Bernoulli **32**(2), 2026, 1356–1382
https://doi.org/10.3150/25-BEJ1909

Causal inference on process graphs: Causal structure and effect identification

NICOLAS-DOMENIC REITER[1,a], JONAS WAHL[2,b], ANDREAS GERHARDUS[3,c] and JAKOB RUNGE[4,d]

[1] *TUM School of Computation, Information and Technology, Technical University of Munich, Munich, Germany,* [a] *nicolas-domenic.reiter@tum.de*
[2] *Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarbrücken, Germany,* [b] *jonas.wahl@dfki.de*
[3] *German Aerospace Center (DLR), Institute of Data Science, Jena, Germany,* [c] *andreas.gerhardus@dlr.de*
[4] *Department of Computer Science, University of Potsdam, Potsdam, Germany,* [d] *jakob.runge@uni-potsdam.de*

A structural vector autoregressive (SVAR) process is a linear causal model for variables that evolve over a discrete set of time points and between which there may be lagged and instantaneous effects. The qualitative causal structure of an SVAR process can be represented by its finite and directed process graph, where a directed link connects two processes whenever there is a lagged or instantaneous effect between them. At the process graph level, the causal structure of SVAR processes is compactly parameterized in the frequency domain. In this paper, we consider the problem of causal discovery and causal effect estimation from the spectral density, the frequency domain analogue of the autocovariance, of the SVAR process. Causal discovery concerns the recovery of the process graph. Causal effect estimation concerns the identification and estimation of frequency domain causal effects. We show that information about the process graph, in terms of d- and t-separation, is generically characterized by algebraic constraints on the spectral density. Furthermore, we introduce a notion of rational identifiability for frequency domain causal effects between processes that may be confounded by exogenous latent processes, and show that the recent graphical latent factor half-trek criterion can be used on the process graph to assess whether a given (confounded) effect can be identified by rational operations on the entries of the spectral density.

Keywords: Causal inference; graphical models; identifiability; spectral density; SVAR processes

1. Introduction

Questions such as "what are the causes of a given phenomenon?" or "how strong is the effect of one phenomenon on another?" are asked and studied in fields as diverse as climate science [13,32,47], social science and economics [4,29], and neuroscience [48,49]. Causal inference [37,38] formalizes these questions mathematically so that they can be analyzed at an abstract level.

Causal models are the mathematical object with which to formalize causal questions. They consist of a (causal) graph and a compatible stochastic functional model, commonly called *Structural Causal Model (SCM)* [37,38]. The vertices on the causal graph represent the entities among which causal relationships are modeled and a directed edge between two vertices indicates an immediate cause-effect relation between the corresponding entities. The functional model assigns a stochastic object to each vertex on the causal graph, along with a set of functions that specify how each stochastic object is influenced by its causes. One goal in causal inference is to understand conditions on the causal model under which certain aspects of the model can be reconstructed from observational data. This goal splits into two main research directions of which the first is known as *causal structure identification* and the second as *causal effect identification*. Causal structure identification concerns the reconstruction of the causal graph from observational data, whereas causal effect identification is about extracting information about the functional model from observations, given at least partial knowledge of the causal graph.

 © 2026 ISI/BS

An important and extensively studied class of causal models are the linear SCMs [5], which are specified by a noise term for each vertex and a coefficient for each link on the causal graph. The observational distribution of a linear SCM is partly described by the covariance of the model. Much research has gone into understanding what information about the the causal model can be recovered from the covariance of the SCM. Since the entries in the covariance are rational functions of the parameters of the linear model, this problem is effectively approached from an algebraic point of view [10,12], where causal structure identification can be seen as the problem of characterizing information about the causal graph in terms of algebraic relations in the covariance matrix. Results in this fashion include the algebraic characterization of d- and t-separation in the underlying causal graph [19,37,38] resp. [52]. Covariance-based causal effect identification is concerned with the question of whether a given causal effect can be expressed as a function of the entries in the covariance matrix. This is an interesting problem when some of the variables are hidden, i.e. not directly included in the covariance matrix. Algebraically, this means that the causal model for the observed variables must be recovered from a submatrix of the covariance. The structure of the causal graph (including the hidden variables) imposes algebraic relations between the entries in the covariance, and these relations allow or prevent certain aspects of the observable causal model to be recovered. In this regard, several graphical criteria have been identified for deciding whether, in a given causal graph, a specific causal effect can be identified from the covariance for almost all linear SCMs compatible with the graph [3,11,17,55]. In order to apply these criteria to identify causal effects from observational data it is necessary that the observations are identically distributed and independent of each other.

However, in many practical applications, observations are time-dependent [18,32,45–47]. In such a situation it is necessary to have a causal model that takes time into account. A simplified way of incorporating temporal dependencies into the causal model is to assume that the system evolves over a discrete set of time points. The causal structure in such systems can be represented graphically by drawing a vertex for each process at every time point, so that a directed edge between vertices expresses a delayed or simultaneous effect. This construction yields an infinite causal graph. Graphs of this particular type are well established in the literature [8,21,38,45] and are, among other names, often referred to as *time series graphs*. An infinite time series graph can be reduced to a finite graph by projecting away the time coordinate [2,38,45]. In this paper, we call this reduced graph the *process graph* of the time series graph.

Structural vector autoregressive (SVAR) processes are theoretically well understood [7,31] and widely used across scientific disciplines to model dynamically interacting quantities [18,48]. SVAR processes can be seen as a linear parametrization of an appropriate causal time series graph. This leads to the question of how to systematically formalize the questions of causal structure identification and causal effect identification for linear causal time series models. The effectiveness and clarity of the algebraic perspective on causal inference for linear causal models with iid observations motivates the search for a similar formulation for linear time series models. The time series graph, together with its linear SVAR parametrization, appears to be the natural choice for a causal model to be identified from observations [21–23,33,38,44–46,53]. Taking into account the specific structure of the time series graph has the advantage that it may be possible to recover causal information from observations that seem unidentifiable when considering the process graph alone, such as cyclic relationships between processes. However, for a given process graph, there are infinitely many time series graphs that collapse into it, meaning that the time series graph can be considerably more complex than its process graph.

In an SVAR process, the causal structure dictated by the process graph, yet independent of the underlying time series, can be captured and analyzed with the *structural equation process (SEP)* representation [40]. In the SEP representation, each vertex is associated with a stochastic process called the internal dynamics, which is similar to the noise term in a linear SCM, and each edge on the process graph is associated with a filter. This linear causal model of stochastic processes can be equivalently

represented in the frequency domain, where causal effects can be computed by means of a generalized path rule. The SEP representation is also relevant to applications because some questions are more naturally formulated in the frequency domain, which has been demonstrated in [41] with an analysis of the causal effect of the 10-11 year solar cycle on the North Atlantic Oscillation (NAO).

In this paper, we use the frequency representation of an SVAR process to give an algebraic formulation of the causal structure and causal effect identification problem in the frequency domain. The structure of this paper is as follows. In Section 2, we recall the necessary concepts on SVAR processes as causal models for dynamically interacting variables. The object from which we seek to identify the frequency representation of the causal model, i.e. the process graph together with the Fourier transformed link filters, is the *spectral density* or *spectrum* of the process. In Section 3, for a fixed time series graph, we parameterize the spectrum of a SVAR process as a matrix over the field of rational functions with real coefficients. Based on this parameterization, we develop the main results of this paper.

Causal structure identification. Section 4 concerns the identification of the process graph of an SVAR process $\mathbb{X}$ from its spectrum $\mathcal{S}$. Our first result is the generic characterization of d-separation on the process graph.

Theorem (d-separation of processes). *Let $G = (V, D)$ be an acyclic process graph, and let $X, Y, Z \subset V$ be three disjoint sets of process indices. For a generic SVAR process compatible with G and every frequency $z \in S^1$ it holds that*

$$\operatorname{rank}([\mathcal{S}]_{Z\cup X, Z\cup Y}) = \operatorname{rank}([\mathcal{S}(z)]_{Z\cup X, Z\cup Y}) = |Z| \iff X \perp_d Y | Z,$$

where $X \perp_d Y|Z$ denotes that X and Y are d-separated by Z.

Our second result generalizes [52][Theorem 2.8] to SVAR processes.

Theorem (t-separation of processes). *Let $G = (V, D)$ be an acyclic process graph and $X, Y \subset V$ be two sets of processes (indices). Then, for a generic SVAR process compatible with G and every frequency $z \in S^1$ it holds that*

$$\operatorname{rank}([\mathcal{S}]_{X,Y}) = \operatorname{rank}([\mathcal{S}(z)]_{X,Y}) = \min\{|C| : \ C \text{ } t\text{-separates } X \text{ and } Y\}.$$

Causal effect identification. Section 5 is devoted to the identification of causal effects in the frequency domain. If a causal effect in the frequency domain can generically be expressed through addition/subtraction and multiplication/division of the entries in the spectrum, then we call that effect rationally identifiable. In this paper, we make this notion more precise and, in so doing, generalize the rational identifiability term that is known for linear Gaussian SCMs. We will conclude this section by showing that the recent latent factor half-trek criterion for linear SCMs [3][Theorem 3.7] can be applied to the process graph to decide whether a given frequency domain causal effect is rationally identifiable.

2. Preliminaries

Let V be a finite set of process indices and $\mathbb{Z}$ the set of integers. A (stationary) time series graph is an infinite graph $\mathcal{G} = (V \times \mathbb{Z}, \mathcal{D})$ such that the set of directed edges $\mathcal{D}$ satisfies the following requirement: If there is a directed edge $(v(t-k), w(t)) \in \mathcal{D}$, then it must hold that $k \geq 0$ and $\{(v(s-k), w(s)) \mid s \in \mathbb{Z}\} \subset \mathcal{D}$. In this case we say that v is causing w at lag k, which we abbreviate with the notation $v \to_k w$. If the lag $k = 0$, then $v \to_0 w$ represents a *direct contemporaneous effect*. The requirement $k \geq 0$ makes

sure that a cause cannot happen after its effect. The *order* p of the time series graph $\mathcal{G}$ is the maximum time lag of a link on $\mathcal{G}$, i.e. $p \coloneqq \max_{v,w\in V}\{k \mid v \to_k w\} \in \mathbb{N}_0 \cup \{\infty\}$. If $p < \infty$, then $\mathcal{G}$ is of finite order, otherwise it is of infinite order. Throughout the main paper we always assume $p < \infty$.

Collapsing the infinite time series graph $\mathcal{G}$ along the time axis yields its finite process graph $G = (V, D)$. This graph contains a directed edge $(v, w) \in D$, which we abbreviate as $v \to w$, if and only if $v \neq w$ and there is some $k \geq 0$ such that $v \to_k w$ is a contemporaneous or lagged causal relation on the time series graph $\mathcal{G}$. For a directed graph $G = (V, D)$, we denote by $\mathcal{T}(G)$ the infinite set of all time series graphs that have process graph G. The parent set of $v \in V$, denoted by $\mathrm{pa}(v)$, consists of all the processes u such that $u \to v$.

A stochastic process $\mathbb{X} = (\mathbb{X}_v(t))_{v\in V, t\in\mathbb{Z}}$ is said to be a structural vector autoregressive (SVAR) process compatible with a time series graph $\mathcal{G}$ if there exists a collection of coefficients $\Phi = (\phi_{v,w}(k))_{v\to_k w}$ indexed by the contemporaneous and lagged relations in $\mathcal{D}$, such that for every $v \in V$ and every $t \in \mathbb{Z}$ the associated random variable satisfies the structural equation

$$\mathbb{X}_v(t) = \sum_{u\in pa(v)} \sum_{k:u\to_k v} \phi_{u,v}(k)\mathbb{X}_u(t-k) + \eta_v(t), \tag{1}$$

where $\eta = (\eta_v(t))_{v\in V, t\in\mathbb{Z}}$ is a mutually independent zero-mean stationary white noise process with positive and finite variance, i.e. all $\eta_v(t)$ are iid zero-mean random variables whose components are mutually independent. In particular, the covariance of the random vector $\eta(t)$ is a positive diagonal matrix $\mathrm{diag}(\mathbb{E}[\eta_v\eta_v])_{v\in V})$. Given a zero-mean random vector or white noise process η with mutually independent components, we get a family of random vectors resp. white noise processes with mutually independent components parameterized by the set of positive diagonal matrices with positive diagonal $\mathbb{R}_+^V$, i.e. the set $\{\Omega\eta \mid \Omega \in \mathbb{R}_+^V\}$.

In order for $\mathbb{X}$ to define a stationary process, the parameter vector Φ needs to satisfy a stability condition. The precise formulation is deferred to Section 3.

Remark 1. For the remainder of this paper, we assume that the time series graph $\mathcal{G}$ has finite order. However, in the supplementary material [42][Appendix B] we define sets of infinite order SVAR processes that are specified by finitely many parameters. We would like to emphasize that all the results we develop in this paper also apply to such infinite order SVAR processes.

2.1. The frequency representation of a SVAR process

In this subsection, we will give a brief overview of how an SVAR process can be represented as an SEP, as developed in [40]. Details on the underlying construction can also be found in the supplementary material [42][Appendix A]. A structural equation process is specified by a linear filter $\lambda_{v,w} = \big(\lambda_{v,w}(s)\big)_{s\in\mathbb{Z}}$ for each link $v \to w$ on the process a graph, and for every $v \in V$ a univariate process $\mathbb{X}_v^{\mathrm{I}} = \big(\mathbb{X}_v^{\mathrm{I}}(t)\big)_{t\in\mathbb{Z}}$ called the internal dynamics of v. These filters and processes are such that the internal dynamics are mutually independent, and for every $v \in V$ it holds that

$$\mathbb{X}_v = \sum_{u\in pa(v)} \lambda_{u,v} * \mathbb{X}_u + \mathbb{X}_v^{\mathrm{I}}, \tag{2}$$

where $*$ denotes convolution of sequences. If we arrange the filters $(\lambda_{u,v})_{u\to v}$ into a matrix valued filter Λ, then we obtain the following system of linear equations for the stochastic processes:

$$\mathbb{X} = \Lambda^\top * \mathbb{X} + \mathbb{X}^{\mathrm{I}} = \left(\sum_{k\geq 0} \Lambda^k\right)^\top * \mathbb{X}^{\mathrm{I}}. \tag{3}$$

This representation exists under certain conditions on Φ, as detailed in [40]. For convenience of the reader we add the proof of this result to the supplementary material [42][Appendix A].

The contemporaneous and lagged covariance information in the processes $\mathbb{X}$ resp. $\mathbb{X}^{\mathrm{I}}$ are captured in their *autocovariance sequence (ACS)*, which we denote by $\Sigma = (\Sigma(k))_{k\in\mathbb{Z}}$ resp. $\Sigma^{\mathrm{I}} = (\Sigma^{\mathrm{I}}(k))_{k\in\mathbb{Z}}$. For $k \in \mathbb{Z}$ the corresponding entries in the ACS are

$$\Sigma(k) \coloneqq \mathbb{E}[\mathbb{X}(t)\mathbb{X}(t-k)^\top], \qquad \Sigma^{\mathrm{I}}(k) \coloneqq \mathbb{E}[\mathbb{X}^{\mathrm{I}}(t)\mathbb{X}^{\mathrm{I}}(t-k)^\top].$$

Since the internal dynamics are mutually independent it holds that $\Sigma^{\mathrm{I}}_{v,w}(k) = 0$ for every $k \in \mathbb{Z}$ if $v \neq w$. The structural process formulation (3) yields the following expression for the ACS:

$$\Sigma = \left(\sum_{k\geq 0} \Lambda^k\right)^\top * \Sigma^{\mathrm{I}} \hat{*} \left(\sum_{k\geq 0} \Lambda^k\right), \tag{4}$$

where $\hat{*}$ denotes tilted convolution of sequences, see [40] or [42][Appendix A].

Every stationary stochastic process admits an equivalent description in the frequency domain [7][Chapter 11.8], and it turns out that the frequency domain offers a convenient perspective to reason about the causal structure of SVAR process at the level of its process graph. The frequency domain analogue of the ACS is the *spectral density*, or *spectrum* for short, $\mathcal{S} \coloneqq \mathcal{F}(\Sigma)$, where $\mathcal{F}$ is the Fourier transformation. Recall that the Fourier transform maps sequences to complex valued functions that are defined on the complex unit circle $S^1 \subset \mathbb{C}$. Elementary properties of the Fourier transformation together with Equation (4) give the identity

$$\mathcal{S} = (I - H)^{-\top} \mathcal{S}^{\mathrm{I}} (I - H^*)^{-1}, \tag{5}$$

where $H = \mathcal{F}(\Lambda)$ is the *direct transfer function*, as shown in [40]. The entries in the direct transfer function are rational functions parameterized by the SVAR coefficients Φ. To write down this parametrization, let us define for a pair of process indices $x, y \in V$ the associated *lag polynomial*

$$\varphi_{x,y}(z) \coloneqq \sum_{k:x\to_k y} \phi_{x,y}(k) z^k.$$

Suppose $v \to w$ is a direct link on the process graph, then its associated *link function* evaluated at $z \in S^1$ is

$$h_{v,w}(z) \coloneqq \mathcal{F}(\lambda_{v,w})(z) = \frac{\varphi_{v,w}(z)}{1 - \varphi_{w,w}(z)}, \tag{6}$$

see [40][Lemma 1]. Also the entries in the spectral density of the internal dynamics $\mathcal{S}^{\mathrm{I}} \coloneqq \mathcal{F}(\Sigma^{\mathrm{I}})$ evaluated at a $z \in S^1$ admit a direct parametrization in terms of the SVAR coefficients. Specifically, the spectral density of the internal dynamics of $v \in V$ is given by

$$\mathcal{S}^{\mathrm{I}}_v(z) = \frac{\omega_v}{(1 - \varphi_{v,v}(z))(1 - \varphi_{v,v}(z^*))}, \tag{7}$$

where ω_v is the variance of the white noise process η_v. The pairwise independence of the internal dynamics implies that $\mathcal{S}^{\mathrm{I}}_{v,w}(z) = 0$ for every $z \in S^1$ if $v \neq w$.

Example 1. Considering the example shown in Figure 1, we illustrate the computation of the direct link function for the link $u \to v$ on the process graph. The coefficients $\phi_{u,v}(0)$ and $\phi_{u,v}(1)$ parameterize

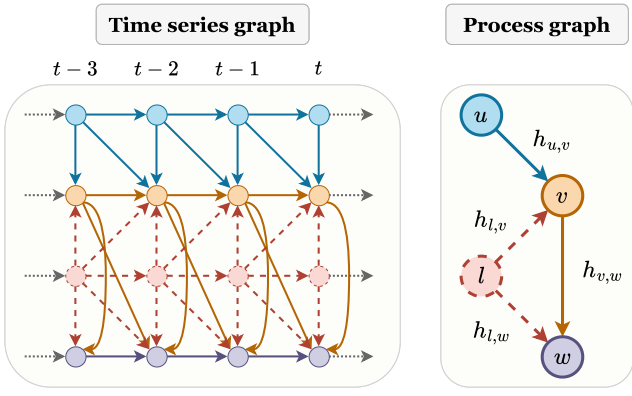

Figure 1. The figure on the left shows an excerpt of an order one time series graph for the processes $V = \{u, v, h, w\}$. The graph on the right displays the associated process graph.

the contemporaneous and lag-one effect of u on v. Similarly, the coefficient $\phi_{v,v}(1)$ quantifies the effect with which the process v carries information from one point in time to the next. So, according to equation (6), the transfer function associated with the edge $u \to v$ is

$$h_{u,v}(z) = \frac{\phi_{u,v}(0) + \phi_{u,v}(1)z}{1 - \phi_{v,v}(1)z}.$$

Regarding the internal spectrum of v, we use equation (7) to compute

$$\mathcal{S}^{\mathrm{I}}_u(z) = \frac{\omega_u}{1 - 2\phi_{u,u}(1)\mathrm{Re}(z) + \phi_{u,u}(1)^2},$$

where $z = \mathrm{Re}(z) + i\mathrm{Im}(z) \in S^1$.

2.2. The spectral trek rule

The algebraic relations between the entries in the spectrum are constrained by the process graph G. In the following, we recall how the process graph structures the spectrum using the *spectral trek rule* [40], a generalization of the classical trek rule [56,57].

Definition 1 (Directed paths and treks). A *directed path* π on G is an ordered sequence of links $\pi = v_1 \to \cdots \to v_n$. The node v_1 is called the starting point of π, and the node v_n is called the endpoint of π. A path π' is said to be a *sub-path* of π, denoted as $\pi' \leq \pi$, if there are $1 \leq k \leq m \leq n$ such that $\pi' = v_k \to \cdots \to v_m$. The empty path at $v \in V$ is the directed path given by the sequence (v). If $v, w \in V$ are process indices, then $\mathbf{P}(v, w)$ refers to the set of directed paths that start at v and end at w.

A *trek* from v to w is a triple $\tau = (\mathrm{Left}(\tau), \mathrm{top}(\tau), \mathrm{Right}(\tau))$, where $\mathrm{top}(\tau) \in V$ is a process index, and $\mathrm{Left}(\tau) \in \mathbf{P}(\mathrm{top}(\tau), v)$ and $\mathrm{Right}(\tau) \in \mathbf{P}(\mathrm{top}(\tau), w)$ are directed paths. The set of all treks from v to w is denoted as $\mathbf{T}(v, w)$.

Given a process graph $G = (V, D)$ and a process index $v \in V$, the *ancestors* of v, denoted $\mathrm{anc}(v)$, is the set of all process indices $u \in V$ such that $\mathbf{P}(u, v) \neq \varnothing$. In particular, this means $v \in \mathrm{anc}(v)$.

Suppose $\pi = (v_i)_{1 \leq i \leq n}$ is a directed path on G, then its associated *path function* [40] is

$$h^{(\pi)} := \prod_{i=1}^{n-1} h_{v_i, v_{i+1}},$$

i.e. the point-wise product of the associated link functions. Note that the path function of the empty path is the constant function with value 1. The *trek function* [40] of a trek τ is the point-wise product of functions

$$\mathcal{S}^{(\tau)} := h^{(\mathrm{Left}(\tau))} \mathcal{S}^{\mathrm{I}}_{\mathrm{top}(\tau)} (h^{(\mathrm{Right}(\tau)})^*.$$

In the following, we recall how the entries in the spectral density can be computed in terms of trek functions.

Proposition 1 (Spectral trek rule [40]). *Suppose $v, w \in V$ are process indices on the process graph G, then the entry $\mathcal{S}_{v,w}$ is equal to the sum of the trek functions that are associated with the treks from v to w, i.e.*

$$\mathcal{S}_{v,w} = \sum_{\tau \in \mathbf{T}(v,w)} \mathcal{S}^{(\tau)}.$$

Example 2. On the process graph in Figure 1 there is the directed path $\pi = u \to v \to w$. The path function of π evaluated at $z \in S^1$ is

$$\begin{aligned} h^{(\pi)}(z) &= h_{u,v}(z) h_{v,w}(z) \\ &= \frac{\phi_{u,v}(0)\phi_{v,w}(0) + (\phi_{u,v}(0)\phi_{v,w}(1) + \phi_{u,v}(1)\phi_{v,w}(0))z + \phi_{u,v}(1)\phi_{v,w}(1)z^2}{1 - (\phi_{v,v}(1) + \phi_{w,w}(1))z - \phi_{v,v}(1)\phi_{w,w}(1)z^2}. \end{aligned}$$

Furthermore, let us compute the entries $\mathcal{S}_{u,w}$ and $\mathcal{S}_{v,w}$ using the spectral trek rule. For $\mathcal{S}_{u,w}$, note that there is exactly one trek connecting v and w, namely the path π. Regarding $\mathcal{S}_{v,w}$, the set of treks is $\mathbf{T}(v, w) = \{v \leftarrow l \to w, v \to w\}$. Consequently, we get the following expressions for the corresponding entries in the spectrum:

$$\mathcal{S}_{w,u} = h^{(\pi)} \mathcal{S}^{\mathrm{I}}_u, \qquad \mathcal{S}_{w,v} = h_{v,w} \mathcal{S}^{\mathrm{I}}_v + h_{l,w} \mathcal{S}^{\mathrm{I}}_l h^*_{l,v}.$$

2.3. Related work

Several research contributions have dealt with the problem of identifying the causal structure and causal effects from observations in the setting of discrete-time processes.

For example, [14,34] showed that d-separation on the process graph implies conditional independence between the associated processes. However, the component processes are not allowed to affect each other contemporaneously. In distinction, our result allows for contemporaneous influences between the modeled processes. In particular, this includes static models where the realizations of the process are independent over time. Our results extend the aforementioned in the sense that we also show the reverse statement. Namely, we will show that conditional (spectral) independence between processes implies d-separation for almost all compatible SVAR processes. Furthermore, we show that d- and t-separation can be identified generically by considering the evaluation of the spectrum at a single arbitrary frequency. Allowing for contemporaneous effects comes at the cost that, in general, less information

about the process graph can be inferred from observations (autocovariance). In [14][Theorem 4.1] or [34][Theorem 3.7 and Theorem 3.9], the authors use the assumption that causal effects must have a lag of at least one to identify the direction of all (possibly cyclic) causal relations. While these approaches involve conditioning on the process's infinite past (i.e., infinite-dimensional random vectors), our approach only uses the spectrum evaluated at a single frequency, that is, a finite-dimensional matrix. However, our approach can only identify causal relations constrained by the Markov equivalence class of an acyclic process graph.

A different approach to study causal relations in discrete-time processes is to analyze them at the level of the time series graph. The Global Markov property, i.e. that d-separation at the level of the time series graph implies conditional independence, has been shown for VAR processes in [53] and for more general models (possibly non-linear and with instantaneous effects) in [28]. Recently, [25][Theorem 3 and Theorem 5] established the Global Markov property for VARMA processes with instantaneous effects parameterizing acyclic time series graphs. In addition, they show that, for almost all parameter configurations consistent with the time series graph, the distribution of the corresponding process is faithful, meaning that conditional independence implies d-separation resp. m-separation (as the MA-part of the process induces bidirectional edges on the time series graph) at the level of the time series graph. In particular, these works show that the classical techniques can be applied to the time series graph in order to identify the causal structure. For acyclic time series graphs, this approach is capable of identifying feedback relations, which makes the time series graph approach attractive for applications. In contrast, recovering feedback relations is beyond the scope of our proposed approach because it requires an acyclic process graph. Nevertheless, it is worthwhile to consider the less complex problem of directly identifying the process graph, since the potentially high complexity of the time series graph can pose serious difficulties for a reliable identification from finite data, especially in high-dimensional cases. To illustrate the difference in complexity between the identification of the time series graph and the process graph, fix an integer $p \geq 0$ and a process graph $G = (V, D)$. The number of time series graphs with G as the process graph and an order of at most p is given by

$$(2^{p+1} - 1)^{|D|} 2^{p|V|}.$$

In this paper, we demonstrate how analyzing an SVAR process in the frequency domain can simplify the identification of its causal structure. However, since this reduction involves temporal aggregation, our approach generally offers fewer identification capabilities than the time series graph approach.

The identification of causal effects in discrete-time systems has been addressed in [15,33,53]. In [15], the front-door and back-door identification of causal effects in stationary time series models is shown. In [53], the authors use a (conditional) instrumental variable approach to identify causal effects in latently confounded VAR processes (without contemporaneous effects), where the causal effects are the parameters of the VAR model. In [33], an instrumental variable approach is used to identify aggregated causal effects between VAR processes, where the aggregated effect between two processes can be understood in terms of the link functions [40] from Definition 6. Specifically, the aggregated effect considered in [33] is equal to the link function evaluated at the frequency $z = 1$.

The estimation of causal effects in SVAR processes, i.e. in the presence of possibly simultaneous causal effects (but without explicit modelling of latent processes), has been addressed in [27,44]. The techniques considered in [27] make additional assumptions about the noise terms, such as heteroskedasticity or non-Gaussianity. In contrast, we focus on relationships that can be identified from the spectral density of a stationary SVAR process. In [44] a constraint based causal discovery algorithm is developed that operates at the level of the time series graph. In the current paper, however, we are concerned with relations and identifiability criteria that are independent of the underlying time series graph.

In a different context, namely that of Hawkes processes [26], i.e. self and mutually exciting point processes, the Fourier and Laplace transform has been used in order to parametrically derive the causal structure and the interaction strengths of linear Hawkes processes from the covariance density [16].

3. An algebraic parameterization of SVAR processes in the frequency domain

The aim of this section is to provide a precise algebraic formulation of how the SVAR parameters map onto the direct transfer function, onto the spectral density of the internal dynamics, and finally onto the spectral density of the process. We will construct all of these objects as matrices over the field of rational functions. For the remainder of this section, we fix a set V of process indices and a random vector $\eta = (\eta_v)_{v \in V}$ whose components are mutually independent and have finite and positive variance.

3.1. An involution on the rational functions

In this section, we will construct a map on the space of rational functions that we need to parameterize the frequency representation of SVAR processes over the field of rational functions with real coefficients. In the following, we denote as $\mathbb{R}[z]$ the ring of polynomials with real coefficients. The degree of a polynomial $f = \sum_{k=0}^{n} \alpha_k z^k \in \mathbb{R}[z]$ is denoted $\deg(f)$ and is equal to n if $\alpha_n \neq 0$. The field of rational functions with real coefficients, which is the field of fractions of $\mathbb{R}[z]$ will be referred to as $\mathbb{R}(z)$. In the supplement [42][Appendix B] we provide more details.

The calculation of the spectral density involves complex conjugation, and since this operation is not a polynomial map, it seems to prevent the parametrization of the spectral density as a matrix with entries in the field $\mathbb{R}(z)$. Nevertheless it is possible due to two ingredients: First, the construction of an involution on the rational functions, i.e. a self-inverting algebra isomorphism $()^* : \mathbb{R}(z) \to \mathbb{R}(z)$, such that $r^*(z) = r(z^*)$ for every $z \in S^1$, and, second, the fact that the relevant domain for the spectrum is only the complex unit circle S^1, where $z^* = z^{-1}$ for every $z \in S^1$.

As a first step in the construction of this map on $\mathbb{R}(z)$, we provide an auxiliary map on the space of polynomials $()^* : \mathbb{R}[z] \to \mathbb{R}[z]$. This map sends a polynomial of degree n, i.e. $f = \sum_{k=0}^{n} \alpha_k z^k \in \mathbb{R}[z]$, to $f^* := \sum_{k=0}^{n} \alpha_{n-k} z^k$. Based on the auxiliary map $()^*$ we define the conjugation $()^* : \mathbb{R}(z) \to \mathbb{R}(z)$ on the space of rational functions as follows:

$$\frac{f}{g} \mapsto \left(\frac{f}{g}\right)^* := \left(z \mapsto \frac{f^*(z)}{g^*(z)} z^{\deg(g) - \deg(f)}\right). \tag{8}$$

Example 3. As an example of conjugation on $\mathbb{R}(z)$ let us consider the link function for $u \to v$ on the process graph in Figure 1. The conjugated rational link function is

$$h^*_{u,v}(z) = \frac{\phi_{u,v}(1) + \phi_{u,v}(0) z}{-\phi_{v,v}(1) + z}.$$

The conjugation map is indeed an automorphism on the space of rational functions $\mathbb{R}(z)$, i.e. it is linear and respects the multiplication of rational functions. The proof of the following proposition is deferred to the supplement [42][Appendix C].

Proposition 2. *The conjugation $()^* : \mathbb{R}(z) \to \mathbb{R}(z)$ is an algebra involution, i.e. conjugation respects multiplication and addition of rational functions, and the conjugation map is its own inverse map, i.e. $(r^*)^* = r$ for all $r \in \mathbb{R}(z)$. Finally, it holds for every $z \in S^1$ and every $r \in \mathbb{R}(z)$ that $r^*(z) = r(z^*)$.*

Since $\mathbb{R}(z)$ is a field (with characteristic zero), we can consider vector spaces over $\mathbb{R}(z)$ with all known concepts from linear algebra, e.g. linear maps, linear independence, rank, bases, Gaussian elimination, determinants. An immediate implication of Proposition 2 is the following.

Corollary 1. *Let $A \in \mathbb{R}(z)^{n\times n}$ be a matrix over the field of rational functions with real coefficients, then it holds that*

$$\det(A^*) = \det(A)^* \in \mathbb{R}(z),$$

where A^* is the entry-wise conjugation of A. In particular, $\det(A) = 0$ *if and only if* $\det(A^*) = 0$.

Throughout this work we use the following matrix notation. Suppose $M \in \mathbb{K}^{V\times V}$ is a matrix over some field $\mathbb{K}$ (of characteristic zero) and let $X, Y \subset V$ be subsets, then we write $[M]_{X,Y}$ for the submatrix of M that one obtains by selecting the entries corresponding to the rows indexed by X and the columns indexed by Y, i.e. $[M]_{X,Y} \coloneqq (M_{x,y})_{x\in X, y\in Y} \in \mathbb{K}^{X\times Y}$.

Remark 2. Suppose $M \in \mathbb{K}^{X\times Y}$ is a matrix. A submatrix $M_r = [M]_{X_r,Y_r} \in k^{X_r\times Y_r}$ of M is called an r-minor if the subsets $X_r \subset X$ and $Y_r \subset Y$ have cardinality r, i.e. $|X_r| = |Y_r| = r$. The matrix M has rank r if M has an ivertibel r-minor M_r, i.e. $\det(M_r)) \neq 0$, and if every $r+1$ minor M_{r+1} of M is not invertible, i.e. $\det(M_{r+1}) = 0$.

3.2. The rational frequency SVAR representation

In this subsection, we use the terminology and constructions from the previous subsection for the frequency domain parametrization of SVAR processes. Our overall goal is to build an algebraic identifiability theory on top of this parametrization. Since identifying causal relationships and effects becomes particularly interesting and challenging when some of the variables are hidden, we extend our graphical terminology to also include latent processes.

Suppose $\mathcal{G}$ is a time series graph over a set of process indices V, some of which are labeled observable and some of which are labeled latent, i.e. $V = O \cup L$ with $O \cap L = \varnothing$. We define the set $\mathcal{D}_O \subset \mathcal{D}$ to be the set of all edges that emerge from some observed $v(t) \in O \times \mathbb{Z}$, and we denote as $\mathcal{D}_L$ the set of edges originating from a latent $l(t) \in L \times \mathbb{Z}$. Similarly, we divide the set of process-level links $D = D_O \cup D_L$, where D_O consists of the links that emerge from an observed process and D_L contains the links emerging from a latent process. For a process $v \in V$, we distinguish between its observed and latent parents by writing $\mathrm{pa}_O(v)$ and $\mathrm{pa}_L(v)$ for $\mathrm{pa}(v) \cap O$ and $\mathrm{pa}(v) \cap L$ respectively. Throughout this work we require any latent structure to satisfy the following.

Assumption 1 (Exogenous latent structure). Suppose $\mathcal{G} = ((O \cup L) \times \mathbb{Z}, \mathcal{D}_O \cup \mathcal{D}_L)$ is a time series graph carrying a latent structure. Then the set of edges is such that every latent process $l \in L$ has zero incoming edges on the process graph $G = (O \cup L, D_O \cup D_L)$.

Example 4. Let us equip the process graph in Figure 1 with a latent structure. We define the observed processes to be $O = \{u, v, w\}$ and the latent processes are $L = \{l\}$. Accordingly, the set of observed edges are $D_O = \{u \to v, v \to w\}$ and the set of latent edges are $L = \{l \to v, l \to w\}$. This aligns with Assumption 1 because there are no edges pointing to l.

To construct the frequency parameterization, we introduce two subsets of $\mathbb{R}(z)^{O\times O}$. The first is the set $\mathbb{R}(z)^D_{\mathrm{stb}}$, which consists of the matrices $A \in \mathbb{R}(z)^{V\times V}$ that satisfy the following three conditions. First,

for any pair $v, w \in V$ such that $v \not\to w$, the entry $A_{v,w}$ is the constant function with value 0. Second, none of the entries of A have poles inside the complex unit disk $D^1 = \{z : |z| \leq 1\}$. Third, the inverse of the submatrix $[I - A]_{O,O} \in \mathbb{R}(z)^{O \times O}$ exists and is such that its entries do not have poles inside D^1 and satisfies the identity

$$[I - A]^{-1}_{O,O} = \sum_{k \geq 0} [A]^k_{O,O}.$$

The second set is $\mathrm{PD}_O(\mathbb{R}(z), D_L)$, which contains all matrices $BB^* + C^\top C^* \in \mathbb{R}(z)^{O \times O}$, where $B \in \mathbb{R}(z)^{O \times O}$ is an invertible diagonal matrix, and $C \in \mathbb{R}(z)^{L \times D}$ is such that $C_{l,v} = 0$ if $l \not\to v$. Note that $\mathrm{PD}_V(\mathbb{R}(z), \varnothing)$ is a subset of the diagonal matrices over $\mathbb{R}(z)$. By virtue of the following lemma, we will from now on regard the internal spectrum as an element in $\mathrm{PD}_V(\mathbb{R}(z), \varnothing)$.

Lemma 1. *There is a diagonal matrix $\tilde{\mathcal{S}}^{\mathrm{I}} \in \mathbb{R}(z)^{O \times O}$ such that its restriction to the complex unit circle is precisely the internal spectrum $\mathcal{S}^{\mathrm{I}}$ from (7), i.e. $\tilde{\mathcal{S}}^{\mathrm{I}}(z) = \mathcal{S}^{\mathrm{I}}(z)$ for all $z \in S^1$.*

Proof. Let $v \in V$ be a process index, then recall that for $z \in S^1$ we defined

$$\mathcal{S}^{\mathrm{I}}_v(z) = \omega_v \frac{1}{(1 - \varphi_{v,v}(z))(1 - \varphi_{v,v}(z^*))},$$

where $\varphi_{v,v} \in \mathbb{R}[z]$ is the lag polynomial of v. Using conjugation on rational functions $\mathbb{R}(z)$ (8) we set

$$\tilde{\mathcal{S}}^{\mathrm{I}}_{v,w} = \begin{cases} \omega_v \dfrac{1}{1 - \varphi_{v,v}} \left(\dfrac{1}{1 - \varphi_{v,v}} \right)^* & \text{if } v = w, \\ 0 & \text{if } v \neq w, \end{cases}$$

so that $\tilde{\mathcal{S}}^{\mathrm{I}} \in \mathbb{R}(z)^{O \times O}_{\mathrm{diag}}$, and by Proposition 2 it holds that $\mathcal{S}^{\mathrm{I}}(z) = \tilde{\mathcal{S}}^{\mathrm{I}}(z)$ for every $z \in S^1$. □

As a final step in the construction of the frequency parametrization, we need to specify its domain, i.e. the set of SVAR parameters whose associated transfer function H lies in $\mathbb{R}(z)^D_{\mathrm{stb}}$ and whose internal spectrum $\mathcal{S}^{\mathrm{I}}$ is an element of $\mathrm{PD}_V(\mathbb{R}(z), \varnothing)$. First of all, let $\mathbb{R}^{\mathcal{D}}$ be the vector space where each dimension corresponds to a unique lagged or contemporaneous relation $v \to_k w$ on the time series graph. Since we assume $\mathcal{G}$ to have finite order, the vector space $\mathbb{R}^{\mathcal{D}}$ is finite dimensional. We define $\mathbb{R}^{\mathcal{D}}_{\mathrm{stb}}$ to be the set of all parameters $\Phi \in \mathbb{R}^{\mathcal{D}}$ such that for every $v \in V$ it holds that

$$\sum_{x,y \in C_v} \sum_{k : x \to_k y} |\phi_{x,y}(k)| < 1, \tag{9}$$

where $C_v = \{w \in V \mid v \in \mathrm{anc}(w) \text{ and } w \in \mathrm{anc}(v)\}$. Note that $C_v = \{v\}$ if and only if v is not part of a cycle on G. Requiring inequality (9) renders the space $\mathbb{R}^{\mathcal{D}}_{\mathrm{stb}}$ an open non empty semi-algebraic subspace of $\mathbb{R}^{\mathcal{D}}$. If the SVAR parameter Φ lies in $\mathbb{R}^{\mathcal{D}}_{\mathrm{stb}}$, then the associated transfer function $H(\Phi)$ is an element in $\mathbb{R}(z)^D_{\mathrm{stb}}$, see the supplementary material [42][Appendix A, Proposition 3]. The decomposition of the links $\mathcal{D} = \mathcal{D}_O \cup \mathcal{D}_L$ factors the set $\mathbb{R}^{\mathcal{D}}_{\mathrm{stb}} \cong \mathbb{R}^{\mathcal{D}_O}_{\mathrm{stb}} \times \mathbb{R}^{\mathcal{D}_L}_{\mathrm{stb}}$ so that a SVAR coefficient vector splits as $\Phi = (\Phi_O, \Phi_L)$. From now on we refer by $\mathbb{R}^{\mathcal{G}}_{\mathrm{stb}}$ to the product space $\mathbb{R}^{\mathcal{D}}_{\mathrm{stb}} \times \mathbb{R}^V_+$, where $\mathbb{R}^V_+$ is the space of all positive definite diagonal matrices (which scale the components of η). We conclude by defining the map

$$\mathcal{F}_{\mathcal{G}} = ([H]_{O,O}, \mathcal{S}^{\mathrm{LI}}) : \mathbb{R}^{\mathcal{G}}_{\mathrm{stb}} \longrightarrow \mathbb{R}(z)^{D_O}_{\mathrm{stb}} \times \mathrm{PD}_O(\mathbb{R}(z), D_L),$$

which sends a SVAR parameter (Φ, Ω) to the associated direct transfer function (of the observed processes) and the *projected internal spectrum*

$$\mathcal{S}^{\mathrm{LI}} \coloneqq \mathcal{S}^{\mathrm{I}}_O + [H]^{\top}_{L,O} \mathcal{S}^{\mathrm{I}}_L [H]^{*}_{L,O},$$

where $\mathcal{S}^{\mathrm{I}}_O$ resp. $\mathcal{S}^{\mathrm{I}}_L$ denotes the submatrix $[\mathcal{S}^{\mathrm{I}}]_{O,O}$ and $[\mathcal{S}^{\mathrm{I}}]_{L,L}$ respectively.

Example 5. Let us consider the projected internal spectrum of the process parameterized by the time series graph in Figure 1. Starting with the diagonal entries, the projected internal spectrum of the process u is equal to its internal spectrum, since u has no latent parents. The processes v and w, on the contrast, both have the process l as a latent parent. So, the contribution of l to v resp. w needs to be added to the internal spectra when computing the projected internal spectra, i.e.

$$\begin{aligned}
\mathcal{S}^{\mathrm{LI}}_u &= \mathcal{S}^{\mathrm{I}}_v \\
&= \omega_u (1 - \varphi_{u,u})^{-1} (1 - \varphi_{u,u})^{-*}, \\
\mathcal{S}^{\mathrm{LI}}_v &= \mathcal{S}^{\mathrm{I}}_v + h_{l,v} h^{*}_{l,v} \mathcal{S}^{\mathrm{I}}_l \\
&= \omega_v (1 - \varphi_{v,v})^{-1} (1 - \varphi_{v,v})^{-*} + \omega_l h_{l,v} h^{*}_{l,v} (1 - \varphi_{l,l})^{-1} (1 - \varphi_{l,l})^{-*}, \\
\mathcal{S}^{\mathrm{LI}}_w &= \mathcal{S}^{\mathrm{I}}_w + h_{l,w} h^{*}_{l,w} \mathcal{S}^{\mathrm{I}}_l \\
&= \omega_w (1 - \varphi_{w,w})^{-1} (1 - \varphi_{w,w})^{-*} + \omega_l h_{l,w} h^{*}_{l,w} (1 - \varphi_{l,l})^{-1} (1 - \varphi_{l,l})^{-*}.
\end{aligned}$$

For the off-diagonal entries, note that only the entry corresponding to the pair (v, w) is non-zero. This is because $\mathrm{pa}_L(u) \cap \mathrm{pa}_L(v) = \varnothing$ and $\mathrm{pa}_L(u) \cap \mathrm{pa}_L(w) = \varnothing$, but $\mathrm{pa}_L(v) \cap \mathrm{pa}_L(w) = \{l\}$. Consequently,

$$\mathcal{S}^{\mathrm{LI}}_{u,v} = \mathcal{S}^{\mathrm{LI}}_{u,w} = 0, \qquad \mathcal{S}^{\mathrm{LI}}_{v,w} = \omega_h h_{l,v} h^{*}_{l,w} (1 - \varphi_{l,l})^{-1} (1 - \varphi_{l,l})^{-*}.$$

Lemma 2. *For two distinct $v, w \in O$ it holds that $\mathcal{S}^{\mathrm{LI}}_{v,w} = 0$ if $\mathrm{pa}_L(v) \cap \mathrm{pa}_L(w) = \varnothing$, and for every $z \in S^1$ the evaluation $\mathcal{S}^{\mathrm{LI}}(z) \in \mathbb{C}^{O \times O}$ is positive definite.*

Proof. First, observe that for two observed processes $v, w \in O$ the associated entry in the projected internal spectrum is given by the formula

$$\mathcal{S}^{\mathrm{LI}}_{v,w} = \begin{cases} \mathcal{S}^{\mathrm{I}}_{v,v} + \displaystyle\sum_{l \in \mathrm{pa}_L(v)} h_{l,v} h^{*}_{l,v} \mathcal{S}^{\mathrm{I}}_l & \text{if } v = w, \\ \displaystyle\sum_{l \in \mathrm{pa}_L(v) \cap \mathrm{pa}_L(w)} \mathcal{S}^{\mathrm{I}}_l h_{l,v} h^{*}_{l,w} & \text{if } v \neq w. \end{cases}$$

It follows that $\mathcal{S}^{\mathrm{LI}}$ is symmetric after entry-wise conjugation and also that $\mathcal{S}^{\mathrm{LI}}_{v,w} = 0$ whenever $\mathrm{pa}_L(v) \cap \mathrm{pa}_L(w) = \varnothing$. It remains to show that, on the unit circle, the projected internal spectrum evaluates to a positive definite matrix. So let $x \in \mathbb{C}^O$ be a non-zero vector and $z \in S^1$ a point on the complex unit circle. Using the formula above and Proposition 2 and that $\mathcal{S}^{\mathrm{I}}_v(z) > 0$ for every $z \in S^1$ we get that

$$(x^*)^{\top} \mathcal{S}^{\mathrm{LI}}(z) x = \sum_{v \in O} \omega_v |x_v|^2 \mathcal{S}^{\mathrm{I}}_v(z) + \sum_{l \in \mathrm{pa}_L(v)} |x_v h_{l,v}(z)|^2 \mathcal{S}^{\mathrm{I}}_l(z) > 0,$$

which shows that $\mathcal{S}^{\mathrm{LI}}(z)$ is positive definite. □

Remark 3. Using the trek rule (Proposition 1) we see that the projected internal spectrum is given by the trek monomials that are associated with the trivial treks and the treks that involve edges in D_L only.

An immediate question concerning the parameterizing map of the observed link functions $[H]_{O,O}$ is whether the SVAR parameters of Φ_O for the lagged and instantaneous effects between the observed nodes can be uniquely recovered from $[H(\Phi_O)]_{O,O}$ almost everywhere. Given the rational function $h_{v,w}(\Phi) = fg^{-1}$, where $f, g \in \mathbb{R}[z]$ are coprime, the coefficients $\phi_{v,w}(k)$ and $\phi_{w,w}(j)$ can be recovered from $h_{v,w}$ if $\varphi_{v,w} = f$ and $g = 1 - \varphi_{w,w}$, i.e. if $1 - \varphi_{w,w}$ and $\varphi_{v,w}$ are coprime. However, this is not always the case, as the following example shows.

Example 6. Consider the time series graph $\mathcal{G} = (\{1,2\} \times \mathbb{Z}, \mathcal{D})$ with the following time lagged relations: $1 \to_0 2$, $1 \to_1 2$, $2 \to_1 2$ and $2 \to_2 2$. Furthermore, we fix some $0 < \alpha < 1$ and consider the parametrization Φ of the links on $\mathcal{G}$, where

$$\phi_{1,2}(0) = 1, \qquad \phi_{1,2}(1) = -\alpha,$$
$$\phi_{2,2}(1) = 2\alpha, \qquad \phi_{2,2}(2) = -\alpha^2.$$

In addition, we consider the parametrization Φ' of $\mathcal{G}$, defined as

$$\phi'_{1,2}(0) = 1, \qquad \phi'_{2,2}(1) = \alpha.$$

Then it is easy to verify that

$$h_{v,w}(\Phi, z) = \frac{1}{1 - \alpha z} = h_{v,w}(\Phi', z).$$

In particular, the coefficient Φ cannot be recovered from the unique representation of $h_{v,w}$ as a fraction of coprime polynomials.

Lemma 3. *The parametrization of the direct transfer function $[H]_{O,O}$ is generically injective. That means, the subset $A \subset \mathbb{R}_{\text{stb}}^{\mathcal{D}_O}$ of parameter vectors Φ_O for which there is another $\Phi'_O \in \mathbb{R}_{\text{stb}}^{\mathcal{D}_O} \setminus \{\Phi_O\}$ such that $[H(\Phi_O)]_{O,O} = [H(\Phi'_O)]_{O,O}$ has Lebesgue measure zero.*

Proof. The recovery of Φ_O from $[H(\Phi_O)]_{O,O}$ could only fail if there is a link $v \to w$ between observed processes $v, w \in O$ such that the denominator and the numerator of the associated link function are not coprime, i.e. if the polynomial $\varphi_{v,w}$ and the polynomial $1 - \varphi_{w,w}$ share a common divisor in $\mathbb{R}[z]$. Recall that two polynomials $f, g \in \mathbb{R}[z]$ have a common divisor in $\mathbb{R}[z]$ if and only if the resultant [20][Chapter 12.1], denoted by $\text{res}(f, g)$, evaluates to zero. Considering $\varphi_{v,w}$ and $\varphi_{w,w}$ as Φ_O-parameterized polynomials exhibits $\text{res}(\varphi_{v,w}, 1 - \varphi_{w,w})$ as a multi-variate polynomial in Φ_O. Thus, a parameter Φ_O can only then not be uniquely recovered from $H(\Phi_O)$ if $\text{res}_{D_O}(\Phi_O) = 0$, where

$$\text{res}_{D_O} \coloneqq \prod_{v \to w \in D_O} \text{res}(\varphi_{v,w}, 1 - \varphi_{w,w}) \in \mathbb{R}[\Phi_O].$$

The polynomial res_{D_O} is non-zero because each of its factors $\text{res}(\varphi_{v,w}, 1 - \varphi_{w,w})$, with $v \to w \in D_O$, is a non-zero polynomial, see the supplementary material [42][Appendix C]. It follows that the zero set $\{\Phi_O \in \mathbb{R}_{\text{stb}}^{\mathcal{D}} \mid \text{res}_{D_O}(\Phi_O) = 0\}$ is a closed lower-dimensional variety (manifold) lying inside the parameter space $\mathbb{R}^{\mathcal{D}_O}$ and is therefore a zero set with respect to the Lebesgue measure. Since the set A is contained in this zero set, the claim follows. □

Finally, using identity (5) and the parametrization $\mathcal{F}_{\mathcal{G}}$ we conclude this section with the following parametrization of the spectrum

$$\mathcal{S}_{\mathcal{G}} : \mathbb{R}^{\mathcal{G}}_{\text{stb}} \xrightarrow{\mathcal{F}_{\mathcal{G}}} \mathbb{R}(z)^{D_O}_{\text{stb}} \times \text{PD}_O(\mathbb{R}(z), D_L) \xrightarrow{\mathcal{S}_G} \text{PD}_O(\mathbb{R}(z)), \tag{10}$$

where $\mathcal{S}_G(H, \mathcal{S}^{\text{LI}}) = (I - H^{\top})^{-1}\mathcal{S}^{\text{LI}}(I - H^*)^{-1}$, and by $\text{PD}_O(\mathbb{R}(z)) \subset \mathbb{R}(z)^{O\times O}$ we refer to the matrices $C^{\top}C^*$, where $C \in \mathbb{R}(z)^{O\times O}$ is invertible. If there is no need to emphasize the time series graph $\mathcal{G}$, we will write $\mathcal{S}$ instead of $\mathcal{S}_{\mathcal{G}}$ to denote the spectral parametrization.

Remark 4. The parametrization map $\mathcal{S}_{\mathcal{G}}$ can also be seen as a rational function in the indeterminates Φ, Ω and z, i.e as a matrix in $\mathbb{R}(\Phi, \Omega, z)^{O\times O}$, where $\mathbb{R}(\Phi, \Omega, z)$ is the field of rational functions in Φ, Ω and z with real coefficients.

The map (10) parameterize the spectrum globally, i.e. each SVAR parameter is associated with a matrix of rational functions. However, it is often not possible to estimate the spectrum of the SVAR process parametrically, for example in the presence of latent processes. In such cases, the discrete time Fourier transform could be used to estimate the evaluation of the spectrum at certain frequencies, also known as the periodogram. To address the question of what can be inferred from evaluations of the spectrum, we define for a frequency $z \in S^1$ the *local spectrum*

$$\mathcal{S}_{\mathcal{G}}(z) : \mathbb{R}^{\mathcal{G}}_{\text{stb}} \xrightarrow{\mathcal{S}_{\mathcal{G}}} \text{PD}_O(\mathbb{R}(z)) \xrightarrow{\text{ev}_z} \text{PD}_O(\mathbb{C}), \tag{11}$$

where $\text{ev}_z : \mathbb{R}(z) \to \mathbb{C} \cup \{\infty\}$ is the map that sends a rational function f to its evaluation $f(z)$. The stability restriction, i.e. inequality (9), together with Proposition 2 and Lemma 2 ensures that the *local spectrum* (11) is well defined, i.e. maps into the set of hermitian positive definite matrices.

Remark 5. Alternatively, the local spectrum at $z \in S^1$ can be seen as matrix over the field of rational functions with complex coefficients and indeterminates Φ and Ω, denoted $\mathbb{C}(\Phi, \Omega)$, i.e. $\mathcal{S}_{\mathcal{G}}(z) \in \mathbb{C}(\Phi, \Omega)^{O\times O}$.

4. Spectral process graph identification

In this section, we use the parametrization of the spectrum as a matrix over the rational functions $\mathbb{R}(z)$ to relate the graphical structure of the process graph to algebraic statements about the spectrum. Specifically, we will see how graphical separation statements about G correspond to rank conditions (which are conditions that can be expressed in terms of determinants, see Remark 2) on certain submatrices of the spectrum $\mathcal{S}$. These characterizations generalize the results for linear SCMs.

Suppose $G = (V, D)$ is a process graph and $\mathcal{G} \in \mathcal{T}(G)$ a compatible time series graph, with two subsets $X, Y \subset V$, then we will refer to $[\mathcal{S}_{\mathcal{G}}]_{X,Y} \in \mathbb{R}(\Phi, \Omega, z)^{X\times Y}$ as the *global subspectrum*, and for a frequency $z \in S^1$ we will call $[\mathcal{S}_{\mathcal{G}}(z)]_{X,Y} \in \mathbb{C}(\Phi, \Omega)^{X\times Y}$ the *local subspectrum* at z.

Definition 2 (Global and local rank of subspectra). The *global rank* of the subspectrum corresponding to X and Y, denoted $\text{rank}([\mathcal{S}_{\mathcal{G}}]_{X,Y})$, is the rank of the matrix $[\mathcal{S}_{\mathcal{G}}]_{X,Y}$ over the field $\mathbb{R}(\Phi, \Omega, z)$. The *local rank* at frequency $z \in S^1$, denoted by $\text{rank}([\mathcal{S}_{\mathcal{G}}(z)]_{X,Y})$, is the rank of $[\mathcal{S}_{\mathcal{G}}(z)]_{X,Y}$ over the field $\mathbb{C}(\Phi, \Omega)$.

Definition 3 (Trek separation [52][Definition 2.7).] Let $G = (V, D)$ be a process graph and $X, Y, Z_X, Z_Y \subset V$ not necessarily disjoint subsets, then (Z_X, Z_Y) is said to *t-separate* X from Y if for every trek τ from a process $x \in X$ to a process $y \in Y$ either Left(τ) intersects Z_X or Right(τ) intersects Z_Y.

The following generalizes [52][Theorem 2.8] for linear SCMs to SVAR processes represented in the frequency domain.

Theorem 1 (Spectral trek separation). *Let G be an acyclic process graph, $\mathcal{G} \in \mathcal{T}(G)$ a compatible time series graph, $X, Y \subset V$ not necessarily disjoint subsets, and $z \in S^1$ a frequency, then*

$$\begin{aligned} \operatorname{rank}([\mathcal{S}_{\mathcal{G}}]_{X,Y}) &= \operatorname{rank}([\mathcal{S}_{\mathcal{G}}(z)]_{X,Y}) \\ &= \min\{|Z_X| + |Z_Y| \mid (Z_X, Z_Y) \text{ t-separates } X \text{ from } Y\}. \end{aligned} \tag{12}$$

Proof. The proof of this theorem follows by exactly the same arguments as in the case of SEM's [52][Theorem 2.8]. The main result on which the proof is based is a spectral version of the Gessel-Viennot Lemma, which we will formulate in the next subsection. □

Remark 6. Theorem 1 has two important implications on the subspectrum of an SVAR process whose parameters are compatible with an acyclic process graph:

1. The rank of any subspectrum only depends on the structure of the process graph, i.e. the rank is independent of the specifics of the time series graph.
2. Generically, the global rank of any subspectrum can be determined by computing the local rank of that subspectrum at any frequency $z \in S^1$. That means, the set of SVAR parameters $(\Phi, \Omega) \in \mathbb{R}^{\mathcal{G}}_{\text{stb}}$ for which the rank of the complex-valued matrix $\mathcal{S}_{\mathcal{G}}(\Phi, \Omega, z) \in \mathbb{C}^{X \times Y}$ is not equal to $\operatorname{rank}([\mathcal{S}_{\mathcal{G}}]_{X,Y})$ is a zero set w.r.t. the Lebesgue measure.

4.1. Spectral characterization of process-level d-separation

It is well known that for linear SCMs conditional independence between variables are generically equivalent to d-separation of the corresponding nodes on the causal graph. We will here see that this observation generalizes to SVAR processes. For the following, let $G = (V, D)$ be an acyclic process graph with underlying time series graph $\mathcal{G} \in \mathcal{T}(G)$. We begin by recalling the graphical notion of d-separation between nodes on directed graphs.

Definition 4 (d-separation [37,38]). Let $G = (V, D)$ be an acyclic process graph and suppose $X, Y, Z \subset V$ are subsets of processes. The set Z d-separates X and Y if every (not necessarily directed) path π between a node $x \in X$ and $y \in Y$ is going through a node $k \in V$ that is either: (1.) a non-collider node on π that belongs to Z, or (2.) a collider node on π that is not ancestral to any process in Z, i.e $k \in V \setminus \operatorname{anc}(Z)$. Given a path π and a node $k \in V$, then k is called a collider node on π if the path π contains a link $v \to k$ and another link $w \to k$. Otherwise k is a non-collider node on π. We write $X \perp_d Y \mid Z$ to denote that the the two sets X and Y are d-separated by the set Z.

We will characterize d-separation statements about the process graph by algebraic relations, namely conditional spectral independence relations between processes, which is based on the conditional spectrum as introduced in [6][Chapter 8 Equation (8.3.4.)]. The conditional spectrum measures the (second-order) dependence between two sets of processes after removing the linear contribution of a third set of processes, i.e. the conditioning set, see [6][Theorem 8.3.1].

Definition 5 (Conditional spectral independence). Let $\mathbb{X}$ be a SVAR process specified by the parameter configuration $(\Phi, \Omega) \in \mathbb{R}^{\mathcal{G}}_{\text{stb}}$, and let $X, Y, Z \subset V$ be disjoint sets of process indices. Then the subprocesses $\mathbb{X}_X$ and $\mathbb{X}_Y$ are said to be *conditionally independent* (in the spectrum) given $\mathbb{X}_Z$ if

$$[\mathcal{S}(\Phi, \Omega)]_{X,Y|Z} \coloneqq [\mathcal{S}(\Phi, \Omega)]_{X,Y} - [\mathcal{S}(\Phi, \Omega)]_{X,Z} [\mathcal{S}(\Phi, \Omega)]^{-1}_{Z,Z} [\mathcal{S}(\Phi, \Omega)]_{Z,Y}$$

is the zero element in $\mathbb{R}(z)^{X \times Y}$.

It turns out that every conditional (spectral) independence statement between subprocesses is equivalent to a rank condition on the spectrum.

Proposition 3. *Conditional spectral independence of $\mathbb{X}_X$ and $\mathbb{X}_Y$ given $\mathbb{X}_Z$ is equivalent to*

$$\operatorname{rank}([\mathcal{S}(\Phi, \Omega)]_{X \cup Z, Y \cup Z}) = |Z|.$$

Proof. The proof goes along the same lines as the proof of [51][Proposition 2.7]. The only difference being that one reasons over the field $\mathbb{R}(z)$ instead of $\mathbb{R}$. □

If the time series graph $\mathcal{G}$ has order zero and the noise term η is Gaussian, then this notion of conditional independence boils down to the definition of conditional independence for normally distributed random variables [51]. We will use the following graphical result to characterize conditional spectral independence of SVAR processes via d-separation on the process graph.

Lemma 4 ([9][Corollary 1.3). *] Let $G = (V, D)$ be a DAG, and $X, Y, Z \subset V$ disjoint sets of vertices. Then it holds that Z d-separates X and Y if and only if there is a partition $Z = Z_X \cup Z_Y$ such that (Z_X, Z_Y) t-separates $X \cup Z$ and $Y \cup Z$.*

Theorem 2. *Let $G = (V, D)$ be an acyclic process graph and let $X, Y, Z \subset V$ be disjoint subsets. Then it holds for generic parameter choices (Φ, Ω) that the sub-processes $\mathbb{X}_X$ and $\mathbb{X}_Y$ are conditionally independent given $\mathbb{X}_Z$ if and only if Z d-separates X from Y. Specifically, for every time series graph $\mathcal{G} \in \mathcal{T}(G)$ and every frequency $z \in S^1$ it holds that*

$$\operatorname{rank}([\mathcal{S}_{\mathcal{G}}]_{Z \cup X, Z \cup Y}) = \operatorname{rank}([\mathcal{S}_{\mathcal{G}}(z)]_{Z \cup X, Z \cup Y}) = |Z| \iff X \perp_d Y | Z.$$

Proof. Suppose $\operatorname{rank}([\mathcal{S}_{\mathcal{G}}]_{X \cup Z, Y \cup Z}) = |Z|$, then Theorem 1 implies that there is a pair of sets (Z_X, Z_Y) that t-separates $X \cup Z$ and $Y \cup Z$, and that satisfies $|Z_X| + |Z_Y| = |Z|$. The requirement $Z \subset Z_X \cup Z_Y$ implies $Z = Z_X \cup Z_Y$. It now follows from Lemma 4 that Z d-separates X and Y. On the other hand, assume X and Y are d-separated by Z, then by Lemma 4 there is (Z_X, Z_Y) which t-separates $X \cup Z$ and $Y \cup Z$ such that $Z = Z_X \cup Z_Y$. Consequently, it holds that $\operatorname{rank}([\mathcal{S}_{\mathcal{G}}]_{X \cup Z, Y \cup Z}) \leq |Z|$. On the other hand it must hold that $Z \subset Z'_X \cup Z'_Y$ for any (Z'_X, Z'_Y) that t-separates $Z \cup X$ and $Z \cup Y$ so that $\operatorname{rank}([\mathcal{S}_{\mathcal{G}}]_{X \cup Z, Y \cup Z}) \geq |Z|$. This finishes the proof. □

For a time series graph $\mathcal{G} \in \mathcal{T}(G)$, a parameter pair $(\Phi, \Omega) \in \mathbb{R}^{\mathcal{G}}_{\text{stb}}$ is said to define a stationary distribution that is faithful with respect to the process graph G if for any three disjoint sets $X, Y, Z \subset V$ it holds that X, Y are d-separated by Z if and only if there is conditional spectral independence between $\mathbb{X}_X$ and $\mathbb{X}_Y$ given $\mathbb{X}_Z$. Otherwise (Φ, Ω) is said to define an unfaithful distribution with respect to G. Theorem 2 implies that the set of parameter pairs defining an unfaithful stationary distribution is a zero set with respect to the Lebesgue measure. In particular, it is generically possible to derive the Markov equivalence class [37,38,50] of the process graph from the spectrum of the process.

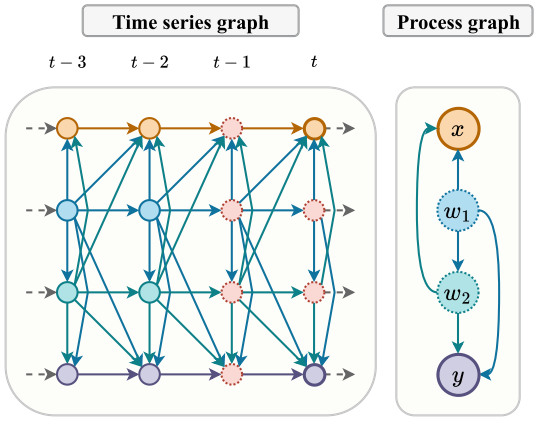

Figure 2. A time series graph of order one together with its process graph. The processes x and y are d-separated by $\{w_1, w_2\}$. The d-separation at the level of the time series graph corresponds to the fact that $x(t)$ and $y(t)$ are independent when conditioned on all the variables that correspond to the nodes colored red.

As mentioned earlier, more information about the causal structure of an SVAR process can be identified by assuming that processes do not influence each other through contemporaneous links [14,34]. In this case, cyclic relationships on the process graph can be identified from the autocovariance sequence of the process. The spectrum, although equivalent to the autocovariance, does not seem to be well suited for recovering information that is critically linked to the temporal order. The reason seems to be that the spectrum evaluated at each frequency is a weighted sum of the time-lagged covariance matrices, i.e. the temporal information is aggregated. In particular, the possibility of resolving cyclic relationships from the evaluation of the spectrum at a single frequency seems unlikely. Instead, it is likely that evaluations at multiple frequencies need to be combined. However, a spectral characterization of cyclic relations remains to be found. This leaves the question of the advantages of analyzing the process in the frequency domain. One advantage is that the algebraic complexity of identifying relationships on the process graph depends only on the process graph and not on the order of the SVAR process, i.e., the complexity of the time series graph. The following example illustrates this advantage.

Example 7. Consider the graph in Figure 2, where x and y are d-separated by $\{w_1, w_2\}$. Theorem 2 implies that this d-spearation relation can be identified generically by veryifing that the 3×3-matrix

$$[\mathcal{S}_{\mathcal{G}}(\Phi, \Omega, z)]_{\{w_1,w_2,x\},\{w_1,w_2,y\}} \in \mathbb{C}^{3\times 3}$$

has rank 2, where $z \in S^1$ is some arbitrary frequency. Alternatively, the d-separation statement could be identified using the time series graph structure together with the assumption that its order is at most one. Let us consider the two random vectors

$$\mathcal{X}' = \begin{bmatrix} \mathbb{X}_{w_1}(t) & \mathbb{X}_{w_1}(t-1) & \mathbb{X}_{w_2}(t) & \mathbb{X}_{w_2}(t-1) & \mathbb{X}_x(t-1) & \mathbb{X}_y(t-1) & \mathbb{X}_x(t) \end{bmatrix},$$
$$\mathcal{X}'' = \begin{bmatrix} \mathbb{X}_{w_1}(t) & \mathbb{X}_{w_1}(t-1) & \mathbb{X}_{w_2}(t) & \mathbb{X}_{w_2}(t-1) & \mathbb{X}_x(t-1) & \mathbb{X}_y(t-1) & \mathbb{X}_y(t) \end{bmatrix},$$

and their covariance matrix $C = \mathbb{E}[\mathcal{X}'(\mathcal{X}'')^\top] \in \mathbb{R}^{7\times 7}$. The d-separation of x and y by $\{w_1, w_2\}$ implies by [52][Theorem 2.17] that the matrix C has rank 6. Now assume that the order of the time series graph is some $p \geq 1$. Then the matrix C to identify the d-separation would have dimension $(4p+3) \times (4p+3)$,

and the rank of C would be $4p+2$ (in case of d-separation). Hence, the dimension of the rank condition due to d-separation depends on the order of the SVAR process. On the other hand, the rank condition characterizing d-separation in the frequency domain is independent of the order of the process, and in the generic case it suffices to check the condition at an arbitrarily chosen frequency.

4.2. Spectral Gessel Viennot lemma

The proof of the trek separation theorem [52][Theorem 2.8] is based on the classical Gessel-Viennot Lemma [24,30]. In the following we will see that this lemma generalizes to SVAR processes represented in the frequency domain, i.e. we establish a graphical description of $\det([(I-H)^{-1}]_{X,Y}) \in \mathbb{R}(z)$, where $X, Y \subset V$ are equal-sized subsets. This description is essential for the proof of Theorem 1 and also for proving graphical identifiability criteria for the link functions of SVAR processes, as we will see in the next section. For the rest of this subsection we assume that the process graph G is acyclic.

Suppose $X = \{x_1, \dots, x_r\} \subset V$ and $Y = \{y_1, \dots, y_r\} \subset V$ are equal-sized sets of process indices. A system of directed paths from X to Y is a collection of directed paths $\Pi = (\pi_1, \dots, \pi_r)$ such that for any two distinct paths $\pi_i, \pi_j \in \Pi$ it holds that they do not have the same starting or end point. In particular, the system Π gives rise to a bijection from X to Y, which maps $x \in X$ to the endpoint $y \in Y$ of the unique path $\pi \in \Pi$ with x as its starting point. Given the ordering of X and Y, Π induces a permutation α on the set $\{1, \dots, r\}$ where $\alpha(i) = j$ if the unique path $\pi \in \Pi$ starting at x_i ends at y_j. In the following, we write $\operatorname{sgn}(\Pi)$ for the sign of the associated permutation α.

A system of directed paths Π is said to be *non-intersecting* if any two paths in Π do not share a single vertex, i.e. for any distinct $\pi, \pi' \in \Pi$ it holds that $V(\pi) \cap V(\pi') = \varnothing$, where $V(\pi)$ is the set of vertices visited by π. In the following we denote as $\mathbf{P}(X, Y)$ the set of all systems of non-intersecting directed paths from X to Y, and for $\Pi \in \mathbf{P}(X, Y)$ we define

$$H^{(\Pi)} \coloneqq \prod_{\pi \in \Pi} h^{(\pi)}.$$

The set $\mathbf{P}(X, Y)$ is finite if G is acyclic but can be infinite if G contains cycles. If $X = \{x\}$ and $Y = \{y\}$, then we abbreviate $\mathbf{P}(\{x\}, \{y\})$ by $\mathbf{P}(x, y)$, which is simply the set of all directed paths starting at x and ending at y.

Proposition 4 (Spectral Gessel-Viennot Lemma). *Let G be a directed acyclic process graph and $\mathcal{G} \in \mathcal{T}(G)$, and $X, Y \subset V$ two subsets of equal size. Then it holds that*

$$\det([I - H]^{-1}_{X,Y}) = \sum_{\Pi \in \mathbf{P}(X,Y)} \operatorname{sgn}(\Pi) H^{(\Pi)}.$$

In particular, $\det([I - H]^{-1}_{X,Y}) \in \mathbb{R}(\Phi, z)$ *if and only if* $\mathbf{P}(X, Y) = \varnothing$. *Furthermore, for every* $z \in S^1$ *it holds that* $\det([I - H(z)]^{-1}_{X,Y}) \in \mathbb{C}(\Phi)$ *is zero if and only if* $\mathbf{P}(X, Y) = \varnothing$.

The classical Gessel-Viennot Lemma is a consequence of the spectral Gessel-Viennot Lemma. In fact, this follows from the restriction to SVAR processes that have order zero. An important implication of the Gessel-Viennot Lemma is a graphical criterion for deciding whether a given submatrix of the covariance of a linear Gaussian model is invertible or not. Analogously, one can apply the same criterion to the process graph to check whether a given subspectrum is generically invertible (as a matrix over the field of rational functions). The next proposition gives the exact criterion. But first we need to introduce the necessary notation.

Suppose $T = (\tau_1, \dots, \tau_n)$ is a system of treks. The right-hand side of T is the system of directed paths $\text{Right}(T) = (\text{Right}(\tau_i))_{i\in[n]}$, and the left-hand side is the system $\text{Left}(T) = (\text{Left}(\tau_i))_{i\in[n]}$. Finally, the top of the system T is given by the vertices $\text{top}(T) = \{\text{top}(v_i)\}_{i\in[n]}$. A system of treks is said to have a sided intersection if its right or left hand side is a system of intersecting directed paths. For subsets $X, Y \subset V$ we denote by $\mathbf{T}(X, Y)$ the set of all such trek systems T that satisfy $\text{Right}(T) \in \mathbf{P}(\text{top}(T), Y)$ and $\text{Left}(T) \in \mathbf{P}(\text{top}(\mathcal{T}), X)$. That means, the right and left sides are systems of non-intersecting directed paths, i.e. $\mathbf{T}(X, Y)$ is the set of trek systems without sided intersection.

Proposition 5 (Spectral version of [52][Proposition 3.4). *] Let X, Y be subsets of V with equal cardinality r and $z \in S^1$ a frequency. It holds that* $\text{rank}([\mathcal{S}]_{X,Y}) = \text{rank}([\mathcal{S}(z)]_{X,Y}) < r$ *if and only if* $\mathbf{T}_0(X, Y) = \varnothing$*, i.e. if there is no trek from X to Y without sided intersection.*

For subsets $X, Y \subset V$ of equal size r, combining the Cauchy-Binet formula with Proposition 4 yields

$$\det([\mathcal{S}]_{X,Y}) = \sum_{T\in\mathbf{T}(X,Y)} \text{sgn}(T)\mathcal{S}^{(T)}, \tag{13}$$

where for $T = (\tau_1, \dots, \tau_r) \in \mathbf{T}(X, Y)$ a system of treks without sided intersection and $\mathcal{S}^{(T)} = \prod_{i=1}^{r} \mathcal{S}^{(\tau_i)}$ the associated product of trek functions and $\text{sgn}(T)$ the sign of the permutation associated with T.

5. Rational identifiability of link functions

In this section, we consider time series graphs equipped with a latent structure that satisfy Assumption 1. A link between two observed processes is said to be identifiable if there is a rational function that maps the spectrum of the process onto the link function of that link, and if the form of this expression is independent of the underlying time series graph. Such a function can always be found if there are no latent processes. In this case, the link functions are identified through regression on the parent processes, i.e. for every $w \in V$ it holds that

$$[H]_{\text{pa}(w),w} = [\mathcal{S}]^{-1}_{\text{pa}(w),\text{pa}(w)}[\mathcal{S}]_{\text{pa}(w),w}.$$

When latent processes confound the effects between observed processes, it depends on the structure of the process graph whether the link functions are identifiable.

Example 8. For the process graph of Figure 1, the confounded link $v \to w$ can be identified from the spectrum if the entry $\mathcal{S}_{v,u}$ is non-zero. In this case we can use the identity:

$$h_{v,w} = \frac{\mathcal{S}_{w,u}}{\mathcal{S}_{v,u}} = \frac{h_{v,w}h_{u,v}\mathcal{S}^{\text{LI}}_u}{h_{u,v}\mathcal{S}^{\text{LI}}_u}.$$

Due to Theorem 1 the term $\mathcal{S}_{u,v}$ is generically non-zero, i.e. $h_{v,w}$ is identifiable for almost all SVAR parameter configurations in $\mathbb{R}^{\mathcal{G}}_{\text{stb}}$. Furthermore, the expression for the link function $h_{v,w}$ in terms of the spectrum $\mathcal{S}$ is independent of the underlying time series graph.

Based on the definition of *rational identifiability* used in [3,10,17], we now define rational identifiability for frequency domain causal effects in SVAR processes represented as SEPs.

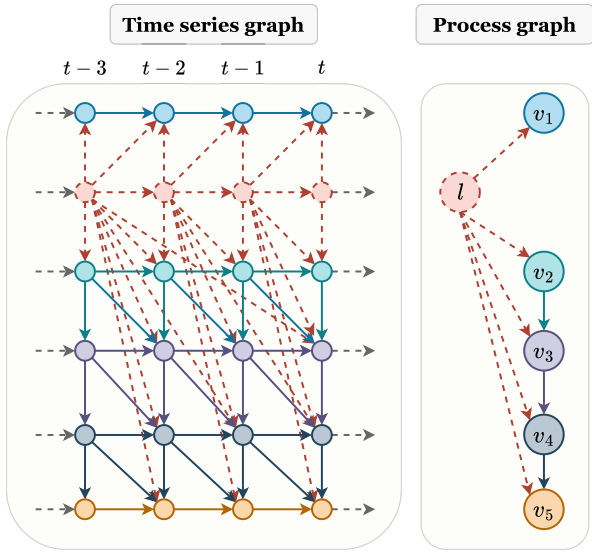

Figure 3. A time series graph (left) together with its process graph (right), see Figure 3.1 a.) in [3]. All link functions between the observed processes (represented by the blue nodes) are rationally identifiable. Furthermore, the graph satisfies the latent factor half-trek criterion

Definition 6 (Rational identifiability). A link $v \to w$ between two observed processes $v, w \in O$ is called *rationally identifiable* if there is a rational function $\psi_{v,w} : \mathrm{PD}_O(\mathbb{R}(z)) \to \mathbb{R}(z)$, and for every time series graph $\mathcal{G} \in \mathcal{T}(G)$ a proper algebraic set $A_{\mathcal{G}} \subset \overline{\mathbb{R}^{\mathcal{G}}_{\mathrm{stb}}}$ such that $\psi \circ \mathcal{S}_{\mathcal{G}}(\Phi, \Omega) = h_{v,w}(\Phi)$ for all $(\Phi, \Omega) \in \mathbb{R}^{\mathcal{G}}_{\mathrm{stb}} \setminus A_{\mathcal{G}}$. In particular, if the link $v \to w$ is rationally identifiable, then it is identifiable for almost all SVAR parameter configurations. Given a frequency $z \in S^1$, the link $v \to w$ is called *locally rationally identifiable* at $z \in S^1$ if there is a rational function $\psi_{v,w}(z) : \mathrm{PD}_O(\mathbb{C}) \to \mathbb{C}$ such that $\psi_{v,w}(z) \circ \mathcal{S}_{\mathcal{G}}(\Phi, \Omega; z) = h_{v,w}(z)$, for every $(\Phi, \Omega) \in \overline{\mathbb{R}^{\mathcal{G}}_{\mathrm{stb}}} \setminus A_{\mathcal{G}}(z)$, where $A_{\mathcal{G}}(z)$ is a proper algebraic subset.

A link $v \to w$ is rationally identifiable if its link function can be expressed through addition, multiplication and division (rational function) of the entries in the spectrum, as in Example 8. The requirement that $\psi_{v,w}$ is a rational function defined on $\mathrm{PD}_O(\mathbb{R}(z))$ makes sure that the expression to identify the link function $h_{v,w}$ is independent of the underlying time series graph. In particular, we do not need to know the specific time series graph in order to identify the link function. But if we happen to know the time series graph $\mathcal{G} \in \mathcal{T}(G)$ as well, then Lemma 3 implies that in the generic case the parameters $\phi_{v,w} = (\phi_{v,w}(k))_{k:v\to_k w}$ and $\phi_{w,w} = (\phi_{w,w}(j))_{j:w\to_j w}$ can also be uniquely recovered.

Before we generalize the latent factor half-trek identifiability criterion to SVAR processes, we revisit an example from [3] to recall the idea behind it.

Example 9 (Time series version of Example 3.1 in [3]). We consider an SVAR process $\mathbb{X}$ consisting of six processes. The causal structure between these processes is specified by the process graph and its underlying time series graph, as shown in Figure 3. Furthermore, the process has a latent structure, i.e. there are five observed processes indexed by $O = \{v_1, \ldots, v_5\}$ and one latent process $L = \{l\}$. Since the latent process has an effect on every observed process, none of the links between observed processes

can be computed by ordinary regression because

$$\mathcal{S}_{v_{i+1},v_i} = h_{v_i,v_{i+1}}\mathcal{S}_{v_i,v_i} + \mathcal{S}^{\mathrm{LI}}_{v_{i+1},v_i}.$$

Nonetheless, every link between observed processes can be identified through rational operations on the spectrum. We begin with the link function h_{v_3,v_4}. The idea is to use the fact that all the observed processes are confounded by a common latent process, which induces relations that can be exploited. To see this, we employ the trek rule and compute

$$\begin{aligned}\mathcal{S}_{v_4,v_2} &= h_{v_4,v_3}\mathcal{S}_{v_3,v_2} + \mathcal{S}^{\mathrm{LI}}_{v_4,v_2},\\ \mathcal{S}_{v_4,v_3} &= h_{v_4,v_3}\mathcal{S}_{v_3,v_3} + \mathcal{S}^{\mathrm{LI}}_{v_4,v_3} + \mathcal{S}^{\mathrm{LI}}_{v_4,v_2}h^*_{v_2,v_3}.\end{aligned}$$

In addition, the trek rule gives the identities

$$\frac{h_{l,v_4}}{h_{l,v_1}}\mathcal{S}_{v_1,v_2} = \mathcal{S}^{\mathrm{LI}}_{v_4,v_2}, \qquad \frac{h_{l,v_4}}{h_{l,v_1}}\mathcal{S}_{v_1,v_3} = \mathcal{S}^{\mathrm{LI}}_{v_4,v_3} + \mathcal{S}^{\mathrm{LI}}_{v_4,v_2}h^*_{v_2,v_3}.$$

Finally, these relations can be summarized by the linear equation system over the rational functions

$$\begin{bmatrix}\mathcal{S}_{v_4,v_2}\\ \mathcal{S}_{v_4,v_3}\end{bmatrix} = \begin{bmatrix}\mathcal{S}_{v_3,v_2} & \mathcal{S}_{v_1,v_2}\\ \mathcal{S}_{v_3,v_3} & \mathcal{S}_{v_1,v_3}\end{bmatrix}\begin{bmatrix}h_{v_3,v_4}\\ \alpha_{v_3,v_4}\end{bmatrix}, \qquad \alpha_{v_3,v_4} = \frac{h_{l,v_4}}{h_{l,v_1}}.$$

The matrix on the right-hand side of the equation is generically invertible as can be seen from Proposition 5 and the fact that $\{v_2 \leftarrow h \rightarrow v_1, v_3\}$ is a system of treks from $\{v_2, v_3\}$ to $\{v_3, v_1\}$. This means that for generic parameter choices it is possible to obtain the vector on the right by matrix inversion, i.e. rational operations on the spectrum. In particular, the link function h_{v_3,v_4} is rationally identifiable. Analogously, it is possible to see that the link function h_{v_4,v_5} is rationally identifiable.

We now consider the link function h_{v_2,v_3}. Once more we use the trek rule to identify the equations

$$\mathcal{S}_{v_3,v_1} = h_{v_2,v_3}\mathcal{S}_{v_2,v_1} + \mathcal{S}^{\mathrm{LI}}_{v_3,v_1}, \qquad \mathcal{S}_{v_3,v_2} = h_{v_2,v_3}\mathcal{S}_{v_2,v_2} + \mathcal{S}^{\mathrm{LI}}_{v_3,v_2}.$$

In order to apply a similar argument as before we would like to express the second terms in the sum as a linear factor of something that is rationally identifiable. To do this we observe that

$$\mathcal{S}_{v_4,v_1} - h_{v_3,v_4}\mathcal{S}_{v_3,v_1} = \mathcal{S}^{\mathrm{LI}}_{v_4,v_1}, \qquad \mathcal{S}_{v_4,v_2} - h_{v_3,v_4}\mathcal{S}_{v_3,v_2} = \mathcal{S}^{\mathrm{LI}}_{v_4,v_2}.$$

From what we have seen above, the expressions on the left hand side, i.e. $\mathcal{S}^{\mathrm{LI}}_{v_4,v_1}$ and $\mathcal{S}^{\mathrm{LI}}_{v_4,v_2}$ respectively, are rationally identifiable, and these two terms are linearly related to $\mathcal{S}^{\mathrm{LI}}_{v_3,v_1}$ and $\mathcal{S}^{\mathrm{LI}}_{v_3,v_2}$ respectively. So we end up with the system of linear equations

$$\begin{bmatrix}\mathcal{S}_{v_3,v_1}\\ \mathcal{S}_{v_3,v_2}\end{bmatrix} = \begin{bmatrix}\mathcal{S}_{v_2,v_1} & \mathcal{S}^{\mathrm{LI}}_{v_4,v_1}\\ \mathcal{S}_{v_2,v_2} & \mathcal{S}^{\mathrm{LI}}_{v_4,v_2}\end{bmatrix}\begin{bmatrix}h_{v_2,v_3}\\ \alpha_{v_2,v_3}\end{bmatrix}, \qquad \alpha_{v_2,v_3} = \frac{h_{l,v_3}}{h_{l,v_4}}.$$

Using similar arguments as above, we conclude that the matrix is generically invertible, so that the link function h_{v_2,v_3} is rationally identifiable. This shows that all links between observed processes are rationally identifiable.

The idea behind the above procedure, i.e. exploiting linear relations between the entries in the projected internal spectrum, can be generalized by using both graphical and algebraic arguments, leading us to

the graphical latent-factor half-trek criterion mentioned earlier. In order to formulate it and generalize it to the time series setting, we need to recall the necessary graphical notions. A *latent factor half-trek* [3] on G is a trek between two observed processes $v, w \in O$ that has one of the two shapes:

$$v \leftarrow l \rightarrow x_1 \rightarrow \cdots \rightarrow x_n \rightarrow w, \qquad v \rightarrow x_1 \rightarrow \cdots \rightarrow x_n \rightarrow w,$$

with $x_1, \ldots, x_n \in O$ and $l \in L$.

Definition 7 (Latent half-trek reachability [3][Definition 3.4).] Let $v, w \in O$ be two observed processes and $L' \subset L$ a set of latent processes. If there exists a latent factor half-trek from v to w that does not pass through any node in L', then w is called *half-trek reachable* from v while avoiding L'. The set of processes that are half-trek reachable from v while avoiding L' is denoted $\mathrm{htr}_{L'}(v)$. If $X \subset O$ is a set of observed processes, then $\mathrm{htr}_{L'}(X) = \bigcup_{x \in X} \mathrm{htr}_{L'}(x)$ is the set of processes that are half-trek reachable while avoiding L' from at least one process $x \in X$.

Definition 8 (LF-HTC [3][Definition 3.6).] Let $v \in O$ be an observed process, and $Y, W \subset O \setminus \{v\}$ two sets of observed processes and $L' \subset L$ a set of latent processes. The triple (Y, W, L') is said to satisfy the *latent factor half trek criterion (LF-HTC)* with respect to v if it satisfies the following three conditions:

1. $|Y| = |\mathrm{pa}_O(v)| + |L'|$ and $|W| = |L'|$ with $W \cap \mathrm{pa}_O(v) = \varnothing$.
2. $Y \cap W = \varnothing$ and $\mathrm{pa}_L(Y) \cap \mathrm{pa}_L(W \cup \{v\}) \subset L'$.
3. There exists a system of latent-factor half-treks from Y to $\mathrm{pa}_O(v) \cup W$ with no sided intersection, such that any half-trek pointing to $w \in W$ has the form $y \leftarrow l \rightarrow w$ for some $y \in Y$ and $l \in L'$.

We will now generalize [3][Theorem 3.7] to SVAR processes represented in the frequency domain. As in [3], we do not assume that the process graph is acyclic. Thus, in principle, the following allows not only for cyclic process graphs but also cyclic time series graphs.

Theorem 3 (spectral LF-HTC identifiability). *Suppose the triple (Y, W, L') of subsets satisfies the LF-HTC with respect to an observed process $v \in O$. If all directed link functions associated with the links $u \rightarrow y \in D_O$ such that $y \in W \cup (Y \cap \mathrm{htr}_{L'}(W \cup \{v\}))$ are rationally identifiable, then all direct link functions $h_{x,v}$, with $x \in \mathrm{pa}_O(v)$, are rationally identifiable, and they are locally rationally identifiable at every frequency $z \in S^1$.*

Note that [3][Theorem 3.7] is generally not directly applicable to the time series graph, since it is an infinite graph with a repeating structure and possible links between the latent vertices (due to auto-dependencies in the latent processes). For similar reasons we cannot apply it to the projection of the time series graph onto a finite time window, see e.g. [21,23,53]. In addition, such a projection typically contains not only latent vertices, but also bidirectional links due to the confounding imposed by the past states of the process.

Example 10. Theorem 3 allows us to prove, by purely graphical arguments, that every link in the process graph shown in Figure 3 is rationally identifiable, as we have already seen in Example 9. First, for v_1 and v_2 the triple $(\varnothing, \varnothing, \varnothing)$ satisfies the LF-HTC because $\mathrm{pa}_O(v_1) = \mathrm{pa}_O(v_2) = \varnothing$. Then we continue with the process v_4, for which we set $Y_{v_4} = \{v_2, v_3\}$ and $W_{v_4} = \{v_1\}$ and $L'_{v_4} = \{l\}$. These sets clearly satisfy conditions 1 and 2 of the definition 8. To see that the third condition is also satisfied, note that $\{v_2 \leftarrow l \rightarrow v_1, v_3, \}$ is a system of treks from $\{v_2, v_3\}$ to $\{v_1, v_3\}$ with no sided intersection. Finally, since $Y_{v_4} \cap \mathrm{htr}_{L'_{v_4}}(\{v_1, v_4)\} = \varnothing$, it follows that the link $v_3 \rightarrow v_4$ is rationally identifiable. For the

process v_5 one can choose exactly the same triple of sets to infer that $v_4 \to v_5$ is rationally identifiable. Finally, for the process v_3 we choose $Y_{v_3} = \{v_2, v_1\}$ and $W_{v_3} = \{v_4\}$ and $L'_{v_3} = \{l\}$. Again, conditions 1 and 2 of Definition 8 are easy to verify. The system of latent-factor half-treks required by condition 3 is $\{v_1 \leftarrow l \to v_4, v_2 \to v_3\}$. Since the link $v_3 \to v_4$ is rationally identifiable, it follows by Theorem 3 that $v_2 \to v_3$ is rationally identifiable.

In this paper, we have explicitly modeled latent confounding by including processes that participate in the overall SVAR process, but for which observations are not available. This is the approach taken in [3] for linear SCMs. However, in the literature on parameter identifiability for linear SCMs, latent confounding between observed variables is often modeled implicitly and represented on the graph by a bidirectional link between the confounded observed nodes. This results in a mixed graph containing both directed and bidirectional links. Such a mixed graph is parameterized by two matrices, one that parametrizes the directed links and another that parametrizes the bidirectional edges by a symmetric positive definite matrix in which an off-diagonal entry is zero if the pair of vertices associated with that entry is not connected by a bidirectional link. The question of interest is whether the matrix parameterizing the directed links can be recovered from the covariance matrix of the linear model. Several papers, such as [11,17,55], have developed criteria for mixed graphs to graphically determine generic identifiability. Since the techniques developed in this paper aim at laying the foundations for generalizing the theory from linear SCMs to SVAR models, one may ask whether the identifiability results for mixed graphs can be generalized to linear time series models, represented as SEPs on the process graph. This generalization would require that each bidirectional edge be parameterized independently of all other bidirectional edges. The frequency representation suggests parameterizing the bidirectional edges by rational functions. However, we cannot parameterize them arbitrarily, as this would give an infinite dimensional space of possible parameterizations for the bidirectional edges, and this would make identifiability results that hold up to a Lebesgue null set hard to establish. One way to parameterize the bidirectional links is the following construction. Given a mixed graph $G = (V, D, B)$, where D is the set of directed links and B is the set of bidirectional links, consider the directed graph $G^L = (V \cup L, D_L)$, where $L = \{l_{v \leftrightarrow w} \mid v \leftrightarrow w \in B\}$, i.e. a set of latent nodes indexed by the bidirectional edges, and $D_L = \{v \leftarrow l_{v,w} \to w \mid v \leftrightarrow w \in B\} \subset L \times V$. Note that the *latent projection*, see [37][Chapter 2], of $G' = (V \cup L, D \cup D_L)$ is the mixed graph G. For a time series graph $\mathcal{G}' \in \mathcal{T}(G')$ compatible with G', the projected internal spectrum $\mathcal{S}^{\mathrm{LI}}_{\mathcal{G}'}$ would be non-zero at an entry on the off-diagonal only if that entry corresponds to a bidirectional edge on G, and the entries corresponding to distinct bidirectional edges could be chosen independently. Suppose C is a criterion that a mixed graph could satisfy, and write $C(G) = 1$ if G satisfies the criterion, and $C(G) = 0$ otherwise. Suppose further that if $C(G) = 1$, then the parameters of the directed links D are rationally identifiable (for linear SCMs), such as [11,17,55]. A generalization of this result to the frequency representation of time series models would be as follows: If G' is the latent factor graph as constructed above for the mixed graph G, then we would have to show that for every $\mathcal{G}' \in \mathcal{T}(G')$, all link functions associated with the links between the observed vertices are rationally identifiable in the sense of Definition 6.

6. Discussion

A structural vector autoregressive (SVAR) process is a linear causal model for processes that evolve over a discrete set of time points. Graphically, the time series graph provides information about the detailed time lags with which the processes drive each other. The time series graph is infinite and, depending on the specific lag structure, highly complex. The complexity of the finite process graph, on the other hand, is independent of the lag structure. That means, analyzing the causal structure in SVAR processes

at the level of the process graph is combinatorially a potentially much simpler problem than reasoning about it at the level of the time series graph. The recently established frequency domain representation serves as a causal model for the process graph. The observational information of an SVAR process is represented by the spectral density, i.e. the frequency domain version of the autocovariance. In causal inference, one is interested in recovering the causal model from the observations. In the context of SVAR processes, this problem reduces to the question of which aspects of the process graph and which link functions can be uniquely identified from the spectral density. In this paper we gave a rigorous algebraic formulation of these questions. We showed that d- and t-separation statements about the process graph are generically characterized by algebraic conditions on the spectral density. Furthermore, we showed that the latent factor half trek criterion can be applied to the process graph to determine whether certain link functions are generically identifiable in terms of rational operations on the spectral density. While all results in this paper are for finite order SVAR processes, it is also possible to show them for for processes that belong to finitely parametrized subsets of infinite order SVAR processes. Details can be found in the supplementary material [42][Appendix B].

In future work it would be interesting to use the spectral characterization of d- and t-separation to estimate the process graph from observational data. The aim would be to identify the Markov equivalence class of the process graph directly, rather than identifying the time series graph first. This could be advantageous from a computational point of view, as the number of conditional independence statements to be checked on the process graph is potentially much smaller than on the underlying time series graph. As we showed in this paper, in the generic case it is possible to identify d-separations in the process graph by analyzing the spectral density evaluated at any given frequency. This observation raises the question of whether there is an optimal frequency at which a given d-separation can be identified with the least uncertainty. A possible topic for future work is to compare the statistical performance of estimating an acyclic process graph using a spectral density approach with a time series causal discovery algorithm, such as [44]. One could also explore whether the spectral density approach can be advantageously combined with time series graph based discovery schemes. A more theoretical problem for future work could be the characterization of cyclic relations on a process graph in the spectrum, assuming that the underlying time series graph does not have contemporaneous links.

It may also be worthwhile to take a closer look at the scenario where the noise terms of the SVAR model are explicitly non-Gaussian. A random variable that does not follow a Gaussian distribution has non-zero cumulants not only in degree two (covariance), but also in the higher order terms. SCMs with non-Gaussian noise terms (i.e. the non-time series case) have been found to exhibit algebraic relationships not only in the covariance, as explained by the trek rule, but also in the higher order terms [1,43]. It is possible to exploit these relationships to extract information about the underlying causal model from the observed distribution [54] under more general general confounding scenarios. Future work could investigate whether SVAR models with non-Gaussian noise terms also exhibit relationships in the higher order spectra [35,36,39] that can be used to recover the frequency domain causal model.

Acknowledgments

The authors thank two anonymous reviewers for their constructive comments and suggestions to improve the paper. N.R. thanks Nils Sturma for encouraging and helpful discussions. N.R. was previously affiliated with the German Aerospace Center (DLR), Institute of Data Science, and the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig. J.W. was previously affiliated with the Technical University of Berlin, Faculty of Electrical Engineering and Computer Science. J.R was previously affiliated with the Technical University of Berlin, Faculty of Electrical Engineering and Computer Science, and the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig.

Funding

J.W. and J.R. received funding from the European Research Council (ERC) Starting Grant CausalEarth under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 948112). N.R., A.G. and J.R. received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100 (IMIRACLI). N.R. received funding from the European Research Council (ERC) Excellent Science Grant GRAPHMODE (Grant Agreement No. 883818). J.W. received support from the German Federal Ministry of Education and Research (BMBF) as part of the project MAC-MERLin (Grant Agreement No. 01IW24007).

Supplementary Material

Supplement to: "Causal inference on process graphs: Causal structure and effect identification" (DOI: 10.3150/25-BEJ1909SUPP; .pdf). In the supplement, we recall basic terminology about stationary stochastic processes, and for the convenience of the reader, we include the proof of the structural equation process representation for SVAR processes, which originally appeared in [40]. In addition, we construct finite-dimensional spaces of infinite-order SVAR processes whose frequency representation remains in the field of rational functions. The results developed in this paper also apply to such infinite-order models. Finally, we give some details on the field of rational functions and the proofs of the results formulated in this paper.

References

[1] Améndola, C., Drton, M., Grosdos, A., Homs, R. and Robeva, E. (2023). Third-order moment varieties of linear non-Gaussian graphical models. *Inf. Inference* **12** 1405–1436. https://doi.org/10.1093/imaiai/iaad007

[2] Assaad, C.K., Devijver, E. and Gaussier, E. (2022). Survey and evaluation of causal discovery methods for time series. *J. Artificial Intelligence Res.* **73** 767–819. https://doi.org/10.1613/jair.1.13428

[3] Barber, R.F., Drton, M., Sturma, N. and Weihs, L. (2022). Half-trek criterion for identifiability of latent variable models. *Ann. Statist.* **50** 3174–3196. https://doi.org/10.1214/22-aos2221

[4] Bernanke, B.S. and Gertler, M. (1995). Inside the black box: The credit channel of monetary policy transmission. *J. Econ. Perspect.* **9** 27–48. https://doi.org/10.1257/jep.9.4.27

[5] Bollen, K.A. (1989). *Structural Equations with Latent Variables* **210**. John Wiley & Sons.

[6] Brillinger, D.R. (2001). Time series. *Soc. Ind. Appl. Math.*. https://doi.org/10.1137/1.9780898719246

[7] Brockwell, P.J. and Davis, R.A. (2009). *Time Series: Theory and Methods*, 2nd ed. New York, NY: Springer. https://doi.org/10.1007/978-1-4419-0320-4

[8] Dahlhaus, R. and Eichler, M. (2003). In *Causality and Graphical Models in Time Series Analysis*. *Oxford Statistical Science Series* 115–137. https://doi.org/10.1093/oso/9780198510550.003.0011

[9] Di, Y. (2009). t-separation and d-separation for Directed Acyclic Graphs. Preprint.

[10] Drton, M. (2018). Algebraic problems in structural equation modeling. In *The 50th Anniversary of Gröbner Bases* **77** 35–87. Tokyo: Mathematical Society of Japan. https://doi.org/10.2969/aspm/07710035

[11] Drton, M., Foygel, R. and Sullivant, S. (2011). Global identifiability of linear structural equation models. *Ann. Statist.* **39** 865–886. https://doi.org/10.1214/10-AOS859

[12] Drton, M., Sturmfels, B. and Sullivant, S. (2009). *Lectures on Algebraic Statistics*. *Oberwolfach Seminars* **39**. Springer. https://doi.org/10.1007/978-3-7643-8905-5

[13] Ebert-Uphoff, I. and Deng, Y. (2012). Causal discovery for climate research using graphical models. *J. Climate* **25** 5648–5665. https://doi.org/10.1175/JCLI-D-11-00387.1

[14] Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *J. Econometrics* **137** 334–353. https://doi.org/10.1016/j.jeconom.2005.06.032

[15] Eichler, M. and Didelez, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Anal.* **16** 3–32. https://doi.org/10.1007/s10985-009-9143-3
[16] Etesami, J., Kiyavash, N., Zhang, K. and Singhal, K. (2016). Learning network of multivariate Hawkes processes: A time series approach. In *32nd Conference on Uncertainty in Artificial Intelligence 2016, UAI 2016* (D. Janzing and A. Ihler, eds.). 162–171.
[17] Foygel, R., Draisma, J. and Drton, M. (2012). Half-trek criterion for generic identifiability of linear structural equation models. *Ann. Statist.* **40** 1682–1713. https://doi.org/10.1214/12-AOS1012
[18] Friston, K., Moran, R. and Seth, A.K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol.* **23** 172–178. Macrocircuits. https://doi.org/10.1016/j.conb.2012.11.010
[19] Geiger, D., Verma, T. and Pearl, J. (1990). Identifying independence in Bayesian networks. *Networks* **20** 507–534. https://doi.org/10.1002/net.3230200504
[20] Gelfand, I.M., Kapranov, M.M. and Zelevinsky, A.V. (1994). *Discriminants, Resultants, and Multidimensional Determinants*. Boston, MA: Birkhäuser Boston. https://doi.org/10.1007/978-0-8176-4771-1_13
[21] Gerhardus, A. (2024). Characterization of causal ancestral graphs for time series with latent confounders. *Ann. Statist.* **52** 103–130. https://doi.org/10.1214/23-AOS2325
[22] Gerhardus, A. and Runge, J. (2020). High-recall causal discovery for autocorrelated time series with latent confounders. *Adv. Neural Inf. Process. Syst.* **33** 12615–12625.
[23] Gerhardus, A., Wahl, J., Faltenbacher, S., Ninad, U. and Runge, J. (2023). Projecting infinite time series graphs to finite marginal graphs using number theory. arXiv preprint. Available at arXiv:2310.05526.
[24] Gessel, I. and Viennot, G. (1985). Binomial determinants, paths, and hook length formulae. *Adv. Math.* **58** 300–321. https://doi.org/10.1016/0001-8708(85)90121-5
[25] González-Pérez, I. (2025). Causality for VARMA processes with instantaneous effects: The global Markov property, faithfulness and instrumental variables. arXiv preprint. Available at arXiv:2501.13037.
[26] HAWKES, A.G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. https://doi.org/10.1093/biomet/58.1.83
[27] Herwartz, H., Lange, A. and Maxand, S. (2022). Data-driven identification in SVARs—when and how can statistical characteristics be used to unravel causal relationships? *Econ. Inq.* **60** 668–693. https://doi.org/10.1111/ecin.13035
[28] Hochsprung, T., Runge, J. and Gerhardus, A. (2024). A global Markov property for solutions of stochastic difference equations and the corresponding full time graphs. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence* (N. Kiyavash and J.M. Mooij, eds.). *Proceedings of Machine Learning Research* **244** 1698–1726. PMLR.
[29] Imbens, G.W. (2024). Causal inference in the social sciences. *Annu. Rev. Stat. Appl.* **11** 123–152. https://doi.org/10.1146/annurev-statistics-033121-114601
[30] Lindström, B. (1973). On the vector representations of induced matroids. *Bull. Lond. Math. Soc.* **5** 85–90. https://doi.org/10.1112/blms/5.1.85
[31] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin, Heidelberg: Springer.
[32] McGraw, M.C. and Barnes, E.A. (2018). Memory matters: A case for Granger causality in climate variability studies. *J. Climate* **31** 3289–3300. https://doi.org/10.1175/JCLI-D-17-0334.1
[33] Mogensen, S.W. (2023). Instrumental processes using integrated covariances. In *Proceedings of the Second Conference on Causal Learning and Reasoning*. *Proceedings of Machine Learning Research* **213** 620–641.
[34] Niemiro, W. and Rajkowski, Ł. (2023). Local dependence graphs for discrete time processes. In *Proceedings of the Second Conference on Causal Learning and Reasoning* (M. van der Schaar, C. Zhang and D. Janzing, eds.). *Proceedings of Machine Learning Research* **213** 772–790. PMLR.
[35] Nikias, C.L. (1993). Higher-order spectral analysis. In *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Societ* 319–319. https://doi.org/10.1109/IEMBS.1993.978564.
[36] Nikias, C.L. and Mendel, J.M. (1993). Signal processing with higher-order spectra. *IEEE Signal Process. Mag.* **10** 10–37.
[37] Pearl, J. (2009). *Causality*, 2nd ed. Cambridge University Press.
[38] Peters, J., Janzing, D. and Schlkopf, B. (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press.

[39] Petropulu, A.P. (1994). Higher-order spectra in biomedical signal processing. In *IFAC Proceedings*. *IFAC Symposium on Modelling and Control in Biomedical Systems* **27** 47–52. Galveston, TX, USA. 27–30 March 1994. https://doi.org/10.1016/S1474-6670(17)46158-1

[40] Reiter, N., Gerhardus, A., Wahl, J. and Runge, J. (2023). Causal inference on process graphs, part I: the structural equation process representation. arXiv preprint. Available at arXiv:2305.11561.

[41] Reiter, N., Wahl, J., Hegerl, G. and Runge, J. (2024). Asymptotic uncertainty in the estimation of frequency domain causal effects for linear processes. arXiv preprint. Available at arXiv:2406.18191.

[42] Reiter, N.-D., Wahl, J., Gerhardus, A. and Runge, J. (2026). Supplement to "Causal inference on process graphs: causal structure and effect identification." https://doi.org/10.3150/25-BEJ1909SUPP

[43] Robeva, E. and Seby, J.-B. (2021). Multi-trek separation in linear structural equation models. *SIAM J. Appl. Algebra Geom.* **5** 278–303. https://doi.org/10.1137/20M1316470

[44] Runge, J. (2020). Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on Uncertainty in Artificial Intelligence* 1388–1397. PMLR.

[45] Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M.D., Muñoz-Marí, J. et al. (2019). Inferring causation from time series in Earth system sciences. *Nat. Commun.* **10** 1–13. https://doi.org/10.1038/s41467-019-10105-3

[46] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. and Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.* **5** eaau4996. https://doi.org/10.1126/sciadv.aau4996

[47] Runge, J., Gerhardus, A., Varando, G., Eyring, V. and Camps-Valls, G. (2023). Causal inference for time series. *Nat. Rev., Earth Environ.* **4** 487–505. https://doi.org/10.1038/s43017-023-00431-y

[48] Seth, A.K., Barrett, A.B. and Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *J. Neurosci.* **35** 3293–3297. https://doi.org/10.1523/JNEUROSCI.4399-14.2015

[49] Siddiqi, S.H., Kording, K.P., Parvizi, J. and Fox, M.D. (2022). Causal mapping of human brain function. *Nat. Rev. Neurosci.* **23** 361–375.

[50] Spirtes, P., Glymour, C.N., Scheines, R. and Heckerman, D. (2001). *Causation, Prediction, and Search*. The MIT Press. https://doi.org/10.7551/mitpress/1754.003.0007

[51] Sullivant, S. (2008). Algebraic geometry of Gaussian Bayesian networks. *Adv. in Appl. Math.* **40** 482–513. https://doi.org/10.1016/j.aam.2007.04.004

[52] Sullivant, S., Talaska, K. and Draisma, J. (2010). Trek separation for Gaussian graphical models. *Ann. Statist.* **38** 1665–1685. https://doi.org/10.1214/09-AOS760

[53] Thams, N., Søndergaard, R., Weichwald, S. and Peters, J. (2024). Identifying causal effects using instrumental time series: Nuisance IV and correcting for the past. *J. Mach. Learn. Res.* **25** 1–51.

[54] Tramontano, D., Monod, A. and Drton, M. (2022). Learning linear non-Gaussian polytree models. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence* (J. Cussens and K. Zhang, eds.). *Proceedings of Machine Learning Research* **180** 1960–1969. PMLR.

[55] Weihs, L., Robinson, B., Dufresne, E., Kenkel, J., McGee, K.K.R. II, Reginald, M.I., Nguyen, N., Robeva, E. and Drton, M. (2018). Determinantal generalizations of instrumental variables. *J. Causal Inference* **6**. https://doi.org/10.1515/jci-2017-0009

[56] Wright, S. (1921). Correlation and causation. *J. Agric. Res.* **20** 557–585.

[57] Wright, S. (1934). The method of path coefficients. *Ann. Math. Statist.* **5** 161–215.

Received September 2024 and revised June 2025