UNIVERSITY OF CALIFORNIA,
IRVINE

Structural Equation Modeling with Latent Variables

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Electrical Engineering and Computer Science

by

Saeed Karimi-Bidhendi

Thesis Committee:
Chancellor's Professor Hamid Jafarkhani, Chair
Assistant Professor Yanning Shen
Associate Professor Aparna Chandramowlishwaran

2021

DEDICATION

To

my parents

for their endless love, support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor professor Hamid Jafarkhani, who has been a great mentor for me throughout my graduate studies. His continuing assistance, indispensable advice and technical guidance has made my journey toward graduate degree to be an invaluable experience. I also would like to thank my committee members professor Yanning Shen and professor Aparna Chandramowlishwaran for their priceless support, time and review of this work.

ABSTRACT OF THE THESIS

Structural Equation Modeling with Latent Variables

By

Saeed Karimi-Bidhendi

MASTER OF SCIENCE in Electrical Engineering and Computer Science

University of California, Irvine, 2021

Chancellor's Professor Hamid Jafarkhani, Chair

Discovering causal relationships between variables is a difficult unsupervised learning task, which becomes more challenging if there are unobserved common causes between pairs of variables. Often it is not feasible to uniquely recover causal relations when only observational data is available. When experimental data is obtainable through interventions, we present a method for guaranteed identification under mild assumptions. We consider a linear structural equation model where there are independent unobserved common causes between pairs of observed variables. The generative process of latent effects is given by the mixing method of blind source separation problem. Our objective is to disentangle the observed causal effects from latent confounders and learn the model parameters that are consistent with observational and experimental data. By exploiting the invariance of latent factors across various interventions, we present matching methods as a way to combine the information across various interventions. Finally, we propose an identification algorithm that uses efficient tensor decomposition for a unique recovery of model parameters and disentangling the latent confounders from observed causal effects.

Chapter 1

Introduction

...out of nothing I have created a strange new universe.

János Bolyai

1.1 Background

While causality implies lawlike necessities, probabilistic models are still the common mathematical language for analysis of causal relations since observations are often plagued with uncertainty. In graphical interpretation of probability, vertices correspond to random variables and edges correspond to certain relationships between pairs of variables. Most of our discussion involves a directed graph G, where each directed edge represents a cause-and-effect relationship between two vertices, with no cycle, sometimes referred to as a directed acyclic graph or DAG. The acyclic property of the graph comes from the fact that a cause precedes its effect in time and no variable is a cause of itself. This graphical representation reduces the chain rule decomposition of the joint probability over variables

to a factorization in which each variable is condition on a specific set of variables. More precisely, for a variable such as X_i in the DAG G, its Markovian parent $\mathrm{Pa}(X_i)$, or parent for short, is the set of variables in G from which there is a direct edge to the variable X_i; therefore, the chain rule factorization is reduced to

$$P(X_1, \cdots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}(X_i)). \tag{1.1}$$

Intuitively, not only $\mathrm{Pa}(X_i)$ represents the set of direct causes of X_i, but also it mediates any association between X_i and other variables since $\mathrm{Pa}(X_i)$ intersects all directed paths to X_i in the DAG G. The distribution P and DAG G that admit the factorization in Eq. (1.1) are called compatible or Markov relative with each other. This correspondence is further captured by both a graphical criterion called d-separation that provides a simple way to characterize the independence relations between variables [27], and the notion of stability, also referred to as DAG-isomorphism or faithfulness, which states that all independence relations embedded in a probability distribution P is entailed in the structure of G.

It is worth noting that the interpretation of DAGs as carrier of independence hypothesis does not necessarily imply causation; however, there are several advantages of building DAG models based on causal rather than correlational information. For instance, judgements required in the construction of the model will be more meaningful and reliable. Axiomatically, if conditional independence relations are result of causal relationships, then representing those relationships would be a more natural and reliable way of expressing what we believe about the world. Another significant advantage of building Bayesian networks based on causal information is their ability to respond to external actions or interventions, which will constitute most of our discussion throughout this work.

The ability of causal Bayesian models comes from the assumption that each parent-child relationship in the graph is an autonomous mechanism, i.e. we can change one such link

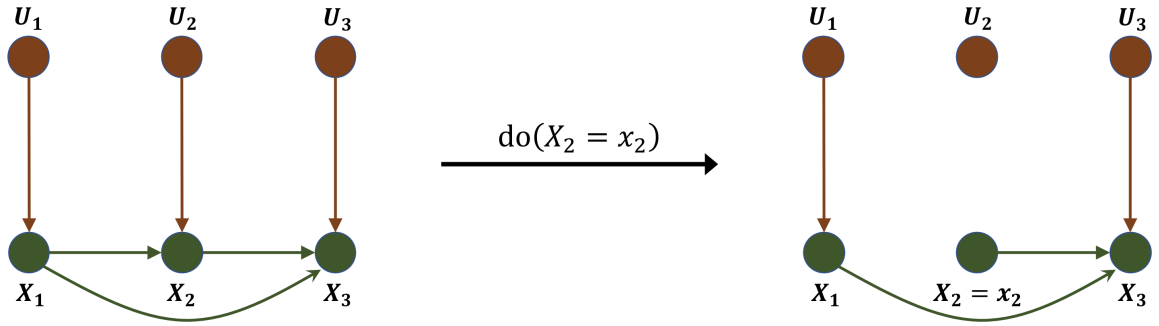


Figure 1.1: An exemplary change in the graphical model under intervention.

without changing other relationships. It can be easily seen that causal models contain more information than probabilistic models since they can tell us how the probability of events would change due to actions while probabilistic models only tell us how probabilities would change due to observations. The assumption of autonomy is essential in causal models since by its virtue we can assume that all changes are local and we only need to modify the factors in Eq. (1.1) corresponding to the action taken, and leave the rest unchanged.

Actions are commonly denoted by the do(.) operator where $\mathrm{do}\,(X_i = x_i)$ represents the action that sets the value of the variable X_i to x_i [27]. Note that while the observation $X_i = x_i$ can lead to conditional inference for the value of the parent set $\mathrm{Pa}\,(X_i)$, the intervention $\mathrm{do}\,(X_i = x_i)$ isolates the variable X_i from its parents and eliminates the mechanisms that used to affect X_i. In particular, the action $\mathrm{do}\,(X_i = x_i)$ removes edges of the graphical model G that are from $\mathrm{Pa}\,(X_i)$ to X_i, as illustrated in Figure 1.1. This in turn changes the factorization in Eq. (1.1) since the terms corresponding to intervened variables are removed from the factorization.

Historically, causal models were introduced using functional equations rather than graphical models. In functional causal models, causal relationships are expressed in the form of deterministic functional equations, and probabilities are defined through the premise that some variables in these equations are hidden. Generally, a functional causal model is a set

of equations of the form:

$$X_i = f_i\left(\mathrm{Pa}\left(X_i\right), U_i\right), \qquad 1 \leq i \leq n \tag{1.2}$$

where $\mathrm{Pa}\left(X_i\right)$ is the set of variables that directly determine the value of X_i, and U_i is the hidden or background factor. If the set of functions in Eq. (1.2) are linear, the resulting model is referred to as a linear structural equation model (SEM) and Eq. (1.2) becomes:

$$X_i = \sum_{j \neq i} a_{i,j} X_j + U_i, \qquad 1 \leq i \leq n. \tag{1.3}$$

Note that in linear models, $\mathrm{Pa}\left(X_i\right)$ corresponds to those variables whose coefficients are non-zero. In a structural causal model, each variable has a distinct autonomous equation associated to it. Equations are autonomous in the sense that alternation in one equation will not change another one. In particular, the action $\mathrm{do}\left(X_i = c\right)$ corresponds to substituting the $i-$th equation in Eq. (1.3) by $X_i = c$. The learning task of function approximation for the functional causal model in Eq. (1.2), also known as causal inference, is crucial in many scientific applications and it is at the heart of this study.

1.2 Causal Discovery

Discovering causal relationship between two variables can be translated to see how sensitive the value of one variable is to a change in the other variable. This can be done through calculating the post-interventional distribution of a variable, performing soft or hard actions etc. Observational data plays a critical role in causal discovery and inference; however, it may not provide sufficient information for a unique identification, especially if the causal model includes latent variables [27].

In principle, there may be several models that would fit a given distribution or observation. When only observational data is available, according to the principle of Occam's razor, one would choose the simplest model that is equally consistent with the data, which we refer to as the minimal model [27]. Although assumptions such as minimality and stability will limit our search to a smaller space of structures, they often lead to a class of observationally equivalent models and do not guarantee a unique or practical solution for the problem. Hence, interventions accompanied by their respective experimental data are essential for inference in these models. Intuitively, each action assigns a set of controlled values to a group of variables and isolates them from their parents in the causal graph. The response of the system to these actions, that comes in the form of experimental data, provides information about the sensitivity of variables to those that are under intervention. While the choice of intervention varies from one application to another, here we consider a class of latent variable models called independent component analysis (ICA), and provide a general framework that guarantees unique recovery of model parameters.

1.3 Related Works

Recently, learning-based methods have been an integral part of predictive modeling [22, 23]. However, for many applications in various fields such as economic, sociology, medicine etc., it is desirable to discover causal relations among variables instead of association between them because predictions based on causal dependencies are more robust and reliable. Latent variables have gradually become an inseparable part of modeling in the past few decades [34]. In many practical applications of empirical sciences, there exist variables that are unobserved but nevertheless affect the visible factors in the model. Utilizing hidden variables not only leads to a more descriptive representation of the model, but also relaxes the computational complexity of many algorithms through dimensionality reduction. However,

parameter learning for latent variable models can lead to many or possibly infinite models that would fit the same observational data due to unobserved effects. Thus, it is sometimes necessary to make some assumptions about hidden variables or the structure of graphical model in order to make the identification process feasible [4]. A method is presented in [27] to calculate the post-interventional distribution when no latent variable exists in the causal model. Moreover, causal learning algorithms such as IC and PC are presented in [27] and [36], respectively, to identify an observationally equivalent class of causal models under the assumption of causal sufficiency where there exists no unobserved common cause between any two observed variables. Recent works on causal model learning in the presence of latent confounding variables generally take only observational data into consideration [17, 30, 39], and they ignore the information embedded in the experimental data obtained through actions; thus, strong modeling assumptions is usually made for the identification process which may not hold in many practical settings. As another example, the linear and acyclic LiNGAM model in [32, 33], and its generalization to the latent confounder setting in [30], are proposed for causal discovery under the assumption of non-Gaussian noise on the observational data in an ICA-based model; however, only a class of models compatible with the data can be recovered and no unique identification is guaranteed. Generalization of LiNGAM to hidden variables and latent Gaussian confounders can be found in [18] and [6], respectively. The equivalence class of models obtained by these methods can be further narrowed down given background knowledge or assumptions such as polytrees for the structure and graph of the causal model [19, 31].

The main characteristic of causal models is their ability to perform interventions and obtain experimental data [27]. There is a rich literature on causal discovery using controlled combinatorics and search experiments [10, 12, 21]. By allowing multi-variable actions, a linear cyclic causal model can be recovered in the presence of confounding effects [20]. When the causal model contains no latent or selection variable, the causal directed acyclic graph can be identified using methods such as PC algorithm. Reference [36] demonstrates that

the PC algorithm is both sound and complete under faithfulness and causal sufficiency assumptions. However, these methods fail in the presence of hidden or selection variables. Reference [36] represents Fast Causal Inference (FCI) algorithm to learn the Markov equivalence class of DAGs with latent and selection variables under faithfulness assumption using conditional independent relations among visible variables. Although [36] and [37] demonstrate the soundness of FCI algorithm, this method is computationally expensive for large graphs. Therefore, several variations of FCI algorithm have been proposed to relax its computational complexity. Reference [35] introduces the Anytime FCI algorithm in which the size of conditioning set is limited by a specific bound. Moreover, [7] represents RFCI algorithm that uses a smaller number of conditional independence tests than FCI. Although both Anytime FCI and RFCI are typically less informative compared to FCI, in many practical scenarios their performance is close to FCI. A method is presented in [38] to recover the causal order in causal latent variable models. Recently, several methods have been proposed to view interventions as active data queries with disparate costs [2, 15], and learn the causal graph with confounding variables in a constrained budget setting [1, 13, 14, 24].

In general, recovering the directed acyclic graph between observed variables in a causal model can be accomplished in various ways such as incorporating the causal knowledge that has been acquired through past observations and experiments, testing a collection of conditional independence relationships to accept or reject a candidate graph based on the d−separation criterion [27] etc.; however, recently various causal discovery methods have been proposed for this purpose. In [11, 20], a method is presented that uses atomic actions to recover the causal ordering. In particular, the method in [11] uses $n-1$ actions to recover the causal directions. The idea behind their method is to examine which variables have changed after performing single-variable actions. Hence, throughout this work we assume that the causal ordering of observed variables in our model is known or recovered using these methods.

1.4 Contributions

In this work, we consider a new model, called structural equation modeling with latent variables (SEM-LV), where a linear SEM governs causal effects on each observed variable, and mutually independent hidden variables affect observables through a latent generative process given by the ICA model. Our aim is to disentangle the causal effects between observed variables from latent confounders, and learn the model parameters using a carefully selected set of interventions. First, we exploit the invariance of latent factors across different actions, and propose a set of matching algorithms to find a unique correspondence between system's response to different actions. Then, we present a framework that not only provides necessary and sufficient conditions of identifiability, but also it can be used to determine all possible sets of actions for a guaranteed and unique recovery of model parameters.

Chapter 2

Problem Formulation

In this section, we study the graphical representation of the structural equation model SEM-LV, as shown in Figure 2.1, that consists of n observed and k latent variables, respectively. Let $X = \{X_1, \cdots, X_n\}$ be the set of observed variables, also referred to as visible variables, and let $H = \{H_1, \cdots, H_k\}$ be the set of latent variables, also referred to as hidden or confounding variables. We denote a sample realization of observed and latent variables by $\vec{\mathbf{x}} = (x_1, \cdots, x_n)^T \in \mathbb{R}^n$ and $\vec{\mathbf{h}} = (h_1, \cdots, h_k)^T \in \mathbb{R}^k$, respectively. The system model in Figure 2.1 can be represented as a directed acyclic graph (DAG) $G = (X \bigcup H, E)$ where vertices consist of observed and latent variables, and E includes all directed edges between vertices, which represents causal relationships between neighboring variables. Note that the acyclic property of the graph G is due to the requirement that a cause precedes its effect in time, and a variable cannot cause itself [27]. In particular, there is no self loop in the graph G. For an observed variable $X_i \in X$, the parent set of X_i is denoted by $\mathrm{Pa}(X_i) \subseteq X \bigcup H$, and it contains all vertices in the graph G from which there is a directed edge to X_i. We assume that the latent variables are independent, and the causality between latent and observed variables is modeled by directed edges from the latent to observed variables, and not the other way around. Hence, the DAG G can be regarded as a quasi-bipartite graph.

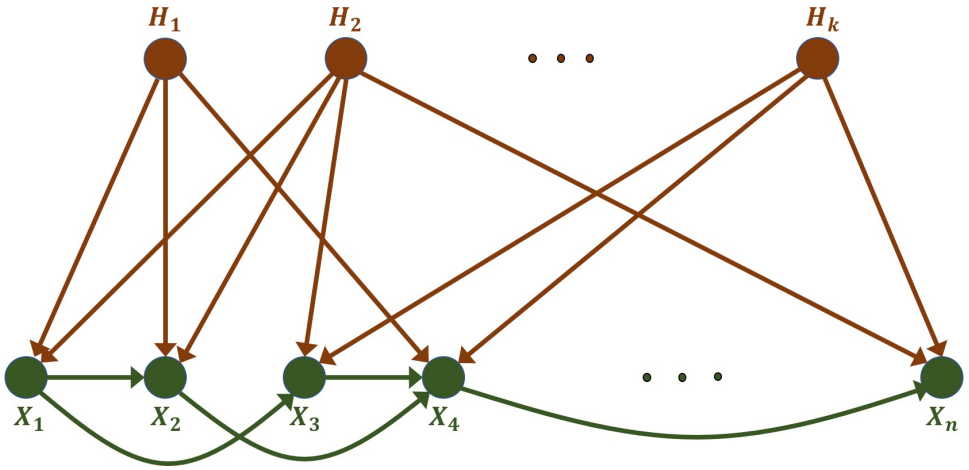


Figure 2.1: Graphical representation of the SEM-LV model.

2.1 Structural Equation Models

Inspired by the classical independent component analysis (ICA) method, we consider a linear model for the structural equations and causal relations between visible and latent variables:

$$\vec{\mathbf{x}} = \mathbf{A}\vec{\mathbf{h}} + \mathbf{B}\vec{\mathbf{x}}, \tag{2.1}$$

where $\mathbf{A} = [a_{i,j}]_{1\leq i\leq n, 1\leq j\leq k} \in \mathbb{R}^{n\times k}$ is referred to as the mixing matrix, and $\mathbf{B} = [b_{i,j}]_{1\leq i,j\leq n} \in \mathbb{R}^{n\times n}$ is referred to as the causal matrix, respectively. The mixing matrix $\mathbf{A}$ corresponds to the edges between latent and observed variables in Figure 2.1, and represents the correlation among them. The causal matrix $\mathbf{B}$ corresponds to the edges between observed variables in Figure 2.1, and encodes the causal relationships among them. By rearranging the terms in Eq. (2.1), we can rewrite the system model as:

$$\vec{\mathbf{x}} = \mathbf{C}\vec{\mathbf{h}}, \tag{2.2}$$

where $\mathbf{C} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{A}$ is the new mixing matrix, and $\mathbf{I} \in \mathbb{R}^{n\times n}$ is the identity matrix. Throughout this work, we assume that the mixing matrix $\mathbf{A}$ is a full rank matrix. Moreover,

we suppose that the faithfulness assumption holds, i.e. if there is a causal path from H_j to X_i then $a_{i,j} \neq 0$, and similarly $b_{i,j} \neq 0$ if there is a causal path from X_j to X_i. The objective of this work is to disentangle causal effects of visible variables from hidden variables, i.e. to recover the mixing matrix $\mathbf{A}$ and the causal matrix $\mathbf{B}$ in Eq. (2.1) by performing interventions on visible variables.

2.2 Interventions

An intervention or action refers to a controlled experiment that sets the value of one or more variables, and isolates those variables from the effect of their direct causes. Note that we can only perform actions on the set of visible variables X since the variables in H are hidden. Intervention changes the structure of the causal graph G in Figure 2.1 in the sense that it removes all the incoming edges to the set of intervened variables. This in turn changes the structural equation modeling, and in particular alters the mixing and causal matrices $\mathbf{A}$ and $\mathbf{B}$ in Eq. (2.1). However, since the parent-child relationships in the graph are assumed to be autonomous mechanisms [27], i.e. we can change one such link without changing other relationships, then performing an action on a particular set of variables does not change the linear models corresponding to other variables.

Here, we narrow our attention to those actions that set the value of variables to zero. Borrowing the do(.) operator notation from [27], the model after the intervention $\text{do}(X_i = 0 | \forall i \in \mathcal{S})$ on the set $\mathcal{S} \subseteq \{1, \cdots, n\}$ of visible variables can be written as:

$$\vec{\mathbf{x}} = \mathbf{A}_{\mathcal{S}}\vec{\mathbf{h}} + \mathbf{B}_{\mathcal{S}}\vec{\mathbf{x}}, \tag{2.3}$$

where $\mathbf{A}_{\mathcal{S}}$ and $\mathbf{B}_{\mathcal{S}}$ are formed by setting the row i of matrices $\mathbf{A}$ and $\mathbf{B}$ to all-zero vectors,

for all $i \in \mathcal{S}$. By rearranging the terms, Eq. (2.3) can be rewritten as:

$$\vec{\mathbf{x}} = \mathbf{C}_{\mathcal{S}}\vec{\mathbf{h}}, \tag{2.4}$$

where the new mixing matrix $\mathbf{C}_{\mathcal{S}} = (\mathbf{I} - \mathbf{B}_{\mathcal{S}})^{-1}\mathbf{A}_{\mathcal{S}}$ is also referred to as the response of the SEM-LV model to the intervention $\text{do}(X_i = 0 | \forall i \in \mathcal{S})$. Let $\mathbf{P}_{\mathcal{S}} = [p_{i,j}]_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$ be a diagonal binary matrix such that $p_{i,i} = 0$ if $i \in S$, and $p_{i,i} = 1$ otherwise. We refer to $\mathbf{P}_{\mathcal{S}}$ as the action matrix since it is in one-to-one correspondence with the intervention $\text{do}(X_i = 0 | \forall i \in \mathcal{S})$. Now, we can model the relationship between matrices $\mathbf{A}_{\mathcal{S}}$ and $\mathbf{A}$ as:

$$\mathbf{A}_{\mathcal{S}} = \mathbf{P}_{\mathcal{S}}\mathbf{A}. \tag{2.5}$$

Similarly, the relation between matrices $\mathbf{B}_{\mathcal{S}}$ and $\mathbf{B}$ can be modeled as:

$$\mathbf{B}_{\mathcal{S}} = \mathbf{P}_{\mathcal{S}}\mathbf{B}. \tag{2.6}$$

Hence, the response to the action $\text{do}(X_i = 0 | \forall i \in \mathcal{S})$ can be rewritten as:

$$\mathbf{C}_{\mathcal{S}} = (\mathbf{I} - \mathbf{P}_{\mathcal{S}}\mathbf{B})^{-1}\mathbf{P}_{\mathcal{S}}\mathbf{A}. \tag{2.7}$$

Our goal is to use the observational data, along with the experimental data from an appropriate set of actions, to recover the mixing matrix $\mathbf{A}$ and the causal matrix $\mathbf{B}$. More precisely, we aim to find the necessary and sufficient condition on the selected set of actions that renders matrices $\mathbf{A}$ and $\mathbf{B}$ identifiable.

2.3 Assumptions

Throughout this work, we make the following mild assumptions about the parameters and the system model in Figure 2.1. Since the mixing and causal matrices are usually random and noisy in practice, these assumptions hold in many practical applications.

Assumption 1. *For any subset $\mathcal{S} \subseteq \{1, \cdots, n\}$ with cardinality of at least 2, every two columns of the submatrix of $\mathbf{A}$ restricted to rows in $\mathcal{S}$ are linearly independent.*

Assumption 2. *Let $\mathbf{A}(i,j)$ be the element in Row i and Column j of the mixing matrix $\mathbf{A}$. For all $i, j \in \{1, 2, ..., k\}$ such that $i \neq j$, and for all $t \in \{1, 2, ..., n\}$ we assume that:*

$$\kappa_i \times \mathbf{A}(t,i)^4 \neq \kappa_j \times \mathbf{A}(t,j)^4, \tag{2.8}$$

where

$$\kappa_i = \mathbb{E}\left[H_i^4\right] - 3, \qquad \forall i \in \{1, \cdots, k\}. \tag{2.9}$$

Assumption 3. *The mixing matrix $\mathbf{A}$ is full rank, and every k rows of $\mathbf{A}$ constitute a full rank $k \times k$ matrix.*

Assumption 4. *Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be two binary diagonal matrices and define $\mathbf{D} = \big(\mathbf{I} - \mathbf{P}_1\mathbf{B}\big)^{-1}\mathbf{P}_1 - \big(\mathbf{I} - \mathbf{P}_2\mathbf{B}\big)^{-1}\mathbf{P}_2$. If $\big|\{i \mid \mathbf{P}_1(i,i) \neq \mathbf{P}_2(i,i)\ ,\ 1 \leq i \leq n\}\big| \leq k$, then the column-wise concatenation of matrices $\mathbf{A}$ and $\mathbf{I} - \mathbf{D}^\dagger\mathbf{D}$ has full row rank; otherwise, the row-wise concatenation of $\mathbf{D}$ and $\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger$ has full column rank, where $\mathbf{A}^\dagger$ and $\mathbf{D}^\dagger$ are any solution to $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$ and $\mathbf{D}\mathbf{D}^\dagger\mathbf{D} = \mathbf{D}$, respectively, such as the pseudoinverse of $\mathbf{A}$ and $\mathbf{D}$. Matrices $\mathbf{A}^\dagger$ and $\mathbf{D}^\dagger$ are also called the weak inverse or g−inverse of $\mathbf{A}$ and $\mathbf{D}$, respectively [26, 28].*

Chapter 3

Tensor Decomposition and the Method of Moments for ICA

The blind source separation problem, first emerged in the framework of neural modeling and signal processing [16, 29], consists of an unknown linear system that mixes a set of hidden source signals and provides a linear combination of them as the observable output. Due to the unknown characteristics and permutation of the hidden variables, the scale and permutation ambiguity is an inherent feature of the blind source separation problem. Even if the scale and permutation ambiguity is permitted, the blind source separation is still an ill-posed problem without additional assumptions. In particular, [9] showed that no solution exists for Gaussian and temporally i.i.d. sources. Hence, several assumptions have been proposed to make the problem well-posed. Two common such assumptions are as follows [5]: (i) Sources can be temporally i.i.d. but non-Gaussian, which corresponds to ICA methods; and (ii) Sources are non-temporally i.i.d. but they can be Gaussian.

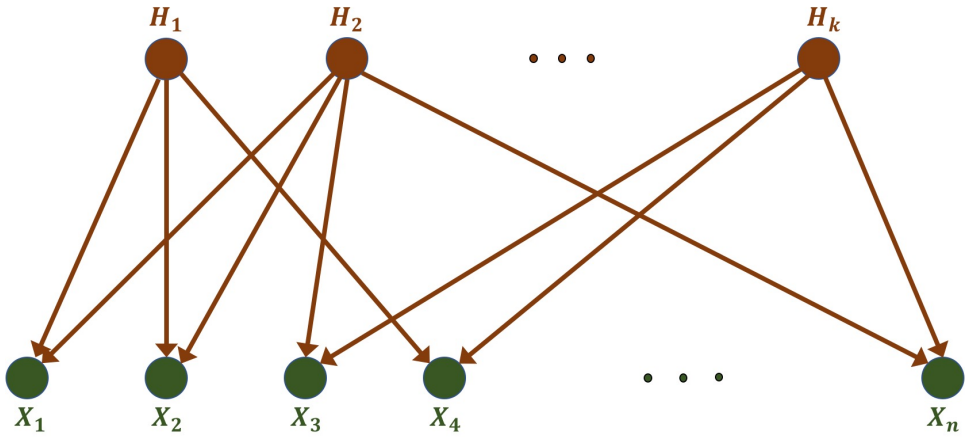


Figure 3.1: Independent Component Analysis (ICA) model.

3.1 Independent Component Analysis

The assumption of non-Gaussian sources that are temporally i.i.d. leads to a class of methods referred to as independent component analysis (ICA), where the unknown sources are jointly independent. More precisely, the ICA model admits the graphical representation in Figure 3.1 and satisfies the following equality:

$$\vec{\mathbf{x}} = \mathbf{A}\vec{\mathbf{h}}, \tag{3.1}$$

where the matrix $\mathbf{A}$ is referred to as the mixing matrix. ICA methods aim at estimating the mixing process to decompose a multivariate random signal into a set of independent non-Gaussian sources. Hence, provided observations over the visible variable $\vec{\mathbf{x}}$, these methods can be applied to the structural equation models in Eqs. (2.1) and (2.2) to recover columns of the mixing matrix $\mathbf{C}$ in Eq. (2.2), which are also referred to as the independent components (ICs). Due to the indeterminacy of permutation and scaling, we can only recover columns of the matrix $\mathbf{C}$ up to the permutation, absolute scale and sign, which we refer to as the PSS indeterminacy for brevity.

While the mixing matrix $\mathbf{C}$ in Eq. (2.2) can be recovered up to the PSS indeterminacy of the

independent components, there is no unique way to decompose it to two matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{C} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{A}$ holds. This implies that the observational data is not enough to identify the mixing and causal matrices $\mathbf{A}$ and $\mathbf{B}$, respectively. Hence, we resort to perform interventions and use the experimental data to make the identification feasible. For an action $\text{do}(X_i = 0 | \forall i \in \mathcal{S})$ on the set $\mathcal{S} \subseteq \{1, \cdots, n\}$ of visible variables, we can use ICA methods to recover the columns of the response matrix $\mathbf{C}_{\mathcal{S}}$ up to the PSS indeterminacy. Our goal is to find an appropriate set of interventions that guarantees identification of the mixing and causal matrices $\mathbf{A}$ and $\mathbf{B}$ despite the inherent indeterminacies in the process.

3.2 Tensor Decomposition and the Method of Moments

Several methods have been proposed to learn the mixing process of an ICA model. Here, we use the method in [3, 8] for guaranteed recovery of the independent components with polynomial sample and computational complexity. This method, which we refer to as tensor method or the method of moments, uses high-order moments of the stochastic vector $\vec{\mathbf{x}}$ to recover columns of $\mathbf{C}$ in Eq. (2.2) up to the PSS indeterminacy. In particular, let $\mathbf{M}_4$ be the $4-$th order cumulant of the zero-mean vector $\vec{\mathbf{x}}$ defined as:

$$\mathbf{M}_4 = \mathbb{E}\left[\vec{\mathbf{x}} \otimes \vec{\mathbf{x}} \otimes \vec{\mathbf{x}} \otimes \vec{\mathbf{x}}\right] - \mathbf{T}, \tag{3.2}$$

where the $4-$th order tensor $\mathbf{T}$ is defined as:

$$[\mathbf{T}]_{i_1,i_2,i_3,i_4} = \mathbb{E}\left[x_{i_1}x_{i_2}\right]\mathbb{E}\left[x_{i_3}x_{i_4}\right] + \mathbb{E}\left[x_{i_1}x_{i_3}\right]\mathbb{E}\left[x_{i_2}x_{i_4}\right] + \mathbb{E}\left[x_{i_1}x_{i_4}\right]\mathbb{E}\left[x_{i_2}x_{i_3}\right], \tag{3.3}$$

for all $1 \leq i_1, i_2, i_3, i_4 \leq n$. If the expected value of the variable x_i is not zero, the term x_i in above equations should be replaced with $x_i - \mathbb{E}\left[x_i\right]$. Using tensor decomposition algorithms, such as the alternating power updates proposed in [3], the $4-$th order moment $\mathbf{M}_4$ in Eq.

(3.2) can be written as:

$$\mathbf{M}_4 = \sum_{i=1}^{k} \kappa_i \vec{\mathbf{c}}^{(i)} \otimes \vec{\mathbf{c}}^{(i)} \otimes \vec{\mathbf{c}}^{(i)} \otimes \vec{\mathbf{c}}^{(i)}, \tag{3.4}$$

where $\vec{\mathbf{c}}^{(i)}$'s are columns of the matrix $\mathbf{C}$ in Eq. (2.2), and κ_i denotes the excess kurtosis defined as:

$$\kappa_i = \mathbb{E}\left[H_i^4\right] - 3, \qquad \forall i \in \{1, \cdots, k\} \tag{3.5}$$

and is a measure of non-Gaussianity. In particular, $\kappa_i = 0$ implies that H_i is a standard normal random variable.

The 4−th order moment $\mathbf{M}_4$ in Eq. (3.2) can be calculated empirically from observations over visible variables, and further decomposed to its rank-one components as in Eq. (3.4) that yields columns of the matrix $\mathbf{C}$. Note that the PSS indeterminacy can be translated to the fact that there is no preferred ordering of the summation terms in Eq. (3.4), and values of the excess kurtosis parameters are unknown. In Section 5, we present an algorithm to alleviate this indeterminacy, which is a crucial aspect of the identification method presented in the next section.

Chapter 4

Causal Effects Identification

In this section, we study the effect of an intervention on the causal graphical model in Figure 2.1, and how the extra information from system's response can be used to recover the mixing and causal matrices $\mathbf{A}$ and $\mathbf{B}$, respectively. Note that the $i-$th row of matrices $\mathbf{A}$ and $\mathbf{B}$ in Eq. (2.1) corresponds to the visible variable X_i, and models the direct causal effect of other variables on the value of X_i. Since $\mathrm{Pa}(X_i)$ is the parent set of the variable X_i in the causal graph, i.e. the set of variables with a directed edge to the variable X_i in Figure 2.1, we have $a_{i,j} \neq 0$ if $H_j \in \mathrm{Pa}(X_i)$, and $b_{i,j} \neq 0$ if $X_j \in \mathrm{Pa}(X_i)$. Hence, one approach to identify the $i-$th row of the mixing and causal matrices is to perform actions on the parent set of the variable X_i and study their effect from the observed experimental data. The intuition for this procedure is that each observed variable takes a value based on the value of its parents in the causal DAG. Thus, to identify one row that corresponds to a particular observed variable, each action rules out a few of the sources from which that observed variable takes its value. Therefore, each action takes a subset of sources and specifies how these sources contribute to value of the observed variable. For the special case where there is no hidden variable in the model, this intuition is also consistent with the Adjustment for Direct Causes [27] which provides a Bayesian perspective on calculating the causal effect based on the

probabilistic properties of the parent set. However, for the SEM-LV model in Figure 2.1 that contains latent variables, not only is it not feasible to perform interventions on the latent variables, but also it is not known a priori which latent variables are in the parent set $\mathrm{Pa}(X_i)$. Therefore, the disentanglement of causal effects in the SEM-LV model demands innovative approaches, as described next.

4.1 Intervention Space and Action Graph

Note that each intervention $\mathrm{do}(X_i = 0 | \forall i \in \mathcal{S})$ corresponds to an action matrix $\mathbf{P}_{\mathcal{S}}$ that sets the value of the visible variable X_i to zero for all $i \in \mathcal{S} \subseteq \{1, \cdots, n\}$. Since there are 2^n such subsets $\mathcal{S}$, the space of all possible interventions $\mathbb{P} = \{\mathbf{P}_{\mathcal{S}} | \mathcal{S} \subseteq \{1, \cdots, n\}\}$ contains 2^n possible actions. As an especial case, the set $\mathcal{S} = \varnothing$ corresponds to the observational data since no action is performed on the visible variables. Also, the set $\mathcal{S} = \{1, \cdots, n\}$ fixes the value of all variables and removes all causal relations in the causal DAG G; thus, no experimental data is available and no inference is possible. However, for any non-empty subset $\mathcal{S} \subsetneqq \{1, \cdots, n\}$, we can infer the response $\mathbf{C}_{\mathcal{S}}$ to the action $\mathbf{P}_{\mathcal{S}}$, up to the PSS indeterminacy of the independent components, using the method of moment in Section 3.2.

For the SEM-LV model in Figure 2.1 that contains n visible and k latent variables, we define the action graph as $\mathbb{G} = (\mathbb{P}, \mathbb{E})$ where the vertices in $\mathbb{P}$ correspond to different action matrices, and $\mathbb{E}$ contains edges between every two distinct vertices in $\mathbb{P}$, i.e. $\mathbb{G}$ is a complete graph. For an edge $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2}) \in \mathbb{E}$ connecting two vertices $\mathbf{P}_{\mathcal{S}_1}$ and $\mathbf{P}_{\mathcal{S}_2}$, let

$$\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2} = \{i \mid \mathbf{P}_{\mathcal{S}_1}(i,i) \neq \mathbf{P}_{\mathcal{S}_2}(i,i),\ i \in \{1, \cdots, n\}\} \tag{4.1}$$

be the set of all diagonal indices where two action matrices are not equal. The set $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}$, assigned to the edge connecting $\mathbf{P}_{\mathcal{S}_1}$ to $\mathbf{P}_{\mathcal{S}_2}$ in the action graph $\mathbb{G}$, is a key factor in deter-

mining the set of interventions that provides identification guarantee. The definition in Eq. (4.1) implies that $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2} = \mathcal{R}_{\mathcal{S}_2,\mathcal{S}_1}$. Moreover, the set $\mathcal{R}$ satisfies the following two lemmas.

Lemma 1. *For the SEM-LV model in Figure 2.1 with n visible and k hidden variables, let $\mathbb{P}$ be the set of all possible actions, and $\mathbb{G}$ be its corresponding action graph. For any three action matrices such as $\mathbf{P}_{\mathcal{S}_1}$, $\mathbf{P}_{\mathcal{S}_2}$ and $\mathbf{P}_{\mathcal{S}_3}$ corresponding to $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3 \in \{1, \cdots, n\}$, we have:*

$$|\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}| + |\mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}| \geq |\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_3}|\,, \tag{4.2}$$

where $|\mathcal{R}|$ denotes the cardinality of the set $\mathcal{R}$. Moreover, the equality holds if and only if two sets $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}$ and $\mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}$ are disjoint. In other words, the cardinality of the set $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}$ defined in Eq. (4.1) satisfies the triangle inequality in the action graph $\mathbb{G}$.

The proof is provided in Appendix A.

Lemma 2. *For the SEM-LV model in Figure 2.1 with n visible and k hidden variables, let $\mathbb{P}$ be the set of all possible actions, and $\mathbb{G}$ be its corresponding action graph. Let $\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_l \subseteq \{1, \cdots, n\}$ be l distinct sets so that edges $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2}), (\mathbf{P}_{\mathcal{S}_2}, \mathbf{P}_{\mathcal{S}_3}), \cdots, (\mathbf{P}_{\mathcal{S}_l}, \mathbf{P}_{\mathcal{S}_1})$ constitute a simple loop in the action graph $\mathbb{G}$. Then, we have:*

$$\mathcal{R}_{\mathcal{S}_l,\mathcal{S}_1} \subseteq \mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2} \cup \mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3} \cup \cdots \cup \mathcal{R}_{\mathcal{S}_{l-1},\mathcal{S}_l}, \tag{4.3}$$

and $\mathcal{R}_{\mathcal{S}_l,\mathcal{S}_1}$ is determined uniquely in terms of sets $\mathcal{R}$ corresponding to edges $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2})$, $(\mathbf{P}_{\mathcal{S}_2}, \mathbf{P}_{\mathcal{S}_3}), \cdots, \big(\mathbf{P}_{\mathcal{S}_{l-1}}, \mathbf{P}_{\mathcal{S}_l}\big)$.

The proof is provided in Appendix B.

4.2 Recovery of the Observed Causal Matrix

Here, we present a method to disentangle observed causal effects from latent confounders, i.e. to recover the mixing and causal matrices $\mathbf{A}$ and $\mathbf{B}$ in Eq. (2.1), respectively. First, we uniquely identify the causal matrix $\mathbf{B}$ one row at a time, then we recover the mixing matrix $\mathbf{A}$ up to the PSS indeterminacy of its columns.

For an action $\text{do}\left(X_i = 0 \mid \forall i \in \mathcal{S}_1\right)$ that corresponds to the action matrix $\mathbf{P}_{\mathcal{S}_1}$, we can rewrite Eq. (2.7) as:

$$\mathbf{P}_{\mathcal{S}_1}\mathbf{A} = \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B}\right)\mathbf{C}_{\mathcal{S}_1}. \tag{4.4}$$

Similarly, we have the following equation for the action $\text{do}\left(X_i = 0 \mid \forall i \in \mathcal{S}_2\right)$:

$$\mathbf{P}_{\mathcal{S}_2}\mathbf{A} = \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B}\right)\mathbf{C}_{\mathcal{S}_2}. \tag{4.5}$$

Note that matrices $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ in Eqs. (4.4) and (4.5) are recovered, using the method of moments in Section 3.2, only up to the absolute scale, sign and permutation of their columns. However, for the rest of this section, we assume that we know the correspondence between columns of the response matrices $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ in terms of scale, sign and permutation. Later in Section 5, we present a novel method to infer this correspondence, which is a crucial aspect of the identification process.

Our strategy to identify the causal matrix $\mathbf{B}$ is to recover one row, say $j \in \{1, \cdots, n\}$, at a time. As mentioned in the beginning of Section 4, in order to identify the $j-$th row of matrices $\mathbf{A}$ and $\mathbf{B}$ which corresponds to the visible variable X_j, we need to perform actions on other visible variables in the SEM-LV model to study their effect on the value of X_j, and no action should be performed on the visible variable X_j itself. This implies that for actions $\mathbf{P}_{\mathcal{S}_1}$ and $\mathbf{P}_{\mathcal{S}_2}$ performed to recover the $j-$th row of the matrix $\mathbf{B}$, we have $j \notin \mathcal{S}_1$ and $j \notin \mathcal{S}_2$.

Therefore, if $\mathbf{Z} = \mathbf{I} - \mathbf{B}$, then the $j-$th row of both matrices $(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B})$ and $(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B})$ are equal to the $j-$th row of $\mathbf{Z}$. Hence, from Eqs. (4.4) and (4.5) we have the following equality:

$$\langle \mathbf{Z}(j,:)\ ,\ \mathbf{C}_{\mathcal{S}_1}(:,l)\rangle = \langle \mathbf{Z}(j,:)\ ,\ \mathbf{C}_{\mathcal{S}_2}(:,l)\rangle \qquad\qquad \forall l \in \{1,\cdots,k\}, \tag{4.6}$$

where $\langle .,. \rangle$ denotes the inner product. Note that Eq. (4.6) holds since both sides are equal to the element $a_{j,l}$ of the mixing matrix $\mathbf{A}$. We can rewrite Eq. (4.6) as:

$$\left(\mathbf{C}_{\mathcal{S}_1} - \mathbf{C}_{\mathcal{S}_2}\right)^T \mathbf{Z}(j,:)^T = \vec{\mathbf{0}}, \tag{4.7}$$

which yields k linear equations for identification of the $j-$th row of $\mathbf{Z}$. Note that the pair of actions used to obtain k linear equations in Eq. (4.7) corresponds to the edge $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2})$ in the action graph $\mathbb{G}$. In other words, for any two sets $\mathcal{S}_1, \mathcal{S}_2 \subseteq \{1,\cdots,n\}$ such that $j \notin \mathcal{S}_1$ and $j \notin \mathcal{S}_2$, the edge $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2})$ in the action graph $\mathbb{G}$ leads to k linear equations as in Eq. (4.7). In order to recover the $j-$th row of $\mathbf{Z}$, we need to determine the number of equations in Eq. (4.7) that are linearly independent.

Theorem 4.1. *For the SEM-LV model in Figure 2.1 with n visible and k hidden variables, let $\mathbb{P}$ be the set of all possible actions, and $\mathbb{G}$ be its corresponding action graph. Let $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ be responses of the SEM-LV model to two action matrices $\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}$ corresponding to the edge $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2})$ in the action graph $\mathbb{G}$, respectively. Then, we have:*

$$\operatorname{rank}\left(\mathbf{C}_{\mathcal{S}_1} - \mathbf{C}_{\mathcal{S}_2}\right) = \min\left(\left|\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}\right|, k\right). \tag{4.8}$$

The proof is provided in Appendix C.

Theorem 4.1 provides a heuristic on the number of linearly independent equations in Eq. (4.7) that can be used to identify the $j-$th row of $\mathbf{Z}$. For an index $i \in \{1,\cdots,n\}$, let $\mathbb{P}_i \subseteq \mathbb{P}$

be the intervention subspace such that $\mathbf{P}_{\mathcal{S}} \in \mathbb{P}_i$ if and only if $i \notin \mathcal{S}$. Therefore, in order to identify the $j-$th row of $\mathbf{Z}$, the pair of action matrices $\mathbf{P}_{\mathcal{S}_1}$ and $\mathbf{P}_{\mathcal{S}_2}$, which corresponds to the edge $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2})$ in $\mathbb{G}$, should be chosen such that $\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_j$. The information retrieved from responses $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ is used to identify the $j-$th row of $\mathbf{Z}$ in the form of a set of linear equations in Eq. (4.7) where $\min\left(\left|\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}\right|, k\right)$ of them are linearly independent according to Theorem 4.1. However, the information provided by a pair of actions in $\mathbb{P}_j$ may not be enough to identify $\mathbf{Z}(j,:)$, i.e. the number of unknown elements in the $j-$th row of $\mathbf{Z}$ can be larger than $\min\left(\left|\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}\right|, k\right)$. Thus, we need to consider several edges of $\mathbb{G}$ with vertices in $\mathbb{P}_j$, and study the number of linearly independent equations that can be obtained from their collective set of equations.

Theorem 4.2. *For the SEM-LV model in Figure 2.1 with n visible and k hidden variables, let $\mathbb{P}$ be the set of all possible actions, and $\mathbb{G}$ be its corresponding action graph. Let $\mathcal{J} = \left\{j_1, \cdots, j_{|\mathcal{J}|}\right\} \subseteq \{1, \cdots, n\}$ be a subset of indices sorted in ascending order, and define $\mathcal{S}_j = \{j\}$ for each $j \in \mathcal{J}$. Consider the collection of edges $\left(\mathbf{P}_{\mathcal{S}_{j_1}}, \mathbf{P}_{\mathcal{S}_{j_2}}\right), \left(\mathbf{P}_{\mathcal{S}_{j_2}}, \mathbf{P}_{\mathcal{S}_{j_3}}\right), \cdots, \left(\mathbf{P}_{\mathcal{S}_{j_{|\mathcal{J}|-1}}}, \mathbf{P}_{\mathcal{S}_{j_{|\mathcal{J}|}}}\right)$ in $\mathbb{G}$, and define $\mathbf{W} \in \mathbb{R}^{n \times (k|\mathcal{J}|-k)}$ to be the column-wise concatenation of matrices $\left(\mathbf{C}_{\mathcal{S}_{j_i}} - \mathbf{C}_{\mathcal{S}_{j_{i+1}}}\right)$ for $1 \leq i < |\mathcal{J}|$. Now, we have:*

$$\operatorname{rank}(\mathbf{W}) = |\mathcal{J}|. \tag{4.9}$$

The proof is provided in Appendix D.

In order to identify the elements in row j of $\mathbf{Z}$, the set $\mathcal{J}$ in Theorem 4.2 should be a subset of $\{1, \cdots, n\} \setminus \{j\}$, and Theorem 4.2 guarantees a certain number of linearly independent equations for the identification process. In particular, if we choose $\mathcal{J} = \{1, \cdots, n\} \setminus \{j\}$, then Theorem 4.2 provides $n-1$ linearly independent equations for recovery of the row j in $\mathbf{Z}$. Since no variable is a cause of itself, we have $\mathbf{Z}(j,j) = 1$, i.e. there are at most $n-1$ unknown elements in the $j-$th row of $\mathbf{Z}$, and these elements can be identified using the aforementioned

$n-1$ linearly independent equations. Note that by choosing $\mathcal{J} = \{1, \cdots, n\} \setminus \{j\}$, we are treating all elements of Row j, except for the diagonal element $\mathbf{Z}(j,j)$, to be unknown and let the model to determine their values based on the experimental data. However, alternative approaches can be used to incorporate properties of these matrices. In particular, the matrix $\mathbf{Z}$ is lower triangular with diagonal entries equal to 1, i.e. there are at most $j-1$ unknown elements in Row j of $\mathbf{Z}$, and we do not necessarily need to perform intervention on all variables, except X_j, to recover Row j of $\mathbf{Z}$. Intuitively, Row j of $\mathbf{Z}$ includes information on how other variables affect the value of X_j; therefore, we only need to conduct actions on ancestors of X_j in the graphical representation of SEM-LV model, i.e. $X_1, \cdots, X_{j-1}$, since these variables are the ones contributing to the value of X_j. Thus, we can choose $\mathcal{J} = \{1, \cdots, j-1\}$ which provides $j-1$ linearly independent equations, according to Theorem 4.2, for recovery of the $j-1$ unknown elements in Row j of $\mathbf{Z}$.

While Theorem 4.2 provides enough theoretical guarantee to design an identification algorithm, it can be generalized to an arbitrary collection of edges in $\mathbb{G}$, which we leave below as a conjecture since its proof and use is beyond the scope of this work.

Conjecture 1. *For the SEM-LV model in Figure 2.1 with n visible and k hidden variables, let the mixing and causal matrices $\mathbf{A}$ and $\mathbf{B}$ be uniform random matrices. For the collection of edges $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2}), (\mathbf{P}_{\mathcal{S}_3}, \mathbf{P}_{\mathcal{S}_4}), \cdots, \left(\mathbf{P}_{\mathcal{S}_{2l-1}}, \mathbf{P}_{\mathcal{S}_{2l}}\right)$ in $\mathbb{G}$, the matrix $\mathbf{W} \in \mathbb{R}^{n \times (kl)}$ is formed by column-wise concatenation of matrices $\left(\mathbf{C}_{\mathcal{S}_{2i-1}} - \mathbf{C}_{\mathcal{S}_{2i}}\right)$, for all $1 \leq i \leq l$. We have:*

$$L \leq \operatorname{rank}(\mathbf{W}) \leq U, \qquad \textit{with probability of 1} \tag{4.10}$$

where

$$L = \min \left[\sum_{i=1}^{l} \min \left(\left| \mathcal{R}_{\mathcal{S}_{2i-1},\mathcal{S}_{2i}} \right|,\ k \right)\ ,\ \left| \bigcup_{i\in[l]} \mathcal{R}_{\mathcal{S}_{2i-1},\mathcal{S}_{2i}} \right| \right], \tag{4.11}$$

$$U = \min \left[\max \left[\sum_{i=1}^{l} \min \left(\left| \mathcal{R}_{\mathcal{S}_{2i-1},\mathcal{S}_{2i}} \right|,\ k \right)\ ,\ \left| \bigcup_{i\in[l]} \mathcal{R}_{\mathcal{S}_{2i-1},\mathcal{S}_{2i}} \right| \right],\ n - |\mathcal{V}| \right], \tag{4.12}$$

and $\mathcal{V} \subseteq \{1, \cdots, n\}$ *is defined as* $\mathcal{V} = \{v \mid \mathbf{P}_{\mathcal{S}_r}(v,v) = \mathbf{P}_{\mathcal{S}_t}(v,v),\ \forall r,t \in \{1,\cdots,2l\}\}$.

Intuitively, Conjecture 1 provides a lower and upper bound on the number of linearly independent equations obtained from a collection of edges in $\mathbb{G}$ with vertices in $\mathbb{P}_j$ that is needed to identify the $j-$th row of $\mathbf{Z}$. Since for all $\mathbf{P}_{\mathcal{S}} \in \mathbb{P}_j$ we have $\mathbf{P}_{\mathcal{S}}(j,j) = 1$, it immediately follows that $j \in \mathcal{V}$. Therefore, Eq. (4.12) implies that $U \leq n-1$, i.e. regardless of the number of edges in $\mathbb{G}$ that we consider for identification of the elements in $\mathbf{Z}(j,:)$, the maximum possible number of linearly independent equations does not exceed $n-1$. However, since no variable can be a cause of itself, all diagonal entries of $\mathbf{Z}$ are equal to 1, i.e. the number of unknown elements in each row of $\mathbf{Z}$ is at most $n-1$ as well.

While Conjecture 1 can directly lead to many heuristics to select the collection of edges in $\mathbb{G}$ that is needed for identification, practical limitations demand to perform as few interventions as possible. In order to reduce the number of actions, the collection of edges in $\mathbb{G}$ that are chosen to identify $\mathbf{Z}$ should be as connected as possible. For instance, two edges $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2})$ and $(\mathbf{P}_{\mathcal{S}_3}, \mathbf{P}_{\mathcal{S}_4})$ corresponding to distinct sets $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{S}_4$, require four actions to take place; however, two edges $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2})$ and $(\mathbf{P}_{\mathcal{S}_2}, \mathbf{P}_{\mathcal{S}_3})$ only require three interventions.

Conjecture 2. *For the SEM-LV model in Figure 2.1 with n visible and k hidden variables, let $\mathbb{P}$ be the set of all possible actions, and $\mathbb{G}$ be its corresponding action graph. Consider the edges $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2}), (\mathbf{P}_{\mathcal{S}_2}, \mathbf{P}_{\mathcal{S}_3}), \cdots, \left(\mathbf{P}_{\mathcal{S}_{l-1}}, \mathbf{P}_{\mathcal{S}_l}\right), (\mathbf{P}_{\mathcal{S}_l}, \mathbf{P}_{\mathcal{S}_1})$ that constitute a simple loop in $\mathbb{G}$ for distinct subsets $\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_l \subseteq \{1, \cdots, n\}$. Let $\mathbf{W} \in \mathbb{R}^{n\times(kl-k)}$ be an $n \times (kl-k)$ matrix*

form by column-wise concatenation of matrices $\left(\mathbf{C}_{\mathcal{S}_i} - \mathbf{C}_{\mathcal{S}_{i+1}}\right)$ *for* $1 \leq i \leq l-1$*, and let* $\overline{\mathbf{W}} \in \mathbb{R}^{n \times (kl)}$ *be an* $n \times (kl)$ *matrix that is formed by column-wise concatenation of* $\mathbf{W}$ *and* $(\mathbf{C}_{\mathcal{S}_l} - \mathbf{C}_{\mathcal{S}_1})$*. Then, we have:*

$$\operatorname{rank}\left(\overline{\mathbf{W}}\right) = \operatorname{rank}\left(\mathbf{W}\right). \tag{4.13}$$

Conjecture 2 indicates that while it is desirable for our collection of edges to be as connected as possible, having a loop in our selected edges does not increase the number of linearly independent equations. This implication is also reinforced by Lemma 2 and Eq. (4.3) since the set $\mathcal{R}_{\mathcal{S}_l,\mathcal{S}_1}$ is a subset of the union of sets $\mathcal{R}$ corresponding to other edges, and it does not change the second term in Eq. (4.11). Therefore, in order to recover each row of $\mathbf{Z}$, the optimal set of edges that lead to minimum number of interventions should be as connected as possible but does not include any loop, i.e. it should constitute a tree within $\mathbb{G}$.

4.3 Recovery of the Mixing Matrix

Guided by Theorem 4.2, identification methods can be developed to recover the causal matrix $\mathbf{B}$ in the SEM-LV model. In Section 5.2, we present one such method that uses only n atomic interventions to identify the matrix $\mathbf{B}$. Here, we assume that the causal matrix $\mathbf{B}$ is recovered, and aim to identify the mixing matrix $\mathbf{A}$ in Eq. (2.1). Note that although the method presented in Section 5 can find the correspondence between columns of two response matrices, the true permutation and dilation of one particular response matrix is always unknown since H_i variables are hidden by nature. Therefore, while the causal matrix $\mathbf{B}$ can be fully recovered, the mixing matrix $\mathbf{A}$ can only be identified up to the permutation, absolute scale and sign of its columns.

Once the causal matrix $\mathbf{B}$ is identified, there is no unique way to recover the mixing matrix

$\mathbf{A}$. Here, we consider two scenarios depending on whether or not observational data is available for the variable $\vec{\mathbf{x}}$.

In the first scenario where the observational data on $\vec{\mathbf{x}}$ is available, the new mixing matrix $\mathbf{C}$ in Eq. (2.2) can be recovered up to the PSS indeterminacy of its columns using the method of moments in Section 3. Then, the mixing matrix $\mathbf{A}$ can be recovered as:

$$\mathbf{A} = (\mathbf{I} - \mathbf{B})\,\mathbf{C}. \tag{4.14}$$

In the second scenario where no observational data exists on $\vec{\mathbf{x}}$, we show that the available information, provided by actions performed to identify $\mathbf{B}$, is enough to recover $\mathbf{A}$ and no further intervention is required. Note that for all actions such as $\mathbf{P}_{\mathcal{S}}$ that were performed to identify Row j of $\mathbf{B}$, we have $\mathbf{P}_{\mathcal{S}}(j,j) = 1$. Therefore, among all actions that were performed to identify all rows of $\mathbf{B}$, there exists a collection of actions such as $\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2}, \cdots, \mathbf{P}_{\mathcal{S}_m}$ that for $\overline{\mathbf{P}} = \sum_{t=1}^{m} \mathbf{P}_{\mathcal{S}_t}$ we have $\overline{\mathbf{P}}(i,i) \geq 1$ for all $i \in \{1, \cdots, n\}$. Now, the mixing matrix $\mathbf{A}$ can be recovered as:

$$\mathbf{A} = \overline{\mathbf{P}}^{-1} \sum_{i=1}^{m} (\mathbf{I} - \mathbf{P}_{\mathcal{S}_i}\mathbf{B})\,\mathbf{C}_{\mathcal{S}_i}. \tag{4.15}$$

Eq. (4.15) comes from the fact that if $\overline{\mathbf{P}}(i,i) > 1$ for a particular index i, then the summation of terms in right-hand-side of Eq. (4.15) results in the $i-$th row of $\mathbf{A}$ to be counted more than once. Hence, the correction factor $\overline{\mathbf{P}}^{-1}$ is used to make the required adjustment.

Chapter 5

Elimination of PSS Indeterminacy

Theorem 4.1 indicates that any pair of action matrices, such as $\mathbf{P}_{\mathcal{S}_1}$ and $\mathbf{P}_{\mathcal{S}_2}$, would yield $\min\left(\left|\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}\right|, k\right)$ number of linearly independent equations for identification of row j of the causal matrix $\mathbf{B}$ provided that $j \notin \mathcal{S}_1 \cup \mathcal{S}_2$. However, as mentioned in Section 4, this statement holds only if the PSS indeterminacy between columns of the response matrices $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ is eliminated. In this section, we provide subroutines to eliminate this indeterminacy based on the selected actions. Then, we present a method accordingly that uses atomic actions to render matrices $\mathbf{A}$ and $\mathbf{B}$ identifiable.

5.1 Independent Component Matching

Based on the set of intervened variables, the intervention space $\mathbb{P} = \{\mathbf{P}_{\mathcal{S}} \mid \mathcal{S} \subseteq \{1, \cdots, n\}\}$ can be partitioned into three disjoint subsets as $\mathbb{P} = \mathbb{P}_\alpha \cup \mathbb{P}_\beta \cup \mathbb{P}_\gamma$ where:

$$\mathbb{P}_\alpha = \{\mathbf{P}_\mathcal{S} \mid \mathcal{S} \subseteq \{3, \cdots, n\}\}, \tag{5.1}$$

$$\mathbb{P}_\beta = \{\mathbf{P}_\mathcal{S} \mid \mathcal{S} \subseteq \{2, \cdots, n\} \ , \ 2 \in \mathcal{S}\}, \tag{5.2}$$

$$\mathbb{P}_\gamma = \{\mathbf{P}_\mathcal{S} \mid \mathcal{S} \subseteq \{1, \cdots, n\} \ , \ 1 \in \mathcal{S}\}. \tag{5.3}$$

In Section 5.2, we present an identification algorithm that uses actions in $\mathbb{P}_\alpha \cup \mathbb{P}_\beta$ to recover the causal matrix $\mathbf{B}$ except for its first column. Then, pairs of actions such as $\mathbf{P}_{\mathcal{S}_1} \in \mathbb{P}_\alpha \cup \mathbb{P}_\beta$ and $\mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\gamma$ are utilized to recover the first column of $\mathbf{B}$. In what follows, we explain in details how the correspondence between columns of two response matrices can be found for three different scenarios of a pair of actions such as $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2})$.

5.1.1 Matching Subroutine 1

In the first scenario, we consider the case where $\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\alpha$, and explain how to find the correspondence between independent components of their respective response matrices. Let $\mathbf{M}_4^{\mathcal{S}_1}$ be the $4-$th order tensor, defined as in Eq. (3.2), that is calculated from the experimental data corresponding to the action matrix $\mathbf{P}_{\mathcal{S}_1}$. The $4-$th order tensor $\mathbf{M}_4^{\mathcal{S}_2}$ is defined similarly. Now, using the method of moments in Section 3.2, we have:

$$\mathbf{M}_4^{\mathcal{S}_1} = \sum_{i=1}^{k} \kappa_i \vec{\mathbf{v}}^{(i)} \otimes \vec{\mathbf{v}}^{(i)} \otimes \vec{\mathbf{v}}^{(i)} \otimes \vec{\mathbf{v}}^{(i)}, \tag{5.4}$$

$$\mathbf{M}_4^{\mathcal{S}_2} = \sum_{i=1}^{k} \eta_i \vec{\mathbf{w}}^{(i)} \otimes \vec{\mathbf{w}}^{(i)} \otimes \vec{\mathbf{w}}^{(i)} \otimes \vec{\mathbf{w}}^{(i)}, \tag{5.5}$$

therefore, the response matrices $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ can be written as:

$$\mathbf{C}_{\mathcal{S}_1} = \left[\vec{\mathbf{v}}^{(1)} \mid \vec{\mathbf{v}}^{(2)} \mid \cdots \mid \vec{\mathbf{v}}^{(k)}\right], \tag{5.6}$$

$$\mathbf{C}_{\mathcal{S}_2} = \left[\vec{\mathbf{w}}^{(1)} \mid \vec{\mathbf{w}}^{(2)} \mid \cdots \mid \vec{\mathbf{w}}^{(k)}\right], \tag{5.7}$$

where the permutation, absolute scale and sign of the columns in Eqs. (5.6) and (5.7) are unknown. The following Lemma is the key to find the correspondence between columns of $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$.

Lemma 3. *For the SEM-LV model in Figure 2.1 with n visible and k hidden variables, let $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ be response matrices to actions $\mathbf{P}_{\mathcal{S}_1}$ and $\mathbf{P}_{\mathcal{S}_2}$ in $\mathbb{P}$, respectively. Then, the first $\min(\mathcal{S}_1 \cup \mathcal{S}_2) - 1$ rows of $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ are the same, i.e. we have:*

$$\mathbf{C}_{\mathcal{S}_1}(1 : \min(\mathcal{S}_1 \cup \mathcal{S}_2) - 1\ ,\ \ :) = \mathbf{C}_{\mathcal{S}_2}(1 : \min(\mathcal{S}_1 \cup \mathcal{S}_2) - 1\ ,\ \ :). \tag{5.8}$$

The proof is provided in Appendix E.

Let $\widetilde{\mathbf{C}}_{\mathcal{S}_1} = \left[\vec{\widetilde{\mathbf{v}}}^{(1)} \middle| \vec{\widetilde{\mathbf{v}}}^{(2)} \middle| \cdots \middle| \vec{\widetilde{\mathbf{v}}}^{(k)}\right]$ be the submatrix of $\mathbf{C}_{\mathcal{S}_1}$ in Eq. (5.6) restricted to its first $\min(\mathcal{S}_1 \cup \mathcal{S}_2) - 1$ rows. Similarly, let $\widetilde{\mathbf{C}}_{\mathcal{S}_2} = \left[\vec{\widetilde{\mathbf{w}}}^{(1)} \middle| \vec{\widetilde{\mathbf{w}}}^{(2)} \middle| \cdots \middle| \vec{\widetilde{\mathbf{w}}}^{(k)}\right]$ be the submatrix of $\mathbf{C}_{\mathcal{S}_2}$ in Eq. (5.7) restricted to its first $\min(\mathcal{S}_1 \cup \mathcal{S}_2) - 1$ rows. Lemma 3 implies that columns of the matrix $\widetilde{\mathbf{C}}_{\mathcal{S}_1}$ can be obtained by a proper rearrangement and dilation of columns of the matrix $\widetilde{\mathbf{C}}_{\mathcal{S}_2}$. From Assumption 1 it can be readily inferred that every two columns of $\widetilde{\mathbf{C}}_{\mathcal{S}_1}$ are linearly independent. Similar observation holds for every pair of columns in $\widetilde{\mathbf{C}}_{\mathcal{S}_2}$. Hence, if we constitute the following matrix:

$$\mathbf{\Theta}_{\widetilde{\mathbf{C}}_{\mathcal{S}_1},\widetilde{\mathbf{C}}_{\mathcal{S}_2}} = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \ldots & \theta_{1,k} \\ \theta_{2,1} & \theta_{2,2} & \ldots & \theta_{2,k} \\ & & \vdots & \\ \theta_{k,1} & \theta_{k,2} & \ldots & \theta_{k,k} \end{bmatrix}, \tag{5.9}$$

such that:

$$\theta_{i,j} = \arccos\left(\frac{\left\langle \vec{\widetilde{\mathbf{v}}}^{(i)} \ , \ \vec{\widetilde{\mathbf{w}}}^{(j)} \right\rangle}{\left\| \vec{\widetilde{\mathbf{v}}}^{(i)} \right\| \times \left\| \vec{\widetilde{\mathbf{w}}}^{(j)} \right\|}\right), \tag{5.10}$$

then, exactly one element in each row and column of the matrix $\mathbf{\Theta}$ is either 0 or π. Now, if for two indices i and j we have $\theta_{i,j} = 0$ or π, then Column i of the matrix $\mathbf{C}_{\mathcal{S}_1}$ corresponds to Column j of the matrix $\mathbf{C}_{\mathcal{S}_2}$. As for the proper sign between these two columns, we keep their current sign in case $\theta_{i,j} = 0$; otherwise, $\theta_{i,j} = \pi$ indicates that the sign for either Column i of $\mathbf{C}_{\mathcal{S}_1}$ or Column j of $\mathbf{C}_{\mathcal{S}_2}$ should be flipped. In what follows, we describe how to eliminate the indeterminacy in absolute scaling between these two columns.

The excess kurtosis associated with each independent component can be uniquely determined based on the characteristics of its corresponding latent variable, as in Eq. (3.5). Since Column i of $\mathbf{C}_{\mathcal{S}_1}$ corresponds to Column j of $\mathbf{C}_{\mathcal{S}_2}$, they are associated with the same hidden variable; hence, their excess kurtosis κ_i and η_j should be equal. Note that the indeterminacy in absolute scaling of the column $\vec{\mathbf{v}}^{(i)}$ in the response matrix $\mathbf{C}_{\mathcal{S}_1}$ is due to the unknown value of the excess kurtosis κ_i. Another way to interpret this is that we can multiply $\vec{\mathbf{v}}^{(i)}$ by a positive constant $c \in \mathbb{R}^+$, and divide its corresponding excess kurtosis κ_i by c^4 such that the $4-$th order cumulant $\mathbf{M}_4^{\mathcal{S}_1}$ in Eq. (5.4) does not change. Thus, we multiply the excess kurtosis κ_i by $\frac{\eta_j}{\kappa_i}$, and divide $\vec{\mathbf{v}}^{(i)}$ by $\sqrt[4]{\frac{\eta_j}{\kappa_i}}$ for a proper scaling between these two columns. Alternatively, we can multiply the excess kurtosis η_j by $\frac{\kappa_i}{\eta_j}$, and divide $\vec{\mathbf{w}}^{(j)}$ by $\sqrt[4]{\frac{\kappa_i}{\eta_j}}$. This procedure is outlined in Algorithm 1.

Note that since variables H_i are hidden by nature, the true scale, sign and permutation of these columns can never be recovered; however, in order to make Eqs. (4.6) and (4.7) valid, we only need to find the right correspondence between columns of $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ to rearrange and rescale these columns accordingly. The intuition for the aforementioned strategy is that once the right correspondence is found, any change in scale, sign and ordering of these

columns translates into changing the order of equations in Eq. (4.7) and multiplying both sides of each equation by a non-zero constant, which does not change the set of linear equations in Eq. (4.7). In other words, if the column $\vec{\mathbf{v}}^{(i)}$ in $\mathbf{C}_{\mathcal{S}_1}$ corresponds to the column $\vec{\mathbf{w}}^{(j)}$ in $\mathbf{C}_{\mathcal{S}_2}$, and they both correspond to the hidden variable H_l, our strategy claims that although the variable H_l is unknown to us, two columns $\vec{\mathbf{v}}^{(i)}$ and $\vec{\mathbf{w}}^{(j)}$ should correspond to each other because they both correspond to an unknown variable. The rationale for this argument is similar to Eq. (4.6) where the equality holds because both sides of the equation are equal to an unknown element in $\mathbf{A}$.

Algorithm 1: Elimination of PSS indeterminacy for pair of actions in $\mathbb{P}_\alpha$

Result: Correspondence between columns of the response matrices;
Input: The SEM-LV model in Fig. 2.1 with n visible and k hidden variables, experimental data for action matrices $\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\alpha$;
— Apply the method of moments in Section 3.2 on the experimental data for $\mathbf{P}_{\mathcal{S}_1}$ to obtain columns of the response $\mathbf{C}_{\mathcal{S}_1}$ up to PSS as in Eqs. (5.4) and (5.6);
— Apply the method of moments in Section 3.2 on the experimental data for $\mathbf{P}_{\mathcal{S}_2}$ to obtain columns of the response $\mathbf{C}_{\mathcal{S}_2}$ up to PSS as in Eqs. (5.5) and (5.7);
— Let $\widetilde{\mathbf{C}}_{\mathcal{S}_1}$ and $\widetilde{\mathbf{C}}_{\mathcal{S}_2}$ be submatrices of $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$, respectively, when restricted to their first $\min(\mathcal{S}_1 \cup \mathcal{S}_2) - 1$ rows;
— Obtain the matrix $\mathbf{\Theta}_{\widetilde{\mathbf{C}}_{\mathcal{S}_1}, \widetilde{\mathbf{C}}_{\mathcal{S}_2}}$ as in Eqs. (5.9) and (5.10);
for $i = 1, \cdots, k$ **do**
  for $j = 1, \cdots, k$ **do**
    if $\theta_{i,j} = 0$;
      • Column i of $\mathbf{C}_{\mathcal{S}_1}$ corresponds to Column j of $\mathbf{C}_{\mathcal{S}_2}$;
      • Divide the $j-$th column of $\mathbf{C}_{\mathcal{S}_2}$ by $\sqrt[4]{\frac{\kappa_i}{\eta_j}}$;
      • Keep the current sign of these two columns;
    if $\theta_{i,j} = \pi$;
      • Column i of $\mathbf{C}_{\mathcal{S}_1}$ corresponds to Column j of $\mathbf{C}_{\mathcal{S}_2}$;
      • Divide the $j-$th column of $\mathbf{C}_{\mathcal{S}_2}$ by $\sqrt[4]{\frac{\kappa_i}{\eta_j}}$;
      • Flip the sign for one of these two columns;
  end
end

5.1.2 Matching Subroutine 2

In the second scenario, we consider the case where $\mathbf{P}_{\mathcal{S}_1} \in \mathbb{P}_\alpha \cup \mathbb{P}_\beta$ and $\mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\beta$. Since $2 \in \mathcal{S}_2$, Lemma 3 indicates that only the first row of the response matrices $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ are the same; therefore, the matrix $\mathbf{\Theta}$ in Eq. (5.9) cannot be constructed. However, if Column i of $\mathbf{C}_{\mathcal{S}_1}$ in Eq. (5.6) corresponds to Column j of $\mathbf{C}_{\mathcal{S}_2}$ in Eq. (5.7), then we have:

$$\mathbf{C}_{\mathcal{S}_1}(1,i) = \mathbf{C}_{\mathcal{S}_2}(1,j), \tag{5.11}$$

$$\kappa_i = \eta_j, \tag{5.12}$$

where Eq. (5.11) follows from Lemma 3, and Eq. (5.12) comes from the fact that both columns correspond to the same hidden variable. Therefore, if the matrix $\mathbf{F}$ is defined as:

$$\mathbf{F} = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,k} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,k} \\ & & \vdots & \\ f_{k,1} & f_{k,2} & \cdots & f_{k,k} \end{bmatrix}, \tag{5.13}$$

such that:

$$f_{i,j} = \frac{\mathbf{C}_{\mathcal{S}_2}(1,j)}{\mathbf{C}_{\mathcal{S}_1}(1,i)}, \tag{5.14}$$

then, for each row of the matrix $\mathbf{F}$ in Eq. (5.13), say i, there is a unique column index $j \in \{1, \cdots, k\}$ such that:

$$f_{i,j} = \pm \sqrt[4]{\frac{\kappa_i}{\eta_j}}. \tag{5.15}$$

Note that this uniqueness comes from Assumption 2 and the fact that the first row of $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ are equal to the first row of $\mathbf{A}$ since the matrix $\mathbf{B}$ is strictly lower triangular. Thus,

if Eq. (5.15) holds for two indices i and j, then column i of $\mathbf{C}_{\mathcal{S}_1}$ corresponds to column j of $\mathbf{C}_{\mathcal{S}_2}$. Moreover, we keep the sign of these two columns in case $f_{i,j} = \sqrt[4]{\frac{\kappa_i}{\eta_j}}$; otherwise, we flip the sign of either Column i of $\mathbf{C}_{\mathcal{S}_1}$ or Column j of $\mathbf{C}_{\mathcal{S}_2}$ if $f_{i,j} = -\sqrt[4]{\frac{\kappa_i}{\eta_j}}$ holds. Finally, we divide Column j of $\mathbf{C}_{\mathcal{S}_2}$ by the factor of $\sqrt[4]{\frac{\kappa_i}{\eta_j}}$ for a proper scaling between these two columns. Alternatively, we can multiply Column i of $\mathbf{C}_{\mathcal{S}_1}$ by this factor. This mechanism is summarized in Algorithm 2.

Algorithm 2: Elimination of PSS indeterminacy for $\mathbf{P}_{\mathcal{S}_1} \in \mathbb{P}_\alpha \cup \mathbb{P}_\beta$ and $\mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\beta$

Result: Correspondence between columns of the response matrices;
Input: The SEM-LV model in Fig. 2.1 with n visible and k hidden variables,
experimental data for action matrices $\mathbf{P}_{\mathcal{S}_1} \in \mathbb{P}_\alpha \cup \mathbb{P}_\beta$ and $\mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\beta$;
— Apply the method of moments in Section 3.2 on the experimental data for $\mathbf{P}_{\mathcal{S}_1}$ to obtain columns of the response $\mathbf{C}_{\mathcal{S}_1}$ up to PSS as in Eqs. (5.4) and (5.6);
— Apply the method of moments in Section 3.2 on the experimental data for $\mathbf{P}_{\mathcal{S}_2}$ to obtain columns of the response $\mathbf{C}_{\mathcal{S}_2}$ up to PSS as in Eqs. (5.5) and (5.7);
— Calculate the matrix $\mathbf{F}$ as in Eqs. (5.13) and (5.14);
for $i = 1, \cdots, k$ **do**
  for $j = 1, \cdots, k$ **do**
    if $f_{i,j} = \sqrt[4]{\frac{\kappa_i}{\eta_j}}$;
      • Column i of $\mathbf{C}_{\mathcal{S}_1}$ corresponds to Column j of $\mathbf{C}_{\mathcal{S}_2}$;
      • Divide the $j-$th column of $\mathbf{C}_{\mathcal{S}_2}$ by $\sqrt[4]{\frac{\kappa_i}{\eta_j}}$;
      • Keep the current sign of these two columns;
    if $f_{i,j} = -\sqrt[4]{\frac{\kappa_i}{\eta_j}}$;
      • Column i of $\mathbf{C}_{\mathcal{S}_1}$ corresponds to Column j of $\mathbf{C}_{\mathcal{S}_2}$;
      • Divide the $j-$th column of $\mathbf{C}_{\mathcal{S}_2}$ by $\sqrt[4]{\frac{\kappa_i}{\eta_j}}$;
      • Flip the sign for one of these two columns;
  end
end

5.1.3 Matching Subroutine 3

In the third scenario, we consider the case where $\mathbf{P}_{\mathcal{S}_1} \in \mathbb{P}_\alpha \cup \mathbb{P}_\beta$ and $\mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\gamma$. We present an identification method in Section 5.2 that uses interventions in $\mathbb{P}_\alpha \cup \mathbb{P}_\beta$ to recover the causal matrix $\mathbf{B}$ except for its first column. Hence, throughout this section, we assume that columns of $\mathbf{B}$, except for the first one, are already identified using interventions in $\mathbb{P}_\alpha \cup \mathbb{P}_\beta$ for which Algorithms 1 and 2 can be used to eliminate the PSS indeterminacy.

Using the method of moments in Section 3.2, we can obtain the response matrices $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ to actions $\mathbf{P}_{\mathcal{S}_1}$ and $\mathbf{P}_{\mathcal{S}_2}$, respectively, where columns are obtained up to permutation, absolute scale and sign. For a proper scaling between these columns, we divide each excess kurtosis κ_i in Eq. (5.4) by $|\kappa_i|$, and multiply the column $\vec{\mathbf{v}}^{(i)}$ by $\sqrt[4]{|\kappa_i|}$. Similarly, we divide each excess kurtosis η_i in Eq. (5.5) by $|\eta_i|$ and multiply the column $\vec{\mathbf{w}}^{(i)}$ by $\sqrt[4]{|\eta_i|}$. Although the correspondence in terms of permutation and sign between columns of $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ is still unknown, these columns are scaled properly after this process that sets all excess kurtosis values to ± 1. Therefore, for each row t such that $t \in \{i \mid \mathbf{P}_{\mathcal{S}_1}(i,i) = \mathbf{P}_{\mathcal{S}_2}(i,i) = 1\}$ we have:

$$\sum_{i=1}^{k} \langle \mathbf{Z}(t,:) \ , \ \mathbf{C}_{\mathcal{S}_1}(:,i) \rangle^2 = \sum_{i=1}^{k} \langle \mathbf{Z}(t,:) \ , \ \mathbf{C}_{\mathcal{S}_2}(:,i) \rangle^2 \tag{5.16}$$

because both sides of Eq. (5.16) are equal to the sum of squares for elements in Row t of $\mathbf{A}$. Note that the only unknown variable in Eq. (5.16) is the element $\mathbf{B}(t,1)$ since other columns of $\mathbf{B}$ are already identified. Moreover, the element $\mathbf{B}(t,1)$ only appears on the left hand side of Eq. (5.16) since the first row of $\mathbf{C}_{\mathcal{S}_2}$ is zero. Therefore, the second order polynomial in Eq. (5.16) yields two solutions $\mathbf{B}(t,1) = b'_{t,1}$ or $b''_{t,1}$.

Lemma 4. *Only one of two solutions for* $\mathbf{B}(t,1)$ *in Eq. (5.16) is consistent with the rank-one tensor decomposition in Eq. (3.4).*

The proof is provided in Appendix F.

Note that both $\mathbf{Z}(t,:)\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{Z}(t,:)\mathbf{C}_{\mathcal{S}_2}$ yield Row t of $\mathbf{A}$ up to permutation and sign. Hence, Column i of $\mathbf{C}_{\mathcal{S}_1}$ corresponds to Column j of $\mathbf{C}_{\mathcal{S}_2}$ if and only if the equation $\mathbf{Z}(t,:)\mathbf{C}_{\mathcal{S}_1}(:,i) = \mathbf{Z}(t,:)\mathbf{C}_{\mathcal{S}_2}(:,j)$ holds and the normalized excess kurtosis κ_i and η_j are equal. Assumption 2 implies the uniqueness of this one-to-one correspondence. Moreover, by repeating this procedure for all indices $t \in \{2, \cdots, n\}$, using actions $\mathbf{P}_{\mathcal{S}_1} \in \mathbb{P}_\alpha \cup \mathbb{P}_\beta$ and $\mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\gamma$ such that $t \in \{i \mid \mathbf{P}_{\mathcal{S}_1}(i,i) = \mathbf{P}_{\mathcal{S}_2}(i,i) = 1\}$, we can fully recover the first column of the causal matrix $\mathbf{B}$. This process is summarized in Algorithm 3.

Algorithm 3: Elimination of PSS indeterminacy for $\mathbf{P}_{\mathcal{S}_1} \in \mathbb{P}_\alpha \cup \mathbb{P}_\beta$ and $\mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\gamma$

Result: Correspondence between columns of the response matrices,
the first column of the causal matrix $\mathbf{B}$;

Input: The SEM-LV model in Fig. 2.1 with n visible and k hidden variables,
the causal matrix $\mathbf{B}$ except for its first column,
experimental data for action matrices $\mathbf{P}_{\mathcal{S}_1} \in \mathbb{P}_\alpha \cup \mathbb{P}_\beta$ and $\mathbf{P}_{\mathcal{S}_2} \in \mathbb{P}_\gamma$;

— Apply the method of moments in Section 3.2 on the experimental data for $\mathbf{P}_{\mathcal{S}_1}$ to obtain columns of the response $\mathbf{C}_{\mathcal{S}_1}$ up to PSS as in Eqs. (5.4) and (5.6);

— Apply the method of moments in Section 3.2 on the experimental data for $\mathbf{P}_{\mathcal{S}_2}$ to obtain columns of the response $\mathbf{C}_{\mathcal{S}_2}$ up to PSS as in Eqs. (5.5) and (5.7);

for $i = 1, \cdots, k$ **do**

— Divide κ_i by $|\kappa_i|$, and multiply $\vec{\mathbf{v}}^{(i)}$ by $\sqrt[4]{|\kappa_i|}$;

— Divide η_i by $|\eta_i|$, and multiply $\vec{\mathbf{w}}^{(i)}$ by $\sqrt[4]{|\eta_i|}$;

end

for $t \in \{i \mid \mathbf{P}_{\mathcal{S}_1}(i,i) = \mathbf{P}_{\mathcal{S}_2}(i,i) = 1\}$ **do**

— Calculate $\mathbf{B}(t,1)$ using Lemma 4 and Eq. (5.16);

for $i = 1, \cdots, k$ **do**

for $j = 1, \cdots, k$ **do**

if $\mathbf{Z}(t,:)\mathbf{C}_{\mathcal{S}_1}(:,i) = \mathbf{Z}(t,:)\mathbf{C}_{\mathcal{S}_2}(:,j)$ and $\kappa_i = \eta_j$;

• Column i of $\mathbf{C}_{\mathcal{S}_1}$ corresponds to Column j of $\mathbf{C}_{\mathcal{S}_2}$;

end

end

end

5.2 Identification Algorithm

Here, we consider n atomic interventions $\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_2}, \cdots, \mathbf{P}_{\mathcal{S}_n}$ such that $\mathcal{S}_i = \{i\}$ for all $1 \leq i \leq n$, and show that the experimental data associated with this set of n actions contains sufficient information to recover the mixing and causal matrices $\mathbf{A}$ and $\mathbf{B}$, respectively.

First, we aim to recover the causal matrix $\mathbf{B}$, except for its first column, one row at a time. For this purpose, we consider the row index $j \in \{2, \cdots, n\}$ and our goal is to recover the elements $\mathbf{B}(j, 2:n)$ using actions $\mathbf{P}_{\mathcal{S}_2}, \mathbf{P}_{\mathcal{S}_3}, \cdots, \mathbf{P}_{\mathcal{S}_{j-1}}, \mathbf{P}_{\mathcal{S}_{j+1}}, \cdots, \mathbf{P}_{\mathcal{S}_n}$. We have excluded the case of $j = 1$ since $\mathbf{B}$ is a strictly lower triangular matrix and its first row is zero. Using Algorithms 1 and 2, we can eliminate the PSS indeterminacy among columns of response matrices for each pair of these actions. Therefore, for each of the edges $(\mathbf{P}_{\mathcal{S}_2}, \mathbf{P}_{\mathcal{S}_3})$, $(\mathbf{P}_{\mathcal{S}_3}, \mathbf{P}_{\mathcal{S}_4}), \cdots,$ $\left(\mathbf{P}_{\mathcal{S}_{j-2}}, \mathbf{P}_{\mathcal{S}_{j-1}}\right)$, $\left(\mathbf{P}_{\mathcal{S}_{j-1}}, \mathbf{P}_{\mathcal{S}_{j+1}}\right)$, $\left(\mathbf{P}_{\mathcal{S}_{j+1}}, \mathbf{P}_{\mathcal{S}_{j+2}}\right), \cdots,$ $\left(\mathbf{P}_{\mathcal{S}_{n-1}}, \mathbf{P}_{\mathcal{S}_n}\right)$ we can constitute the Eq. (4.7) accordingly. Column-wise concatenation of the matrices $\mathbf{C}_{\mathcal{S}_r} - \mathbf{C}_{\mathcal{S}_t}$ corresponding to these edges yields the matrix $\mathbf{W} \in \mathbb{R}^{n \times (k(n-3))}$. According to the Theorem 4.2, the rank of this matrix can be calculated as:

$$\operatorname{rank}[\mathbf{W}] = |\{1, \cdots, n\} \backslash \{1, j\}| = n - 2. \tag{5.17}$$

Note that this is also consistent with Conjecture 1 where we have:

$$L \leq \operatorname{rank}[\mathbf{W}] \leq U, \tag{5.18}$$

such that:

$$L = \min\left[(n-3) \times \min(2,\ k)\ ,\ n-2\right], \tag{5.19}$$

$$U = \min\left[\max\left[(n-3) \times \min(2,\ k)\ ,\ n-2\right]\ ,\ n-2\right]. \tag{5.20}$$

Since $k \geq 2$ and $n \geq 4$, from straightforward calculations it follows that $L = U = n - 2$;

thus, we have rank $[\mathbf{W}] = n - 2$.

Let $\mathbf{W}_f \in \mathbb{R}^{n\times(n-2)}$ be an $n \times (n-2)$ matrix formed by keeping the $n-2$ columns of $\mathbf{W}$ that are linearly independent. Now, we have:

$$\mathbf{W}_f^T \mathbf{Z}(j,:)^T = \vec{\mathbf{0}}. \tag{5.21}$$

According to Lemma 3, the first row of $\mathbf{W}$, which is multiplied by the element $\mathbf{B}(j,1)$ in Eq. (5.21), is zero. Moreover, since $\mathbf{B}$ is a causal matrix, we have $\mathbf{Z}(j,j) = 1$. Therefore, if $\overline{\mathbf{W}}_f \in \mathbb{R}^{(n-2)\times(n-2)}$ is an $(n-2)\times(n-2)$ matrix formed by removing rows 1 and j of $\mathbf{W}_f$, and $\overline{\mathbf{Z}}(j,:) \in \mathbb{R}^{1\times(n-2)}$ is formed by removing the 1st and $j-$th elements of $\mathbf{Z}(j,:)$, then we have:

$$\overline{\mathbf{W}}_f^T \overline{\mathbf{Z}}(j,:)^T = -\mathbf{W}_f(j,:)^T, \tag{5.22}$$

thus, the unknown elements in Row j of $\mathbf{Z}$ can be recovered as:

$$\overline{\mathbf{Z}}(j,:)^T = -\left(\overline{\mathbf{W}}_f^T\right)^{-1} \mathbf{W}_f(j,:)^T. \tag{5.23}$$

In other words, Eq. (5.23) yields the elements in $\mathbf{Z}(j, 2:n)$ except for the element $\mathbf{Z}(j,j)$ which is known a priori to be 1. This renders Row j of $\mathbf{B}$ to be identifiable except for the element $\mathbf{B}(j,1)$ in the first column. By repeating this process for each row, the causal matrix $\mathbf{B}$ can be fully recovered, except for its first column, using the $n-1$ atomic actions $\mathbf{P}_{\mathcal{S}_2}, \mathbf{P}_{\mathcal{S}_3}, \cdots, \mathbf{P}_{\mathcal{S}_n}$.

Intuitively, the first column of the causal matrix $\mathbf{B}$ indicates how the variable X_1 affects other visible variables, and to capture this causal effect, we need to intervene on the variable X_1. By using actions $\mathbf{P}_{\mathcal{S}_2}, \mathbf{P}_{\mathcal{S}_3}, \cdots, \mathbf{P}_{\mathcal{S}_n}$, no intervention is performed on the variable X_1; hence, its effect on other variables is cancelled out from both sides of the Eq. (4.6). The intuition

for this cancellation effect is captured in Lemma 3 which implies that the first row of the matrix $\mathbf{W}_f$ in Eq. (5.21) is zero and the variable X_1 does not appear in the identification process outlined in Eqs. (5.21) to (5.23).

Now, we aim to recover the first column of $\mathbf{B}$, one element at a time. For this purpose, we consider the row index $j \in \{2, \cdots, n\}$ and our goal is to identify the element $\mathbf{B}(j, 1)$. Note that for any atomic action such as $\mathbf{P}_{\mathcal{S}_t}$ where $t \neq j$, we have $j \in \{i \mid \mathbf{P}_{\mathcal{S}_1}(i, i) = \mathbf{P}_{\mathcal{S}_t}(i, i) = 1\}$. Therefore, the element $\mathbf{B}(j, 1)$ can be uniquely calculated from the pair of actions $\mathbf{P}_{\mathcal{S}_1}$ and $\mathbf{P}_{\mathcal{S}_t}$ using Eq. (5.16) and Lemma 4 as outlined in Algorithm 3. Now that the causal matrix $\mathbf{B}$ is fully recovered, the mixing matrix $\mathbf{A}$ can be calculated up to the PSS indeterminacy of independent components from Eq. (4.15). This procedure is summarized in Algorithm 4.

Note that causal relations among visible variables makes the matrix $\mathbf{B}$ to be strictly lower triangular after topological ordering. Therefore, there are at most $j - 1$ unknown elements in Row j of $\mathbf{B}$. However, in Eqs. (5.21) to (5.23) we have only used the fact that $\mathbf{B}(j, j) = 0$, i.e. we have treated all other elements to be unknown, and let the algorithm to recover them. Since the algorithm uses the experimental data that satisfies the causal relations between visible variables, it will calculate all elements on or above the diagonal of $\mathbf{B}$ to be zero. However, as discussed in Section 4.2, similar procedure as described above can be carried out that uses the fact that $\mathbf{B}$ is strictly lower triangular, and employs atomic actions on $X_1, \cdots, X_{j-1}$, i.e. $\mathbf{P}_{\mathcal{S}_1}, \cdots, \mathbf{P}_{\mathcal{S}_{j-1}}$, to recover Row j of $\mathbf{B}$.

Algorithm 4: Identification Method

Result: The mixing matrix $\mathbf{A}$ and the causal matrix $\mathbf{B}$;

Input: The SEM-LV model in Fig. 2.1 with n visible and k hidden variables, experimental data for each atomic action $\mathbf{P}_{\mathcal{S}_i}$ where $\mathcal{S}_i = \{i\}$ for $1 \leq i \leq n$;

// Recover the response matrices;

for $i = 1, \cdots, n$ **do**

— Apply the method of moments in Section 3.2 on the experimental data for $\mathbf{P}_{\mathcal{S}_i}$ to obtain columns of the response matrix $\mathbf{C}_{\mathcal{S}_i}$ up to PSS;

end

// Recover the causal matrix $\mathbf{B}$ one row at a time, except for its first column;

for $t = 2, \cdots, n$ **do**

— Initialize the matrix $\mathbf{W} \in \mathbb{R}^{n \times 0}$;

for $\left(\mathbf{P}_{\mathcal{S}_i}, \mathbf{P}_{\mathcal{S}_j}\right) \in \big\{ \left(\mathbf{P}_{\mathcal{S}_2}, \mathbf{P}_{\mathcal{S}_3}\right), \cdots, \left(\mathbf{P}_{\mathcal{S}_{t-2}}, \mathbf{P}_{\mathcal{S}_{t-1}}\right), \left(\mathbf{P}_{\mathcal{S}_{t-1}}, \mathbf{P}_{\mathcal{S}_{t+1}}\right), \left(\mathbf{P}_{\mathcal{S}_{t+1}}, \mathbf{P}_{\mathcal{S}_{t+2}}\right), \cdots, \left(\mathbf{P}_{\mathcal{S}_{n-1}}, \mathbf{P}_{\mathcal{S}_n}\right) \big\}$ **do**

— Eliminate the PSS indeterminacy between columns of response matrices $\mathbf{C}_{\mathcal{S}_i}$ and $\mathbf{C}_{\mathcal{S}_j}$ using Algorithms 1 and 2;

— Perform column-wise matrix concatenation $\mathbf{W} = \left[\mathbf{W} \ ; \ \mathbf{C}_{\mathcal{S}_i} - \mathbf{C}_{\mathcal{S}_j}\right]$;

end

— Obtain the matrices $\mathbf{W}_f$ and $\overline{\mathbf{W}}_f$ from $\mathbf{W}$ as in Eqs. (5.21) and (5.22);

— Recover the row t of $\mathbf{B}$ except for the element $\mathbf{B}(t, 1)$ using Eq. (5.23);

end

// Recover the first column of the causal matrix $\mathbf{B}$;

— Select two distinct indices $i, j \in \{2, \cdots, n\}$;

— Identify the first column of $\mathbf{B}$ by applying Algorithm 3 on the pairs of actions $(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_i})$ and $\left(\mathbf{P}_{\mathcal{S}_1}, \mathbf{P}_{\mathcal{S}_j}\right)$;

// Recover the mixing matrix A up to the PSS of its columns;

— Eliminate the PSS indeterminacy between columns of the response matrices $\mathbf{C}_{\mathcal{S}_1}, \cdots, \mathbf{C}_{\mathcal{S}_n}$ using Algorithms 1, 2 and 3;

— Using Eq. (4.15), recover the mixing matrix as $\mathbf{A} = \frac{1}{n-1} \sum_{i=1}^{n} \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_i} \mathbf{B}\right) \mathbf{C}_{\mathcal{S}_i}$

Chapter 6

Results and Discussions

We validate our proposed identification method using an exemplary SEM-LV model with $n = 5$ visible and $k = 3$ hidden variables. Latent variables are i.i.d. and admit a Laplacian distribution with the location parameter $\mu = 0$, and the scale parameter $b = 0.2$. The ground truth mixing and causal matrices $\mathbf{A}$ and $\mathbf{B}$ are provided in Table 6.1.

Table 6.1: The ground-truth mixing and causal matrices.

mixing matrix A			**causal matrix B**				
4.2641	3.1823	2.3416	0	0	0	0	0
0.2616	1.9111	3.0209	0.4486	0	0	0	0
−0.3516	−0.9527	−0.5902	0.9099	0.3969	0	0	0
−2.8339	−4.4090	−3.1865	0.8436	0.7028	0.4585	0	0
−10.7413	−7.2342	−7.3736	1.4372	0.6046	0.7382	0.7976	0

Each realization of visible variables is generated using a random sampling of latent variables according to their distribution, and the ground truth mixing and causal matrices, i.e. the matrices $\mathbf{A}$ and $\mathbf{B}$ in Table 6.1 and their modifications according to Eqs. (2.5) and (2.6). In what follows, we vary the number of generated data over visible variables, and study its effect on the estimation performance of Algorithm 4.

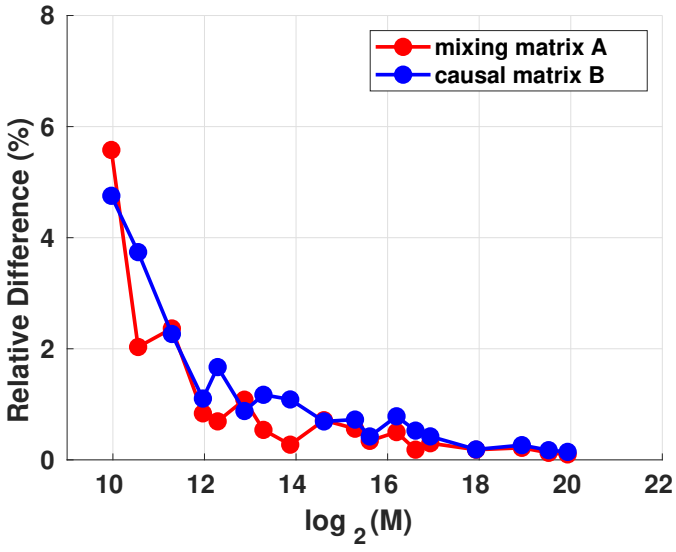


Figure 6.1: Estimation error versus sample size.

Assume that $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ are the estimated mixing and causal matrices using Algorithm 4. In order to measure the efficacy of the estimation, we calculate the ratio between Frobenius norm of the difference between predicted and ground-truth, and Frobenius norm of the ground-truth. In other words, we compute $\frac{\|\mathbf{A}-\widehat{\mathbf{A}}\|_F}{\|\mathbf{A}\|_F}$ and $\frac{\|\mathbf{B}-\widehat{\mathbf{B}}\|_F}{\|\mathbf{B}\|_F}$ versus M, i.e. the number of observed data for each atomic action used for matrix recovery, as shown in Figure 6.1. Since the estimated mixing matrix is always up to the permutation, scale and sign of its columns, we eliminate the PSS indeterminacy between $\mathbf{A}$ and $\widehat{\mathbf{A}}$ using the aforementioned matching algorithms before calculating their relative difference in Frobenius norm.

As shown in Figure 6.1, the estimation error decreases with an increase in the number of observed data for each atomic action. Note that our analysis and algorithms are based on infinite sample regime, where the rank one components in tensor decomposition can be recovered perfectly. In this scenario, our estimation of $\mathbf{A}$ and $\mathbf{B}$ will have zero error due to theoretical guarantees by Theorems 4.1 and 4.2. This is in agreement with Figure 6.1 where our identification algorithm achieves a relative difference of 0.1% and 0.14% for the mixing and causal matrices, respectively, given $M = 10^6$ experimental data for each intervention. The matrices $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ in this case are provided in Table 6.2.

Table 6.2: The estimated mixing and causal matrices for $M = 10^6$.

mixing matrix $\widehat{\mathbf{A}}$			**causal matrix $\widehat{\mathbf{B}}$**				
4.2640	3.1779	2.3466	0	0	0	0	0
0.2610	1.9121	3.0064	0.4509	0	0	0	0
−0.3495	−0.9472	−0.5849	0.9091	0.3957	0	0	0
−2.8373	−4.4026	−3.1822	0.8427	0.7021	0.4584	0	0
−10.7358	−7.2342	−7.3756	1.4351	0.6061	0.73811	0.7972	0

When M is not infinite, elements of the matrix $\mathbf{\Theta}$ in Eq. (5.9) are not exactly 0 or π for the corresponding columns, due to the approximation error in rank one tensor decomposition. Therefore, first we keep track of the change in sign and replace $\theta_{i,j}$ with $\pi - \theta_{i,j}$ if $\theta_{i,j} > \frac{\pi}{2}$, then we apply Hungarian algorithm to find the corresponding columns. Similar approach is carried out for the matrix $\mathbf{F}$ in Eq. (5.13). Figure 6.1 demonstrates the robustness of our method to this error in tensor decomposition for sample sizes as small as $M = 10^3$.

An observation in Figure 6.1 reveals that the error in estimation of the mixing matrix is usually less than that of the causal matrix. This is partially due to the averaging effect in Eq. (4.15) and the last step of Algorithm 4, where the effect of estimation error in one particular response matrix is mitigated by averaging over all response matrices. In particular, simulation results show that if we consider a modification of Eq. (4.15) where $\mathbf{A}$ is recovered using response matrices to only two atomic actions, the relative error in Frobenius norm of $\mathbf{A}$ and $\widehat{\mathbf{A}}$ increases, although it is still less than that of the causal matrix.

In this work, we focused on the space of all possible actions that sets the value of a collection of visible variables to zero, and presented an identification algorithm to recover the mixing and causal matrices. However, generalization of this framework can be considered such that not only intervened variables can have non-zero values, but also the observational and experimental data are noisy. In [25], a preliminary study of this scenario is provided along with the theoretical analysis of the matrix recovery in finite sample regime.

Chapter 7

Conclusion

In this work, a linear structural equation model is discussed where latent variables follow the mixing process of the blind source separation problem to affect observed variables. We modeled this identification task as a disentanglement problem, and studied the necessary and sufficient conditions for recovery of the model parameters. Since observational data can only be used to recover a combination of latent and observed causal effects, we employed a series of interventions that would guarantee unique identification. A novel matching process is proposed that uses the invariance property of latent effects across observed variables to combine the information corresponding to different interventions. Then, an identification method is proposed to disentangle the observed causal effects from hidden confounders. Simulation results show that our proposed algorithm can efficiently recover the mixing and causal effects in both finite and infinite sample regime.

Bibliography

[1] J. Acharya, A. Bhattacharyya, C. Daskalakis, and S. Kandasamy. Learning and testing causal models with interventions. *arXiv preprint arXiv:1805.09697*, 2018.

[2] R. Addanki, S. Kasiviswanathan, A. McGregor, and C. Musco. Efficient intervention design for causal discovery with latents. In *International Conference on Machine Learning*, pages 63–73. PMLR, 2020.

[3] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15:2773–2832, 2014.

[4] A. Anandkumar, D. Hsu, A. Javanmard, and S. Kakade. Learning linear bayesian networks with latent variables. In *International Conference on Machine Learning*, pages 249–257. PMLR, 2013.

[5] J.-F. Cardoso. The three easy routes to independent component analysis; contrasts and geometry. In *Proc. ICA*, volume 2001, pages 1–6. Citeseer, 2001.

[6] Z. Chen and L. Chan. Causality in linear nongaussian acyclic models in the presence of latent gaussian confounders. *Neural Computation*, 25(6):1605–1641, 2013.

[7] D. Colombo, M. H. Maathuis, M. Kalisch, and T. S. Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pages 294–321, 2012.

[8] P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[9] G. Darmois. Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire. *Revue de l'Institut International de Statistique*, pages 2–8, 1953.

[10] F. Eberhardt. Causation and intervention. *Unpublished doctoral dissertation, Carnegie Mellon University*, page 93, 2007.

[11] F. Eberhardt, C. Glymour, and R. Scheines. N-1 experiments suffice to determine the causal relations among n variables. In *Innovations in machine learning*, pages 97–112. Springer, 2006.

[12] F. Eberhardt and R. Scheines. Interventions and causal inference. *Philosophy of science*, 74(5):981–995, 2007.

[13] A. Ghassami, S. Salehkaleybar, and N. Kiyavash. Interventional experiment design for causal structure learning. *arXiv preprint arXiv:1910.05651*, 2019.

[14] A. Ghassami, S. Salehkaleybar, N. Kiyavash, and E. Bareinboim. Budgeted experiment design for causal structure learning. In *International Conference on Machine Learning*, pages 1724–1733. PMLR, 2018.

[15] Y.-B. He and Z. Geng. Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9(Nov):2523–2547, 2008.

[16] J. Hérault, C. Jutten, and B. Ans. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In *10 Colloque sur le traitement du signal et des images, FRA, 1985*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 1985.

[17] P. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf. Nonlinear causal discovery with additive noise models. *Advances in neural information processing systems*, 21:689–696, 2008.

[18] P. O. Hoyer, S. Shimizu, A. J. Kerminen, and M. Palviainen. Estimation of causal effects using linear non-gaussian causal models with hidden variables. *International Journal of Approximate Reasoning*, 49(2):362–378, 2008.

[19] F. Huang, N. U. Naresh, I. Perros, R. Chen, J. Sun, and A. Anandkumar. Guaranteed scalable learning of latent tree models. In *Uncertainty in Artificial Intelligence*, pages 883–893. PMLR, 2020.

[20] A. Hyttinen, F. Eberhardt, and P. O. Hoyer. Learning linear cyclic causal models with latent variables. *The Journal of Machine Learning Research*, 13(1):3387–3439, 2012.

[21] A. Hyttinen, F. Eberhardt, and M. Järvisalo. Constraint-based causal discovery: Conflict resolution with answer set programming. In *UAI*, pages 340–349, 2014.

[22] S. Karimi-Bidhendi, A. Arafati, A. L. Cheng, Y. Wu, A. Kheradvar, and H. Jafarkhani. Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases. *Journal of Cardiovascular Magnetic Resonance*, 22(1):1–24, 2020.

[23] S. Karimi-Bidhendi, F. Munshi, and A. Munshi. Scalable classification of univariate and multivariate time series. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 1598–1605. IEEE, 2018.

[24] M. Kocaoglu, K. Shanmugam, and E. Bareinboim. Experimental design for learning causal graphs with latent variables. In *Nips*, 2017.

[25] A. Liu, H. Liu, T. Li, S. Karimi-Bidhendi, Y. Yue, and A. Anandkumar. Disentangling observed causal effects from latent confounders using method of moments. *arXiv preprint arXiv:2101.06614*, 2021.

[26] G. Matsaglia and G. PH Styan. Equalities and inequalities for ranks of matrices. *Linear and multilinear Algebra*, 2(3):269–292, 1974.

[27] J. Pearl. *Causality*. Cambridge university press, 2009.

[28] C. R. Rao and S. K. Mitra. Further contributions to the theory of generalized inverse of matrices and its applications. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 289–300, 1971.

[29] J.-P. Roll. *Contribution à la Proprioception Musculaire, à la Perception et au Contrôle du Mouvement Chez l'Homme*. PhD thesis, Éditeur inconnu, 1981.

[30] S. Salehkaleybar, A. Ghassami, N. Kiyavash, and K. Zhang. Learning linear non-gaussian causal models in the presence of latent variables. *Journal of Machine Learning Research*, 21(39):1–24, 2020.

[31] F. Sepehr and D. Materassi. An algorithm to learn polytree networks with hidden nodes. In *NeurIPS*, pages 15084–15093, 2019.

[32] S. Shimizu, P. O. Hoyer, A. Hyvärinen, A. Kerminen, and M. Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.

[33] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer, and K. Bollen. Directlingam: A direct method for learning a linear non-gaussian structural equation model. *The Journal of Machine Learning Research*, 12:1225–1248, 2011.

[34] A. Skrondal and S. Rabe-Hesketh. Latent variable modelling: a survey. *Scandinavian Journal of Statistics*, 34(4):712–745, 2007.

[35] P. Spirtes. An anytime algorithm for causal inference. In *AISTATS*, 2001.

[36] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman. *Causation, prediction, and search*. MIT press, 2000.

[37] P. L. Spirtes, C. Meek, and T. S. Richardson. Causal inference in the presence of latent variables and selection bias. *arXiv preprint arXiv:1302.4983*, 2013.

[38] T. Tashiro, S. Shimizu, A. Hyvärinen, and T. Washio. Parcelingam: a causal ordering method robust against latent confounders. *Neural computation*, 26(1):57–83, 2014.

[39] K. Zhang and A. Hyvarinen. On the identifiability of the post-nonlinear causal model. *arXiv preprint arXiv:1205.2599*, 2012.

Appendix A

Proof of the Lemma 1

Let $i \in \mathcal{R}_{\mathcal{S}_1,\mathcal{S}_3}$, i.e. $\mathbf{P}_{\mathcal{S}_1}(i,i) \neq \mathbf{P}_{\mathcal{S}_3}(i,i)$. Now, if $\mathbf{P}_{\mathcal{S}_2}(i,i) = \mathbf{P}_{\mathcal{S}_3}(i,i)$, then $\mathbf{P}_{\mathcal{S}_1}(i,i) \neq \mathbf{P}_{\mathcal{S}_2}(i,i)$, i.e. $i \in \mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}$. However, if $\mathbf{P}_{\mathcal{S}_2}(i,i) \neq \mathbf{P}_{\mathcal{S}_3}(i,i)$, then $i \in \mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}$. Therefore, $i \in \mathcal{R}_{\mathcal{S}_1,\mathcal{S}_3}$ implies that i either belongs to $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}$ or $\mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}$, which indicates that $|\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_3}| \leq |\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}| + |\mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}|$.

The above argument implies that if $i \in \mathcal{R}_{\mathcal{S}_1,\mathcal{S}_3}$, then i either belongs to $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}$ or $\mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}$, but it cannot belong to both. In other words, $i \in \mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2} \cup \mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}$ but $i \notin \mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2} \cap \mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}$. Therefore, the equality in Eq. (4.2) holds if and only if $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2} \cap \mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3} = \varnothing$, i.e. if two sets $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}$ and $\mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}$ are disjoint, and the proof is complete. ■

Appendix B

Proof of the Lemma 2

For any index $t \in \{1, \cdots, n\}$, let $g(t)$ be the cardinality of the set

$$\big\{i \mid \mathbf{P}_{\mathcal{S}_i}(t,t) \neq \mathbf{P}_{\mathcal{S}_{i+1}}(t,t), \quad i \in \{1, \cdots, l-1\}\big\}. \tag{B.1}$$

Since diagonal entries of action matrices are either 0 or 1, it directly follows from the definition of the set $\mathcal{R}$ that $t \in \mathcal{R}_{\mathcal{S}_l,\mathcal{S}_1}$ if and only if $g(t)$ is an odd number. Note that if $g(t)$ is an odd number, then $g(t) \geq 1$; hence, there exists at least one index such as $j \in \{1, \cdots, l-1\}$ for which we have $\mathbf{P}_{\mathcal{S}_j}(t,t) \neq \mathbf{P}_{\mathcal{S}_{j+1}}(t,t)$, i.e. $t \in \mathcal{R}_{\mathcal{S}_j,\mathcal{S}_{j+1}}$. Therefore, any element of $\mathcal{R}_{\mathcal{S}_l,\mathcal{S}_1}$, such as t, appears in at least one of the sets $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}, \mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}, \cdots, \mathcal{R}_{\mathcal{S}_{l-1},\mathcal{S}_l}$ and Eq. (4.3) is proved.

The uniqueness of the set $\mathcal{R}_{\mathcal{S}_l,\mathcal{S}_1}$ comes from the fact that for each $t \in \mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2} \cup \mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3} \cup \cdots \cup \mathcal{R}_{\mathcal{S}_{l-1},\mathcal{S}_l}$, we have $t \in \mathcal{R}_{\mathcal{S}_l,\mathcal{S}_1}$ if and only if the number of sets $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}, \mathcal{R}_{\mathcal{S}_2,\mathcal{S}_3}, \cdots, \mathcal{R}_{\mathcal{S}_{l-1},\mathcal{S}_l}$ to which t belongs is an odd number, and the proof is complete. ■

Appendix C

Proof of the Theorem 4.1

Using Eq. (2.7), we can write $\mathbf{C}_{\mathcal{S}_1} - \mathbf{C}_{\mathcal{S}_2}$ as follows:

$$\mathbf{C}_{\mathcal{S}_1} - \mathbf{C}_{\mathcal{S}_2} = \left[\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1} \mathbf{B} \right)^{-1} \mathbf{P}_{\mathcal{S}_1} - \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2} \mathbf{B} \right)^{-1} \mathbf{P}_{\mathcal{S}_2} \right] \mathbf{A}. \tag{C.1}$$

Since $\mathbf{A}$ is a full rank matrix, i.e. $\operatorname{rank}[\mathbf{A}] = \min(n, k) = k$, then we need to prove that:

$$\operatorname{rank}[\mathbf{D}] = \left| \mathcal{R}_{\mathcal{S}_1, \mathcal{S}_2} \right|. \tag{C.2}$$

where $\mathbf{D} = \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1} \mathbf{B} \right)^{-1} \mathbf{P}_{\mathcal{S}_1} - \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2} \mathbf{B} \right)^{-1} \mathbf{P}_{\mathcal{S}_2}$. For an action matrix $\mathbf{P}_{\mathcal{S}}$, we define the matrix $\mathbf{Q}_{\mathcal{S}}$ to be the submatrix of $\mathbf{P}_{\mathcal{S}}$ restricted to rows in $\{i \mid 1 \leq i \leq n, i \notin \mathcal{S}\}$. It readily follows that $\mathbf{Q}_{\mathcal{S}}^T \mathbf{Q}_{\mathcal{S}} = \mathbf{P}_{\mathcal{S}}$, and $\mathbf{Q}_{\mathcal{S}} \mathbf{Q}_{\mathcal{S}}^T$ is equal to the identity matrix. Moreover, using straightforward algebraic calculations, we can rewrite the matrix on the left-hand-side of Eq. (C.2) as:

$$\begin{aligned} \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1} \mathbf{B} \right)^{-1} \mathbf{P}_{\mathcal{S}_1} - \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2} \mathbf{B} \right)^{-1} \mathbf{P}_{\mathcal{S}_2} = {} & \mathbf{Q}_{\mathcal{S}_1}^T \left[\mathbf{Q}_{\mathcal{S}_1} \mathbf{P}_{\mathcal{S}_1} \left(\mathbf{I} - \mathbf{B} \right) \mathbf{P}_{\mathcal{S}_1} \mathbf{Q}_{\mathcal{S}_1}^T \right]^{-1} \mathbf{Q}_{\mathcal{S}_1} \\ & - \mathbf{Q}_{\mathcal{S}_2}^T \left[\mathbf{Q}_{\mathcal{S}_2} \mathbf{P}_{\mathcal{S}_2} \left(\mathbf{I} - \mathbf{B} \right) \mathbf{P}_{\mathcal{S}_2} \mathbf{Q}_{\mathcal{S}_2}^T \right]^{-1} \mathbf{Q}_{\mathcal{S}_2}. \end{aligned} \tag{C.3}$$

Let $\mathbf{Z} = \mathbf{I} - \mathbf{B}$, and denote the complement of the set $\mathcal{S}$ by $\overline{\mathcal{S}} = \{1, \cdots, n\} \backslash \mathcal{S}$. Given a matrix $\mathbf{U}$ and two index sets $\mathcal{S}_r$ and $\mathcal{S}_c$, let $\mathbf{U}(\mathcal{S}_r, \mathcal{S}_c)$ be the submatrix of $\mathbf{U}$ restricted to rows in $\mathcal{S}_r$ and columns in $\mathcal{S}_c$. From Eq. (C.3) and properties of the matrix $\mathbf{Q}_{\mathcal{S}}$, we have:

$$\left((\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B})^{-1}\mathbf{P}_{\mathcal{S}_1}\right)\left(\overline{\mathcal{S}}_1, \overline{\mathcal{S}}_1\right) = \left(\mathbf{Z}\left(\overline{\mathcal{S}}_1, \overline{\mathcal{S}}_1\right)\right)^{-1}, \tag{C.4}$$

$$\left((\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B})^{-1}\mathbf{P}_{\mathcal{S}_2}\right)\left(\overline{\mathcal{S}}_2, \overline{\mathcal{S}}_2\right) = \left(\mathbf{Z}\left(\overline{\mathcal{S}}_2, \overline{\mathcal{S}}_2\right)\right)^{-1}. \tag{C.5}$$

Now, we need to following lemma to proceed with the proof.

Lemma 5. *Let $\mathcal{S} \subsetneq \{1, \cdots, n\}$ be an index set, and consider an invertible $n \times n$ matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ such that $\mathbf{U}(\mathcal{S}, \mathcal{S})$ is also invertible. Then, we have:*

$$\det\left[\left(\mathbf{U^{-1}}\right)\left(\overline{\mathcal{S}}, \overline{\mathcal{S}}\right)\right] = \frac{1}{\det[\mathbf{U}]} \times \det\left[\mathbf{U}(\mathcal{S}, \mathcal{S})\right]. \tag{C.6}$$

Proof: Without the loss of generality, we assume that $\max(\mathcal{S}) < \min\left(\overline{\mathcal{S}}\right)$ since rearranging rows of the matrix $\mathbf{U}$ to make it so only changes the sign in both sides of Eq. (C.6). Now, from the properties of determinant for block matrices, we have:

$$\det[\mathbf{U}] = \det\left[\mathbf{U}(\mathcal{S}, \mathcal{S})\right] \times \det\left[\mathbf{U}\left(\overline{\mathcal{S}}, \overline{\mathcal{S}}\right) - \mathbf{U}\left(\overline{\mathcal{S}}, \mathcal{S}\right) \times \left(\mathbf{U}(\mathcal{S}, \mathcal{S})\right)^{-1} \times \mathbf{U}\left(\mathcal{S}, \overline{\mathcal{S}}\right)\right]. \tag{C.7}$$

Moreover, from the inversion formula for block matrices, we have:

$$\left(\mathbf{U}^{-1}\right)\left(\overline{\mathcal{S}}, \overline{\mathcal{S}}\right) = \left(\mathbf{U}\left(\overline{\mathcal{S}}, \overline{\mathcal{S}}\right) - \mathbf{U}\left(\overline{\mathcal{S}}, \mathcal{S}\right) \times \left(\mathbf{U}(\mathcal{S}, \mathcal{S})\right)^{-1} \times \mathbf{U}\left(\mathcal{S}, \overline{\mathcal{S}}\right)\right)^{-1}. \tag{C.8}$$

Since the determinant of the inverse matrix is equal to the inverse of the determinant for that matrix, Eq. (C.6) follows from Eqs. (C.7) and (C.8), and the proof is complete.

For the ease of argument and notations, we assume that $\max\left(\overline{\mathcal{S}}_1 \cap \overline{\mathcal{S}}_2\right) < \min\left(\overline{\mathcal{S}}_1 \cap \mathcal{S}_2\right)$ and $\max\left(\overline{\mathcal{S}}_1 \cap \mathcal{S}_2\right) < \min\left(\mathcal{S}_1 \cap \overline{\mathcal{S}}_2\right)$. Similar reasoning can be carried out for the general case.

Now, using Lemma 5 and Eqs. (C.4) and (C.5), we have:

$$\det\left[\mathbf{D}\left(\overline{\mathcal{S}}_1 \cap \mathcal{S}_2, \overline{\mathcal{S}}_1 \cap \mathcal{S}_2\right)\right] = \frac{1}{\det\left[\mathbf{Z}\left(\overline{\mathcal{S}}_1, \overline{\mathcal{S}}_1\right)\right]} \times \det\left[\mathbf{Z}\left(\overline{\mathcal{S}}_1 \cap \overline{\mathcal{S}_2}, \overline{\mathcal{S}}_1 \cap \overline{\mathcal{S}_2}\right)\right], \tag{C.9}$$

$$\det\left[\mathbf{D}\left(\overline{\mathcal{S}}_2 \cap \mathcal{S}_1, \overline{\mathcal{S}}_2 \cap \mathcal{S}_1\right)\right] = \frac{1}{\det\left[\mathbf{Z}\left(\overline{\mathcal{S}}_2, \overline{\mathcal{S}}_2\right)\right]} \times \det\left[\mathbf{Z}\left(\overline{\mathcal{S}}_2 \cap \overline{\mathcal{S}_1}, \overline{\mathcal{S}}_2 \cap \overline{\mathcal{S}_1}\right)\right]. \tag{C.10}$$

Since $\mathbf{Z}$ is a lower triangular matrix with diagonal entries equal to 1, it follows from Eqs. (C.9) and (C.10) that:

$$\det\left[\mathbf{D}\left(\overline{\mathcal{S}}_1 \cap \mathcal{S}_2, \overline{\mathcal{S}}_1 \cap \mathcal{S}_2\right)\right] = 1 \neq 0, \tag{C.11}$$

$$\det\left[\mathbf{D}\left(\overline{\mathcal{S}}_2 \cap \mathcal{S}_1, \overline{\mathcal{S}}_2 \cap \mathcal{S}_1\right)\right] = 1 \neq 0. \tag{C.12}$$

Note that $\mathbf{D}\left(\overline{\mathcal{S}}_1 \cap \mathcal{S}_2, \overline{\mathcal{S}}_2 \cap \mathcal{S}_1\right) = \mathbf{0}$ and $\mathbf{D}\left(\overline{\mathcal{S}}_2 \cap \mathcal{S}_1, \overline{\mathcal{S}}_1 \cap \mathcal{S}_2\right) = \mathbf{0}$. Therefore the submatrix of $\mathbf{D}$ restricted to rows and columns in $\left(\overline{\mathcal{S}}_2 \cap \mathcal{S}_1\right) \cup \left(\overline{\mathcal{S}}_1 \cap \mathcal{S}_2\right)$ is a block diagonal matrix, where the two diagonal blocks have non-zero determinants according to Eqs. (C.11) and (C.12). Therefore, the rows of $\mathbf{D}$ in $\left(\overline{\mathcal{S}}_2 \cap \mathcal{S}_1\right) \cup \left(\overline{\mathcal{S}}_1 \cap \mathcal{S}_2\right)$ are all linearly independent. Note that all rows of $\mathbf{D}$ in $\mathcal{S}_1 \cap \mathcal{S}_2$ are zero since these rows are set to zero by both action matrices. Moreover, using straightforward algebraic calculations, and the fact that both matrices $(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B})$ and $(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B})$ are lower triangular, it follows that the element in row i and column j of both matrices is equal to $(-1)^{i+j} \times b_{j+1,j} \times b_{j+2,j+1} \times \cdots \times b_{i,i-1}$, for $j < i < \min(\mathcal{S}_1 \cup \mathcal{S}_2)$. Hence, the rows of $\mathbf{D}$ in $\overline{\mathcal{S}}_1 \cap \overline{\mathcal{S}}_2$ are all zero too. Therefore, the rank of the matrix $\mathbf{D}$, which is equal to the number of linearly independent rows, is equal to $\left|\left(\overline{\mathcal{S}}_2 \cap \mathcal{S}_1\right) \cup \left(\overline{\mathcal{S}}_1 \cap \mathcal{S}_2\right)\right|$. Since $\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2} = \left(\overline{\mathcal{S}}_2 \cap \mathcal{S}_1\right) \cup \left(\overline{\mathcal{S}}_1 \cap \mathcal{S}_2\right)$, it follows that $\operatorname{rank}[\mathbf{D}] = |\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}|$; thus, we have:

$$\operatorname{rank}[\mathbf{C}_{\mathcal{S}_1} - \mathbf{C}_{\mathcal{S}_2}] = \operatorname{rank}[\mathbf{D}\mathbf{A}] \leq \min(\operatorname{rank}[\mathbf{D}], \operatorname{rank}[\mathbf{A}]) = \min(|\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}|, k). \tag{C.13}$$

According to Corollary 6.1 in [26], in case $|\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}| \leq k$, the equality in Eq. (C.13) holds if and only if the partitioned matrix $\left(\mathbf{A}, \mathbf{I} - \mathbf{D}^\dagger\mathbf{D}\right)$ has full row rank. However, if $|\mathcal{R}_{\mathcal{S}_1,\mathcal{S}_2}| > k$,

then the equality in Eq. (C.13) holds if and only if the row-wise concatenation of $\mathbf{D}$ and $\mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}$ has full column rank. These conditions are met according to Assumption 4; hence, the equality in Eq. (C.13) holds, and the proof is complete. ■.

Note: Since the mixing and causal matrices are usually random and noisy in practical applications, the necessary and sufficient condition for the equality in Eq. (C.13) holds with high probability given the properties of random matrices. In general, the matrix $\mathbf{B}$ does not have to be strictly lower triangular for Theorem 4.1 to hold. Moreover, we have used the key property that $\mathbf{Z}\left(\overline{\mathcal{S}}_1 \cap \overline{\mathcal{S}_2}, \overline{\mathcal{S}}_1 \cap \overline{\mathcal{S}_2}\right) \neq 0$ in Eqs. (C.9) and (C.10) to derive Eqs. (C.11) and (C.12); however, this property is not necessary for Theorem 4.1 to hold as well. The proof for the general case is beyond the scope of this work; however, in what follows, we provide a series of conjectures useful for proving the general scenario.

Conjecture 3. *Let $\mathcal{S} \subsetneq \{1, \cdots, n\}$ be an index set, and consider a uniformly random and invertible $n \times n$ matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$. Then, with probability of 1, we have:*

$$|\mathcal{S}| - \left|\overline{\mathcal{S}}\right| = \operatorname{rank}\left[\mathbf{U}\left(\mathcal{S}, \mathcal{S}\right)\right] \qquad \textit{if and only if} \qquad \left(\mathbf{U}^{-1}\right)\left(\overline{\mathcal{S}}, \overline{\mathcal{S}}\right) = \mathbf{0}. \tag{C.14}$$

Conjecture 4. *Let $\mathcal{S} \subsetneq \{1, \cdots, n\}$ be an index set, and consider a uniformly random and invertible $n \times n$ matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$. Then, with probability of 1, we have:*

$$|\mathcal{S}| - \operatorname{rank}\left[\mathbf{U}\left(\mathcal{S}, \mathcal{S}\right)\right] = \left|\overline{\mathcal{S}}\right| - \operatorname{rank}\left[\left(\mathbf{U}^{-1}\right)\left(\overline{\mathcal{S}}, \overline{\mathcal{S}}\right)\right]. \tag{C.15}$$

Corollary 1. *If $\operatorname{rank}\left[\mathbf{U}\left(\mathcal{S}, \mathcal{S}\right)\right] < |\mathcal{S}| - \left|\overline{\mathcal{S}}\right|$, then the matrix $\mathbf{U}$ is not invertible.*

Conjecture 5. *Let $\mathcal{S} \subsetneq \{1, \cdots, n\}$ be an index set, and consider a uniformly random and invertible $n \times n$ matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$. If the condition $|\mathcal{S}| - \left|\overline{\mathcal{S}}\right| = \operatorname{rank}\left[\mathbf{U}\left(\mathcal{S}, \mathcal{S}\right)\right]$ holds, then with probability of 1, we have:*

$$\sum_{i,j \in \mathcal{S}} \mathbf{U}(j,i) \times \mathbf{U}^{-1}(i,j) = \operatorname{rank}\left[\mathbf{U}\left(\mathcal{S}, \mathcal{S}\right)\right]. \tag{C.16}$$

Appendix D

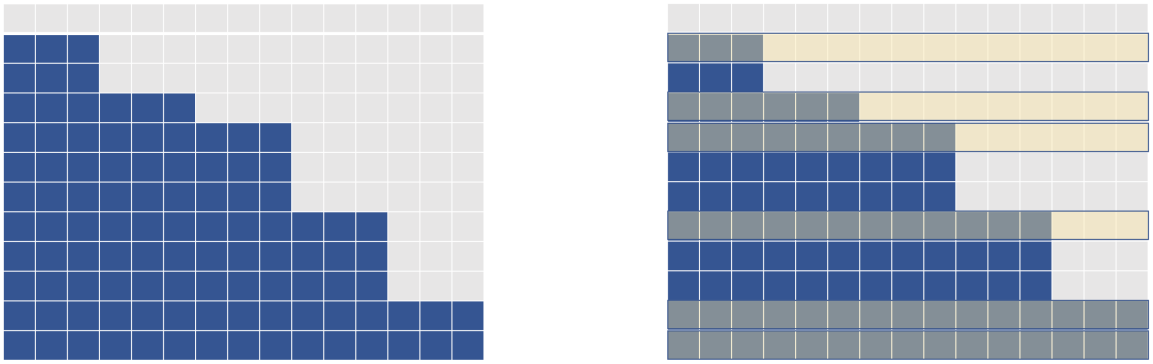
Proof of the Theorem 4.2

Using Theorem 4.1, we have:

$$\operatorname{rank}\left[\mathbf{C}_{\mathcal{S}_{j_i}} - \mathbf{C}_{\mathcal{S}_{j_{i+1}}}\right] = \min\left(\left|\mathcal{R}_{\mathcal{S}_{j_i},\mathcal{S}_{j_{i+1}}}\right|,\ k\right) = \min\left(\left|\{j_i, j_{i+1}\}\right|,\ k\right) = 2. \tag{D.1}$$

According to the proof of Theorem 4.1, not only we have $\operatorname{rank}\left[\mathbf{D}_{j_i,j_{i+1}}\right] = \left|\mathcal{R}_{\mathcal{S}_{j_i},\mathcal{S}_{j_{i+1}}}\right| = |\{j_i, j_{i+1}\}| = 2$, but also rows j_i and j_{i+1} of $\mathbf{D}_{j_i,j_{i+1}}$ are linearly independent. Hence, it readily follows that rows j_i and j_{i+1} of $\mathbf{C}_{\mathcal{S}_{j_i}} - \mathbf{C}_{\mathcal{S}_{j_{i+1}}} = \mathbf{D}_{j_i,j_{i+1}}\mathbf{A}$ are also linearly independent. This is because if we employ the proof by contradiction and use the fact that all other rows of $\mathbf{D}_{j_i,j_{i+1}}$ can be written as a linear combination of rows j_i and j_{i+1} of $\mathbf{D}_{j_i,j_{i+1}}$, then we obtain $\operatorname{rank}\left[\mathbf{C}_{\mathcal{S}_{j_i}} - \mathbf{C}_{\mathcal{S}_{j_{i+1}}}\right] = 1$ which contradicts Eq. (D.1).

Using Lemma 3 and the fact that elements of $\mathcal{J}$ are sorted in ascending order, it follows that the matrix $\mathbf{W}$ has the form shown via an example in Figure D.1(a), where the zero and non-zero elements are denoted via gray and blue colors, respectively, and non-zero elements follow a step-wise property. Since rows j_i and j_{i+1} of $\mathbf{C}_{\mathcal{S}_{j_i}} - \mathbf{C}_{\mathcal{S}_{j_{i+1}}}$ are linearly independent, and $\mathbf{W}$ is obtained by column-wise concatenation of matrices $\mathbf{C}_{\mathcal{S}_{j_i}} - \mathbf{C}_{\mathcal{S}_{j_{i+1}}}$ for all $i \in \{1, \cdots, |\mathcal{J}| - 1\}$, it follows that all rows of $\mathbf{W}$ that are within $\mathcal{J}$ are linearly



(a) (b)

Figure D.1: (a) The matrix $\mathbf{W}$ for a SEM-LV model with $n = 12$, $k = 3$, and $\mathcal{J} = \{2, 4, 5, 8, 11, 12\}$. (b) The set of six linearly independent rows of $\mathbf{W}$.

independent. These rows are shown in the exemplary model of Figure D.1(b). Therefore:

$$\text{rank}\left[\mathbf{W}\right] \geq \left|\mathcal{J}\right|. \tag{D.2}$$

Lemma 6. *Rank of the matrix* $\overline{\mathbf{D}} = \left[\mathbf{D}_{j_i,j_{i+1}}\ ,\ \mathbf{D}_{j_{i+1},j_{i+2}}\right]$*, that is formed by column-wise concatenation of matrices* $\mathbf{D}_{j_i,j_{i+1}}$ *and* $\mathbf{D}_{j_{i+1},j_{i+2}}$*, is equal to 3.*

Proof: According to the proof of Theorem 4.1*, we have* $\text{rank}\left[\mathbf{D}_{j_i,j_{i+1}}\right] = 2$*, and two columns* $\mathbf{D}_{j_i,j_{i+1}}(:, j_i)$ *and* $\mathbf{D}_{j_i,j_{i+1}}(:, j_{i+1})$ *are linearly independent. Similarly,* $\text{rank}\left[\mathbf{D}_{j_{i+1},j_{i+2}}\right] = 2$*, and two columns* $\mathbf{D}_{j_{i+1},j_{i+2}}(:, j_{i+1})$ *and* $\mathbf{D}_{j_{i+1},j_{i+2}}(:, j_{i+2})$ *are linearly independent. In order to show that the rank of* $\overline{\mathbf{D}}$ *is equal to three, we find three linearly independent columns such that their span includes all other columns of* $\overline{\mathbf{D}}$*.*

Note that all columns of $\mathbf{D}_{j_i,j_{i+1}}$ *can be expressed as a linear combination of two columns* $\vec{\mathbf{d}}_1 = \mathbf{D}_{j_i,j_{i+1}}(:, j_i)$ *and* $\vec{\mathbf{d}}_2 = \mathbf{D}_{j_i,j_{i+1}}(:, j_{i+1})$*. Similarly, all columns of* $\mathbf{D}_{j_{i+1},j_{i+2}}$ *are representable as a linear combination of columns* $\vec{\mathbf{d}}_3 = \mathbf{D}_{j_{i+1},j_{i+2}}(:, j_{i+1})$ *and* $\vec{\mathbf{d}}_4 = \mathbf{D}_{j_{i+1},j_{i+2}}(:, j_{i+2})$*. Therefore, the span of these four columns contains all columns of* $\overline{\mathbf{D}}$*, and we need to show that only three of these four columns are linearly independent to complete the proof. Since* $\mathbf{B}$ *is strictly lower triangular matrix, it follows from simple algebraic calculations that the first* $j_i - 1$ *and* $j_{i+1} - 1$ *elements of* $\vec{\mathbf{d}}_1$ *and* $\vec{\mathbf{d}}_2$ *are zero, respectively, and* $\vec{\mathbf{d}}_1(j_i) = -1$ *and* $\vec{\mathbf{d}}_2(j_{i+1}) = 1$*.*

Similarly, the first $j_{i+1}-1$ and $j_{i+2}-1$ elements of $\vec{\mathbf{d}}_3$ and $\vec{\mathbf{d}}_4$ are zero, respectively, and $\vec{\mathbf{d}}_3(j_{i+1}) = -1$ and $\vec{\mathbf{d}}_4(j_{i+2}) = 1$. Using this property and the fact that $j_i < j_{i+1} < j_{i+2}$, it readily follows that $\vec{\mathbf{d}}_1$, $\vec{\mathbf{d}}_2$ and $\vec{\mathbf{d}}_4$ are linearly independent, and $\vec{\mathbf{d}}_1$, $\vec{\mathbf{d}}_3$ and $\vec{\mathbf{d}}_4$ are also linearly independent. Hence, we need to demonstrate that $\vec{\mathbf{d}}_3$ is in the span of three linearly independent vectors $\vec{\mathbf{d}}_1$, $\vec{\mathbf{d}}_2$ and $\vec{\mathbf{d}}_4$. Similar argument can be carried out to show that $\vec{\mathbf{d}}_2$ is in the span of three linearly independent vectors $\vec{\mathbf{d}}_1$, $\vec{\mathbf{d}}_3$ and $\vec{\mathbf{d}}_4$.

From the definition, we have:

$$\begin{aligned}
\mathbf{D}_{j_i,j_{i+1}} + \mathbf{D}_{j_{i+1},j_{i+2}} &= \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_{j_i}}\mathbf{B}\right)^{-1}\mathbf{P}_{\mathcal{S}_{j_i}} - \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_{j_{i+1}}}\mathbf{B}\right)^{-1}\mathbf{P}_{\mathcal{S}_{j_{i+1}}} \\
&\quad + \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_{j_{i+1}}}\mathbf{B}\right)^{-1}\mathbf{P}_{\mathcal{S}_{j_{i+1}}} - \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_{j_{i+2}}}\mathbf{B}\right)^{-1}\mathbf{P}_{\mathcal{S}_{j_{i+2}}} \\
&= \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_{j_i}}\mathbf{B}\right)^{-1}\mathbf{P}_{\mathcal{S}_{j_i}} - \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_{j_{i+2}}}\mathbf{B}\right)^{-1}\mathbf{P}_{\mathcal{S}_{j_{i+2}}} \\
&= \mathbf{D}_{j_i,j_{i+2}}, \qquad \text{(D.3)}
\end{aligned}$$

and from the proof of Theorem 4.1 *we know that* $\operatorname{rank}\left[\mathbf{D}_{j_i,j_{i+2}}\right] = 2$ *and columns j_i and j_{i+2} are linearly independent. Therefore, we can write:*

$$\begin{aligned}
\vec{\mathbf{d}}_2 + \vec{\mathbf{d}}_3 &= \mathbf{D}_{j_i,j_{i+2}}(:, j_{i+1}) \\
&= t_1 \times \mathbf{D}_{j_i,j_{i+2}}(:, j_i) + t_2 \times \mathbf{D}_{j_i,j_{i+2}}(:, j_{i+2}) \\
&= t_1 \times \left(\vec{\mathbf{d}}_1 + \mathbf{D}_{j_{i+1},j_{i+2}}(:, j_i)\right) + t_2 \times \vec{\mathbf{d}}_4 \\
&= t_1 \times \left(\vec{\mathbf{d}}_1 + \left[s_1 \times \vec{\mathbf{d}}_3 + s_2 \times \vec{\mathbf{d}}_4\right]\right) + t_2 \times \vec{\mathbf{d}}_4, \qquad \text{(D.4)}
\end{aligned}$$

where we have used the property that column c of $\mathbf{D}_{j_i,j_{i+1}}$ is zero for all $j_{i+1} < c \leq n$. This property follows from simple algebraic calculations and the fact that $\mathbf{B}$ is a strictly lower triangular matrix. Eq. (D.4) implies that $\vec{\mathbf{d}}_3$ can be written as a linear combination of $\vec{\mathbf{d}}_1$, $\vec{\mathbf{d}}_2$, and $\vec{\mathbf{d}}_4$, and the proof for Lemma 6 is complete.

Note that we can rewrite $\mathbf{W}$ as:

$$\mathbf{W} = \Big[\mathbf{D}_{j_1,j_2}\ ,\ \mathbf{D}_{j_2,j_3}\ ,\ \cdots\ ,\ \mathbf{D}_{j_{|\mathcal{J}|-1},j_{|\mathcal{J}|}}\Big] \times \begin{bmatrix} \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \cdots & \mathbf{0} \\ \vdots & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A} \end{bmatrix} = \overline{\mathbf{D}} \times \overline{\mathbf{A}}. \tag{D.5}$$

By successively applying Lemma 6 to the matrix $\overline{\mathbf{D}} = \Big[\mathbf{D}_{j_1,j_2}\ ,\ \mathbf{D}_{j_2,j_3}\ ,\ \cdots\ ,\ \mathbf{D}_{j_{|\mathcal{J}|-1},j_{|\mathcal{J}|}}\Big]$, we obtain $\operatorname{rank}\left[\ \overline{\mathbf{D}}\ \right] = |\mathcal{J}|$. Therefore, Eq. (D.5) implies that:

$$\begin{aligned} \operatorname{rank}[\mathbf{W}] &\leq \min\left(\operatorname{rank}\left[\ \overline{\mathbf{D}}\ \right], \operatorname{rank}\left[\ \overline{\mathbf{A}}\ \right]\right) \\ &= \min\left(|\mathcal{J}|\,,\ k \times (|\mathcal{J}| - 1)\right) \\ &= |\mathcal{J}|\,. \end{aligned} \tag{D.6}$$

Eqs. (D.2) and (D.6) indicate that $\operatorname{rank}[\mathbf{W}] = |\mathcal{J}|$ and the proof is complete. ■

Appendix E

Proof of the Lemma 3

Using Eq. (2.7), we have:

$$\mathbf{C}_{\mathcal{S}_1} = \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B}\right)^{-1}\mathbf{P}_{\mathcal{S}_1}\mathbf{A}, \tag{E.1}$$

$$\mathbf{C}_{\mathcal{S}_2} = \left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B}\right)^{-1}\mathbf{P}_{\mathcal{S}_2}\mathbf{A}. \tag{E.2}$$

First, we show that the first $\min\left(\mathcal{S}_1 \cup \mathcal{S}_2\right)-1$ rows of matrices $\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B}\right)^{-1}$ and $\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B}\right)^{-1}$ are the same. Since $\mathbf{B} = [b_{i,j}]_{1\leq i,j\leq n}$ is a strictly lower triangular matrix, then both $\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B}\right)$ and its inverse, i.e. $\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B}\right)^{-1}$, are also lower triangular matrices. Similarly, both $\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B}\right)$ and its inverse, i.e. $\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B}\right)^{-1}$, are lower triangular matrices. Therefore, we only need to compare the elements in row i and column j of $\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B}\right)^{-1}$ and $\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B}\right)^{-1}$, where $j < i \leq \min\left(\mathcal{S}_1 \cup \mathcal{S}_2\right) - 1$. Using Cramer's rule, we have:

$$\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B}\right)^{-1} = \frac{1}{\det\left[\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B}\right)\right]} \times \operatorname{adj}\left[\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B}\right)\right], \tag{E.3}$$

$$\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B}\right)^{-1} = \frac{1}{\det\left[\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B}\right)\right]} \times \operatorname{adj}\left[\left(\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B}\right)\right]. \tag{E.4}$$

Note that since both $(\mathbf{I}-\mathbf{P}_{\mathcal{S}_1}\mathbf{B})$ and $(\mathbf{I}-\mathbf{P}_{\mathcal{S}_2}\mathbf{B})$ are lower triangular matrices with diagonal entries equal to 1, we have: $\det\left[(\mathbf{I}-\mathbf{P}_{\mathcal{S}_1}\mathbf{B})\right]=\det\left[(\mathbf{I}-\mathbf{P}_{\mathcal{S}_2}\mathbf{B})\right]=1$. Using straightforward algebraic calculations and the fact that $j<i\leq\min\left(\mathcal{S}_1\cup\mathcal{S}_2\right)-1$, it readily follows that the element in row i and column j of both adjugate matrices $\text{adj}\left[(\mathbf{I}-\mathbf{P}_{\mathcal{S}_1}\mathbf{B})\right]$ and $\text{adj}\left[(\mathbf{I}-\mathbf{P}_{\mathcal{S}_2}\mathbf{B})\right]$ is equal to $(-1)^{i+j}\times b_{j+1,j}\times b_{j+2,j+1}\times\cdots\times b_{i,i-1}$. Hence, the first $\min\left(\mathcal{S}_1\cup\mathcal{S}_2\right)-1$ rows of matrices $(\mathbf{I}-\mathbf{P}_{\mathcal{S}_1}\mathbf{B})^{-1}$ and $(\mathbf{I}-\mathbf{P}_{\mathcal{S}_2}\mathbf{B})^{-1}$ in Eqs. (E.3) and (E.4) are the same, which concludes our first aim.

Second, we show that the elements in row i and column j of response matrices $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ are the same, where $i\leq\min\left(\mathcal{S}_1\cup\mathcal{S}_2\right)-1$. As we showed above, the $i-$th row of matrices $(\mathbf{I}-\mathbf{P}_{\mathcal{S}_1}\mathbf{B})^{-1}$ and $(\mathbf{I}-\mathbf{P}_{\mathcal{S}_2}\mathbf{B})^{-1}$ in Eqs. (E.1) and (E.2) are the same. Moreover, the first i elements in the $j-$th column of matrices $\mathbf{P}_{\mathcal{S}_1}\mathbf{A}$ and $\mathbf{P}_{\mathcal{S}_2}\mathbf{A}$ are equal, for each $j\in\{1,\cdots,k\}$. Since elements above the diagonal of matrices $(\mathbf{I}-\mathbf{P}_{\mathcal{S}_1}\mathbf{B})^{-1}$ and $(\mathbf{I}-\mathbf{P}_{\mathcal{S}_2}\mathbf{B})^{-1}$ are zero, it directly follows that the first $\min\left(\mathcal{S}_1\cup\mathcal{S}_2\right)-1$ rows of both response matrices are the same, and the proof is complete. ■

Appendix F

Proof of the Lemma 4

We can rewrite the Eq. (2.7) for the two actions as follows:

$$\mathbf{P}_{\mathcal{S}_1}\mathbf{A} = (\mathbf{I} - \mathbf{P}_{\mathcal{S}_1}\mathbf{B})\,\mathbf{C}_{\mathcal{S}_1}, \tag{F.1}$$

$$\mathbf{P}_{\mathcal{S}_2}\mathbf{A} = (\mathbf{I} - \mathbf{P}_{\mathcal{S}_2}\mathbf{B})\,\mathbf{C}_{\mathcal{S}_2}. \tag{F.2}$$

Since excess kurtosis values for columns of response matrices are normalized to ± 1, elements in row t_1 of $\mathbf{C}_{\mathcal{S}_1}$ and $\mathbf{C}_{\mathcal{S}_2}$ are the same up to permutation and sign, where $t_1 \in \{i \mid \mathbf{P}_{\mathcal{S}_1}(i,i) = \mathbf{P}_{\mathcal{S}_2}(i,i) = 1\}$. Therefore, elements in row t_1 of $\mathbf{P}_{\mathcal{S}_1}\mathbf{A}$ and $\mathbf{P}_{\mathcal{S}_2}\mathbf{A}$ should be the same up to their permutation and sign. Note that since the first row of $\mathbf{C}_{\mathcal{S}_2}$ is zero, Eq. (F.2) yields the elements in row t_1 of $\mathbf{P}_{\mathcal{S}_2}\mathbf{A}$, which we refer to as the true row t_1 of $\mathbf{A}$ and denote it by $\vec{\mathbf{a}}_{t_1}$, regardless of the value of $\mathbf{B}(t_1, 1)$. Now, for a fixed ordering of the columns of $\mathbf{C}_{\mathcal{S}_1}$, two solutions $b'_{t_1,1}$ and $b''_{t_1,1}$ yield two candidates for row t_1 of $\mathbf{A}$, which we denote by $\vec{\mathbf{a}}'_{t_1}$ and $\vec{\mathbf{a}}''_{t_1}$, respectively. Note that since the ordering for columns of $\mathbf{C}_{\mathcal{S}_1}$ is fixed, $i-$th element of $\vec{\mathbf{a}}'_{t_1}$ corresponds to the $i-$th element of $\vec{\mathbf{a}}''_{t_1}$. Now, if both solutions $b'_{t_1,1}$ and $b''_{t_1,1}$ are valid, since there exist an index $j \in \{1, \cdots, k\}$ such that both these elements correspond to the $j-$th element of $\vec{\mathbf{a}}_{t_1}$, it follows that $\vec{\mathbf{a}}'_{t_1}(i) = \pm\vec{\mathbf{a}}''_{t_1}(i)$ for all $i \in \{1, \cdots, k\}$.

Now, if $\vec{\mathbf{a}}'_{t_1}(i) = \vec{\mathbf{a}}''_{t_1}(i)$ for some index i, it follows that either $b'_{t_1,1} = b''_{t_1,1}$, which concludes the proof, or $\mathbf{C}_{\mathcal{S}_1}(1,i) = 0$. Furthermore, for any two indices i_1 and i_2 such that $\vec{\mathbf{a}}'_{t_1}(i_1) = -\vec{\mathbf{a}}''_{t_1}(i_1)$ and $\vec{\mathbf{a}}'_{t_1}(i_2) = -\vec{\mathbf{a}}''_{t_1}(i_2)$, it follows from simple algebraic calculations that $\frac{\mathbf{C}_{\mathcal{S}_1}(1,i_1)}{\vec{\mathbf{a}}'_{t_1}(i_1)} = \frac{\mathbf{C}_{\mathcal{S}_1}(1,i_2)}{\vec{\mathbf{a}}'_{t_1}(i_2)}$ and $\frac{\mathbf{C}_{\mathcal{S}_1}(1,i_1)}{\vec{\mathbf{a}}''_{t_1}(i_1)} = \frac{\mathbf{C}_{\mathcal{S}_1}(1,i_2)}{\vec{\mathbf{a}}''_{t_1}(i_2)}$. Therefore, both candidate rows $\vec{\mathbf{a}}'_{t_1}$ and $\vec{\mathbf{a}}''_{t_1}$ are proportional to the first row of $\mathbf{C}_{\mathcal{S}_1}$. Similar observation is true for the two candidate rows corresponding to another index $t_2 \in \{i \mid \mathbf{P}_{\mathcal{S}_1}(i,i) = \mathbf{P}_{\mathcal{S}_2}(i,i) = 1\}$. An immediate implication of this is that if both pairs of solutions $\left\{b'_{t_1,1}, b''_{t_1,1}\right\}$ and $\left\{b'_{t_2,1}, b''_{t_2,1}\right\}$ are valid, then any of the four possible pairs of rows t_1 and t_2 of $\mathbf{A}$ will be linearly dependent which is in contradiction with Assumption 1. Therefore, there exists at most one index, say t_1, for which both solutions $b'_{t_1,1}$ and $b''_{t_1,1}$ are valid.

Now, if we restrict matrices $\mathbf{P}_{\mathcal{S}_1}\mathbf{A}$ and $\mathbf{P}_{\mathcal{S}_2}\mathbf{A}$ in Eqs. (F.1) and (F.2) to row indices in $\{i \mid \mathbf{P}_{\mathcal{S}_1}(i,i) = \mathbf{P}_{\mathcal{S}_2}(i,i) = 1\}$, we obtain two submatrices which we denote by $\widehat{\mathbf{P}_{\mathcal{S}_1}\mathbf{A}}$ and $\widehat{\mathbf{P}_{\mathcal{S}_2}\mathbf{A}}$, respectively. Similar to Eq. (5.9), we construct the matrix $\mathbf{\Theta} = [\theta_{i,j}]_{1 \le i,j \le k}$ where $\theta_{i,j}$ is the angle between columns i of $\widehat{\mathbf{P}_{\mathcal{S}_1}\mathbf{A}}$ and column j of $\widehat{\mathbf{P}_{\mathcal{S}_2}\mathbf{A}}$. Note that the two solutions $b'_{t_1,1}$ and $b''_{t_1,1}$ can only alter the row t'_1 of $\widehat{\mathbf{P}_{\mathcal{S}_2}\mathbf{A}}$, in the sense of negating the sign of some elements, that corresponds to row t_1 in $\mathbf{P}_{\mathcal{S}_2}\mathbf{A}$. Therefore, if one of the solutions, say $b'_{t_1,1}$ is valid, then the other solution, i.e. $b''_{t_1,1}$, that flips the sign of some elements in row t'_1 of $\widehat{\mathbf{P}_{\mathcal{S}_2}\mathbf{A}}$ but keeps all other rows unchanged, cannot lead to a valid solution, i.e. exactly one 0 or π in each row and column of $\mathbf{\Theta}$, and the proof is complete. ■