq \mathcal{I}$ and value $\boldsymbol{\xi}_I \in \mathfrak{X}_I$ on an SCM $\mathcal{M} = (\mathfrak{X}, F, \mathbf{E})$ maps it to the *intervened* SCM $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)} = (\mathfrak{X}, \widetilde{F}, \mathbf{E})$ with $\widetilde{F}$ the family of measurable functions:

$$\tilde{f}_i(\mathbf{x}_{\mathrm{pa}(i)\cap\mathcal{I}}, \mathbf{e}_{\mathrm{pa}(i)\cap\mathcal{J}}) = \begin{cases} \xi_i & i \in I, \\ f_i(\mathbf{x}_{\mathrm{pa}(i)\cap\mathcal{I}}, \mathbf{e}_{\mathrm{pa}(i)\cap\mathcal{J}}) & i \in \mathcal{I}\backslash I. \end{cases}$$

Note that the solvability of an SCM may change after a perfect intervention, e.g. a uniquely solvable SCM may no longer be so after certain interventions.

3.1.2 Dynamical systems

We consider dynamical systems $\mathcal{D}$ describing $p = |\mathcal{I}|$ (random) variables $\mathbf{X}(t)$ taking value in $\mathfrak{X} = \mathbb{R}^p$. They consist of a set of coupled first-order ordinary differential equations (ODEs) where the initial conditions $\mathbf{X}(0)$ are determined by exogenous random variables $\mathbf{E} = (E_i)_{i\in\mathcal{J}}$ taking value in $\mathfrak{E} = \mathbb{R}^p$. That is,

$$\begin{aligned} \dot{X}_i(t) &= f_i(\mathbf{X}(t)), && \forall i \in \mathcal{I}, \\ X_i(0) &= E_i, && \forall i \in \mathcal{I}, \end{aligned}$$

where we only consider systems for which each f_i is a locally Lipschitz continuous function.[4] Throughout this chapter, we will assume for any dynamical system we encounter that for $\mathbb{P}^{\mathbf{E}}$-almost every $\mathbf{e} \in \mathbb{R}^p$ the initial value problem with $\mathbf{X}(0) = \mathbf{e}$ has a unique solution $\mathbf{X}(t, \mathbf{e})$ for all $t \geq 0$, given by

$$\mathbf{X}(t, \mathbf{e}) = \mathbf{X}(0, \mathbf{e}) + \int_0^t \mathbf{f}\big(\mathbf{X}(s, \mathbf{e})\big) ds. \tag{3.2}$$

This solution $\mathbf{X}(t, \mathbf{e})$ can be trivially extended to $\mathfrak{E}$ and it is measurable in $\mathbf{e}$ for all t (Han et al., 2017).

[4] If the dynamics depends on (random) parameters, they can be modelled as additional endogenous variables with vanishing time derivatives and initial conditions corresponding to the (random) parameters. Therefore, without loss of generality, we may assume that the functions f_i only depend on $\mathbf{X}$.

A *fixed point* (or equilibrium point) of a dynamical system $\mathcal{D}$ is a point $\mathbf{x}^* \in \mathbb{R}^p$ for which $\mathbf{f}(\mathbf{x}^*) = \mathbf{0}$. For $\mathbf{e} \in \mathbb{R}^p$, the dynamical system converges to an equilibrium $\mathbf{X}^*(\mathbf{e}) \in \mathbb{R}^p$ if

$$\lim_{t\to\infty} \mathbf{X}(t, \mathbf{e}) = \mathbf{X}^*(\mathbf{e}). \tag{3.3}$$

If for $\mathbb{P}^{\mathbf{E}}$-almost every $\mathbf{e}$ the limit in equation (3.3) exists, then we say that $\mathcal{D}$ converges to the equilibrium solution $\mathbf{X}^* = \lim_{t\to\infty} \mathbf{X}(t, \mathbf{E})$.

Interventions on dynamical systems can be modelled in different ways. One could for example fix the value of targeted values at one time-point. Alternatively, one could fix the trajectory of the targeted values as in Rubenstein et al. (2018). Here, we follow Mooij, Janzing, and Schölkopf (2013) and define interventions as operations that fix the value of the targeted variables to a constant (for all time).

Definition 3.4. A *perfect intervention* $\mathrm{do}(I, \boldsymbol{\xi}_I)$ where $I \subseteq \mathcal{I}$ and $\boldsymbol{\xi}_I \in \boldsymbol{\mathcal{X}}_I$ results in the *intervened* dynamical system $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ specified by

$$\begin{aligned} \dot{X}_i(t) &= 0, & X_i(0) &= \xi_i, & \forall i &\in I, \\ \dot{X}_i(t) &= f_i(\mathbf{X}(t)), & X_i(0) &= E_i, & \forall i &\in \mathcal{I}\backslash I, \end{aligned}$$

and the independent exogenous random variables $\mathbf{E}$ are unaffected by the intervention.

We say that a causal model $\mathcal{M}$ *completely* captures the causal semantics of the stationary behaviour of a dynamical system $\mathcal{D}$ if for all $I \subseteq \mathcal{I}$ and all $\boldsymbol{\xi}_I \in \boldsymbol{\mathcal{X}}_I$: the equilibrium solutions of $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ coincide with the solutions of $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ (up to $\mathbb{P}^{\mathbf{E}}$-null sets).

The construction of SCMs from dynamical systems in Mooij, Janzing, and Schölkopf (2013) relies on the fact that for systems that converge to a fixed point independent of initial conditions (i.e. globally asymptotically stable systems), the fixed point directly gives a complete description of its stationary behaviour. A much weaker stability assumption is that of *(global) semi-stability* (Bhat et al., 1999; Campbell et al., 1979), where solutions of a system converge to a stable equilibrium determined by initial conditions. The following definition is adapted from Haddad et al. (2010).

Definition 3.5. Let $\mathcal{D}$ be a dynamical system and $\boldsymbol{\mathcal{U}} \subseteq \mathbb{R}^p$ an invariant subset (i.e. if $\mathbf{x}(0) \in \boldsymbol{\mathcal{U}}$ then $\mathbf{x}(t) \in \boldsymbol{\mathcal{U}}$ for all $t \geq 0$). Then

(i) A fixed point $\mathbf{x}^* \in \boldsymbol{\mathcal{U}}$ is *Lyapunov stable* with respect to $\boldsymbol{\mathcal{U}}$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $\mathbf{x}(0) \in \boldsymbol{\mathcal{U}}$: if $\|\mathbf{x}(0) - \mathbf{x}^*\| < \delta$ then for all $t \geq 0$, $\|\mathbf{x}(t) - \mathbf{x}^*\| < \epsilon$.

(ii) A fixed point $\mathbf{x}^* \in \boldsymbol{\mathcal{U}}$ is *semistable* w.r.t. $\boldsymbol{\mathcal{U}}$ if it is Lyapunov stable and, additionally, there exists a relatively open subset[5] $\boldsymbol{\mathcal{N}}$ of $\boldsymbol{\mathcal{U}}$ that contains $\mathbf{x}^*$ such that $\mathbf{x}(t)$ converges to a Lyapunov stable fixed point for all $\mathbf{x}(0) \in \boldsymbol{\mathcal{N}}$.

[5] $\boldsymbol{\mathcal{N}}$ is a relatively open subset of $\boldsymbol{\mathcal{U}}$ if there is an open set $\boldsymbol{\mathcal{N}}' \subseteq \mathbb{R}^p$ such that $\boldsymbol{\mathcal{N}} = \boldsymbol{\mathcal{N}}' \cap \boldsymbol{\mathcal{U}}$.

(iii) A semistable fixed point $\mathbf{x}^* \in \boldsymbol{\mathcal{U}}$ is globally semistable with respect to $\boldsymbol{\mathcal{U}}$ if $\boldsymbol{\mathcal{N}} = \boldsymbol{\mathcal{U}}$.

Finally, we say that a dynamical system $\mathcal{D}$ is globally semistable with respect to $\boldsymbol{\mathcal{U}}$ if all its fixed points are globally semistable with respect to $\boldsymbol{\mathcal{U}}$.

In other words: $\mathcal{D}$ is globally semistable with respect to an invariant subset $\boldsymbol{\mathcal{U}}$ if all its fixed points are Lyapunov stable, lie in $\boldsymbol{\mathcal{U}}$, and for all initial conditions in $\boldsymbol{\mathcal{U}}$, the system converges to a fixed point. In Definition 3.6 we also take stability of intervened systems into account.

Definition 3.6. A dynamical system $\mathcal{D}$ is *structurally semistable* if for all $I \subseteq \mathcal{J}$ there exists $\boldsymbol{\mathcal{U}} \subseteq \mathbb{R}^p$ with $\mathbb{P}^{\mathbf{E}_{\mathcal{J}\setminus I}}(\boldsymbol{\mathcal{U}}_{\mathcal{J}\setminus I}) = 1$ such that: $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ is globally semistable w.r.t. $\boldsymbol{\mathcal{U}}$ (for any $\boldsymbol{\xi}_I \in \boldsymbol{\mathcal{X}}_I$).

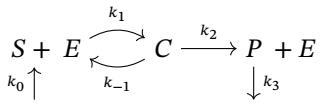
Whether a dynamical system converges to a certain fixed point depends on initial conditions. This dependence can often be described by constants of motion, and there exists a vast literature on how and when these can be derived from differential equations. The notion of semi-stability is appropriate in many real-world applications in chemical kinetics, environmental, and economic systems (Haddad et al., 2010). For chemical reaction networks, there exist convenient criteria on the network structure that guarantee global semi-stability (Chellaboina et al., 2009), and for mechanical systems semi-stability characterizes the motion of rigid bodies subject to damping (Bhat et al., 1999).

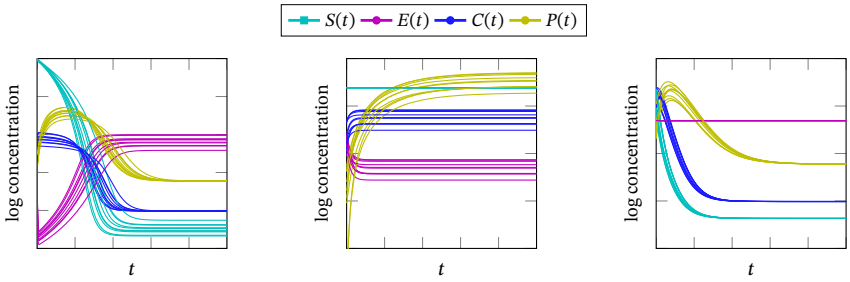
3.2 Dynamical systems as SCMs

In this section, we consider SCM representations of the equilibria in a chemical reaction and conclude that, generally, SCMs are not flexible enough to completely capture the causal semantics of stationary behaviour in dynamical systems.

3.2.1 Basic enzyme reaction

The basic enzyme reaction is a well-known example of a biochemical reaction network. It describes a system where a substrate S reacts with an enzyme E to form a complex C which is then converted into a product P and the enzyme (Murray, 2002). In the open enzyme reaction a constant influx of substrate and an efflux of product are added (Belgacem et al., 2012). The process can be presented by a reaction graph with strictly positive rate parameters $\mathbf{k} = [k_0, k_{-1}, k_1, k_2, k_3]$:





(a) No intervention. (b) Intervention on S. (c) Intervention on E.

Figure 3.1: Temporal dependence of concentrations in the basic enzyme reaction in equations (3.4) to (3.7) with random initial conditions and $\mathbf{k} = [0.4, 0.3, 1.0, 1.1, 0.5]$. Other choices for the rate parameters give qualitatively similar results. In Figure (a) we see that both S and P converge to an equilibrium that depends on initial conditions in the observed system. Figure (b) shows that C, E, and P converge to an equilibrium that depends on the initial conditions after an intervention on S. Finally, from Figure (c) we conclude that S, C, and P converge to an equilibrium that is independent of the initial conditions after an intervention on E.

Differential equations that describe the concentrations of each molecule in the system over time can be obtained by application of the law of mass-action, which states that the rate of a reaction is proportional to the product of the concentrations of the reactants (Murray, 2002), yielding:

$$\dot{S}(t) = k_0 - k_1 S(t)E(t) + k_{-1}C(t), \tag{3.4}$$
$$\dot{E}(t) = -k_1 S(t)E(t) + (k_{-1} + k_2)C(t), \tag{3.5}$$
$$\dot{C}(t) = k_1 S(t)E(t) - (k_{-1} + k_2)C(t), \tag{3.6}$$
$$\dot{P}(t) = k_2 C(t) - k_3 P(t), \tag{3.7}$$
$$(S(0), E(0), C(0), P(0)) = (s_0, e_0, c_0, p_0). \tag{3.8}$$

We simulated the system in equations (3.4) to (3.7) with random initial conditions and also under interventions on S and E. Figure 3.1 shows how and whether the time trajectories of concentrations depend on initial conditions in different interventional settings.

3.2.2 Equilibrium solutions

By explicit calculation we can verify that given strictly positive initial conditions, the dynamical system converges to an equilibrium (S^*, C^*, E^*, P^*) if it exists, under any perfect intervention (Belgacem et al., 2012). In the supplementary material of this chapter (see Section 3.A) we give the details of this calculation and show that the system is structurally semistable. The equilibria can be found by deriving constraints

on solutions of the system:

(i) At equilibrium the system is at rest and therefore all time derivatives vanish. From the equations of motion associated with the variables in the system we can derive constraints that are invariant under all interventions that do not target the variable associated with the constraint. For example, the equations of motion in (3.4) yields the equilibrium equation

$$k_0 - k_1 S^* E^* + k_{-1} C^* = 0,$$

which constrains the equilibrium state unless S is targeted by an intervention.

(ii) Symmetries or (linear) dependences between the time derivatives lead to conservation laws (i.e. constants of motion), which are relations between variables that are time-invariant but that are typically invariant under fewer interventions than constraints of the first type. For example, since $\dot{C}(t) + \dot{E}(t) = 0$ for all t, we have that

$$C(t) + E(t) = c_0 + e_0, \qquad \forall\, t, \tag{3.9}$$

unless C, E or both C and E are targeted by an intervention.

(iii) A system may contain (derived) variables whose time-derivative does not depend on itself. Since

$$\dot{S}(t) - \dot{E}(t) = k_0 - k_2 C(t), \tag{3.10}$$

the variable C cannot be 'freely manipulated', in the sense that $S(t) - E(t)$ does not converge to equilibrium under interventions $\mathrm{do}(C = \xi_C)$ when $\xi_C \neq {}^{k_0}/_{k_2}$. For $\xi_C = {}^{k_0}/_{k_2}$ a new constant of motion is introduced so that $S(t) - E(t) = s_0 - e_0$ unless S, E or both S and E are targeted by an intervention.

It can be shown, through explicit calculations, that for any perfect intervention these constraints have no solution when the dynamical system does not converge to an equilibrium and they have a unique solution when the system does converge to an equilibrium. A complete description of the equilibrium solutions of this system under perfect interventions (if it exists) is given in Table 3.2 of the supplementary material in Section 3.A. A subset of the results is presented in Table 3.1 which, together with Figure 3.1 in the previous section, illustrates the rich causal semantics of this system at equilibrium (e.g. an intervention on S makes C^* dependent on the initial conditions, while an intervention on E makes S^* independent of the initial conditions).

3.2.3 SCM representation

Globally asymptotically stable dynamical systems converge to a unique fixed point and for these systems Mooij, Janzing, and Schölkopf (2013) have shown that SCMs can be constructed from ordinary differential equations by equilibration. For the basic enzyme reaction (which is not globally asymptotically stable) their construction

Table 3.1: Equilibrium solutions S^*, C^*, and E^* for the basic enzyme reaction dynamics in (3.4) to (3.7) under various interventions. We let $y = \frac{1}{2}\sqrt{(e_0 - s_0)^2 + 4\frac{k_0(k_{-1}+k_2)}{k_1 k_2}}$.

Intervention	S^*	C^*	E^*
$\mathrm{do}(\emptyset)$	$\frac{k_0+k_{-1}\frac{k_0}{k_2}}{k_1(e_0+c_0-\frac{k_0}{k_2})}$	$\frac{k_0}{k_2}$	$e_0 + c_0 - \frac{k_0}{k_2}$
$\mathrm{do}(S = \xi_S)$	ξ_s	$\frac{k_1\xi_s(e_0+c_0)}{k_{-1}+k_2+k_1\xi_s}$	$\frac{(k_{-1}+k_2)(e_0+c_0)}{k_{-1}+k_2+k_1\xi_s}$
$\mathrm{do}(C = \frac{k_0}{k_2})$	$\frac{(e_0-s_0)}{2} + y$	$\frac{k_0}{k_2}$	$\frac{-(e_0-s_0)}{2} + y$
$\mathrm{do}(E = \xi_E)$	$\frac{k_0+k_{-1}\frac{k_0}{k_2}}{k_1\xi_e}$	$\frac{k_0}{k_2}$	ξ_e

method would yield the structural equations:

$$S^* = \frac{k_0+k_{-1}C^*}{k_1E^*}, \tag{3.11}$$

$$E^* = \frac{(k_{-1}+k_2)C^*}{k_1S^*}, \tag{3.12}$$

$$C^* = \frac{k_1S^*E^*}{k_{-1}+k_2}, \tag{3.13}$$

$$P^* = \frac{k_2}{k_3}C^*. \tag{3.14}$$

While this SCM represents the causal semantics of the system's fixed points, it would be underspecified as an SCM for the stationary behaviour of the basic enzyme reaction. Indeed, this SCM has multiple solutions, corresponding to different possible initial conditions of the dynamical system and it does not contain any information on which of its solutions is realized. Theorem 3.1 shows that a complete SCM representation of the stationary behaviour in the basic enzyme reaction does not exist.

Theorem 3.1. *The causal semantics of the stationary behaviour of the basic enzyme reaction, and its dependence on initial states, cannot be completely represented by an SCM with endogenous variables S^*, E^*, C^*, P^*.*

Proof. In Section 3.A.2 of the supplementary material we show that the system converges to an equilibrium under the intervention $\mathrm{do}(E^* = e, C^* = c, P^* = p)$. Setting $\dot{S} = 0$ in (3.4) and then solving for S, we find that $S^* = (k_0+k_{-1}c)/(k_1e)$ and therefore any SCM that models the effect of this intervention correctly must have a structural equation that is equivalent to equation (3.11). Analogously, considering the remaining three interventions on three out of four variables, we find that an SCM that correctly models the effects of those interventions must have structural equations for E^*, C^* and P^* that are equivalent to the structural equations (3.12) to (3.14), respectively. Table 3.1 shows that the system converges to an equilibrium that depends on the initial conditions c_0 and e_0 under the null intervention. This equilibrium is a solution

of the structural equations in (3.11) to (3.14). However, these structural equations do not depend on initial conditions and admit other solutions as well. Therefore they do not completely represent the stationary behaviour of the system. □

3.3 Causal constraints models

In this section, we first introduce Causal Constraints Models (CCMs) and then we prove that they completely capture the causal semantics of the stationary behaviour of dynamical systems that can be described in terms of continuous first-order differential equations and initial conditions.

Causal constraints models can be seen as a generalization of SCMs, which are specified in terms of structural equations that constrain the solutions of the SCM unless the corresponding variable(s) are targeted by interventions. Analogously, CCMs are specified in terms of *causal constraints*: relations between variables that constrain the solutions of the model under a set of explicitly defined intervention targets.

Definition 3.7. Let $\mathcal{I}$, $\mathcal{J}$ and $\mathcal{K}$ be index sets. A *Causal Constraints Model (CCM)* is a triple $(\mathfrak{X}, \Phi, \mathbf{E})$, with:

(i) $\mathfrak{X}$ representing the domain of endogenous variables and $\mathbf{E}$ a tuple of exogenous random variables as in Definition 3.1,

(ii) a set $\Phi = \{\phi_k : k \in \mathcal{K}\}$ of causal constraints, each of which is a triple $\phi_k = (f_k, c_k, A_k)$ where,

- $f_k : \mathfrak{X}_{\mathrm{pa}(k)\cap\mathcal{I}} \times \mathfrak{E}_{\mathrm{pa}(k)\cap\mathcal{J}} \to \mathcal{Y}_k$ is a measurable function, $\mathcal{Y}_k$ a standard measurable space and $\mathrm{pa}(k) \subseteq \mathcal{I} \cup \mathcal{J}$,
- $c_k \in \mathcal{Y}_k$ is a constant,
- $A_k \subseteq \mathcal{P}(\mathcal{I})$ specifies the set of intervention targets under which ϕ_k is *active*.

Example 3.1. Consider the price, supply, and demand of a certain product, denoted by P, S, and D respectively, related by the following causal constraint:

$$(f, c, A) = (S - D,\ 0,\ \{\emptyset, \{D\}, \{S\}, \{D, S\}\})\,. \tag{3.15}$$

The constraint $S - D = 0$ is active in the observational setting because $\emptyset$ is in the activation set $A = \{\emptyset, \{D\}, \{S\}, \{D, S\}\}$. It is also active when either D, S or both D and S are targeted by an intervention. The constraint becomes inactive after an intervention on P. In other words, supply equals demand unless the price of the product is intervened upon (e.g. through price fixing). △

3.3.1 CCM solutions

The solution of a CCM is defined in complete analogy with that of SCM solutions. Similar to SCMs, the solution of a CCM may not always exist or there may exist multiple solutions with different distributions.

Definition 3.8. Let $\mathcal{M} = (\mathfrak{X}, \Phi, \mathbf{E})$ be a CCM and let $\Phi^{\emptyset} := \{\phi_k = (f_k, c_k, A_k) \in \Phi : \emptyset \in A_k\}$. A random variable $\mathbf{X}$ taking value in $\mathfrak{X}$ is a *solution* of $\mathcal{M}$ if

$$f_k(\mathbf{X}_{\mathrm{pa}(k)\cap\mathcal{I}}, \mathbf{E}_{\mathrm{pa}(k)\cap\mathcal{J}}) = c_k \text{ a.s., } \forall (f_k, c_k, A_k) \in \Phi^{\emptyset}.$$

3.3.2 CCM interventions

The notion of an intervention targeting the causal constraints in a CCM is a generalization of interventions that act on the structural equations of an SCM. Roughly speaking, the activation sets of the causal constraints in the model are updated and additional causal constraints describe the intervention.

Definition 3.9. Let $\mathcal{M} = (\mathfrak{X}, \Phi, \mathbf{E})$ be a CCM and let $I \subseteq \mathcal{I}$ be the intervention target and $\boldsymbol{\xi}_I \in \mathfrak{X}_I$ the target value. The *intervened CCM* is given by $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)} = (\mathfrak{X}, \widetilde{\Phi}, \mathbf{E})$ where:

- for each $i \in I$ we add a causal constraint describing the intervened value of the targets, $(x_i, \xi_i, \mathcal{P}(\mathcal{I}\backslash\{i\})) \in \widetilde{\Phi}$,
- for each causal constraint $(f, c, A) \in \Phi$ we get a modified causal constraint $(f, c, A_{\mathrm{do}(I)}) \in \widetilde{\Phi}$ if $A_{\mathrm{do}(I)} \neq \emptyset$, where

$$A_{\mathrm{do}(I)} = \{A_i \setminus J : A_i \in A, \quad J \subseteq I \subseteq A_i\}.$$

Definition 3.9 says that for any $A_i \in A$, and for any combination of two subsequent interventions such that $I_1 \cup I_2 = A_i$, the constraint will be active. So after I_1 (which needs to be a subset of A_i), any I_2 that adds the remaining elements $A_i \setminus I_1$ (plus possibly any elements that were already in I_1) will activate the constraint. For example, the effect of different interventions on a set $A_{\mathrm{do}(\emptyset)} = \{\emptyset, \{1,2\}, \{2,3\}\}$:

$$\begin{aligned}
A_{\mathrm{do}(1)} &= \{\{2\}, \{1,2\}\}, \\
A_{\mathrm{do}(2)} &= \{\{1\}, \{1,2\}, \{3\}, \{2,3\}\}, \\
A_{\mathrm{do}(\{1,2\})} &= A_{\mathrm{do}(1)\mathrm{do}(2)} = A_{\mathrm{do}(2)\mathrm{do}(1)} \\
&= \{\emptyset, \{1\}, \{2\}, \{1,2\}\}, \\
A_{\mathrm{do}(\{1,2,3\})} &= \emptyset.
\end{aligned}$$

Lemma 3.1 shows that the effect of multiple interventions on a CCM depends neither on whether the interventions are performed simultaneously or sequentially nor on the order in which they are performed.

Lemma 3.1. *Let $\mathcal{M}$ be a CCM for variables indexed by $\mathcal{I}$ and let $I, J \subseteq \mathcal{I}$ be two disjoint sets of intervention targets with intervention values $\boldsymbol{\xi}_I \in \boldsymbol{\mathcal{X}}_I$ and $\boldsymbol{\xi}_J \in \boldsymbol{\mathcal{X}}_J$ respectively. Then*

$$\begin{aligned}\left(\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}\right)_{\mathrm{do}(J,\boldsymbol{\xi}_J)} &= \left(\mathcal{M}_{\mathrm{do}(J,\boldsymbol{\xi}_J)}\right)_{\mathrm{do}(I,\boldsymbol{\xi}_I)} \\ &= \mathcal{M}_{\mathrm{do}(I\cup J,\boldsymbol{\xi}_{I\cup J})}.\end{aligned}$$

Proof. The result follows directly from Definition 3.9. □

Example 3.1 (continuing from p. 54)**.** Suppose that the supply of a product, if it is not targeted by an intervention, is determined by a function f_S, which takes as input the price of the product P and an exogenous random variable E (e.g. the cost of production). The system for price, supply, and demand can be represented by an (underspecified) CCM $\mathcal{M} = (\mathbb{R}^3, \Phi, E)$, where Φ consists of two causal constraints:

$$\begin{aligned}&(S - D, && 0, && \{\emptyset, \{D\}, \{S\}, \{D, S\}\}),\\ &(S - f_S(P, E), && 0, && \{\emptyset, \{D\}, \{P\}, \{D, P\}\}).\end{aligned}$$

After an intervention on P we get $\mathcal{M}_{\mathrm{do}(P,\xi_P)} = (\mathbb{R}^3, \widetilde{\Phi}, E)$, where the updated set of causal constraints is given by

$$\begin{aligned}&(S - f_S(P, E), && 0, && \{\emptyset, \{D\}, \{P\}, \{D, P\}\}),\\ &(P, && \xi_P, && \{\emptyset, \{D\}, \{S\}, \{D, S\}\}).\end{aligned}$$

Note that after an intervention on P, there would be no intervention under which the causal constraint $(S - D,\ 0,\ \{\emptyset, \{D\}, \{S\}, \{D, S\}\})$ is still active (not even for the null intervention), so it is discarded from $\widetilde{\Phi}$. △

3.3.3 From SCM to CCM

Structural equations in SCMs are constraints that are active as long as their corresponding variables are not targeted by interventions. This can be used to demonstrate how, for real-valued SCMs, an equivalent CCM with the same solutions under interventions can be constructed.[6]

Proposition 3.1. *Let $\mathcal{M}^{SCM} = (\mathbb{R}^p, F, \mathbf{E})$ be a real-valued SCM and $\mathcal{I} = \{1, \dots, p\}$ an index set. The CCM $\mathcal{M}^{CCM} = (\mathbb{R}^p, \Phi, \mathbf{E})$ with causal constraints Φ:*

$$\left(f_j(\mathbf{x}_{\mathrm{pa}(j)}, \mathbf{e}_{\mathrm{pa}(j)}) - x_j,\ 0,\ A_j = \mathcal{P}(\mathcal{I}\backslash\{j\})\right),\ \forall j \in \mathcal{I},$$

[6]The general case, where variables take value in a standard measurable space, requires an additive structure on the variable domains with a zero-element.

has the same solutions as $\mathcal{M}^{SCM}$ under any intervention.

Proof. The result follows from Definitions 3.8 and 3.9. □

3.3.4 Equilibrium causal models

We have seen that SCMs may fail to completely capture the causal semantics of stationary behaviour in dynamical systems. Here we prove that CCMs can always completely represent the causal semantics of dynamical systems at equilibrium.

Theorem 3.2. *Let $\mathcal{D}$ be a dynamical system such that for all $I \subseteq \mathcal{I}$ and all $\boldsymbol{\xi}_I \in \mathfrak{X}_I$, $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ has a unique solution of the form* (3.2)*. Then there exists a CCM $\mathcal{M}(\mathcal{D})$ such that for all $I \subseteq \mathcal{I}$ and all $\boldsymbol{\xi}_I \in \mathfrak{X}_I$:*

- *the equilibrium solutions of $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ coincide with the solutions of $\big(\mathcal{M}(\mathcal{D})\big)_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$,*
- *the following diagram commutes:*

$$\begin{array}{ccc} \mathcal{D} & \longmapsto & \mathcal{M}(\mathcal{D}) \\ \big\downarrow & & \big\downarrow \\ \mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)} & \longmapsto & \big(\mathcal{M}(\mathcal{D})\big)_{\mathrm{do}(I,\boldsymbol{\xi}_I).} \end{array}$$

Proof. By assumption, the intervened system $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ has a unique solution $\mathbf{X}_t(\boldsymbol{\xi}_I, \mathbf{e}_{\mathcal{J}\setminus I})$ given by $\mathbf{X}(t,(\boldsymbol{\xi}_I, \mathbf{e}_{\mathcal{J}\setminus I}))$ which is measurable in $(\boldsymbol{\xi}_I, \mathbf{e}_{\mathcal{J}\setminus I})$ for all t. For $I \subseteq \mathcal{I}$ define:

$$\mathbf{C}_I = \{(\boldsymbol{\xi}_I, \mathbf{e}_{\mathcal{J}\setminus I}) \in \mathbb{R}^{|I|} \times \mathbb{R}^{|\mathcal{J}\setminus I|} : \quad \mathbf{X}_t(\boldsymbol{\xi}_I, \mathbf{e}_{\mathcal{J}\setminus I}) \text{ converges as } t \to \infty\}. \tag{3.16}$$

Consider the measurable function $\mathbf{g}^I : \mathbb{R}^{\mathcal{J}} \times \mathbb{R}^{\mathcal{J}\setminus I} \to \mathbb{R}^{\mathcal{J}}$ defined by

$$\mathbf{g}^I(\mathbf{x}, \mathbf{e}_{\mathcal{J}\setminus I}) := \mathbf{X}^*\big((\mathbf{x}_I, \mathbf{e}_{\mathcal{J}\setminus I})\big)\mathbf{1}_{\mathbf{C}_I}\big((\mathbf{x}_I, \mathbf{e}_{\mathcal{J}\setminus I})\big) + (\mathbf{x}+\mathbf{1})\big(\mathbf{1} - \mathbf{1}_{\mathbf{C}_I}\big((\mathbf{x}_I, \mathbf{e}_{\mathcal{J}\setminus I})\big)\big) - \mathbf{x}, \tag{3.17}$$

where $\mathbf{X}^*\big((\mathbf{x}_I, \mathbf{e}_{\mathcal{J}\setminus I})\big) = \lim_{t\to\infty} \mathbf{X}_t(\boldsymbol{\xi}_I, \mathbf{e}_{\mathcal{J}\setminus I})$. The constraint $\mathbf{g}^I(\mathbf{x}, \mathbf{e}_{\mathcal{J}\setminus I}) = \mathbf{0}$ gives a contradiction if and only if $(\mathbf{x}_I, \mathbf{e}_{\mathcal{J}\setminus I}) \notin \mathbf{C}_I$, and reduces to the equation $\mathbf{x} = \mathbf{X}^*((\mathbf{x}_I, \mathbf{e}_{\mathcal{J}\setminus I}))$ otherwise. Therefore, the equilibrium solutions of $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ coincide with the solutions of the equation $\mathbf{g}^I(\mathbf{x}, \mathbf{e}_{\mathcal{J}\setminus I}) = \mathbf{0}$. The CCM $\mathcal{M}(\mathcal{D}) := (\mathfrak{X}, \Phi, \mathbf{E})$ with $\Phi = \{(\mathbf{g}^I, \mathbf{0}, A^I = \{I\}) : I \subseteq \mathcal{I}\}$ satisfies the properties of the theorem by construction. □

Theorem 3.2 proves that a CCM representation always exists that completely characterizes the causal semantics of a dynamical system at equilibrium. Although we construct a CCM in the proof of the theorem, it does not give a parsimonious representation of the system.[7] In the next section, we will outline an intuitive and more convenient construction method in the context of ODEs.

[7] Interestingly, the CCM construction in the proof of Theorem 3.2 can be applied to dynamical systems at finite time t.

3.4 From ODE to CCM

We consider how and when parsimonious CCM representations can be derived from ODEs and initial conditions in a dynamical system. We demonstrate how causal constraints, unlike SCMs can completely capture the stationary behaviour of the basic enzyme reaction and that they have no solution if the system does not converge.

3.4.1 Causal constraints from differential equations

When modelling the stationary behaviour of a system of a set of first-order ODEs, setting the time-derivatives equal to zero constrains the solution space of the equilibrium model to the fixed points of the system. A CCM allows us to interpret such constraints as *causal* by explicitly specifying under which interventions they put constraints on the equilibrium solutions of the system.

Example 3.2. For the basic enzyme reaction, some of the causal constraints are obtained by setting the time derivatives of the four variables of the system in equations (3.4) to (3.7) equal to zero. The resulting constraints are active as long as the corresponding variables are not targeted by an intervention. This leads to the causal constraints in equations (3.18) to (3.21) below,

$$(k_0 + k_{-1}C^* - k_1S^*E^*, \quad 0, \quad \mathcal{P}(\mathcal{I}\backslash\{S\})), \tag{3.18}$$
$$(-k_1S^*E^* + (k_{-1} + k_2)C^*, \quad 0, \quad \mathcal{P}(\mathcal{I}\backslash\{E\})), \tag{3.19}$$
$$(k_1S^*E^* - (k_{-1} + k_2)C^*, \quad 0, \quad \mathcal{P}(\mathcal{I}\backslash\{C\})), \tag{3.20}$$
$$(k_2C^* - k_3P^*, \quad 0, \quad \mathcal{P}(\mathcal{I}\backslash\{P\})), \tag{3.21}$$

with $\mathcal{I} = \{S, C, E, P\}$. At this stage, the CCM is equivalent to the underspecified SCM of the dynamical system that we discussed in Section 3.2.3. We will proceed in the next section by adding more causal constraints. △

Example 3.3. The Lotka-Volterra model (Murray, 2002) is a set of differential equations that is often used to describe the dynamics of a system where prey (e.g. deer) and predators (e.g. wolves), denoted by X_1 and X_2 respectively, interact. The dynamics of the biological model are given by

$$\dot{X_1} = X_1(t)(k_{11} - k_{12}X_2(t)), \tag{3.22}$$
$$\dot{X_2} = -X_2(t)(k_{22} - k_{21}X_1(t)), \tag{3.23}$$

with initial values $X_1(0) > 0, X_2(0) > 0$ and strictly positive rate parameters. The system has two fixed points $(X_1^*, X_2^*) = (0, 0)$ and $(X_1^*, X_2^*) = (k_{22}/k_{21}, k_{11}/k_{12})$, which

can be represented either by causal constraints,

$$(X_1^*(k_{11} - k_{12}X_2^*), 0, \{\emptyset, \{2\}\}), \tag{3.24}$$
$$(X_2^*(k_{22} - k_{21}X_1^*), 0, \{\emptyset, \{1\}\}), \tag{3.25}$$

or (equivalently) by structural equations:

$$X_1^* = X_1^* + X_1^*(k_{11} - k_{12}X_2^*), \tag{3.26}$$
$$X_2^* = X_2^* - X_2^*(k_{22} - k_{21}X_1^*). \tag{3.27}$$

These (structural) equations do not describe the stable steady state behavior of the model, because the system displays undamped oscillations around the positive fixed point, as was pointed out by Mooij, Janzing, and Schölkopf (2013) and Murray (2002). In the next section we proceed by adding additional relevant constraints to the CCM. △

3.4.2 Causal constraints from constants of motion

For dynamical systems that admit a constant of motion (i.e. a conserved quantity), the trajectories of its solutions are confined to a space that is constrained by its initial conditions. Hence, the solutions for the equilibrium must be similarly constrained. In a CCM we interpret these constraints as causal by specifying under which interventions they constrain the solution space.

Example 3.2 (continuing from p. 58)**.** For the basic enzyme reaction, we include the conservation law that results from the linear dependence between the time derivative of the free enzyme E and the complex C in equation (3.9). Since this relation holds as long as the 'cycle' between C and E is not broken, we obtain the following causal constraint

$$(C^* + E^* - (c_0 + e_0), \qquad 0, \qquad \mathcal{P}(\mathcal{I}\backslash\{C, E\}))\,. \tag{3.28}$$

Another conservation law appeared after intervention on the variable C. The resulting conservation law $S(t) - E(t) = s_0 - e_0$ applies as long as the 'cycle' between S and E is not broken by another intervention on the system. This leads to the final causal constraint:

$$(S^* - E^* - (s_0 - e_0), \qquad 0, \qquad \{\{C\}, \{C, P\}\})\,. \tag{3.29}$$

Let Φ be the set of causal constraints in (3.18) to (3.21) and (3.28) and (3.29). In Section 3.2.2 we showed that the active constraints in Φ have a unique solution under any intervention. If $\mathbf{E} = (s_0, e_0, c_0, p_0)$ is a set of exogenous random variables then

the CCM $\mathcal{M} = (\mathbb{R}^4_{>0}, \Phi, \mathbf{E})$ completely captures the stationary behaviour of the basic enzyme reaction. △

Remark 3.1. Interestingly, if we treat C as a latent endogenous variable that cannot be intervened upon, the equilibrium to which the dynamics of the basic enzyme reaction converges can be completely described by the following marginal CCM (see Section 3.A.3 for details):

$$\left(\frac{k_0+k_{-1}k_0/k_2}{k_1E^*} - S^*, \qquad 0, \qquad \mathcal{P}(\mathcal{I}'\backslash\{S\})\right), \tag{3.30}$$

$$\left(\frac{(k_{-1}+k_2)(c_0+e_0)}{k_{-1}+k_2+k_1S^*} - E^*, \qquad 0, \qquad \mathcal{P}(\mathcal{I}'\backslash\{E\})\right), \tag{3.31}$$

$$\left(\frac{k_2}{k_3}\frac{k_1S^*E^*}{k_{-1}+k_2} - P^*, \qquad 0, \qquad \mathcal{P}(\mathcal{I}'\backslash\{P\})\right), \tag{3.32}$$

where $\mathcal{I}'$ is an index set for $\{S, E, P\}$. From Proposition 3.1 it can be seen that there exists an equivalent SCM that does completely capture the causal semantics of S, E, and P, as long as one does not intervene on C. △

Example 3.3 (continuing from p. 58)**.** The Lotka-Volterra model provides an example of a system that admits a non-linear conservation law:

$$\begin{aligned} &-k_{21}X_1 + k_{22}\log(X_1) - k_{12}X_2 + k_{11}\log(X_2) = \\ &-k_{21}X_1(0) + k_{22}\log(X_1(0)) - k_{12}X_2(0) + k_{11}\log(X_2(0)), \end{aligned} \tag{3.33}$$

which represents a constraint that is only active in the observational setting. If the system would converge to an equilibrium (X_1^*, X_2^*) the causal constraints derived from the differential equations should hold simultaneously. These constraints are only satisfied when the system starts out in one of the fixed points (e.g. $(X_1(0), X_2(0)) = (k_{22}/k_{21}, k_{11}/k_{12})$). Otherwise the dynamical system exhibits steady-state oscillations and the set of causal constraints has no solution.

A complete causal description of how the equilibrium behaviour changes under perfect interventions can be obtained by adding the following two causal constraints:

$$(X_1^* - X_1(0)\mathbf{1}_{\{k_{11}-k_{12}X_2^*\geq 0\}}, \qquad 0, \qquad \{\{2\}\}), \tag{3.34}$$

$$(X_2^* - X_2(0)\mathbf{1}_{\{k_{22}-k_{21}X_1^*\leq 0\}}, \qquad 0, \qquad \{\{1\}\}). \tag{3.35}$$

Addition of the causal constraint in equation (3.34) ensures that after an intervention on the amount of predators X_2: a) when the amount of predators X_2 is forced to take on a constant value k_{11}/k_{12} the amount of prey X_1 remains constant, b) the amount of prey X_1 goes extinct when there are more predators, and c) the model has no solution if there are fewer predators. This can easily be verified by attempting to

solve the active constraints in equations (3.26) and (3.34). The causal constraint in equation (3.35) can be interpreted similarly. Together, the causal constraints in equations (3.24), (3.25), (3.33), (3.34), and (3.35) capture the stationary behaviour of the predator-prey model.[8] The SCM on the other hand has the fixed points of the system as a solution and does not predict the non-convergent behaviour. △

3.4.3 Constructing CCMs

Causal constraints (or structural equations) derived from differential equations result in a causal description of the fixed points in a system. Clearly, the set of solutions to a CCM which is constructed from the ODEs and constants of motion in a dynamical system contains the fixed points of that system. For structurally semistable systems the addition of causal constraints derived from constants of motion results in a complete causal description of the system's stationary behaviour when the constraints specify the equilibria in terms of initial conditions, as can be seen from Theorem 3.3 and Corollary 3.1 below.

Theorem 3.3. *Let $\mathcal{D}$ be a dynamical system that converges to a fixed point if it has at least one. Let $\mathcal{M}$ be a CCM constructed from the ODEs and constants of motion in $\mathcal{D}$ for which all solutions, if they exist, are unique up to $\mathbb{P}^{\mathbf{E}}$-zero sets. $\mathcal{D}$ converges to an equilibrium $\mathbf{X}^*$ if and only if $\mathbf{X}^*$ is a solution of $\mathcal{M}$.*

Proof. First assume that $\mathcal{D}$ has a fixed point, so that $\mathcal{D}$ converges to an equilibrium $\mathbf{X}^*(\mathbf{e})$ for almost every $\mathbf{e} \in \mathbb{R}^p$. We have that a) $\mathbf{X}^*(\mathbf{e})$ satisfies the constants of motion in the dynamical system and b) for $\mathbf{X}^*(\mathbf{e})$ the time-derivatives appearing in the ODEs are equal to zero. Hence if $\mathcal{D}$ converges to $\mathbf{X}^*$ then $\mathbf{X}^*$ is a solution of $\mathcal{M}$. Since $\mathcal{M}$ has no more than one solution (up to zero sets), the reverse statement is also true. Now assume that $\mathcal{D}$ has no fixed point. In that case $\mathcal{M}$ has no solutions, and $\mathcal{D}$ cannot converge to an equilibrium. □

Corollary 3.1. *Let $\mathcal{D}$ be structurally semistable and $\mathcal{M}$ a CCM constructed from the ODEs and constants of motion in $\mathcal{D}$ for which under any intervention, all solutions, if they exist, are unique up to $\mathbb{P}^{\mathbf{E}}$-zero sets. Then for all $I \subseteq \mathcal{I}$ and $\boldsymbol{\xi}_I \in \mathbb{R}^{|I|}$: $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ converges to an equilibrium $\mathbf{X}^*(I, \boldsymbol{\xi}_I)$ if and only if $\mathbf{X}^*(I, \boldsymbol{\xi}_I)$ is a solution of $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$.*

Proof. If $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ has a solution then $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ has a fixed point with $\mathbf{x}_I^* = \boldsymbol{\xi}_I$ and it converges because $\mathcal{D}$ is structurally semistable. If $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ has no solution then $\mathcal{D}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ does not converge to a fixed point. The result follows from Theorem 3.3 and Definition 3.9. □

The basic enzyme reaction in Example 3.2 is structurally semistable, while the Lotka-Volterra model in Example 3.3 is not. Corollary 3.1 tells us that for structurally

[8] This can be verified by explicitly calculating the solutions of the model under all interventions.

semistable systems, if a CCM constructed from ODEs and constants of motions has at most one solution under any intervention, then the CCM completely captures the causal semantics of the stationary behaviour of the system.

3.5 Functional laws

CCMs can also represent *functional laws*, which are relations between variables that are invariant under *all* interventions. Causal constraints allow one to explicitly state under which interventions a constraint is active. Therefore a CCM never admits a solution that violates the functional law, where an SCM would.

Example 3.4. It is well-known that the pressure P and temperature T for N particles of an ideal gas in a fixed volume V are related by the ideal gas law. In absence of any knowledge about the environment, this system can be represented by the (underspecified) CCM $\mathcal{M} = (\mathbb{R}^2, \{(PV - Nk_BT, 0, \mathcal{P}(\mathcal{I}))\}, \mathbb{P}_\emptyset)$, where k_B is Boltzmann's constant, and $\mathcal{I}$ is an index set for the variables (P, T) in the system. If we were to describe the same system using an SCM, then we would need two copies of this causal constraint as structural equations:

$$P = \frac{Nk_BT}{V}, \qquad T = \frac{PV}{Nk_B}.$$

Indeed, considering interventions on one of the variables leaves no choice for the structural equation of the other one. Furthermore, a simultaneous intervention on P and T always has a solution in the SCM representation, even when this means that the ideal gas law is violated. The CCM representation typically does not have a solution under such an intervention (unless the target values satisfy the ideal gas law constraint). Therefore, the CCM representation of functional laws like the ideal gas law is more parsimonious and more natural than any SCM representation can be. △

A functional law can be any relation that is invariant under all interventions. For example, a transformation of a (set of) variables to another (set of) variables describing the same system can also be modelled as a functional law.

Example 3.5. Let $\mathcal{I}$ be an index set of (T, V, O). Suppose that the viscosity T of a salad dressing, consisting of a certain amount of oil O and a certain amount of vinegar V is determined by a causal constraint $\phi = (f, 0, \mathcal{P}(\mathcal{I}\backslash\{T\}))$ where f is a function depending on the amount of oil and vinegar. By adding causal constraints

$$\begin{aligned} &(O_r - O/(O+V), && 0, && \mathcal{P}(\mathcal{I})),\\ &(V_r - V/(O+V), && 0, && \mathcal{P}(\mathcal{I})), \end{aligned}$$

a CCM allows us to have the relative amounts of oil and vinegar O_r and V_r in the model without running into logical contradictions. △

3.6 Conclusion

While Structural Causal Models (SCMs) form a very popular modeling framework in many applied sciences, we have shown that they are neither powerful enough to model the rich equilibrium behavior of simple dynamical systems such as the basic enzyme reaction, nor simple functional laws of nature like the ideal gas law. This raises the question whether the common starting point in causal discovery—that the data-generating process can be modeled with an SCM—is tenable in certain application domains, for example, for biochemical systems.

We believe that the examples presented in this paper form a compelling motivation to extend the common causal modeling framework to potentially broaden the impact of causal modeling in dynamical systems. In this work, we introduced Causal Constraints Models (CCMs). We showed how they can be 'constructed' from differential equations and initial conditons and proved that they can completely capture the causal semantics of functional laws and stationary behavior in dynamical systems.

One intuitively appealing aspect of SCMs is their graphical interpretation. In Chapters 4 and 5, we investigate graphical representations of the causal and independence structure implied by sets of (causal) constraints. This allows us to better understand the causal and independence properties of certain dynamical systems at equilibrium.

3.A Explicit calculations for the basic enzyme reaction

In this section we present additional results concerning the basic enzyme reaction that were already highlighted throughout this chapter. First we present a table with all fixed points of the basic enzyme reaction. We proceed with proving that the dynamical system of a basic enzyme reaction converges to a fixed point whenever a fixed point exists. Finally, we derive a simple marginal model from the CCM representation of the basic enzyme reaction.

3.A.1 Fixed points

The fixed points of the basic enzyme reaction, for all intervened systems, are given in Table 3.2. For any intervention, these are obtained by solving the system of equations that one gets by considering the causal constraints of the CCM for the basic enzyme reaction that are active under that specific intervention. In other words, we take all equations for which a particular intervention is in the activation set.

Table 3.2: Fixed points of the basic enzyme reaction, where $y = \frac{1}{2}\sqrt{(e_0 - s_0)^2 + 4\frac{k_0(k_{-1}+k_2)}{k_1 k_2}}$. The symbol $\emptyset$ is used to indicate that the CCM has no fixed points.

intervention	S^*	C^*	E^*	P^*
none	$\frac{k_0+k_{-1}\frac{k_0}{k_2}}{k_1(e_0+c_0-\frac{k_0}{k_2})}$	$\frac{k_0}{k_2}$	$e_0 + c_0 - \frac{k_0}{k_2}$	$\frac{k_0}{k_3}$
$\mathrm{do}(S = s)$	s	$\frac{k_1 s(e_0+c_0)}{k_{-1}+k_2+k_1 s}$	$\frac{(k_{-1}+k_2)(e_0+c_0)}{k_{-1}+k_2+k_1 s}$	$\frac{k_2}{k_3}\frac{k_1 s(e_0+c_0)}{k_{-1}+k_2+k_1 s}$
$\mathrm{do}(C = c), c = \frac{k_0}{k_2}$	$\frac{(s_0-e_0)}{2} + y$	c	$\frac{-(s_0-e_0)}{2} + y$	$\frac{k_2}{k_3}c$
$\mathrm{do}(C = c), c \neq \frac{k_0}{k_2}$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$\mathrm{do}(E = e)$	$\frac{k_0+k_{-1}\frac{k_0}{k_2}}{k_1 e}$	$\frac{k_0}{k_2}$	e	$\frac{k_0}{k_3}$
$\mathrm{do}(P = p)$	$\frac{k_0+k_{-1}\frac{k_0}{k_2}}{k_1(e_0+c_0-\frac{k_0}{k_2})}$	$\frac{k_0}{k_2}$	$e_0 + c_0 - \frac{k_0}{k_2}$	p
$\mathrm{do}(S = s, C = c)$	s	c	$\frac{k_{-1}+k_2}{k_1}\frac{c}{s}$	$\frac{k_2}{k_3}c$
$\mathrm{do}(S = s, E = e)$	s	$\frac{k_1}{k_{-1}+k_2}se$	e	$\frac{k_2}{k_3}\frac{k_1}{k_{-1}+k_2}se$
$\mathrm{do}(S = s, P = p)$	s	$\frac{k_1 s(e_0+c_0)}{k_{-1}+k_2+k_1 s}$	$\frac{(k_{-1}+k_2)(e_0+c_0)}{k_{-1}+k_2+k_1 s}$	p
$\mathrm{do}(C = c, E = e)$	$\frac{k_0+k_{-1}c}{k_1 e}$	c	e	$\frac{k_2}{k_3}c$
$\mathrm{do}(C = c, P = p), c = \frac{k_0}{k_2}$	$\frac{(s_0-e_0)}{2} + y$	c	$\frac{-(s_0-e_0)}{2} + y$	p
$\mathrm{do}(C = c, P = p), c \neq \frac{k_0}{k_2}$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$\mathrm{do}(E = e, P = p)$	$\frac{k_0+k_{-1}\frac{k_0}{k_2}}{k_1 e}$	$\frac{k_0}{k_2}$	e	p
$\mathrm{do}(S = s, C = c, E = e)$	s	c	e	$\frac{k_2}{k_3}c$
$\mathrm{do}(S = s, C = c, P = p)$	s	c	$\frac{k_{-1}+k_2}{k_1}\frac{c}{s}$	p
$\mathrm{do}(S = s, E = e, P = p)$	s	$\frac{k_1}{k_{-1}+k_2}se$	e	p
$\mathrm{do}(C = c, E = e, P = p)$	$\frac{k_0+k_{-1}c}{k_1 e}$	c	e	p
$\mathrm{do}(S = s, C = c, E = e, P = p)$	s	c	e	p

3.A.2 Convergence results for the basic enzyme reaction

In this section, we show that the basic enzyme reaction always converges to its fixed point, as long as it exists. We also show that the intervened basic enzyme reaction has the same property. To prove this result we rely on both explicit calculations and a convergence property of so-called cooperative systems that we obtained from Belgacem et al. (2012). To prove convergence for the observed system and the system after interventions on P and E, we use the latter technique. Convergence to the equilibrium solution after interventions on S and C can be shown by explicit calculation. The convergence results for combinations of interventions can be obtained by a trivial

extension of the arguments that were used in the other cases.

3.A.2.1 Cooperativity in the basic enzyme reaction

To show that the basic enzyme reaction converges to a unique equilibrium, if it exists, we make use of a result that we obtained from Belgacem et al. (2012): cooperative systems (see Definition 3.10) have the attractive convergence property (see Proposition 3.2).

Definition 3.10. A system of ODEs $\dot{\mathbf{X}}$ is *cooperative* if the Jacobian matrix has non-negative off-diagonal elements, or there exists an integer k such that the Jabobian has $(k \times k)$ and $(n-k) \times (n-k)$ main diagonal matrices with non-negative off-diagonal entries and the rectangular off-diagonal sub-matrices have non-positive entries.

Proposition 3.2. *Let $\dot{\mathbf{X}} = \mathbf{f}(\mathbf{X})$ be a cooperative system with a fixed point $\mathbf{x}^*$. If there exist two points $\mathbf{x}_{min}, \mathbf{x}_{max} \in \mathfrak{X}$ such that $\mathbf{x}_{min} \leq \mathbf{x}^* \leq \mathbf{x}_{max}$ and $\mathbf{f}(\mathbf{x}_{min}) \geq 0$ and $\mathbf{f}(\mathbf{x}_{max}) \leq 0$, then the hyper-rectangle betweeen $\mathbf{x}_{min}$ and $\mathbf{x}_{max}$ is invariant*[9] *and for almost all initial conditions inside this rectangle the solution converges to $\mathbf{x}^*$.*

3.A.2.2 Convergence of the observed system

Recall that the dynamics of the basic enzyme reaction are given by

$$\dot{S}(t) = k_0 - k_1 S(t)E(t) + k_{-1}C(t), \tag{3.36}$$
$$\dot{E}(t) = -k_1 S(t)E(t) + (k_{-1} + k_2)C(t), \tag{3.37}$$
$$\dot{C}(t) = k_1 S(t)E(t) - (k_{-1} + k_2)C(t), \tag{3.38}$$
$$\dot{P}(t) = k_2 C(t) - k_3 P(t), \tag{3.39}$$
$$S(0) = s_0, \quad E(0) = e_0, \quad C(0) = c_0, \quad P(0) = p_0, \tag{3.40}$$

where $\mathbf{x}_0 = (s_0, e_0, c_0, p_0)$ are the initial conditions of the system. The analysis in Belgacem et al. (2012) applies to a slightly different model that also includes feedback from P to C. Here, we will adapt their analysis which is based on Proposition 3.2. Notice that the arguments that will be given in this section can be applied to the dynamics of a system where P is intervened upon as well.

We start by rewriting the system of ODEs in equation (3.36) to (3.39) by using the fact that $\dot{E}(t) + \dot{C}(t) = 0$ so that $E(t) = e_0 + c_0 - C(t)$:

$$\dot{S}(t) = k_0 - k_1 S(t)(e_0 + c_0 - C(t)) + k_{-1}C(t), \tag{3.41}$$
$$\dot{C}(t) = k_1 S(t)(e_0 + c_0 - C(t)) - (k_{-1} + k_2)C(t), \tag{3.42}$$
$$\dot{P}(t) = k_2 C(t) - k_3 P(t). \tag{3.43}$$

[9] An invariant set is a set with the property that once a trajectory of a dynamical set enters it, it cannot leave.

Cooperativity The corresponding Jacobian matrix is given by,

$$J(S,C,P) = \begin{pmatrix} -k_1(e_0+c_0-C(t)) & k_{-1}+k_1S(t) & 0 \\ k_1(e_0+c_0-C(t)) & -(k_{-1}+k_2)-k_1S(t) & 0 \\ 0 & k_2 & -k_3 \end{pmatrix}. \tag{3.44}$$

Since all off-diagonal elements in the Jacobian matrix are non-negative, the observational system is a cooperative system by Definition 3.10.

Convergence From Table 3.2 we find that the observed system has a unique (positive) fixed point as long as $e_0 + c_0 > \frac{k_0}{k_2}$. We want to use Proposition 3.2 to show that the system converges to this fixed point, so we need to find $\mathbf{x}_{\min}$ and $\mathbf{x}_{\max}$ so that all three derivatives are non-negative and non-positive respectively.

For $\mathbf{x}_{\min} = (0,0,0)$, then $\dot{S} = k_0 > 0$ and $\dot{C} = \dot{P} = 0$ so all derivatives are nonnegative. The upper vertex must be chosen so that all derivative are non-positive:

$$\begin{aligned} \dot{S} \leq 0 &\iff S \geq \frac{k_0 + k_{-1}C}{k_1(e_0+c_0-C)}, \\ \dot{C} \leq 0 &\iff S \geq \frac{(k_{-1}+k_2)C}{k_1(e_0+c_0-C)}, \\ \dot{P} \leq 0 &\iff P \geq \frac{k_2}{k_3}C. \end{aligned}$$

The basic enzyme reaction only has a fixed point as long as $C < e_0 + c_0$ (otherwise $\dot{S}(t) > 0$). If we let C approach $e_0 + c_0$, then the inequality constraints on the derivatives are satisfied as S and P go to infinity. More formally we can choose

$$\mathbf{x}_{\max} = \left(S = \max\left(\frac{k_0 + k_{-1}C}{k_1(e_0+c_0-C)}, \frac{(k_{-1}+k_2)C}{k_1(e_0+c_0-C)}\right), C = e_0 + c_0 - \epsilon, P = \frac{k_2}{k_3}C + \frac{1}{\epsilon}\right).$$

When ϵ approaches zero, both S and P go to infinity and all derivatives are non-positive. Hence, by Proposition 3.2, the system converges to its fixed point for almost all valid initial values of S, C, and P (for which the fixed point exists).

3.A.2.3 Intervention on E

Similarly, we can also show that the system where E is targeted by an intervention that sets it equal to e, converges to the (unique) equilibrium in Table 3.2. The intervened system of ODEs is given by

$$\begin{aligned} \dot{S} &= k_0 - k_1 eS + k_{-1}C, \\ \dot{C} &= k_1 eS - (k_{-1}+k_2)C, \\ \dot{P} &= k_2 C - k_3 P. \end{aligned}$$

The Jacobian is given by

$$J(S,C,P) = \begin{pmatrix} -k_1 e & k_{-1} & 0 \\ k_1 e & -(k_{-1}+k_2) & 0 \\ 0 & k_2 & -k_3 \end{pmatrix}. \tag{3.45}$$

Since all off-diagonal elements are non-negative this is a cooperative system by Definition 3.10.

All derivatives are non-negative at the point $(S,C,P) = (0,0,0)$, and all derivatives are non-positive at the point (s,c,p) where

$$s = \max\left(\frac{k_{-1}c + k_0}{k_1 e}, \frac{(k_{-1}+k_2)c}{k_1 e}\right),$$
$$p = \frac{k_2}{k_3}c,$$

where $c \to \infty$. We then apply Proposition 3.2 to show that the intervened system converges to the equilibrium value from all valid initial values.

3.A.2.4 Intervention on S

We show that the system converges to the equilibrium solution after an intervention on S by explicit calculation. The intervened system of ODEs is given by

$$\begin{aligned} \dot{S}(t) &= 0, \\ \dot{E}(t) &= -k_1 s E(t) + (k_{-1}+k_2)C(t), \\ \dot{C}(t) &= k_1 s E(t) - (k_{-1}+k_2)C(t), \\ \dot{P}(t) &= k_2 C(t) - k_3 P(t). \end{aligned}$$

Since $\dot{C}(t) + \dot{E}(t) = 0$, we can write $E(t) = e_0 + c_0 - C(t)$, resulting in the following differential equation

$$\begin{aligned} \dot{C}(t) &= k_1 s(e_0 + c_0 - C(t)) - (k_{-1}+k_2)C(t), & (3.46) \\ &= -(k_1 s + k_{-1} + k_2)C(t) + k_1 s(e_0 + c_0). & (3.47) \end{aligned}$$

We take the limit $t \to \infty$ of the solution to the initial value problem to obtain

$$C^* = \lim_{t\to\infty} \frac{k_1 s(e_0+c_0)}{(k_1 s + k_1 + k_2)} + e^{-(k_1 s + k_{-1} + k_2)t} = \frac{k_1 s(e_0+c_0)}{(k_1 s + k_{-1} + k_2)}. \tag{3.48}$$

The result for E follows from the fact that $E(t) = e_0 + c_0 - C(t)$. The result for P follows by explicitly solving the differential equation and taking the limit $t \to \infty$.

3.A.2.5 Intervention on C

There is no equilibrium solution when the intervention targeting C does not have value $\frac{k_0}{k_2}$, as can be seen from Table 3.2. To show that the system converges when the equilibrium solution exists, we can explicitly solve the initial value problem and take the limit $t \to \infty$. The intervened system of ODEs after an intervention $\text{do}(C = \frac{k_0}{k_2})$ is given by

$$\begin{aligned}
\dot{S}(t) &= -k_1 S(t)E(t) + (k_{-1} + k_2)k_0/k_2 = -k_1 S(t)E(t) + k,\\
\dot{E}(t) &= -k_1 S(t)E(t) + (k_{-1} + k_2)k_0/k_2 = -k_1 S(t)E(t) + k,\\
\dot{C}(t) &= 0,\\
\dot{P}(t) &= k_0 - k_3 P(t),
\end{aligned}$$

where we set $k = (k_{-1} + k_2)k_0/k_2$ for brevity. The initial value problem for P can be solved explicitly, and by taking the limit $t \to \infty$ we obtain

$$P^* = \lim_{t\to\infty} P(t) = \lim_{t\to\infty} \frac{k_0}{k_3} + c \cdot e^{-k_3 t} = \frac{k_0}{k_3},$$

which is the same as the equilibrium solution in Table 3.2. The solution for S is more involved. First we substitute $E(t) = S(t) - (s_0 - e_0)$ (since $\dot{S}(t) - \dot{E}(t) = 0$) which gives us the following differential equation

$$\dot{S}(t) = -k_1 S(t)(S(t) - (s_0 - e_0)) + k = -k_1 S(t)^2 + (s_0 - e_0)k_1 S(t) + k.$$

To solve this differential equation we first divide both sides by $(-k_1(S(t))^2 + (s_0 - e_0)k_1 S(t) + k)$, and integrate both sides with respect to t,

$$\int \frac{dS(t)/dt}{-k_1 S(t)^2 + (s_0 - e_0)k_1 S(t) + k} dt = \int 1 dt \tag{3.49}$$

$$\int \frac{dS(t)}{-k_1 S(t)^2 + (s_0 - e_0)k_1 S(t) + k} = (t + c) \tag{3.50}$$

To evaluate the left-hand side of this equation we want to apply the following standard integral:

$$\int \frac{1}{ax^2 + bx + c} dx = \begin{cases} -\frac{2}{\sqrt{b^2-4ac}} \tanh^{-1}\left(\frac{2ax+b}{\sqrt{b^2-4ac}}\right) + C, & \text{if } |2ax + b| < \sqrt{b^2 - 4ac}, \\ -\frac{2}{\sqrt{b^2-4ac}} \coth^{-1}\left(\frac{2ax+b}{\sqrt{b^2-4ac}}\right) + C, & \text{else.} \end{cases} \tag{3.51}$$

for $b^2 - 4ac > 0$. We first check the condition:

$$b^2 - 4ac = (s_0 - e_0)^2 k_1^2 + 4k_1 k > 0.$$

We now take the first solution to the standard integral (the second solution gives the same limiting result for S, as we will see later on). We apply the first solution in (3.51) to (3.50) to obtain the following equivalent equations:

$$\frac{2\tanh^{-1}\left(\frac{2k_1S(t)-(s_0-e_0)k_1}{\sqrt{4k_1k+(s_0-e_0)^2k_1^2}}\right)}{\sqrt{4k_1k+(s_0-e_0)^2k_1^2}} = t+c \tag{3.52}$$

$$\tanh^{-1}\left(\frac{2k_1S(t)-(s_0-e_0)k_1}{\sqrt{4k_1k+(s_0-e_0)^2k_1^2}}\right) = \frac{1}{2}(t+c)\sqrt{4k_1k+(s_0-e_0)^2k_1^2} \tag{3.53}$$

$$\frac{2k_1S(t)-(s_0-e_0)k_1}{\sqrt{4k_1k+(s_0-e_0)^2k_1^2}} = \tanh\left(\frac{1}{2}(t+c)\sqrt{4k_1k+(s_0-e_0)^2k_1^2}\right), \tag{3.54}$$

Solving (3.54) for S gives,

$$S(t) = \frac{1}{2k_1}\left(\tanh\left(\frac{1}{2}(t+c)\sqrt{4k_1k+(s_0-e_0)^2k_1^2}\right)\sqrt{4k_1k+(s_0-e_0)^2k_1^2}+k_1(s_0-e_0)\right).$$

By taking the limit $t\to\infty$, plugging in $k=(k_{-1}+k_2)\frac{k_0}{k_2}$, and rewriting we obtain the equilibrium solution in Table 3.2:

$$\begin{aligned}\lim_{t\to\infty}S(t) &= \frac{k_1(s_0-e_0)+\sqrt{4k_1k+(s_0-e_0)^2k_1^2}}{2k_1}\\ &= \frac{k_1(s_0-e_0)+\sqrt{4k_1(k_{-1}+k_2)\frac{k_0}{k_2}+(s_0-e_0)^2k_1^2}}{2k_1}\\ &= \frac{1}{2}\left((s_0-e_0)+\sqrt{(s_0-e_0)^2+4\frac{k_0(k_{-1}+k_2)}{k_1k_2}}\right).\end{aligned}$$

Note that if we take the second solution to the standard integral in (3.51), then we would have ended up with the same solution for $S(t)$ with tanh replaced by coth, but the limit $\lim_{t\to\infty}S(t)$ would still be the same. The solution for E follows from the fact that $E(t)=S(t)-(s_0-e_0)$. The solutions for all joint interventions were found by combining the arguments that were given for the single interventions.

3.A.3 Marginal model

In the paper we presented a marginal model for the basic enzyme reaction. Here we show how it can be derived from the causal constraints in the CCM, which are given

by

$$k_0 + k_{-1}C - k_1SE = 0, \qquad \mathcal{P}(\mathcal{I}\backslash\{S\}), \tag{3.55}$$
$$k_1SE - (k_{-1} + k_2)C = 0, \qquad \mathcal{P}(\mathcal{I}\backslash\{C\}), \tag{3.56}$$
$$-k_1SE + (k_{-1} + k_2)C = 0, \qquad \mathcal{P}(\mathcal{I}\backslash\{E\}), \tag{3.57}$$
$$k_2C - k_3P = 0, \qquad \mathcal{P}(\mathcal{I}\backslash\{P\}), \tag{3.58}$$
$$C + E - (c_0 + e_0) = 0, \qquad \mathcal{P}(\mathcal{I}\backslash\{C, E\}), \tag{3.59}$$
$$S - E - (s_0 - e_0) = 0, \qquad \{\{C\}, \{C, P\}\}. \tag{3.60}$$

We obtain the marginal model as follows:

(i) Reduce the number of variables that can be targeted by an intervention: $\mathcal{I}' = \{S, E, P\}$.

(ii) Rewrite the causal constraint in (3.56) to $C = \frac{k_1SE}{k_{-1}+k_2}$. Note that this equation holds under any intervention in $\mathcal{P}(\mathcal{I}') = \mathcal{P}(\mathcal{I}\backslash\{C\})$. Then substitute this expression for C into equation (3.55) to obtain

$$\frac{k_0 + k_{-1}\frac{k_0}{k_2}}{k_1E} - S = 0, \qquad \mathcal{P}(\mathcal{I}'\backslash\{S\}),$$

where the activation set of the causal constraint is given by the intersection $\mathcal{P}(\mathcal{I}\backslash\{S\}) \cap \mathcal{P}(\mathcal{I}')$. Then substitute this expression for C into equation (3.58) to obtain

$$\frac{k_2}{k_3}\frac{k_1SE}{k_{-1} + k_2} - P = 0, \qquad \mathcal{P}(\mathcal{I}'\backslash\{P\}),$$

where the activation set of the causal constraint is given by the intersection $\mathcal{P}(\mathcal{I}\backslash\{P\}) \cap \mathcal{P}(\mathcal{I}')$.

(iii) Rewrite the causal constraint in (3.59) to $C = e_0 + c_0 - E$ and note that this equation holds under interventions in $\mathcal{P}(\mathcal{I}'\backslash\{E\})$. Then substitute this expression for C into equation (3.57) to obtain

$$\frac{(k_{-1} + k_2)(c_0 + e_0)}{k_{-1} + k_2 + k_1S} - E = 0, \qquad \mathcal{P}(\mathcal{I}'\backslash\{E\}),$$

where the activation set of the causal constraint is given by the intersection $\mathcal{P}(\mathcal{I}\backslash\{C, E\}) \cap \mathcal{P}(\mathcal{I}'\backslash\{E\})$.

This procedure results in the following marginal model

$$\frac{k_0 + k_{-1}\frac{k_0}{k_2}}{k_1E} - S = 0, \qquad \mathcal{P}(\mathcal{I}'\backslash\{S\}),$$

$$\frac{(k_{-1}+k_2)(c_0+e_0)}{k_{-1}+k_2+k_1S} - E = 0, \qquad \mathcal{P}(\mathcal{I}'\backslash\{E\}),$$
$$\frac{k_2}{k_3}\frac{k_1SE}{k_{-1}+k_2} - P = 0, \qquad \mathcal{P}(\mathcal{I}'\backslash\{P\}).$$

Because we kept track of the interventions under which each equation is active when we substituted C into the equations of other causal constraints, we preserved the causal structure of the model. That is, the marginal CCM model has the same solutions as the original CCM under interventions in $\mathcal{P}(\mathcal{I}')$.

Chapter 4

Conditional independences and causal relations implied by sets of equations

Based on:
Conditional independences and causal relations implied by sets of equations,
T. Blom, M.M. van Diepen, and J.M. Mooij,
Journal of Machine Learning Research 22.178 (2021), pp. 1-62.

Real-world complex systems are often modelled by sets of equations with endogenous and exogenous variables. What can we say about the causal and probabilistic aspects of variables that appear in these equations without explicitly solving the equations? We make use of Simon's causal ordering algorithm (Simon, 1953) to construct a *causal ordering graph* and prove that it expresses the effects of soft and perfect interventions on the equations under certain unique solvability assumptions. We further construct a *Markov ordering graph* and prove that it encodes conditional independences in the distribution implied by the equations with independent random exogenous variables, under a similar unique solvability assumption. We discuss how this approach reveals and addresses some of the limitations of existing causal modelling frameworks, such as causal Bayesian networks and structural causal models.

4.1 Introduction

The discovery of causal relations is a fundamental objective in many scientific endeavours. The process of the scientific method usually involves a hypothesis, such as a causal graph or a set of equations, that explains observed phenomena. Such

a graph structure can be learned automatically from conditional independences in observational data via causal discovery algorithms, e.g. the well-known PC and FCI algorithms (Spirtes, Glymour, et al., 2000; J. Zhang, 2008). The crucial assumption in *causal discovery* is that directed edges in this learned graph express causal relations between variables. However, an immediate concern is whether directed mixed graphs actually can simultaneously encode the causal semantics and the conditional independence constraints of a system.[1] We explicitly define soft and perfect interventions on sets of equations and demonstrate that, for some models, a single graph expressing conditional independences between variables via d-separations does not seem to represent the effects of these interventions in an unambiguous way, while graphs that also have vertices representing equations do encode both the dependence and causal structure implied by these models. In particular, we show that the output of the PC algorithm does not have a straightforward causal interpretation when it is applied to data generated by a simple dynamical model with feedback at equilibrium.

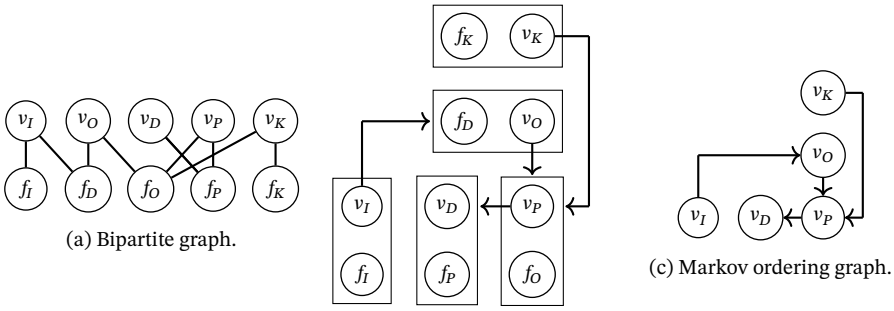
It is often said that the "gold standard" in causal discovery is controlled experimentation. Indeed, the main principle of the scientific method is to derive predictions from a hypothesis, such as a causal graph or set of equations, that are then verified or rejected through experimentation. We show how testable predictions can be derived automatically from sets of equations via the *causal ordering algorithm*, introduced by Simon (1953). We adapt and extend the algorithm to construct a *directed cluster graph* that we call the *causal ordering graph*. From this, we can construct a directed graph that we call the *Markov ordering graph*. We prove that, under a certain unique solvability assumption, the latter implies conditional independences between variables which can be tested in observational data and the former represents the effects of soft and certain perfect interventions which can be verified through experimentation. We believe that the technique of causal ordering is a useful and scalable tool in our search for and understanding of causal relations.

In this work, we also shed new light on differences between the causal ordering graph and the graph associated with a Structural Causal Model (SCM) (see Bongers, Forré, et al. (2020) and Pearl (2009)), which are also commonly referred to as Structural Equation Models (SEMs).[2] Specifically, we demonstrate that the two graphical representations may model different sets of interventions. Furthermore, we show that a stronger Markov property can sometimes be obtained by applying causal ordering to the structural equations of an SCM. By explicitly defining interventions and by distinguishing between the Markov ordering graph and the causal ordering graph we gain new insights about the correct interpretation of results in Dash (2005) and Iwasaki et al. (1994). Throughout this work, we discuss an example in Iwasaki et al.

[1] See, for example, (Dawid, 2010) and references therein for a discussion.

[2] The latter term has been used by econometricians since the 1950s. Note that, in the past some econometricians have used (cyclic/non-recursive) "structural models" without requiring that there is a specified one-to-one correspondence between endogenous variables and equations; see e.g. Basmann (1963). Recent usage is consistent with the implication that there is a specified variable on the left-hand side for each equation as is common in the SCM framework.

(1994) to illustrate our ideas. Here, we use it to highlight the contributions of this paper and to provide an overview of its central concepts.



(a) Bipartite graph.

(b) Causal ordering graph.

(c) Markov ordering graph.

Figure 4.1: Three graphical representations for the bathtub system in equilibrium. The bipartite graph in Figure (a) is a representation of the structure of equations $F = \{f_K, f_I, f_P, f_O, f_D\}$ where the vertices $V = \{v_K, v_I, v_P, v_O, v_D\}$ correspond to endogenous variables and there is an edge $(v - f)$ if and only if the variable v appears in equation f. The outcome of the causal ordering algorithm is the directed cluster graph in Figure (b), in which rectangles represent a partition of the variable and equation vertices into clusters. The corresponding Markov ordering graph for the variable vertices is given in Figure (c).

Example 4.1. Let us revisit a physical model of a filling bathtub in equilibrium that is presented in Iwasaki et al. (1994). Consider a system where water flows from a faucet into a bathtub at a constant rate X_{v_I} and it flows out of the tub through a drain with diameter X_{v_K}. An ensemble of such bathtubs that have faucets and drains with different (unknown) rates and diameters can be modelled by the equations f_K and f_I below:

$$f_K : \quad X_{v_K} = U_{w_K}, \tag{4.1}$$
$$f_I : \quad X_{v_I} = U_{w_I}, \tag{4.2}$$

where U_{w_K} and U_{w_I} are independent random variables both taking value in $\mathbb{R}_{>0}$. When the faucet is turned on the water level X_{v_D} in the bathtub increases as long as the inflow X_{v_I} of the water exceeds the outflow X_{v_O} of water. The differential equation $\dot{X}_{v_D}(t) = U_{w_1}(X_{v_I}(t) - X_{v_O}(t))$ defines the mechanism for the rate of change in $X_{v_D}(t)$, where U_{w_1} is a constant or a random variable taking value in $\mathbb{R}_{>0}$. At equilibrium the rate of change is equal to zero, resulting in the equilibrium equation

$$f_D : \quad U_{w_1}(X_{v_I} - X_{v_O}) = 0. \tag{4.3}$$

As the water level X_{v_D} increases, the pressure X_{v_P} that is exerted by the water increases as well. The mechanism for the change in pressure is defined by the differential equation $\dot{X}_{v_P}(t) = U_{w_2}(g\, U_{w_3} X_{v_D}(t) - X_{v_P}(t))$, where g is the gravitational constant and U_{w_2}, U_{w_3} are constants or random variables both taking value in $\mathbb{R}_{>0}$. After equilibration, we obtain

$$f_P : \qquad U_{w_2}(g\, U_{w_3} X_{v_D} - X_{v_P}) = 0. \tag{4.4}$$

The higher the pressure X_{v_P} or the bigger the size of the drain X_{v_K}, the faster the water flows through the drain. The differential equation $\dot{X}_{v_O}(t) = U_{w_4}(U_{w_5} X_{v_K} X_{v_P}(t) - X_{v_O}(t))$ models the outflow rate of the water, where U_{w_4}, U_{w_5} are constants or random variables both taking value in $\mathbb{R}_{>0}$. The equilibrium equation f_O is given by

$$f_O : \qquad U_{w_4}(U_{w_5} X_{v_K} X_{v_P} - X_{v_O}) = 0. \tag{4.5}$$

We will study the conditional independences that are implied by equilibrium equations (4.1) to (4.5). In Sections 4.5.1 and 4.5.2 we will define the notion of soft and perfect interventions on sets of equations as a generalization of soft and perfect interventions on SCMs. The causal properties of sets of equilibrium equations are examined by comparing the equilibrium distribution before and after an intervention. Our approach is related to the comparative statics analysis that is used in economics to study the change in equilibrium distribution after changing exogenous variables or parameters in the model, see also Simon and Iwasaki (1988). In this work, we will additionally consider the effects on the equilibrium distribution of perfect interventions targeting endogenous variables in the equilibrium equations.

Graphical representations. A set of equations can be represented by a bipartite graph. In the case of the filling bathtub, the structure of equilibrium equations (4.1) to (4.5) is represented by the bipartite graph in Figure 4.1(a). The set $V = \{v_K, v_I, v_P, v_O, v_D\}$ consists of vertices that correspond to variables and the vertices in the set $F = \{f_K, f_I, f_P, f_O, f_D\}$ correspond to equations. There is an edge between a variable vertex v_i and an equation vertex f_j if the variable labelled v_i appears in the equation with label f_j. A formal definition of a system of constraints and its associated bipartite graph will be provided in Section 4.1.1. The causal ordering algorithm, introduced by Simon (1953) and reformulated by us in Section 4.2, takes a *self-contained* bipartite graph as input and returns a *causal ordering graph*. A causal ordering graph is a *directed cluster graph* which consists of variable vertices v_i and equation vertices f_j that are partitioned into clusters. Directed edges go from variable vertices to clusters. For the filling bathtub, the causal ordering graph is given in Figure 4.1(b). In Section 4.4 we will show how the *Markov ordering graph* can be constructed from a causal ordering graph. For the equilibrium equations of the filling bathtub, the Markov ordering graph is given in Figure 4.1(c). The causal ordering algorithm of Simon (1953) can only be applied to bipartite graphs that have

the property that they are *self-contained*. In Section 4.3 we introduce an extended causal ordering algorithm that can also be applied to bipartite graphs that are not self-contained.

Markov property. The Markov ordering graph in Figure 4.1(c) encodes conditional independences between the equilibrium solutions X_{v_K}, X_{v_I}, X_{v_P}, X_{v_O}, and X_{v_D} of the equilibrium equations. In particular, d-separations between variable vertices in the Markov ordering graph imply conditional independences between the corresponding variables under certain solvability conditions, as we will prove in Theorem 4.3 in Section 4.4. In Figure 4.1(c), the variable vertices v_I and v_D are d-separated by v_O. It follows that at equilibrium the inflow rate X_{v_I} and the water level X_{v_D} are independent given the outflow rate X_{v_O}. In Sections 4.3 and 4.4.4 we show how we can use *a perfect matching* for a bipartite graph to construct a directed graph that implies conditional independences between variables via σ-separations.[3]

Soft interventions. The causal ordering graph in Figure 4.1(b) encodes the effects of *soft* interventions targeting (equilibrium) *equations*. This type of intervention is often also referred to as a mechanism change. We assume that the variables in each cluster can be solved uniquely from the equations in their cluster both before and after the intervention.[4] A soft intervention has no effect on a variable if there is no directed path from the intervention target to the cluster containing the variable, as we will prove in Theorem 4.5 in Section 4.5.1. Consider an experiment where the value of the gravitational constant g is altered (e.g. by moving the bathtub to the moon) resulting in an alteration of the equation f_P. This is a soft intervention on f_P. There is no directed path from f_P to clusters that contain the vertices $\{v_K, v_I, v_P, v_O\}$ in the causal ordering graph in Figure 4.1(b). Since the conditions of Theorem 4.5 are satisfied, the soft intervention on f_P has no effect on $\{X_{v_K}, X_{v_I}, X_{v_P}, X_{v_O}\}$ but it may have an effect on X_{v_D} (depending on the precise functional form of the equations and the values of the parameters).

Perfect interventions. The causal ordering graph in Figure 4.1(b) also encodes the effects of *perfect* interventions on *clusters*, under the assumption that variables can be solved uniquely from the equations of their clusters in the causal ordering graph before and after intervention. We will formally prove this in Theorem 4.6 in Section 4.5.2. Consider a perfect intervention on the cluster $\{f_K, v_K\}$ (i.e. fixing the diameter X_{v_K} of the drain by altering the equation f_K) in Figure 4.1(b). This intervention

[3] Forré et al. (2017) introduced the notion of σ-separations to replace d-separations in directed graphs that may contain cycles. See Section 1.2 in Chapter 1 for more details.

[4] For the underlying dynamical model this assumption means that we assume that the equations of the model define a unique equilibrium to which the system converges and that the system also converges to a unique equilibrium that is defined by the model equations after an intervention on one of the parameters or exogenous variables in the model. For some dynamical systems (e.g., a biochemical reaction network) extra equations are required that describe the dependence of the equilibrium distribution on initial conditions and cannot be modelled in the standard SCM framework (Blom, Bongers, et al., 2019). For these systems, Blom, Bongers, et al. (2019) introduced the more general class of Causal Constraints Models.

generically changes the solution for $\{X_{v_K}, X_{v_P}, X_{v_D}\}$ because v_K is targeted by the intervention and there are directed paths from the cluster of v_K to the clusters of v_P and v_D. It has no effect on $\{X_{v_I}, X_{v_O}\}$ because there are no directed paths from the cluster of v_K to the clusters of v_I and v_O. △

4.1.1 System of constraints

Our formal treatment of sets of equations mirrors the definition of a structural causal model in the sense that we separate the model from the endogenous random variables that solve it. An introduction to cyclic SCMs was provided in Section 1.2.1). The related graph terminology and Markov properties can be found in Section 4.A.1 and 4.A.2. Here, we introduce a mathematical object that we call a *system of constraints* to represent equations and their structure as a bipartite graph.

Definition 4.1. A *system of constraints* is a tuple $\langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} = \langle V, F, E \rangle \rangle$ where

(i) $\mathfrak{X} = \bigotimes_{v \in V} \mathcal{X}_v$, where each $\mathcal{X}_v$ is a standard measurable space and the domain of a variable X_v,

(ii) $\mathbf{X}_W = (X_w)_{w \in W}$ is a family of independent random variables taking value in $\mathfrak{X}_W$ with $W \subseteq V$ a set of indices corresponding to exogenous variables,[5]

(iii) $\boldsymbol{\Phi} = (\Phi_f)_{f \in F}$ is a family of constraints, each of which is a tuple $\Phi_f = \langle \phi_f, c_f, V(f) \rangle$, with:
   (i) $V(f) \subseteq V$
   (ii) c_f a constant taking value in a standard measurable space $\mathcal{Y}_f$,
   (iii) $\phi_f : \mathfrak{X}_{V(f)} \to \mathcal{Y}_f$ a measurable function,

(iv) $\mathcal{B} = \langle V, F, E \rangle$ is a bipartite graph with:
   (i) V a set of nodes corresponding to variables,
   (ii) F a set of nodes corresponding to constraints,
   (iii) $E = \{(f - v) : f \in F, v \in V(f)\}$ a set of edges.

Henceforth we will use the terms 'variables' and 'vertices corresponding to variables' interchangeably. We will also use the terms 'constraints', 'equations', and 'vertices corresponding to constraints' interchangeably. We will often refer to the bipartite graph in a system of constraints as the 'associated bipartite graph'. A constraint is formally defined as a triple consisting of a measurable function, a constant, and a subset of the variables. For the sake of convenience we will often write constraints as equations instead. Note that the notation for adjacencies in the associated bipartite graph is equivalent to the notation for the variables that belong to a constraint: $V(f) = \mathrm{adj}_{\mathcal{B}}(f)$. For a set $S_F \subseteq F$, we will let $\mathrm{adj}_{\mathcal{B}}(S_F) = V(S_F) = \cup_{f \in S_F} V(f)$ denote the adjacencies of the vertices $f \in S_F$.

When modelling some system with a system of constraints, we are implicitly assuming that the constraints are *reversible* in the sense that the causal relations

[5] This means that the nodes $V \setminus W$ correspond to endogenous variables.

between the endogenous variables are flexible and may depend in principle on the entire set of constraints in the system. However, there is an important modelling choice regarding which of the variables to consider as *endogenous* ("internal" to the system) and which variables to consider as *exogenous* ("external" to the system). The implicit assumption here is that *the endogenous variables cannot cause the exogenous variables*. This is the (only) causal "background knowledge" that is expressed formally by a system of constraints. As Simon (1953) showed, and as we will explicate in later sections, the causal relations between the endogenous variables can then be obtained by applying Simon's causal ordering algorithm.

Example 4.2. Consider two variables: the temperature in a room (X_1) and the reading of a thermometer in the same room (X_2). One can think of different systems of constraints to model these variables. One possibility is the single constraint ($X_1 - X_2 = 0$) in which both X_1 and X_2 are considered to be endogenous variables. As it turns out, we will then not be able to draw any conclusion regarding the causal relation between X_1 and X_2. Another possibility would be to use the same constraint, but now considering X_1 to be exogenous and X_2 to be endogenous. Then, one will find that X_1 causes X_2, but not vice versa, which may appear to be a realistic model. Yet another possibility with the same constraint would be to consider X_2 to be the exogenous variable and X_1 to be endogenous. This model would be considered less realistic in most situations (except perhaps in somewhat unnatural settings where the thermometer would be broken, but its reading would be used by some agent to adjust the heating in order to control the room temperature).

Thus, the constraint $X_1 - X_2 = 0$ on its own does not lead to any conclusions regarding the causal relations between variables X_1 and X_2; it is only through the additional background knowledge (represented by the distinction between endogenous and exogenous variables) that the causal directionality is fixed. In cases with more than one endogenous variable (like in Example 4.1), the causal ordering algorithm can be used to "propagate" the causal directionality from exogenous to endogenous variables. △

4.1.2 Related work and contributions

Graphical models are a popular statistical tool to model probabilistic aspects of complex systems. They represent a set of conditional independences between random variables that correspond to vertices which allows us to learn their graphical structure from data (Lauritzen, 1996). These models are often interpreted causally, so that directed edges between vertices are interpreted as direct causal relations between corresponding variables (Pearl, 2009). The strong assumptions that are necessary for this viewpoint have been the topic of debate (Dawid, 2010). This work contributes to this discussion by revisiting an example in Iwasaki et al. (1994), and discussing how it

seems that, in this case, the presence of vertices representing equations is required to simultaneously express both conditional independences and the effects of interventions in a single graph.

Throughout this work, we discuss the application of the causal ordering algorithm to the equilibrium equations of the bathtub model that we discussed in Example 4.1. In the literature, feedback processes that have reached equilibrium have been represented by e.g. chain graphs (Lauritzen and Richardson, 2002) and cyclic directed graphs (Bongers and Mooij, 2018; Mooij, Janzing, and Schölkopf, 2013; Spirtes and Richardson, 1995). For the latter it was shown that they imply conditional independences in the equilibrium distribution via the d-separation criterion in the linear or discrete case (Forré et al., 2017) but that the directed global Markov property may fail if the underlying model is neither linear nor discrete (Spirtes and Richardson, 1995). The alternative criterion that Spirtes and Richardson (1995) formulated for the "collapsed graph" was recently reformulated in terms of *σ-separations* and shown to hold in very general settings (Forré et al., 2017). Constraint-based causal discovery algorithms for the cyclic setting under various assumptions are given in Forré et al. (2018), Mooij and Claassen (2020), Mooij, Magliacane, et al. (2020), Richardson (1996), and Strobl (2018). The causal properties of dynamical systems at equilibrium were previously studied by Blom, Bongers, et al. (2019), Bongers and Mooij (2018), F. M. Fisher (1970), Hyttinen et al. (2012), Lauritzen and Richardson (2002), Mooij, Janzing, Heskes, et al. (2011), and Mooij, Janzing, and Schölkopf (2013) (see also Chapter 3), who consider graphical and causal models that arise from studying the stationary behaviour of dynamical models. For the deterministic case, Mooij, Janzing, and Schölkopf (2013) propose to map first-order differential equations to labelled equilibrium equations and then to the structural equations of an SCM. This idea was recently generalized to the stochastic case and higher order differential equations (Bongers and Mooij, 2018). For certain systems, such as the bathtub model in Example 4.1, this construction may lead to a cyclic SCM with self-cycles (Bongers and Mooij, 2018). The causal and conditional independence properties of cyclic SCMs (possibly with self-cycles) have been studied by Bongers, Forré, et al. (2020). In other work assumptions on the underlying dynamical model have been made to avoid the complexities of SCMs with self-cycles. Here, we will consider potential benefits (e.g. obtaining a stronger Markov property) of applying the technique of causal ordering to the structural equations of the cyclic SCM for the equilibrium equations of dynamical systems such as the bathtub system.

Our work generalizes the causal ordering algorithm which was introduced by Simon (1953). Following Dash and Druzdzel (2008), we formally prove that the causal ordering graph that is constructed by the algorithm is unique. One of the novelties of this work is that we also prove that it encodes the effects of soft and certain perfect interventions and, moreover, we show how it can be used to construct a DAG that implies conditional independences via the d-separation criterion. There also exists a different, computationally more efficient, algorithm for causal ordering (Gonçalves

et al., 2016; Nayak, 1995). We formally prove that this algorithm is equivalent to the one in Simon (1953). This approach motivates an alternative representation of the system as a directed graph that may contain cycles. We prove that the *generalized directed global Markov property*, as formulated by Forré et al. (2017), holds for this graphical representation. Using methods to determine the upper-triangular form of a matrix in Pothen and Fan (1990), we further extend the causal ordering algorithm so that it can be applied to any bipartite graph.

In Section 4.6 we will present a detailed discussion of how our work relates to that of Bongers, Forré, et al. (2020), Bongers and Mooij (2018), Dash (2005), and Iwasaki et al. (1994). We show that what Iwasaki et al. (1994) call the "causal graph" coincides with the Markov ordering graph in our work. We take a closer look at the intricacies of (possible) causal implications of the Markov ordering graph and notice that it neither represents the effects of soft interventions nor does it have a straightforward interpretation in terms of perfect interventions. Because Simon and Iwasaki (1988) assume that a one-to-one correspondence between variables and equations is known in advance, they can use the Markov ordering graph to read off the effects of soft interventions. We argue that the causal ordering graph, and not the Markov ordering graph, should be used to represent causal relations when the matching between variables and equations is not known before-hand. This sheds some new light on the work of Dash (2005) on (causal) structure learning and equilibration in dynamical systems. We further discuss the advantages and disadvantages of our causal ordering approach compared to the SCM framework.

4.2 Causal ordering

In this section, we adapt the causal ordering algorithm of Simon (1953), rephrase it in terms of *self-contained* bipartite graphs, and define the output of the algorithm as a *directed cluster graph*.[6] We then prove that Simon's causal ordering algorithm is well-defined and has a unique output.

Definition 4.2. A *directed cluster graph* is an ordered pair $\langle \mathcal{V}, \mathcal{E} \rangle$, where $\mathcal{V}$ is a partition $V^{(1)}, V^{(2)}, \dots, V^{(n)}$ of a set of vertices V and $\mathcal{E}$ is a set of directed edges $v \to V^{(i)}$ with $v \in V$ and $V^{(i)} \in \mathcal{V}$. For $x \in V$ we let $\mathrm{cl}(x)$ denote the cluster in $\mathcal{V}$ that contains x. We say that there is a *directed* path from $x \in V$ to $y \in V$ if either $\mathrm{cl}(x) = \mathrm{cl}(y)$ or there is a sequence of clusters $V_1 = \mathrm{cl}(x), V_2, \dots, V_{k-1}, V_k = \mathrm{cl}(y)$ so that for all $i \in \{1, \dots, k-1\}$ there is a vertex $z_i \in V_i$ such that $(z_i \to V_{i+1}) \in \mathcal{E}$.

[6] The notion of a directed cluster graph corresponds to the box representation of a collapsed graph in Richardson (1996), Chapter 4.

4.2.1 Self-contained bipartite graphs

The causal ordering algorithm in Simon (1953) is presented in terms of a *self-contained* set of equations and variables that appear in them. For bipartite graphs, the notion of *self-containedness* corresponds to the conditions in Definition 4.3.

Definition 4.3. Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph. A subset $F' \subseteq F$ is said to be *self-contained* if

(i) $|F'| = |\mathrm{adj}_{\mathcal{B}}(F')|$,

(ii) $|F''| \leq |\mathrm{adj}_{\mathcal{B}}(F'')|$ for all $F'' \subseteq F'$.[7]

The bipartite graph $\mathcal{B}$ is said to be *self-contained* if $|F| = |V|$ and F is self-contained. A non-empty self-contained set $F' \subseteq F$ is said to be a *minimal self-contained set*[8] if all its non-empty strict subsets are not self-contained.

Example 4.3. In Figure 4.2 a bipartite graph is shown with self-contained sets

$$\{f_1\}, \{f_1, f_2, f_3, f_4\}, \{f_1, f_2, f_3, f_4, f_5\}$$

where $\{f_1\}$ is a minimal self-contained set. Since the set $\{f_1, f_2, f_3, f_4, f_5\}$ is self-contained and $|V| = |F| = 5$, we say that this bipartite graph is self-contained. △

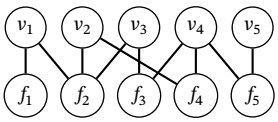


Figure 4.2: A self-contained bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ with $V = \{v_1, v_2, v_3, v_4, v_5\}$ and $F = \{f_1, f_2, f_3, f_4, f_5\}$. The sets $\{f_1\}, \{f_1, f_2, f_3, f_4\}$, and $\{f_1, f_2, f_3, f_4, f_5\}$ are self-contained, and $\{f_1\}$ is the only minimal self-contained set.

Sets of equations that model systems in the real world often include both *endogenous* and *exogenous* variables. The distinction is that exogenous variables are assumed to be determined outside the system and function as inputs to the model, whereas the endogenous variables are part of the system. The following example illustrates that the associated bipartite graph for a set of equations with both endogenous and exogenous variables is usually not self-contained.

Example 4.4. Let $V = \{v_1, v_2, w_1, w_2\}$ be an index set for endogenous and exogenous variables $\mathbf{X} = (X_i)_{i\in V}$, $W = \{w_1, w_2\}$ a subset that is an index set for exogenous variables only, and $F = \{f_1, f_2\}$ an index set for equations:

$$\Phi_{f_1} : \qquad X_{v_1} - X_{w_1} = 0,$$

[7] This condition is also called the Hall Property (M. Hall, 1986).

[8] In this case the Strong Hall Property holds, that is $|F''| < |\mathrm{adj}_{\mathcal{B}}(F'')|$ for all $\emptyset \subsetneq F'' \subsetneq F'$ (M. Hall, 1986).

$$\Phi_{f_2}: \qquad X_{v_2} - X_{v_1} - X_{w_2} = 0.$$

The associated bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ is given in Figure 4.3(a). It has vertices V that correspond to both endogenous variables X_{v_1}, X_{v_2} and exogenous variables X_{w_1}, X_{w_2}. The vertices F correspond to constraints Φ_{f_1} and Φ_{f_2}. Edges between vertices $v \in V$ and $f \in F$ are present whenever $v \in V(f)$ (i.e. when the variable X_v appears in the constraint Φ_f). Since $|V| \neq |F|$, the associated bipartite graph is not self-contained. △

(a) Bipartite graph. (b) Directed cluster graph.

Figure 4.3: The bipartite graph in Figure (a) is associated with the constraints in Example 4.4. Exogenous variables are indicated by dashed circles. The directed cluster graph that is obtained by applying Algorithm 1 is shown in Figure (b).

4.2.2 Causal ordering algorithm

The causal ordering algorithm, as formulated by Simon (1953), has as input a self-contained set of equations and as output it has an ordering on clusters of variables that appear in these equations. We reformulate the algorithm in terms of bipartite graphs and minimal self-contained sets. The input of the algorithm is then a self-contained bipartite graph and its output a directed cluster graph that we call the *causal ordering graph*.

The causal ordering algorithm below has been adapted for systems of constraints with exogenous variables. The input is a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ and a set of vertices $W \subseteq V$ (corresponding to exogenous variables) such that the subgraph $\mathcal{B}' = \langle V', F', E'\rangle$ induced by $(V \setminus W) \cup F$ is self-contained. The algorithm starts out by adding the exogenous vertices as singleton clusters to a cluster set $\mathcal{V}$ during an initialization step. Subsequently, the algorithm searches for a minimal self-contained set $S_F \subseteq F$ in $\mathcal{B}'$. Together with the set of adjacent variable vertices $S_V = \mathrm{adj}_{\mathcal{B}'}(S_F)$ a cluster $S_F \cup S_V$ is formed and added to $\mathcal{V}$. For each $v \in V$, an edge $(v \to (S_F \cup S_V))$ is added to $\mathcal{E}$ if $v \notin S_V$ and $v \in \mathrm{adj}_{\mathcal{B}}(S_F)$. In other words, the cluster has an incoming edge from each variable vertex that is adjacent to the cluster but not in it. These steps are then repeated for the subgraph induced by the vertices $(V' \cup F') \setminus (S_V \cup S_F)$ that are not in the cluster, as long as this is not the null graph. The order in which the self-contained sets are obtained is represented by one of the topological orderings of the clusters in the causal ordering graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E}\rangle$.

Algorithm 1: Causal ordering using minimal self-contained sets.

Input: a set of exogenous vertices W, a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ such that its subgraph induced by $(V \setminus W) \cup F$ is self-contained
Output: directed cluster graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E}\rangle$

```
E ← ∅                                                  // initialization
V ← {{w} : w ∈ W}                                      // initialization
B' ← ⟨V', F', E'⟩ subgraph induced by (V \ W) ∪ F       // initialization
while B' is not the null graph do
    S_F ← a minimal self-contained set of F'
    C ← S_F ∪ adj_{B'}(S_F)                            // construct cluster
    V ← V ∪ {C}                                        // add cluster
    for v ∈ adj_B(S_F) \ adj_{B'}(S_F) do
        E ← E ∪ {(v → C)}                              // add edges to cluster
    B' ← subgraph of B' induced by (V' ∪ F') \ C       // remove cluster
```

Theorem 4.1 shows that the output of causal ordering via minimal self-contained sets is well-defined and unique.

Theorem 4.1. *The output of Algorithm 1 is well-defined and unique.*

The following example shows how the causal ordering algorithm works on the self-contained bipartite graph in Figure 4.2 and the bipartite graph in Figure 4.3(a).

Example 4.5. Consider the set of equations in Example 4.4 and its associated bipartite graph in Figure 4.3(a). The subgraph induced by the endogenous variables v_1, v_2 and the constraints f_1, f_2 is self-contained. We initialize Algorithm 1 with $\mathcal{E}$ the empty set, $\mathcal{V} = \{\{w_1\}, \{w_2\}\}$, and $\mathcal{B}'$ the subgraph induced by $\{v_1, v_2, f_1, f_2\}$. We then first find the minimal self-contained set $\{f_1\}$. Its adjacencies are $\{v_1\}$ in $\mathcal{B}'$ and $\{v_1, w_1\}$ in $\mathcal{B}$. We add $\{v_1, f_1\}$ to $\mathcal{V}$ and add the edge $(w_1 \to \{v_1, f_1\})$ to $\mathcal{E}$. Finally, we add $\{v_2, f_2\}$ to $\mathcal{V}$ and the edges $(v_1 \to \{v_2, f_2\})$ and $(w_2 \to \{v_2, f_2\})$ to $\mathcal{E}$. The output of the causal ordering algorithm is the directed cluster graph in Figure 4.3(b). This reflects how one would solve the system of equations Φ_{f_1}, Φ_{f_2} with respect to X_{v_1}, X_{v_2} in terms of X_{w_1}, X_{w_2} by hand. △

4.3 Extending the causal ordering algorithm

In this section we present an adaptation of an alternative, computationally less expensive, algorithm for causal ordering which uses perfect matchings instead of minimal self-contained sets, similar to the algorithm suggested by Nayak (1995). Gonçalves et al. (2016) proved that Simon's classic algorithm makes use of a subroutine that solves an NP-hard problem, whereas the computational complexity of

Nayak's algorithm is bounded by $\mathcal{O}(|V|\,|E|)$, where $|V|$ is the number of nodes and $|E|$ is the number of edges in the bipartite graph. Here, we provide a proof for the fact that causal ordering via minimal self-contained sets is equivalent to causal ordering via perfect matchings. There are many systems of equations with a unique solution that consist of more equations than there are endogenous variables, most notably in the case of non-linear equations, or in the presence of cycles. In that case the bipartite graph associated with these equations may not be self-contained. In this section, we show how Nayak's algorithm can be extended using maximum matchings so that it can be applied to any bipartite graph.

4.3.1 Causal ordering via perfect matchings

Given a bipartite graph $\mathcal{B}$, the *associated directed graph* can be constructed from a matching $\mathcal{M}$ by *orienting edges*. A directed cluster graph can then be constructed via the operations that *construct clusters* and *merge clusters* in Definition 4.4 below.

Definition 4.4. Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph and $\mathcal{M}$ a perfect matching for $\mathcal{B}$.

(i) *Orient edges*: For each $(v - f) \in E$ the edge set E_{dir} has an edge $(v \leftarrow f)$ if $(v - f) \in \mathcal{M}$ and an edge $(v \rightarrow f)$ if $(v - f) \notin \mathcal{M}$. E_{dir} has no additional edges. The *associated directed graph* is $\mathcal{G}(\mathcal{B}, \mathcal{M}) = \langle V \cup F, E_{\text{dir}}\rangle$.

(ii) *Construct clusters*: Let $\mathcal{V}'$ be a partition of vertices $V \cup F$ into strongly connected components in $\mathcal{G}(\mathcal{B}, \mathcal{M})$. For each $(x \rightarrow w) \in E_{\text{dir}}$ the edge set $\mathcal{E}'$ has an edge $(x \rightarrow \text{cl}(w))$ if $x \notin \text{cl}(w)$, where $\text{cl}(w) \in \mathcal{V}'$ is the strongly connected component of w in $\mathcal{G}(\mathcal{B}, \mathcal{M})$. The edge set $\mathcal{E}'$ has no additional edges. The *associated clustered graph* is $\text{clust}(\mathcal{G}(\mathcal{B}, \mathcal{M})) = \langle \mathcal{V}', \mathcal{E}'\rangle$.

(iii) *Merge clusters*: Let $\mathcal{V} = \{S \cup \mathcal{M}(S) : S \in \mathcal{V}'\}$. For each $(x \rightarrow S) \in \mathcal{E}'$ with $x \notin \mathcal{M}(S)$ the edge set $\mathcal{E}$ contains an edge $(x \rightarrow S \cup M(S))$. The edge set $\mathcal{E}$ has no additional edges. The *associated clustered and merged graph* is $\text{merge}(\text{clust}(\mathcal{G}(\mathcal{B}, \mathcal{M}))) = \langle \mathcal{V}, \mathcal{E}\rangle$.[9]

For *causal ordering via perfect matching* we require as input a set of exogenous vertices W and a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$, for which the subgraph $\mathcal{B}'$ induced by the vertices $(V \cup F) \setminus W$ is self-contained. The output is a directed cluster graph. The details can be found in Algorithm 2. We see that the algorithm starts out by finding a perfect matching[10] for $\mathcal{B}'$,[11] which is then used to orient edges in the bipartite graph $\mathcal{B}$. The algorithm then proceeds by partitioning vertices in the resulting directed graph into strongly connected components to construct the associated

[9] In Theorem 4.2 we will show that this is the causal ordering graph $\text{CO}(\mathcal{B})$.

[10] Note that a bipartite graph has a perfect matching if and only if it is self-contained (M. Hall, 1986). See also Theorem 4.7 and Corollary 4.2 in Appendix 4.B.4.

[11] The Hopcraft-Karp-Karzanov algorithm, which runs in $\mathcal{O}(|E|\sqrt{|V \cup F|})$, can be used to find a perfect matching (Hopcroft et al., 1973; Karzanov, 1973).

clustered graph.[12] Finally, the merge operation is applied to construct the causal ordering graph. Theorem 4.2 below shows that causal ordering via perfect matchings is equivalent to causal ordering via minimal self-contained sets.

Algorithm 2: Causal ordering via perfect matchings.

Input: a set of exogenous vertices W, a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ such that the subgraph induced by $(V \cup F) \setminus W$ is self-contained
Output: directed cluster graph $\langle \mathcal{V}, \mathcal{E}\rangle$
$\mathcal{B}' \leftarrow$ subgraph induced by $(V \setminus W) \cup F$ // initialization
$\mathcal{M} \leftarrow$ perfect matching for $\mathcal{B}'$ // initialization
$E_{\text{dir}} \leftarrow \emptyset$ // orient edges
for $(v - f) \in E$ **with** $f \in F$ **do**
  if $(v - f) \in \mathcal{M}$ **then**
    Add $(v \leftarrow f)$ to E_{dir}
  else
    Add $(v \rightarrow f)$ to E_{dir}
$\mathcal{V}' \leftarrow$ strongly connected components of $\langle V \cup F, E_{\text{dir}}\rangle$ // clustering
$\mathcal{E}' \leftarrow \emptyset$
for $(x \rightarrow w) \in E_{\text{dir}}$ **do**
  for $S \in \mathcal{V}'$ **do**
    if $w \in S$ **and** $x \notin S$ **then**
      Add $(x \rightarrow S)$ to $\mathcal{E}'$
$\mathcal{V}, \mathcal{E} \leftarrow \emptyset$ // merge clusters
for $S \in \mathcal{V}'$ **do**
  Add $S \cup M(S)$ to $\mathcal{V}$
  for $(x \rightarrow S) \in \mathcal{E}'$ **do**
    if $x \notin M(S)$ **then**
      Add $(x \rightarrow S \cup M(S))$ to $\mathcal{E}$

Theorem 4.2. *The output of Algorithm 2 coincides with the output of Algorithm 1.*

The following example illustrates that the output of causal ordering via perfect matchings does not depend on the choice of perfect matching and coincides with the output of Algorithm 1.

Example 4.6. Consider the bipartite graph $\mathcal{B}$ in Figure 4.4(a). The subgraph induced by the vertices $V = \{v_1, \dots, v_5\}$ and $F = \{f_1, \dots, f_5\}$ is the self-contained bipartite graph in Figure 4.2. We will follow the steps in both Algorithm 1 and 2 to construct the causal ordering graph.

[12] Tarjan's algorithm, which runs in $\mathcal{O}(|V| + |E|)$ time, can be used to find the strongly connected components in a directed graph (Tarjan, 1972).

For causal ordering with minimal self-contained sets we first add the exogenous variables to the cluster set $\mathcal{V}$ as the singleton clusters $\{w_1\}, \{w_2\}, \{w_3\}, \{w_4\}, \{w_5\}$, and $\{w_6\}$. The only minimal self-contained set in the subgraph induced by the vertices $V = \{v_1, \ldots, v_5\}$ and $F = \{f_1, \ldots, f_5\}$ is $\{f_1\}$. Since f_1 is adjacent to v_1 we add $C_1 = \{v_1, f_1\}$ to $\mathcal{V}$. Since f_1 is adjacent to w_1 in $\mathcal{B}$ we add $(w_1 \to C_1)$ to $\mathcal{E}$. The subgraph $\mathcal{B}' = \langle V', F', E' \rangle$ induced by the remaining nodes $V' = \{v_2, v_3, v_4, v_5\}$ and $F' = \{f_2, f_3, f_4, f_5\}$ has $\{f_2, f_3, f_4\}$ as its only minimal self-contained set. Since the set $\{f_2, f_3, f_4\}$ is adjacent to $\{v_2, v_3, v_4\}$ in $\mathcal{B}'$, we add $C_2 = \{v_2, v_3, v_4, f_2, f_3, f_4\}$ to $\mathcal{V}$. Since v_1, w_2, w_3, w_4, and w_5 are adjacent to $\{f_2, f_3, f_4\}$ in $\mathcal{B}$ but not part of C_2, we add the edges $(v_1 \to C_2)$, $(w_2 \to C_2)$, $(w_3 \to C_2)$, $(w_4 \to C_2)$, and $(w_5 \to C_2)$ to $\mathcal{E}$. The subgraph induced by the remaining nodes v_5 and f_5 has $\{f_5\}$ as its minimal self-contained subset. We add $C_3 = \{v_5, f_5\}$ to $\mathcal{V}$ and the edges $(v_4 \to C_3)$ and $(w_6 \to C_3)$ to $\mathcal{E}$. The directed cluster graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ is given in Figure 4.4(e).

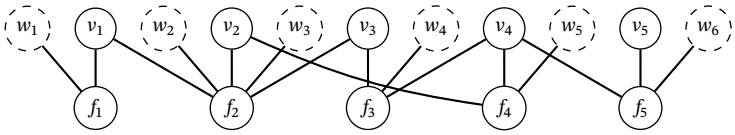
For causal ordering via perfect matchings, we consider the following two perfect matchings of the self-contained bipartite graph in Figure 4.2:

$$\mathcal{M}_1 = \{(v_1 - f_1), (v_2 - f_2), (v_3 - f_3), (v_4 - f_4), (v_5 - f_5)\},$$
$$\mathcal{M}_2 = \{(v_1 - f_1), (v_2 - f_4), (v_3 - f_2), (v_4 - f_3), (v_5 - f_5)\}.$$

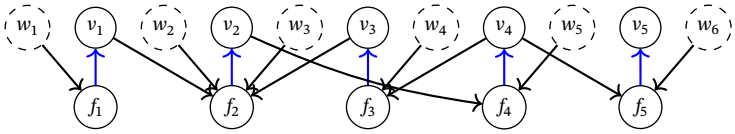
We use these one-to-one correspondences between endogenous variable vertices and constraint vertices in the orientation step in Definition 4.4 to obtain the associated directed graphs $\mathcal{G}(\mathcal{B}, \mathcal{M}_1)$ and $\mathcal{G}(\mathcal{B}, \mathcal{M}_2)$ in Figures 4.4(b) and 4.4(c) respectively. Application of the clustering step in Definition 4.4 to either $\mathcal{G}(\mathcal{B}, \mathcal{M}_1)$ or $\mathcal{G}(\mathcal{B}, \mathcal{M}_2)$ results in the clustered graph $\mathrm{clust}(\mathcal{G}(\mathcal{B}, \mathcal{M}_2)) = \mathrm{clust}(\mathcal{G}(\mathcal{B}, \mathcal{M}_1))$ in Figure 4.4(d). The final step is to merge clusters in this directed cluster graph. We find that the causal ordering graph $\mathrm{merge}(\mathrm{clust}(\mathcal{G}(\mathcal{B}, \mathcal{M}_1))) = \mathrm{merge}(\mathrm{clust}(\mathcal{G}(\mathcal{B}, \mathcal{M}_2)))$ in Figure 4.4(e) does not depend on the choice of perfect matching, as is implied by Theorem 4.2. Note that the output of causal ordering with minimal self-contained sets coincides with the output of causal ordering via perfect matchings. △

4.3.2 Coarse decomposition via maximum matchings

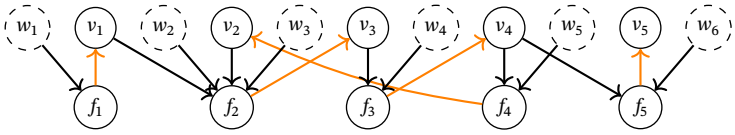
The extension that we propose relies on the *coarse decomposition* of bipartite graphs in Pothen and Fan (1990), which was originally proposed by Dulmage et al. (1958). The main idea is that a set of equations (i.e. a system of constraints) can be divided into an *incomplete* part that has fewer equations than variables, an *over-complete* part that has more equations than variables, and a part that is self-contained. The *coarse decomposition* in Definition 4.5 below uses the notions of a *maximum matching* and an *alternating path* for a maximum matching. The former is a matching so that there are no matchings with a greater cardinality, while the latter is a sequence of distinct vertices and edges $(v_1, e_1, v_2, e_2, \ldots, e_{n-1}, v_n)$ so that edges e_i are alternately in and out



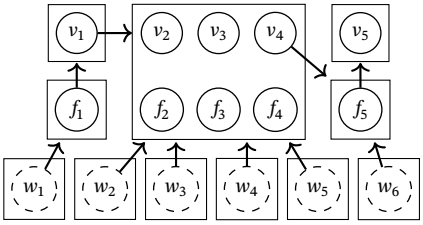
(a) Bipartite graph $\mathcal{B}$ where dashed vertices indicate exogenous variables.



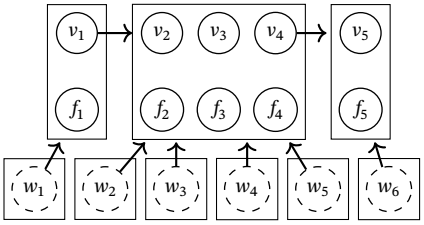
(b) Associated directed graph $\mathcal{G}(\mathcal{B}, \mathcal{M}_1)$.



(c) Associated directed graph $\mathcal{G}(\mathcal{B}, \mathcal{M}_2)$.



(d) Clustered graph clust($\mathcal{G}(\mathcal{B}, \mathcal{M}_1)$).



(e) Causal ordering graph CO($\mathcal{B}$).

Figure 4.4: Causal ordering with two different perfect matchings M_1 and M_2 applied to the bipartite graph in Figure (a). The results of subsequently orienting edges, constructing clusters, and merging clusters as in Definition 4.4 are given in Figures (b) to (e). The edges in M_1 that are oriented from variables to equations in Figure (b) are indicated with blue edges. Likewise, edges in M_2 are indicated with orange edges in Figure (c). The clustered graph in Figure (d) coincides with clust($\mathcal{G}(\mathcal{B}, \mathcal{M}_2)$) and for the causal ordering graph in Figure (e) we have that CO($\mathcal{B}$) = merge(clust($\mathcal{G}(\mathcal{B}, \mathcal{M}_1)$)) = merge(clust($\mathcal{G}(\mathcal{B}, \mathcal{M}_2)$)).

a maximum matching M. Proposition 4.1 by Pothen (1985) shows that the coarse decomposition is unique.[13] In this section we loosely follow the exposition of the coarse decomposition in Van Diepen (2019) and Pothen and Fan (1990).

Definition 4.5. Let M be a maximum matching for a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ and let V_{un} and F_{un} denote the unmatched vertices in V and F respectively. The *incomplete set* $T_I \subseteq V \cup F$ and *overcomplete set* $T_O \subseteq V \cup F$ are given by:

$$T_I := \{x \in V \cup F : \text{ there is an alternating path between } x \text{ and some } y \in V_{\text{un}}\},$$
$$T_O := \{x \in V \cup F : \text{ there is an alternating path between } x \text{ and some } y \in F_{\text{un}}\}.$$

The *complete set* is given by $T_C = V \cup F \setminus (T_I \cup T_O)$. The *coarse decomposition* $\text{CD}(\mathcal{B}, M)$ is given by $\langle T_I, T_C, T_O\rangle$. The *incomplete graph* $\mathcal{B}_I$ is the subgraph of $\mathcal{B}$ induced by vertices T_I, the *complete graph* $\mathcal{B}_C$ is the subgraph of $\mathcal{B}$ induced by vertices T_C, and the *overcomplete graph* $\mathcal{B}_O$ is the subgraph of $\mathcal{B}$ induced by vertices T_O.

Note that T_I and T_O are necessarily disjoint, for more details see Lemma 4.8 in Appendix 4.B.2.

Proposition 4.1. *[Pothen (1985)] The coarse decomposition of a bipartite graph $\mathcal{B}$ is independent of the choice of the maximum matching.*

There exist fast algorithms that are able to find a maximum matching in a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$, such as the Hopcraft-Karp-Karzanov algorithm, which runs in $\mathcal{O}(|E|\sqrt{|V \cup F|})$ time (Hopcroft et al., 1973; Karzanov, 1973). In the following example we manually searched for maximum matchings to illustrate the result in Proposition 4.1 that the coarse decomposition is unique.

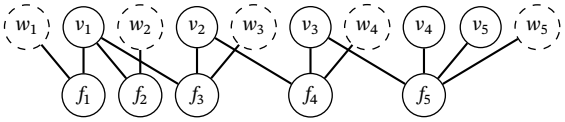
Example 4.7. Consider the bipartite graph $\mathcal{B}'$ in Figure 4.5(b), which has the following four maximum matchings.

$$M_1 = \{(v_1 - f_2), (v_2 - f_3), (v_3 - f_4), (v_4 - f_5)\}, \tag{4.6}$$
$$M_2 = \{(v_1 - f_1), (v_2 - f_3), (v_3 - f_4), (v_5 - f_5)\}, \tag{4.7}$$
$$M_3 = \{(v_1 - f_2), (v_2 - f_3), (v_3 - f_4), (v_5 - f_5)\}, \tag{4.8}$$
$$M_4 = \{(v_1 - f_1), (v_2 - f_3), (v_3 - f_4), (v_4 - f_5)\}. \tag{4.9}$$

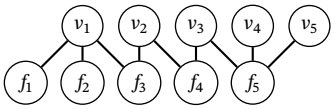
By Proposition 4.1 we know that the coarse decomposition $\text{CD}(\mathcal{B}', M)$, with $M \in \{M_1, M_2, M_3, M_4\}$, does not depend on the choice of maximum matching. The coarse decomposition is displayed in Figure 4.5(c). It is a straightforward exercise to verify that applying Definition 4.5 to each of the maximum matchings results in the same coarse decomposition. Note that if the vertices $\{f_1, \ldots f_5\}$ are associated with equations, and the vertices $\{v_1, \ldots, v_5\}$ are associated with variables, then the incomplete

[13] For completeness, we have included a proof of this theorem in Appendix 4.B.2.

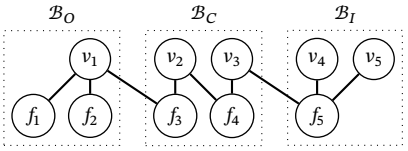
graph $\mathcal{B}_I$ has fewer equations than variables, whereas the over-complete graph $\mathcal{B}_O$ has more equations than variables. The complete graph $\mathcal{B}_C$ is self-contained. △



(a) Associated bipartite graph $\mathcal{B}$.



(b) Subgraph $\mathcal{B}'$ induced by $V \cup F$.



(c) Coarse decomposition of $\mathcal{B}'$.

Figure 4.5: The bipartite graph $\mathcal{B}$ associated with the system of equations in Example 4.8 is given in Figure (a). Its subgraph $\mathcal{B}'$ induced by $V = \{v_1, \dots, v_5\}$ and $F = \{f_1, \dots, f_5\}$ in Figure (b) is not self-contained. The coarse decomposition of $\mathcal{B}'$ is given in Figure (c).

4.3.3 Causal ordering via coarse decomposition

Here we present the extended causal ordering algorithm. It relies on the unique coarse decomposition of a bipartite graph into its incomplete, complete, and over-complete parts. Lemma 4.1, due to Pothen (1985), shows that the complete graph has a perfect matching. Together, Lemma 4.1 and Lemma 4.2 justify the steps in Algorithm 3 to construct a causal ordering graph. The proofs are provided in Appendix 4.B.2.

Lemma 4.1. *[Pothen (1985)] Let $\mathcal{B}$ be a bipartite graph with coarse decomposition $\langle T_I, T_C, T_O \rangle$. The subgraph $\mathcal{B}_C$ of $\mathcal{B}$ induced by vertices in T_C has a perfect matching and is self-contained.*

Lemma 4.2. *[Pothen (1985)] Let $\mathcal{B} = \langle V, F, E \rangle$ be a bipartite graph with a maximum matching M. Let $\mathrm{CD}(\mathcal{B}, M) = \langle T_I, T_C, T_O \rangle$ be the associated coarse decomposition. No edge joins a vertex in $T_I \cap V$ with a vertex in $(T_C \cup T_O) \cap F$ and no edge joins a vertex in $T_C \cap V$ with a vertex in $T_O \cap F$.*

Algorithm 3 takes a set of exogenous vertices $W \subseteq V$ and a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ as input. The output is a causal ordering graph $\langle \mathcal{V}, \mathcal{E}\rangle$. The algorithm first uses a maximum matching M for the subgraph $\mathcal{B}'$ of $\mathcal{B}$ induced by $(V \setminus W) \cup F$ to construct the coarse decomposition $\langle T_I, T_C, T_O\rangle$ of $\mathcal{B}'$. Since the complete graph $\mathcal{B}_C$ is self-contained (by Lemma 4.1) the causal ordering algorithm for self-contained bipartite graphs can be applied to obtain the directed cluster graph $\mathrm{CO}(\mathcal{B}_C) = \langle \mathcal{V}_C, \mathcal{E}_C\rangle$. The cluster set $\mathcal{V}$ consists of the clusters in $\mathcal{V}_C$ and the connected components in $\mathcal{B}_I$ and $\mathcal{B}_O$. The edge set $\mathcal{E}$ contains all edges in $\mathcal{E}_C$. For edges between vertices $v \in T_O \cap V$ and $f \in T_C \cap F$ in $\mathcal{B}$ an edge $(v \to \mathrm{cl}_{\mathcal{V}}(f))$ is added to $\mathcal{E}$.[14] Similarly, for edges between vertices $v \in (T_O \cup T_C) \cap V$ and $f \in T_I \cap F$ an edge $(v \to \mathrm{cl}_{\mathcal{V}}(f))$ is also added to $\mathcal{E}$. By Lemma 4.2 there are no other edges between the incomplete, complete, and overcomplete graphs. Finally, edges from exogenous vertices W are added to the causal ordering graph. The details can be found in Algorithm 3.

Algorithm 3: Causal ordering via coarse decomposition.

Input: a set of exogenous vertices W, a bipartite graph $\mathcal{B} = \langle V \cup W, F, E\rangle$.
Output: directed cluster graph $\langle \mathcal{V}, \mathcal{E}\rangle$
$\mathcal{B}' \leftarrow$ subgraph of $\mathcal{B}$ induced by $(V \setminus W) \cup F$
$\mathcal{M} \leftarrow$ maximum matching for $\mathcal{B}'$
$\langle T_I, T_C, T_O\rangle \leftarrow \mathrm{CD}(\mathcal{B}', \mathcal{M})$ `// coarse decomposition`
$\mathcal{B}_C \leftarrow$ subgraph of $\mathcal{B}'$ induced by T_C
$\mathcal{B}_I \leftarrow$ subgraph of $\mathcal{B}'$ induced by T_I
$\mathcal{B}_O \leftarrow$ subgraph of $\mathcal{B}'$ induced by T_O
$\langle \mathcal{V}_C, \mathcal{E}_C\rangle \leftarrow$ causal ordering graph for $\mathcal{B}_C$ `// construct clusters`
$\mathcal{V}_I \leftarrow$ partition of T_I into connected components in $\mathcal{B}_I$
$\mathcal{V}_O \leftarrow$ partition of T_O into connected components in $\mathcal{B}_O$
$\mathcal{V} \leftarrow \mathcal{V}_I \cup \mathcal{V}_C \cup \mathcal{V}_O \cup \{\{w\} : w \in W\}$
$\mathcal{E} \leftarrow \mathcal{E}_C$ `// find edges`
for $(v - f) \in E$ **do**
  if $v \in (T_O \cup T_C) \cap V$ **and** $f \in T_I \cap F$ **then**
    Add $(v \to \mathrm{cl}_{\mathcal{V}}(f))$ to $\mathcal{E}$
  else if $v \in T_O \cap V$ **and** $f \in T_C \cap F$ **then**
    Add $(v \to \mathrm{cl}_{\mathcal{V}}(f))$ to $\mathcal{E}$
for $w \in W$ **do**
  add $(w \to \mathrm{cl}_{\mathcal{V}}(f))$ to $\mathcal{E}$ for all $f \in \mathrm{adj}_{\mathcal{B}}(w)$ `// exogenous vertices`

Corollary 4.1. *The output of Algorithm 3 is well-defined and unique.*

Proof. This follows directly from Theorem 4.1 and Proposition 4.1. □

[14] Note that $\mathrm{cl}_{\mathcal{V}}(x)$ denotes the cluster in the partition $\mathcal{V}$ that contains the vertex x.

Corollary 4.1 shows that the output of causal ordering via coarse decomposition does not depend on the choice of the maximum matching (i.e. the output is unique). The following example provides a manual demonstration of the causal ordering algorithm via the coarse decomposition.

Example 4.8. We apply the causal ordering algorithm via coarse decomposition (i.e. Algorithm 3) to the bipartite graph in Figure 4.5(a). Its subgraph induced by endogenous variables and equations is the bipartite graph in Figure 4.5(b) and its coarse decomposition is given in Figure 4.5(c). Since, by Lemma 4.1, $\mathcal{B}_C$ is self-contained we can apply the causal ordering algorithm (Algorithm 1) to the complete subgraph resulting in the directed cluster graph $\mathrm{CO}(\mathcal{B}_C) = \langle \mathcal{V}_C, \mathcal{E}_C \rangle$ where $\mathcal{V}_C = \{\{v_2, f_3\}, \{v_3, f_4\}\}$ and $\mathcal{E}_C = \{(v_2 \to \{v_3, f_4\})\}$. The cluster set is then given by $\mathcal{V} = \mathcal{V}_C \cup \{\{v_4, v_5, f_5\}\} \cup \{\{v_1, f_1, f_2\}\}$. We then add singleton clusters $\{w_1\}$, $\{w_2\}$, $\{w_3\}$, $\{w_4\}$, $\{w_5\}$ for each exogenous vertex. Next we add the edges $\mathcal{E}_C$, $(v_1 \to \{\{v_2, f_3\})$ and $(v_3 \to \{v_4, v_5, f_5\})$ to the edge set $\mathcal{E}$. Finally, we add edges $(w_1 \to \{v_1, f_1, f_2\})$, $(w_2 \to \{v_1, f_1, f_2\})$, $(w_3 \to \{v_2, f_3\})$, $(w_4 \to \{v_3, f_4\})$ and $(w_5 \to \{v_4, v_5, f_5\})$ to the edge set $\mathcal{E}$. The resulting causal ordering graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ is given in Figure 4.6. △

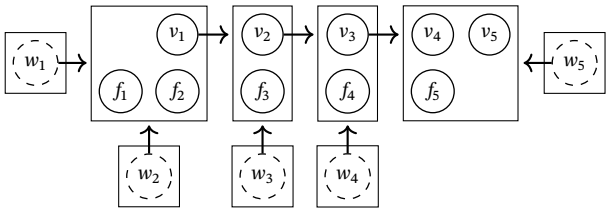


Figure 4.6: Causal ordering graph for the bipartite graph in Figure 4.5(a).

4.4 Markov ordering graph

First we consider (unique) solvability assumptions for systems of constraints. We will then construct the *Markov ordering graph* and prove that it implies conditional independences between variables that appear in constraints. We also apply our method to the model for the filling bathtub in Example 4.1. Finally, we present a novel result regarding the generalized directed global Markov property for solutions of systems of constraints and an *associated directed graph*.

4.4.1 Solvability for systems of constraints

In this section, we consider (unique) solutions of systems of constraints with exogenous random variables, and give a sufficient condition under which the output of the

causal ordering algorithm can be interpreted as the order in which sets of (endogenous) variables can be solved in a set of equations (i.e. constraints).

Definition 4.6. We say that a measurable mapping $\mathbf{g} : \mathfrak{X}_W \mapsto \mathfrak{X}_{V\setminus W}$ that maps values of the exogenous variables to values of the endogenous variables is a *solution to a system of constraints* $\langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B}\rangle$ if

$$\phi_f(\mathbf{g}_{V(f)\setminus W}(\mathbf{X}_W), \mathbf{X}_{V(f)\cap W}) = c_f, \quad \forall f \in F, \quad \mathbb{P}_{\mathbf{X}_W}\text{-a.s.}$$

We say that the system of constraints is *uniquely* solvable (or "has a unique solution") if all its solutions are $\mathbb{P}_{\mathbf{X}_W}$-a.s. equal.

The system of constraints in the example below is solvable but not uniquely solvable. The example illustrates that the dependence or independence between solution components (i.e. endogenous variables) is not the same for all solutions.

Example 4.9. Consider a system of constraints $\langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B}\rangle$ with $\mathfrak{X} = \mathbb{R}^4$ and independent exogenous random variables $\mathbf{X}_W = (X_w)_{w\in\{w_1,w_2\}}$ taking value in $\mathbb{R}^2$. Suppose that $\boldsymbol{\Phi}$ consists of the constraints

$$\Phi_{f_1} = \langle X_{V(f_1)} \mapsto X_{v_1} - X_{w_1}, 0, \{v_1, w_1\}\rangle, \tag{4.10}$$
$$\Phi_{f_2} = \langle X_{V(f_2)} \mapsto X_{v_2}^2 - |X_{w_2}|, 0, \{v_2, w_2\}\rangle. \tag{4.11}$$

This system of constraints has solutions with different distributions. One solution is given by $(X^*_{v_1}, X^*_{v_2}) = (X_{w_1}, \sqrt{|X_{w_2}|})$ and another solution is $(X'_{v_1}, X'_{v_2}) = (X_{w_1}, \operatorname{sgn}(X_{w_1})\sqrt{|X_{w_2}|})$. Note that the solution components $X^*_{v_1}$ and $X^*_{v_2}$ are independent, whereas the solution components X'_{v_1} and X'_{v_2} may be dependent. △

Underspecified (and overspecified) systems of constraints can be avoided by the requirement that it is *uniquely solvable*. In Definition 4.7 below we give a sufficient condition under which a unique solution can be obtained by solving variables in clusters from equations in these clusters.

Definition 4.7. A system of constraints $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B}\rangle$ is *solvable w.r.t. constraints* $S_F \subseteq F$ *and endogenous variables* $S_V \subseteq V(S_F)\setminus W$ if there exists a measurable function $\mathbf{g}_{S_V} : \mathfrak{X}_{V(S_F)\setminus S_V} \to \mathfrak{X}_{S_V}$ s.t. $\mathbb{P}_{\mathbf{X}_W}$-a.s., for all $\mathbf{x}_{V(S_F)\setminus W} \in \mathfrak{X}_{V(S_F)\setminus W}$:

$$\phi_f(\mathbf{x}_{V(f)\setminus W}, \mathbf{X}_{V(f)\cap W}) = c_f, \ \ \forall f \in S_F \iff \mathbf{x}_{S_V} = \mathbf{g}_{S_V}(\mathbf{x}_{V(S_F)\setminus(S_V\cup W)}, \mathbf{X}_{V(S_F)\cap W}).$$

$\mathcal{M}$ is *uniquely solvable w.r.t. constraints* S_F *and endogenous variables* S_V if the converse implication also holds.

The following condition suffices for the existence of a unique solution that can be obtained by solving for variables from equations in their cluster along a topolo-

gical ordering of the clusters in the causal ordering graph. This weakens the assumptions made in Simon (1953) who requires both unique solvability w.r.t. every subset of equations (and the endogenous variables that appear in them) and self-containedness of the bipartite graph.

Definition 4.8. We say that $\mathcal{M}$ is *uniquely solvable w.r.t. the causal ordering graph* $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ if it is uniquely solvable w.r.t. $S \cap F$ and $S \cap V$ for all $S \in \mathcal{V}$ with $S \cap W = \emptyset$.

For systems of constraints for cyclic models or with non-linear equations, for which the incomplete subgraph is not the empty graph, the condition of unique solvability with respect to the causal ordering graph is not always satisfied. This is illustrated by Example 4.10 below.

Example 4.10. Let $V = \{v_1, \dots v_5\}$ be an index set for endogenous variables $X_{v_1}, \dots, X_{v_5}$ taking value in $\mathbb{R}$, $W = \{w_1, \dots, w_5\}$ an index set for independent exogenous random variables $U_{w_1}, \dots, U_{w_5}$ taking value in $\mathbb{R}$, and p_1, p_2 parameters with values in $\mathbb{R}$. Consider the following non-linear system of constraints:

$$\Phi_{f_1} : \quad X_{v_1}^2 - U_{w_1} = 0, \tag{4.12}$$
$$\Phi_{f_2} : \quad \mathrm{sgn}(X_{v_1}) - \mathrm{sgn}(U_{w_2}) = 0, \tag{4.13}$$
$$\Phi_{f_3} : \quad X_{v_2} - p_1 X_{v_1} - U_{w_3} = 0, \tag{4.14}$$
$$\Phi_{f_4} : \quad X_{v_3} - p_2 X_{v_2} - U_{w_4} = 0, \tag{4.15}$$
$$\Phi_{f_5} : \quad X_{v_3} + X_{v_4} + X_{v_5} - U_{w_5} = 0. \tag{4.16}$$

The associated bipartite graph $\mathcal{B}$ is given in Figure 4.5(a) and the corresponding causal ordering graph is given in Figure 4.6. It is easy to check that the system of constraints is uniquely solvable with respect to the clusters $\{v_1, f_1, f_2\}$, $\{v_2, f_3\}$, and $\{v_3, f_4\}$ in the causal ordering graph. Equation f_5 does not provide a unique solution for the variables v_4 and v_5 and hence the system is not uniquely solvable with respect to the cluster $\{v_4, v_5, f_5\}$. △

Generally speaking, systems of constraints are not uniquely solvable with respect to the clusters in the incomplete set of vertices in the associated bipartite graph. In order to derive a Markov property for the complete and overcomplete sets of vertices in the associated bipartite graph, we use the condition in Definition 4.9 below, which is slightly weaker than the one in Definition 4.7. Since self-contained bipartite graphs do not have an incomplete part there is no difference between the two conditions in that case.

Definition 4.9. Let $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} \rangle$ be a system of constraints. Denote its coarse decomposition by $\mathrm{CD}(\mathcal{B}) = \langle T_I, T_C, T_O \rangle$ and its causal ordering graph by $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$. We say that $\mathcal{M}$ is *maximally uniquely solvable* if it is

(i) uniquely solvable w.r.t. $S \cap F$ and $S \cap V$ for all $S \in \mathcal{V}$ with $S \cap W = \emptyset$ and $S \cap T_I = \emptyset$, and
(ii) solvable with respect to $T_I \cap F$ and $(T_I \cap V) \setminus W$.

This condition suffices to guarantee the existence of a solution, and that it is unique on the (over)complete part $(T_O \cup T_C) \cap V \setminus W$.

4.4.2 Directed global Markov property via causal ordering

The *Markov ordering graph* is constructed from a causal ordering graph by *declustering* and then marginalizing out the vertices that correspond to constraints.

Definition 4.10. Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a directed cluster graph. The *declustered* graph is given by $D(\mathcal{G}) = \langle V, E \rangle$ with $V = \cup_{S \in \mathcal{V}} S$ and $(v \to w) \in E$ if and only if $(v \to \mathrm{cl}(w)) \in \mathcal{E}$. For a system of constraints $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} \rangle$ with $\mathcal{B} = \langle V, F, E \rangle$, we say that $\mathrm{MO}(\mathcal{B}) = D(\mathrm{CO}(\mathcal{B}))_{\mathrm{mar}(F)}$ is the *Markov ordering graph*, where $\mathrm{MO}(\mathcal{B}) = D(\mathrm{CO}(\mathcal{B}))_{\mathrm{mar}(F)}$ is the *latent projection* of $D(\mathrm{CO}(\mathcal{B}))$ onto $V \setminus F$, see Section 4.A.1 for more details.[15]

Under the assumption that systems of constraints are uniquely solvable with respect to the (over)complete part of their causal ordering graph, Theorem 4.3 relates d-separations between vertices in the Markov ordering graph to conditional independences between the corresponding components of a solution of the system of constraints.

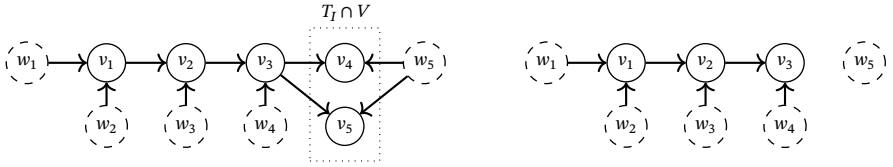
Theorem 4.3. *Let* $\mathbf{X}^* = \mathbf{h}(\mathbf{X}_W)$ *with* $\mathbf{h} : \mathfrak{X}_W \to \mathfrak{X}_{V \setminus W}$ *be a solution of a system of constraints* $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} \rangle$ *with coarse decomposition* $\mathrm{CD}(\mathcal{B}) = \langle T_I, T_C, T_O \rangle$. *Let* $\mathrm{MO}_{\mathrm{CO}}(\mathcal{B})$ *denote the subgraph of the Markov ordering graph induced by* $T_C \cup T_O$ *and let* $\mathbf{X}^*_{\mathrm{CO}}$ *denote the corresponding solution components. If* $\mathcal{M}$ *is maximally uniquely solvable then the pair* $(\mathrm{MO}_{\mathrm{CO}}(\mathcal{B}), \mathbb{P}_{\mathbf{X}^*_{\mathrm{CO}}})$ *satisfies the directed global Markov property (see Definition 4.15).*

In particular, when the incomplete and overcomplete sets are empty (i.e. when $T_I = \emptyset$ and $T_C = \emptyset$) and the system is uniquely solvable with respect to the causal ordering graph, Theorem 4.3 tells us that the pair $(\mathrm{MO}(\mathcal{B}), \mathbb{P}_{\mathbf{X}^*})$ satisfies the directed global Markov property.

Example 4.11. Consider the system of constraints in Example 4.10. The Markov ordering graph for the associated bipartite graph in Figure 4.5(a) can be constructed from the causal ordering graph in Figure 4.6 and is given in Figure 4.7. One can check that the system of constraints is uniquely solvable with respect to the clusters in the complete and overcomplete sets. The Markov ordering graph can be used to read off

[15] Throughout this work, we sometimes add exogenous vertices in the Markov ordering graph to emphasize their presence. At other times, we exclude them for brevity.

conditional independences from d-separations between vertices that are not in the incomplete part. For example, since v_1 is d-separated from v_3 given v_2, we deduce that $X_{v_1} \perp\!\!\!\perp X_{v_3} \,|\, X_{v_2}$, for any solution of the constraints. △



(a) $\mathrm{MO}(\mathcal{B})$. (b) $\mathrm{MO}_{\mathrm{CO}}(\mathcal{B})$.

Figure 4.7: (a) The Markov ordering graph associated with the system of constraints in Example 4.10. It can be constructed from the causal ordering graph in Figure 4.6. The vertices in the incomplete graph are indicated by the dashed rectangle. (b) Its subgraph induced by $T_C \cap T_O$. Theorem 4.3 shows that d-separations in $\mathrm{MO}_{\mathrm{CO}}(\mathcal{B})$ imply conditional independences.

4.4.3 Application to the filling bathtub

In Example 4.1 we informally described an equilibrium model for a filling bathtub. The endogenous variables of the system are the diameter X_{v_K} of the drain, the rate X_{v_I} at which water flows from the faucet, the water pressure X_{v_P}, the rate X_{v_O} at which the water goes through the drain and the water level X_{v_D}. The model is formally represented by a system of constraints $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} \rangle$ where:

(i) $\mathfrak{X} = \mathbb{R}_{>0}^{12}$ is a product of standard measurable spaces corresponding to the domain of variables that are indexed by $\{v_K, v_I, v_P, v_O, v_D, w_K, w_I, w_1, \dots, w_5\}$,

(ii) $\mathbf{X}_W = \{U_{w_I}, U_{w_K}, U_{w_1}, \dots, U_{w_5}\}$ is a family of independent exogenous random variables,

(iii) $\boldsymbol{\Phi}$ is a family of constraints:

$$
\begin{aligned}
\Phi_{f_K} &= \langle X_{V(f_K)} \mapsto X_{v_K} - U_{w_K}, && 0, && V(f_K) = \{v_K, w_K\} \rangle, \\
\Phi_{f_I} &= \langle X_{V(f_I)} \mapsto X_{v_I} - U_{w_I}, && 0, && V(f_I) = \{v_I, w_I\} \rangle, \\
\Phi_{f_P} &= \langle X_{V(f_P)} \mapsto U_{w_1}(g U_{w_2} X_{v_D} - X_{v_P}), && 0, && V(f_P) = \{v_D, v_P, w_1, w_2\} \rangle, \\
\Phi_{f_O} &= \langle X_{V(f_O)} \mapsto U_{w_3}(U_{w_4} X_{v_K} X_{v_P} - X_{v_O}), && 0, && V(f_O) = \{v_K, v_P, v_O, w_3, w_4\} \rangle, \\
\Phi_{f_D} &= \langle X_{V(f_D)} \mapsto U_{w_5}(X_{v_I} - X_{v_O}), && 0, && V(f_D) = \{v_I, v_O, v_5\} \rangle,
\end{aligned}
$$

(iv) The associated bipartite graph $\mathcal{B} = \langle V, F, E \rangle$ is as in Figure 4.8. The vertices $F = \{f_K, f_I, f_P, f_O, f_D\}$ correspond to constraints and the vertices $V \setminus W = \{v_K, v_I, v_P, v_O, v_D\}$ and $W = \{w_I, w_K, w_1, \dots, w_5\}$ correspond to endogenous and exogenous variables respectively. Note that the subgraph induced by the endogenous vertices $V \setminus W$ is the self-contained bipartite graph presented in Figure 4.1(a).

Solvability with respect to the causal ordering graph: Applying Algorithm 1 to the bipartite graph results in the causal ordering graph $\mathrm{CO}(\mathcal{B})$ in Figure 4.9. Since the

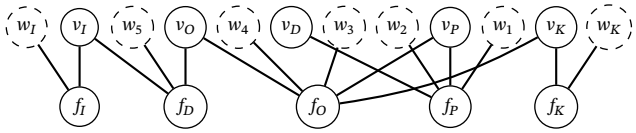


Figure 4.8: The bipartite graph associated with the equilibrium equations of the bathtub.

bipartite graph induced by the endogenous variables and equations is self-contained, there is no incomplete or overcomplete subgraph. The assumption of maximal unique solvability in Theorem 4.3 then reduces to the assumption of unique solvability with respect to the causal ordering graph. Through explicit calculations, it is easy to verify that $\mathcal{M}$ is (maximally) uniquely solvable with respect to $\mathrm{CO}(\mathcal{B})$, whenever $g \neq 0$:

(i) For the cluster $\{f_K, v_K\}$ we have that $X_{v_K} - U_{w_K} = 0 \iff X_{v_K} = U_{w_K}$.

(ii) For the cluster $\{f_I, v_I\}$ we have that $X_{v_I} - U_{w_I} = 0 \iff X_{v_I} = U_{w_I}$.

(iii) For $\{f_O, v_P\}$ we have that $U_{w_3}(U_{w_4}X_{v_K}X_{v_P} - X_{v_O}) = 0 \iff X_{v_P} = \dfrac{X_{v_O}}{U_{w_3}X_{v_K}}$.

(iv) For $\{f_D, v_O\}$ we have that $U_{w_5}(X_{v_I} - X_{v_O}) \iff X_{v_O} = X_{v_I}$.

(v) For $\{f_P, v_D\}$ we have that $U_{w_1}(gU_{w_2}X_{v_D} - X_{v_P}) \iff X_{v_D} = \dfrac{X_{v_P}}{gU_{w_2}}$.

In practice, we do not always need to manually check the assumption of unique solvability with respect to the causal ordering graph. For example, in linear systems of equations of the form $\mathbf{AX} = \mathbf{Y}$, we may use the fact that this assumption is satisfied when the matrix of coefficients $\mathbf{A}$ is invertible. More generally, global implicit function theorems give conditions under which (non-linear) systems of equations have a unique solution (Krantz et al., 2013).[16] We consider detailed analysis of conditions under which (maximal) unique solvability is guaranteed to be outside the scope of this paper. Note that, under the assumption of (maximal) unique solvability, the conditional independences can be read off from the Markov ordering graph without the requirement of calculating explicit solutions.

Markov ordering graph: Application of declustering and marginalization of vertices in F, as in Definition 4.10, to the causal ordering graph in Figure 4.9 results in the Markov ordering graph in Figure 4.10(a). Since $\mathcal{M}$ is uniquely solvable with respect to $\mathrm{CO}(\mathcal{B})$, Theorem 4.3 tells us that the pair $(\mathrm{MO}(\mathcal{B}), \mathbb{P}_{\mathbf{X}^*})$ satisfies the directed global Markov property, where $\mathbf{X}^*$ is a solution of $\mathcal{M}$.[17]

[16] In particular, Hadamard's global implicit function theorem in Krantz et al. (2013) states the following (Hadamard, 1906). Let $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^n$ be a C^2 mapping. Suppose that $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ and that the Jacobian determinant is non-zero at each point. Further suppose that whenever $K \subseteq \mathbb{R}^n$ is compact then $\mathbf{f}^{-1}(K)$ is compact (i.e. $\mathbf{f}$ is proper). Then $\mathbf{f}$ is one-to-one and onto. In the literature, several conditions have been formulated yielding global inverse theorems in different or more general settings, see for example Gutú (2017) and Idczak (2016).

[17] Recall that there is no incomplete and overcomplete part of the bipartite graph. Therefore we have that $\mathrm{MO}_{\mathrm{CO}}(\mathcal{B}) = \mathrm{MO}(\mathcal{B})$.

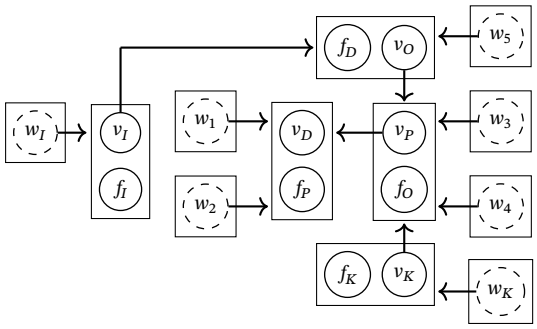
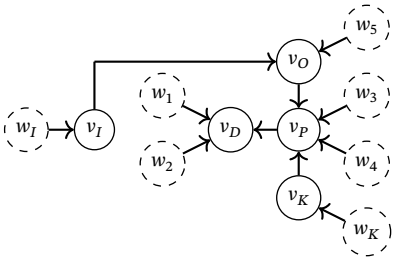
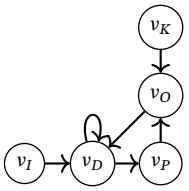


Figure 4.9: The causal ordering graph for the equilibrium equations of the bathtub system.



(a) Markov ordering graph.



(b) Graph of the SCM.

Figure 4.10: The Markov ordering graph for the equilibrium equations of the filling bathtub system, obtained by applying Definition 4.10 to the causal ordering graph in Figure 4.9 is given in Figure 4.10(a). The d-separations in the Markov ordering graph imply conditional independences between corresponding endogenous variables. Most of these conditional independences cannot be read off from the graph for the SCM of the bathtub system in Figure 4.10(b), except for $X_{v_I} \perp\!\!\!\perp X_{v_K}$.

Encoded conditional independences: Since the assumption of unique solvability with respect to the causal ordering graph holds for this particular example, we can read off conditional independences between endogenous variables from the Markov ordering graph. More precisely, the d-separations in $\mathrm{MO}(\mathcal{B})$ between vertices in $V \setminus W$ imply conditional independences between the corresponding endogenous variables. For example:

$$
\begin{aligned}
v_K \overset{d}{\underset{\mathrm{MO}(\mathcal{B})}{\perp}} v_O \quad &\Longrightarrow \quad X_{v_K} \perp\!\!\!\perp X_{v_O}, \\
v_K \overset{d}{\underset{\mathrm{MO}(\mathcal{B})}{\perp}} v_D \,|\, v_P \quad &\Longrightarrow \quad X_{v_K} \perp\!\!\!\perp X_{v_D} \,|\, X_{v_P}, \\
v_I \overset{d}{\underset{\mathrm{MO}(\mathcal{B})}{\perp}} v_P \,|\, v_O \quad &\Longrightarrow \quad X_{v_I} \perp\!\!\!\perp X_{v_P} \,|\, X_{v_O},
\end{aligned}
$$

$$v_O \overset{d}{\underset{\mathrm{MO}(\mathcal{B})}{\perp}} v_D \,|\, v_P \implies X_{v_O} \perp\!\!\!\perp X_{v_D} \,|\, X_{v_P}.$$

For $g > 0$, every solution to the system of constraints has the same distribution, and this distribution is d-faithful to the Markov ordering graph. When $g = 0$, the system of constraints only has a solution if $U_{w_I} = 0$ almost surely; in that case the corresponding distribution is not d-faithful w.r.t. the Markov ordering graph in Figure 4.10(a).

Comparison to SCM representation: The (random) differential equations that describe the system of a bathtub can be *equilibrated to an SCM* that has a self-cycle. Bongers, Forré, et al. (2020) show that the model has the following structural equations:

$$\begin{aligned}
X_{v_K} &= U_{w_K}, \\
X_{v_I} &= U_{w_I}, \\
X_{v_P} &= gU_{w_3}X_{v_D}, \\
X_{v_O} &= U_{w_5}X_{v_K}X_{v_P}, \\
X_{v_D} &= X_{v_D} + U_{w_1}(X_{v_I} - X_{v_O}).
\end{aligned}$$

The graph of this SCM is depicted in Figure 4.10(b). Because the SCM is uniquely solvable w.r.t. the strongly connected components $\{v_I\}, \{v_D, v_P, v_O\}$ and $\{v_K\}$, the σ-separations in this graph imply conditional independences (Theorem 6.3 in Bongers, Forré, et al., 2020). Most of the conditional independences implied by the Markov ordering graph cannot be read off from the graph of this SCM in Figure 4.10(b) via the σ-separation criterion, except for $X_{v_I} \perp\!\!\!\perp X_{v_K}$. Clearly the distribution of a solution to the system of constraints is not faithful to the graph of the SCM and causal ordering on the equilibrium equations provides a stronger Markov property than equilibration to an SCM.

An important difference between SCMs and systems of constraints is that while the former require a particular one-to-one correspondence between endogenous variables and structural equations, the latter do not require a similar correspondence between endogenous variables and constraints. Interestingly, in the case of the bathtub model, a one-to-one correspondence between variables and constraints is obtained automatically by the causal ordering algorithm. In general, the bipartite graph of a set of structural equations is self-contained and perfect matchings connect each variable to an equation. If the SCM is acyclic then the associated bipartite graph has a unique perfect matching that retrieves the correspondence between variables and equations in the SCM. We further discuss applications of the technique of causal ordering to structural equations in Section 4.6.2.

4.4.4 Generalized directed global Markov property

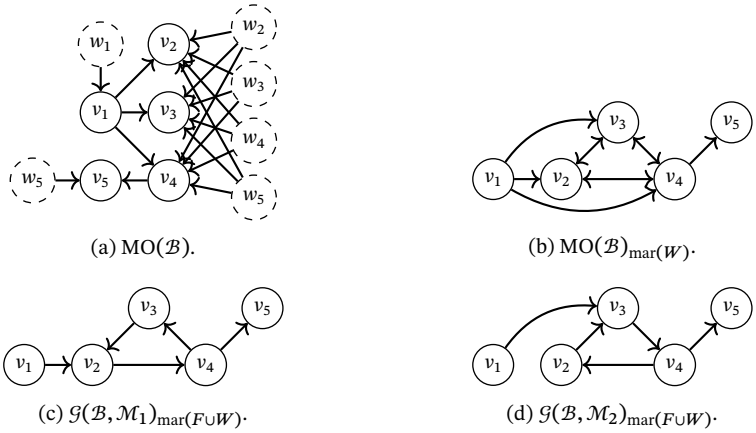
For systems of constraints with no over- or incomplete parts, the associated directed graph that is constructed in the causal ordering algorithm via perfect matchings also yields a Markov property. Theorem 4.4 below shows that for systems that are uniquely solvable with respect to the causal ordering graph, the σ-separations between variable vertices in the directed graph $\mathcal{G}(\mathcal{B},\mathcal{M})_{\mathrm{mar}(F)}$ imply conditional independences between the corresponding solution components.

Theorem 4.4. *Let $\mathbf{X}^* = \mathbf{g}(\mathbf{X}_W)$ be a solution of a system of constraints $\langle \mathcal{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B}\rangle$, where the subgraph of $\mathcal{B} = \langle V, F, E\rangle$ induced by $(V \cup F) \setminus W$ has a perfect matching $\mathcal{M}$. If for each strongly connected component S in $\mathcal{G}(\mathcal{B},\mathcal{M})$ with $S \cap W = \emptyset$, the system $\mathcal{M}$ is uniquely solvable w.r.t. $S_V = (S \cup \mathcal{M}(S)) \cap V$ and $S_F = (S \cup \mathcal{M}(S)) \cap F$ then the pair $(\mathcal{G}(\mathcal{B},\mathcal{M})_{\mathrm{mar}(F)}, \mathbb{P}_{\mathbf{X}^*})$ satisfies the generalized directed global Markov property (Definition 4.15).*

Example 4.12. Consider a system of constraints $\mathcal{M} = \langle \mathcal{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B}\rangle$ with $W = \{w_1, \dots, w_6\}$, $V \setminus W = \{v_1, \dots, v_5\}$, $F = \{f_1, \dots, f_5\}$, and $\mathcal{B} = \langle V, F, E\rangle$ as in Figure 4.4(a). Suppose that $\mathcal{X} = \mathbb{R}^{11}$ and consists of constraints:

$$\begin{aligned}
\Phi_{f_1} &: & X_{v_1} - X_{w_1} &= 0,\\
\Phi_{f_2} &: & X_{v_2} - X_{v_1} + X_{v_3} + X_{w_2} - X_{w_3} &= 0,\\
\Phi_{f_3} &: & X_{w_4} - X_{v_3} + X_{v_4} &= 0,\\
\Phi_{f_4} &: & X_{w_5} + X_{v_2} - X_{v_4} &= 0,\\
\Phi_{f_5} &: & X_{w_6} - X_{v_4} + X_{v_5} &= 0.
\end{aligned}$$

It is easy to check that this linear system of equations can be uniquely solved in the order prescribed by the causal ordering graph $\mathrm{CO}(\mathcal{B})$ in Figure 4.4(e). Therefore, according to Theorem 4.3 the d-separations among endogenous variables in the corresponding Markov ordering graph $\mathrm{MO}(\mathcal{B})$ imply conditional independences between the corresponding endogenous variables. It follows that d-separations in the Markov ordering graph $\mathrm{MO}(\mathcal{B})_{\mathrm{mar}(W)}$ for the endogenous variables in Figure 4.11(b) imply conditional independences between the corresponding variables. For example, we see that v_1 and v_5 are d-separated by v_4 and deduce that for a solution $\mathbf{X}^*$ to the system of constraints it holds that $X^*_{v_1} \perp\!\!\!\perp X^*_{v_5} \,|\, X^*_{v_4}$. One may note that d-separations in $\mathrm{MO}(\mathcal{B})_{\mathrm{mar}(W)}$ coincide with σ-separations in both associated directed graphs $\mathcal{G}(\mathcal{B},\mathcal{M}_1)_{\mathrm{mar}(F\cup W)}$ and $\mathcal{G}(\mathcal{B},\mathcal{M}_2)_{\mathrm{mar}(F\cup W)}$, which are depicted in Figures 4.11(c) and 4.11(d) respectively. It can be seen from the proof of Theorem 4.4 in Appendix 4.B.5 that this result holds in general. It follows from Theorem 4.4 that the σ-separations in $\mathcal{G}(\mathcal{B},\mathcal{M}_1)_{\mathrm{mar}(F\cup W)}$ and $\mathcal{G}(\mathcal{B},\mathcal{M}_2)_{\mathrm{mar}(F\cup W)}$ imply conditional independences between the corresponding variables. For example, we see that v_1 and v_5 are σ-separated by v_4 in both graphs, and hence $X^*_{v_1} \perp\!\!\!\perp X^*_{v_5} \,|\, X^*_{v_4}$ for a solution $\mathbf{X}^*$. △



(a) $\mathrm{MO}(\mathcal{B})$. (b) $\mathrm{MO}(\mathcal{B})_{\mathrm{mar}(W)}$.

(c) $\mathcal{G}(\mathcal{B}, \mathcal{M}_1)_{\mathrm{mar}(F \cup W)}$. (d) $\mathcal{G}(\mathcal{B}, \mathcal{M}_2)_{\mathrm{mar}(F \cup W)}$.

Figure 4.11: The Markov ordering graph of the causal ordering graph in Figure 4.4(e) is given in Figure 4.11(a). Marginalization of the exogenous vertices W results in the directed mixed graph in Figure 4.11(b). The directed graphs in Figures 4.11(c) and 4.11(d) are obtained by marginalizing out the constraint vertices F and exogenous vertices W from the directed graphs $\mathcal{G}(\mathcal{B}, \mathcal{M}_1)$ and $\mathcal{G}(\mathcal{B}, \mathcal{M}_2)$ in Figures 4.4(b) and 4.4(c) respectively. Note that d-separations in the Markov ordering graph correspond to σ-separations in the associated directed graphs in Figures 4.4(b) and 4.4(c).

4.5 Causal implications of sets of equations

Nowadays, it is common to relate causation directly to the effects of manipulation (Pearl, 2009; Woodward, 2003). In the context of sets of equations there are many ways to model manipulations on these equations. Assuming that the manipulations correspond to feasible actions in the real world that is modelled by the equations, the effects of manipulations correspond to causal relations. In order to derive causal implications from systems of constraints, we explicitly define two types of manipulation. We consider the notions of both *soft* and *perfect* interventions on sets of equations.[18] We prove that the causal ordering graph represents the effects of both *soft interventions on equations* and *perfect interventions on clusters* in the causal ordering graph. We also show that these interventions commute with causal ordering.

4.5.1 The effects of soft interventions

A *soft intervention*, also known as a "mechanism change", acts on a constraint. It replaces the targeted constraint by a constraint in which the same variables appear

[18]Our definitions in the context of systems of constraints may deviate from conventional definitions of interventions on SCMs. In an SCM, each variable is associated with a single structural equation. The notion of a perfect intervention on an SCM does not carry over to systems of constraints because there is no imposed one-to-one correspondence between equations and variables.

as in the original one. This type of intervention does not change the bipartite graph that represents the structure of the constraints.

Definition 4.11. Let $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} \rangle$ be a system of constraints, let $\Phi_f = \langle \phi_f, c_f, V(f) \rangle \in \boldsymbol{\Phi}$ be a constraint, c'_f a constant taking value in a measurable space $\mathcal{Y}$, and $\phi'_f : \mathfrak{X}_{V(f)} \to \mathcal{Y}$ a measurable function. A *soft intervention* $\mathrm{si}(f, \phi'_f, c'_f)$ *targeting* Φ_f results in the intervened system $\mathcal{M}_{\mathrm{si}(f,\phi'_f,c'_f)} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}_{\mathrm{si}(f,\phi'_f,c'_f)}, \mathcal{B} \rangle$ where $\boldsymbol{\Phi}_{\mathrm{si}(f,\phi'_f,c'_f)} = \big(\boldsymbol{\Phi} \setminus \{\Phi_f\}\big) \cup \{\Phi'_f\}$ with $\Phi'_f = \langle \phi'_f, c'_f, V(f) \rangle$.

For systems of constraints that are maximally uniquely solvable w.r.t. the causal ordering graph, both before and after a soft intervention, Theorem 4.5 shows that such a soft intervention does not have an effect on variables that cannot be reached by a directed path from that constraint in the causal ordering graph, while it may have an effect on other variables.[19]

Theorem 4.5. *Let $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} \rangle$ be a system of constraints with coarse decomposition $\mathrm{CD}(\mathcal{B}) = \langle T_I, T_C, T_O \rangle$. Suppose that $\mathcal{M}$ is maximally uniquely solvable w.r.t. the causal ordering graph $\mathrm{CO}(\mathcal{B})$ and let $\mathbf{X}^* = \mathbf{g}(\mathbf{X}_W)$ be a solution of $\mathcal{M}$. Let $f \in (T_C \cup T_O) \cap F$ and assume that the intervened system $\mathcal{M}_{\mathrm{si}(f,\phi'_f,c'_f)}$ is also maximally uniquely solvable w.r.t. $\mathrm{CO}(\mathcal{B})$. Let $\mathbf{X}' = \mathbf{h}(\mathbf{X}_W)$ be a solution of $\mathcal{M}_{\mathrm{si}(f,\phi'_f,c'_f)}$. If there is no directed path from f to $v \in (T_C \cup T_O) \cap V$ in $\mathrm{CO}(\mathcal{B})$ then $X^*_v = X'_v$ almost surely. On the other hand, if there is a directed path from f to v in $\mathrm{CO}(\mathcal{B})$ then X^*_v may have a different distribution than X'_v, depending on the details of the model $\mathcal{M}$.*

Example 4.13 shows that the presence of a directed path in the causal ordering graph for the equilibrium equations of the bathtub system implies a causal effect for almost all parameter values. This illustrates that non-effects and *generic* effects can be read off from the causal ordering graph.[20]

Example 4.13. Recall the system of constraints for the filling bathtub in Section 4.4.3. Think of an experiment where the gravitational constant g is changed so that it takes on a different value g' without altering the other equations that describe the bathtub system. Such an experiment is, at least in theory, feasible. For example, it can be accomplished by accelerating the bathtub system or by moving the bathtub system to another planet. We can model the effect on the equilibrium distribution in such an experiment by a soft intervention targeting f_P that replaces the constraint Φ_{f_P} by

$$\langle X_{V(f_P)} \mapsto U_{w_1}(g' U_{w_2} X_{v_D} - X_{v_P}),\ 0,\ V(f_P) = \{v_D, v_P, w_1, w_2\} \rangle. \tag{4.17}$$

[19] Our result generalizes Theorem 6.1 in Simon (1953) for linear self-contained systems of equations. The proof of our theorem is similar.

[20] If a directed path from an equation vertex f to a variable vertex v implies that an intervention on f changes the distribution of the solution component X_v for almost all values (w.r.t. Lebesgue measure) of the parameters, then we say that there is a *generic* causal effect of f on v.

Table 4.1: The effects of soft interventions on constraints in the causal ordering graph for the bathtub system in Figure 4.9.

target	generic effect	non-effect
f_K	$X_{v_K}, X_{v_P}, X_{v_D}$	X_{v_I}, X_{v_O}
f_I	$X_{v_I}, X_{v_P}, X_{v_O}, X_{v_D}$	X_{v_K}
f_P	X_{v_D}	$X_{v_K}, X_{v_I}, X_{v_P}, X_{v_O}$
f_O	X_{v_P}, X_{v_D}	$X_{v_K}, X_{v_I}, X_{v_O}$
f_D	$X_{v_P}, X_{v_O}, X_{v_D}$	X_{v_K}, X_{v_I}

Which variables are and which are not affected by this soft intervention? We can read off the effects of this soft intervention from the causal ordering graph in Figure 4.9. There is no directed path from f_P to v_K, v_I, v_P or v_O. Therefore, perhaps surprisingly, Theorem 4.5 tells us that the soft intervention targeting f_P neither has an effect on the pressure X_{v_P} at equilibrium nor on the outflow rate X_{v_O} at equilibrium. Since there is a directed path from f_P to v_D, the water level X_{v_D} at equilibrium may be different after a soft intervention on f_P. If the gravitational constant g is equal to zero, then the system of constraints for the bathtub is not maximally uniquely solvable w.r.t. the causal ordering graph (except if $U_{w_I} = 0$ almost surely). For all other values of the parameter g the generic effects and non-effects of soft interventions on other constraints of the bathtub system can be read off from the causal ordering graph and are presented in Table 4.1. △

4.5.2 The effects of perfect interventions

A *perfect intervention* acts on a variable and a constraint. Definition 4.12 shows that it replaces the targeted constraint by a constraint that sets the targeted variable equal to a constant. Note that this definition of perfect interventions is very general and allows interventions for which the intervened system of constraints is not maximally uniquely solvable w.r.t. the causal ordering graph. In this work, we will only consider the subset of perfect interventions that target clusters in the causal ordering graph, for which the intervened system is also maximally uniquely solvable w.r.t. the causal ordering graph. We consider an analysis of necessary conditions on interventions for the intervened system to be consistent beyond the scope of this work.

Definition 4.12. Let $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} = \langle V, F, E \rangle \rangle$ be a system of constraints and let $\xi_v \in \mathcal{X}_v$. A *perfect intervention* $\mathrm{do}(f, v, \xi_v)$ *targeting the variable* $v \in V \setminus W$ *and the constraint* $f \in F$ results in an intervened system, denoted as $\mathcal{M}_{\mathrm{do}(f,v,\xi_v)} = \langle \mathcal{X}, \mathbf{X}_W, \boldsymbol{\Phi}_{\mathrm{do}(f,v,\xi_v)}, \mathcal{B}_{\mathrm{do}(f,v)} \rangle$ where

(i) $\boldsymbol{\Phi}_{\mathrm{do}(f,v,\xi_v)} = \big(\boldsymbol{\Phi} \setminus \Phi_f\big) \cup \{\Phi'_f\}$ with $\Phi'_f = \langle X_v \mapsto X_v, \xi_v, \{v\}\rangle$,
(ii) $\mathcal{B}_{\mathrm{do}(f,v)} = \langle V, F, E'\rangle$ with $E' = \{(i-j) \in E : i, j \neq f\} \cup \{(v-f)\}$.

Perfect interventions on a set of variable-constraint pairs $\{(f_1, v_1), \dots, (f_n, v_n)\}$ in a system of constraints are denoted by $\mathrm{do}(S_F, S_V, \boldsymbol{\xi}_{S_V})$ where $S_F = \langle f_1, \dots, f_n\rangle$ and $S_V = \langle v_1, \dots, v_n\rangle$ are tuples. For a bipartite graph $\mathcal{B}$ so that its subgraph induced by $(V \cup F) \setminus W$ is self-contained, Lemma 4.3 shows that the subgraph of the intervened bipartite graph $\mathcal{B}_{\mathrm{do}(S_F,S_V)}$ induced by $(V \cup F) \setminus W$ is also self-contained when $S = (S_F \cup S_V)$ is a cluster in $\mathrm{CO}(\mathcal{B})$ with $S \cap W = \emptyset$.

Lemma 4.3. *Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph and $W \subseteq V$, so that the subgraph of $\mathcal{B}$ induced by $(V \cup F) \setminus W$ is self-contained. Consider an intervention $\mathrm{do}(S_V, S_F)$ on a cluster $S = S_F \cup S_V$ with $S \cap W = \emptyset$ in the causal ordering graph $\mathrm{CO}(\mathcal{B})$. The subgraph of $\mathcal{B}_{\mathrm{do}(S_F,S_V)}$ induced by $(V \cup F) \setminus W$ is self-contained.*

Theorem 4.6 shows how the causal ordering graph can be used to read off the (generic) effects and non-effects of *perfect interventions on clusters* in the complete and overcomplete sets of the associated bipartite graph under the assumption of unique solvability with respect to the complete and overcomplete sets in the causal ordering graph.

Theorem 4.6. *Let $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} = \langle V, F, E\rangle\rangle$ be a system of constraints with coarse decomposition $\mathrm{CD}(\mathcal{B}) = \langle T_I, T_C, T_O\rangle$. Assume that $\mathcal{M}$ is maximally uniquely solvable w.r.t. $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E}\rangle$ and let $\mathbf{X}^*$ be a solution of $\mathcal{M}$. Let $S_F \subseteq (T_C \cup T_O) \cap F$ and $S_V \subseteq (T_C \cup T_O) \cap (V \setminus W)$ be such that $(S_F \cup S_V) \in \mathcal{V}$. Consider the intervened system $\mathcal{M}_{\mathrm{do}(S_F,S_V,\boldsymbol{\xi}_{S_V})}$ with coarse decomposition $\mathrm{CD}(\mathcal{B}_{\mathrm{do}(S_F,S_V)}) = \langle T'_I, T'_C, T'_O\rangle$. Let $\mathbf{X}'$ be a solution of $\mathcal{M}_{\mathrm{do}(S_F,S_V,\boldsymbol{\xi}_{S_V})}$. If there is no directed path from any $x \in S_V$ to $v \in (T_C \cup T_O) \cap V$ in $\mathrm{CO}(\mathcal{B})$ then $X^*_v = X'_v$ almost surely. On the other hand, if there is $x \in S_V$ such that there is a directed path from x to $v \in (T_C \cup T_O) \cap V$ in $\mathrm{CO}(\mathcal{B})$, then X^*_v may have a different distribution than X'_v.*

One way to determine whether a perfect intervention has an effect on a certain variable is to explicitly solve the system of constraints before and after the intervention and check which solution components are altered. In particular, when the distribution of a solution component is different for almost all parameter values, then we say that there is a generic effect. This way, we can establish the generic effects of a perfect intervention without solving the equations by relying on a solvability assumption. Example 4.14 illustrates this notion of perfect intervention on the system of constraints for the filling bathtub that we first introduced in Example 4.1 and shows how the generic effects and non-effects of perfect interventions on clusters can be read off from the causal ordering graph.

Example 4.14. Recall the system of constraints $\mathcal{M}$ for the filling bathtub at equilibrium in Section 4.4.3. Consider the perfect interventions $\mathrm{do}(f_P, v_D, \xi_D)$, $\mathrm{do}(f_D, v_O, \xi_O)$,

Table 4.2: Solutions for system of constraints describing the bathtub system in Section 4.4.3 without interventions (i.e. the observed system) and after perfect interventions $\mathrm{do}(f_P, v_D, \xi_D)$, $\mathrm{do}(f_D, v_O, \xi_O)$, and $\mathrm{do}(f_D, v_D, \xi_D)$.

	observed	$\mathrm{do}(f_P, v_D, \xi_D)$	$\mathrm{do}(f_D, v_O, \xi_O)$	$\mathrm{do}(f_D, v_D, \xi_D)$
$X^*_{v_K}$	U_{w_K}	U_{w_K}	U_{w_K}	U_{w_K}
$X^*_{v_I}$	U_{w_I}	U_{w_I}	U_{w_I}	U_{w_I}
$X^*_{v_P}$	$\frac{U_{w_I}}{(U_{w_4}U_{w_K})}$	$\frac{U_{w_I}}{(U_{w_4}U_{w_K})}$	$\frac{\xi_O}{(U_{w_4}U_{w_K})}$	$gU_{w_2}\xi_D$
$X^*_{v_O}$	U_{w_I}	U_{w_I}	ξ_O	$U_{w_4}U_{w_K}gU_{w_2}\xi_D$
$X^*_{v_D}$	$\frac{U_{w_I}}{(U_{w_4}U_{w_K}gU_{w_2})}$	ξ_D	$\frac{\xi_O}{(U_{w_4}U_{w_K}gU_{w_2})}$	ξ_D

Table 4.3: The effects of perfect interventions on clusters of variables and constraints in the causal ordering graph for the bathtub system in Figure 4.9 obtained by Theorem 4.6. Since $\{f_D, v_D\}$ is not a cluster in the causal ordering graph, the effects of this intervention cannot be read off from the causal ordering graph.

target	generic effect	non-effect
f_K, v_K	$X_{v_K}, X_{v_P}, X_{v_D}$	X_{v_I}, X_{v_O}
f_I, v_I	$X_{v_I}, X_{v_P}, X_{v_O}, X_{v_D}$	X_{v_K}
f_P, v_D	X_{v_D}	$X_{v_K}, X_{v_I}, X_{v_P}, X_{v_O}$
f_O, v_P	X_{v_P}, X_{v_D}	$X_{v_K}, X_{v_I}, X_{v_O}$
f_D, v_O	$X_{v_P}, X_{v_O}, X_{v_D}$	X_{v_K}, X_{v_I}
$f_P, f_D, f_O, v_P, v_D, v_O$	$X_{v_P}, X_{v_O}, X_{v_D}$	X_{v_K}, X_{v_I}

and $\mathrm{do}(f_D, v_D, \xi_D)$. These interventions model experiments that can, at least in principle, be conducted in practice:

(i) The intervention $\mathrm{do}(f_P, v_D, \xi_D)$ replaces the constraint f_P by a constraint that sets the water level X_{v_D} equal to a constant and leaves all other constraints unaffected. This could correspond to an experimental set-up where the constant g in the constraint Φ_{f_P} is controlled by accelerating and decelerating the bathtub system precisely in such a way that the water level X_{v_D} is forced to take on a constant value ξ_D both in time and across the ensemble of bathtubs. We observe the system once it has reached equilibrium.

(ii) The interventions $\mathrm{do}(f_D, v_O, \xi_O)$ and $\mathrm{do}(f_D, v_D, \xi_D)$ may correspond to an experiment where a hose is added to the system that can remove or add water precisely in such a way that either the outflow rate X_{v_O} or the water level X_{v_D} is kept at a constant level both in time and across the ensemble of bathtubs. The system is observed when it has reached equilibrium.

Note that the cluster $\{f_D, v_D\}$ is not a cluster in the causal ordering graph in Figure 4.9. However, the system of constraints $\mathcal{M}_{\mathrm{do}(f_D,v_D,\xi_D)}$ is maximally uniquely solvable with respect to the causal ordering graph $\mathrm{CO}(\mathcal{B}_{\mathrm{do}(f_D,v_D)})$, and therefore the effects of the intervention are well-defined.[21] By explicit calculation we obtain the (unique) solutions in Table 4.2 for the observed and intervened bathtub systems. By comparing with the solutions in the observed column we read off that the perfect intervention $\mathrm{do}(f_P, v_D, \xi_D)$ does not change the solution for the variables $X_{v_K}, X_{v_I}, X_{v_P}, X_{v_O}$, but it generically does change the solution for X_{v_D}. We further find that $\mathrm{do}(f_D, v_D, \xi_D)$ and $\mathrm{do}(f_D, v_O, \xi_D)$ affect the solution for the variables $X_{v_P}, X_{v_O}, X_{v_D}$ but not of X_{v_K} and X_{v_I}.

The causal ordering graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ for the bathtub system is given in Figure 4.9. It has clusters $\mathcal{V} = \{\{f_K, v_K\}, \{f_I, v_I\}, \{f_P, v_D\}, \{f_O, v_P\}, \{f_D, v_O\}\}$. Under the assumption that the (intervened) system is maximally uniquely solvable w.r.t. its causal ordering graph, we can apply Theorem 4.6 and read off the generic effects and non-effects of perfect interventions on clusters, which are presented in Table 4.3. This illustrates the fact that we can establish the generic effects and non-effects of the perfect interventions $\mathrm{do}(f_P, v_D, \xi_D)$ and $\mathrm{do}(f_D, v_O, \xi_O)$, which act on clusters in the causal ordering graph, without explicitly solving the system of equations. We will discuss differences between causal implications of the causal ordering graph and the graph of the SCM in Figure 4.10(b) in Section 4.6. △

4.5.3 Interventions commute with causal ordering

Given a system of constraints we can obtain the causal ordering graph after a perfect intervention on one of its clusters in the original causal ordering graph by running the causal ordering algorithm on the bipartite graph in the intervened system of constraints. In this section we will define an operation of "perfect intervention" directly on the clusters in a causal ordering graph and show that the causal ordering graph that is obtained after a perfect intervention coincides with the causal ordering graph of the intervened system (i.e. perfect interventions on clusters in the causal ordering graph commute with the causal ordering algorithm). Roughly speaking, a perfect intervention on a cluster in a directed cluster graph removes all incoming edges to that cluster and separates all variable vertices and constraint vertices in the targeted cluster into separate clusters in a specified way.

Definition 4.13. Let $\mathcal{B} = \langle V, F, E \rangle$ be a bipartite graph and W a set of exogenous variables. Let $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ be the corresponding causal ordering graph and consider $S \in \mathcal{V}$ with $S \cap W = \emptyset$. Let $S_F = \langle f_i : i = 1, \dots, n \rangle$ and $S_V = \langle v_i : i = 1, \dots, n \rangle$

[21] The intervention on $\{f_D, v_D\}$ is interesting because it removes the constraint that the water flowing through the faucet X_{v_I} must be equal to the water flowing through the drain X_{v_O}. This can be accomplished by adding a hose to the system through which additional water can flow in and out of the bathtub to ensure that X_{v_D} remains at a constant level. Notice that, in this example, the *total* inflow and *total* outflow of water remain equal, while the inflow *through the faucet* and the outflow *through the drain* may differ.

with $n = |S \cap V| = |S \cap F|$ be tuples consisting of all the vertices in $S \cap F$ and $S \cap V$ respectively. A *perfect intervention* $\mathrm{do}(S_F, S_V)$ *on a cluster* $\{S_F, S_V\}$ results in the directed cluster graph $\mathrm{CO}(\mathcal{B})_{\mathrm{do}(S_F,S_V)} = \langle \mathcal{V}', \mathcal{E}' \rangle$ where[22]

(i) $\mathcal{V}' = (\mathcal{V} \setminus \{S\}) \cup \{\{v_i, f_i\} : i = 1, \ldots, n\}$,

(ii) $\mathcal{E}' = \{(x \to T) \in \mathcal{E} : T \neq S\}$.

A soft intervention on a system of constraints has no effect on the bipartite graphical structure of the constraints and the variables that appear in them. Since the bipartite graph of the system is the same before and after soft interventions, it trivially follows that soft interventions commute with causal ordering. The following proposition shows that perfect interventions on clusters also commute with causal ordering.

Proposition 4.2. *Let $\mathcal{B} = \langle V, F, E \rangle$ be a bipartite graph and W a set of exogenous variables. Let $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ be the corresponding causal ordering graph. Let $S_F \subseteq F$ and $S_V \subseteq V \setminus W$ be such that $(S_F \cup S_V) \in \mathcal{V}$. Then:*

$$\mathrm{CO}(\mathcal{B}_{\mathrm{do}(S_F,S_V)}) = \mathrm{CO}(\mathcal{B})_{\mathrm{do}(S_F,S_V)}.$$

The bipartite graph in Figure 4.12 (a) has the causal ordering graph depicted in Figure 4.12 (b). The perfect intervention $\mathrm{do}(S_F, S_V)$ with $S_F = \langle f_2, f_3 \rangle$ and $S_V = \{v_2, v_3\}$ on this causal ordering graph results in the directed cluster graph in Figure 4.12 (d). Since perfect interventions on clusters commute with causal ordering, this graph can also be obtained by applying the causal ordering algorithm to the intervened bipartite graph in Figure 4.12 (c). Proposition 4.2 shows that perfect interventions on the graphical level can be used to draw conclusions about dependencies and causal implications of the underlying intervened system of constraints. We will use this result in Section 4.6.3 to elucidate the commutation properties of equilibration and interventions in dynamical models as defined in Dash (2005) and Bongers and Mooij (2018).

4.6 Discussion

In this section we give a detailed account of how our work relates to some of the existing literature on causal ordering and causal modelling.

4.6.1 "The causal graph": A misnomer?

Our work extends the work of Simon (1953) who introduced the causal ordering algorithm. We extensively discussed the example of a bathtub that first appeared in Iwasaki et al. (1994), in which the authors refer to the Markov ordering graph as "the

[22] A perfect intervention $\mathrm{do}(S_F, S_V, \boldsymbol{\xi}_V)$ replaces constraints Φ_{f_i} with causal constraints $\Phi'_{f_i} = \langle X_{v_i} \mapsto X_{v_i}, \xi_{v_i}, \{v_i\} \rangle$. Notice that the labels f_i of the constraints are unaltered, and therefore only the edges in the bipartite graph and causal ordering graph change after an intervention, as well as the clusters in the causal ordering graph, while the labels of vertices are preserved.

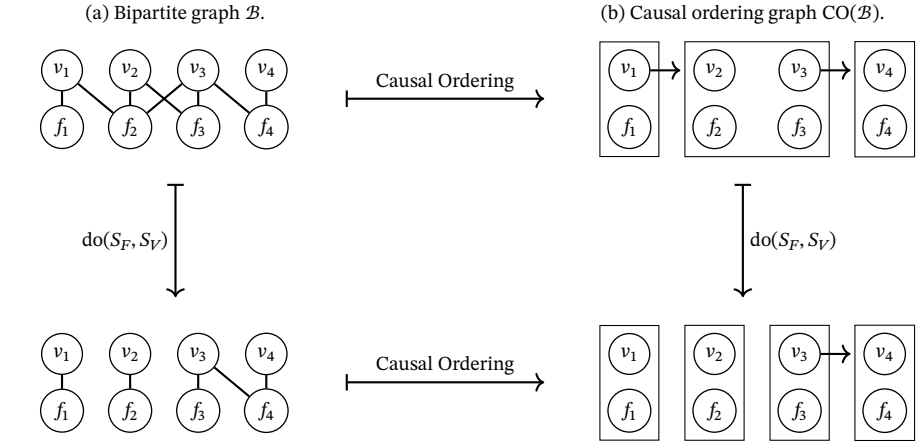


Figure 4.12: The intervention do(S_F, S_V) with ordered sets $S_F = \langle f_2, f_3 \rangle$ and $S_V = \langle v_2, v_3 \rangle$ commutes with causal ordering. Application of causal ordering and the intervention to the bipartite graph (a) results in the causal ordering graph (b) and the intervened bipartite graph (c) respectively. The directed cluster graph (d) can be obtained either by applying causal ordering to the intervened bipartite graph or by intervening on the causal ordering graph.

causal graph" and claim that this graph represents the effects of "manipulations". We observe here that the Markov ordering graph in Figure 4.10(a) does not have an unambiguous causal interpretation, contrary to claims in the literature. In this work we have formalized soft and perfect interventions, which are two common types of manipulation. This allows us to show that the Markov ordering graph, unlike the causal ordering graph, neither represents the effects of soft interventions nor does it have a straightforward interpretation in terms of perfect interventions. Iwasaki et al. (1994) do not clarify what the correct causal interpretation of the Markov ordering graph should be and therefore we believe that the term "causal graph" is a misnomer from a contemporary perspective on interventions and causality.

Markov ordering. To support this claim, we consider the bathtub system in Iwasaki et al. (1994) that we presented in Example 4.1. The structure of the equations and the endogenous variables that appear in them can be represented by the bipartite graph in Figure 4.13(a). The corresponding Markov ordering graph in Figure 4.13(c) corresponds to the graph that Iwasaki et al. (1994) call the "causal graph" for the bathtub system. Note that Iwasaki et al. (1994) do not make a distinction between variable vertices and equation vertices like we do. Their "causal graph" therefore has vertices K, I, P, O, D instead of v_K, v_I, v_P, v_O, v_D. An aspect that is not discussed at all by Iwasaki et al. (1994), is that the Markov ordering graph implies conditional

independences between components of solutions of equations.[23]

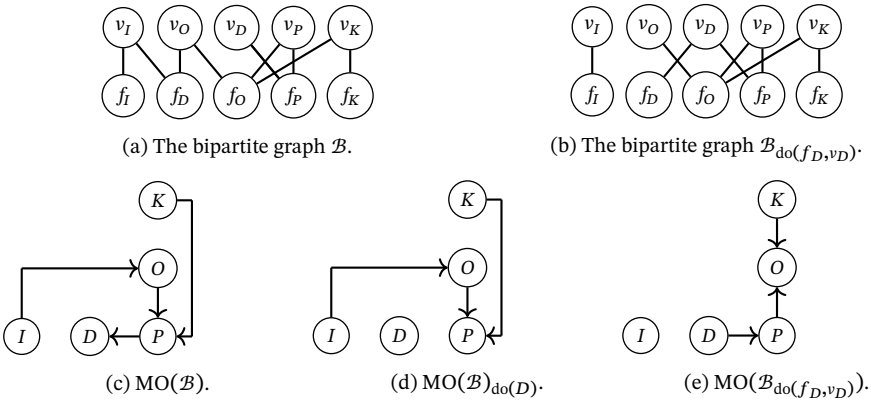
Soft interventions. We first consider the representation of soft interventions. Table 4.1 shows that a soft intervention on f_D has a generic effect on the solution for the variables v_P, v_O, and v_D. This soft intervention cannot be read off from the Markov ordering graph in Figure 4.13(c) because there is no vertex f_D. Since Iwasaki et al. (1994) make no distinction between variable vertices and equation vertices, a manipulation on D should perhaps be interpreted as a soft intervention on the vertex D in the Markov ordering graph in Figure 4.13(c) instead. However, the graphical structure would lead us to erroneously conclude that the soft intervention on D *only* has an effect on the variable D. In earlier work, Simon and Iwasaki (1988) assumed that a matching between variable and equation vertices is known in advance, allowing them to read off effects of soft interventions. We conclude that the Markov ordering graph, by itself, does not represent the effects of soft interventions on equations in general.

Perfect interventions. In Example 4.14 we found that a perfect intervention $\text{do}(f_D, v_D, \xi_D)$ has an effect on the solution of the variables v_P, v_O and v_D. If we would interpret this manipulation as a perfect intervention on D in the Markov ordering graph in Figure 4.13(c) then we would mistakenly find that this intervention only affects the variable D. Since Iwasaki et al. (1994) do not make a distinction between variable vertices and equation vertices we could also interpret a manipulation on D as the perfect intervention $\text{do}(f_P, v_D, \xi_D)$ or $\text{do}(f_D, v_O, \xi_O)$. From Table 4.2 we see that these perfect interventions would change the solution of the variables $\{v_D\}$ and $\{v_P, v_O, v_D\}$ respectively. Only the perfect intervention $\text{do}(f_P, v_D, \xi_D)$ which targets the cluster containing v_D corresponds to a perfect intervention on D in the Markov ordering graph in Figure 4.13(c). Since it is not clear from the Markov ordering graph what type of experiment a perfect intervention on one of its vertices should correspond to, we conclude that the Markov ordering graph cannot be used to read off the effects of perfect interventions.

Causal ordering graph. The causal ordering graph for the bathtub system is given in Figure 4.1(b). We proved that the causal ordering graph, contrary to the Markov ordering graph, represents the effects of soft interventions on equations and perfect interventions on clusters (see Theorems 4.3 and 4.6). To derive causal implications from sets of equations we therefore propose to use the notion of the causal ordering graph instead. The distinction between variable vertices and equation vertices is also made by Simon (1953) who shows how, for linear systems of equations, the principles of causal ordering can be used to qualitatively assess the effects of soft interventions on equations. A different, but closely related, notion of the causal ordering graph is

[23] Iwasaki et al. (1994) consider deterministic systems of equations and therefore it would not have made sense to consider Markov properties. In earlier work, the vanishing partial correlations implied by linear systems with three variables and normal errors were studied by Simon (1954).

used by Hautier et al. (2004) in the context of control systems modelling.



(a) The bipartite graph $\mathcal{B}$.

(b) The bipartite graph $\mathcal{B}_{\mathrm{do}(f_D, v_D)}$.

(c) $\mathrm{MO}(\mathcal{B})$.

(d) $\mathrm{MO}(\mathcal{B})_{\mathrm{do}(D)}$.

(e) $\mathrm{MO}(\mathcal{B}_{\mathrm{do}(f_D, v_D)})$.

Figure 4.13: The bipartite graph for the bathtub system without exogenous variables is given in Figure (a). The intervened bipartite graph is given in Figure (b). The Markov ordering graphs for the observed and intervened bathtub system are given in Figures (c) and (e) respectively. Figure (d) shows the graph that we obtain by intervening on the Markov ordering graph. Note that this does not correspond with the Markov ordering graph of the intervened bathtub system in Figure 4.13(e).

4.6.2 Relation to other causal models

The results in this work are easily applicable to other modelling frameworks, such as the popular SCM framework (Bongers, Forré, et al., 2020; Pearl, 2009). Application of causal ordering to the structural equations of an SCM with self-cycles may result in a different ordering than the one implied by the SCM. In particular, causal ordering may lead to a stronger Markov property and a representation of effects of a different set of (perfect) interventions. Even though the causal ordering graph itself may not allow us to read off non-effects of arbitrary perfect interventions, one can still obtain those by first intervening on the bipartite graph, then applying the causal ordering algorithm, and finally reading off the descendants of the intervention targets (under appropriate maximal unique solvability conditions).

Structural Causal Models. In an SCM, each endogenous variable is on the left-hand side of exactly one structural equation and perfect interventions always act on a structural equation and its corresponding variable. In comparison, a system of constraints consists of symmetric equations and the asymmetric relations between variables are derived automatically by the causal ordering algorithm. Consider, for example, the following structural equations:

$$X_1 = U_1 \tag{4.18}$$

$$X_2 = aX_1 + U_2, \tag{4.19}$$

where X_1, X_2 are endogenous variables, U_1, U_2 are exogenous random variables, and a is a constant. The ordering $X_1 \to X_2$ can also be obtained by causal ordering of the following set of equations:

$$X_1 - U_1 = 0, \tag{4.20}$$
$$X_2 - aX_1 - U_2 = 0. \tag{4.21}$$

Note that any set of structural equations implies a self-contained set of equations.[24] We can thus always apply the causal ordering algorithm to structural equations. Interestingly, since the output of the causal ordering algorithm is unique (see Theorem 4.1), the structure that is provided by the structural equations is actually redundant if the structural equations contain no cycles.

SCM for the bathtub. Recall that at equilibrium, the bathtub system can be described by the following structural equations:

$$\begin{aligned} f_K &: \quad X_{v_K} = U_{w_K}, & f_O &: \quad X_{v_O} = U_{w_5} X_{v_K} X_{v_P}, \\ f_I &: \quad X_{v_I} = U_{w_I}, & f_D &: \quad X_{v_D} = X_{v_D} + U_{w_1}(X_{v_I} - X_{v_O}), \\ f_P &: \quad X_{v_P} = g U_{w_3} X_{v_D}. \end{aligned}$$

The graph of this SCM is depicted in Figure 4.10(b), and the descendants and non-descendants of vertices are given in Table 4.4. Can we use this table to read off generic causal effects of perfect interventions targeting $\{f_K, v_K\}$, $\{f_I, v_I\}$, $\{f_P, v_P\}$, $\{f_O, v_O\}$, and $\{f_D, v_D\}$? The graph of the SCM contains (self-)cycles and the SCM does not have a (unique) solution under each of these perfect interventions.[25] Therefore, the graph of this SCM may not have a straightforward causal interpretation. Indeed, Bongers, Forré, et al. (2020) pointed out that for SCMs with cycles or self-cycles, the absence (presence) of directed edges and directed paths between vertices may not correspond one-to-one to the absence (generic presence) of direct and indirect causal effects, as it does in DAGs. (Self-)cycles may even lead to (in)direct causal effects without a corresponding directed edge or path being present in the graph of the SCM. For the bathtub example, that unusual behaviour does not occur, but instead it illustrates another behaviour: certain causal effects are absent, even though one would naïvely expect these to be generically present based on the graph of the SCM.[26] For example,

[24] In a set of structural equations each variable is matched to a single equation. Since the set of equations has a perfect matching it is self-contained by Hall's marriage theorem (see Theorem 4.7 in Appendix 4.B.4).

[25] There is no (unique) solution if one fixes the outflow rate of the system X_{v_O} to a value that is not equal to X_{v_I} for the perfect interventions targeting $\{f_O, v_O\}$ and $\{f_P, v_P\}$. In the dynamical model for the bathtub, these perfect interventions would correspond with the water level becoming (plus or minus) infinity.

[26] Such behaviour is a characteristic of perfectly adaptive dynamical systems Blom and Mooij, 2021. The causal and probabilistic properties of these systems are the topic of Chapter 5 of this thesis.

Table 4.4 shows that v_O is a descendant of v_K in the graph of the SCM while the solution for the outflow rate X_{v_O} does *not* change after the perfect intervention $\mathrm{do}(f_K, v_K)$. That this causal relation is absent can actually be read off from the causal ordering graph in Figure 4.1(b).

For the bathtub system, the causal ordering algorithm can exploit the fact that equation f_D can be replaced by $f'_D : 0 = U_{w_1}(X_{v_I} - X_{v_O})$, which does not involve v_D, whereas for the SCM this self-cycle cannot be removed. This causes the following differences in the results of the two approaches:

(i) The d-separations in the Markov ordering graph in Figure 4.10(a) imply more conditional independences than those implied by the σ-separations in the graph of the SCM in Figure 4.10(b) (as was discussed in detail in Section 4.4.3).

(ii) The graph of the SCM and the causal ordering graph represent different perfect intervention targets. In the graph of the SCM, we have minimal perfect intervention targets of the form $\{f_i, v_i\}$ with $i \in \{K, I, P, O, D\}$, while the causal ordering graph represents minimal perfect interventions on clusters $\{f_K, v_K\}$, $\{f_I, v_I\}$, $\{f_P, v_D\}$, $\{f_O, v_P\}$, and $\{f_D, v_O\}$. In both cases, the set of all perfect intervention targets that are represented by the graph are obtained by taking unions of minimal perfect intervention targets.

(iii) The causal ordering graph of the bathtub has a straightforward causal interpretation because the bathtub system still has a unique solution under interventions on clusters in the causal ordering graph. In contrast, the graph of the SCM for the bathtub system does not have a straightforward causal interpretation and the bathtub system does not have a solution under each perfect intervention on the SCM.

We conclude that the causal ordering approach yields a more "faithful" representation of the bathtub than the SCM framework.

Other frameworks. Since the causal ordering algorithm can be applied to any set of equations, the results that we developed here are generally applicable to sets of equations in other modelling frameworks. For example, we introduced Causal Constraint Models (CCMs) in Chapter 3 without a graphical representation for the independence structure between the variables. The causal ordering algorithm can be directly applied to a set of active constraints to obtain a Markov ordering graph.

4.6.3 Equilibration in dynamical models

In this subsection we will discuss in more detail the relation between our work and other closely related work, in particular to that of Dash (2005).

Dynamical models in terms of first order differential equations can be *equilibrated to a set of equations* by equating each time-derivative to zero (Bongers and Mooij, 2018; Mooij, Janzing, and Schölkopf, 2013). They can be *equilibrated and mapped to a causal ordering graph* by applying the causal ordering algorithm to the resulting set of equilibrium equations. They can also be *equilibrated and mapped to*

Table 4.4: The descendants and non-descendants of intervention targets in the graph of the SCM for the bathtub system in Figure 4.10(b).

target	descendants	non-descendants
f_K, v_K	v_K, v_P, v_O, v_D	v_I
f_I, v_I	v_I, v_P, v_O, v_D	v_K
f_P, v_P	v_P, v_O, v_D	v_K, v_I
f_O, v_O	v_P, v_O, v_D	v_K, v_I
f_D, v_D	v_P, v_O, v_D	v_K, v_I

a Markov ordering graph by subsequently applying Definition 4.10 to this causal ordering graph. The bathtub system provides an example of what Dash (2005) calls a "violation of the Equilibration Manipulation Commutability property".[27]

Consider the dynamical system version of the filling bathtub, with dynamical equations

$$\begin{aligned}
f_K &: \quad X_{v_K} = U_{w_K}, \\
f_I &: \quad X_{v_I} = U_{w_I}, \\
f_D &: \quad \dot{X}_{v_D}(t) = U_{w_1}(X_{v_I}(t) - X_{v_O}(t)), \\
f_P &: \quad \dot{X}_{v_P}(t) = U_{w_2}(g\, U_{w_3} X_{v_D}(t) - X_{v_P}(t)), \\
f_O &: \quad \dot{X}_{v_O}(t) = U_{w_4}(U_{w_5} X_{v_K} X_{v_P}(t) - X_{v_O}(t)).
\end{aligned}$$

Equilibration yields the equilibrium equations f_K, f_I, f_D, f_P, and f_O in equations (4.1) to (4.5). It is clear in this particular case that any perfect intervention $\mathrm{do}(S_F, S_V, _V)$ (where we extended Definition 4.12 to dynamical equations) commutes with equilibration (substituting zeroes for all first-order derivatives).[28] This type of commutation relation actually holds also in more general settings (see Mooij, Janzing, and Schölkopf (2013) and Bongers and Mooij (2018)).

On the other hand, mapping a set of equations to the corresponding Markov ordering graph does not necessarily commute with perfect interventions. For example, for the perfect intervention $\mathrm{do}(f_D, v_D)$, the Markov ordering graphs $\mathrm{MO}(\mathcal{B})_{\mathrm{do}(v_D)}$ and

[27] We argue that this is confusing terminology in two ways. First, what Dash calls "equilibration" is what we would call equilibration to a set of equations, composed with the mapping to the Markov ordering graph. Second, Dash follows Iwasaki et al. (1994) in referring to the Markov ordering graph as the "causal graph". We argued in Section 4.6.1 that this is a misnomer, as in general there is no straightforward one-to-one correspondence between the Markov ordering graph and the causal semantics of the system. This terminological confusion explains the apparent contradiction with the result of Bongers and Mooij (2018), who prove that equilibration to an SCM commutes with manipulation (for perfect interventions).

[28] Note that it is crucially important here to ensure that the *labelling* of the equations is not changed by the equilibration operation.

$\mathrm{MO}(\mathcal{B}_{\mathrm{do}(f_D,v_D)})$ are wildly different, as can be seen by comparing Figures 4.13(d) and 4.13(e) respectively. Since perfect interventions do commute with equilibration, one can conclude that also the composition of equilibration followed by mapping to the Markov ordering graph fails to commute with this perfect intervention. This is the phenomenon that Dash (2005) pointed out.

This lack of commutability does not hold for *all* perfect interventions. For example, one can easily check that the perfect intervention $\mathrm{do}(f_K, v_K)$ commutes with the composition of equilibration followed by mapping to the Markov ordering graph. More generally, Proposition 4.2 tells us that for the bathtub, the clusters in the causal ordering graph ($\{f_K, v_K\}$, $\{f_I, v_I\}$, $\{f_P, v_D\}$, $\{f_O, v_P\}$, and $\{f_D, v_O\}$) represent the minimal perfect interventions targets for which both operations do commute. This means that of the perfect interventions that Dash (2005) considers ($\mathrm{do}(\{v_K, f_K\})$, $\mathrm{do}(\{v_I, f_I\})$, $\mathrm{do}(\{v_D, f_D\})$, $\mathrm{do}(\{v_O, f_O\})$, $\mathrm{do}(\{v_P, f_P\})$, and combinations thereof), exactly three commute with the mapping to the Markov ordering graph (namely $\mathrm{do}(\{f_K, v_K\})$, $\mathrm{do}(\{f_I, v_I\})$, $\mathrm{do}(\{f_P, v_P, f_O, v_O, f_D, v_D\})$, and combinations thereof). Hence, these are also the three minimal perfect interventions in that set that commute with equilibration followed by mapping to the Markov ordering graph.

As pointed out by Dash (2005), this lack of commutability has important implications when one tries to discover causal relations through structure learning, which we will briefly discuss in the next subsection.

4.6.4 Structure learning

We have shown that, under a solvability assumption, d-separations in the Markov ordering graph (or σ-separations in the directed graph associated with a particular perfect matching) imply conditional independences between variables in a system of constraints (see Theorem 4.3 and Theorem 4.4). Constraint-based causal discovery algorithms relate conditional independences in data to graphs under the Markov condition and the corresponding d- or σ-faithfulness assumption. Roughly speaking, the equivalence class of the Markov ordering graph (or the directed graph associated with a particular perfect matching) can be learned from data under the assumption that all conditional independences in the data are implied by the graph. The bathtub system in Example 4.1 is used by Dash (2005), who simulates data from the dynamical model until it reaches equilibrium, and then applies the PC-algorithm to learn the graphical structure of the system. It is no surprise that the learned structure is the Markov ordering graph in Figure 4.13(c). The usual assumption is then that the Markov ordering graph equals the causal graph, where directed edges express direct causal relations between variables. In this work we have shown that this learned Markov ordering graph does *not* have such a straightforward causal interpretation.

4.7 Conclusion

In this work, we reformulated Simon's causal ordering algorithm and demonstrated that it is a convenient and scalable tool to study causal and probabilistic aspects of models consisting of equations. In particular, we showed how the technique of causal ordering can be used to construct a *causal ordering graph* and a *Markov ordering graph* from a set of equations, without calculating explicit global solutions to the system of equations. The novelties of this paper include an extension of the causal ordering algorithm for general bipartite graphs, and proving that the corresponding Markov ordering graph implies conditional independences between variables, whereas the corresponding causal ordering graph encodes the effects of soft and perfect interventions.

To model causal relations between variables in sets of equations unambiguously, we generalized existing notions of perfect interventions on SCMs. The main idea is that a perfect intervention on a set of equations targets variables and specified equations, whereas a perfect intervention on a Structural Causal Model (SCM) targets variables and their associated structural equations. We considered a simple dynamical model with feedback and demonstrated that, contrary to claims in the literature, the Markov ordering graph does not generally have any obvious causal interpretation in terms of soft or perfect interventions. We showed that the causal ordering graph, on the other hand, does encode the effects of soft and certain perfect interventions. The main take-away is that we need to make a distinction between variables and equations in graphical representations of the probabilistic and causal aspects of models with feedback. By making this distinction, we clarified the correct interpretation of some existing results in the literature. Additionally, we shed new light on discussions in causal discovery about the justification of using a single directed graph with endogenous variables as vertices to simultaneously represent causal relations and conditional independences. We believe that the phenomenon where the Markov ordering graph does not encode causal semantics in the usual way manifests itself in certain biological or econometric models with feedback at equilibrium. In Chapter 5 and Chapter 6 we investigate these occurrences further.

4.A Preliminaries

4.A.1 Graph terminology

A *bipartite graph* is an ordered triple $\mathcal{B} = \langle V, F, E\rangle$ where V and F are disjoint sets of vertices and E is a set of undirected edges $(v - f)$ between vertices $v \in V$ and $f \in F$. For a vertex $x \in V \cup F$ we write $\mathrm{adj}_{\mathcal{B}}(x) = \{y \in V \cup F : (x - y) \in E\}$ for its *adjacencies*, and for $X \subseteq V \cup F$ we write $\mathrm{adj}_{\mathcal{B}}(X) = \bigcup_{x \in X} \mathrm{adj}_{\mathcal{B}}(x)$ to denote the adjacencies of X in $\mathcal{B}$. A *matching* $\mathcal{M} \subseteq E$ for a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ is a subset of edges that have no common endpoints. We say that two vertices x and y are *matched* when

$(x - y) \in \mathcal{M}$. We let $\mathcal{M}(x)$ denote the set of vertices to which x is matched. Note that if $(x - y) \in \mathcal{M}$ then $\mathcal{M}(x) = \{y\}$ and if x is not matched then $\mathcal{M}(x) = \emptyset$. We let $\mathcal{M}(X) = \bigcup_{x \in X} \mathcal{M}(x)$ denote the set of vertices to which the set of vertices $X \subseteq V \cup F$ is matched. A matching is *perfect* if all vertices $V \cup F$ are matched.

A *directed graph* is an ordered pair $\mathcal{G} = \langle V, E \rangle$ where V is a set of vertices and E is a set of directed edges $(v \to w)$ between vertices $v, w \in V$. A *directed mixed graph* is an ordered triple $\mathcal{G} = \langle V, E, B \rangle$ where $\langle V, E \rangle$ is a directed graph and B is a set of bi-directed edges between vertices in V. If a directed mixed graph $\langle V, E, B \rangle$ has an edge $(v \to v) \in E$ then we say that it has a *self-cycle*. We say that a vertex v is a *parent* of w if $(v \to w) \in E$ and write $v \in \mathrm{pa}_{\mathcal{G}}(w)$. Similarly we say that w is a *child* of v if $(v \to w) \in E$ and write $w \in \mathrm{ch}_{\mathcal{G}}(v)$. A *path* is a sequence of distinct vertices and edges $(v_1, e_1, v_2, e_2, \ldots, e_{n-1}, v_n)$ where for $i = 1, \ldots, n-1$ we have that $e_i = (v_i \to v_{i+1})$, $e_i = (v_i \leftarrow v_{i+1})$, or $e_i = (v_i \leftrightarrow v_{i+1})$. The path is called *open* if there is no $v_i \in \{v_2, \ldots v_{n-1}\}$ such that there are two arrowheads at v_i on the path (i.e. there is no collider on the path). A *directed path* $(v \to \ldots \to w)$ from v to w is a path where all arrowheads point in the direction of w. We say that v is an *ancestor* of w if there is a directed path from v to w and write $v \in \mathrm{an}_{\mathcal{G}}(w)$. We say that w is a *descendant* of v if there is a directed path from v to w and write $w \in \mathrm{de}_{\mathcal{G}}(v)$.

Let $\mathcal{G} = \langle V, E, B \rangle$ be a directed mixed graph and consider the relation:

$$v \sim w \iff w \in \mathrm{an}_{\mathcal{G}}(v) \cap \mathrm{de}_{\mathcal{G}}(v) = \mathrm{sc}_{\mathcal{G}}(v).$$

Since the relation is reflexive, symmetric, and transitive this is an equivalence relation. The equivalence classes $\mathrm{sc}_{\mathcal{G}}(v)$ are called the *strongly connected components* of $\mathcal{G}$. A directed graph without self-cycles is *acyclic* if and only if all of its strongly connected components are singletons. A directed graph with no directed cycles is called a Directed Acyclic Graph (DAG).

A *perfect intervention* $\mathrm{do}(I)$ on a directed mixed graph $\mathcal{G} = \langle V, E, B \rangle$ removes all edges with an arrowhead at any of the nodes $i \in I \subseteq V$. That is, $\mathcal{G}_{\mathrm{do}(I)} = \langle V, E', B' \rangle$ where $E' = \{(x \to y) \in E : y \notin I\}$ and $B' = \{(x \leftrightarrow y) \in B : x \notin I, y \notin I\}$. *Marginalizing* out a set of nodes $W \subseteq V$ from a directed mixed graph $\mathcal{G} = \langle V, E, B \rangle$ results in a directed mixed graph $\mathcal{G}_{\mathrm{mar}(W)} = \langle V \setminus W, E_{\mathrm{mar}(W)}, B_{\mathrm{mar}(W)} \rangle$ (also known as the *latent projection*) where:

(i) $E_{\mathrm{mar}(W)}$ consists of edges $(x \to y)$ such that $x, y \in V \setminus W$ and there exist $w_1, \ldots, w_k \in W$ such that the directed path $x \to w_1 \to \ldots \to w_k \to y$ is in $\mathcal{G}$.

(ii) $B_{\mathrm{mar}(W)}$ consists of edges $(x \leftrightarrow y)$ such that $x, y \in V \setminus W$ and there exist $w_1, \ldots, w_k \in W$ such that at least one of the following paths is in $\mathcal{G}$: (i) $x \leftrightarrow y$, or (ii) $x \leftarrow w_1 \leftarrow \ldots \leftarrow w_i \to \ldots \to w_k \to y$, or (iii) $x \leftarrow w_1 \leftarrow \ldots \leftarrow w_i \leftrightarrow w_{i+1} \to \ldots \to w_k \to y$.

The operations of marginalization and intervention commute (Forré et al., 2017).

4.A.2 Graph separation and Markov properties

In the literature, several versions of Markov properties for graphical models and corresponding probability distributions have been put forward, see e.g. Forré et al. (2017), Lauritzen, Dawid, et al. (1990), Pearl (2009), and Spirtes, Glymour, et al. (2000). For DAGs and Acyclic Directed Mixed Graphs (ADMGs), the d-separation criterion is often used to relate conditional independences between variables in a model to the underlying (acyclic) graphical structure of the model (Pearl, 2009). For graphs that contain cycles the 'collapsed graph' representation of Spirtes and Richardson (1995) inspired Forré et al. (2017) to introduce the σ-separation criterion.

Definition 4.14. For a directed mixed graph $\mathcal{G} = \langle V, E, B\rangle$ we say that a path $(v_1, \dots, v_n)$ *is σ-blocked by $Z \subseteq V$* if

(i) $v_1 \in Z$ and/or $v_n \in Z$, or

(ii) there is a vertex $v_i \notin \mathrm{an}_{\mathcal{G}}(Z)$ on the path such that the adjacent edges both have an arrowhead at v_i, or

(iii) there is a vertex $v_i \in Z$ on the path such that: $v_i \to v_{i+1}$ with $v_{i+1} \notin \mathrm{sc}_{\mathcal{G}}(v_i)$, or $v_{i-1} \leftarrow v_i$ with $v_{i-1} \notin \mathrm{sc}_{\mathcal{G}}(v_i)$, or both.

The path is *d-blocked by Z* if it is σ-blocked or if there is a vertex $v_i \in Z$ on the path such that at least one of the adjacent edges does not have an arrowhead at v_i. We say that $X \subseteq V$ and $Y \subseteq V$ are *σ-separated* by $Z \subseteq V$ if every path in $\mathcal{G}$ with one end-vertex in X and one end-vertex in Y is σ-blocked by Z, and write

$$X \overset{\sigma}{\underset{\mathcal{G}}{\perp}} Y \,|\, Z.$$

If every such path is d-blocked by Z then we say that *X and Y are d-separated by Z*, and write

$$X \overset{d}{\underset{\mathcal{G}}{\perp}} Y \,|\, Z.$$

It can be shown that σ-separation implies d-separation and that the two are equivalent for acyclic graphs (Forré et al., 2017). In general, d-separation does not imply σ-separation. The d-separations or σ-separations in a probabilistic graphical model may imply conditional independences via the Markov properties in Definition 4.15 below.

Definition 4.15. For a directed mixed graph $\mathcal{G} = \langle V, E, B\rangle$ and a probability distribution $\mathbb{P}_{\mathbf{X}}$ on a product $\mathfrak{X} = \bigotimes_{v\in V} \mathcal{X}_v$ of standard measurable spaces $\mathcal{X}_v$, we say that the pair $(\mathcal{G}, \mathbb{P}_{\mathbf{X}})$ satisfies the *directed global Markov property* if for all subsets $W, Y, Z \subseteq V$:

$$W \overset{d}{\underset{\mathcal{G}}{\perp}} Y \,|\, Z \implies \mathbf{X}_W \overset{}{\underset{\mathbb{P}_{\mathbf{X}}}{\perp\!\!\!\perp}} \mathbf{X}_Y \,|\, \mathbf{X}_Z.$$

The pair $(\mathcal{G}, \mathbb{P}_{\mathbf{X}})$ satisfies the *generalized directed global Markov property* if for all sub-

sets $W, Y, Z \subseteq V$:

$$W \overset{\sigma}{\underset{\mathcal{G}}{\perp}} Y \,|\, Z \implies \mathbf{X}_W \perp\!\!\!\perp_{\mathbb{P}_\mathbf{X}} \mathbf{X}_Y \,|\, \mathbf{X}_Z.$$

Since σ-separations imply d-separations but not the other way around, the generalized directed global Markov property is strictly weaker than the directed global Markov property (Bongers, Forré, et al., 2020). For acyclic SCMs the induced probability distribution on endogenous variables and the corresponding DAG satisfy the directed global Markov property (Lauritzen, Dawid, et al., 1990). The variables that solve a simple SCM obey the generalized directed global Markov property relative to the graph of the SCM (Bongers, Forré, et al., 2020), while d-separation is limited to more specific settings such as acyclic models, discrete variables, or continuous variables with linear relations (Forré et al., 2017). A comprehensive account of different Markov properties for graphical models is provided by Forré et al. (2017).

Constraint-based causal discovery algorithms require some additional faithfulness assumption. A probability distribution is *d-faithful* to a directed mixed graph when each conditional independence implies a d-separation in that graph. Similarly, a probability distribution is *σ-faithful* to a directed mixed graph when each conditional independence implies a σ-separation in that graph. In non-linear, non-discrete, cyclic settings the σ-faithfulness assumption is a natural extension of the notion of the common d-faithfulness assumption with σ-separation replacing d-separation. Under the additional assumption of causal sufficiency (i.e., no latent confounding variables), the NL-CCD algorithm was shown to be sound under the generalized directed Markov property and the weaker d-faithfulness assumption (Chapter 4 in Richardson (1996)). Recently, Forré et al. (2018), Mooij and Claassen (2020), and Mooij, Magliacane, et al. (2020) proved soundness for a variety of causal discovery algorithms under the generalized directed Markov property and the σ-faithfulness assumption. Strobl (2018) proved soundness of a causal discovery algorithm under the directed Markov property and the d-faithfulness assumption, allowing for latent confounding and selection bias.

4.B Proofs

In this section of the appendix, all proofs are provided.

4.B.1 Causal ordering via minimal self-contained sets

In this section we prove Theorem 4.1 below.

Theorem 4.1. *The output of Algorithm 1 is well-defined and unique.*

Lemma 4.4 below shows that the minimal self-contained sets in a self-contained bipartite graph are disjoint. Lemma 4.5 shows that the induced subgraph after

one iteration of Algorithm 1, with a self-contained bipartite graph as input, is self-contained. The minimal self-contained sets in the graph which are not used in the iteration are minimal self-contained sets of the induced subgraph. This shows that the output of Algorithm 1 is well-defined. We then use Lemma 4.4 and 4.5 to prove Lemma 4.6 which states that the output of Algorithm 1, with a self-contained bipartite graph as input, is unique. This implies that the output of Algorithm 1, which has an initialization that is uniquely determined by the specification of exogenous variables W, must also be unique.

Lemma 4.4. *Let $\mathcal{B} = \langle V, F, E\rangle$ be a self-contained bipartite graph. Let $\mathcal{S}_F$ be the set of minimal self-contained sets in $\mathcal{B}$. The sets in $\mathcal{S}_F$ are pairwise disjoint, and, likewise, the sets of adjacent nodes*

$$\mathcal{S}_V = \{\mathrm{adj}_{\mathcal{B}}(S) : S \in \mathcal{S}_F\},$$

of the minimal self-contained sets in $\mathcal{S}_F$ are pairwise disjoint.

Proof. Let $S_1 \subseteq F$ and $S_2 \subseteq F$ be non-empty distinct minimal self-contained sets in $\mathcal{S}_F$. For the sake of contradiction, assume that $S_1 \cap S_2 \neq \emptyset$. Since S_1 is minimal self-contained, we know that $S_1 \cap S_2 \subset S_1$ is not self-contained. Hence, by Definition 4.3, we have that

$$|S_1 \cap S_2| < |\mathrm{adj}_{\mathcal{B}}(S_1 \cap S_2)|. \tag{4.22}$$

Consider the following equations:

$$\begin{aligned}
&|\mathrm{adj}_{\mathcal{B}}(S_1)| + |\mathrm{adj}_{\mathcal{B}}(S_2)| - |S_1 \cap S_2| && (4.23)\\
&\quad = |S_1| + |S_2| - |S_1 \cap S_2| && (4.24)\\
&\quad = |S_1 \cup S_2|\\
&\quad \leq |\mathrm{adj}_{\mathcal{B}}(S_1 \cup S_2)| && (4.25)\\
&\quad = |\mathrm{adj}_{\mathcal{B}}(S_1) \cup \mathrm{adj}_{\mathcal{B}}(S_2)|\\
&\quad = |\mathrm{adj}_{\mathcal{B}}(S_1)| + |\mathrm{adj}_{\mathcal{B}}(S_2)| - |\mathrm{adj}_{\mathcal{B}}(S_1) \cap \mathrm{adj}_{\mathcal{B}}(S_2)|\\
&\quad \leq |\mathrm{adj}_{\mathcal{B}}(S_1)| + |\mathrm{adj}_{\mathcal{B}}(S_2)| - |\mathrm{adj}_{\mathcal{B}}(S_1 \cap S_2)|, && (4.26)
\end{aligned}$$

where equality (4.24) holds by condition (i) of Definition 4.3, since $\mathcal{B}$ is self-contained inequality (4.25) holds by condition (ii) of Definition 4.3, and inequality (4.26) holds because $\mathrm{adj}_{\mathcal{B}}(S_1 \cap S_2) \subseteq \mathrm{adj}_{\mathcal{B}}(S_1) \cap \mathrm{adj}_{\mathcal{B}}(S_2)$. It follows that

$$|S_1 \cap S_2| \geq |\mathrm{adj}_{\mathcal{B}}(S_1) \cap \mathrm{adj}_{\mathcal{B}}(S_2)| \geq |\mathrm{adj}_{\mathcal{B}}(S_1 \cap S_2)| \geq 0.$$

This is in contradiction with equation (4.22), and hence $S_1 \cap S_2 = \emptyset$. This implies that $|S_1 \cap S_2| = 0$ and therefore by the inequalities above we have that $|\mathrm{adj}_{\mathcal{B}}(S_1) \cap \mathrm{adj}_{\mathcal{B}}(S_2)| = 0$. Thus $\mathrm{adj}_{\mathcal{B}}(S_1) \cap \mathrm{adj}_{\mathcal{B}}(S_2) = \emptyset$. □

Lemma 4.5. *Let $\mathcal{B} = \langle V, F, E\rangle$ be a self-contained bipartite graph. Suppose that F has minimal self-contained sets $\mathcal{S}_F$. Let $\mathcal{B}'$ be the subgraph of $\mathcal{B}$ induced by*

$$V' := V \backslash \mathrm{adj}_{\mathcal{B}}(S), \quad \textit{and} \quad F' := F \backslash S,$$

with $S \in \mathcal{S}_F$. Then the following two properties hold:
(i) $\mathcal{B}'$ is self-contained, and
(ii) the sets in $\mathcal{S}_F \backslash \{S\}$ are minimal self-contained in $\mathcal{B}'$.

Proof. Let $S \in \mathcal{S}_F$ be a minimal self-contained subset in $\mathcal{B}$. Since $\mathcal{B}$ and S are self-contained we have that $|V| = |F|$ and $|S| = |\mathrm{adj}_{\mathcal{B}}(S)|$ respectively. Therefore

$$|V'| = |V \setminus \mathrm{adj}_{\mathcal{B}}(S)| = |V| - |\mathrm{adj}_{\mathcal{B}}(S)| = |F| - |S| = |F \setminus S| = |F'|.$$

This shows that condition (i) of Definition 4.3 is satisfied for $\mathcal{B}'$. Assume, for the sake of contradiction, that F' does not satisfy condition (ii) of Definition 4.3 in the induced subgraph $\mathcal{B}'$. Then there exists $S' \subseteq F'$ such that $|S'| > |\mathrm{adj}_{\mathcal{B}'}(S')|$. Consider the following equations:

$$\begin{aligned}
|S \cup S'| &= |S| + |S'| \\
&> |\mathrm{adj}_{\mathcal{B}}(S)| + |\mathrm{adj}_{\mathcal{B}'}(S')| \\
&= |\mathrm{adj}_{\mathcal{B}}(S)| + |\mathrm{adj}_{\mathcal{B}}(S')| - |\mathrm{adj}_{\mathcal{B}}(S) \cap \mathrm{adj}_{\mathcal{B}}(S')| \\
&= |\mathrm{adj}_{\mathcal{B}}(S) \cup \mathrm{adj}_{\mathcal{B}}(S')| \\
&= |\mathrm{adj}_{\mathcal{B}}(S \cup S')| \\
&\geq |S \cup S'|,
\end{aligned}$$

where the last inequality holds because $\mathcal{B}$ is self-contained by assumption. This is a contradiction, and we conclude that both conditions of Definition 4.3 are satisfied for $\mathcal{B}'$. This shows that $\mathcal{B}'$ is self-contained.

Let $S_1 \in \mathcal{S}_F$ and $S_2 \in S_F$ be two distinct minimal self-contained sets in $\mathcal{B}$. Suppose that $\mathcal{B}_1$ is a subgraph of $\mathcal{B}$ induced by $V \setminus \mathrm{adj}_{\mathcal{B}}(S_1)$ and $F \setminus S_1$. By Lemma 4.4 we know that $S_1 \cap S_2 = \emptyset$ and $\mathrm{adj}_{\mathcal{B}}(S_1) \cap \mathrm{adj}_{\mathcal{B}}(S_2) = \emptyset$. It follows that for all $S' \subseteq S_2$ we have that $\mathrm{adj}_{\mathcal{B}}(S') = \mathrm{adj}_{\mathcal{B}_1}(S')$. We find that

$$\begin{aligned}
|S_2| &= |\mathrm{adj}_{\mathcal{B}}(S_2)| = |\mathrm{adj}_{\mathcal{B}_1}(S_2)|, \\
|S'| &\leq |\mathrm{adj}_{\mathcal{B}}(S')| = |\mathrm{adj}_{\mathcal{B}_1}(S')|,
\end{aligned}$$

for all $S' \subseteq S_2$. This shows that S_2 satisfies the conditions of Definition 4.3 in the bipartite graph $\mathcal{B}_1$. Since S_2 has no non-empty strict subsets that are self-contained in $\mathcal{B}$ we have that S_2 has no non-empty strict subsets that are self-contained in $\mathcal{B}_1$. We conclude that S_2 is a minimal self-contained subset in $\mathcal{B}_1$. This shows that the

sets $\mathcal{S}_F \setminus \{S\}$ are minimal self-contained in $\mathcal{B}'$. □

Lemma 4.6. *Let $\mathcal{B} = \langle V, F, E\rangle$ be a self-contained bipartite graph. The output* $\mathrm{CO}(\mathcal{B})$ *of Algorithm 1 is unique.*

Proof. Suppose $\mathcal{G}_1 = \langle \mathcal{V}_1, \mathcal{E}_2\rangle$ and $\mathcal{G}_2 = \langle \mathcal{V}_2, \mathcal{E}_2\rangle$ are directed cluster graphs that are obtained by running Algorithm 1. Let $A = (1, 2, \dots, |\mathcal{V}_1|)$ be an ordered set that indicates the order in which clusters $S^{(a)}$ (with $a \in A$) are added to $\mathcal{V}_1$ in the first run of the algorithm. Similarly, $B = (1, 2, \dots, |\mathcal{V}_2|)$ is an ordered set that indicates the order in which clusters $T^{(b)}$ (with $b \in B$) are added to $\mathcal{V}_2$ in the second run of the algorithm. With a slight abuse of notation we define $\mathcal{B} \setminus (S^{(k)})_{k<i}$ as the subgraph of $\mathcal{B}$ induced by the nodes $(S^{(k)})_{k\geq i}$. Similarly, $\mathcal{B} \setminus (T^{(k)})_{k<i}$ denotes the subgraph of $\mathcal{B}$ induced by the nodes $(T^{(k)})_{k\geq i}$.

Intermediate result: We will prove that for $i \in (1, 2, \dots, |\mathcal{V}_1|)$ there exists $b_i \in B$ such that $S^{(i)} = T^{(b_i)}$ by induction.

Base case: The algorithm adds the cluster $S^{(1)}$ to $\mathcal{V}_1$ in the first step of the first run. Therefore, we know that the set of nodes $F \cap S^{(1)}$ must be minimal self-contained in $\mathcal{B}$. Let $1 \leq k \leq |\mathcal{V}_2|$ be arbitrary. By Lemma 4.5 it follows that $F \cap S^{(1)}$ is minimal self-contained in $\mathcal{B} \setminus (T^{(j)})_{j<k}$ provided $S^{(1)} \neq T^{(j)}$ for all $j < k$. Since $\mathcal{B}$ is finite, the minimal self-contained set $S^{(1)}$ must be chosen eventually, and hence there exists $b_1 \in B$ such that $S^{(1)} = T^{(b_1)}$.

Induction hypothesis: Let $1 \leq i < |\mathcal{V}_1|$ be arbitrary and assume that for all $j \leq i$ there exists $b_j \in B$ such that $S^{(j)} = T^{(b_j)}$. We want to show that there exists $b_{i+1} \in B$ such that $S^{(i+1)} = T^{(b_{i+1})}$.

Induction step: Let $B' = B \setminus (b_1, \dots, b_i) = (b'_1, \dots, b'_{|\mathcal{V}_2|-i})$ be an ordered set such that $b'_j \prec b'_{j+1}$ for all $j = 1, \dots, |\mathcal{V}_2| - (i + 1)$.

(i) In the second run of the algorithm, the cluster $T^{(b'_1)}$ is added to $\mathcal{V}_2$ right after the clusters $T^{(b_j)}$ with $b_j \prec b'_1$ are added to $\mathcal{V}_2$ and removed from the bipartite graph. Therefore, the set $F \cap T^{(b'_1)}$ is minimal self-contained in $\mathcal{B} \setminus (T^{(b_j)})_{j\leq i, b_j \prec b'_1}$. In the first run of the algorithm, the clusters $S^{(1)} = T^{(b_1)}, \dots, S^{(i)} = T^{(b_i)}$ are subsequently added to $\mathcal{V}_1$ and removed from the bipartite graph. Therefore, by Lemma 4.4 and Lemma 4.5, we have that $F \cap T^{(b'_1)}$ is minimal self-contained in $\mathcal{B}' = \mathcal{B} \setminus (T^{(b_j)})_{j\leq i} = \mathcal{B} \setminus (S^{(k)})_{k\leq i}$. Hence, both $F \cap T^{(b'_1)}$ and $F \cap S^{(i+1)}$ are minimal self-contained in $\mathcal{B}'$. Therefore, by Lemma 4.4 and Lemma 4.5, either $T^{(b'_1)} = S^{(i+1)}$ (in which case we are done) or $F \cap S^{(i+1)}$ is minimal self-contained in $\mathcal{B}' \setminus T^{(b'_1)}$.

(ii) Let $k \leq |\mathcal{V}_2| - i$ be arbitrary. By iteration of the argument in the previous step we find that $F \cap T^{(b'_k)}$ is minimal self-contained in $(\mathcal{B} \setminus (T^{(b_j)})_{j\leq i, b_j \prec b'_k}) \setminus (T^{(b'_j)})_{j<k}$ and hence in $\mathcal{B}' \setminus (T^{(b'_j)})_{j<k}$, so that either $T^{(b'_k)} = S^{(i+1)}$ or $F \cap S^{(i+1)}$ is minimal

self-contained in $\mathcal{B}' \setminus (T^{(b'_j)})_{j\leq k}$. Since the bipartite graph is finite, there exists $m \in 1, \ldots, |\mathcal{V}_2| - i$ such that $T^{(b'_m)} = S^{(i+1)}$. By definition of B' there exists $b_{i+1} \in B$ such that $S^{(i+1)} = T^{(b_{i+1})}$.

This proves that the clusters in $\mathcal{V}_1$ are also clusters in $\mathcal{V}_2$. By symmetry we find that the clusters $S^{(a)}$ in $\mathcal{V}_1$ and the clusters $T^{(b)}$ in $\mathcal{V}_2$ coincide. Since $\mathcal{V}_1 = \mathcal{V}_2$ it follows immediately from the construction of edges in the algorithm that $\mathcal{E}_1 = \mathcal{E}_2$ and hence $\mathcal{G}_1 = \mathcal{G}_2$. □

4.B.2 Coarse decomposition

For completeness, we include the proofs of the results in Pothen and Fan (1990) that are necessary to show that the output of the extended causal ordering algorithm (Algorithm 3) is unique. The presentation in this section is based on the exposition of Van Diepen (2019). In order to prove the statements in Lemma 4.2 and Proposition 4.1, we require additional results. Lemma 4.7 and 4.8 show that the incomplete, complete, and over-complete set are disjoint. The former uses the notion of an *augmented path* for a bipartite graph $\mathcal{B}$ and a matching M, which is an alternating path for M that starts and ends with an unmatched vertex.

Lemma 4.7. *[Berge (1957)] M is a maximum matching for a bipartite graph $\mathcal{B}$ if and only if $\mathcal{B}$ does not contain any augmenting paths for M.*

Proof. The proof can be found in Berge (1957). □

Lemma 4.8. *[Pothen (1985)] Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph with a maximum matching M. The incomplete set T_I and the overcomplete set T_O in Definition 4.5 are disjoint.*

Proof. For the sake of contradiction, assume that there is a vertex $v \in T_I \cap T_O$. Then there is an alternating path from an unmatched vertex in V to v and there is also an alternating path from an unmatched vertex in F to v. By sticking these two paths together we obtain an augmented path. It follows from Lemma 4.7 that M is not maximum. This is a contradiction and therefore T_I and T_O must be disjoint. □

Lemma 4.9 and Lemma 4.1 show that for a bipartite graph and a maximum matching with coarse decomposition $\mathrm{CD}(\mathcal{B}, M)$, the vertices in T_I, T_C, T_O are matched to vertices in T_I, T_C, T_O respectively. Furthermore the subgraph of $\mathcal{B}$ induced by T_C is self-contained, so that Algorithm 1 can be applied.

Lemma 4.9. *[Pothen (1985)] Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph with a maximum matching M. Let $\mathrm{CD}(\mathcal{B}, M) = \langle T_I, T_C, T_O\rangle$ be the associated coarse decomposition. A matched vertex in T_I is matched to a vertex in T_I and a matched vertex in T_O is matched to a vertex in T_O.*

Proof. For a matched vertex $x \in T_I$ there is an alternating path starting from an unmatched vertex $u_v \in V$ to x. When $x \in V$, this alternating path ends with a matched edge and hence x is matched to a vertex in T_I. When $x \in F$ the alternating path ends with an unmatched edge. We may extend the alternating path with the edge adjacent to x that is in M, and hence is matched to a vertex in T_I. For a matched vertex $x \in T_O$ there is an alternating path starting from an unmatched vertex $u_f \in F$ to x. When $x \in F$, this alternating path ends with a matched edge and hence x is matched to a vertex in T_O. When $x \in V$, the alternating path ends with an unmatched edge. The alternating path may be extended with the edge adjacent to x that is in M, and hence x is matched to a vertex in T_O. □

Lemma 4.1. *[Pothen (1985)] Let $\mathcal{B}$ be a bipartite graph with coarse decomposition $\langle T_I, T_C, T_O\rangle$. The subgraph $\mathcal{B}_C$ of $\mathcal{B}$ induced by vertices in T_C has a perfect matching and is self-contained.*

Proof. By Lemma 4.9 we know that vertices in T_I and T_O can only be matched to a vertex in T_I and T_O, respectively. There are no unmatched vertices in T_C, so vertices in $T_C \cap V$ are perfectly matched to vertices in $T_C \cap F$. It follows from Hall's marriage theorem that $\mathcal{B}_C$ is self-contained (M. Hall, 1986). □

The following lemma restricts edges that can be present between the incomplete, complete and overcomplete sets. This shows that clusters of the causal ordering graph that are in the overcomplete set are never descendants of clusters in the incomplete or complete set. Similarly, it also shows that clusters in the incomplete set are never ancestors of the complete or overcomplete sets. Lemma 4.2 is then used to prove Proposition 4.1.

Lemma 4.2. *[Pothen (1985)] Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph with a maximum matching M. Let $\mathrm{CD}(\mathcal{B}, M) = \langle T_I, T_C, T_O\rangle$ be the associated coarse decomposition. No edge joins a vertex in $T_I \cap V$ with a vertex in $(T_C \cup T_O) \cap F$ and no edge joins a vertex in $T_C \cap V$ with a vertex in $T_O \cap F$.*

Proof. Suppose that there is an edge $e = (v - f)$ between a vertex $v \in T_I \cap V$ to a vertex $f \in (T_C \cup T_O) \cap F$. By Lemma 4.9 the edge is not in the maximum matching. Note that there is an alternating path from an unmatched vertex in $T_I \cap V$ to v that starts with an unmatched edge and ends with a matched edge. By adding the edge $(v - f)$, we obtain again an alternating path so that $f \in T_I$. This is a contradiction, and hence there is no edge between $(v - f)$. The second part of the lemma follows by symmetry. □

Proposition 4.1. *[Pothen (1985)] The coarse decomposition of a bipartite graph $\mathcal{B}$ is independent of the choice of the maximum matching.*

Proof. Let M be an arbitrary matching and let $\mathrm{CD}(\mathcal{B}, M) = \langle T_I, T_C, T_O\rangle$. Note that all vertices in $(T_I \cap V) \setminus U_V$ are M-matched to vertices in $T_I \cap F$ (by construction and

Lemma 4.9). Also, all vertices in $(T_O \cap F) \setminus U_F$ are M-matched with vertices in $T_O \cap V$. Finally, all vertices in $T_C \cap V$ are M-matched with vertices in $T_C \cap F$ and vice versa by Lemma 4.1. By Lemma 4.2 we have $\mathrm{adj}_{\mathcal{B}}(T_I \cap V) = T_I \cap F$ and $\mathrm{adj}_{\mathcal{B}}(T_O \cap F) = T_O \cap V$, so *any* matching for $\mathcal{B}$ can only match vertices in $T_I \cap V$ with vertices in $T_I \cap F$ and vertices in $T_O \cap F$ with vertices in $T_O \cap V$.

For the sake of contradiction, assume that there exists a maximum matching M' that matches a vertex in $T_I \cap F$ with a vertex in $(T_C \cup T_O) \cap V$. Write:

$$M_V = \{v \in V : \exists f \in F : v - f \in M\}, \qquad M'_V = \{v \in V : \exists f \in F : v - f \in M'\},$$
$$M_F = \{f \in F : \exists v \in V : v - f \in M\}, \qquad M'_F = \{f \in F : \exists v \in V : v - f \in M'\}.$$

Note that the number of edges in matching M' is bounded by

$$\begin{aligned} |M'| &= |M'_V| \\ &= |M'_V \cap T_I| + |M'_V \cap T_C| + |M'_V \cap T_O| \\ &\leq (|F \cap T_I| - 1) + |V \cap T_C| + |V \cap T_O| \\ &= (|M_V \cap T_I| - 1) + |M_V \cap T_C| + |M_V \cap T_O| \\ &= |M_V| - 1 = |M| - 1, \end{aligned}$$

where we used that (i) vertices in $T_I \cap V$ can only be matched with vertices in $T_I \cap F$, (ii) all nodes in $T_I \cap F$ are M-matched with vertices in $M_V \cap T_I$, (iii) all variable vertices in T_C are M-matched, and (iv) all vertices in $T_O \cap V$ are M-matched. This contradicts the assumption that M' is a maximum matching.

In a similar way, one obtains a contradiction when assuming the existence of a maximum matching M'' that matches a vertex in $T_O \cap V$ with a vertex in $(T_I \cup T_C) \cap F$. Hence any maximum matching of $\mathcal{B}$ must match all vertices in $T_I \cap F$ with vertices in $T_I \cap V$, and all vertices in $T_O \cap V$ with vertices in $T_O \cap F$. We conclude that T_O and T_I do not depend on the choice of maximum matching. By definition T_C is uniquely determined by T_O and T_I. Therefore the coarse decomposition is independent of the choice of maximum matching. □

4.B.3 Markov property via d-separation

In this section we prove Theorem 4.3 below.

Theorem 4.3. *Let $\mathbf{X}^* = \mathbf{h}(\mathbf{X}_W)$ with $\mathbf{h} : \mathfrak{X}_W \to \mathfrak{X}_{V \setminus W}$ be a solution of a system of constraints $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} \rangle$ with coarse decomposition $\mathrm{CD}(\mathcal{B}) = \langle T_I, T_C, T_O \rangle$. Let $\mathrm{MO}_{\mathrm{CO}}(\mathcal{B})$ denote the subgraph of the Markov ordering graph induced by $T_C \cup T_O$ and let $\mathbf{X}^*_{\mathrm{CO}}$ denote the corresponding solution components. If $\mathcal{M}$ is maximally uniquely solvable then the pair $(\mathrm{MO}_{\mathrm{CO}}(\mathcal{B}), \mathbb{P}_{\mathbf{X}^*_{\mathrm{CO}}})$ satisfies the directed global Markov property (see Definition 4.15).*

Proof. Let $v \in (T_C \cup T_O) \cap (V \setminus W)$ be arbitrary and define $S_V = \mathrm{cl}(v) \cap V$ and

$S_F = \mathrm{cl}(v) \cap F$. First, we will show that $V(S_F) \setminus S_V = \mathrm{pa}_{\mathrm{MO}(\mathcal{B})}(v)$. The following equivalences hold for $x \in V$:

$$\begin{aligned}
x \in V(S_F) \setminus S_V &\iff x \in \mathrm{adj}_{\mathcal{B}}(S_F) \setminus S_V && \text{(by Definition 4.1)} \\
&\iff (x \to \mathrm{cl}(v)) \text{ in } \mathrm{CO}(\mathcal{B}) && \text{(by definition of Algorithm 3)} \\
&\iff (x \to v) \text{ in } D(\mathrm{CO}(\mathcal{B})) && \text{(by Definition 4.10)} \\
&\iff (x \to v) \text{ in } D(\mathrm{CO}(\mathcal{B}))_{\mathrm{mar}(F)} \\
&\iff (x \to v) \text{ in } \mathrm{MO}(\mathcal{B}) && \text{(by Definition 4.10)} \\
&\iff x \in \mathrm{pa}_{\mathrm{MO}(\mathcal{B})}(v).
\end{aligned}$$

By assumption, the system of constraints is maximally uniquely solvable w.r.t. $\mathrm{CO}(\mathcal{B})$. Note that $S_V \subseteq V(S_F)$. Hence, there exist measurable functions $g_i : \mathfrak{X}_{\mathrm{pa}_{\mathrm{MO}(\mathcal{B})}(v)} \to \mathcal{X}_i$ for all $i \in S_V$ such that $\mathbb{P}_{\mathbf{x}_W}$-a.s., for all $\mathbf{x}_{V(S_F)\setminus W} \in \mathfrak{X}_{V(S_F)\setminus W}$:

$$\begin{aligned}
&\forall f \in S_F : \ \phi_f(\mathbf{x}_{V(f)\setminus W}, \mathbf{X}_{V(f)\cap W}) = c_f \iff \\
&\forall i \in S_V : \ x_i = g_i\left(\mathbf{x}_{\mathrm{pa}_{\mathrm{MO}(\mathcal{B})}(v)\setminus W}, \mathbf{X}_{\mathrm{pa}_{\mathrm{MO}(\mathcal{B})}(v)\cap W}\right).
\end{aligned}$$

Since $v \in (T_C \cup T_O) \cap (V \setminus W)$ was chosen arbitrarily and $\mathbf{X}^* = \mathbf{h}(\mathbf{X}_W)$ with $\mathbf{h}$ a solution of $\mathcal{M}$, it follows that

$$X_v^* = g_v\left(\mathbf{X}^*_{\mathrm{pa}_{\mathrm{MO}(\mathcal{B})}(v)}\right) \quad \mathbb{P}_{\mathbf{X}_W}\text{-a.s.},$$

for all $v \in (T_C \cup T_O) \cap (V \setminus W)$. The directed global Markov property was already shown to hold for pairs $(\mathcal{G}, \mathbb{P}_{\mathbf{X}})$ where $\mathcal{G}$ is a DAG and $\mathbf{X}$ is a solution to a set of structural equations with functional dependences corresponding to the DAG (Lauritzen, 1996; Pearl, 2009). Because the Markov ordering graphs $\mathrm{MO}(\mathcal{B})$ and $\mathrm{MO}_{\mathrm{CO}}(\mathcal{B})$ are acyclic by construction, and $\mathrm{MO}_{\mathrm{CO}}(\mathcal{B})$ is the graph corresponding to this set of structural equations, this completes the proof. □

4.B.4 Causal ordering via perfect matchings

In this section we prove Theorem 4.2 below.

Theorem 4.2. *The output of Algorithm 2 coincides with the output of Algorithm 1.*

The following result gives a necessary and sufficient condition for the existence of a perfect matching for a bipartite graph and can be found in M. Hall (1986).

Theorem 4.7 (Hall's Marriage Theorem)**.** *Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph with $|V| = |F|$. Then $\mathcal{B}$ has a perfect matching if and only if $|F'| \leq |\mathrm{adj}_{\mathcal{B}}(F)|$ for all $F' \subseteq F$.*

From Hall's Marriage Theorem it trivially follows that a bipartite graph has a perfect matching if and only if it is self-contained.

Corollary 4.2. *Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph. Then $\mathcal{B}$ has a perfect matching if and only if $\mathcal{B}$ is self-contained.*

Proof. If $\mathcal{B}$ has a perfect matching then $|V| = |F|$. By Definition 4.3 we know that if $\mathcal{B}$ is self-contained then $|V| = |F|$. Hence, the statement follows from Definition 4.3 and Theorem 4.7. □

The following technical lemma is used to prove Lemma 4.11, which shows that the output of Algorithm 1 coincides with that of Algorithm 2 in the case that the input of the algorithm is a self-contained bipartite graph and $W = \emptyset$.

Lemma 4.10. *Let $\mathcal{M}$ be a perfect matching for a self-contained bipartite graph $\mathcal{B} = \langle V, F, E\rangle$. Let $S_V^{(1)}, \dots, S_V^{(n)}$ be a topological ordering of the strongly connected components in the graph $\mathcal{G}(\mathcal{B}, \mathcal{M})_{\mathrm{mar}(F)}$. Let $\mathcal{B}^{(i)}$ be the subgraph of $\mathcal{B}$ induced by $\bigcup_{j=i}^{n}(S_V^{(j)} \cup \mathcal{M}(S_V^{(j)}))$. Then $\mathcal{B}^{(i)}$ is self-contained and $\mathcal{M}(S_V^{(i)})$ is a minimal self-contained set in $\mathcal{B}^{(i)}$.*

Proof. We use the notation $\mathcal{G}^{(k)} := \mathcal{G}(\mathcal{B}^{(k)}, \mathcal{M}^{(k)})$ and $S_F^{(k)} := \mathcal{M}^{(k)}(S_V^{(k)})$, where $\mathcal{M}^{(1)} = \mathcal{M}$ (we will define $\mathcal{M}^{(i)}$ with $i > 1$ later). First we show that $S_F^{(1)}$ is self-contained in $\mathcal{B}^{(1)}$. We proceed by proving that $S_F^{(1)}$ is minimal self-contained in $\mathcal{B}^{(1)}$ and that $\mathcal{B}^{(2)}$ is a self-contained bipartite graph. Finally, we consider how these arguments can be iterated to prove the lemma.

By definition of a perfect matching and the fact that $\mathcal{B}^{(1)} = \mathcal{B}$ is self-contained, we know that:

$$|S_V^{(1)}| = |S_F^{(1)}| \leq |\mathrm{adj}_{\mathcal{B}^{(1)}}(S_F^{(1)})|. \tag{4.27}$$

By definition of topological ordering and the orientation step in Definition 4.4 we know that:

$$\mathrm{adj}_{\mathcal{B}^{(1)}}(S_F^{(1)}) \subseteq S_V^{(1)}.$$

Together, these two inequalities show that $|S_F^{(1)}| = |\mathrm{adj}_{\mathcal{B}^{(1)}}(S_F^{(1)})|$. Because $\mathcal{B}^{(1)}$ is self-contained, the set $S_F^{(1)}$ satisfies both conditions of Definition 4.3. We conclude that $S_F^{(1)}$ is self-contained in $\mathcal{B}^{(1)}$.

Assume, for the sake of contradiction, that $S_F^{(1)}$ is not *minimal* self-contained. Then there exists a non-empty strict subset $F' \subset S_F^{(1)}$ that is self-contained in $\mathcal{B}^{(1)}$. First note that, by Definition 4.3, we have that $|F'| = |\mathrm{adj}_{\mathcal{B}^{(1)}}(F')|$ and $|S_V^{(1)}| = |S_F^{(1)}|$ so that $S_V^{(1)} \setminus \mathrm{adj}_{\mathcal{B}^{(1)}}(F') \neq \emptyset$ and $\mathrm{adj}_{\mathcal{B}^{(1)}}(F') \neq \emptyset$. Furthermore, by Definition 4.4 (orientation step), we must have that:

$$\mathrm{pa}_{\mathcal{G}^{(1)}}(\mathrm{adj}_{\mathcal{B}^{(1)}}(F')) = \mathcal{M}^{(1)}(\mathrm{adj}_{\mathcal{B}^{(1)}}(F')) = F'. \tag{4.28}$$

Therefore there is no directed edge from any vertex in $F \setminus F'$ to any vertex in $\mathrm{adj}_{\mathcal{B}^{(1)}}(F')$. Clearly, there can be no edge in $\mathcal{G}^{(1)}$ between any vertex $v \in S_V^{(1)} \setminus \mathrm{adj}_{\mathcal{B}^{(1)}}(F')$ and any vertex $f' \in F'$ and hence

$$\mathrm{pa}_{\mathcal{G}^{(1)}}(S_V^{(1)} \setminus \mathrm{adj}_{\mathcal{B}^{(1)}}(F')) = \mathcal{M}^{(1)}(S_V^{(1)} \setminus \mathrm{adj}_{\mathcal{B}^{(1)}}(F')) = F \setminus F'. \tag{4.29}$$

Therefore, there can be no directed path from any $v \in S_V^{(1)} \setminus \mathrm{adj}_{\mathcal{B}^{(1)}}(F')$ to any $f \in F'$ in $\mathcal{G}^{(1)}$. This contradicts the assumption that $S_V^{(1)}$ is a strongly connected component in $\mathcal{G}^{(1)}_{\mathrm{mar}(F)}$. We conclude that $S_F^{(1)}$ is minimal self-contained in $\mathcal{B}^{(1)}$.

Clearly, the set $\mathcal{M}^{(2)} := \{(i - j) \in \mathcal{M}^{(1)} : i, j \notin S_V^{(1)} \cup S_F^{(1)}\}$ is a perfect matching for $\mathcal{B}^{(2)}$. By Corollary 4.2 we therefore know that $\mathcal{B}^{(2)}$ is self-contained. Since $S_V^{(2)}, \dots, S_V^{(n)}$ is a topological ordering for the strongly connected components in $\mathcal{G}^{(2)}_{\mathrm{mar}(F)}$ the above argument can be repeated to show that $S_F^{(2)}$ is minimal self-contained in $\mathcal{B}^{(2)}$. For arbitrary $i \in \{1, \dots, n\}$ this entire argument can be iterated to show that $S_F^{(i)}$ is minimal self-contained in the self-contained bipartite graph $\mathcal{B}^{(i)}$. □

Lemma 4.11. *Let $\mathcal{M}$ be an arbitrary perfect matching for a self-contained bipartite graph $\mathcal{B} = \langle V, F, E\rangle$. The directed cluster graph $\mathcal{G}_1 = \langle \mathcal{V}_1, \mathcal{E}_1\rangle$ that is obtained by application of Definition 4.4 coincides with the output $\mathcal{G}_2 = \langle \mathcal{V}_2, \mathcal{E}_2\rangle$ of Algorithm 1.*

Proof. Let $S^{(1)}, \dots, S^{(n)}$ be a topological ordering of the strongly connected components in $\mathcal{G}(\mathcal{M}, \mathcal{B})_{\mathrm{mar}(F)}$. By Definition 4.4 the cluster set $\mathcal{V}_1$ consists of clusters $S^{(i)} \cup \mathcal{M}(S^{(i)})$ with $i \in \{1, \dots, n\}$. By Lemma 4.10, Algorithm 1 can be run in such a way that the clusters $S^{(i)} \cup \mathcal{M}(S^{(i)})$ are added to $\mathcal{V}_2$ in the order specified by the topological ordering. By Theorem 4.1 the output of Algorithm 1 is unique and therefore $\mathcal{V}_1 = \mathcal{V}_2$. By Definition 4.4 the following equivalences hold for $C \in \mathcal{V}_1 = \mathcal{V}_2$ and $v \in V \setminus C$:

$$\begin{aligned}
(v \to C) \in \mathcal{E}_1 &\iff \exists w \in C \text{ s.t. } (v \to w) \text{ in } \mathcal{G}(\mathcal{M}, \mathcal{B}) \\
&\iff \exists w \in C \text{ s.t. } (v - w) \in E \text{ and } (v - w) \notin \mathcal{M} \\
&\iff v \in \mathrm{adj}_{\mathcal{B}}(C \cap F) \setminus \mathcal{M}(C \cap F) \\
&\iff v \in \mathrm{adj}_{\mathcal{B}}(C \cap F) \setminus (C \cap V) \\
&\iff (v \to C) \in \mathcal{E}_2.
\end{aligned}$$

Let $C \in \mathcal{V}_1 = \mathcal{V}_2$ and $f \in F \cap (\mathrm{adj}_{\mathcal{B}}(C) \setminus C)$. By definition of Algorithm 1 we know that $(f \to C) \notin \mathcal{E}_2$. Note that $\mathcal{M}(C \cap F) = C \cap V$. By Definition 4.4 there is no edge $(f \to v)$ with $v \in C \cap V$ in $\mathcal{G}(\mathcal{B}, \mathcal{M})$ and hence by Definition we know that $(f \to C) \notin \mathcal{E}_2$. By construction, edges $(x \to C)$ with $x \in C$ are neither in $\mathcal{E}_1$ nor in $\mathcal{E}_2$. We conclude that $\mathcal{E}_1 = \mathcal{E}_2$ and consequently $\mathcal{G}_1$ coincides with $\mathcal{G}_2$. □

Lemma 4.11 shows that the output of Algorithm 1 coincides with the output

of Algorithm 2 if the input is a self-contained bipartite graph. Otherwise, both Algorithm 1 and 2 have an initialization that is determined by the specification of exogenous variables. The exogenous variables are placed into separate clusters and there are directed edges from each exogenous variable to the clusters of its adjacencies for both algorithms. The output of the two algorithms coincides for any valid input.

4.B.5 Markov property via σ-separation

Here, we prove the following theorem.

Theorem 4.4. *Let $\mathbf{X}^* = \mathbf{g}(\mathbf{X}_W)$ be a solution of a system of constraints $\langle \mathcal{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} \rangle$, where the subgraph of $\mathcal{B} = \langle V, F, E \rangle$ induced by $(V \cup F) \setminus W$ has a perfect matching $\mathcal{M}$. If for each strongly connected component S in $\mathcal{G}(\mathcal{B}, \mathcal{M})$ with $S \cap W = \emptyset$, the system $\mathcal{M}$ is uniquely solvable w.r.t. $S_V = (S \cup \mathcal{M}(S)) \cap V$ and $S_F = (S \cup \mathcal{M}(S)) \cap F$ then the pair $(\mathcal{G}(\mathcal{B}, \mathcal{M})_{\mathrm{mar}(F)}, \mathbb{P}_{\mathbf{X}^*})$ satisfies the generalized directed global Markov property (Definition 4.15).*

The proof of this theorem relies on results by Forré et al. (2017), who define the notion of an *acyclic augmentation* for a class of graphical models that they call *HEDGes*. They define the *augmentation* of a HEDG as a directed graph where hyperedges are represented by vertices with additional edges. The acyclic augmentation of a HEDG is obtained by *acyclification* of the edge set of it augmentation (Forré et al., 2017). The acyclification of a directed graph is given in Definition 4.16.

Definition 4.16. Let $\mathcal{G} = \langle V, E \rangle$ be a directed graph. The *acyclification* of E, denoted by E^{acy}, has edges $(i \to j) \in E^{\mathrm{acy}}$ if and only if $i \notin \mathrm{sc}_{\mathcal{G}}(j)$ and there exists $k \in \mathrm{sc}_{\mathcal{G}}(j)$ such that $(i \to k) \in E$.

Lemma 4.12 shows that the clustering operation in Definition 4.4 on directed graphs, followed by the declustering operation in Definition 4.10, results in the same directed graph as the one that is obtained by applying the acyclification operation to its edge set.

Lemma 4.12. *Let $\mathcal{G} = \langle V, E \rangle$ be a directed graph. It holds that $\mathcal{G}^{\mathrm{acy}} = \langle V, E^{\mathrm{acy}} \rangle = D(\mathrm{clust}(\mathcal{G}))$.*

Proof. This follows from Definitions 4.10, 4.4, and 4.16. □

The following proposition shows that σ-separations in a directed graph coincide with d-separations in the graph that is obtained by clustering and subsequently declustering that directed graph.

Proposition 4.3. *Let $\mathcal{G} = \langle V, E \rangle$ be a directed graph with nodes V and $\mathcal{G}^{\mathrm{acy}} = \langle V, E^{\mathrm{acy}} \rangle$. Then for all subsets $A, B, C \subseteq V$:*

$$A \overset{\sigma}{\underset{\mathcal{G}}{\perp}} B \,|\, C \iff A \overset{d}{\underset{\mathcal{G}^{\mathrm{acy}}}{\perp}} B \,|\, C \iff A \overset{d}{\underset{D(\mathrm{clust}(\mathcal{G}))}{\perp}} B \,|\, C.$$

Proof. The first equivalence is Proposition A.19 in Bongers, Forré, et al. (2020). The second equivalence follows directly from Lemma 4.12. □

We now have all ingredients to finish the proof of Theorem 4.4. First note that, since the subgraph of $\mathcal{B} = \langle V, F, E\rangle$ induced by $(V \cup F) \setminus W$ has a perfect matching, $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E}\rangle$ is well-defined by Corollary 4.2. Let $S_V^{(1)}, \dots, S_V^{(n)}$ be the strongly connected components in $\mathcal{G}_{\mathrm{dir}}$, where $\mathcal{G}_{\mathrm{dir}} := \mathcal{G}(\mathcal{B}, \mathcal{M})_{\mathrm{mar}(F)}$. By Lemma 4.10 and the definition of Algorithm 1 we know that $\mathcal{V}$ consists of the clusters $S_V^{(i)} \cup \mathcal{M}(S_V^{(i)})$ with $i = 1, \dots, n$. Therefore, $\mathcal{M}$ is uniquely solvable with respect to $\mathrm{CO}(\mathcal{B})$. By Theorem 4.3 we have that for subsets $A, B, C \subseteq V \setminus W$:

$$A \overset{d}{\underset{\mathrm{MO}(\mathcal{B})}{\perp}} B \,|\, C \implies \mathbf{X}_A \underset{\mathbb{P}_{\mathbf{X}}}{\perp\!\!\!\perp} \mathbf{X}_B \,|\, \mathbf{X}_C. \tag{4.30}$$

By Proposition 4.3 we have that:

$$A \overset{\sigma}{\underset{\mathcal{G}_{\mathrm{dir}}}{\perp}} B \,|\, C \iff A \overset{d}{\underset{\mathcal{G}_{\mathrm{dir}}^{\mathrm{acy}}}{\perp}} B \,|\, C \iff A \overset{d}{\underset{D(\mathrm{clust}(\mathcal{G}_{\mathrm{dir}}))}{\perp}} B \,|\, C. \tag{4.31}$$

The desired result follows from implications (4.30) and (4.31) when $D(\mathrm{clust}(\mathcal{G}_{\mathrm{dir}})) = \mathrm{MO}(\mathcal{B})$. Consider the cluster set $\mathcal{V}_{\mathrm{mar}(F)} = \{S \cap V : S \in \mathcal{V}\}$ and note that edges in $\mathrm{CO}(\mathcal{B})$ go from vertices in V to clusters in $\mathcal{V}$. By Definition 4.10 and 4.4 we have that:

$$D(\langle \mathcal{V}_{\mathrm{mar}(F)}, \mathcal{E}\rangle) = D(\langle \mathcal{V}, \mathcal{E}\rangle)_{\mathrm{mar}(F)} \quad \text{and} \quad \mathrm{clust}(\mathcal{G}_{\mathrm{dir}}) = \langle \mathcal{V}_{\mathrm{mar}(F)}, \mathcal{E}\rangle, \tag{4.32}$$

respectively. It follows that

$$D(\mathrm{clust}(\mathcal{G}_{\mathrm{dir}})) = D(\mathrm{CO}(\mathcal{B}))_{\mathrm{mar}(F)} = \mathrm{MO}(\mathcal{B}). \tag{4.33}$$

Note that both d-separations and σ-separations are preserved under marginalization of exogenous vertices W (Bongers, Forré, et al., 2020; Forré et al., 2017). This finishes the proof.

4.B.6 Effects of interventions

This section is devoted to the proofs of the results that were presented in Section 4.5.

Theorem 4.5. *Let $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B}\rangle$ be a system of constraints with coarse decomposition $\mathrm{CD}(\mathcal{B}) = \langle T_I, T_C, T_O\rangle$. Suppose that $\mathcal{M}$ is maximally uniquely solvable w.r.t. the causal ordering graph $\mathrm{CO}(\mathcal{B})$ and let $\mathbf{X}^* = \mathbf{g}(\mathbf{X}_W)$ be a solution of $\mathcal{M}$. Let $f \in (T_C \cup T_O) \cap F$ and assume that the intervened system $\mathcal{M}_{\mathrm{si}(f, \phi_f', c_f')}$ is also maximally uniquely solvable w.r.t. $\mathrm{CO}(\mathcal{B})$. Let $\mathbf{X}' = \mathbf{h}(\mathbf{X}_W)$ be a solution of $\mathcal{M}_{\mathrm{si}(f, \phi_f', c_f')}$. If there is no directed path from f to $v \in (T_C \cup T_O) \cap V$ in $\mathrm{CO}(\mathcal{B})$ then $X_v^* = X_v'$ almost surely.*

On the other hand, if there is a directed path from f to v in $\mathrm{CO}(\mathcal{B})$ then X_v^* may have a different distribution than X_v', depending on the details of the model $\mathcal{M}$.

Proof. The directed cluster graph $\mathrm{CO}(\mathcal{B})$ is acyclic by construction and therefore there exists a topological ordering of its clusters. When there is no directed path from f to v in $\mathrm{CO}(\mathcal{B})$ then there exists a topological ordering $V^{(1)}, \dots, V^{(n)}$ of the clusters such that $\mathrm{cl}(v)$ comes before $\mathrm{cl}(f)$. Note that clusters of vertices in the incomplete set T_I are never ancestors of clusters in $T_C \cup T_O$ by Lemma 4.2 (the proof of this lemma was given in Appendix 4.B.2). Therefore there exists a topological ordering of clusters so that no cluster in T_I precedes a cluster in $T_C \cup T_O$. By the assumption of unique solvability w.r.t. the clusters $T_C \cup T_O$ in $\mathrm{CO}(\mathcal{B})$ we know that the solution component for any variable $v \in V^{(i)} \subseteq T_C \cup T_O$ can be solved from the constraints in $V^{(i)}$ after plugging in the relevant solution components $\bigcup_{j=1}^{i-1} V^{(j)}$. By the solvability assumption, the solution components X_v^* and X_v' are equal almost surely.

By assumption, the variables in $\mathrm{cl}(f)$ can be solved from the constraints in $\mathrm{cl}(f)$. Hence, a soft intervention on a constraint in $\mathrm{cl}(f)$ may change the distribution of the solution components $\mathbf{X}^*_{\mathrm{cl}(f)\cap V}$ that correspond to the variable vertices in $\mathrm{cl}(f)$. Suppose that there exists a sequence of clusters $V_1 = \mathrm{cl}(f), V_2, \dots, V_{k-1}, V_k = \mathrm{cl}(v)$ such that for all $V_i \in \{V_1, \dots, V_{k-1}\}$ there is a vertex $z_i \in V_i$ such that $(z_i \to V_{i+1})$ in $\mathrm{CO}(\mathcal{B})$. In that case we know that $V_i \cup T_I = \emptyset$ for $i = 1, \dots, k$. By the assumption of maximal unique solvability w.r.t. $\mathrm{CO}(\mathcal{B})$ the solution components for the variables in $V_2, \dots V_k$ may depend on the distribution of the unique solution components $\mathbf{X}^*_{\mathrm{cl}(f)\cap V}$ that correspond to the variable vertices in $\mathrm{cl}(f)$. It follows that the solution $\mathbf{X}_v^*$ may be different from that of $\mathbf{X}_v'$, if there is a directed path from f to v in $\mathrm{CO}(\mathcal{B})$. □

Lemma 4.3. *Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph and $W \subseteq V$, so that the subgraph of $\mathcal{B}$ induced by $(V \cup F) \setminus W$ is self-contained. Consider an intervention $\mathrm{do}(S_V, S_F)$ on a cluster $S = S_F \cup S_V$ with $S \cap W = \emptyset$ in the causal ordering graph $\mathrm{CO}(\mathcal{B})$. The subgraph of $\mathcal{B}_{\mathrm{do}(S_F,S_V)}$ induced by $(V \cup F) \setminus W$ is self-contained.*

Proof. By definition of Algorithm 2 we know that the subgraph of $\mathcal{B}$ induced by $(V \cup F) \setminus W$ has a perfect matching $\mathcal{M}$ such that $\mathcal{M}(S_F) = S_V$. By definition of a perfect intervention on the bipartite graph we know that $\mathcal{M}$ is also a perfect matching for the subgraph of $\mathcal{B}_{\mathrm{do}(S_F,S_V)}$ induced by $(V \cup F) \setminus W$. The result follows from Corollary 4.2. □

Proposition 4.2. *Let $\mathcal{B} = \langle V, F, E\rangle$ be a bipartite graph and W a set of exogenous variables. Let $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E}\rangle$ be the corresponding causal ordering graph. Let $S_F \subseteq F$ and $S_V \subseteq V \setminus W$ be such that $(S_F \cup S_V) \in \mathcal{V}$. Then:*

$$\mathrm{CO}(\mathcal{B}_{\mathrm{do}(S_F,S_V)}) = \mathrm{CO}(\mathcal{B})_{\mathrm{do}(S_F,S_V)}.$$

Proof. Let $S_V = \langle s_v^1, \dots, s_v^m\rangle$ and $S_F = \langle s_f^1, \dots s_f^m\rangle$ denote the targeted variables and constraints. We consider the output $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E}\rangle$ of the causal ordering algorithm.

Suppose that the order in which clusters $V^{(i)}$ are added to $\mathcal{V}$ is given by

$$V^{(1)}, \dots, V^{(k)} = (S_F \cup S_V), \dots, V^{(n)}. \tag{4.34}$$

Consider $\mathrm{CO}(\mathcal{B}_{\mathrm{do}(S_F,S_V)}) = \langle \mathcal{V}', \mathcal{E}' \rangle$. It follows from Definition 4.12, Lemma 4.4, Lemma 4.5, and the definition of Algorithm 3 (i.e. the extended causal ordering algorithm) that

$$V^{(1)}, \dots, V^{(k-1)}, \{s_f^1, s_v^1\}, \dots, \{s_f^m, s_v^m\}, V^{(k+1)}, \dots V^{(n)} \tag{4.35}$$

is an order in which clusters could be added to $\mathcal{V}'$. This shows that there are two differences between $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ and $\mathrm{CO}(\mathcal{B}_{\mathrm{do}(S_F,S_V)}) = \langle \mathcal{V}', \mathcal{E}' \rangle$: first $(S_F \cup S_V) \in \mathcal{V}$ whereas $\{\{s_f^i, s_v^i\} : i = 1, \dots, m\} \subseteq \mathcal{V}'$ and second the clusters $(S_F \cup S_V)$ may have parents in $\mathrm{CO}(\mathcal{B})$ but the clusters $\{s_f^i, s_v^i\}$ (with $i \in \{1, \dots, m\}$) have no parents in $\mathrm{CO}(\mathcal{B}_{\mathrm{do}(S_F,S_V)})$. The result follows directly from Definition 4.13. □

Theorem 4.6. *Let $\mathcal{M} = \langle \mathfrak{X}, \mathbf{X}_W, \boldsymbol{\Phi}, \mathcal{B} = \langle V, F, E \rangle \rangle$ be a system of constraints with coarse decomposition $\mathrm{CD}(\mathcal{B}) = \langle T_I, T_C, T_O \rangle$. Assume that $\mathcal{M}$ is maximally uniquely solvable w.r.t. $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ and let $\mathbf{X}^*$ be a solution of $\mathcal{M}$. Let $S_F \subseteq (T_C \cup T_O) \cap F$ and $S_V \subseteq (T_C \cup T_O) \cap (V \setminus W)$ be such that $(S_F \cup S_V) \in \mathcal{V}$. Consider the intervened system $\mathcal{M}_{\mathrm{do}(S_F,S_V,\xi_{S_V})}$ with coarse decomposition $\mathrm{CD}(\mathcal{B}_{\mathrm{do}(S_F,S_V)}) = \langle T_I', T_C', T_O' \rangle$. Let $\mathbf{X}'$ be a solution of $\mathcal{M}_{\mathrm{do}(S_F,S_V,\xi_{S_V})}$. If there is no directed path from any $x \in S_V$ to $v \in (T_C \cup T_O) \cap V$ in $\mathrm{CO}(\mathcal{B})$ then $X_v^* = X_v'$ almost surely. On the other hand, if there is $x \in S_V$ such that there is a directed path from x to $v \in (T_C \cup T_O) \cap V$ in $\mathrm{CO}(\mathcal{B})$, then X_v^* may have a different distribution than X_v'.*

Proof. First note that $T_C \cup T_O = T_C' \cup T_O'$ by Definition 4.12. Let $v \in S_V$. Since the variable vertices S_V are targeted by the perfect intervention, we have that $X_v' = \xi_v$, which may be different from the solution component X_v^*. Consider $v \in V \setminus S_V$ and its cluster $\mathrm{cl}(v)$ in $\mathrm{CO}(\mathcal{B})$. Since the causal ordering graph is acyclic by construction, there exists a topological ordering $V^{(1)}, \dots, V^{(i)} = \mathrm{cl}(v), \dots V^{(n)}$ of the clusters in $\mathrm{CO}(\mathcal{B})$ (where n is the amount of clusters in $\mathrm{CO}(\mathcal{B})$) such that $V^{(j)} \prec \mathrm{cl}(v)$ implies that there is a directed path from some vertex in $V^{(j)}$ to the cluster $\mathrm{cl}(v)$ in $\mathrm{CO}(\mathcal{B})$. Note that clusters in T_I are never ancestors of clusters in $T_C \cup T_O$ and that the ordering $V^{(1)}, \dots, V^{(n)}$ is such that no cluster in T_I precedes a cluster in $T_C \cup T_O$. By assumption, the solution component X_v^* can be solved from the constraints and variables in $V^{(i)} = \mathrm{cl}(v)$ by plugging in the solution for variables in $V^{(1)}, \dots, V^{(i-1)}$. Let $s_f^1, \dots s_f^m$ and $s_v^1, \dots s_v^m$ denote the ordered vertices in S_F and S_V respectively and suppose that $S_V \cup S_F = V^{(k)}$ for some $k \in \{1, \dots, n\}$. By definition of a perfect intervention on a cluster we know that $V^{(1)}, \dots, V^{(k-1)}, \{s_f^1, s_v^1\}, \dots, \{s_f^m, s_v^m\}, V^{(k+1)}, \dots V^{(n)}$ is a topological ordering of clusters in $\mathrm{CO}(\mathcal{B})_{\mathrm{do}(S_F,S_V)} = \mathrm{CO}(\mathcal{B}_{\mathrm{do}(S_F,S_V)})$ (by Proposition 4.2). Furthermore, maximal unique solvability w.r.t. $\mathrm{CO}(\mathcal{B})$ implies maximal unique solvability w.r.t. $\mathrm{CO}(\mathcal{B}_{\mathrm{do}(S_F,S_V)})$.

Suppose that $V^{(k)} \succ \mathrm{cl}(v)$ in the topological ordering for $\mathrm{CO}(\mathcal{B})$. By maximal unique solvability w.r.t. $\mathrm{CO}(\mathcal{B})_{\mathrm{do}(S_F,S_V)}$, X'_v can be solved from the constraints and variables in $\mathrm{cl}(v)$ by plugging in the solution for variables in $V^{(1)}, \ldots, V^{(i-1)}$. It follows that $X^*_v = X'_v$ almost surely and by construction of the topological ordering there is no directed path from any $x \in S_V$ to v in $\mathrm{CO}(\mathcal{B})$. Suppose that $V^{(k)} \prec \mathrm{cl}(v)$ in the topological ordering for $\mathrm{CO}(\mathcal{B})$. By maximal unique solvability w.r.t. $\mathrm{CO}(\mathcal{B})_{\mathrm{do}(S_F,S_V)}$, we know that X'_v can be solved from the constraints and variables in $V^{(i)}$ by plugging in the solution for variables in $V^{(1)}, \ldots, V^{(k-1)}, \{s^1_f, s^1_v\}, \ldots, \{s^m_f, s^m_v\}, V^{(k+1)}, \ldots V^{(i-1)}$. It follows that X^*_v and X'_v may have a different distribution, and by construction of the topological ordering there is a directed path from a vertex in S_V to the cluster $\mathrm{cl}(v)$ in $\mathrm{CO}(\mathcal{B})$. □

Chapter 5

Causality and independence in perfectly adapted dynamical systems

Based on:
Causality and independence in perfectly adapted dynamical systems,
T. Blom, and J.M. Mooij,
Submitted to Journal of Causal Inference.

Perfect adaptation in a dynamical system is the phenomenon that one or more variables have an initial transient response to a persistent change in an external stimulus but revert to their original value as the system converges to equilibrium. The causal ordering algorithm can be used to construct an *equilibrium causal ordering graph* that represents causal relations and a *Markov ordering graph* that implies conditional independences from a set of equilibrium equations. Based on this, we formulate sufficient graphical conditions to identify perfect adaptation from a set of first-order differential equations. Furthermore, we give sufficient conditions to test for the presence of perfect adaptation in experimental equilibrium data. We apply our ideas to a simple model for a protein signalling pathway and test its predictions both in simulations and on real-world protein expression data. We demonstrate that perfect adaptation in this model can explain why the presence and orientation of edges in the output of causal discovery algorithms does not always appear to agree with the presence and orientation of edges in biological consensus networks.

5.1 Introduction

Understanding causal relations is an objective that is central to many scientific endeavours. It is often said that 'the gold standard' for causal discovery is a randomized controlled trial, but practical experiments can be too expensive, unethical, or otherwise infeasible. The promise of causal discovery is that we can, under certain assumptions, learn about causal relations by using a combination of data and background knowledge (Mooij, Magliacane, et al., 2020; Spirtes, Glymour, et al., 2000; J. Zhang, 2008). Roughly speaking, causal discovery algorithms construct a graphical representation that encodes certain aspects of the data, such as conditional independences in the case of constraint-based causal discovery, given some constraints that are imposed by background knowledge. Under additional assumptions on the underlying causal mechanisms (e.g. the causal Markov condition, faithfulness, acyclicity) these graphical representations have a causal interpretation as well (Lauritzen, Dawid, et al., 1990; Mooij and Claassen, 2020; Mooij, Magliacane, et al., 2020; Spirtes, Glymour, et al., 2000). In this work, we specifically consider the equilibrium distribution of *perfectly adapted dynamical systems* that have the property that the class of graphs that encode the conditional independences in the distribution does not have a straightforward causal interpretation in terms of the changes in distribution induced by soft or perfect interventions. Systems with this property were already discussed in Chapter 4, but we will study them in more detail in the present chapter and in Chapter 6.

Perfect adaptation in a dynamical system is the phenomenon that one or more variables initially respond to a persistent external stimulus but ultimately revert to their original value. As a consequence, variables in the system change due to an external input, but they become independent of the stimulus change after the system reaches equilibrium again. We study the differences between the causal structure implied by the dynamic equations and the conditional dependence structure of the equilibrium distribution. To do so, we make use of the technique of *causal ordering*, introduced by Simon (1953), which can be used to construct a *Markov ordering graph* that encodes conditional independences between variables, as well as a *causal ordering graph* that represents causal relations, as we proved in Chapter 4. We introduce the notion of a *dynamic causal ordering graph* to represent transient causal effects in a dynamical model. We use these graphs to provide a sufficient graphical condition, for dynamical systems to achieve perfect adaptation, which does not require simulations or explicit calculations. Furthermore, we provide sufficient conditions to test for the presence of perfect adaptation in real-world data with the help of the Markov ordering graph and we elucidate the appropriate causal interpretation of the output of causal discovery algorithms when applied to (perfectly adapted) dynamical systems at equilibrium. Finally, we discuss how the notions of the causal Markov condition and the causal faithfulness condition, which are often used to tie graphs that represent conditional independences in a probability distribution to the causal properties of

the system that generated the data, become ambiguous in the case of perfectly adapted dynamical systems where the equilibrium and dynamical causal ordering graph are different.

We illustrate our ideas on three simple dynamical systems with feedback: the bathtub model in Dash (2005) and Iwasaki et al. (1994), the viral infection model in De Boer (2012) (a detailed presentation and analysis of this model will be presented in Chapter 6), and a chemical reaction network in Ma et al. (2009). We discuss how perfect adaptation may also manifest itself in applications of causal discovery algorithms to a popular protein expression data set (Sachs, Perez, et al., 2005). The output of causal discovery algorithms applied to this data sometimes appears to be at odds with the biological consensus presented in Sachs, Perez, et al. (2005), see for example Mooij, Magliacane, et al. (2020) and Ramsey et al. (2018). We present a model for the Ras-Raf-Mek-Erk signalling pathway, based on a model in Shin et al. (2009), under saturation conditions and test its predictions both in simulations and on real-world data. We demonstrate that perfect adaptation in this model can explain why the presence and orientation of edges in the output of causal discovery algorithms does not always appear to agree with the direction of edges in biological consensus networks that are based on a partial representation of the underlying dynamical mechanisms.

5.2 Background

In this section we consider the assumptions underpinning popular constraint-based causal discovery algorithms and give a brief description of a simple local causal discovery algorithm, introduced by Cooper (1997). We proceed with a concise introduction to the causal ordering algorithm, which was first introduced by Simon (1953) and conclude with a discussion of related work.

5.2.1 Causal discovery

The main objective in causal discovery is to infer causal relations from experimental and observational data. The most common causal discovery algorithms can be roughly divided into score-based and constraint-based approaches, where the latter are more generally applicable. The idea of constraint-based causal discovery algorithms (e.g PC or FCI and variants thereof, see Colombo et al. (2012), Forré et al. (2018), Spirtes, Glymour, et al. (2000), and J. Zhang (2008)), which we focus on in the remainder of this section, is that causal relations can be inferred by exploiting conditional independences in the data. These algorithms attempt to construct an equivalence class of graphs that encode a set of conditional independence relations in a probability distribution via a graphical separation criterion. A *d-separation* is a relation between three sets of vertices in a graph that indicates whether all paths between two sets of vertices are blocked by the vertices in a third, see Pearl (2009)

or Spirtes, Glymour, et al. (2000) for more details. If every d-separation in a graph implies a conditional independence in the probability distribution then we say that it satisfies the *directed global Markov property* w.r.t. that graph. Conversely, if every conditional independence in the probability distribution is due to a d-separation in a graph then we say that it is *d-faithful* to that graph. When a probability distribution satisfies the Markov property w.r.t. a graph and is also faithful to the graph, then this graph is a compact representation of the conditional independences in the probability distribution and we say that it *encodes* its independence relations.

A lot of work has been done to understand the various conditions (e.g. linearity, Gaussianity, discreteness, causal sufficiency, acyclicity) under which a graph that encodes all conditional independences and dependences in a probability distribution has a certain causal interpretation, see Colombo et al. (2012), Forré et al. (2018), Hyttinen et al. (2012), Lacerda et al. (2008), Mooij and Claassen (2020), Mooij, Magliacane, et al. (2020), Richardson and Spirtes (1999), Spirtes, Glymour, et al. (2000), Strobl (2018), and J. Zhang (2008). Perhaps the simplest assumption is that the data was generated by a causal DAG[1] (Spirtes, Glymour, et al., 2000). In that case, the *causal Markov condition*, which states that variables are independent of their non-effects conditional on all their direct causes, and the *causal faithfulness condition*, which states that there are no other conditional independences than those implied by the causal Markov condition, ensure that there exists a single DAG that represents both conditional independences and causal relations (Lauritzen, Dawid, et al., 1990; Pearl, 2009). For the acyclic setting, powerful constraint-based causal discovery algorithms such as PC (under the assumption of causal sufficiency) and FCI (when latent confounders may be present) have been developed (Spirtes, Glymour, et al., 2000).

However, many systems of interest in various scientific disciplines (e.g. biology, econometrics, physics) include feedback mechanisms. Cyclic Structural Causal Models (SCMs) (Bongers, Forré, et al., 2020) can be used to model causal features and conditional independence relations of systems that contain cyclic causal relationships. For linear SCMs with causal cycles, several causal discovery algorithms have been developed (Hyttinen et al., 2012; Lacerda et al., 2008; Richardson and Spirtes, 1999; Strobl, 2018) that are based on d-separations. The d-separation criterion is applicable to acyclic settings and to cyclic SCMs with either discrete variables or linear relations between continuous variables, but it is too strong in general (Spirtes and Richardson, 1995). Forré et al. (2017), inspired by the 'collapsed graph' in Spirtes and Richardson (1995), developed the alternative *σ-separation* criterion for graphs that may contain cycles. If every σ-separation in a graph implies a conditional independence in the probability distribution then we say that it satisfies the *generalized directed global Markov property* w.r.t. that graph. Conversely, if every conditional independence in the probability distribution is due to a σ-separation in a graph then we say that it is σ-

[1] A Directed Acyclic Graph (DAG) is a pair $\langle V, E \rangle$ where V is a set of vertices and E a set of directed edges between vertices such that there are no directed cycles.

faithful to that graph. Forré et al. (2018) propose a sound and complete causal discovery algorithm based on σ-separations and the assumption of σ-faithfulness for data that is generated by a cyclic SCM with non-linear relations between continuous variables. Recently, Mooij and Claassen (2020) proved that the PC and FCI algorithms are sound and complete in this setting and showed how to read off causal relations and other features from the output of the algorithm. In earlier work, Richardson (1996) proved soundness of a causal discovery algorithm under the generalized directed Markov property and the d-faithfulness assumption, under the additional assumption of causal sufficiency. At the end of this section we will consider the LCD (i.e. Local Causal Discovery) algorithm in Cooper (1997), which was proven to be sound in both the σ- and d-separation settings (Mooij, Magliacane, et al., 2020).

In this chapter, we consider equilibrium distributions that are generated by dynamical models. The causal relations in an equilibrium model are defined through the effects of persistent interventions (i.e. interventions that are constant over time) on the equilibrium solution of variables that are endogenous to the model, assuming that the system again converges to equilibrium. In Chapter 4, we showed that directed graphs encoding the conditional independences between endogenous variables in the equilibrium distribution of dynamical systems with feedback do not have a straightforward and intuitive causal interpretation. As a consequence, the output of algorithms such as LCD, PC, or FCI applied to equilibrium data of dynamical systems with feedback at equilibrium cannot always be interpreted causally in a naïve way. One issue is that the equilibrium distribution of certain (perfectly adapted) dynamical systems can also be generated by a causal DAG (consider e.g. the bathtub example in Chapter 4, (Dash, 2005; Iwasaki et al., 1994), or Section 5.3.1.1), while the causal mechanisms of the true underlying system are provided by the dynamics of a model that includes feedback. This example illustrates some of the arguments made by Dawid (2010) against the use of causal DAGs. The in-depth analysis of causality and independence in perfectly adapted dynamical systems in this paper contributes to this discussion. Representations of dynamical systems at equilibrium as cyclic SCMs may not have a unique solution under perfect interventions (Bongers and Mooij, 2018) and in Chapter 3 we demonstrated that the causal semantics of the system may not be fully captured by a cyclic SCM. Here, we will present methods that supplement existing methods for SCMs to study the properties of perfectly adapted dynamical systems in more detail.

In this chapter we will, for the sake of simplicity, limit our attention to one of the simplest causal ordering algorithms, LCD. This algorithm is a straightforward and efficient search method to detect one specific (causal) structure from background knowledge and observations or experimental data (Cooper, 1997). The algorithm looks for triples of variables (C, X, Y) for which (a) C is a context variable that is not caused by any other observed variable and (b) the following (in)dependences hold: $C \not\perp\!\!\!\perp X$, $X \not\perp\!\!\!\perp Y$, and $C \perp\!\!\!\perp Y \mid X$. Figure 2.2 shows the graphs that correspond to the LCD triple (C, X, Y). Note that, in the absence of latent confounders, there are no bi-

directed edges, and the graph structure of an LCD triple is a DAG. Under the causal Markov and causal faithfulness assumptions, directed edges in the graph of an LCD triple represent causal relations. For simplicity, we only consider DAGs to encode the conditional independence relations in equilibrium distributions of dynamical models. The ideas in this paper can be extended to a setting with latent variables and more advanced causal discovery algorithms.

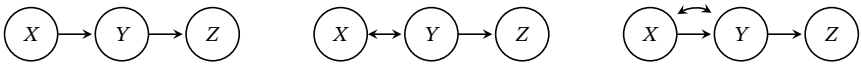


Figure 5.1: Possible graph structures of an LCD triple. In the absence of latent confounders the triple has the structure of the DAG in the figure on the left.

5.2.2 Causal ordering

The causal ordering algorithm, which was first introduced by Simon (1953), applies to sets of equations and returns an ordering of the variables and equations. A thorough treatment of the causal ordering was already given in Chapter 4. To keep this chapter self-contained, we give a brief introduction to the causal ordering algorithm of Nayak (1995), which is based on the block triangular form of matrices in Pothen and Fan (1990), as it was present in the previous chapter. This algorithm is equivalent but computationally more efficient than the original causal ordering algorithm (Gonçalves et al., 2016). It is applicable to sets of equations that can be represented by a bipartite graph with a *perfect matching* (i.e. there exists a subset $M \subseteq E$ of the edges in the bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ so that every vertex in $V \cup F$ is adjacent to exactly one edge in M). Although we extended the causal ordering algorithm to general bipartite graphs in the previous chapter, we will in the present chapter, for the most part, assume that a perfect matching exists for the sake of simplicity.

The structure of a set of equations and the variables that appear in them can be represented by a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$, where vertices F correspond to the equations and vertices V correspond to the endogenous variables that appear in these equations. For each endogenous variable $v \in V$ that appears in an equation $f \in F$ there is an edge $(v - f) \in E$. The output of the causal ordering algorithm is a directed cluster graph $\langle \mathcal{V}, \mathcal{E}\rangle$, consisting of a partition $\mathcal{V}$ of the vertices $V \cup F$ into clusters and edges $(v \to S) \in \mathcal{E}$ that go from vertices $v \in V$ to clusters $S \in \mathcal{V}$.

Application of the causal ordering algorithm to a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ results in the directed cluster graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E}\rangle$, which we will call the *causal ordering graph*. It is constructed in four steps:

(i) Find a perfect matching $M \subseteq E$ and let $M(S)$ denote the vertices in $V \cup F$ that are joined to vertices in $S \subseteq V \cup F$ by an edge in M.

(ii) For each $(v - f) \in E$ with $v \in V$ and $f \in F$: if $(v - f) \in M$ orient the edge as $(v \leftarrow f)$ and if $(v - f) \notin M$ orient the edge as $(v \to f)$. Let $\mathcal{G}(\mathcal{B}, M)$ denote the

resulting directed graph.

(iii) Partition vertices $V \cup F$ into strongly connected components $\mathcal{V}'$ of $\mathcal{G}(\mathcal{B}, M)$. Create the cluster set $\mathcal{V}$ consisting of clusters $S \cup M(S)$ for each $S \in \mathcal{V}'$. For each edge $(v \to f) \in E$ add an edge $(v \to \mathrm{cl}(f))$ to $\mathcal{E}$ when $v \notin \mathrm{cl}(f)$, where $\mathrm{cl}(f)$ denotes the cluster in $\mathcal{V}$ that contains f.

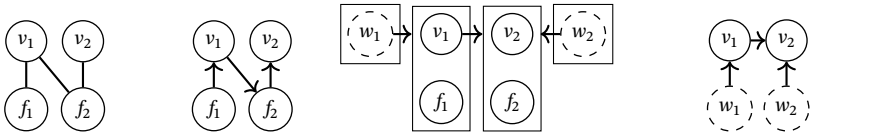
(iv) Optionally, exogenous variables appearing in the equations can be added as singleton clusters to $\mathcal{V}$, with edges towards the clusters of the equations in which they appear in $\mathcal{E}$.

Example 5.1. Consider the following set of equations with index set $F = \{f_1, f_2\}$ that contain endogenous variables with index set $V = \{v_1, v_2\}$:

$$f_1 : \quad X_{v_1} - U_{w_1} = 0, \tag{5.1}$$

$$f_2 : \quad X_{v_2} + X_{v_1} - U_{w_2} = 0, \tag{5.2}$$

where U_{w_1} and U_{w_2} are exogenous (random) variables indexed by $W = \{w_1, w_2\}$. Figure 5.2(a) shows the associated bipartite graph $\mathcal{B} = \langle V, F, E \rangle$. This graph has exactly one perfect matching $M = \{(v_1 - f_1), (v_2 - f_2)\}$, which is used in step (ii) of the causal ordering algorithm to construct the directed graph $\mathcal{G}(\mathcal{B}, M)$ in Figure 5.2(b). The causal ordering graph that is obtained after applying steps (iii) and (iv) of the causal ordering algorithm is given in Figure 5.2(c). △



(a) Bipartite graph. (b) Oriented graph. (c) Causal ordering graph. (d) Markov ordering graph.

Figure 5.2: The bipartite graph $\mathcal{B}$ associated with equations (5.1) and (5.2) is given in Figure 5.2(a). The oriented graph $\mathcal{G}(\mathcal{B}, M)$ obtained in step (ii) of the causal ordering algorithm, with perfect matching M, in Example 5.1 is shown in Figure (b). The causal ordering graph $\mathrm{CO}(\mathcal{B})$, with added exogenous variables, is given in Figure (c). The corresponding Markov ordering graph $\mathrm{MO}(\mathcal{B})$ is displayed in Figure (d).

Throughout this chapter, we will assume that sets of equations are *uniquely solvable with respect to the causal ordering graph*, as in Definition 4.8. Roughly speaking, this means that the endogenous variables in the model can be solved from the equations in their clusters along a topological ordering of the causal ordering graph. In Chapter 4 we showed that the causal ordering graph represents the effects of soft and certain perfect interventions under the assumption of unique solvability w.r.t. the causal ordering graph, see Theorem 4.5 and Theorem 4.6 respectively. Soft interventions target equations; they do not change which variables appear in the targeted equation and may only alter the parameters or form of the equation. Perfect

interventions target clusters in the causal ordering graph and replace the equations in the targeted cluster with equations that set the variables in the cluster equal to constant values. We say that there is a direct path from a vertex x to a vertex y in a directed cluster graph $\langle \mathcal{V}, \mathcal{E} \rangle$ if either $\mathrm{cl}(x) = \mathrm{cl}(y)$ or there is a sequence of clusters $V_1 = \mathrm{cl}(x), V_2, \dots, V_{k-1}, V_k = \mathrm{cl}(y)$ so that for all $i \in \{1, \dots, k-1\}$ there is a vertex $z_i \in V_i$ such that $(z_i \to V_{i+1}) \in \mathcal{E}$. A soft intervention on an equation or a perfect intervention on a cluster has no effect on a variable in the causal ordering graph whenever there is no directed path to that variable from the intervention target (i.e. the targeted equation or an arbitrary vertex in the targeted cluster, respectively). Since the equations in Example 5.1 are uniquely solvable w.r.t. the causal ordering graph in Figure 5.2(c) we can use it to read off that, for example, a soft intervention targeting f_1 may have an effect on X_{v_2} and that a perfect intervention targeting the cluster $\{v_2, f_2\}$ has no effect on X_{v_1}.

Given the probability distribution of exogenous random variables, one gets a unique probability distribution on the endogenous variables under the assumption of unique solvability w.r.t. the causal ordering graph. The *Markov ordering graph* is a directed graph $\mathrm{MO}(\mathcal{B})$ that implies conditional independences between the endogenous random variables that solve the system via d-separations, see Theorem 4.3. The Markov ordering graph $\langle V, E \rangle$ is obtained from a causal ordering graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ by putting $V = \bigcup_{S \in \mathcal{V}} S$ and constructing edges $(v \to w) \in E$ if and only if $(v \to \mathrm{cl}(w)) \in \mathcal{E}$. The Markov ordering graph for the set of equations in Example 5.1 is given in Figure 5.2(d). The d-separations in this graph imply conditional independences between the corresponding variables. For instance, since v_1 and w_2 are d-separated we know that X_{v_1} and X_{w_2} are independent.

Assuming that the probability distribution is d-faithful to the Markov ordering graph and that we have a conditional independence oracle, we know that the output of the PC-algorithm is the Markov equivalence class of the Markov ordering graph. However, as we already demonstrated in Chapter 4, for certain dynamical systems, the directed edges in the Markov ordering graph should not be interpreted as causal relations. Likewise, we will discuss three examples of perfectly adapted systems at equilibrium for which the Markov ordering graph does not have a straightforward causal interpretation in Section 5.3.3.2. In Section 5.6.1 we provide a brief discussion about the ambiguity of the causal Markov and faithfulness conditions in these examples.

5.2.3 Related work

Causal ordering is a technique that can be used to relate the (equilibrium) equations in a dynamical model to causal properties and conditional independence relations (Blom and Mooij, 2020; Blom, Van Diepen, et al., 2021; Simon and Iwasaki, 1988). The relationship between dynamical models and causal models has already received much attention over the years. The works of F. M. Fisher (1970), Mogensen et al.

(2018), Rubenstein et al. (2018), Sokol et al. (2014), and Voortman et al. (2010) considered causal relations in dynamical systems that are not at equilibrium, while Blom, Bongers, et al. (2019), Hyttinen et al. (2012), Lacerda et al. (2008), Lauritzen and Richardson (2002), Mooij, Janzing, Heskes, et al. (2011), and Mooij, Janzing, and Schölkopf (2013) considered graphical and causal models that arise from studying the stationary behaviour of dynamical models. An adaptation of Blom, Bongers, et al. (2019) can be found in Chapter 3 of this thesis. In Chapter 6, which is based on Blom and Mooij (2020), we study the robustness of model predictions when two subsystems in a dynamical model at equilibrium are combined, and consider opportunities for using causal discovery to detect feedback loops and the presence of variables that are not self-regulating using both models and experimental data for a subsystem. The causal behaviour of dynamical models and their equilibration to an SCM is studied by Bongers and Mooij (2018) and Dash (2005). In the previous chapters, we have noted subtleties regarding the use of a single graphical model to represent both conditional independence properties and causal properties of the variables in certain dynamical systems at equilibrium. This was also noted in Blom, Bongers, et al. (2019), Dash (2005), Dawid (2010), Lacerda et al. (2008), and Lauritzen and Richardson (2002). Often, restrictive assumptions on the underlying dynamical models are made to avoid these subtleties. In this thesis we directly address these issues by using the causal ordering algorithm to construct separate graphical representations for the causal properties and conditional independence relations implied by these systems. Our approach can be used in the equilibrium setting, but can also be employed to model transient causal effects in non-equilibrium settings, as we will discuss in Section 5.3.2.1. In this chapter, we focus on using these ideas to study the properties of perfectly adapted systems and applying this in particular to better understand the causal mechanisms that drive protein signalling networks.

In Chapter 3, we showed that the popular SCM framework (Bongers, Forré, et al., 2020; Pearl, 2009) is not flexible enough to fully capture the causal semantics in terms of perfect interventions targeting variables of certain dynamical systems at equilibrium. Therefore we proposed to use Causal Constraints Models (CCMs) instead. The drawback of this approach is that the causal constraints do not possess some of the attractive properties of SCMs, although this is greatly alleviated by the causal ordering technique presented in Chapter 4, which can be used to construct graphical representations of causal relations and conditional independences. In the discussion section we consider how this technique can be used to obtain graphical presentations and a Markov property for the dynamical model of the basic enzyme reaction that we considered in Chapter 3.

The analysis of network topologies that can achieve perfect adaptation is a topic of interest in cell biology, see for example (Araujo et al., 2018; Ferrell, 2016; Krishnan et al., 2019; Ma et al., 2009; Muzzey et al., 2009). The present work provides a method that facilitates the analysis of perfectly adapted dynamical systems by providing a principled method to identify perfect adaptation either from model equations or from

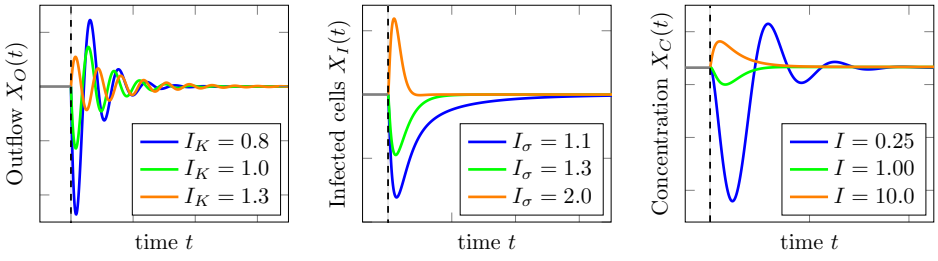
experimental data and background knowledge. It is our hope that the ideas presented in this paper contribute to increasing the impact of causal inference in cell biology and dynamical modelling.

5.3 Perfect adaptation

The ability of a system to converge to its original state when a constant and persistent external stimulus is added or changed is referred to as *perfect adaptation*. If the adaptive behaviour does not depend on the precise setting of parameters then we say that the adaptation is *robust*. In the literature, the most interesting of the two is robust perfect adaptation, which is also commonly referred to as perfect adaptation. Henceforth, we will use the term *perfect adaptation* to refer to *robust perfect adaptation*. In this section, we take a look at several examples of simple dynamical systems that can achieve perfect adaptation and consider how we can identify models that are capable of perfect adaptation. Finally, we discuss the correct interpretation of the output of some constraint-based causal discovery algorithms applied to perfectly adapted dynamical systems and possibilities for the identification of perfect adaptation from (equilibrium) data.

5.3.1 Examples

In this section we present three dynamical systems and show that they are capable of achieving perfect adaptation. The details of simulations that are presented in this section are given in Appendix 5.A.



(a) Filling bathtub model. (b) Viral infection model. (c) Reaction network model.

Figure 5.3: Simulations of the outflow rate $X_O(t)$ in the bathtub model, the amount of infected cells $X_I(t)$ in the viral infection model, and the concentration $X_C(t)$ in the biochemical reaction network with a negative feedback loop after a change in the input signal. The timing of this change is indicated by a vertical dashed line. The three systems started with input signals $I_K = 1.2$, $I_\sigma = 1.6$, and $I = 1.5$. After a transient response $X_O(t), X_I(t)$, and $X_C(t)$ all converge to their original equilibrium value (i.e. they perfectly adapt to the input signal).

5.3.1.1 Filling bathtub

We consider the example of a filling bathtub from Iwasaki et al. (1994), that was discussed at length in Chapter 4. Let $I_K(t)$ be an input signal that represents the size of a drain in the bathtub. The inflow rate $X_I(t)$, water level $X_D(t)$, water pressure $X_P(t)$, and outflow rate $X_O(t)$ are modelled by the following static and dynamic equations:

$$X_I(t) = U_I, \tag{5.3}$$
$$\dot{X}_D(t) = U_1(X_I(t) - X_O(t)), \tag{5.4}$$
$$\dot{X}_P(t) = U_2(g\, U_3 X_D(t) - X_P(t)), \tag{5.5}$$
$$\dot{X}_O(t) = U_4(U_5 I_K(t) X_P(t) - X_O(t)), \tag{5.6}$$

where g is the gravitational constant, and $U_I, U_1, U_2, U_3, U_4, U_5$ are independent exogenous random variables taking value in $\mathbb{R}_{>0}$. Let X_D, X_P, and X_O denote the respective equilibrium solutions for the water level, water pressure, and outflow rate. The equilibrium equations associated with this model can easily be constructed by setting the time derivatives equal to zero and assuming the input signal $I_K(t)$ to have a constant value I_K:

$$f_I : \quad X_I - U_I = 0, \tag{5.7}$$
$$f_D : \quad U_1(X_I - X_O) = 0, \tag{5.8}$$
$$f_P : \quad U_2(g\, U_3 X_D - X_P) = 0, \tag{5.9}$$
$$f_O : \quad U_4(U_5 I_K X_P - X_O) = 0, \tag{5.10}$$

We call the labelling f_D, f_P, f_O that we choose for the equilibrium equations that are constructed from the time-derivatives the *natural labelling* for this dynamical system, which means that the equilibrium equation constructed from $\dot{X}_i(t)$ of variable v_i is labelled as f_i. A solution (X_I, X_D, X_P, X_O) to the system of equilibrium equations satisfies $X_I = U_I$ and $X_O = X_I$ almost surely. From this we conclude that, at equilibrium, the outflow rate is independent of the size of the drain I_K, assuming that U_I is independent of I_K. We recorded the changes in the system after we changed the input signal I_K of the bathtub system in equilibrium. The results in Figure 5.3(a) show that the outflow rate has a transient response to changes in the input signal I_K, but it ultimately converges to its original value. The outflow rate X_O in the bathtub model *perfectly adapts* to changes in I_K.

5.3.1.2 Viral infection model

We consider the example of a simple dynamical model for a viral infection and immune response in De Boer (2012). This example will be treated in more detail in Chapter 6, which is based on Blom and Mooij (2020). The model describes target cells $X_T(t)$, infected cells $X_I(t)$, and an immune response $X_E(t)$. We will treat $I_\sigma(t)$

as an exogenous input signal that represents the production rate of target cells. The system is defined by the following dynamic equations:

$$\dot{X}_T(t) = I_\sigma(t) - d_T X_T(t) - \beta X_T(t) X_I(t), \tag{5.11}$$

$$\dot{X}_I(t) = (f\beta X_T(t) - d_I - kX_E(t)) X_I(t), \tag{5.12}$$

$$\dot{X}_E(t) = (aX_I(t) - d_E) X_E(t). \tag{5.13}$$

We have that $\beta = \frac{bp}{c}$ where b is the infection rate, p is the number of virus particles produced per infected cell, and c is the clearance rate of viral particles. Furthermore, d_T is the death rate of target cells, a is an activation rate, d_E and d_I are turnover rates and k is a mass-action killing rate. We assume that a, k are constants and that d_T, d_I, d_E, and β are independent exogenous random variables. We use the natural labelling for the equilibrium equations that are constructed from the differential equations:[2]

$$f_T : \quad I_\sigma - d_T X_T - \beta X_T X_I = 0, \tag{5.14}$$

$$f_I : \quad f\beta X_T - d_I - kX_E = 0, \tag{5.15}$$

$$f_E : \quad aX_I - d_E = 0, \tag{5.16}$$

assuming a constant value I_σ of the input signal. We initialized the model in an equilibrium state and simulated the response of the model after changing the input signal I_σ to three different values. Figure 5.3(b) shows that the amount of infected cells $X_I(t)$ has a transient response to a change in the input signal, but then returns to its original value, it perfectly adapts to changes in I_σ.

5.3.1.3 Reaction networks with a negative feedback loop

The phenomenon of perfect adaptation is a common feature in biochemical reaction networks and there exist many network topologies that can achieve (near) perfect adaptation (Araujo et al., 2018; Ferrell, 2016). For networks consisting of only three nodes Ma et al. (2009) found by an exhaustive search that there exist two major classes of network topologies that produce robust adaptive behaviour. The reaction diagrams for these networks are given in Figure 5.4. Here we will only analyse Negative Feedback with a Buffer Node (NFBN), we will examine the other network in the discussion section and in Appendix 5.D. The NFBN system can be described by the following first-order differential equations:

$$\dot{X}_A(t) = I(t) k_{IA} \frac{(1 - X_A(t))}{K_{IA} + (1 - X_A(t))} - F_A k_{F_A A} \frac{X_A(t)}{K_{F_A A} + X_A(t)}, \tag{5.17}$$

[2] Following De Boer (2012), we are only interested in strictly positive solutions of this dynamical system. Therefore, we use the equilibrium equation f_I instead of $(f\beta X_T - d_I - kX_E)X_I = 0$ and f_E instead of $(aX_I - d_E)X_E = 0$.

$$\dot{X}_B(t) = X_C(t)k_{CB}\frac{(1-X_B(t))}{K_{CB}+(1-X_B(t))} - F_B k_{F_BB}\frac{X_B(t)}{K_{F_BB}+X_B(t)}, \tag{5.18}$$

$$\dot{X}_C(t) = X_A(t)k_{AC}\frac{(1-X_C(t))}{K_{AC}+(1-X_C(t))} - X_B(t)k_{BC}\frac{X_C(t)}{K_{BC}+X_C(t)}, \tag{5.19}$$

where $X_A(t)$, $X_B(t)$, $X_C(t)$ are concentrations of three compounds A, B, and C, while $I(t)$ represents an external input into the system. Assume that k_{IA}, k_{CB}, and k_{AC} are independent exogenous random variables, that we will denote as U_A, U_B, U_C respectively, and that the other parameters are constants. Ma et al. (2009) show that perfect adaptation is achieved under saturation conditions, $(1-X_B(t)) \gg K_{CB}$ and $X_B(t) \gg K_{F_BB}$, in which case the following approximation can be made:

$$\dot{X}_B(t) \approx X_C(t)k_{CB} - F_B k_{F_BB}. \tag{5.20}$$

Under the assumption that $I(t)$ has a constant value, the system converges to an equilibrium. We will denote the equilibrium equations that are associated with the time derivatives $\dot{X}_A(t)$ and $\dot{X}_C(t)$ using the natural labelling f_A and f_C. The equilibrium equation f_B is obtained by setting the approximation of the time derivative $\dot{X}_B(t)$ equal to zero. We initialized this model in an equilibrium state and then simulated its response after changing the input signal I to three different values. Figure 5.3(c) shows that $X_C(t)$ perfectly adapts to changes in the input signal I.

(a) Negative feedback with a buffer node. (b) Incoherent feedforward loop with proportioner node.

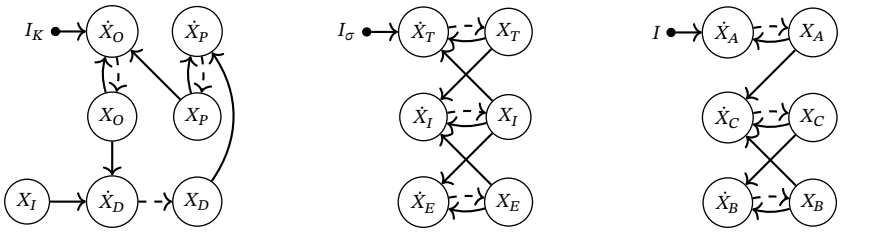
Figure 5.4: The two three-node network topologies that can achieve perfect adaptation in Ma et al. (2009). The motif in Figure 5.4(a) shows Negative Feedback with a Buffer Node (NFBN) B, while Figure 5.4(b) shows an Incoherent Feedforward Loop with a Proportioner Node (IFFLP) B. Orange edges represent saturated reactions, blue edges represent linear reactions, and black edges are unconstrained reactions. Arrowheads represent a positive influence and edges ending with a circle represent negative influence.

5.3.2 Identification of perfect adaptation

In this section, we consider graphical representations of dynamical systems that can achieve perfect adaptation. We use these representations to formulate a sufficient graphical criterion to identify perfect adaptation in first-order dynamical models.

5.3.2.1 Graphical representations

The *functional graph* is a compact representation of the structure of a set of first-order differential equations that are written in canonical form (i.e. derivatives are on the left-hand side of each equation and functions of variables on the right-hand side). Vertices in this graph represent variables or derivatives of variables and directed dashed edges from derivatives to their corresponding variables indicate integration links. Additionally, there is an edge from a variable to a derivative whenever that variable appears in the corresponding differential equation. Figures 5.5(a), 5.5(b), and 5.5(c) show the functional graphs for the bathtub model, the viral infection model, and the reaction network respectively.



(a) Filling bathtub model. (b) Viral infection model. (c) Reaction network model.

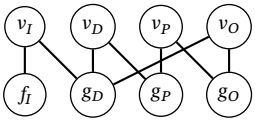
Figure 5.5: The functional graphs of the dynamics of the bathtub model, the viral infection model, and the reaction network with negative feedback. The input vertices I_K, I_σ, and I are represented by black dots.

Contrary to Iwasaki et al. (1994), we do not interpret the functional graphs in Figure 5.5 causally, because they may not have an intuitive causal interpretation in terms of regular interventions (e.g. we will not say that X_O *causes* $\dot{X}_O$ nor that $\dot{X}_O$ causes X_O even though there are directed edges between these vertices in the graph).[3] Here, we will consider a graphical representation that represents causal relations in a system of first-order differential equations in canonical form. To do so, we associate both the derivative $\dot{X}_i(t)$ and the corresponding variable $X_i(t)$ with the same vertex v_i. We use the natural labelling for the differential equations, so that a vertex g_i is associated with the differential equation for $\dot{X}_i(t)$. We then construct a *dynamical bipartite graph* $\mathcal{B}_{\text{dyn}} = \langle V, F, E\rangle$ with variable vertices v_i in V and the corresponding dynamical equation vertices $g_i \in F$ and additionally static equation vertices $f_i \in F$. The edge set E has an edge $(v_i - f_j)$ whenever $X_i(t)$ appears in the static equation f_j. Furthermore, there are edges $(v_i - g_j)$ whenever $X_i(t)$ or $\dot{X}_i(t)$ appears in the dynamic equation g_j (which includes the cases $i = j$ due to the natural labelling used).

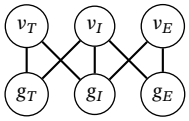
The dynamical bipartite graphs for the dynamics of the bathtub model, the viral infection, and the reaction network with feedback are given in Figures 5.6(a), 5.6(b),

[3] Bongers and Mooij (2018) provide an alternative notion of the functional graph that does have an intuitive causal interpretation.

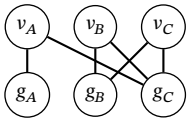
and 5.6(c), respectively. Henceforth, we will assume that the dynamical bipartite graph has a perfect matching that extends the natural labelling of the dynamic equations, i.e. such that all pairs (v_i, g_i) are matched. Application of the causal ordering algorithm to the associated dynamical bipartite graph for the model of a filling bathtub, the viral infection model, and the reaction network results in the *dynamical causal ordering graphs* in Figures 5.7(a), 5.7(b), and 5.7(c), respectively.[4]



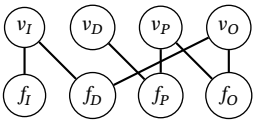
(a) Filling bathtub, dynamic bipartite graph.



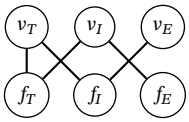
(b) Viral infection, dynamic bipartite graph.



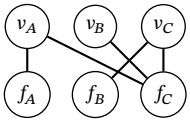
(c) Reaction network, dynamic bipartite graph.



(d) Filling bathtub, equilibrium bipartite graph.



(e) Viral infection, equilibrium bipartite graph.

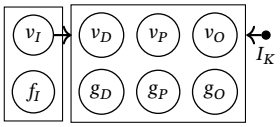


(f) Reaction network, equilibrium bipartite graph.

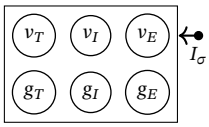
Figure 5.6: The dynamical bipartite graphs for the bathtub model, the viral infection, and the reaction network with negative feedback are presented in Figures (a), (b), and (c), respectively. The equilibrium bipartite graphs for the bathtub model, the viral infection, and the reaction network with negative feedback are given in Figures (d), (e), and (f), respectively. Comparing the equilibrium bipartite graphs with the dynamic bipartite graphs we note that that there is no edge $(v_D - f_D)$ in Figure (d) while $(v_D - g_D)$ is present in Figure (a), the edges $(v_I - f_I)$ and $(v_E - f_E)$ are not present in Figure (e) whilst the edges $(v_I - g_I)$ and $(v_E - g_E)$ are present in Figure (b), and there is no edge $(v_B - f_B)$ in Figure (f) while the edge $(v_B - g_B)$ is present in Figure (c).

The structure of the equilibrium equations can be used to construct an *equilibrium causal ordering graph* that represents the causal structure of dynamical models at equilibrium. The *equilibrium bipartite graphs* for the equilibrium equations of the filling bathtub, the viral infection, and the reaction network with feedback are given in Figures 5.6(d), 5.6(e), and 5.6(f), respectively. Application of the causal ordering algorithm to these equilibrium bipartite graphs results in the equilibrium causal ordering graphs in Figures 5.7(d), 5.7(e), and 5.7(f), respectively. Notice that variables v_i do not always end up in the same cluster with the equilibrium equation f_i of the natural labelling. For example, we see in Figure 5.7(d) that a soft intervention targeting the equilibrium equation f_O constructed from the time derivative of the outflow rate $X_O(t)$ (e.g. a change in the value of U_5) does *not* affect the value of the outflow

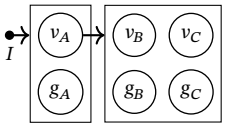
[4]Our approach here differs from dynamic causal ordering in Iwasaki et al. (1994), who include separate vertices for derivatives and variables that are linked by 'definitional' integration links. Their result is similar to the functional graph in Figure 5.5(a).



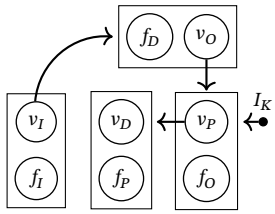
(a) Filling bathtub, dynamic causal ordering graph.



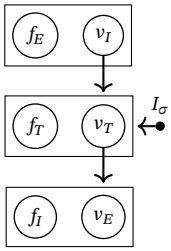
(b) Viral infection, dynamic causal ordering graph.



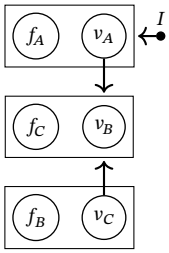
(c) Reaction network, dynamic causal ordering graph.



(d) Filling bathtub, equilibrium causal ordering graph.



(e) Viral infection, equilibrium causal ordering graph.



(f) Reaction network, equilibrium causal ordering graph.

Figure 5.7: The dynamical causal ordering graphs for the bathtub model, the viral infection, and the reaction network with negative feedback are given in Figures (a), (b), and (c), respectively. The equilibrium causal ordering graphs for the equilibrium equations of the bathtub model, the viral infection and the reaction network with negative feedback are given in Figures (d), (e), and (f), respectively. The input vertices I_K, I_σ, and I are denoted by black dots. The absence or presence of a directed path from a cluster, equation vertex, or input vertex to a variable vertex implies that a causal effect is absent or generically present, respectively.

rate X_O at equilibrium. In Section 4.6 of the previous chapter we showed that, consequently, equations and clusters that may be targeted by interventions should be clearly distinguished from the variables that could be affected by those interventions to preserve an unambiguous causal interpretation.

5.3.2.2 Identification of perfect adaptation via causal ordering

With the help of the dynamic causal ordering graph and the equilibrium causal ordering graph we can identify perfect adaptation without requiring simulations or explicit calculations. To do so, we require that Assumption 5.1 below holds.

Assumption 5.1. If there is a directed path from an input vertex to a variable vertex in the dynamic causal ordering graph of a set of first-order differential equations in canonical form, possibly with static equations as well, then there is a response of that variable to changes in the input signal some (small) time-step later.

We believe that, at least for a large class of dynamical systems, this assumption is satisfied for almost all parameter values w.r.t. the Lebesgue measure on a suitable parameter space (i.e. the property holds *generically*). It seems reasonable to assume

that a change in one of the variables or parameters that appear on the right-hand side of a first-order differential equation in canonical form at time t results in a generic change in the value of the variable on the left-hand side of that differential equation at a time $t + \Delta t$. Consider a perfect matching M for the dynamical bipartite graph $\mathcal{B}_{\text{dyn}}$ that extends the natural labelling. By construction, directed paths in $\mathcal{G}(\mathcal{B}_{\text{dyn}}, M)$, which coincide with directed paths in the dynamic causal ordering graph $\text{CO}(\mathcal{B}_{\text{dyn}})$, then correspond to transient causal effects (which may persist at equilibrium).

Under Assumption 5.1, a directed path from the input vertex to a variable vertex in the dynamical causal ordering graph implies a response to a change in the input signal. Lemma 5.1, which follows directly from Theorem 4.6, shows that at equilibrium, a change in the input signal has no effect on the value of a variable if there is no directed path from the input vertex to that variable in the equilibrium causal ordering graph. Theorem 5.1 then formulates sufficient graphical conditions for the identification of perfect adaptation.

Lemma 5.1. *Consider a model consisting of static equations, a set of first-order differential equations in canonical form, and an input signal. Assume that the equilibrium bipartite graph has a perfect matching and that the static equations and equilibrium equations derived from the first-order differential equations are uniquely solvable w.r.t. the equilibrium causal ordering graph for all relevant values of the input signal. If there is no directed path from an input vertex to a variable vertex in the equilibrium causal ordering graph then a change in the input signal has no effect on the equilibrium solution of that variable.*

These observations directly lead to our first main result. The apparent simplicity of Theorem 5.1 is due to it relying on appropriate powerful definitions and concepts such as causal ordering.

Theorem 5.1. *Consider a model that satisfies the conditions of Lemma 5.1 and assume that the associated dynamic causal ordering graph has a perfect matching that extends the natural labelling. Under Assumption 5.1, the presence of a direct path from the input signal I to a variable X_v in the dynamical causal ordering graph and the absence of such a path in the equilibrium causal ordering graph, implies that X_v perfectly adapts to changes in the input signal I.*

We see that there is a directed path from the input signal I_K to v_O in the dynamical causal ordering graph in Figure 5.7(a), while no such path exists in the equilibrium causal ordering graph in Figure 5.7(d). It follows from Theorem 5.1 that X_O perfectly adapts to changes in the input signal I_K. This is in agreement with the simulation in Figure 5.3(a). Similarly, we can verify that the amount of infected cells X_I in the viral infection model perfectly adapts to changes in the input signal I_σ and that X_C perfectly adapts to I in the reaction network with negative feedback. Clearly, it is easy to verify that perfect adaptation in the bathtub model, the viral infection model, and

the reaction network with negative feedback can be identified using the graphical criteria in Theorem 5.1 using Figure 5.7.

In Section 5.6 we construct graphical representations for a dynamical model of a basic enzymatic reaction that achieves perfect adaptation but does not satisfy the conditions in Theorem 5.1. In Appendix 5.D we will show that the biochemical reaction network in Figure 5.4(b), which Ma et al. (2009) identified as being capable of achieving perfect adaptation, does not satisfy the conditions in Theorem 5.1 either. This shows that these conditions are not necessary for the identification of perfect adaptation in dynamical systems at equilibrium. The further development of methods to analyse perfectly adapted dynamical systems that do not satisfy the conditions of Theorem 5.1 remains a challenge for future work. We believe that the methods presented in this section are a useful tool for the characterization of a large class of network topologies that are able to achieve perfect adaptation and for the automated analysis of the behaviour of certain perfectly adapted dynamical systems.

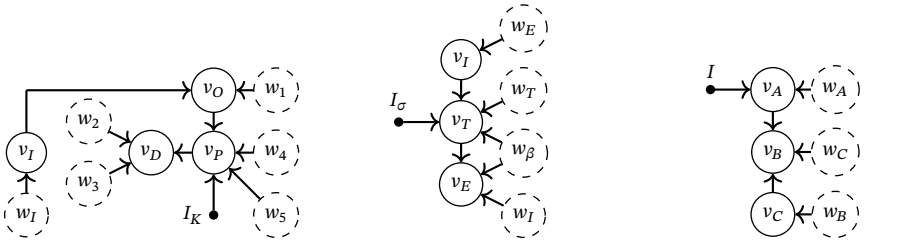
5.3.3 Recognizing perfect adaptation in data

So far we have only considered how perfect adaptation can be identified in mathematical models. In this section we focus on methods for model selection from data that is generated by perfectly adapted dynamical systems. We also discuss how the output of certain constraint-based causal discovery algorithms can be correctly interpreted for such systems.

5.3.3.1 Conditional independences

The Markov ordering graph can be used to derive conditional independences that are implied by a model at equilibrium and that can be tested in equilibrium data. The Markov ordering graphs for the equilibrium distribution of the dynamical models in the previous sections are constructed after including independent exogenous random variables to the equilibrium causal ordering graph. For the bathtub model, we let vertices $\{w_I, w_1, \dots, w_5\}$ represent independent exogenous random variables $U_I, U_1, \dots, U_5$. For the viral infection model we let w_T, w_I, w_E, w_β represent independent exogenous random variables d_T, d_I, d_E, and β in equations (5.11), (5.12), and (5.13). Finally, for the reaction network with negative feedback, we let w_A, w_B, and w_C represent independent exogenous random variables that appear in the differential equations for $X_A(t)$, $X_B(t)$, and $X_C(t)$ respectively.

The Markov ordering graphs for the filling bathtub model, the viral infection model, and the model of a reaction network with a negative feedback loop are given in Figures 5.8(a), 5.8(b), and 5.8(c) respectively. Note that the Markov ordering graph for the bathtub model coincides with the result in Dash and Druzdzel (2008), who simulated data from the bathtub model until the system reached equilibrium and then applied the PC algorithm to the equilibrium data. Although Dash (2005) inter-



(a) Filling bathtub model. (b) Viral infection model. (c) Reaction network model.

Figure 5.8: The Markov ordering graphs for the bathtub, the viral infection, and the reaction network with a negative feedback loop are given in Figures 5.8(a), 5.8(b), and 5.8(c) respectively. Exogenous variables are denoted by dashed circles and input vertices are denoted by black dots.

prets the learned graphical representation as the 'causal graph', this graph does not have a straightforward causal interpretation, see Section 5.3.3.2 and Section 4.6 in the previous chapter for more details. Instead, the d-separations in these graphs imply conditional independences in the equilibrium distribution between the corresponding variables. For example, since v_I is d-separated from v_D given v_P in the Markov ordering graph of the bathtub model at equilibrium, X_I will be independent of X_D given X_P. The implied conditional independences can for instance be used in the process of model selection. A demonstration of selecting immune responses for a viral infection model using the Markov ordering graph will be given in Chapter 6.

5.3.3.2 Interpretation of the Markov ordering graph

In this section, we will demonstrate that the Markov ordering graphs in Figures 5.8(a), 5.8(b), and 5.8(c) do not have a straightforward causal interpretation in terms of interventions, contrary to what is sometimes claimed (Dash, 2005; Iwasaki et al., 1994). To see this, we first explicitly state what we mean when we talk about 'causal relations'. In contemporary literature, the common interpretation is that, in the context of a model, an intervention on the cause brings about a change in the effect.

So let us consider an intervention on a dynamical model of a filling bathtub at equilibrium that manipulates the time-derivative $\dot{X}_D(t)$, and consequently the associated equilibrium equation f_D (e.g. by changing one of the parameters that appear in that differential equation). Assuming that the system converges to equilibrium after the intervention, the equilibrium causal ordering graph in Figure 5.7(d) tells us that this intervention on f_D generically changes the equilibrium distributions of X_O, X_P, and X_D. Since f_D is not included in Figure 5.8(a), it is not possible to read off the effect of this intervention from the Markov ordering graph of the equilibrium distribution. Clearly, if we would interpret a soft intervention on f_D as an intervention on v_D in the Markov ordering graph, then we would wrongly conclude that the interven-

tion has no effect on X_O and X_P, if we were to interpret the Markov ordering graph causally. Similarly, the equilibrium causal ordering graph in Figure 5.7(d) tells us that an intervention targeting f_P only affects X_D, whereas the Markov ordering graph in Figure 5.8(a) would incorrectly suggest that an intervention targeting v_P affects both X_P and X_D, if we were to interpret it causally. We conclude that the directed edges in the Markov ordering graph do not represent causal relations in terms of soft interventions.

Analogously, we find that the Markov ordering graph cannot be interpreted in terms of perfect ("surgical") interventions either. The correct interpretation of a directed edge $(v_i \to v_j)$ in the Markov ordering graph for the equilibrium distribution of a set of first-order differential equations is that an intervention targeting equations in the cluster of v_i (as encoded in the more informative equilibrium causal ordering graph) has an effect on the equilibrium distribution of v_j. In many systems, equilibrium equations f_i derived from differential equations for variables $X_i(t)$ end up in the same cluster as the associated variable v_i. In that case, the Markov ordering graph has an unambiguous causal interpretation. In the next chapter, which is based on Blom and Mooij (2020), we will show that the Markov ordering graph for the equilibrium distribution of dynamical models in which each variable is self-regulating does have this straightforward causal interpretation. However, perfectly adaptive systems do not have this property.

5.3.3.3 Detecting perfect adaptation

The most straightforward approach to detect perfect adaptation is to collect time-series data while experimentally changing the input signal to the system. One can then simply observe whether the variables in the system revert to their original values. Unfortunately, this type of data is not always available. Another way to identify feedback loops that achieve perfect adaptation uses a combination of observational equilibrium data, background knowledge, and experimental data. Our second main result, Theorem 5.2, gives sufficient conditions under which we can identify a system that is capable of perfect adaptation from experimental equilibrium data.

Theorem 5.2. *Consider a set of first-order dynamical equations in canonical form, satisfying the conditions of Theorem 5.1, for variables V that has equilibrium equations F with the natural labelling and consider a soft intervention targeting an equation $f_i \in F$. Assume that the system is uniquely solvable w.r.t. the equilibrium causal ordering graph both before and after the intervention and that the intervention alters the equilibrium distribution of all descendants of f_i in the equilibrium causal ordering graph. If either*

(i) the soft intervention does not change the equilibrium distribution of X_i, or

(ii) the soft intervention alters the equilibrium distribution of a variable corresponding to a non-descendant of v_i in the Markov ordering graph,

or both, then the system is capable of perfect adaptation.

Proof. If condition (i) holds there is no directed path in the causal ordering graph

from f_i to v_i in the equilibrium causal ordering graph, by the assumption that the soft intervention on f_i changes the equilibrium distribution of all its descendants. By definition of the dynamical bipartite graph there is a directed path from g_i to v_i in the dynamical causal ordering graph, because g_i and v_i end up in the same cluster (note that this follows by using the natural labelling as perfect matching and the result that the causal ordering graph does not depend on the chosen perfect matching, see Theorem 4.1 and Theorem 4.2). It follows from Theorem 5.1 that X_i perfectly adapts to an input signal I_{f_i} on f_i (i.e. a soft intervention targeting $\dot{X}_i(t)$ and thus the equilibrium equation f_i).

Suppose that (i) does not hold while (ii) does hold. By Theorem 4.5 (which roughly states that the presence of a causal effect at equilibrium implies the presence of a corresponding directed path in the equilibrium causal ordering graph) we have that f_i is an ancestor of v_i and some v_h in the equilibrium causal ordering graph, while v_i is not an ancestor of v_h in the Markov ordering graph. For a perfect matching M of the equilibrium bipartite graph let $v_j = M(f_i)$. Then v_j is in the same cluster as f_i in the equilibrium causal ordering graph by construction. Note that $j = i$ would give a contradiction, as then v_i would be an ancestor of v_h in the Markov ordering graph. Suppose that the vertex f_j, that is associated with v_j through the natural labelling, is matched to a non-ancestor of v_j in the equilibrium causal ordering graph. Because of the edge $(g_j - v_j)$ in the dynamical bipartite graph, it follows from Theorem 5.1 that X_j perfectly adapts to an input signal I_{f_j} on f_j. Therefore the system is able to achieve perfect adaptation. Now suppose that f_j is matched to an ancestor v_k of v_j, and consider the vertex f_k. The previous argument can be repeated to show perfect adaptation for X_k is present when f_k is matched to a non-ancestor of v_k in the equilibrium causal ordering graph. Otherwise, f_k must be matched to an ancestor of v_k. Note that the ancestors of v_k are a subset of the ancestors of v_j, which in turn are a subset of the ancestors of v_i. In a finite system of equations, v_i has a finite set of ancestors and therefore we eventually find, by repeating our argument, a vertex f_m that cannot be matched to an ancestor of v_m because v_m has no ancestors that are not matched to one of the vertices $f_i, f_j, f_k, \ldots$ that were considered up to that point. Because f_m is matched to a non-ancestor we then find that X_m perfectly adapts to an input signal on I_{f_m} as before. □

Based on the result in Theorem 5.2 we can device the following scheme to detect perfectly adapted dynamical systems from data and background knowledge. We start by collecting observational equilibrium data and use the PC or LCD algorithm to learn a (partial) representation of the Markov ordering graph, assuming the observational distribution to be faithful w.r.t. the Markov ordering graph. We then consider a soft intervention that changes a known equation in the first-order differential equation model (i.e. it targets a known equilibrium equation). If this intervention does not change the distribution of the variable corresponding to this target using the natural labelling, or if it changes the distribution of identifiable non-descendants of the vari-

able corresponding to the target according to the learned Markov equivalence class, we can apply Theorem 5.2 to identify the perfectly adapted dynamical system. Note that this procedure relies on several assumptions, including faithfulness.

5.4 Application to a protein signalling model

In cell biology, dynamical systems for protein signalling networks are used to model processes where information is transmitted between and into cells. The underlying dynamics of such models may have unexpected consequences for causal discovery efforts using structure learning methods, see also (Sachs, Itani, et al., 2013). Here, we specifically consider the phenomenon of perfect adaptation in a simple model of a well-studied molecular pathway. Using the technique of causal ordering to analyse the conditional independences and causal relations that are implied by the model, we elucidate the causal interpretation of the output of constraint-based causal discovery algorithms like LCD when they are applied to protein expression data.

We do not claim that the model that we analyse here is a realistic model of the protein signalling pathway. Although we will show that the model is able to explain certain observations in real-world data, this is not that surprising for a model with that many parameters.[5] Instead, our goal is to demonstrate that in systems with perfect adaptation our standard intuitions regarding the output of causal discovery algorithms might fail. Furthermore, we explain the discrepancies between the graphical representations that are produced by causal ordering for equilibrium equations and causal discovery from equilibrium data. In combination, these two techniques help us to better understand causal properties of dynamical systems at equilibrium.

5.4.1 Dynamical model

We consider the mathematical model for the Ras-Raf-Mek-Erk signalling cascade in Shin et al. (2009). Let $V = \{v_s, v_r, v_m, v_e\}$ be an index set for endogenous variables that represent the equilibrium concentrations X_s, X_r, X_m, and X_e of active Ras, Raf, Mek, and Erk proteins respectively. The dynamics are given by:

$$\dot{X}_s(t) = \frac{I(t)k_{Is}\,(T_s - X_s(t))}{(K_{Is} + (T_s - X_s(t)))\left(1 + \left(\frac{X_e(t)}{K_e}\right)^{\frac{3}{2}}\right)} - F_s k_{F_s s}\frac{X_s(t)}{K_{F_s s} + X_s(t)} \tag{5.21}$$

$$\dot{X}_r(t) = \frac{X_s(t)k_{sr}(T_r - X_r(t))}{K_{sr} + (T_r - X_r(t))} - F_r k_{F_r r}\frac{X_r(t)}{K_{F_r r} + X_r(t)} \tag{5.22}$$

[5] As mathematician John von Neumann once put it: "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk".

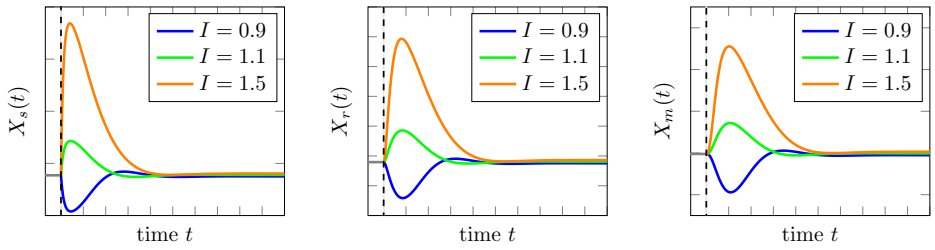
$$\dot{X}_m(t) = \frac{X_r(t)k_{rm}(T_m - X_m(t))}{K_{rm} + (T_m - X_m(t))} - F_m k_{F_m m}\frac{X_m(t)}{K_{F_m m} + X_m(t)} \tag{5.23}$$

$$\dot{X}_e(t) = \frac{X_m(t)k_{me}(T_e - X_e(t))}{K_{me} + (T_e - X_e(t))} - F_e k_{F_e e}\frac{X_e(t)}{K_{F_e e} + X_e(t)}, \tag{5.24}$$

where we assume that $I(t)$ is an external stimulus or perturbation.[6] Roughly speaking, there is a signalling pathway that goes from $I(t)$ to $X_s(t)$ to $X_r(t)$ to $X_m(t)$ to $X_e(t)$ with negative feedback from $X_e(t)$ on $X_s(t)$. As we did for the reaction network with negative feedback in Section 5.3, we consider the system under saturation conditions. For $(T_e - X_e(t)) \gg K_{me}$ and $X_e(t) \gg K_{F_e e}$ the following approximation holds:

$$\dot{X}_e(t) \approx X_m(t)k_{me} - F_e k_{F_e e}. \tag{5.25}$$

We let f_s, f_r, f_m, and f_e represent the equilibrium equations corresponding to the dynamical equations in (5.21), (5.22), (5.23), and (5.24) respectively, where we assume the input signal to have a constant value I. We simulated the model under saturation conditions until it reached equilibrium, and then we recorded the changes in the concentrations $X_s(t)$, $X_r(t)$, and $X_m(t)$ after a change in the input signal I. The results in Figure 5.3 show that Ras, Raf, and Mek revert to their original values after an initial response. Clearly the equilibrium concentrations X_s, X_r, and X_m perfectly adapt to the input signal I. The details of this simulation can be found in Appendix 5.A. In the next section we will show that the concentration of active Erk does not perfectly adapt to changes in the input signal.



(a) Concentration of active Ras. (b) Concentration of active Raf. (c) Concentration of active Mek.

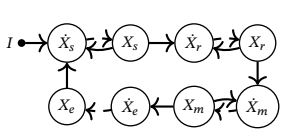
Figure 5.9: Perfect adaptation in the Ras-Raf-Mek-Erk signalling pathway. After an initial response to a change of input signal the equilibrium concentrations of active Ras, Raf, and Mek revert to their original values. The concentration of active Erk does not adapt to changes in the input signal. The details of the simulation can be found in Appendix 5.A.

[6] For simplicity, we slightly adapted the model so that the feedback mechanism through Raf Kinase Inhibitor Protein (RKIP) is not included. In the differential equation for activated Mek we therefore discarded the dependence on RKIP. The goal here is not to give the most realistic model but to elucidate the phenomenon of perfect interpretation and the causal interpretation of the Markov ordering graph for perfectly adapted dynamical systems.

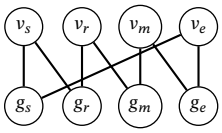
5.4.2 Graphical representations

We consider graphical representations of the protein signalling pathway. A compact representation of the structure of differential equations (5.21), (5.22), (5.23), and (5.24) is given in Figure 5.10(a). Using the natural labelling, we construct the dynamical bipartite graph in Figure 5.10(b) from the first-order differential equations. The associated dynamic causal ordering graph, with the input signal I included, is given in Figure 5.10(c).

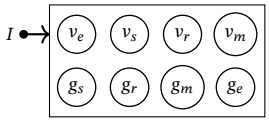
Under saturation conditions, the equilibrium equations f_s, f_r, f_m, and f_e obtained by setting equations (5.21), (5.22), (5.23), and (5.25) to zero have the bipartite structure in Figure 5.10(d). Note that there is no edge $(f_e - v_e)$ in the equilibrium bipartite graph because $X_E(t)$ does not appear in the approximation (5.25) of (5.24). The associated equilibrium causal ordering graph is given in Figure 5.10(e), where the cluster $\{I\}$ is added with an edge towards the cluster $\{v_e, f_s\}$ because I appears in equation (5.21) and in no other equations. So far we have treated all symbols in equations (5.21), (5.22), (5.23), and (5.24) as deterministic parameters. Let w_s, w_r, w_m, and w_e represent independent exogenous random variables appearing in the equilibrium equations f_s, f_r, f_m, and f_e respectively. After adding them to the causal ordering graph with edges to their respective clusters we construct the Markov ordering graph for the equilibrium distribution in Figure 5.10(f).



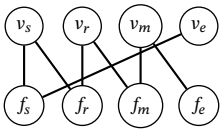
(a) Functional graph.



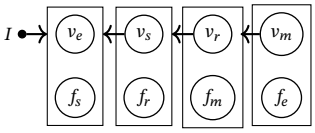
(b) Dynamic bipartite graph.



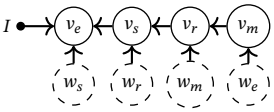
(c) Dynamic causal ordering graph.



(d) Equilibrium bipartite graph.



(e) Equilibrium causal ordering graph.



(f) Markov ordering graph.

Figure 5.10: Six graphs associated with the protein signalling pathway model under saturation conditions where indices s, r, m, e correspond to concentrations of active Ras, Raf, Mek, and Erk respectively. The functional graph, dynamic bipartite graph, and equilibrium bipartite graph are compact representations of the model. The dynamic causal ordering graph encodes the presence of transient (generic) causal effects. The equilibrium causal ordering graph represents the effects of manipulations to the equilibrium equations of the model. The Markov ordering graph implies conditional independences in the equilibrium distribution of the variables in the model via d-separations.

5.4.3 Model predictions and causal discovery

We discuss some of the predictions that can be read off from the equilibrium causal ordering graph and the Markov ordering graph for the equilibrium distribution of the model. In Section 5.5 we will test these predictions in simulated equilibrium data and real-world protein expression data.

5.4.3.1 Conditional independences and correlations

The d-separations in the Markov ordering graph imply conditional independences between the corresponding variables (according to Theorem 4.3). From the graph in Figure 5.10(f) we read off the following (implied) conditional independences:

$$I \perp\!\!\!\perp X_s, \quad I \perp\!\!\!\perp X_r, \quad I \perp\!\!\!\perp X_m, \quad X_e \perp\!\!\!\perp X_r \,|\, X_s, \quad X_e \perp\!\!\!\perp X_m \,|\, X_s, \quad X_s \perp\!\!\!\perp X_m \,|\, X_r.$$

A more extensive overview of d-separations and predicted conditional independences can be found in Appendix 5.B. Under the faithfulness assumption, the vertices that are not d-separated in the Markov ordering graph are dependent in the equilibrium distribution. Both the noise that is introduced into the model by exogenous random variables and the model parameters affect the strength of these dependencies. The Markov ordering graph in Figure 5.10(f) suggests that the correlation between Mek (i.e. X_m) and Raf (i.e. X_r) should be stronger than the correlation between Mek and Erk (i.e. X_e) because extra noise is introduced along the longer pathway Mek-Raf-Ras-Erk.

5.4.3.2 Inhibition of MEK activity

A common biological experiment that is used to study protein signalling pathways is the use of an inhibitor that decreases the activity of a protein on the pathway. Such an inhibitor slows down the rate at which the active protein is able to activate another protein. Here, we consider inhibition of Mek activity. Therefore, an experiment where the activity of Mek is inhibited has an effect on parameters in the differential equations in which $X_m(t)$ appears. Since $\dot{X}_e$ is the only child of X_m in the functional graph in Figure 5.10(a), we can interpret this experiment as a soft intervention on g_e in the dynamic bipartite graph and on f_e in the equilibrium causal ordering graph, where the rate k_{me} at which Erk is activated is decreased. Since there is a directed path from f_e to v_m, v_r, v_s, and v_e in the causal ordering graph in Figure 5.10(e), we expect that a change in an input signal I_e on f_e (e.g. a change in the parameter k_{me}) affects the equilibrium concentrations of active Mek, Raf, Ras, and Erk respectively. Note that Ras, Raf, and Mek are non-descendants of Erk in the Markov ordering graph in Figure 5.10(f), so that under the assumptions in Theorem 5.2 we can use this experiment to detect perfect adaptation in the protein pathway.

5.4.3.3 Causal discovery

Suppose that we have observational equilibrium data from the protein signalling pathway model and also experimental equilibrium data from a setting where Mek activity is inhibited. The context variable C that indicates from which setting the data was collected (observation or experimental), is not caused by any observed variable, since this variable is set externally by the experimenter at the start of the experiment. This set-up satisfies the conditions of a context variable in the LCD algorithm. In the case of Mek inhibition, this context variable represents a soft intervention on the equation f_e in the causal ordering graph in Figure 5.10(e). The Markov ordering graph that includes the context variable C (but where the independent exogenous random variables have been marginalized out) is given in Figure 5.11. To construct this graph, the context variable C is first added to the equilibrium causal ordering graph in Figure 5.10(e) as a singleton cluster with an edge towards the cluster $\{v_m, f_e\}$. The Markov ordering graph is then constructed from the resulting directed cluster graph in the usual way. From this, we can read off (conditional) independences to find the LCD triples that are implied by the equilibrium equations of the model. We find that (C, v_m, v_r), (C, v_m, v_s), (C, v_m, v_e), (C, v_r, v_s), (C, v_r, v_e), and (C, v_s, v_e) are all LCD triples.

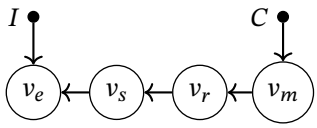


Figure 5.11: Markov ordering graph of the protein signalling pathway with the context variable C included. This context variable indicates whether a cell was treated with a Mek inhibitor or not. The input signal I is taken to be constant over time, but is varied across experiments.

With a conditional independence oracle, and under the faithfulness assumption, the output of complete causal discovery algorithms (like the PC algorithm if causal sufficiency is assumed, or more generally, the FCI algorithm) would be the Markov equivalence class of the Markov ordering graph. Here, it is important to note that the Markov ordering graph does not have a straightforward *causal* interpretation for this perfectly adapted dynamical system. The reasoning is similar to the discussion in Section 5.3.3.2.

For the protein signalling model, the common biological understanding of the underlying causal mechanism is that Raf activates Mek, Mek activates Erk, and that it is very likely that there is negative feedback from a protein downstream of Erk on Raf (Fritsche-Guenther et al., 2011). Therefore, even though Raf is a direct cause of Mek (see Figure 5.10(c)), in line with the biological consensus, Mek is also an indirect cause of Raf. At equilibrium, Raf is no longer a cause of Mek due to the perfect adaptation. This leads to a situation where there is a directed path from Raf to Mek

in biological consensus networks (like the one in Sachs, Perez, et al. (2005) where the feedback loop from Erk to Raf has not been included) while there is a directed path in the *opposite* direction in the Markov ordering graph for the equilibrium equations. If we were to apply the LCD algorithm to the experimental Mek inhibition equilibrium data, we would detect a directed path from Mek to Raf but not from Raf to Mek. The conclusion that 'Mek is a cause of Raf' while no causal relation from Raf on Mek can be detected could, at first glance, appear to be at odds with expert knowledge. Similar observations of an apparent "causal reversal" in protein interaction networks have been observed more often, see also Mooij and Heskes (2013), Mooij, Magliacane, et al. (2020), and Triantafillou et al. (2017). The phenomenon of perfect adaptation can help to explain differences between biological consensus networks and the output of causal discovery algorithms. We have shown that a simple model that is capable of perfect adaptation can explain some of the differences between the output of standard constraint-based causal discovery algorithms and biological consensus networks that represent other aspects of the underlying mechanisms. Confusion about causal relations can be avoided by explicitly specifying the interventions that correspond to the claimed causal effects, by distinguishing between statements about the equilibrium distribution and the dynamical model, and analysing models with our approach based on the technique of causal ordering.

In Chapter 6, which is based on Blom and Mooij (2020), we will show that the causal relations and conditional independences that are implied by the equilibrium equations of a dynamical model may not be preserved when it is combined with another model. They discuss how, for dynamical systems at equilibrium that are only partially modelled and observed, one can reason about the presence of unobserved feedback loops and variables that are not self-regulating in the whole system. In Appendix 5.C, we show that these ideas can also be applied when only $X_s(t)$, $X_r(t)$, and $X_m(t)$ are included as endogenous variables in the perfectly adapted protein signalling model that we presented in this section.

5.5 Experiments

In this section we present simulations to confirm the qualitative model predictions for the protein signalling model in Section 5.4. We then consider data from real-world experiments in order to test the validity of the protein signalling model.

5.5.1 Simulations

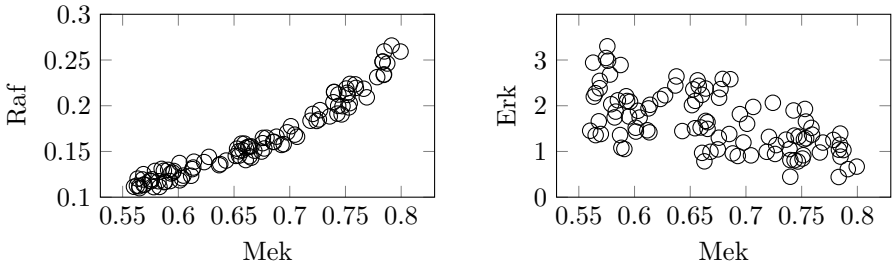
We took as input signal $I(t) = i$, with i sampled from a uniform distribution on the interval $(0.5, 1.5)$. We also drew random samples for the parameters k_{Is}, k_{sr}, k_{rm}, and k_{me} from uniform distributions on the intervals $(1.2, 1.5)$, $(2.4, 3.0)$, $(1.7, 2.0)$, and $(0.7, 1.0)$ respectively. We then simulated the dynamical model in equations (5.21) to (5.24) with parameter settings: $K_{Is} = 1.0$, $K_e = 1.5$, $F_s = 1.0$, $k_{F_{ss}} = 1.0$, $K_{F_{ss}} = 0.9$,

$K_{sr} = 1.0$, $F_r = 0.3$, $k_{F_r r} = 1.0$, $K_{F_r r} = 0.8$, $K_{rm} = 0.9$, $F_m = 0.2$, $k_{F_m m} = 1.0$, $K_{F_m m} = 1.2$, $K_{me} = 0.0001$, $F_e = 0.7$, $k_{F_e e} = 1.2$, and $K_{F_e e} = 0.0001$. The parameters were chosen in such a way that the approximation in equation (5.25) of equation (5.24) is valid and so that the system converges to an equilibrium where the concentrations of active proteins are strictly between 0 and $T_s = 1.0$, $T_r = 1.0$, $T_m = 1.0$, and $T_e = 5.0$ respectively. We experimented with other parameter values as well, and observed that the analysis of the qualitative behaviour of the model that we present here is valid for many values of the parameters.

5.5.1.1 Conditional independences and correlation strength

To test whether the conditional independences in Section 5.4.3.1 hold when the system is at equilibrium, we ran the simulation $n = 500$ times until it reached equilibrium and recorded the equilibrium concentrations X_s, X_r, X_m, and X_e. We tested all (conditional) independences with a maximum of one conditioning variable using Spearman's rank correlation test with a p-value threshold of 0.01. This way, we retrieved all predicted (conditional) independences and all predicted (conditional) dependences. Table 5.1 in Appendix 5.B provides a list of the estimated correlations and the corresponding p-values.[7]

In Section 5.4.3 we discussed how the Markov ordering graph for the simple model of a protein signalling pathway suggests that the correlation between Mek and Raf should be stronger than the correlation between Mek and Erk. The scatter plots in Figure 5.12 below confirm this prediction.



(a) Strong correlation between active Raf and Mek. (b) Weaker correlation between active Mek and Erk.

Figure 5.12: Two scatter plots of the Mek-Raf and Mek-Erk concentrations of 100 samples of the simulation experiment of the protein signalling pathway in Section 5.5.1.1. Note the difference in the signal to noise ratio. The correlation between Mek and Raf is clearly stronger than the correlation between Mek and Erk. The estimate of the rank correlation between Mek and Raf is 0.98 and between Mek and Erk it is -0.51.

[7] Because the LCD algorithm only uses conditional independence test with a maximum of one variable in the conditioning test, we do not consider conditional independence tests with larger conditioning sets in this work. We did experiment with larger conditioning sets but we were not able to retrieve all predicted conditional dependences with our parameter settings and only $n = 500$ samples.

Different parameter regimes correspond to different qualitative behaviour of the protein pathway model. For example, when almost all of the Erk molecules are activated we have that $X_e \approx T_e$. If we repeat the experiment in Section 5.5.1.1 with $K_e = 100$ and with K_{Is} drawn from a uniform distribution on the interval $(1.9, 2.5)$ then we find that the correlation between X_m and X_e is 0.054 with a p-value of 0.087. The correlation between X_m and X_r is 0.76 with a p-value smaller than $2.2e^{-16}$. The dependence between the concentrations of active Mek and Erk thus disappears under saturation conditions for Erk, while the correlation between Mek and Raf remains strong.

5.5.1.2 Inhibition of MEK activity and LCD

We assessed the effect of decreasing the activity of Mek on the equilibrium concentrations of Ras, Raf, Mek, and Erk. To that end, we simulated the model with fixed parameters $I = 1.0$, $k_{Is} = 1.0$, $k_{sr} = 1.0$, $k_{rm} = 1.0$, and $k_{me} = 1.1$ until it reached equilibrium. We then decreased the parameter that controls the activity of Mek to $k_{me} = 1.0$. The recorded responses of the concentrations of active Ras, Raf, Mek, and Erk are displayed in Figure 5.13. From this we confirm our prediction that inhibition of Mek activity affects the equilibrium concentrations of Ras, Raf, Mek, and Erk.

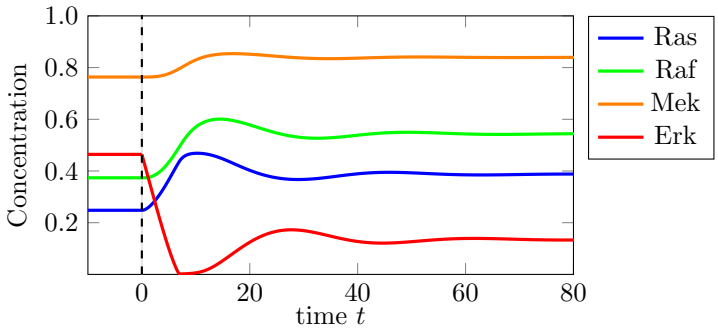


Figure 5.13: Simulation of the response of the concentrations of active Ras, Raf, Mek, and Erk after inhibition of the activity of Mek. The system starts out in equilibrium with $k_{me} = 1.1$. The concentrations of Ras, Raf, Mek, and Erk are recorded after the parameter controlling Mek activity is decreased to $k_{me} = 1.0$ from $t = 0$ on.

We also simulated a scenario where the inhibition of Mek activity is treated as a context variable, that can be used to apply the LCD algorithm. We ran the simulation $n = 500$ times with $k_{me} = C$, where the context variable C is drawn from a uniform distribution on the interval $(0.98, 1.1)$. To avoid deterministic relations, we drew the parameter $k_{F_e e}$ from a uniform distribution on $(0.7, 1.0)$. We ran the simulations until the system reached equilibrium and recorded the equilibrium values of the variables.

We then applied the LCD algorithm to search for LCD triples in this equilibrium data with context variable C. For the conditional independence tests we used Spearman's rank correlation with a p-value threshold of 0.01. We found the expected LCD triples (C, v_m, v_r), (C, v_m, v_s), (C, v_m, v_e), (C, v_r, v_s), (C, v_r, v_e), (C, v_s, v_e) and no others.

5.5.2 Protein expression data

In this section we test the predictions of our model on protein signalling data from real-world experiments. For a thorough description of these experiments we refer to Sachs, Perez, et al. (2005) and Lun et al. (2017).

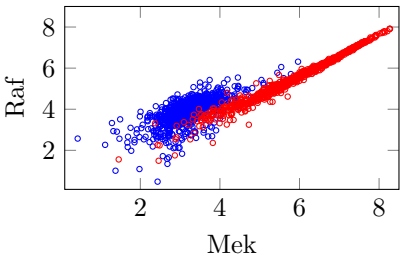
5.5.2.1 Correlation strength

In the simulations of the simple protein signalling pathway model we demonstrated that, as predicted, the correlation between Raf and Mek was much stronger than the correlation between Mek and Erk, and the latter correlation completely disappeared in a setting where Erk was saturated. We test these correlations in a multivariate single-cell protein expression dataset that was used in Sachs, Perez, et al. (2005). We considered data that was pooled from different experimental settings, in which cells were exposed to stimulatory and/or inhibitory interventions.[8] Using Spearman's rank correlation we found a correlation of 0.78 with a p-value smaller than $2.2e^{-16}$ between Raf and Mek. The correlation between Mek and Erk was -0.023 with a p-value of 0.065. The biological consensus according to Sachs, Perez, et al. (2005) is that there is a signalling pathway from Raf to Mek to Erk. Note that the simple model in Section 5.4.1 provides an explanation as to why we are not able to reject the null hypothesis of zero correlation between Mek and Erk.

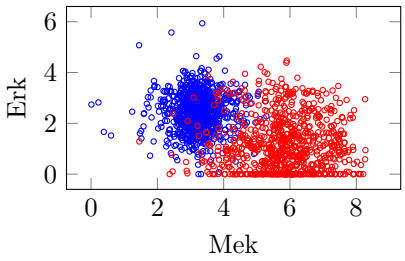
5.5.2.2 Inhibition of MEK activity

The experimental protein expression data used in Sachs, Perez, et al. (2005) includes data from an experiment where cells were perturbed with U0126, which is a known inhibitor of Mek activity. Figures 5.14(a) and 5.14(b) show the log-transformed concentrations of active Raf, Mek, and Erk proteins after treatment with and without the U0126 perturbation. In both cases the sample was treated with anti-CD3 and anti-CD28, see Sachs, Perez, et al. (2005) for more details on the dataset. These plots clearly show that inhibition of Mek activity results in an increase in the concentrations of active Raf and active Mek and a reduction in the concentration of active Erk. This is in agreement with observations in the simulation study.

[8] In particular, we used specific perturbation conditions with the following reagents: β2cAMP, AKT inhibitor, U0126, PMA, G06976, Psitectorigenin, LY294002. In some conditions the general perturbation by the reagent anti-CD3/CD28 also was included. See Sachs, Perez, et al. (2005) for more details.



(a) Scatter plot for concentrations of active Mek and Raf.



(b) Scatter plot for concentrations of active Mek and Erk.

Figure 5.14: Two scatter plots of the active Mek, Raf, and Erk concentrations in the Sachs data. The red circles correspond to the sample treated with anti-CD3/CD28 and the Mek-activity inhibitor U0126. The blue circles correspond to the sample treated only with anti-CD3/CD28. The inhibition of Mek results in an increase in the concentration of active Mek and Raf, whereas the concentration of active Erk is reduced.

5.5.2.3 Detecting LCD Triples

The perturbations with different inhibitors and stimulants in the data that was used in Sachs, Perez, et al. (2005) can be treated as context variables (Mooij, Magliacane, et al., 2020). For the context variable associated with a specific perturbation (e.g. AKT inhibitor) data points collected from the condition with that perturbation were indicated with a 1, while data points collected from other experimental conditions were indicated with a 0. We then searched for LCD triples involving Raf, Mek, and Erk. We detected the following LCD triples using Spearman's rank correlation test with a p-value threshold of 0.01: (AKT inhibitor, Raf, Mek), (LY294002, Raf, Mek), (Psitectorigenin, Raf, Mek), (AKT inhbitor, Raf, Erk), and (β2cAMP, Raf, Erk). This suggests that, if the system was at equilibrium under saturation conditions, there should be directed paths from Raf to Mek in the Markov ordering graph. Although this observation agrees with the biological consensus networks that include negative feedback from Erk on Raf, there is no directed path from Raf to Mek in the Markov ordering graph for the saturated equilibrium model in Figure 5.10(f). However, the results of LCD seem to strongly depend on the implementation details. For example, both Mooij, Magliacane, et al. (2020) and Boeken et al. (2020) report LCD triples in this dataset that imply a directed path from Mek to Raf, as was predicted by the saturated equilibrium model. Furthermore, the assumption that the system was saturated and at equilibrium may have been violated. We also found LCD triples that imply a directed path from Raf to Erk in the Markov ordering graph, if the system was at equilibrium under saturation conditions. Such LCD triples were also reported by Boeken et al. (2020), but not by Mooij, Magliacane, et al. (2020). The detected LCD triples agree with the direction of edges in the Markov ordering graph for the saturated equilibrium model in Figure 5.10(f).

We also searched for LCD triples involving Mek and Erk in the protein signalling data in Lun et al. (2017). This data was collected at different time-points after the abundance of certain proteins was over-expressed in an experiment. We treated the measured expression levels of targeted proteins as context variables in the LCD algorithm, as Blom, Klimovskaia, et al. (2018) do. We followed the pre-processing steps in Blom, Klimovskaia, et al. (2018), and selected a subset of the perturbations for our analysis. There were three replica's of each experiment and we searched for LCD triples that consistently appeared in all replicas, using Spearman's rank correlation test with a p-value threshold of 0.01. This way, we found the triple (p70RSK, Mek, Erk) at $t = 5$ from the data of experiments where p70RSK was over-expressed and the triple (p38, Erk, Mek) at $t = 60$ from the data of experiments where p38 was over-expressed. The LCD triple (p38, Erk, Mek) suggests that, under saturation conditions and at equilibrium, there is a directed path from Erk to Mek in the Markov ordering graph. The fact that this does not agree with the Markov ordering graph in Figure 5.10(f) could be due to a violation of the assumptions of saturation or equilibrium. The LCD triple (p70RSK, Mek, Erk) suggests that there should be a directed path from Mek to Erk in the Markov ordering graph. This is in agreement with the predictions of the protein pathway model at equilibrium and under saturation conditions.

In conclusion, LCD results on real-world data depend on implementation details. In some cases, they agree with the Markov ordering graph in Figure 5.10(f), in other cases they don't.

5.6 Discussion

In this section, we will discuss that the notions of the causal Markov and faithfulness conditions, which are used to tie causal relations to conditional independences in the setting of causal DAGs, are ambiguous in the context of perfectly adapted systems. We also give an example of a dynamical system for which rewriting of the equilibrium equations reveals a stronger Markov property. We believe these are interesting topics for future work, because understanding the conditions under which the output of constraint-based causal discovery algorithms has a straightforward and intuitive causal interpretation may increase the impact of causal discovery in application domains where perfectly adapted systems frequently occur, if the observed lack of robustness of these methods can be overcome.

5.6.1 Ambiguity of causal Markov and faithfulness conditions

The causal faithfulness condition and causal Markov condition can be used to relate graphs that represent causal relations between variables to properties of the probability distribution on the space of these variables. In this work, we explicitly differentiate between causal relations in a dynamical model and in the equilibrium model. Fur-

thermore, we also make a clear distinction between the Markov ordering graph (representing conditional independences) and the causal ordering graph (which encodes causal relations). In this context, the commonly used notions of causal faithfulness and the causal Markov condition become ambiguous.

To see this, consider the dynamical causal ordering graph for the viral infection model in Figure 5.7(b). Note that v_T, v_I, v_E share a cluster and that I_σ is a cause of $X_T(t)$, $X_I(t)$, and $X_E(t)$. At the same time, the Markov ordering graph for the equilibrium equations implies that X_I is independent of I_σ, see Figure 5.7(e). If we put a probability distribution on I_σ, we could say that the equilibrium distribution of the variables in the viral infection model is not *faithful* to cause-effect relations implied by the dynamic causal ordering graph.

Additionally, in the equilibrium causal ordering graph for the reaction network with a feedback loop in Figure 5.7(f) we see that the only direct cause, in terms of interventions on the equilibrium model, of X_A is I and that X_A is not a cause of X_C. However, the dynamical causal ordering graph in Figure 5.7(c) indicates that we cannot expect $X_A(t)$ to be independent of $X_C(t)$ given I, when the system has not yet reached equilibrium. Roughly speaking, the distribution of a system that is initialized with certain initial conditions and that has not yet reached equilibrium at time t is not Markov with respect to the cause-effect relations in the equilibrium model. We consider a study into more generally applicable formulations of these concepts that could be used also for perfectly adaptive systems to be outside the scope of the current paper.

5.6.2 Rewriting equations may reveal additional structure

Theorem 5.1 specifies sufficient but not necessary conditions for the presence of perfect adaptation. The equilibrium distribution of some systems is not faithful to the Markov ordering graph associated with the equilibrium equations in the model. Here, we will discuss the dynamical model for the basic enzymatic reaction in Chapter 3 and we will demonstrate that this model is capable of perfect adaptation, does not satisfy the conditions in Theorem 5.1, and that the presence of directed paths in the equilibrium causal ordering graph does not imply the presence of a causal effect at equilibrium. The basic enzyme reaction models a substrate S that reacts with an enzyme E to form a complex C, which is converted into a product P and the enzyme E. The dynamical equations for the concentrations $X_S(t), X_E(t), X_C(t)$, and $X_P(t)$ are given by:

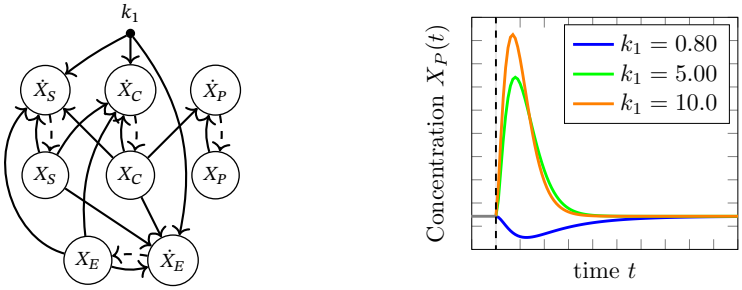
$$\dot{X}_S(t) = k_0 - k_1 X_S(t) X_E(t) + k_{-1} X_C(t), \tag{5.26}$$

$$\dot{X}_C(t) = k_1 X_S(t) X_E(t) - (k_{-1} + k_2) X_C(t), \tag{5.27}$$

$$\dot{X}_E(t) = -k_1 X_S(t) X_E(t) + (k_{-1} + k_2) X_C(t), \tag{5.28}$$

$$\dot{X}_P(t) = k_2 X_C(t) - k_3 X_P(t), \tag{5.29}$$

where $k_{-1}, k_0, k_1, k_2, k_3$ and the initial conditions are independent exogenous random variables S_0, C_0, E_0, and P_0 taking value in $\mathbb{R}_{>0}$ (Murray, 2002). We included the parameter k_1 into the functional graph of this system in Figure 5.15(a). Since there is a path from k_1 to $X_P(t)$ we would expect that a change in k_1 would generically lead to a transient response of $X_P(t)$. We verified this by simulating this model with $k_{-1} = 1.0$, $k_0 = 1.0$, $k_1 = 1.0$, $k_2 = 0.8$, $k_3 = 2.5$ and with initial conditions $X_S(0) = 1.0$, $X_E(0) = 0.5$, $X_C(0) = 0.5$, and $X_P(0) = 1.0$ until the system reached equilibrium. We then recorded the response after changing the input signal k_1. Figure 5.15(b) shows that X_P perfectly adapts to changes in the input signal k_1.



(a) Functional graph of basic enzyme reaction. (b) Perfect adaptation in basic enzyme reaction.

Figure 5.15: The functional graph of the basic enzyme reaction modelled by equations (5.26), (5.27), (5.28), and (5.29) in Figure 5.15(a) shows that there is a directed path from an input signal that controls the parameter k_1 to all endogenous variables X_S, X_C, X_E, X_P. Figure 5.15(b) shows that the concentration X_P perfectly adapts after an initial transient response to a persistent change in the parameter k_1.

The equilibrium equations of the model are given by:

$$f_S : \quad k_0 - k_1 X_S X_E + k_{-1} X_C = 0, \tag{5.30}$$
$$f_C : \quad k_1 X_S X_E - (k_{-1} + k_2) X_C = 0, \tag{5.31}$$
$$f_E : \quad -k_1 X_S X_E + (k_{-1} + k_2) X_C = 0, \tag{5.32}$$
$$f_P : \quad k_2 X_C - k_3 X_P = 0, \tag{5.33}$$
$$f_{CE} : \quad X_C + X_E - (C_0 + E_0) = 0, \tag{5.34}$$

where the last equation is derived from the constant of motion $X_C(t) + X_E(t)$, as was discussed in detail in Section 3.2.1. Via the extended causal ordering algorithm (see Algorithm 3 in the previous chapter) the equilibrium causal ordering graph in Figure 5.16 can be constructed from the equilibrium equations in the model. There is a directed path from k_1 to v_P in the Markov ordering graph. Therefore, even though the basic enzyme reaction does achieve perfect adaptation, we see that it does not satisfy the conditions of Theorem 5.1. The simulation in Figure 5.15(b) indicates that there

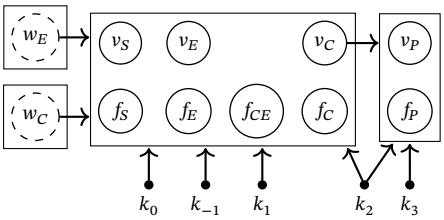


Figure 5.16: The equilibrium causal ordering graph constructed from the equilibrium equations of basic enzyme reaction modelled by equilibrium equations f_S, f_C, f_E, f_p, and f_{CE}.

is no causal effect of k_1 on X_P at equilibrium. The basic enzyme reaction is an example of a system for which directed paths in the equilibrium causal ordering graph do not imply generic causal relations between variables.

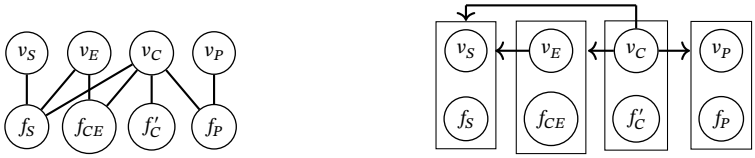
By combining equilibrium equations we can achieve stronger conclusions for this particular case. For instance, we could consider the equation f'_C, obtained from summing equations f_S and f_C:

$$f'_C: \qquad k_0 - k_2 X_C = 0, \tag{5.35}$$

in combination with f_S, f_P, and f_{CE}. The equilibrium equations f_C and f_E can be dropped because they are linear combinations of the other equations. The equilibrium bipartite graph and equilibrium causal ordering graph associated with f_S, f_{CE}, f'_C, and f_P are given in Figure 5.17. An intervention targeting k_1 would correspond to a soft intervention targeting the equation f_S. Furthermore, in Figure 5.17(b) there is no directed path from f_S to v_P, while in Figure 5.16 such a path does exist. Using Theorem 4.5 we conclude that this soft intervention has no effect on the concentration of v_P are equilibrium. Clearly, the equilibrium causal ordering graph in Figure 5.17(b) for the rewritten equilibrium equations reveals more structure than the one in Figure 5.16 for the original equilibrium equations. Furthermore, the two causal ordering graphs do not model the same set of perfect interventions. For example, the (non)effects of an intervention targeting the cluster $\{v_S, f_S\}$ in the causal ordering graph in Figure 5.17(b), where f_S is replaced by an equation $v_S = \xi_S$ setting v_S equal to a constant $\xi_S \in \mathbb{R}_{>0}$, cannot be read off from the equilibrium causal ordering graph in Figure 5.16.

5.7 Conclusion

Perfect adaptation is the phenomenon that a dynamical system initially responds to a change of input signal but reverts back to its original value as the system converges to equilibrium. We used the technique of causal ordering to obtain sufficient graphical conditions to identify perfect adaptation in a set of first-order differential equations.



(a) Equilibrium bipartite graph.

(b) Equilibrium causal ordering graph.

Figure 5.17: The equilibrium bipartite graph and equilibrium causal ordering graph associated with the basic enzyme reaction after rewriting the equilibrium equations. The absence of a directed path from f_S to v_E, v_C, v_P indicates that a soft intervention targeting f_S has no effect on those variables at equilibrium.

The notion of a *dynamical causal ordering graph* was introduced to support our explanation of the differences between the equilibrium and dynamical causal structure. Moreover, we showed how perfect adaptation can be detected in equilibrium observational and experimental data of soft interventions with known targets.

Constraint-based causal discovery algorithms operate by constructing a graphical representation of the conditional independences in data or a probability distribution, and then reason back about what this implies for the causal relations between variables. Under additional assumptions (such as the causal Markov and faithfulness conditions) the learned graph can be interpreted causally, but these assumptions cannot generally be tested in real-world data. We demonstrated that for perfectly adapted dynamical systems the output of causal discovery algorithms applied to equilibrium data may appear to be at odds with our understanding of the mechanisms that drive the system, suggesting that the standard causal Markov and causal faithfulness conditions are not appropriate for such systems. Therefore, in practical applications of causal discovery to equilibrium data, we should avoid ambiguous terminology that obscures the possible differences between causal relations that manifest themselves in dynamical and equilibrium settings.

We illustrated our ideas on a variety of dynamical models and corresponding equilibrium equations. We applied the technique that we presented in this paper to a model for a well-studied protein signalling pathway and tested our predictions both in simulations and on real-world protein expression data. This turned out to be beneficial for explanation of the differences between the causal interpretation of the results of local causal discovery in real-world data and biological consensus networks. We hope that the results presented in this work will bring the world of causal inference closer to application domains that use dynamical models, and vice versa.

5.A Perfect adaptation simulations

For the simulations in Figures 5.3 and 5.9 of the model of a filling bathtub, the viral infection model, the reaction network with a feedback loop, and the protein pathway we used the settings listed below. Since we only simulated a single response, we used constant values for the exogenous random variables as well.

(i) Filling bathtub: First we recorded the behaviour of the system for the parameters $I_K = 1.2$, $U_I = 5.0$, $U_1 = 1.1$, $U_2 = 1.0$, $U_3 = 1.2$, $U_4 = 1.0$, $U_5 = 0.8$, $g = 1.0$ until it reached equilibrium. We then changed the input parameter I_K to 0.8, 1.0, and 1.3 and recorded the response until the system reverted to equilibrium.

(ii) Viral infection: For the parameter settings $I_\sigma = 1.6$, $d_T = 0.9$, $\beta = 0.9$, $f = 1.0$, $d_I = 0.3$, $k = 1.5$, $a = 0.1$, $d_E = 0.25$, we simulated the model until it reached equilibrium. We changed the input parameter I_σ to 1.1, 1.3, and 2.0 and recorded the response until equilibrium was reached.

(iii) Reaction Network: We simulated the model until it reached equilibrium with parameters $I = 1.5$, $k_{IA} = 1.4$, $K_{IA} = 0.8$, $F_A = 1.1$, $k_{F_AA} = 0.9$, $K_{F_AA} = 1.2$, $k_{CB} = 0.6$, $K_{CB} = 0.0001$, $F_B = 0.7$, $k_{F_BB} = 0.7$, $K_{F_BB} = 0.0001$, $k_{AC} = 2.1$, $K_{AC} = 1.5$, $k_{BC} = 0.7$, $K_{BC} = 0.6$. The settings were chosen in such a way that the saturation conditions $(1 - X_B(t)) \gg K_{CB}$ and $X_B(t) \gg K_{F_BB}$ were satisfied. We then changed the input signal to 0.25, 1.0, and 10.0 and recorded the response.

(iv) Protein pathway: The parameter settings of the simulation were $I = 1.0$, $k_{Is} = 1.0$, $T_s = 1.0$, $K_{Is} = 1.0$, $K_e = 1.5$, $F_s = 1.0$, $k_{F_ss} = 1.0$, $K_{F_ss} = 0.9$, $k_{sr} = 1.0$, $K_{sr} = 1.0$, $T_r = 1.0$, $F_r = 0.3$, $k_{F_rr} = 1.0$, $K_{F_rr} = 0.8$, $k_{rm} = 1.0$, $K_{rm} = 0.9$, $T_m = 1.0$, $F_m = 0.2$, $k_{F_mm} = 1.0$, $K_{F_mm} = 1.2$, $k_{me} = 1.0$, $K_{me} = 0.0001$, $T_e = 1.0$, $F_e = 0.7$, $k_{F_ee} = 1.2$, $K_{F_ee} = 0.0001$. This ensured that the saturation conditions $(T_e - X_e(t)) \gg K_{me}$ and $X_e(t) \gg K_{F_ee}$ were satisfied. After the system reached equilibrium we changed the input signal to 0.9, 1.1, and 1.5 and recorded the response.

The qualitative behaviour that we presented in Figure 5.3 can be observed for a range of parameter values and does not require exact tuning of the parameters.

5.B Conditional independences

The Markov ordering graph in Figure 5.10(f) was derived from the equilibrium equations of the protein pathway model under saturation conditions. From this we can read off the following d-separations:

$$I \overset{d}{\perp} v_s, \quad I \overset{d}{\perp} v_s \,|\, v_r, \quad I \overset{d}{\perp} v_s \,|\, v_m, \quad I \overset{d}{\perp} v_s \,|\, \{v_r, v_m\},$$

$$I \overset{d}{\perp} v_r, \quad I \overset{d}{\perp} v_r \,|\, v_s, \quad I \overset{d}{\perp} v_r \,|\, v_m, \quad I \overset{d}{\perp} v_r \,|\, \{v_s, v_m\},$$

$$I \overset{d}{\perp} v_m, \quad I \overset{d}{\perp} v_m \,|\, v_s, \quad I \overset{d}{\perp} v_m \,|\, v_r, \quad I \overset{d}{\perp} v_m \,|\, \{v_s, v_r\},$$

$$v_e \overset{d}{\perp} v_r \,|\, v_s, \quad v_e \overset{d}{\perp} v_r \,|\, \{v_s, v_m\}, \quad v_e \overset{d}{\perp} v_r \,|\, \{v_s, I\}, \quad v_e \overset{d}{\perp} v_r \,|\, \{v_s, v_m, I\},$$

$$v_e \overset{d}{\perp} v_m \,|\, v_s, \quad v_e \overset{d}{\perp} v_m \,|\, v_r, \quad v_e \overset{d}{\perp} v_m \,|\, \{v_s, v_r\},$$

$$v_e \overset{d}{\perp} v_m \,|\, \{v_s, I\}, \quad v_e \overset{d}{\perp} v_m \,|\, \{v_r, I\}, \quad v_e \overset{d}{\perp} v_m \,|\, \{v_s, v_r, I\},$$

$$v_s \overset{d}{\perp} v_m \,|\, v_r, \quad v_s \overset{d}{\perp} v_m \,|\, \{v_e, v_r\}, \quad v_s \overset{d}{\perp} v_m \,|\, \{v_r, I\}, \quad v_s \overset{d}{\perp} v_m \,|\, \{v_e, v_r, I\}.$$

It is easy to check that the equilibrium equations and endogenous variables in this model are uniquely solvable w.r.t. the causal ordering graph (see Definition 4.8). Therefore, Theorem 4.3 tells us that the d-separations above imply conditional independences between the variables in the model. Table 5.1 shows that the conditional independences with a maximum conditioning set of size one that are implied by the Markov ordering graph are also present in the simulated data.

5.C Reasoning about feedback loops

Consider the protein signalling model under saturation conditions that is defined by equations (5.21), (5.22), (5.23), and (5.25). Suppose that the system is only partially modelled and that $X_e(t)$ is treated as a latent exogenous variable U_e in the submodel for $X_s(t)$, $X_r(t)$, and $X_m(t)$ defined by equations:

$$\dot{X}_s(t) = \frac{I(t)k_{Is}\,(T_s - X_s(t))}{(K_{Is} + (T_s - X_s(t)))\left(1 + \left(\frac{U_e}{K_e}\right)^{\frac{3}{2}}\right)} - F_s k_{F_s s}\frac{X_s(t)}{K_{F_s s} + X_s(t)}, \tag{5.36}$$

$$\dot{X}_r(t) = \frac{X_s(t)k_{sr}(T_r - X_r(t))}{K_{sr} + (T_r - X_r(t))} - F_r k_{F_r r}\frac{X_r(t)}{K_{F_r r} + X_r(t)}, \tag{5.37}$$

$$\dot{X}_m(t) = \frac{X_r(t)k_{rm}(T_m - X_m(t))}{K_{rm} + (T_m - X_m(t))} - F_m k_{F_m m}\frac{X_m(t)}{K_{F_m m} + X_m(t)}. \tag{5.38}$$

Application of the causal ordering technique to the equilibrium equations associated with these differential equations results in the Markov ordering graph in Figure 5.18. Assuming faithfulness, the d-connections in this graph indicate that the input signal I is dependent on the equilibrium distributions of X_s, X_r, and X_m. However, if X_s, X_r, and X_m were generated by the larger model with the Markov ordering graph in Figure 5.10(f), we know that a statistical test would indicate that they are in-

dependent. The discrepancy between the Markov ordering graph for the equilibrium equations of the submodel and these statistical tests would not be due to a faithfulness violation but could be wholly explained by a holistic modelling approach (i.e. by not assuming all unobserved causes to be exogenous to the observed variables). According to Corollary 6.1 and Proposition 6.1 in the next chapter, which is based on Blom and Mooij (2020), the discrepancy between the observed and predicted conditional independences implies the presence of a non-selfregulating variable and an unobserved dynamical feedback loop involving X_s, X_r, and X_m, if we assume faithfulness. This is in agreement with the fact that the dynamic variable $X_e(t)$ is not self-regulating and that there is a feedback loop in the dynamical causal ordering graph (indicated by the cluster $\{v_s, v_r, v_m, v_e, f_s, f_r, f_m, f_e\}$) in the saturated protein signalling model in Section 5.4. Interestingly, we can infer the presence of feedback without modelling or observing $X_e(t)$.

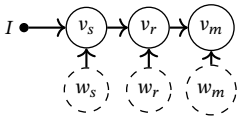


Figure 5.18: Markov ordering graph for the partial model of the protein signalling pathway model given by equations (5.36), (5.37), and (5.38).

5.D IFFLP Network

The IFFLP topology in Ma et al. (2009) that we discussed in Section 5.3.1.3 could be a graphical representation of the following differential equations:

$$\dot{X}_A(t) = I(t)k_{IA}\frac{(1-X_A(t))}{K_{IA}+(1-X_A(t))} - F_A k_{F_AA}\frac{X_A(t)}{K_{F_AA}+X_A(t)}, \tag{5.39}$$

$$\dot{X}_B(t) = X_A(t)k_{AB}\frac{(1-X_B(t))}{K_{AB}+(1-X_B(t))} - F_B k_{F_BB}\frac{X_B(t)}{K_{F_BB}+X_B(t)}, \tag{5.40}$$

$$\dot{X}_C(t) = X_A(t)k_{AC}\frac{(1-X_C(t))}{K_{AC}+(1-X_C(t))} - X_B(t)k_{BC}\frac{X_C(t)}{K_{BC}+X_C(t)}, \tag{5.41}$$

where $I(t)$ represents an external input into the system. This network is capable of perfect adaptation if the first term of $\dot{X}_B(t)$ is in the saturated region $(1-X_B(t)) \gg K_{AB}$ and the second term is in the linear region $X_B(t) \ll K_{F_BB}$, which allows us to make the following approximation:

$$\frac{dX_B(t)}{dt} \approx X_A(t)k_{AB} - \frac{F_B k_{F_BB}}{K_{F_BB}}X_B(t). \tag{5.42}$$

Therefore, a steady state solution X_B for B is proportional to the steady state solution X_A for A. Since both terms in equation (5.41) are proportional to X_A we find that the steady state solution X_C for C is a function of only the parameters k_{AC}, K_{AC}, k_{BC}, K_{BC}, k_{AB}, F_B, k_{F_BB}, and K_{F_BB} (note that X_A factors out of the equilibrium equation corresponding to (5.41)), and hence it does not depend on the input parameter I. Since a change in the input signal I changes $\dot{X}_A(t)$ there is a transient effect on $X_A(t)$. Similarly there must also be a transient effect on both $X_B(t)$ and $X_C(t)$. It follows that the system achieves perfect adaptation.

The equilibrium equations associated with equations (5.39), the approximation (5.42) to (5.40), and (5.41) are given by:

$$f_A: \quad Ik_{IA}\frac{(1-X_A)}{K_{IA}+(1-X_A)} - F_A k_{F_AA}\frac{X_A}{K_{F_AA}+X_A} = 0, \tag{5.43}$$

$$f_B: \quad X_A k_{AB} - \frac{F_B k_{F_BB}}{K_{F_BB}} X_B = 0, \tag{5.44}$$

$$f_C: \quad X_A k_{AC}\frac{(1-X_C)}{K_{AC}+(1-X_C)} - X_B k_{BC}\frac{X_C}{K_{BC}+X_C} = 0. \tag{5.45}$$

The associated equilibrium causal ordering graph in Figure shows that there is a directed path from the input signal I to the cluster $\{v_A, v_B, v_C\}$. Therefore, the conditions of Theorem 5.1 are not satisfied for the system with input signal I.

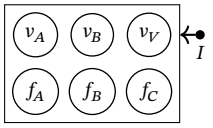


Figure 5.19: The equilibrium causal ordering graph for the IFFLP network modelled by equations (5.43), (5.44), and (5.45).

Table 5.1: The conditional independences in the simulation of the protein pathway described in Section 5.5.1 were assessed using Spearman's rank correlations. With a p-value threshold of 0.01, d-separations with a separating set of size 0 or 1 coincide with conditional independences with conditioning sets of size 0 or 1.

Independence test	Correlation	p-value	d-separation
$I \perp\!\!\!\perp X_s$	0.029	0.51	yes
$I \perp\!\!\!\perp X_r$	0.020	0.66	yes
$I \perp\!\!\!\perp X_m$	0.021	0.64	yes
$I \perp\!\!\!\perp X_e$	0.777	$< 2.2e^{-16}$	no
$X_s \perp\!\!\!\perp X_r$	0.957	$< 2.2e^{-16}$	no
$X_s \perp\!\!\!\perp X_m$	0.933	$< 2.2e^{-16}$	no
$X_s \perp\!\!\!\perp X_e$	−0.561	$< 2.2e^{-16}$	no
$X_r \perp\!\!\!\perp X_m$	0.977	$< 2.2e^{-16}$	no
$X_r \perp\!\!\!\perp X_e$	−0.542	$< 2.2e^{-16}$	no
$X_m \perp\!\!\!\perp X_e$	−0.524	$< 2.2e^{-16}$	no
$I \perp\!\!\!\perp X_s \mid X_r$	0.037	0.83	yes
$I \perp\!\!\!\perp X_s \mid X_m$	0.027	0.61	yes
$I \perp\!\!\!\perp X_r \mid X_s$	−0.030	0.51	yes
$I \perp\!\!\!\perp X_r \mid X_m$	−0.005	0.91	yes
$I \perp\!\!\!\perp X_m \mid X_s$	−0.018	0.69	yes
$I \perp\!\!\!\perp X_m \mid X_r$	0.010	0.83	yes
$X_e \perp\!\!\!\perp X_r \mid X_s$	−0.019	0.67	yes
$X_e \perp\!\!\!\perp X_m \mid X_s$	$-2.1 \cdot 10^{-4}$	0.99	yes
$X_e \perp\!\!\!\perp X_m \mid X_r$	0.031	0.49	yes
$X_s \perp\!\!\!\perp X_m \mid X_r$	−0.031	0.48	yes
$I \perp\!\!\!\perp X_e \mid X_s$	0.959	$6.0 \cdot 10^{-275}$	no
$I \perp\!\!\!\perp X_e \mid X_r$	0.937	$1.6 \cdot 10^{-229}$	no
$I \perp\!\!\!\perp X_e \mid X_m$	0.925	$1.2 \cdot 10^{-211}$	no
$I \perp\!\!\!\perp X_s \mid X_e$	0.894	$4.7 \cdot 10^{-175}$	no
$I \perp\!\!\!\perp X_r \mid X_e$	0.832	$1.5 \cdot 10^{-129}$	no
$I \perp\!\!\!\perp X_m \mid X_e$	0.799	$1.1 \cdot 10^{-111}$	no
$X_e \perp\!\!\!\perp X_s \mid X_r$	−0.176	$8.0 \cdot 10^{-5}$	no
$X_e \perp\!\!\!\perp X_s \mid X_m$	−0.236	$9.3 \cdot 10^{-8}$	no
$X_e \perp\!\!\!\perp X_s \mid I$	−0.928	$8.7 \cdot 10^{-216}$	no
$X_e \perp\!\!\!\perp X_r \mid X_m$	−0.164	$2.2 \cdot 10^{-4}$	no
$X_e \perp\!\!\!\perp X_r \mid I$	−0.885	$5.3 \cdot 10^{-167}$	no
$X_e \perp\!\!\!\perp X_m \mid I$	−0.859	$2.8 \cdot 10^{-146}$	no
$X_s \perp\!\!\!\perp X_r \mid I$	0.957	$1.7 \cdot 10^{-269}$	no
$X_s \perp\!\!\!\perp X_r \mid X_e$	0.939	$5.3 \cdot 10^{-232}$	no
$X_s \perp\!\!\!\perp X_r \mid X_m$	0.590	$3.2 \cdot 10^{-48}$	no
$X_s \perp\!\!\!\perp X_m \mid I$	0.933	$3.0 \cdot 10^{-223}$	no
$X_s \perp\!\!\!\perp X_m \mid X_e$	−0.907	$1.2 \cdot 10^{-188}$	no
$X_r \perp\!\!\!\perp X_m \mid I$	−0.977	0	no
$X_r \perp\!\!\!\perp X_m \mid X_e$	0.968	$2.8 \cdot 10^{-302}$	no
$X_r \perp\!\!\!\perp X_m \mid X_s$	0.807	$1.1 \cdot 10^{-115}$	no

Chapter 6

Robustness of model predictions under extension

Based on:
Robustness of model predictions under extension,
T. Blom, and J.M. Mooij,
To appear in: Proceedings of Causal Discovery & Causality-Inspired Machine Learning Workshop at Neural Information Processing Systems, 2020.
Arxiv preprint: 2012.04723v1

Often, mathematical models of the real world are simplified representations of complex systems. A caveat to using models for analysis is that predicted causal effects and conditional independences may not be robust under model extensions, and therefore applicability of such models is limited. In this work, we consider conditions under which qualitative model predictions are preserved when two models are combined. We show how to use the technique of *causal ordering* to efficiently assess the robustness of qualitative model predictions and characterize a large class of model extensions that preserve these predictions. For dynamical systems at equilibrium, we demonstrate how novel insights help to select appropriate model extensions and to reason about the presence of feedback loops. We apply our ideas to a viral infection model with immune responses.

6.1 Introduction

Key aspects of the scientific method include generating a model or hypothesis that explains a phenomenon, deriving testable predictions from this model or hypothesis, and designing an experiment to test these predictions in the real world. There are quite some interesting statistical systems for which simple Structural Causal Models

(Bongers, Forré, et al., 2020; Pearl, 2009) do not model all causal and Markov properties of the system, as we demonstrated in Chapter 3, Chapter 4, and Chapter 5. In those cases the causal ordering algorithm, first introduced by Simon (1953), can be used to better understand these properties. In this paper we consider what happens when two systems are combined and we give conditions under which the properties of the whole system can be understood in terms of properties of its parts. We discuss how a holistic approach towards causal modelling may result in novel insights when we derive and test the predictions of systems for which new properties emerge from the combination of its parts.

We consider the practical issue of assessing whether qualitative model predictions are robust under model extensions. We revisit the observations of De Boer (2012) concerning a viral infection model and demonstrate that the qualitative causal predictions of this model can change dramatically when the model is extended with extra equations describing simple immune responses. To assess the robustness of predicted causal relations or conditional independences, it would be useful to gain a better understanding of the class of model extensions that lead to changes in these predictions. We propose the technique of causal ordering (Simon, 1953) as an efficient method to assess the robustness of qualitative causal predictions. This allows us to characterize a large class of model extensions under which these predictions are preserved. We also consider the class of models that are obtained from the equilibrium equations of dynamical models where each variable is *self-regulating*. For this class, we show that the predicted presence of causal relations and absence of conditional independences is robust when the model is extended with new equations.

The promise of causal discovery algorithms is that they are able to learn causal relations from a combination of background knowledge and data. The general idea of many constraint-based approaches (e.g. PC or FCI and variants thereof (Colombo et al., 2012; Mooij and Claassen, 2020; Spirtes, Glymour, et al., 2000; J. Zhang, 2008)) is to exploit information about conditional independences in a probability distribution to construct an equivalence class of graphs that encode certain aspects of the probability distribution, and then draw conclusions about the causal relations from the graphs. There is a large amount of literature concerning particular algorithms for which the learned structure expresses causal relations under certain conditions (e.g. linearity, causal sufficiency, absence of feedback loops), see for example (Colombo et al., 2012; Forré et al., 2018; Hyttinen et al., 2012; Lacerda et al., 2008; Mooij and Claassen, 2020; Mooij, Magliacane, et al., 2020; Richardson and Spirtes, 1999; Spirtes, Glymour, et al., 2000; Strobl, 2018; J. Zhang, 2008). In this chapter, our main interest is in dynamical models with the property that graphs of variables that encode the conditional independences of their equilibrium distribution should not be interpreted causally at all. Given a model for a subsystem, we present novel insights that enable us to reject model extensions based on conditional independences in equilibrium data of the subsystem. We demonstrate how, for the equilibrium distribution of certain dynamical models, this approach allows us to reason about the presence of variables

that are not self-regulating and feedback mechanisms that involve unobserved variables. We hope that, in future work, existing algorithms that are designed for causal discovery could be useful for reasoning about appropriate model extensions from a combination of partial models and observational data of a subsystem.

6.1.1 Causal ordering graph and the effects of interventions

To keep the contents of this chapter self-contained, we give a concise introduction to the technique of causal ordering, introduced by Simon (1952), and presented in Chapter 4.[1] In short, the causal ordering algorithm takes a set of equations as input and returns a *causal ordering graph* that encodes the effects of interventions and a *Markov ordering graph* that implies conditional independences between variables in the model, see Theorem 4.3. Compared with the popular framework of Structural Causal Models (Pearl, 2009), the distinction between the causal ordering and Markov ordering graphs does not provide new insights for acyclic models but it results in non-trivial conclusions for models with feedback, as suggested in the discussion in Section 6.2.4 and Chapter 4.

We consider models consisting of equations F that contain endogenous variables V, independent exogenous random variables W, and constant parameters P. The structure of equations and the endogenous variables that appear in them can be represented by the *associated bipartite graph* $\mathcal{B} = \langle V, F, E\rangle$, where each endogenous variable is associated with a distinct vertex in V, and each equation is associated with a distinct vertex in F. There is an edge $(v - f) \in E$ if and only if variable $v \in V$ appears in equation $f \in F$. The causal ordering algorithm constructs a *directed cluster graph* $\langle \mathcal{V}, \mathcal{E}\rangle$, where $\mathcal{V}$ is a partition of vertices V into clusters and $\mathcal{E}$ is a set of directed edges from vertices in V to clusters in $\mathcal{V}$. Given a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ with a perfect matching M, the causal ordering algorithm proceeds with the following three steps:[2]

(i) For $v \in V$, $f \in F$ orient edges $(v - f)$ as $(v \leftarrow f)$ when $(v - f) \in M$ and as $(v \rightarrow f)$ otherwise; this yields a directed graph $\mathcal{G}(\mathcal{B}, M)$.

(ii) Find all strongly connected components $S_1, S_2, \dots, S_n$ of $\mathcal{G}(\mathcal{B}, M)$. Let $\mathcal{V}$ be the set of clusters $S_i \cup M(S_i)$ for $i \in \{1, \dots, n\}$, where $M(S_i)$ denotes the set of vertices that are matched to vertices in S_i in matching M.

(iii) Let $\mathrm{cl}(f)$ denote the cluster in $\mathcal{V}$ containing f. For each $(v \rightarrow f)$ such that $v \notin \mathrm{cl}(f)$ add an edge $(v \rightarrow \mathrm{cl}(f))$ to $\mathcal{E}$.

Optionally, independent exogenous random variables and parameters can be added as singleton clusters with edges towards the clusters of the equations in which they appear. It was shown that the resulting directed cluster graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E}\rangle$,

[1] Actually, we consider an equivalent algorithm for causal ordering that was shown to be more computationally efficient by (Gonçalves et al., 2016; Nayak, 1995), see Chapter 4 for more details.

[2] A perfect matching M is a subset of edges in a bipartite graph so that every vertex is adjacent to exactly one edge in M. Note that not every bipartite graph has a perfect matching.

which we refer to as the *causal ordering graph*, is independent of the choice of perfect matching, see Theorem 4.1. Example 6.1 shows how the algorithm works and a graphical illustration of the algorithm for a more elaborate cyclic model can be found in the supplement.

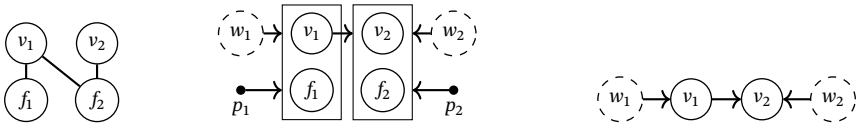
Example 6.1. Let $V = \{v_1, v_2\}$, $W = \{w_1, w_2\}$, and $P = \{p_1, p_2\}$ be index sets. Consider model equations f_1 and f_2 with endogenous variables $(X_v)_{v \in V}$, exogenous random variables $(U_w)_{w \in W}$ and constant parameters C_p with $p \in P$ below.

$$f_1 : \quad C_{p_1} X_{v_1} - U_{w_1} = 0, \tag{6.1}$$

$$f_2 : \quad C_{p_2} X_{v_2} + X_{v_1} + U_{w_2} = 0. \tag{6.2}$$

The bipartite graph $\mathcal{B} = \langle V, F, E \rangle$ in Figure 6.1(a), with $E = \{(v_1 - f_1), (v_1 - f_2), (v_2 - f_2)\}$ is a compact representation of the model structure. This graph has a perfect matching $M = \{(v_1 - f_1), (v_2 - f_2)\}$. By orienting edges in $\mathcal{B}$ according to the rules in step (i) of the causal ordering algorithm we obtain the directed graph $\langle V \cup F, E_{\text{dir}} \rangle$ with $E_{\text{dir}} = \{(f_1 \to v_1), (f_2 \to v_2), (v_1 \to f_2)\}$. The clusters $C_1 = \{v_1, f_1\}$ and $C_2 = \{v_2, f_2\}$ are added to $\mathcal{V}$ in step (ii) of the algorithm, and the edge $(v_1 \to C_2)$ is added to $\mathcal{E}$ in step (iii). Finally, we may add the parameters P and independent exogenous random variables W as singleton clusters to $\mathcal{V}$, and the edges $(p_1 \to C_1)$, $(w_1 \to C_1)$, $(p_2 \to C_2)$, and $(w_2 \to C_2)$ to $\mathcal{E}$. The resulting causal ordering graph is given in Figure 6.1(b). △

Throughout this chapter, we will assume that models are uniquely solvable with respect to the causal ordering graph, which roughly means that for each cluster, the equations in that cluster can be solved uniquely for the endogenous variables in that cluster (see Definition 4.8 for details). A perfect intervention on a cluster that contains equation vertices represents a model change where the equations in the targeted cluster are replaced by equations that set the endogenous variables in that cluster equal to constant values. A soft intervention targets an equation, parameter, or exogenous variable, but does not affect which variables appear in the equations. We say that there is a directed path from a vertex x to a vertex y in a causal ordering graph $\langle \mathcal{V}, \mathcal{E} \rangle$ if either $\text{cl}(x) = \text{cl}(y)$ or there is a sequence of clusters $V_1 = \text{cl}(x), V_2, \ldots, V_{k-1}, V_k = \text{cl}(y)$ so that for all $i \in \{1, \ldots, k-1\}$ there is a vertex $z_i \in V_i$ such that $(z_i \to V_{i+1}) \in \mathcal{E}$. It can be shown that a) the presence of a directed path from a cluster, equation, parameter, or exogenous variable that is targeted by a soft intervention towards a certain variable in the causal ordering graph implies that the intervention has a generic effect on that variable and b) if no such path exists there is no causal effect of the intervention on that variable, see Theorem 4.5 for more details.



(a) Bipartite graph. (b) Causal ordering graph. (c) Markov ordering graph.

Figure 6.1: The bipartite graph in Figure (a) is a compact representation of the model in Example 6.1. The corresponding causal ordering graph and Markov ordering graph are given in Figures (b) and (c) respectively. Exogenous variables are denoted by dashed circles and parameters by black dots.

6.1.2 Markov ordering graph and causal discovery

The causal ordering graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ of model equations F with endogenous variables V, exogenous random variables W, constant parameters P, and bipartite graph $\mathcal{B}$ can be used to construct the *Markov ordering graph*, which is a DAG $\mathrm{MO}(\mathcal{B}) = \langle V \cup W, E \rangle$, with $(x \to y) \in E$ if and only if $(x \to \mathrm{cl}(y)) \in \mathcal{E}$. The Markov ordering graph for the model equations in Example 6.1 is given in Figure 6.1(c). It has been shown that, under the assumption of unique solvability w.r.t. the causal ordering graph, d-separations in the Markov ordering graph imply conditional independences between the corresponding variables (Blom, Van Diepen, et al., 2021). Henceforth, we will assume that the probability distribution of the solution $(X_v)_{v \in V}$ to a set of model equations is faithful to the Markov ordering graph. In other words, each conditional independence in the distribution implies a d-separation in the Markov ordering graph. Under the assumption that data is generated from such a model, some causal discovery algorithms, such as the PC algorithm (Spirtes, Glymour, et al., 2000), aim to construct the Markov equivalence class of the Markov ordering graph. In this work, we will specifically focus on feedback models for which the Markov ordering graph of the equilibrium distribution, and consequently the output of many causal discovery algorithms, does not have a straightforward causal interpretation.

6.2 Causal ordering for a viral infection model

This work was inspired by a viral infection model in De Boer (2012), who showed through explicit calculations that the predictions of the model are not robust under addition of an immune response. This sheds doubt on the correct interpretation of variables and parameters in the model. For many systems it is intrinsically difficult to study their behaviour in detail. The use of simplified mathematical models that capture key characteristics aids in the analysis of a certain properties of the system. The hope is that the explanations inferred from model equations are legitimate accounts of the true underlying system (De Boer, 2012). In reality, a modeller must take into account that the outcome of these studies may be contingent on the specif-

ics of the model design. Here, we demonstrate how causal ordering can be used as a scalable tool to assess the robustness of model predictions without requiring explicit calculations.

6.2.1 Viral infection without immune response

Let U_σ be a production term for target cells, d_T the death rate for target cells, U_f the fraction of successful infections, and U_δ the death rate of productively infected cells. Define $\beta = \frac{bp}{c}$, where b is the infection rate, p the amount of virus produced per infected cell, and c the clearance rate of viral particles. The following first-order differential equations describe how the amount of target cells $X_T(t)$ and the amount of infected cells $X_I(t)$ evolve over time (De Boer, 2012):

$$\dot{X}_T(t) = U_\sigma - d_T X_T(t) - \beta X_T(t) X_I(t), \tag{6.3}$$

$$\dot{X}_I(t) = (U_f \beta X_T(t) - U_\delta) X_I(t), \tag{6.4}$$

Suppose that we want to use this simple viral infection model to explain why the *set-point viral load* (i.e. the total amount of virus circulating in the bloodstream) of chronically infected HIV-patients differs by several orders of magnitude, as De Boer (2012) does. To analyse this problem we look at the equilibrium equations that are implied by equations (6.3) and (6.4):[3]

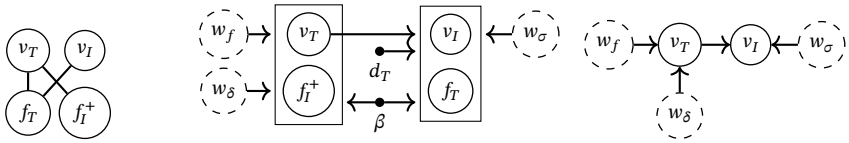
$$f_T : \quad U_\sigma - d_T X_T - \beta X_T X_I = 0, \tag{6.5}$$

$$f_I^+ : \quad U_f \beta X_T - U_\delta = 0. \tag{6.6}$$

Throughout the remainder of this work we will use this *natural labelling* of equilibrium equations, where the equation derived from the derivative $\dot{X}_i(t)$ is labelled f_i. For first-order differential equations that are written in canonical form, $\dot{X}_i(t) = g_i(X(t))$, the natural labelling always exists.

Suppose that U_σ, U_f and U_δ are independent exogenous random variables taking values in $\mathbb{R}_{>0}$ and d_T, β are strictly positive parameters. The associated bipartite graph, causal ordering graph, and Markov ordering graph are given in Figure 6.2. The causal ordering graph tells us that soft interventions targeting U_σ, U_f, U_δ, d_T, or β generically have an effect on the equilibrium distribution of the amount of infected cells X_I. From here on, we say that the causal ordering graph of a model predicts the *generic presence* or *absence* of causal effects. The Markov ordering graph shows that v_T and w_σ are d-separated. This implies that the amount of target cells X_T should be independent of the production rate U_σ when the system is at equilibrium. Henceforth, we will say that the Markov ordering graph predicts the *generic presence* or *absence* of conditional dependences.

[3] Since we are only interested in strictly positive solutions we removed X_I from the equilibrium equation $f_I : \ (U_f \beta X_T - U_\delta) X_I = 0$ to obtain f_I^+.



(a) Bipartite graph. (b) Causal ordering graph. (c) Markov ordering graph.

Figure 6.2: Graphical representations of the viral infection model in equations (6.5) and (6.6). Vertices v_i and w_j correspond to variables X_i and U_j, respectively. The causal ordering graph represents generic effects of interventions. The d-separations in Figure 6.2(c) imply conditional independences.

6.2.2 Viral infection with a single immune response

The viral infection model in equations (6.3) and (6.4) can be extended with a simple immune response $X_E(t)$ by adding the following dynamic and static equations:

$$\dot{X}_E(t) = (U_a X_I(t) - d_E)X_E(t), \tag{6.7}$$

$$X_\delta(t) = d_I + U_k X_E(t), \tag{6.8}$$

where U_a is an activation rate, d_E and d_I are turnover rates and U_k is a mass-action killing rate (De Boer, 2012). Note that the exogenous random variable U_δ is now treated as an endogenous variable $X_\delta(t)$ instead. We derive the following equilibrium equations using the natural labelling provided by equations (6.7) and (6.8):[4]

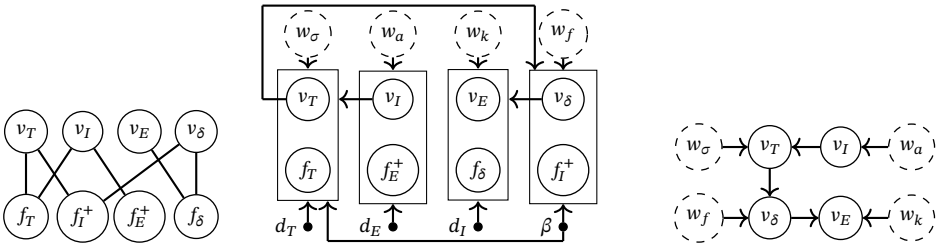
$$f_E^+ : \quad U_a X_I - d_E = 0, \tag{6.9}$$

$$f_\delta : \quad X_\delta - d_I - U_k X_E = 0, \tag{6.10}$$

Henceforth, we will call the addition of equations F_+ to F a *model extension*. Notice that, when two sets of equations are combined, there may exist variables that were exogenous in the submodel (i.e. the original model) but that are endogenous within the whole model (i.e. the extended model). Generally, equations F_+ may contain endogenous variables in V and exogenous variables in W but they may also contain additional endogenous variables V_+ and additional exogenous variables W_+. Parameters and exogenous variables that appear in equations F can appear as endogenous variables in V_+ and in the extended model $F_{\text{ext}} = F \cup F_+$. In that case, these variables are no longer considered to be parameters of exogenous variables within the extended model.

Suppose that U_a and U_k are independent exogenous random variables taking values in $\mathbb{R}_{>0}$ and d_E, d_I are parameters taking value in $\mathbb{R}_{>0}$. The bipartite graph, causal ordering graph, and Markov ordering graph associated with equations (6.5), (6.6), (6.9), and (6.10) (with X_δ replacing U_δ) are given in Figure 6.3. The causal ordering

[4] Analogous to changing f_I to f_I^+ for strictly positive solutions, we will look at f_E^+ instead of f_E.



(a) Bipartite graph. (b) Causal ordering graph. (c) Markov ordering graph.

Figure 6.3: Graphical representations of the viral infection model with a single immune response. The presence or absence of causal relations and d-connections implied by the graphs in Figure 6.2 are not preserved if a single immune response is added.

graph predicts a causal effect of U_σ and d_T on X_T but not on X_I. By comparing with the predictions of the causal ordering graph in Figure 6.2(b), we find that effects of interventions targeting U_σ and d_T are not robust under the model extension. The Markov ordering graph of the extended model shows that w_σ is d-connected to v_T, and hence U_σ and X_T are dependent. We conclude that the independence between U_σ and X_T that was implied by the Markov ordering graph of the viral infection model without immune response is not robust under the model extension.

The systematic graphical procedure followed here easily leads to the same causal conclusions as De Boer (2012) obtained by explicitly solving the equilibrium equations. In addition, it leads to predictions regarding the conditional (in)dependences in the equilibrium distribution.

6.2.3 Viral infection with multiple immune responses

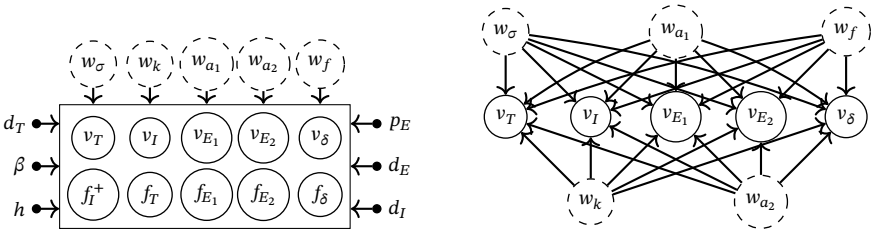
The following static and dynamical equations describe multiple immune responses:

$$\dot{X}_{E_i}(t) = \frac{p_E X_{E_i}(t) U_{a_i} X_I(t)}{h + X_{E_i}(t) + U_{a_i} X_I(t)} - d_E X_{E_i}(t), \qquad i = 1, 2, \ldots, n \tag{6.11}$$

$$X_\delta(t) = d_I + U_k \sum_i^n U_{a_i} X_{E_i}(t), \tag{6.12}$$

where there are n immune responses, U_{a_i} is the avidity of immune response i, p_E is the maximum division rate, and h is a saturation constant (De Boer, 2012). For $n = 2$ we can derive equilibrium equations f_{E_1}, f_{E_2}, and f_δ using the natural labelling as we did for the equilibrium equations in the previous section. Together with the equilibrium equations (6.5) and (6.6) (with X_δ replacing U_δ) for the viral infection model this is another extended model. The bipartite graph of this extended model is given in Figure 6.5(a), while the causal ordering graph can be found in Figure 6.4(a). By

comparing the directed paths in this causal ordering graph with that of the original viral infection model (i.e. the model without an immune response) in Figure 6.2(b), it can be seen that the predicted presence of causal relations is preserved under extension of the model with multiple immune responses, while the predicted absence of causal relations is not. Similarly, by comparing d-separations in the Markov ordering graphs in Figure 6.2(c) with those in Figure 6.4(b), we find that predicted conditional dependences are preserved under the extensions, while the predicted conditional independences are not.



(a) Causal ordering graph.

(b) Markov ordering graph.

Figure 6.4: Graphical representations of the viral infection model with multiple immune responses. The presence of causal relations and d-connections in Figure 6.2 is preserved.

6.2.4 Markov ordering graphs and causal interpretations

Here, we will demonstrate that the Markov ordering graphs for the equilibrium equations of the viral infection models neither have a straightforward causal interpretation in terms of soft interventions targeting parameters, exogenous variables, or equations *nor* in terms of perfect interventions on variables in the dynamical model. To see this, consider the Markov ordering graph in Figure 6.3(c) for the viral infection with a single immune response. The edge $(v_I \to v_T)$ cannot correspond to the effect of a soft intervention targeting f_I^+, because the causal ordering graph in Figure 6.3(b) shows that there is no such effect. Clearly, directed paths in the Markov ordering graph do not necessarily represent the effects of soft interventions. The natural way to model a perfect intervention targeting a variable in the Markov ordering graph is to replace the (differential) equation of that variable with an equation setting that variable equal to a certain value in the underlying dynamical model (Mooij, Janzing, and Schölkopf, 2013). By explicitly solving equilibrium equations it is easy to check that replacing f_δ with an equation setting X_δ equal to a constant generically changes the distribution of X_I. Since there is no directed path from v_δ to v_I in the Markov ordering graph, the effect of this perfect intervention would not have been predicted by the Markov ordering graph, if it would have been interpreted causally. Therefore, contrary to the causal ordering graph, the Markov ordering graph does not have a causal interpretation in terms of soft or perfect interventions on the true underlying

dynamical model.

6.3 Robust causal predictions under model extensions

One way to gauge the robustness of model predictions is to check to what extent they depend on the model design. The example of a viral infection with different immune responses in the previous section indicates that qualitative causal predictions entailed by the causal ordering graph of a mathematical model may strongly depend on the particulars of the model. Both the implied presence or absence of causal relations at equilibrium and the implied presence or absence of conditional independences at equilibrium may change under certain model extensions. Under what conditions are these qualitative model predictions preserved under extensions? In this section, we characterize a large class of model extensions under which qualitative equilibrium predictions are preserved.

Theorem 6.1 gives a sufficient condition on model extensions under which the predicted generic presence of causal relations and predicted generic presence of conditional dependences at equilibrium is preserved. The proof is given in the supplement.

Theorem 6.1. *Consider model equations F containing endogenous variables V with bipartite graph $\mathcal{B}$. Suppose F is extended with equations F_+ containing endogenous variables in $V \cup V_+$, where V_+ contains endogenous variables that are added by the model extension.*[5] *Let $\mathcal{B}_{\text{ext}}$ be the bipartite graph associated with $F_{\text{ext}} = F \cup F_+$ and $V_{\text{ext}} = V \cup V_+$, and $\mathcal{B}_+$ the bipartite graph associated with the extension F_+ and V_+, where variables in V appearing in F^+ are treated as exogenous variables (i.e. they are not added as vertices in $\mathcal{B}_+$). If $\mathcal{B}$ and $\mathcal{B}_+$ both have a perfect matching then:*

(i) $\mathcal{B}_{\text{ext}}$ has a perfect matching,
(ii) ancestral relations in $\text{CO}(\mathcal{B})$ are also present in $\text{CO}(\mathcal{B}_{\text{ext}})$,
(iii) d-connections in $\text{MO}(\mathcal{B})$ are also present in $\text{MO}(\mathcal{B}_{\text{ext}})$.

This result characterizes a large set of extensions under which the implied causal effects and conditional dependences of a model are preserved. Consider again the equilibrium behaviour of the viral infection models in Section 6.2. We already showed explicitly that the extension of the viral infection model with multiple immune responses preserved the predicted presence of causal relations and conditional dependences, but with the help of Theorem 6.1 we only would have needed to check whether the bipartite graph in Figure 6.5(c) has a perfect matching to arrive at the same conclusion. The bipartite graph for the extension with a single immune response in Figure 6.5(b) does not have a perfect matching and hence the conditions of Theorem 6.1 do not hold. Recall that this model extension did not preserve the predicted presence of causal relations.

[5] V_+ may also contain parameters or exogenous variables that appear in F and become endogenous in the extended model.

The theorem below gives a stronger condition under which (conditional) independence relations and the absence of causal relations that are implied by a model are also predicted by the extended model. The proof is provided in the supplement.

Theorem 6.2. *Let F, F_+, F_{ext}, V, V_+, V_{ext}, $\mathcal{B}$, $\mathcal{B}_+$, and $\mathcal{B}_{\text{ext}}$ be as in Theorem 6.1. If $\mathcal{B}$ and $\mathcal{B}_+$ both have perfect matchings and no vertex in V_+ is adjacent to a vertex in F in $\mathcal{B}_{\text{ext}}$ then:*[6]

(i) ancestral relations absent in $\text{CO}(\mathcal{B})$ *are also absent in* $\text{CO}(\mathcal{B}_{\text{ext}})$*,*

(ii) d-connections absent in $\text{MO}(\mathcal{B})$ *are also absent in* $\text{MO}(\mathcal{B}_{\text{ext}})$*.*

This result characterizes a large class of model extensions under which all qualitative model predictions are preserved. Consider again the equilibrium models for the viral infection in Section 6.2. The bipartite graph for the extension with a single immune response, which we obtain by adding equations (6.9) and (6.10), does not have a perfect matching. In the bipartite graph associated with the viral infection model with multiple immune responses the additional endogenous variable v_δ is adjacent to f_I. Neither of the model extensions satisfies the conditions of Theorem 6.2. We already demonstrated that neither of the model extensions preserves all qualitative model predictions. An example of a model extension that does satisfy the conditions in Theorem 6.1 and 6.2 is an acyclic structural causal model that is extended with another acyclic structural causal model such that the additional variables are non-ancestors of the original ones. Together, Theorem 6.1 and 6.2, can be used to understand when the properties of a system can be understood by studying the properties of its parts.

6.4 Selection of model extensions

So far, we have considered methods to assess the robustness of qualitative model predictions. In this section we will show how this idea results in novel opportunities regarding causal discovery. In particular, if we assume that the systems that we observe are part of a larger partially observed system, then we can use the methods in this paper to reason about causal mechanisms of unobserved variables. Consider, for example, the viral infection model for which we have demonstrated that extensions with different immune responses imply different (conditional) independences between variables in the original model. The Markov ordering graphs in Figures 6.2(c), 6.3(c), and 6.4(b) imply the following (in)dependences:

(i) Viral infection without immune response: $U_\sigma \perp\!\!\!\perp X_T$, $U_\sigma \not\perp\!\!\!\perp X_I$.

(ii) Viral infection with single immune response: $U_\sigma \not\perp\!\!\!\perp X_T$, $U_\sigma \perp\!\!\!\perp X_I$.

(iii) Viral infection with multiple immune responses: $U_\sigma \not\perp\!\!\!\perp X_T$, $U_\sigma \not\perp\!\!\!\perp X_I$.

[6] A vertex in V_+ is considered adjacent to F if it corresponds with one of the exogenous random variables or parameters in F that become endogenous in the model extension.

Given a model for variables X_T and X_I only, we can reject model extensions based on the (conditional) independences for variables X_T, X_I, and U_σ. Using this holistic modelling approach, we can reason about an unknown model extension without observing the new mechanisms or variables. In the remainder of this section, we further discuss how this idea can be applied to equilibrium data of dynamical systems.

6.4.1 Reasoning about self-regulating variables

We say that a variable in a set of first-order differential equations in canonical form is *self-regulating* if it can be solved uniquely from the equilibrium equation that is constructed from its derivative. For models in which every variable is self-regulating there exists a perfect matching where each variable v_i is matched to its associated equilibrium equation f_i according to the natural labelling, for more details see Lemma 6.1 in the supplement.[7] It then follows from Theorem 6.1 that the presence of ancestral relations and d-connections is robust under dynamical model extensions in which each variable is self-regulating, as is stated more formally in Corollary 6.1 below.

Corollary 6.1. *Consider a first-order dynamical model in canonical form for endogenous variables V and an extension consisting of canonical first-order differential equations for additional endogenous variables V_+. Let F and $F_{\text{ext}} = F \cup F_+$ be the equilibrium equations of the original and extended model respectively. If all variables in $V \cup V_+$ are self-regulating then (ii) and (iii) of Theorem 6.1 hold.*

Corollary 6.1 characterizes a class of models under which qualitative predictions for the equilibrium distribution are robust, but the result can also be interpreted from a different angle. Suppose that we have equilibrium data that is generated by an extended dynamical model with equilibrium equations F_{ext}, but we only have a *partial* model consisting of equations in F for a subset $V \subseteq V_{\text{ext}} = V \cup V_+$ of variables that appear in $F_{\text{ext}} = F \cup F_+$. If we would find conditional independences between variables in V that do not correspond to d-separations in the Markov ordering graph of the partial model, this does not necessarily mean that the model equations are wrong. It could also be the case, for example, that we are wrong to assume that the system can be studied in a reductionist manner and that the model should be extended. Furthermore, under the assumption that data is generated from the equilibrium distribution

[7] Interestingly, the Markov ordering graph for the equilibrium equations of such a model always has a causal interpretation. By construction of the causal ordering graph from the bipartite graph and the perfect matching provided by the natural labelling, we know that a vertex v_i always appears in a cluster with f_i in the causal ordering graph. The presence or absence of directed paths in the Markov ordering graph can then easily be associated with the presence or absence of directed paths in the causal ordering graph. Consequently, the Markov ordering graph can be interpreted in terms of both soft interventions targeting equations and perfect interventions that set variables equal to a constant by replacement of the associated dynamical and equilibrium equations. Note that dynamical systems with only self-regulating variables were also considered by Mooij, Janzing, and Schölkopf (2013), where it was shown that their equilibria can be modelled as Structural Causal Models without self-loops.

of a dynamical model, Corollary 6.1 tells us that conditional independences in the data that are not predicted by the equations of a partial model imply the presence of variables that are not self-regulating, if we assume faithfulness. This shows that, given a model for a subsystem, we can reason about the properties of unobserved and unknown variables in the whole system. Consider, for example, the model of the viral infection without immune response and assume that this is a submodel of a larger system. Suppose that we observe a conditional independence between U_σ and X_I and assume that the model equations of the submodel are correct. Since the Markov ordering graph in Figure 6.2(c) implies that U_σ and X_I are dependent, Corollary 6.1 tells us that there must be variables that are not self-regulating in the extended system. If the extended system can be described by the strictly positive solutions of the viral infection model with a single immune response, so that U_σ and X_I are independent, then we see from equations (6.5), (6.6), (6.9), and (6.10) that both $X_E(t)$ and $X_I(t)$ are not self-regulating.

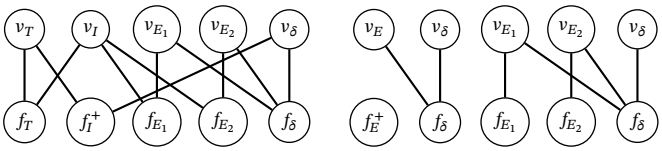
6.4.2 Reasoning about feedback loops

We say that an extension of a dynamical model *introduces a new feedback loop with the original dynamical model* when there is feedback in the extended dynamical model that involves variables in both the original model and the model extension. To make this definition more precise, consider the set E_{nat} of edges $(v_i - f_i)$ that are associated with the natural labelling of the equilibrium equations of the extended dynamical model. The feedback loops in the dynamical model coincide with cycles in the directed graph $\mathcal{G}(\mathcal{B}_{\text{nat}}, M_{\text{nat}})$ that is obtained by applying step (i) of the causal ordering algorithm to the bipartite graph $\mathcal{B}_{\text{nat}} = \langle V_{\text{ext}}, F_{\text{ext}}, E_{\text{ext}} \cup E_{\text{nat}} \rangle$ using the perfect matching $M_{\text{nat}} = E_{\text{nat}}$.[8] The following proposition can be used to reason about the presence of partially unobserved feedback loops given a model and observations for a subsystem.

Proposition 6.1. *Consider a first-order dynamical model in canonical form for endogenous variables V and an extension consisting of canonical first-order differential equations for additional endogenous variables V_+. Let F and $F_{\text{ext}} = F \cup F_+$ be the equilibrium equations of the original and extended model respectively. Let $\mathcal{B} = \langle V, F, E \rangle$ be the bipartite graph associated with F and $\mathcal{B}_{\text{ext}} = \langle V_{\text{ext}}, F_{\text{ext}}, E_{\text{ext}} \rangle$ the bipartite graph associated with F_{ext}. Assume that $\mathcal{B}$ and $\mathcal{B}_{\text{ext}}$ both have perfect matchings. If the model extension does not introduce a new feedback loop with the original dynamical model, then d-connections in* $\text{MO}(\mathcal{B})$ *are also present in* $\text{MO}(\mathcal{B}_{\text{ext}})$.

Proposition 6.1 characterizes a class of model extensions under which qualitative model predictions are robust, but it also shows how we can reason about the existence

[8] Note that a feedback loop in the dynamical model does not imply a feedback loop in the equilibrium equations as well. For example, there is feedback in the dynamical equations (6.3), (6.4), but there is no feedback in the causal ordering graph of the equilibrium equations in Figure 6.2(b) nor in the directed graph that is constructed in step (i) of the causal ordering algorithm.



(a) Extended model. (b) Single. (c) Multiple.

Figure 6.5: The bipartite graphs associated with the viral infection model with multiple immune responses, the single immune response extension, and the multiple immune response extension are given in Figures 6.5(a), 6.5(b), and 6.5(c), respectively.

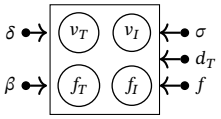


Figure 6.6: Causal ordering graph for positive and non-positive solutions of the viral infection model.

of unobserved feedback loops. To be more precise, it shows that, given a submodel for a subsystem, the presence of conditional independences that are not predicted by the submodel imply the existence of an unobserved feedback loop, if we assume faithfulness. If, for example, we assume that the viral infection model without an immune response is a submodel of the system that is described by the strictly positive equilibrium solutions of the viral infection model with a single immune response, then we would observe an independence between U_σ and X_T that is not predicted by the model equations of the submodel. Proposition 6.1 would then imply that there is an unobserved feedback loop. Indeed, it can be seen from equations (6.3), (6.4), (6.7), (6.8) that there is an unobserved feedback loop from $X_I(t)$ to $X_E(t)$ to $X_\delta(t)$ and back to $X_I(t)$, while the Markov ordering graphs in Figures 6.2(c) and 6.3(c) imply that U_σ and X_I are dependent in the original model and independent in the extended model. We consider the use of existing structure learning algorithms for the detection of feedback loops in models with variables that are not self-regulating from a combination of background knowledge and observational equilibrium data to be an interesting topic for future work.

6.5 Discussion

In this work we revisited several models of viral infections and immune responses. In our treatment of these models we closely followed the approach in De Boer (2012) and therefore we only considered strictly positive solutions. If we would have modelled all solutions then, for example, we would have considered the equilibrium equation $f_I : (U_f \beta X_T - U_\delta) X_I = 0$ instead of f_I^+ in equation (6.6). In that case, we would have obtained the causal ordering graph in Figure 6.6 instead of that in Figure 6.2(b). Clearly, the model predictions of the causal ordering graph for the positive solutions in Figure 6.2(b) are more informative. The choice of only modelling strictly positive solutions depends on the application.

In many application domains mathematical models are used to predict the equi-

librium behaviour of complex systems. An important issue is that (causal) predictions may strongly depend on the specifics of the model design. We revisited an example of a viral infection model (De Boer, 2012), in which implied causal relations and conditional independences change dramatically when equations, describing immune reactions, are added. Analysis of this behaviour through explicit calculations is neither insightful nor scalable. We showed how the technique of causal ordering can be used to efficiently analyse the robustness of implied causal effects and conditional independences. Using key insights provided by this approach we characterized large classes of model extensions under which predicted causal relations and conditional independences are robust. We hope that the results presented in this paper are a step towards bringing the world of causal inference closer to practical applications.

Our results for the characterization of the robustness of model extensions can also be used to reason about the properties of models that are the combination of two submodels. This way, we can study systems whose causal and Markov properties can be understood in a reductionistic manner by considering the properties of its parts. When the properties of the whole model differ from those of its parts, a holistic modelling approach would be required. For models of the equilibrium distribution of dynamical systems, we proved that extensions of dynamical models where each variable is self-regulating preserve the predicted presence of causal effects and d-connections in the original model. Based on those insights, we proposed a novel approach to model selection, where information about conditional independences can be used in combination with model equations to reason about possible model extensions or the presence of feedback mechanisms. For dynamical models with feedback, the output of structure learning algorithms does not always have a causal interpretation in terms of soft or perfect interventions for the equilibrium distribution. We have shown that in dynamical systems where each variable is self-regulating the identifiable directed edges in the learned graph do express causal relations between variables.

6.A Causal ordering algorithm applied to a cyclic model

In this section we demonstrate how the causal ordering algorithm works on a set of equations for a cyclic model. The algorithm is also presented graphically. Consider the following equations for endogenous variables $\mathbf{X}$ and exogenous random variables $\mathbf{U}$:

$$f_1: \quad g_1(X_{v_1}, U_{w_1}) = 0, \tag{6.13}$$
$$f_2: \quad g_2(X_{v_2}, X_{v_1}, X_{v_4}, U_{w_2}) = 0, \tag{6.14}$$
$$f_3: \quad g_3(X_{v_3}, X_{v_2}, U_{w_3}) = 0, \tag{6.15}$$
$$f_4: \quad g_4(X_{v_4}, X_{v_3}, U_{w_4}) = 0, \tag{6.16}$$
$$f_5: \quad g_5(X_{v_5}, X_{v_4}, U_{w_5}) = 0. \tag{6.17}$$

The associated bipartite graph in Figure 6.7(a) consists of variable vertices $V = \{v_1, \dots, v_5\}$ and equation vertices $F = \{f_1, \dots, f_5\}$. There is an edge between a variable vertex and an equation vertex whenever that variable appears in the equation. The associated bipartite graph has exactly two perfect matchings:

$$M_1 = \{(v_1 - f_1), (v_2 - f_2), (v_3 - f_3), (v_4 - f_4), (v_5 - f_5)\},$$
$$M_2 = \{(v_1 - f_1), (v_2 - f_3), (v_3 - f_4), (v_4 - f_2), (v_5 - f_5)\}.$$

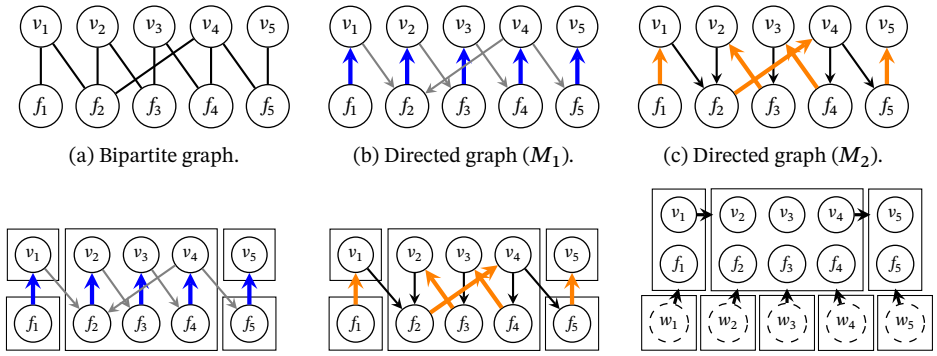
Application of the first step of the causal ordering algorithm results either in the directed graph in Figure 6.7(b) or that in Figure 6.7(c), depending on the choice of the perfect matching. The segmentation of vertices into strongly connected components, which takes place in the second step of the algorithm, results in the clusters $\{v_1\}$, $\{f_1\}$, $\{v_2, v_3, v_4, f_2, f_3, f_4\}$, $\{v_5\}$, and $\{f_5\}$. To construct the clusters of the causal ordering graph we add $S_i \cup M(S_i)$ to a cluster set $\mathcal{V}$ for each S_i in the segmentation. The segmentation of vertices into strongly connected components is displayed in Figures 6.7(d) and 6.7(e). Notice that the segmentation in Figure 6.7(d) is the same as that in Figure 6.7(e). It is known that the segmentation into strongly connected components is unique (i.e. it does not depend on the choice of the perfect matching), a result that can be found in Chapter 4. The cluster set $\mathcal{V}$ for the causal ordering graph in Figure 6.7(f) is constructed by merging clusters in the segmented graph whenever two clusters contain vertices that are matched and by adding exogenous variables as singleton clusters. The edge set $\mathcal{E}$ for the causal ordering graph is obtained by adding edges $(v \to C)$ from an endogenous vertex v to a cluster C, whenever $v \notin C$ and there is an edge from v to $f \in C$ in the directed graph. Finally, we also add edges from exogenous vertices to clusters that contain equations in which the corresponding exogenous random variables appear.

6.B Proofs

Theorem 6.1. *Consider model equations F containing endogenous variables V with bipartite graph $\mathcal{B}$. Suppose F is extended with equations F_+ containing endogenous variables in $V \cup V_+$, where V_+ contains endogenous variables that are added by the model extension.*[9] *Let $\mathcal{B}_{\text{ext}}$ be the bipartite graph associated with $F_{\text{ext}} = F \cup F_+$ and $V_{\text{ext}} = V \cup V_+$, and $\mathcal{B}_+$ the bipartite graph associated with the extension F_+ and V_+, where variables in V appearing in F^+ are treated as exogenous variables (i.e. they are not added as vertices in $\mathcal{B}_+$). If $\mathcal{B}$ and $\mathcal{B}_+$ both have a perfect matching then:*

(i) $\mathcal{B}_{\text{ext}}$ has a perfect matching,
(ii) ancestral relations in $\text{CO}(\mathcal{B})$ are also present in $\text{CO}(\mathcal{B}_{\text{ext}})$,
(iii) d-connections in $\text{MO}(\mathcal{B})$ are also present in $\text{MO}(\mathcal{B}_{\text{ext}})$.

[9] V_+ may also contain parameters or exogenous variables that appear in F and become endogenous in the extended model.



(a) Bipartite graph. (b) Directed graph (M_1). (c) Directed graph (M_2).

(d) Segmentation (M_1) (e) Segmentation (M_2) (f) Causal ordering graph.

Figure 6.7: Graphical illustration of the causal ordering algorithm that was described in Section 6.1.1. Figure 6.7(a) shows the bipartite graph that is associated with equations (6.13) to (6.17). Application of the first step of the causal ordering algorithm results in the directed graph in Figure 6.7(b) for perfect matching M_1 and that in Figure 6.7(c) for perfect matching M_2. The blue and orange edges correspond to the edges in the perfect matchings M_1 and M_2, respectively. Figures 6.7(d) and 6.7(e) show that the segmentation into strongly connected components does not depend on the choice of the perfect matching. Exogenous vertices and edges from these vertices to clusters were added to the causal ordering graph in Figure 6.7(f).

Proof. The causal ordering graph $\mathrm{CO}(\mathcal{B})$ is constructed from a perfect matching M for the bipartite graph $\mathcal{B} = \langle V, F, E\rangle$. Let M_+ be a perfect matching for $\mathcal{B}_+$. Note that $M_{\text{ext}} = M \cup M_+$ is a perfect matching for $\mathcal{B}_{\text{ext}} = \langle V \cup V_+, F \cup F_+, E_{\text{ext}}\rangle$. Following the causal ordering algorithm for $\mathcal{B}, M$ and $\mathcal{B}_{\text{ext}}, M_{\text{ext}}$, we note that $\mathcal{G}(\mathcal{B}, M)$ is a subgraph of $\mathcal{G}(\mathcal{B}_{\text{ext}}, M_{\text{ext}})$ and hence clusters in $\mathrm{CO}(\mathcal{B})$ are fully contained in clusters in $\mathrm{CO}(\mathcal{B}_{\text{ext}})$. Therefore ancestral relations in $\mathrm{CO}(\mathcal{B})$ are also present in $\mathrm{CO}(\mathcal{B}_{\text{ext}})$.

It follows directly from the definition (see Forré et al. (2017)) that σ-connections in a graph remain present if the graph is extended with additional vertices and edges. The directed graphs $\mathcal{G}(\mathcal{B}, M)$ and $\mathcal{G}(\mathcal{B}_{\text{ext}}, M_{\text{ext}})$ can be augmented with exogenous variables by adding exogenous vertices to these graphs with directed edges towards the equations in which they appear. The σ-connections in the augmentation of $\mathcal{G}(\mathcal{B}, M)$ must also be present in the augmentation of $\mathcal{G}(\mathcal{B}_{\text{ext}}, M_{\text{ext}})$. By Corollary 2.8.4 in Forré et al. (2017) and Lemma 4.12 in Blom, Van Diepen, et al. (2021) we have that d-connections in $\mathrm{MO}(\mathcal{B})$ must also be present in $\mathrm{MO}(\mathcal{B}_{\text{ext}})$. □

Theorem 6.2. *Let F, F_+, F_{ext}, V, V_+, V_{ext}, $\mathcal{B}$, $\mathcal{B}_+$, and $\mathcal{B}_{\text{ext}}$ be as in Theorem 6.1. If $\mathcal{B}$ and $\mathcal{B}_+$ both have perfect matchings and no vertex in V_+ is adjacent to a vertex in F in $\mathcal{B}_{\text{ext}}$ then:*[10]

(i) ancestral relations absent in $\mathrm{CO}(\mathcal{B})$ are also absent in $\mathrm{CO}(\mathcal{B}_{\text{ext}})$,

[10] A vertex in V_+ is considered adjacent to F if it corresponds with one of the exogenous random variables or parameters in F that become endogenous in the model extension.

(ii) d-connections absent in $\mathrm{MO}(\mathcal{B})$ *are also absent in* $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$.

Proof. Since $\mathcal{B}$ and $\mathcal{B}_+$ both have perfect matchings the results of Theorem 6.1 hold. Let $\mathcal{G}(\mathcal{B}, M)$, and $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$ be as in the proof of Theorem 6.1. Note that in M_{ext} vertices in F_+ are matched to vertices in V_+ and therefore edges between $f_+ \in F_+$ and $v \in \mathrm{adj}_{\mathcal{B}_{\mathrm{ext}}}(F_+) \setminus V_+$ are oriented as $(f_+ \leftarrow v)$ in $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$. By assumption, we therefore have that vertices in V_+ are non-ancestors of vertices in $V \cup F$ in $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$. Since $M \subseteq M_{\mathrm{ext}}$ we know that the same directed edges between vertices in V and F appear in both $\mathcal{G}(\mathcal{B}, M)$ and $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$. Notice that the subgraph of $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$ induced by the vertices $V \cup F$ coincides with $\mathcal{G}(\mathcal{B}, M)$. Hence $\mathrm{CO}(\mathcal{B})$ is the induced subgraph of $\mathrm{CO}(\mathcal{B}_{\mathrm{ext}})$ and $\mathrm{MO}(\mathcal{B})$ is the induced subgraph of $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$. □

Lemma 6.1. *Consider a first-order dynamical model in canonical form for endogenous variables V and let F be the equilibrium equations of the model. If all variables in V are self-regulating then* $\mathcal{B}$ *has a perfect matching.*

Proof. Recall that the equilibrium equation constructed from the derivative of a variable i is labelled f_i according to the natural labelling. When a variable in $v_i \in V$ is self-regulating then it can be matched to its equilibrium equation f_i. If this holds for all variables in V then $\mathcal{B}$ has a perfect matching. □

Lemma 6.2. *Let* $\mathcal{B}$ *be a bipartite graph and let M and* M' *be two distinct perfect matchings. The associated directed graphs* $\mathcal{G}(\mathcal{B}, M)$ *and* $\mathcal{G}(\mathcal{B}, M')$ *that are obtained in step (i) of the causal ordering algorithm differ only in the direction of cycles.*

Proof. This follows directly from the fact that the output of the causal ordering algorithm does not depend on the choice of the perfect matching. This result is a direct consequence of Theorem 1 and Theorem 3 in Blom, Van Diepen, et al. (2021). □

Proposition 6.1. *Consider a first-order dynamical model in canonical form for endogenous variables V and an extension consisting of canonical first-order differential equations for additional endogenous variables* V_+. *Let F and* $F_{\mathrm{ext}} = F \cup F_+$ *be the equilibrium equations of the original and extended model respectively. Let* $\mathcal{B} = \langle V, F, E \rangle$ *be the bipartite graph associated with F and* $\mathcal{B}_{\mathrm{ext}} = \langle V_{\mathrm{ext}}, F_{\mathrm{ext}}, E_{\mathrm{ext}} \rangle$ *the bipartite graph associated with* F_{ext}. *Assume that* $\mathcal{B}$ *and* $\mathcal{B}_{\mathrm{ext}}$ *both have perfect matchings. If the model extension does not introduce a new feedback loop with the original dynamical model, then d-connections in* $\mathrm{MO}(\mathcal{B})$ *are also present in* $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$.

Proof. Let E_{nat} be the set of edges $(v_i - f_i)$ associated with the natural labelling of the equilibrium equations of the extended dynamical model. Note that the feedback loops in the dynamical model coincide with cycles in the directed graph $\mathcal{G}(\mathcal{B}_{\mathrm{nat}}, M_{\mathrm{nat}})$ that is obtained by applying step (i) of the causal ordering algorithm to the bipartite graph $\mathcal{B}_{\mathrm{nat}} = \langle V_{\mathrm{ext}}, F_{\mathrm{ext}}, E_{\mathrm{ext}} \cup E_{\mathrm{nat}} \rangle$ using the perfect matching $M_{\mathrm{nat}} = E_{\mathrm{nat}}$.

By Theorem 6.1, we know that if $\mathcal{B}$ and $\mathcal{B}_+$ (the subgraph of $\mathcal{B}_{\mathrm{ext}}$ induced by $V_+ \cup F_+$) both have perfect matchings then d-connections in $\mathrm{MO}(\mathcal{B})$ must also be

present in MO($\mathcal{B}_{\text{ext}}$). Therefore, if there exists a perfect matching M_{ext} for $\mathcal{B}_{\text{ext}}$ so that each $f \in F$ is M_{ext}-matched to a vertex $v \in V$ and each $f_+ \in F_+$ is M_{ext}-matched to a vertex $v_+ \in V_+$ in $\mathcal{B}_{\text{ext}}$, d-connections in MO($\mathcal{B}$) are also present in MO($\mathcal{B}_{\text{ext}}$).

We will prove the contrapositive of the proposition, so we start with the assumption that the d-connections in MO($\mathcal{B}$) are not preserved in MO($\mathcal{B}_{\text{ext}}$). In that case, there must exist a perfect matching M_{ext} for $\mathcal{B}_{\text{ext}}$ so that there is an $f \in F$ that is M_{ext}-matched to a $v_+ \in V_+$ and a $v \in V$ that is M_{ext}-matched to a $f_+ \in F_+$. Note that since $\mathcal{B}_{\text{ext}}$ is a subgraph of $\mathcal{B}_{\text{nat}}$, this perfect matching M_{ext} is also a perfect matching for $\mathcal{B}_{\text{nat}}$. Lemma 6.2 says that $\mathcal{G}(\mathcal{B}_{\text{nat}}, M_{\text{nat}})$ and $\mathcal{G}(\mathcal{B}_{\text{nat}}, M_{\text{ext}})$ only differ in the direction of cycles. We know that vertices in V are only M_{nat}-matched to vertices in F, while vertices in V_+ are only M_{nat}-matched to vertices in F_+. Therefore, the vertices v_+ and f must be on a directed cycle in both directed graphs, as well as v and f_+. Hence the model extension F_+ introduced a new feedback loop that includes variables in the original model. □

Summary

The technique of causal ordering is used to study causal and probabilistic aspects of model equations. Causal discovery algorithms are used to learn causal and dependence structure from data. In this thesis, *Causality and independence in systems of equations*, we explored the relationship between the causal ordering and the output of causal discovery algorithms. By combining these techniques we bridged the gap between the world of dynamical systems at equilibrium and the literature about causal methods for static systems. In a nutshell, this resulted in novel insights about models with feedback which lead to a better understanding of observed phenomena in certain biological systems. Based on these ideas, we also outlined a novel approach towards causal discovery for dynamical systems at equilibrium.

This work was inspired by a desire to understand why the output of causal discovery algorithms sometimes appears to be at odds with expert knowledge. In particular, we were interested in explaining the results of causal discovery methods applied to protein expression data. For this application we found that the orientation of learned causal pathways is sometimes opposite to the causal directions in biological consensus networks. Initially, we proposed the unknown presence of measurement error as an explanation. We also presented a partial solution to this issue under additional assumptions on the true underlying causal mechanisms. We demonstrated that a phenomenon called perfect adaptation provides another plausible explanation for observed edge reversals in the context of dynamical systems at equilibrium. Roughly speaking, this entails that there is a feedback loop that ensures that certain transient causal effects are not observable at equilibrium. We proved that the application of causal discovery methods to perfectly adapted dynamical systems may lead to edge reversal in (protein signalling) models.

Additionally, our ideas can be applied to assess the robustness of qualitative model predictions. We provided examples where dependences in a model change dramatically after additional equations introduce a new feedback loop that achieves perfect adaptation. Based on this, we outlined a holistic approach to model selection so that we can reason about the existence of feedback loops in a larger system by only studying the model and observations for a subsystem. Furthermore, we proposed conditions under which the presence of a perfectly adapting feedback loop can

be detected from model equations or from a combination of experimental data and background knowledge.

This line of research was made possible by novel theoretical results and interpretations of the causal ordering algorithm. To formalize this technique we explicitly defined the notion of different types of interventions, and thereby the concept of a certain kind of causation, in the context of a system of equations. Furthermore, we extended the causal ordering algorithm so that it can be applied to a very general class of models. We proved that its output, which we call the *causal ordering graph*, encodes the effects of soft and certain perfect interventions. We introduced the notion of the *Markov ordering graph*, showed how it can be constructed from the causal ordering graph, and proved that it represents conditional independences. We demonstrated that the Markov ordering graph does not have a straightforward causal interpretation. Although graphs representing conditional independences also represent causality in static, acyclic systems, the examples in this work illustrate that the distinction between the Markov ordering graph and causal ordering graph is needed to avoid ambiguity in the context of dynamical systems with feedback at equilibrium.

The key assumption of many causal discovery algorithms is that the underlying system can be modelled by the well-known class of structural causal models. We showed that this framework is not flexible enough to fully capture the causal semantics of the equilibrium distribution of basic chemical reaction networks. To overcome these limitations, we proposed a generalization that we call causal constraints models for which we prove that they do generally capture the equilibrium distributions of dynamical systems under perfect interventions. Additional causal structure and the dependence structure of these models can be obtained through application of the causal ordering algorithm. In this thesis we looked beyond existing causal modelling frameworks to establish a relationship between dynamical models at equilibrium and causal discovery.

Samenvatting

De causale ordeningstechniek wordt gebruikt om de causale en probabilistische eigenschappen van vergelijkingsstelsel te bestuderen. Causale ontdekkingsalgoritmes worden gebruikt om de causale afhankelijkheidsstructuur van een systeem te leren uit data. In dit proefschrift, getiteld *Causaliteit en onafhankelijkheid in stelsels van vergelijkingen*, verkennen we de relatie tussen causale ordening en de uitkomsten van causale ontdekkingsalgoritmes. Door deze methodes te combineren slaan we een brug tussen de wereld van dynamische systemen in evenwicht en de literatuur over causale ontdekkingen voor statische systemen. Kortgezegd heeft dit geleid tot zowel nieuwe inzichten over dynamische modellen met terugkoppelingsmechanismes als ook tot een beter begrip van geobserveerde fenomenen in bepaalde biologische systemen. Op basis van deze ideeën stippelen we ook een nieuwe benaderingswijze uit voor causale ontdekkingen in dynamische systemen in evenwicht.

Dit proefschrift is geïnspireerd door onze wens om beter te begrijpen waarom de uitkomsten van causale ontdekkingsalgoritmes soms niet overeen lijken te komen met kennis van domeinexperts. We waren in het bijzonder geïnteresseerd in het verklaren van uitkomsten van causale ontdekkingsalgoritmes toegepast op expressiedata van eiwitten. In deze toepassing zagen we dat de oriëntatie van geleerde causale paden soms tegenovergesteld was aan de causale richting in biologische consensusnetwerken. In eerste instantie stellen we de verborgen aanwezigheid van onzekerheid door metingen voor als mogelijke verklaring. Voor dit probleem presenteren we ook een gedeeltelijke oplossing onder extra aannames over het werkelijke onderliggende causale mechanisme. Daarnaast laten we zien dat een fenomeen genaamd perfecte adaptatie een plausibele verklaring is voor de geobserveerde omkering van causale richtingen in dynamische systemen in evenwicht. Ruwweg is er dan sprake van een terugkoppeling die er voor zorgt dat kortstondige effecten niet te observeren zijn wanneer het systeem naar een evenwichtstoestand is teruggekeerd. We bewijzen dat de toepassing van causale ontdekkingsalgoritmes op perfect geadapteerde dynamische systemen kan leiden tot omkering van causale richting in modellen die gebruikt worden voor the modelleren van eiwitexpressie.

Onze ideeën kunnen bovendien toegepast worden om de robuustheid van kwalitatieve modelvoorspellingen te beoordelen. We geven voorbeelden waarin de afhan-

kelijkheden tussen variabelen in een model in sterke mate veranderen wanneer extra vergelijkingen worden toegevoegd die een terugkoppelingsmechanisme introduceren dat voor perfecte adaptatie zorgt. In het verlengde hiervan stippelen we een holistische methode uit die het mogelijk maakt om te redeneren over het bestaan van een terugkoppelingsmechanisme in een groter systeem, gegeven een model en observaties van een kleiner subsysteem. Daarnaast stellen we condities op waaronder de aanwezigheid van terugkoppeling en perfecte adaptatie gedetecteerd kan worden ofwel in modelvergelijkingen dan wel uit een combinatie van experimentele data en achtergrondkennis.

Deze onderzoeksrichting werd mogelijk gemaakt door nieuwe theoretische resultaten en interpretaties op het gebied van de causale ordeningstechniek. Om deze methode te formaliseren definiëren we expliciet de notie van verschillende types interventies. Daarmee leggen we ook het concept van een specifiek soort causaliteit vast in de context van een vergelijkingsstelsel. Om ervoor te zorgen dat de methode toepasbaar is op een algemenere modelklasse, breiden we het bestaande causale ordeningsalgoritme uit. We bewijzen dat de uitkomst, die we de *causale ordeningsgraaf* noemen, de effecten van zachte en bepaalde perfecte interventies weergeeft. We introduceren ook de notie van een *Markoviaanse ordeningsgraaf*. We laten zien hoe deze geconstrueerd kan worden uit de causale ordeningsgraaf en bewijzen dat deze conditionele onafhankelijkheden tussen variabelen representeert. We geven voorbeelden om te illustreren dat de Markoviaanse ordeningsgraaf niet altijd een eenduidige causale interpretatie heeft maar dat de causale ordeningsgraaf deze eigenschap wel bezit. Voor statische, acyclische systemen bestaan er grafen met variabelen als knopen die ook causale relaties weergeven. Echter, de voorbeelden in dit proefschrift illustreren dat een onderscheid tussen de causale ordeningsgraaf en de Markoviaanse ordeningsgraaf ambiguïteit wegneemt in de context van dynamische systemen met terugkoppeling in evenwicht.

De cruciale aanname waar vele causale ontdekkingsalgoritmes op gestoeld zijn, is dat de onderliggende mechanismes gemodelleerd kunnen worden door de bekende klasse van structurele causale modellen. We laten in dit proefschrift echter zien dat dit raamwerk niet flexibel genoeg is om de causale semantiek van een simpel chemisch reactie netwerk in evenwicht volledig te omvatten. Om deze beperkingen te omzeilen, introduceren we een generalisatie die we causale vergelijkingsmodellen noemen. We bewijzen dat deze in het algemeen wel de evenwichtsverdeling van dynamische systemen onder perfecte interventies beschrijven, terwijl toepassing van causale ordening de afhankelijkheidsstructuur en een uitgebreidere causale structuur expliciet maakt.

In dit proefschrift hebben we gepoogd om voorbij de bestaande causale modellen te kijken om de relatie tussen dynamische systemen in evenwicht en causale ontdekkingen in kaart te brengen.

Bibliography

Angrist, J. D. and Evans, W. N. (1998). 'Children and their parents labor supply: evidence from exogenous variation in family size'. *The American Economic Review* 88.3, pp. 450–477.

Araujo, R. P. and Liotta, L. A. (2018). 'The topological requirements for robust perfect adaptation in networks of any size'. *Nature Communications* 9. Article number: 1757.

Basmann, R. L. (1963). 'The causal interpretation of non-triangular systems of economic relations'. *Econometrica* 31.3, pp. 439–448.

Behbehani, G., Bendall, S., Clutter, M., Fantl, W., and Nolan, G. (2012). 'Single-cell mass cytometry adapted to measurements of the cell cycle'. *Cytometry A* 81.7, pp. 552–566.

Belgacem, I. and Gouzé, J.-L. (2012). 'Global stability of full open reversible Michaelis-Menten reactions'. *IFAC Proceedings Volumes* 45.15, pp. 591–596.

Berge, C. (1957). 'Two theorems in graph theory'. *Proceedings of the National Academy of Sciences of the United States of America* 43.9, pp. 842–844.

Berndsen, R. (1995). 'Causal ordering in economic models'. *Decision Support Systems* 15.2, pp. 157–165.

Bhat, S. P. and Bernstein, D. S. (1999). 'Lyapunov analysis of semistability'. *Proceedings of the American Control Conference* 3.

Blom, T., Bongers, S., and Mooij, J. M. (2019). 'Beyond structural causal models: causal constraints models'. *Proceedings of the 35th Annual Conference on Uncertainty in Artificial Intelligence (UAI-19)*.

Blom, T., Klimovskaia, A., Magliacane, S., and Mooij, J. M. (2018). 'An upper bound for random measurement error in causal discovery'. *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*.

Blom, T. and Mooij, J. M. (2020). 'Robustness of model predictions under extension'. arXiv: 2012.04723. A different version was presented during the Causal Discovery & Causality-Inspired Machine Learning Workshop at Neural Information Processing Systems, 2020 under the title 'Robust model predictions via causal ordering'. The present paper is to appear in the proceedings.

Blom, T. and Mooij, J. M. (2021). 'Causality and independence in perfectly adapted dynamical systems'. arXiv: 2101.11885v1. Submitted to Journal of Causal Inference.

Blom, T., Van Diepen, M. M., and Mooij, J. M. (2021). 'Conditional independences and causal relations implied by sets of equations'. *Journal of Machine Learning Research* 22.178, pp. 1–62.

Boeken, P. A. and Mooij, J. M. (2020). 'A Bayesian nonparametric conditional two-sample test with an application to local causal discovery'. arXiv: 2008.07382.

Bollen, K. A. (1989). 'Structural equations with latent variables'. John Wiley & Sons, Inc.

Bollen, K. A. and Ting, K.-F. (1993). 'Confirmatory Tetrad Analysis'. *Sociological Methodology* 23, pp. 147–175.

Bongers, S., Forré, P., Peters, J., Schölkopf, B., and Mooij, J. M. (2020). 'Foundations of structural causal models with cycles and latent variables'. arXiv: 1611.06221.

Bongers, S. and Mooij, J. M. (2018). 'From random differential equations to structural causal models: the stochastic case'. arXiv: 1803.08784.

Campbell, S. and Rose, N. (1979). 'Singular perturbation of autonomous linear systems'. *SIAM Jorunal of Math. Analysis* 10, pp. 542–551.

Cartwright, N. (1983). 'How the laws of physics lie'. Oxford University Press.

Cartwright, N. (2007). 'Hunting causes and using them. Approaches in philosophy and economics'. Cambridge University Press.

Chellaboina, V., Bhat, S. P., Haddad, W. P., and Bernstein, D. S. (2009). 'Modeling and analysis of mass-action kinetics'. *IEEE Control Systems* 29.4.

Colombo, D., Maathuis, M., Kalisch, H., and Richardson, T. S. (2012). 'Learning high-dimensional directed acyclic graphs with latent and selection variables'. *The Annals of Statistics* 40, pp. 294–321.

Cooper, G. (1997). 'A simple constraint-based algorithm for efficiently mining observational databases for causal relationships'. *Data Mining and Knowledge Discovery* 1.2, pp. 203–224.

Dash, D. (2005). 'Restructuring dynamic causal systems in equilibrium'. *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AIStats 2005)*, pp. 81–88.

Dash, D. and Druzdzel, M. J. (2008). 'A note on the correctness of the causal ordering algorithm'. *Artificial Intelligence* 172.15, pp. 1800–1808.

Dawid, A. P. (2010). 'Beware of the DAG!' *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*. Vol. 6. Proceedings of Machine Learning Research, pp. 59–86.

De Boer, R. J. (2012). 'Which of our modeling predictions are robust?' *PLOS Computational Biology* 8.7, e1002593.

De Kleer, J. and Brown, J. (1986). 'Theories of causal ordering'. *Artificial Intelligence* 29.1, pp. 33–61.

Dowe, P. (1992). 'Wesley Salmon's process theory of causality and the conserved quantity theory'. *Philosophy of Science* 59.2, pp. 195–216.

Drton, M., Massam, H., and Olkin, I. (2008). 'Moments of minors of Wishart matrices'. *The Annals of Statistics* 36.5, pp. 2261–2283.

Dulmage, A. L. and Mendelsohn, N. S. (1958). 'Coverings of bipartite graphs'. *Canadian Journal of Mathematics* 10, pp. 517–534.

Duncan, O. D. (1975). 'Introduction to structural equation models'.

Eberhardt, F. and Scheines, R. (2007). 'Interventions and causal inference'. *Philosophy of Science* 74.5, pp. 981–995.

Einstein, A. (1927). 'Newtons Mechanik und ihr Einfluß auf die Gestaltung der theoretischen Physik'. *Die Naturwissenschaften* 15. [reprinted in A. Einstein: Ideas and Opinions (Alvin Redman, London 1954)].

Evans, M. G. (1959). 'Causality and explanation in the logic of Aristotle'. *Philosophy and Phenomenological Research* 19.4, pp. 466–485.

Ferrell, J. E. (2016). 'Perfect and near-perfect adaptation in cell signaling'. *Cell Systems* 2.2, pp. 62–67.

Fisher, F. M. (1970). 'A correspondence principle for simultaneous equation models'. *Econometrica* 38.1, pp. 73–92.

Fisher, R. A. (1935). 'Design of experiments'. Hafner.

Forré, P. and Mooij, J. M. (2017). 'Markov properties for graphical models with cycles and latent variables'. arXiv: 1710.08775.

Forré, P. and Mooij, J. M. (2018). 'Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders'. *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*.

Forré, P. and Mooij, J. M. (2019). 'Causal calculus in the presence of cycles, latent confounders and selection bias'. *Proceedings of the 35th Annual Conference on Uncertainty in Artificial Intelligence (UAI-19)*.

Fritsche-Guenther, R., Witzel, F., Sieber, A., Herr, R., Schmidt, N., Braun, S., Brummer, T., Sers, C., and Blüthgen, N. (2011). 'Strong negative feedback from Erk to Raf confers robustness to MAPK signalling'. *Molecular Systems Biology* 7.1. PMID: 21613978 PMCID: PMC3130559.

Gibbs, W. (1902). 'Elementary Principles of Statistical Mechanics'. New Haven: Yale University Press.

Goldberger, A. S. and Duncan, O. D. (1973). 'Structural equation models in the social sciences'. Seminar Press.

Gonçalves, B. and Porto, F. (2016). 'A note on the complexity of the causal ordering problem'. *Artificial Intelligence* 238, pp. 154–165.

Gutú, O. (2017). 'On global inverse theorems'. *Topological Methods in Nonlinear Analysis* 49.2, pp. 401–444.

Haavelmo, T. (1943). 'The statistical implications of a system of simultaneous equations'. *Econometrica* 11.1, pp. 1–12.

Hadamard, J. (1906). 'Sur les transformations ponctuelles'. *Bulletin de la Société Mathématique de France* 34, pp. 71–84.

Haddad, W. M., Chellaboina, V., and Hui, Q. (2010). 'Nonnegative and Compartmental Dynamical Systems'. Princeton University Press.

Hall, M. (1986). 'Combinatorial theory'. 2nd edition. John Wiley & Sons.

Hall, N. (2004). 'Two concepts of causation'. *Causation and Conterfactuals*. Ed. by J. Collins, N. Hall, and L. Paul. MIT press, pp. 225–276.

Han, X. and Kloeden, P. E. (2017). 'Random Ordinary Differential Equations and Their Numerical Solution'. Springer Singapore.

Harris, N. and Drton, M. (2013). 'PC-algorithm for nonparanormal graphical models'. *Journal of Machine Learning Research* 14, pp. 3365–3383.

Hautier, J. and Barre, P. (2004). 'The causal ordering graph - A tool for modelling and control law synthesis'. *Studies in Informatics and Control Journal* 13.4, pp. 265–283.

Hitchcock, C. (2003). 'Of Humean Bondage'. *British Journal for the Philosophy of Science* 54, pp. 1–25.

Hoover, K. D. (1988). 'The new classical macroeconomics'. John Wiley & Sons.

Hopcroft, J. E. and Karp, R. M. (1973). 'An n5/2 Algorithm for Maximum Matchings in Bipartite Graphs'. *SIAM Journal on Computing* 2, pp. 225–231.

Hume, D. (1928). 'A treatise of human nature'. Ed. by A. Selby-Bigge. Oxford University Press.

Hyttinen, A., Eberhard, F., and Hoyer, P. O. (2012). 'Learning linear cyclic causal models with latent variables'. *The Journal of Machine Learning Research* 13.1, pp. 3387–3439.

Idczak, D. (2016). 'On a generalization of a global implicit function theorem'. *Advanced Nonlinear Studies* 16.1, pp. 87–94.

Iwasaki, Y. and Simon, H. A. (1994). 'Causality and model abstraction'. *Artificial intelligence* 67.1, pp. 143–194.

Kalisch, M. and Bühlmann, P. (2007). 'Estimating high-dimensional directed acyclic graphs with the PC-algorithm'. *Journal of Machine Learning Research* 8, pp. 613–636.

Kant, I. (1783). 'Prolegomena to any future metaphysics'. Trans. by P. Carus.

Karzanov, A. (1973). 'An exact estimate of an algorithm for finding a maximum flow, applied to the problem "on representatives"'. *Problems in Cybernetics*, pp. 66–70.

Kloeden, P. E. and Platen, E. (1992). 'Numerical solution of stochastic differential equations'. Springer-Verlag.

Krantz, S. G. and Parks, H. R. (2013). 'The Implicit Function Theorem: History, Theory, and Applications'. Birkhäuser Basel.

Krishnan, J. and Floros, I. (2019). 'Adaptive information processing of network modules to dynamic and spatial stimuli'. *BMC Systems Biology* 13. PMID: 30866946 PMCID: PMC6417070.

Kuroki, M. and Pearl, J. (2014). 'Measurement bias and effect restoration in causal inference'. *Biometrika* 101.2, pp. 423–437.

Lacerda, G., Spirtes, P., Ramsey, J., and Hoyer, P. O. (2008). 'Discovering Cyclic Causal Models by Independent Components Analysis'. *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*. PMID: 10213331.

Langsam, H. (1994). 'Kant, Hume, and Our Ordinary Concept of Causation'. *Philosophy and Phenomenological Research* 54.3.

Lauritzen, S. L. (1996). 'Graphical Models'. Oxford University Press.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H. (1990). 'Independence properties of directed Markov fields'. *Networks* 20.5, pp. 491–505.

Lauritzen, S. L. and Richardson, T. S. (2002). 'Chain graph models and their causal interpretations'. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64.3, pp. 321–361.

Lewis, D. (1973). 'Counterfactuals'. Blackwell.

Licata, G. (2019). 'Aristotle's Doctrine of Causes and the Manipulative Theory of Causality'. *Axiomathes* 29, pp. 635–666.

Lun, X., Zanotelli, V., Wade, J., Schapiro, D., Tognetti, M., Dobberstein, N., and Bodenmiller, B. (2017). 'Influence of node abundance on signaling network state and dynamics analyzed by mass cytometry'. *Nature Biotechnology* 35.2, pp. 164–172.

Ma, W., Trusina, A., El-Samad, H., Lim, W. A., and Tang, C. (2009). 'Defining network topologies that can achieve biochemical adaptation'. *Cell* 138.4, pp. 760–773.

Maathuis, M., Colombo, D., Kalisch, M., and Bühlmann, P. (2010). 'Predicting causal effects in large-scale systems from observational data'. *Nature Methods* 7, pp. 247–248.

Mackie, J. (1974). 'The Cement of the Universe: A Study of Causation'. Oxford: Clarendon.

Mason, S. J. (1953). 'Feedback theory – some properties of signal flow graphs'. *Proceedings of the IRE* 41.9, pp. 1144–1156.

Mason, S. J. (1956). 'Feedback theory – further properties of signal flow graphs'. *Proceedings of the IRE* 44.7, pp. 920–926.

Mogensen, S. W., Malinsky, D., and Hansen, N. R. (2018). 'Causal learning for partially observed stochastic dynamical systems'. *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*.

Mooij, J. M. and Claassen, T. (2020). 'Constraint-based causal discovery using partial ancestral graphs in the presence of cycles'. *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI-20)*. Ed. by J. Peters and D. Sontag. Vol. 124. PMLR, pp. 1159–1168.

Mooij, J. M. and Heskes, T. (2013). 'Cyclic causal discovery from continuous equilibrium data'. *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*. Ed. by A. Nicholson and P. Smyth. Corvallis, Oregon: AUAI Press, pp. 431–439.

Mooij, J. M., Janzing, D., Heskes, T., and Schölkopf, B. (2011). 'On causal discovery with cyclic additive noise models'. *Advances in Neural Information Processing Systems*, pp. 639–647.

Mooij, J. M., Janzing, D., and Schölkopf, B. (2013). 'From ordinary differential equations to structural causal models: the deterministic case'. *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*, pp. 440–448.

Mooij, J. M., Magliacane, S., and Claassen, T. (2020). 'Joint causal inference from multiple contexts'. *Journal of Machine Learning Research* 21.99, pp. 1–108.

Moravcsik, J. M. E. (1974). 'Aristotle on adequate explanations'. *Synthese* 28.1, pp. 3–17.

Murray, J. D. (2002). 'Mathematical Biology I: An Introduction'. Third edition. Springer-Verlag New York.

Muzzey, D., Gómez-Uribe, C. A., Mettetal, J. T., and Oudenaarden, A. van (2009). 'A systems-level analysis of perfect adaptation in yeast osmoregulation'. *Cell* 138.1, pp. 160–171.

Nayak, P. (1995). 'Automated Modeling of Physical Systems'. Springer-Verlag Berlin Heidelberg.

Pearl, J. (2010). 'On measurement bias in causal inference'. *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*.

Pearl, J. (2009). 'Causality. Models, Reasoning and Inference'. Second edition. Cambridge University Press.

Piaget, J. (1930). 'The Child's Conception of Causality'. *Journal of Philosophical Studies* 5.20, pp. 638–642.

Pothen, A. (1985). 'Sparse Null Bases and Marriage Theorems'. PhD thesis. Cornell University.

Pothen, A. and Fan, C.-J. (1990). 'Computing the block triangular form of a sparse matrix'. *ACM Transactions on Mathematical Software (TOMS)* 16.4, pp. 303–324.

Ramsey, J. and Andrews, B. (2018). 'FASK with interventional knowledge recovers edges from the Sachs model'. arXiv: 1805.03108.

Reichenbach, H. (1956). 'The Direction of Time'. Dover Publications.

Reiss, J. (2007). 'Causation: an opiniated introduction'. *draft available at:http://jreiss.org/Causality%20Manuscript.pdf*.

Richardson, T. S. (1996). 'Models of feedback: interpretation and discovery'. PhD dissertation. Carnegie-Mellon University.

Richardson, T. S. and Spirtes, P. (1999). 'Automated discovery of linear feedback models'. *Computation, Causation, and Discovery*, pp. 254–304.

Ringer, F. (2002). 'Max Weber on causal analysis, interpretation, and comparison'. *History and Theory* 41.2, pp. 163–178.

Rubenstein*, P. K., Weichwald*, S., Bongers, S., Mooij, J. M., Janzing, D., Grosse-Wentrup, M., and Schölkopf, B. (2017). 'Causal consistency of structural equation models'. *Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence (UAI-17)*. *equal contribution.

Rubenstein, P. K., Bongers, S., Schölkopf, B., and Mooij, J. M. (2018). 'From deterministic ODEs to dynamic structural causal models'. *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*.

Russell, B. (1903). 'The Principles of Mathematics'. Cambridge University Press.

Sachs, K., Itani, S., Fitzgerald, J., Schoeberl, B., Nolan, G., and Tomlin, C. (2013). 'Single timepoint models of dynamic systems'. *Interface Focus* 3 (4), p. 24511382.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A., and Nolan, G. P. (2005). 'Causal protein-signaling networks derived from multiparameter single-cell data'. *Science* 308 (5721), pp. 523–529.

Saeed, B., Belyaeva, A., Wang, Y., and Uhler, C. (2020). 'Anchored causal inference in the presence of measurement error'. *Proceedings of the 64th Annual Conference on Uncertainty in Artificial Intelligence (UAI-20)*.

Salmon, W. C. (1984). 'Scientific Explanation and the Causal Structure of the World'. Princeton University Press.

Scheines, R. and Ramsey, J. (2016). 'Measurement error and causal discovery'. *CEUR workshop proceedings* 1792, pp. 1–7.

Scheines, R., Spirtes, P., Glymour, C., Meek, C., and Richardson, T. (1998). 'The TETRAD project: constraint based aids to causal model specification'. *Multivariate Behavioral Research* 33, pp. 65–117.

Semmelweis, I. and Carter, K. C. (1983). 'The Etiology, Concept, and Prophylaxis of Childbed Fever'. University of Wisconsin Press.

Shin, S. Y., Rath, O., Choo, S. M., Fee, F., McFerran, B., Kolch, W., and Cho, K. H. (2009). 'Positive- and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-Mek-Erk signal transduction pathway'. *Journal of Cell Science* 122, pp. 425–435.

Silva, R., Scheines, R., Glymour, C., and Spirtes, P. (2006). 'Learning the structure of linear latent variable models'. *Journal of Machine Learning Research* 7, pp. 191–246.

Simon, H. A. (1952). 'On the definition of the causal relation'. *The Journal of Philosophy* 49 (16), pp. 517–528.

Simon, H. A. (1953). 'Causal ordering and identifiability'. *Studies in Econometric Methods*. John Wiley & Sons, pp. 49–74.

Simon, H. A. (1954). 'Spurious correlation: a causal interpretation'. *Journal of the American Statistical Association* 49.267, pp. 467–479.

Simon, H. A. and Iwasaki, Y. (1988). 'Causal ordering, comparative statistics, and near decomposability'. *Journal of Econometrics* 39, pp. 149–173.

Sokol, A. and Hansen, N. R. (2014). 'Causal interpretation of stochastic differential equations'. *Electronic Journal of Probability* 19, pp. 1–24.

Spirtes, P. (1989). 'A necessary and sufficient condition for conditional independencies to imply a vanishing tetrad difference'. Laboratory for Computational Linguistics, Carnegie Mellon University, Pittsburgh, PA.

Spirtes, P., Glymour, C., and Scheines, R. (2000). 'Causation, Prediction, and Search'. MIT press.

Spirtes, P. and Richardson, T. S. (1995). 'Directed cyclic graphical representations of feedback models'. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*.

Strobl, E. V. (2018). 'A constraint-based algorithm for causal discovery with cycles, latent variables and selection bias'. *International Journal of Data Science and Analytics* 8, pp. 33–56.

Sullivant, S., Talaska, K., and Draisma, J. (2010). 'Trek separation for Gaussian graphical models'. *The Annals of Statistics* 38.3, pp. 1665–1685.

Suppes, P. (1970). 'A Probabilistic Theory of Causality'. Amsterdam: North-Holland Publishing Company.

Tarjan, R. E. (1972). 'Depth-first search and linear graph algorithms'. *SIAM Journal on Computing* 1.2, pp. 146–160.

Thoemmes, F., Rosseel, Y., and Textor, J. (2018). 'Local fit evaluation of structural equation models using graphical criteria'. *Psychological Methods* 23.1, pp. 27–41.

Triantafillou, S., Lagani, V., Heinze-Deml, C., Schmidt, A., Tegner, J., and Tsamardinos, I. (2017). 'Predicting causal relationships from biological data: applying automated causal discovery on mass cytometry data of human immune cells'. *Scientific reports* 7.1, p. 12724.

Van Diepen, M. M. (2019). 'Extending causal ordering algorithms for general systems of equations'. Written under the supervision of T. Blom and Joris M. Mooij. MA thesis. University of Amsterdam.

Voortman, M., Dash, D., and Druzdzel, M. J. (2010). 'Learning why things change: the difference-based causality learner'. *Proceedings of the Twenty-Sixth Annual Conference on Uncertainty in Artificial Intelligence (UAI-10)*.

Wishart, J. (1928). 'Sampling errors in the variance of two factors'. *British Journal of Psychology* 19, pp. 180–187.

Woodward, J. (2003). 'Making things happen'. Oxford University Press.

Wright, P. G. (1928). 'The tariff on animal and vegetable oils'. The MacMillan Company.

Wright, S. (1921). 'Correlation and causation'. *Journal of Agricultural Research* 20, pp. 557–585.

Zhang, J. (2008). 'On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias'. *Artificial Intelligence* 172.16–17, pp. 1873–1896.

Zhang, K., Gong, M., Ramsey, J., Batmanghelich, K., Spirtes, P., and Glymour, C. (2017). 'Causal discovery in the presence of measurement error: identifiability conditions'. *UAI workshop on causality*.

Zhang, K., Gong, M., Ramsey, J., Batmanghelich, K., Spirtes, P., and Glymour, C. (2018). 'Causal discovery with linear non-Gaussian models under measurement error: structural identifiability results'. *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*.

Overview of publications

The introduction in Chapter 1 has not been previously published and was entirely written by me for this thesis. The content of Chapter 2 is based on:

Blom, T., Klimovskaia, A., Magliacane, S., and Mooij, J. M. (2018). 'An upper bound for random measurement error in causal discovery'. *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*.

My personal contributions were the original idea to use an upper bound for measurement error to improve the accuracy of local causal discovery on the protein expression data, the mathematical theory, simulations, most of the writing, and implementation of the application to protein expression data. Klimovskaia first suggested to the protein expression dataset that we used and proposed the preprocessing steps. Magliacane and Mooij contributed through discussions and editing of the papers. Chapter 3 is based on:

Blom, T., Bongers, S., and Mooij, J. M. (2019). 'Beyond structural causal models: causal constraints models'. *Proceedings of the 35th Annual Conference on Uncertainty in Artificial Intelligence (UAI-19)*.

The original idea to model dynamical systems at equilibrium with causal constraints arose during discussions about the basic enzyme reaction with both Bongers and Mooij. My personal contributions are the idea to consider semi-stable dynamical systems and most of the writing. The mathematical theory was developed by me and was inspired by discussions with Bongers and Mooij. The contents of Chapter 4 are based on:

Blom, T., Van Diepen, M. M., and Mooij, J. M. (2021). 'Conditional independences and causal relations implied by sets of equations'. *Journal of Machine Learning Research* 22.178, pp. 1–62.

The original idea to use the extended causal ordering algorithm to find a Markov property for a set of equations was initiated by me. The formalization of the extended causal ordering algorithm and the proofs for its uniqueness properties were developed and written by Van Diepen during her master's thesis project under supervision of Mooij and me. The notion of the causal ordering graph was first proposed

by Mooij. Patrick Forré helped us work out the relationship between d-separations in the Markov ordering graph and the σ-separation criterion in the oriented graph. The mathematical theory and proofs about probabilistic and causal properties of the Markov ordering graph and the causal ordering graph were developed by me and aided by continuous discussions with Mooij. I was responsible for most of the writing. Chapter 5 is based on:

Blom, T. and Mooij, J. M. (2021). 'Causality and independence in perfectly adapted dynamical systems'. arXiv: 2101.11885v1. Submitted to Journal of Causal Inference.

The idea to use causal ordering to study perfectly adapted dynamical systems was my personal contribution as well as the development of theoretical results and most of the writing. The notions of the equilibrium causal ordering graph and the dynamic causal ordering graph were proposed by Mooij. Chapter 6 is based on:

Blom, T. and Mooij, J. M. (2020). 'Robustness of model predictions under extension'. arXiv: 2012.04723. A different version was presented during the Causal Discovery & Causality-Inspired Machine Learning Workshop at Neural Information Processing Systems, 2020 under the title 'Robust model predictions via causal ordering'. The present paper is to appear in the proceedings.

My personal contributions include the idea to apply causal ordering to assess the robustness of model predictions, as well as the idea to use it for model selection. The majority of the writing and development of mathematical theory were done by me.

Acknowledgements

Nu is de tijd aangebroken om mijn promotietraject af te sluiten door terug te blikken op de bijzondere jaren die ik heb mogen beleven. Voor mij was het een onvergetelijke en leerzame periode van ongekende mogelijkheden. Het aanstekelijke enthousiasme van collega's zorgde ervoor dat ik me met veel plezier op het onderzoek heb gestort. De steun van familie, vrienden, collega's, en anderen heeft mij geholpen om ook door te zetten op de momenten dat het allemaal tegen leek te zitten. In dit hoofdstuk wil ik graag iedereen bedanken die de afgelopen jaren, direct of indirect, heeft bijgedragen aan het tot stand komen van dit boek.

Allereerst wil ik mijn promotor Joris Mooij bedanken voor de fantastische samenwerking en zijn persoonlijke betrokkenheid bij alle aspecten van dit project. Een betere begeleider kan ik mij niet voorstellen. Ik kreeg alle ruimte om te onderzoeken wat mijn interesse wekte terwijl mij toch altijd een kompas werd aangeboden om de nodige richting te geven aan de ideeën die hieruit voortkwamen. Ik denk vooral vaak terug aan het vrijuit filosoferen en de levendige discussies die ons tot verrassende nieuwe inzichten brachten. Voor de inhoudelijke begeleiding en de ondersteuning bij het schrijven van artikelen zal ik altijd dankbaar blijven. Ik dank ook mijn co-promotor Max Welling voor de inspirerende omgeving die hij heeft gecreëerd in het machine learning lab. Het was een waar genoegen om als PhD student deel uit te mogen maken van dit getalenteerde en gezellige team.

I would like to express my gratitude to the members of the defense committee: Philip Dawid, Thomas Richardson, Sonja Smets, and Peter Sloot. Daarnaast wil ik Johannes Textor niet alleen bedanken voor deelname aan de commissie maar vooral ook voor de ontmoetingen die stof tot nadenken gaven en een inspiratiebron vormden voor het onderzoek. In het bijzonder wil ik Patrick Forré bedanken voor nieuwe wiskundige inzichten, zijn heldere manier van uitleggen, en de vrolijkheid tijdens de lunch en conferenties.

Het doen van mijn PhD onderzoek was niet half zo leuk geweest zonder de mensen in onze causaliteitsgroep. I would like to thank Sara for making me feel right at home from the first day that I started and for supporting me with my first paper submissions. Ik denk met veel vreugde terug aan de gezellige tijd dat ik een kantoor mocht delen met Thijs. Verder wil ik graag Stephan bedanken voor de suc-

cesvolle samenwerkingen en het vele sparren. The group would not have been the same without Philip, Noud, en Theodora. Thank you all so much for all the fruitful discussions and laughter over the years. De samenwerking met Mirthe heb ik ervaren als een ontzettend fijne periode doordat we echt op een lijn zaten en daar wil haar graag voor bedanken. Furthermore, I would like to thank Stephan, Tim, Alexander, and Qi for being great office mates. Finally, I am grateful to Shihan, Zeynep, Rianne, and many other (former) members of the machine learning lab for good times and interesting conversations. Tot slot wil ik Félice bedanken voor haar betrokkenheid en administratieve ondersteuning.

Ik had niet zo ver kunnen komen zonder de vrolijkheid van vrienden die op verschillende momenten in mijn leven een rol van betekenis hebben gespeeld. In het bijzonder wil ik Alexandru, Lennart, Anna, Luuk, Peter, Gijs, Christiaan, Philip, en Sanne bedanken, maar ook de vriendengroep uit Maarssen: Vincent, Martyn & Kai, Folkert, Harmen & Merel, Jouke, Bastiaan & Merel, Vincent & Michel, Daan, en Tessa. Bedankt voor jullie vriendschap en voor de gezellige tijden die we hebben mogen beleven.

Mijn schoonfamilie heeft mij veel steun geboden tijdens het schrijven van dit proefschrift. Jan, Yvonne, Susanne, en Joey hebben mij met open armen in de familie ontvangen. Ik ben ze veel dank verschuldigd voor de liefde, steun, en wijze raad die ze mij de afgelopen jaren hebben gegeven. De passie van Jan voor het doen van onderzoek was voor mij een stimulans om ook als onderzoeker te willen werken en Yvonne heeft me geholpen om meer inzicht te krijgen in mijzelf.

Op uitdagende momenten gaan mijn gedachten uit naar mijn oma Janny die op z'n Gronings gezegd zou hebben dat het allemaal wel weer 'goud' zou komen. Van mijn pake Rienk herinner ik me dat hij zorgde voor een vrolijke noot; erg belangrijk om de moed erin te houden. Wanneer het nodig was om voet bij stuk te houden dacht ik aan de volharding van mijn opa Harm. Mijn oma Tineke heeft mij van jongs af aan bijgebracht om elke dag weer dankbaar te zijn voor alle mooie dingen die ons gegeven zijn. Ik wil haar bedanken voor deze wijze les en haar niet aflatende steun. Eggie, Hyke, Tsjitske, Gretha, en Frans hebben mij geleerd om je eigen ding te doen en dat je kansen moet grijpen wanneer ze je gegeven worden, maar vooral ook om met beide benen op de grond moet blijven staan. Bedankt voor deze belangrijke lessen.

Ik ben mijn vader Aeilko dankbaar voor het feit dat hij ons huis vulde met prachtige boeken en muziek waar ik helemaal in op kon gaan. Zijn ingetogen liefde voor de exact wetenschappen is op mij overgeslagen. Ik heb altijd tegen hem opgekeken. Mijn moeder Janke is mijn steun en toeverlaat en zonder haar was ik niet de persoon geweest die ik vandaag de dag ben. Zij heeft mij door dik en dun gesteund en stond altijd voor mij klaar. Het is niet in woorden uit te drukken hoe belangrijk zij voor mij is. Dank je wel, mama.

Tot slot wil ik nog enkele woorden richten tot mijn echtgenoot Kay. Ons avontuur begon twaalf jaar geleden toen jij mijn hart veroverde. We ontdekten samen de wiskunde tijdens onze studie en de wereld tijdens onze reizen. Het leven met jou is

één groot avontuur en er is niemand met wie ik het liever zou willen delen dan met jou. Of we nu eindeloos discussiëren over levensvragen, duiken in prachtige zeeën, sparren over wiskundige vraagstukken, wandelen in de wildernis, elkaar helpen met artikelen en presentaties, ronddwalen in vreemde landen, of gewoon lekker gek doen, het mooiste is dat ik bij jou kan zijn. Dank je wel voor je liefde en steun.

About the author

Tineke Blom was born in 1990 in Dokkum, Friesland, the Netherlands. She completed her secondary education at CSG Comenius in Leeuwarden in 2008, and subsequently she started her studies at Utrecht University. After obtaining her bachelor's degrees in physics and mathematics, she pursued a master's degree in mathematics, specializing in probability and statistics. She obtained her master's degree in 2015.

She started a PhD project at the University of Amsterdam under supervision of Joris Mooij and Max Welling in 2016. Her research focussed on causality and independence in systems of equations, in particular dynamical systems at equilibrium. This resulted in several publications, which provided the basis for the contents of this dissertation.

Tineke will defend her thesis at the Agnietenkapel, Amsterdam on September 23rd, 2021.