Department of Mathematics
TUM School of Computation, Information and Technology
Technical University of Munich

Conditional Independence in Graphical Continuous Lyapunov Models

Sarah Lumpp

Thesis for the attainment of the academic degree

Master of Science

at the TUM School of Computation, Information and Technology of the Technical University of Munich

Supervisor:
Prof. Dr. Mathias Drton

Advisor:
Daniela Schkoda

Submitted:
Munich, 28.07.2023

I hereby declare that this thesis is entirely the result of my own work except where otherwise indicated. I have only used the resources given in the list of references.

Munich, 28.07.2023

Zusammenfassung

Das graphische stetige Lyapunov-Modell ist ein neuer Ansatz zur statistischen Modellierung von Abhängigkeitsstrukturen in multivariaten Daten, wobei die Strukturen Rückkopplungsschleifen enthalten können. Die Kovarianzmatrizen der Verteilungen in einem solchen Modell werden als Lösungen der stetigen Lyapunov-Gleichung über geeignete Drift- und Volatilitätsmatrizen parametrisiert, die aus einem multivariaten Ornstein-Uhlenbeck-Prozess stammen. Es wurde kürzlich gezeigt, dass zwei Variablen im Lyapunov-Modell unabhängig sind, wenn die entsprechenden Knoten im Graphen nicht durch einen Treck verbunden sind. Wir vermuten, dass auch die umgekehrte Implikation gilt, d.h. wenn zwei Knoten durch einen Treck im Graphen verbunden sind, können die entsprechenden Variablen im Lyapunov-Modell nicht bedingt unabhängig sein. In dieser Arbeit beginnen wir die Untersuchung der Hypothese mit der Betrachtung des Lyapunov-Modells eines gerichteten Pfades beliebiger Länge. Wir beweisen, dass im Pfadmodell keine bedingten Unabhängigkeiten gelten, die höchstens 100 konditionierende Variablen zwischen den beiden betrachteten Knoten beinhalten. Darüber hinaus entwickeln wir eine Methode, um jedes Gegenbeispiel für die Aussage, dass der erste und letzte Knoten des Pfades bedingt unabhägig gegeben alle Zwischenknoten sind, zu Gegenbeispielen für solche Aussagen auf jedem längeren Pfad zu erweitern. Zusätzlich illustrieren wir die Herausforderungen im Umgang mit singulären Kovarianzmatrizen, die auf dem Weg zu einem vollständigen Beweis der Vermutung für das Pfadmodell auftreten.

Abstract

The graphical continuous Lyapunov model is a new approach to statistically model dependence structures that may include feedback loops in multivariate data. The covariance matrices of the distributions in such a model are parametrized as solutions of the continuous Lyapunov equation via suitable drift and volatility matrices that stem from a multivariate Ornstein-Uhlenbeck process. It was recently shown that two variables are independent in the Lyapunov model if the corresponding nodes in the graph are not connected by a trek. We conjecture that the opposite implication also holds, meaning that if two nodes are connected by a trek in the graph, then the corresponding variables cannot be conditionally independent in the Lyapunov model. In this thesis, we start the investigation of the conjecture by considering the Lyapunov model of a directed path of arbitrary length. We prove that no conditional independence statements that involve at most 100 conditioning variables between the two considered nodes hold in the path model. We further devise a way to extend any counterexample for the statement where the first and last node of the path are conditionally independent given all intermediate nodes to counterexamples for statements on any longer path. Additionally, we illustrate the challenges of working with singular covariance matrices that arise on the way to a full proof of the conjecture for the path model.

Acknowledgements

I want to thank Prof. Mathias Drton for proposing such an exciting topic and for guiding me through the process with support and encouragement every step of the way.

I am very grateful to my advisor Daniela Schkoda for her dedicated support throughout this thesis, for instance, in the form of insightful ideas, helpful suggestions, and detailed feedback.

I am thankful for the considerable time they both invested in discussing the material with me and providing new ideas to explore.

Contents

1 Introduction

Discovering causal relationships and understanding the underlying structures in complex systems is one of the main objectives in science. *Graphical models* are a powerful tool to model, infer, and interpret these relationships (Pearl, 2009). They allow us to compactly encode multivariate distributions by means of a graph-based representation (Koller and Friedman, 2009). In such a graphical representation, every node in the graph represents a variable of the distribution while the edges model the dependence structure between the variables. A graphical model is then a set of probability distributions whose independence pattern fits the graph.

Directed graphs provide a natural representation for many types of real-world applications from medicine to economics. A *directed graph* is defined as an ordered pair $G = (V, E)$ of a set of vertices $V = \{1, \dots, p\} = [p]$ and an edge set $E \subseteq V \times V$, where $(i, j) \in E$ denotes the directed edge $i \to j$. Note that this definition allows for self-loops, i.e., edges of the form $i \to i$ for $i \in V$ can occur in G. If we restrict the edge set to $E \subseteq V \times V \setminus \{(i, i) \mid i \in V\}$ and additionally impose the condition that the graph does not have any directed cycles, G is a *directed acyclic graph*, short DAG.

Assume that we have a random vector $\mathbf{X} = (X_1, \dots, X_p)^T$ of jointly continuous random variables. Then we can use a DAG $G = (V, E)$ to model the joint distribution of the random variables by associating every node with a random variable and by requiring that the joint density of $\mathbf{X}$ factorizes with respect to the edge structure of the graph. It is useful to work with a DAG, as the acyclicity assumption makes the model easy to handle and to interpret.

One main aspect of graphical models based on DAGs is encoding the conditional independence patterns that all distributions in the model adhere to. Two random variables X and Y are conditionally independent given a third random variable Z if they are independent in the conditional distribution given Z. In the following, we denote this by $X \perp\!\!\!\perp Y \mid Z$. In a slight abuse of notation, we will often refer to random variables by their indices and therefore by their corresponding nodes. Hence, "$1 \perp\!\!\!\perp p \mid 2$" is a short notation for "$X_1 \perp\!\!\!\perp X_p \mid X_2$". The conditional independence properties of the distributions in a graphical model can be exploited, for example, in structure learning algorithms that infer the graphical structure from data.

However, in many applications, the acyclicity assumption cannot be made confidently: from complex biological systems to models of supply and demand, feedback loops occur in many structures describing parts of our world. Fitch (2019) and Varando and Hansen (2020) recently introduced a new type of graphical model – the *graphical continuous Lyapunov model*. Similar to the structural equation model, the Lyapunov model parametrizes the covariance matrix Σ of a multivariate Gaussian variable via an equation – here the continuous Lyapunov equation

$$M\Sigma + \Sigma M^T + C = 0$$

with the matrices $M, C \in \mathbb{R}^{p \times p}$ as parameters. By assuming a random vector to arise from a dynamic process in equilibrium, namely the Ornstein-Uhlenbeck process, we include a temporal perspective in the model that allows the modeling and interpretation of feedback loops. While some properties of the Lyapunov model such as the identifiability and estimation of parameters have already been considered by Fitch (2019), Varando and Hansen (2020), and Dettling et al. (2022a,b), the question of the conditional independence structure of the Lyapunov model is still open.

For our further considerations, we require the notion of a trek. A *trek* from node i to node j is a walk in a graph consisting of two directed walks starting at the same node, one ending in i and the other one in j (Varando and Hansen, 2020). These directed walks can be each of length zero, so a directed path or even a single directed edge is also a trek. Varando and Hansen (2020) found that, if there is no trek from i to j in a graph G, then X_i and X_j are marginally independent in all distributions in the Lyapunov model of G. We conjecture that these independencies – as well as all conditional independencies they induce – are the only conditional independencies that hold in the graphical continuous Lyapunov model. In other words, we conjecture that if two nodes i and j are connected by a trek, then X_i and X_j cannot be conditionally independent. Our goal is to establish a universal way of constructing distributions in the Lyapunov model as counterexamples to contradict all such conditional independence statements of nodes that are connected by a trek.

With this thesis, we start the investigation by considering directed paths – the simplest substructure of a directed graph. The *directed path* of length $p \in \mathbb{N}_{>0}$ with self-loops is given by $G_p := (V_p, E_p)$ where $V_p := [p]$ and $E_p := \{(i,j) \in V_p \times V_p \mid j = i \text{ or } j = i+1\}$. We are interested in conditional independence statements such as

$$2 \perp\!\!\!\perp 5 \mid 1, 3 \tag{1}$$

in the path model on five nodes illustrated in Figure 1. We conjecture that no such conditional independence statements hold in the model.

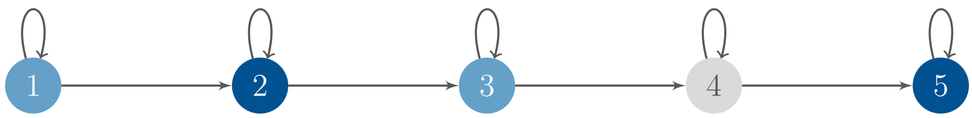


Figure 1: Directed path G_5 with self-loops on $p = 5$ nodes. The nodes are colored according to the conditional independence statement (1): the conditioning nodes (nodes 1,3) are marked in light blue and the node not appearing in the statement (node 4) is grey.

For the path model, we can formulate our conjecture and the overarching goal more formally.

Conjecture. Assume $p \in \mathbb{N}_{\geq 2}$. Let $i, j \in V_p$, $i < j$, and $S \subseteq V_p \setminus \{i, j\}$. Then, there is no conditional independence statement of the form

$$i \perp\!\!\!\perp j \mid S \tag{2}$$

that holds for all distributions in the Lyapunov model of the directed path of length p.

Goal. For every such statement as (2), find a counterexample, i.e., a distribution in the model, where the statement does not hold.

The complexity of the task at hand is evident: we will see that the number N_p of potential conditional independence relations with $S \neq \emptyset$ is exponential in the number of nodes p. In lower-dimensional settings, this is still manageable. For instance, for $p = 3$ and $p = 4$, there are $N_3 = 3$ and $N_4 = 18$ statements, respectively, for which we require counterexamples. However, for $p = 10$, there are already $N_{10} = 11475$ conditional independence statements that we want to contradict. In this thesis, we aim to find a way to generate counterexamples that generalize to varying numbers of nodes and different conditional independence patterns in the Lyapunov model of the directed path.

The thesis is structured as follows. In Chapter 2, we illustrate the limitations of the classical directed Gaussian graphical model when it comes to cyclic graphs, present the thought process leading to the new model, and formally introduce the graphical continuous Lyapunov model. Additionally, the chapter contains an excursion into the world of stochastic processes for readers who are not familiar with the subject. In Chapter 3, we recap the for our purposes relevant aspects of conditional independence in multivariate normal distributions. We formally define the path model and present a few low-dimensional examples of conditional independence relationships as well as first counterexamples.

Chapter 4 constitutes the main contribution of this thesis. We provide the theoretical foundation to construct a counterexample for (2) in the Lyapunov model of the path of length p by extending a counterexample for the statement

$$1 \perp\!\!\!\perp q \mid 2, \ldots, q-1 \tag{3}$$

in the Lyapunov model of the path of length q, where $q < p$ is suitably chosen. Using these results, we prove the conjecture postulated above for the subset of statements (2) with at most 100 conditioning variables occurring between i and j. In Chapter 5, we illustrate the challenges of working with singular covariance matrices that arise on the way to proving the conjecture for any number of conditioning variables occurring between i and j. We provide a first approach toward a general proof of the conjecture by aiming to construct a counterexample for (3).

2 The Lyapunov model

In this chapter, we give an introduction to the graphical continuous Lyapunov model. We start by revisiting the classical directed Gaussian graphical model and discussing its limitations regarding cyclic graphs. Then, we retrace the thought process in previous works to extend the model. To gain the required vocabulary and understanding of the mathematical background for the extended model, we take a short excursion into the world of stochastic processes.

Notation. First, we establish some notation and basic definitions. For easier notation, we denote vectors by bold letters; for example, $\mathbf{X} = (X_1, \ldots, X_p)^T$ is a p-dimensional random vector.

Let

$$\mathbb{R}^E := \{M := (m_{ij}) \in \mathbb{R}^{p \times p} \mid m_{ji} = 0 \text{ if } (i,j) \notin E\}$$

be the space of matrices whose sparsity patterns are given by a graph $G = (V, E)$. In other words, a non-zero entry m_{ji} of M indicates an edge $i \to j$ in the graph. We can interpret the entries of M as the edge weights in the graph G. This definition allows for zero edge weights, i.e., we can have $(i,j) \in E$ even though $m_{ji} = 0$. When we consider the matrix of edge weights $B \in \mathbb{R}^E$ of a DAG G, we can always permute the nodes in a topological order such that B is a strictly lower triangular matrix (Drton, 2018). Additionally, we define $D_+ := \{\Omega \in \mathbb{R}^{p \times p} \mid \Omega \text{ diagonal }, \Omega_{ii} > 0 \ \forall i \in [p]\}$.

Further, we write PD_p for the set of positive definite $p \times p$ matrices. They are symmetric by definition. It can be shown that the closure of PD_p is the set of positive semi-definite $p \times p$ matrices, which is convex and forms a cone (Boyd and Vandenberghe, 2004). As its interior, the set PD_p is often called the *cone of positive definite matrices* in literature.

2.1 The classical directed Gaussian graphical model

Let $\mathbf{X} = (X_1, \ldots, X_p)^T$ be a random vector consisting of jointly continuous random variables and let $G = (V, E)$ be a directed graph with $V = [p]$. Without loss of generality, we assume that $\mathbf{X}$ has mean zero. A convenient way to model the distribution of $\mathbf{X}$ is via a structural equation model. A *structural equation model* associated with the graph G models each variable as a function of its parents in the graph G together with a random noise component. Here, we assume that $\mathbf{X}$ is the solution of a system of linear structural equations

$$\mathbf{X} = B\mathbf{X} + \boldsymbol{\varepsilon} \tag{4}$$

with $B \in \mathbb{R}^E$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(0, \Omega)$ for some $\Omega \in D_+$. The structural equations are given by the edge weights of G and by the noise variables ε_i that model noise as independent Gaussian errors with mean zero.

If G is a DAG, the matrix B can always be permuted to be lower-triangular with zeros on the diagonal. Then, the matrix $I_p - B$ is lower-triangular with only ones on the diagonal and thus invertible. Consequently, we can rewrite (4) to

$$\mathbf{X} = (I_p - B)^{-1}\boldsymbol{\varepsilon} \sim \mathcal{N}_p(0, (I_p - B)^{-1}\Omega(I_p - B)^{-T}).$$

Since we assume the error vector $\boldsymbol{\varepsilon}$ to be Gaussian, the random vector $\mathbf{X}$ follows a multivariate normal distribution as well. The covariance matrix of $\mathbf{X}$ is the unique matrix Σ that solves the matrix equation

$$(I_p - B)\Sigma(I_p - B)^T = \Omega. \tag{5}$$

It depends only on the matrix of edge weights given by the graph G and the covariance matrix Ω of the Gaussian error terms. These deliberations give rise to the following formal definition (Drton, 2018).

Definition 2.1. *The* directed Gaussian graphical model *given by a DAG $G = (V, E)$ is the family of all multivariate normal distributions $\mathcal{N}(0, \Sigma)$ with covariance matrix Σ in the set*

$$\mathcal{M}_G = \{\Sigma \mid (I_p - B)\Sigma(I_p - B)^T = \Omega \text{ with } B \in \mathbb{R}^E,\ \Omega \in D_+\}.$$

From the definition, it is clear that we can identify the model with the set $\mathcal{M}_G$ of covariance matrices of distributions belonging to the model. Thus, we also refer to $\mathcal{M}_G$ as the model (Drton, 2018).

The acyclicity assumption on the graph G induces many favorable properties of the model $\mathcal{M}_G$. First, the acyclic structure of the graph admits a natural causal interpretation (Drton and Maathuis, 2017). Moreover, the density of a distribution in the model permits a convenient factorization into a product of conditional densities involving only subsets of the variables. This renders the distribution much more tractable and enables the estimation of the parameters of the model (Koller and Friedman, 2009; Drton and Maathuis, 2017; Maathuis et al., 2019). The factorization property can as well be reformulated in terms of a relationship between the conditional independence characteristics of the distribution and separation properties of the graph – the so-called *global Markov property* (Lauritzen, 1996). In other words, the graph G encodes the conditional independence pattern of the distributions in the model $\mathcal{M}_G$. This property is exploited in many applications, for example, structure learning algorithms like the PC algorithm that can infer a causal structure based on conditional independence properties of the given data (Spirtes et al., 2000).

2.2 Allowing directed cycles

The classical directed Gaussian graphical model does not allow directed cycles in the associated graphical representation. In practice, however, there is an abundance of problems and corresponding data sets that entail feedback loops among the variables and therefore directed cycles in a corresponding graph. Can we model them in a similar way? We start with an example.

Example 2.2. We consider a graph on three nodes that includes a cycle given by the following edge weight matrix

$$B = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then the matrix

$$I_3 - B = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

has determinant $\det(I_3 - B) = 1 + (-1) = 0$, so there is no unique solution Σ to equation (5). Thus, a direct extension of the model is complicated by the fact that the structural equations need not always have a unique solution.

Another aspect to consider apart from the missing density factorization properties is the matter of interpretation (Dettling et al., 2022a). In the acyclic case, the variables can be brought into a topological order that can be interpreted as a causal or at least temporal order for some variables. In the cyclic case, this is not possible since there is no universal starting point of the cycle. Both cases are depicted in Figure 2 for a graph on two nodes X_A and X_B.

(a) Acyclic graph. **(b)** Cyclic graph with self-loops.

Figure 2: Two versions of a directed graph on two nodes X_A and X_B.

Assume for instance that X_A encodes the supply of a product and X_B encodes the demand for this product. The acyclic graph assigns a clear temporal and possibly causal order to the variables: there was first supply of the product and afterward demand, perhaps created by the novelty of the product. The demand does not influence the supply in this model. The cyclic graph on the right, however, incorporates both supply influencing demand, demand influencing supply, as well as both variables influencing themselves.

It is obvious that in many applications, the cyclic model on the right is much more realistic. The causal reasoning is, however, much harder, since we cannot resort to a topological order of the variables – like in the classic “chicken-or-egg” problem we do not know what came first. A one-time measurement of X_A cannot simultaneously be taken before and after a measurement of X_B. This problem arises in scenarios where the changes occur on a much faster time scale than the measurements. For example, supply and demand are usually averaged over a certain time period (Hyttinen et al., 2012). Similar examples arise in the medical or biological context, where measurements of vitals like blood pressure and heart rate change on a fast time scale.

It is intuitively clear that a cause always precedes its effect, so the true causal structure has to be acyclic over time (Hyttinen et al., 2012). This consideration gives rise to the idea of unrolling the cyclic graph into an acyclic graph of the variables at different discrete time points t as depicted in Figure 3.

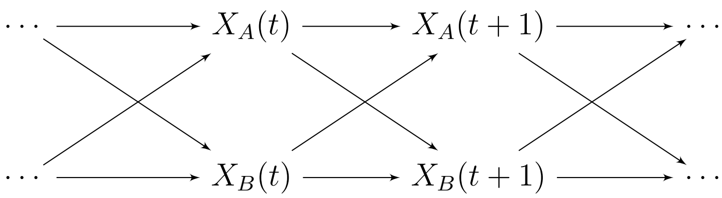


Figure 3: Cyclic graph with self-loops on two nodes X_A and X_B unrolled as an acyclic graph.

Following this way of interpretation, we adapt the model as proposed by Hyttinen et al. (2012). Given some initial values $\mathbf{X}(0)$, adapting (4) to include a temporal perspective results in

$$\begin{aligned}\mathbf{X}(t) :&= B\mathbf{X}(t-1) + \boldsymbol{\varepsilon} \\ &= B^t\mathbf{X}(0) + \sum_{i=0}^{t-1} B^i \boldsymbol{\varepsilon},\end{aligned} \tag{6}$$

where we recursively inserted $\mathbf{X}(t-1)$ into the equation. The sequence B^t as well as the series $\sum_{i=0}^{t-1} B^i$ converge for $t \to \infty$ if and only if the spectral radius of B is smaller than 1. Then, assuming that the eigenvalues of B have absolute value smaller than 1, the first term of (6) converges to zero due to $B^t \to 0$ for $t \to \infty$. Under the same assumption, the partial sum $\sum_{i=0}^{t-1} B^i$ converges to $(I_p - B)^{-1}$ for $t \to \infty$ yielding

$$\mathbf{X}(t) \xrightarrow{t\to\infty} \mathbf{X} := (I_p - B)^{-1}\boldsymbol{\varepsilon}.$$

In other words, $\mathbf{X}(t)$ converges to an equilibrium, where the value of $\mathbf{X}$ is fully determined by B and $\boldsymbol{\varepsilon}$ and is independent of $\mathbf{X}(0)$. Note that the matrix $I_p - B$ is invertible due to the restriction on the eigenvalues of B. Given a fixed value $\boldsymbol{\varepsilon}$, the equilibrium is deterministic. When considering a graph without cycles, this new interpretation coincides with the interpretation of the classical Gaussian graphical model via structural equations in the previous section.

We can extend the model by assuming the error terms $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \Omega)$ to be time-dependent as well, i.e., let $\boldsymbol{\varepsilon}(t) \sim \mathcal{N}(0, \Omega)$ be independent and identically distributed. The resulting equation that defines the model is then

$$\mathbf{X}(t) := B\mathbf{X}(t-1) + \boldsymbol{\varepsilon}(t).$$

This is now a process with stochastic dynamics that no longer has a deterministic equilibrium. We can interpret it as a VAR(1) model (Young et al., 2019), which is a first-order vector auto-regressive model for multivariate time series data with independent errors. It generalizes the first-order univariate auto-regressive model AR(1).

Still assuming that the spectral radius of B is smaller than 1, recursive computation of the covariance matrix yields

$$\mathrm{Var}(\mathbf{X}(t)) = B\mathrm{Var}(\mathbf{X}(t-1))B^T + \Omega = \ldots = \sum_{i=0}^{t} B^i\Omega(B^i)^T \to \sum_{i=0}^{\infty} B^i\Omega(B^i)^T =: \Sigma$$

for $t \to \infty$. Consequently, we have the equilibrium distribution

$$\mathbf{X} \sim \mathcal{N}(0, \Sigma)$$

with equilibrium covariance matrix Σ (Hyttinen et al., 2012; Young et al., 2019).

In equilibrium, the variance is independent of the time t, so we can give a recursive equation for Σ as

$$\Sigma = B\Sigma B^T + \Omega. \tag{7}$$

Note that Ω itself is a covariance matrix and therefore symmetric. Equation (7) is called the *discrete Lyapunov equation*. Thus, the equilibrium covariance matrix of the VAR(1) model solves the discrete Lyapunov equation. Young et al. (2019) derive some first identifyability results for a specific subset of VAR(1) models. However, they still need to impose the acyclicity constraint for asymptotic results.

Taking on a more general perspective, stochastic processes in discrete time $t = 0, 1, 2, \ldots$ are often of the form

$$X(t+1) = f(X(t), s(t)) + \varepsilon(t).$$

That is, the next state of the process depends on the current state $X(t)$ and other varying parameters $s(t)$ through a function f and is perturbed by an additional noise term $\varepsilon(t)$. Analyzing such a model can be facilitated by passing to an analogous diffusion or stochastic differential equation model. They are often more tractable while retaining the same relevant information (Karlin and Taylor, 1981). This gives rise to the idea – first proposed by Fitch (2019), then further developed by (Varando and Hansen, 2020) – of employing a more general continuous-time model to accommodate for directed cycles in a graphical model.

2.3 Stochastic processes

Before we introduce the new model, we take a brief detour into the area of stochastic processes to establish the needed vocabulary and mathematical concepts. This overview is mainly based on the comprehensive works by Karlin and Taylor (1975), Karlin and Taylor (1981), Parzen (1999), and Lawler (2006) as well as the lecture notes by Borghini (2012).

A *stochastic process* is a collection $\{X(t) \mid t \in T\}$ of random variables where the index set T is either discrete, for example $T = \mathbb{N}_0$, yielding a *discrete time process*, or otherwise an uncountable subset of $\mathbb{R}$, for example $T = \mathbb{R}_{\geq 0}$, yielding a *continuous time process*. The set of possible values of $X(t)$ is called the *state space* of the stochastic process. If the state space is finite or countable, it is called *discrete*, otherwise *continuous*. The values may be one-dimensional or multidimensional.

2.3.1 General properties

A way of describing a stochastic process is to specify the joint distribution of

$$X(t_1), \ldots, X(t_n)$$

for all $n \in \mathbb{N}$ and for all $t_1, \ldots, t_n \in T$. One example are *Gaussian processes*.

Definition 2.3. *A* Gaussian process *is a stochastic process for which, for every $n \in \mathbb{N}$ and for every finite set $\{t_1, \ldots, t_n\}$ of time points, the random vector*

$$(X(t_1), \ldots, X(t_n))^T$$

has a multivariate normal distribution.

Another way of describing a stochastic process is to give a formula for $X(t)$ at each time point t in terms of a family of random variables with known probability distribution. Stochastic processes can also arise as solutions to stochastic differential equations.

There are many properties that characterize and distinguish stochastic processes; one of them is *stationarity*.

Definition 2.4. *A stochastic process $\{X(t) \mid t \in T\}$ is said to be* stationary *if for any $n \in \mathbb{N}$, $h > 0$, and any points $t_1, \ldots, t_n \in T$ where $t_i + h \in T$ for all $i \in [n]$, the joint distribution of*

$$X(t_1), \ldots, X(t_n)$$

coincides with the joint distribution of

$$X(t_1 + h), \ldots, X(t_n + h).$$

That means the unconditional probability distribution is invariant under an arbitrary shift of all considered time points. This implies that the random mechanism of the process remains the same as time progresses. It implies in particular that the distribution of $X(t)$ is the same for all t.

2.3.2 Markov processes in general

An important subset of stochastic processes are so-called *Markov processes*. A Markov process is a stochastic process where the change at a time t is only determined by the value $X(t)$ of the process at this time and not by any values before t.

Definition 2.5. *A process is* Markov *if for any set of n time points $t_1 < \cdots < t_n \in T$ and any values $x_1, \ldots, x_n \in \mathbb{R}$*

$$P(X(t_n) \leq x_n \mid X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}) = P(X(t_n) \leq x_n \mid X(t_{n-1}) = x_{n-1})$$

holds.

In other words, the conditional distribution of $X(t_n)$ given some previous known values depends only on the most recent one. Then the probability of any future event depends only on the present state and is not altered by additional knowledge of past behavior.

Both discrete and continuous time processes can be Markov. For discrete-time processes, we assume that $T = \mathbb{N}_0$; and for continuous processes, we assume $T = \mathbb{R}_{\geq 0}$. If the state space is finite or countable, the process is called a *Markov chain*. We assume that the state space S of a Markov chain is either $S = \{0, 1, \ldots, N\}$ or $S = \mathbb{N}_0$. A continuous time Markov process where every sample path $X(t)$ is a continuous function in t is called a *diffusion process*.

A Markov process is fully determined by its initial unconditional distribution at $t = 0$ and the *transition probability function* $P(X(t) \in E \mid X(t_0) = s)$ specifying the conditional probability that the state of the system at time t is in a set E, given that at a time $t_0 < t$ it is in state s. If the transition probability function depends only through $t - t_0$ on t and t_0, the process is said to have *time-homogeneous* transition probabilities.

2.3.3 Markov chains

The probability of a discrete-time Markov chain changing from a state $X(n) = i$ to a state $X(n+1) = j$ in one time step at time n

$$P_{ij}^{(n,n+1)} := P(X(n+1) = j \mid X(n) = i)$$

is called the *one-step transition probability.* For simplicity, we assume from now on that these probabilities are time-homogeneous, i.e., independent of the time parameter n so that we can write $P_{ij} := P_{ij}^{(n,n+1)}$. The theory can, however, be easily extended to the more general non-homogeneous case. The one-step transition probabilities can be arranged in a (possibly infinite) *transition probability matrix*

$$P := (P_{ij})_{i,j=0,1,\dots}.$$

The i-th row of P is the probability distribution of the values of $X(n+1)$ given that $X(n) = i$. Therefore, the sum of all entries in each row is 1. The probability distribution of a Markov chain with time-homogeneous transition probabilities is fully determined by the initial distribution of the values of $X(0)$ given by $p_i := P(X(0) = i)$ for every $i \in S$ together with the one-step transition probabilities P_{ij} for all $i, j \in S$.

Taking this idea further, we define the n-step transition probabilities

$$P_{ij}^{(n)} := P(X(m+n) = j \mid X(m) = i),$$

i.e., the probability of the process changing from state i to state j in n steps. Due to the time-homogeneous transition probabilities, we can form the (possibly infinite) n-step transition probability matrix

$$P^{(n)} := (P_{ij}^{(n)})_{ij=0,1,\dots}.$$

It can be shown that the *Chapman Kolmogorov equation*

$$P_{ij}^{(n)} = \sum_{k \in S} P_{ik}^{(r)} P_{kj}^{(s)} \tag{8}$$

holds for any $i, j \in S$ and any $r, s \in \mathbb{N}_0$ with $r + s = n$ and $P_{ij}^{(0)} = \delta_{ij}$ (Klenke, 2014). Then, we have $P^{(1)} = P$. The probability of the process being in state j at time point n is

$$p_j^{(n)} := P(X(n) = j) = \sum_{i \in S} p_i P_{ij}^{(n)}.$$

If the state space S of the considered Markov chain is finite, the matrices $P^{(n)}$ are finite as well. Consequently, we can rewrite the previous statements in terms of matrix multiplication. We obtain

$$P^{(n)} = P^n,$$

so the n-step transition probability $P_{ij}^{(n)} = (P^n)_{ij}$ is the (i,j)-th entry of the n-th power of the one-step transition probability matrix. With the initial probability distribution given by a vector $\mathbf{p} = (p_0, p_1, \ldots, p_N)$, the unconditional distribution of $X(n)$ is given by

$$\mathbf{p^{(n)}} = \mathbf{p}P^{(n)}.$$

The concepts we have seen so far can also be extended to continuous time Markov chains where $T = \mathbb{R}_{\geq 0}$. Again, we can define transition probabilities

$$P_{ij}^{(s,s+t)} := P(X(s+t) = j \mid X(s) = i)$$

for times $s < s+t \in T$ and states $i, j \in S$. We assume the transition probabilities to be time-homogeneous so that we can write $P_{ij}^{(t)}$. As before, the theory can be easily extended to the general case.

We assume that the transition probabilities are continuous at $t = 0$ with $\lim_{t\to 0} P_{ij}^{(t)} = \delta_{ij}$. It can be shown that they satisfy the Chapman-Kolmogorov equation (8) as in the discrete-time case and that they are uniformly continuous as functions of $t > 0$. The one-step transition probabilities we defined for discrete-time Markov chains are replaced by their infinitesimal analogs that are defined through the derivatives of the transition probability functions at zero.

2.3.4 Stationary distributions

A point of interest is the asymptotic behavior of a Markov chain and the influence of the initial state of the process over time. We still assume the considered chains to have time-homogeneous transition probabilities. A discrete-time Markov chain is said to have a *limiting distribution* or *long-run distribution* if there is a probability distribution $\boldsymbol{\pi}$ such that for every $i, j \in S$ we have

$$\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j,$$

where the limit is independent of i. This implies

$$\lim_{n\to\infty} p_j^{(n)} = \lim_{n\to\infty} \sum_{i\in S} p_i P_{ij}^{(n)} = \sum_{i\in S} p_i \pi_j = \pi_j,$$

so the unconditional probability $p_j^{(n)}$ of the chain being in state j at time point n converges to a probability π_j regardless of the chosen initial unconditional distribution. Note that limiting distributions exist only in specific cases.

Definition 2.6 (Lawler (2006)). *A discrete time Markov chain has a* stationary, equilibrium, steady-state*, or* invariant *probability distribution if there exists a probability vector π such hat for every $j \in S$ we have*

$$\pi_j = \sum_{i\in S} \pi_i P_{ij}.$$

If S is finite, this equation is equivalent to $\boldsymbol{\pi} = \boldsymbol{\pi} P$ in matrix notation. The notation shows that a stationary probability distribution of a finite state space Markov chain is a left eigenvector to the eigenvalue 1 of the transition probability matrix. A stationary distribution need not exist nor be unique (Klenke, 2014). The following example illustrates that a stationary distribution is not necessarily a limiting distribution.

Example 2.7. A chain that periodically alternates between two states with transition probability matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

has $\pi = (\frac{1}{2}, \frac{1}{2})$ as a stationary distribution but $P^{(n)} = P^n$ does not converge. Therefore, no limiting distribution exists (Parzen, 1999; Lawler, 2006).

For a finite state space S, it is easy to see that a stationary distribution $\boldsymbol{\pi}$ satisfies

$$\boldsymbol{\pi} = \boldsymbol{\pi} P = (\boldsymbol{\pi} P) P = \boldsymbol{\pi} P^2 = \cdots = \boldsymbol{\pi} P^n = \boldsymbol{\pi} P^{(n)}$$

for every $n \in \mathbb{N}_{>0}$ as well. Choosing a stationary distribution $\boldsymbol{\pi}$ as the initial distribution at time point $n = 0$ yields the unconditional distribution

$$\mathbf{p}^{(n)} = \boldsymbol{\pi} P^{(n)} = \boldsymbol{\pi},$$

in the case that S is finite. This equation implies that the unconditional distribution of the Markov chain is the same at each time step n – in other words, it is stationary.

It can be shown that a time-homogeneous discrete-time Markov chain is stationary if and only if its initial distribution is a stationary distribution (Karlin and Taylor, 1975). If a limiting distribution $\boldsymbol{\pi}$ exists, it is also a stationary distribution as

$$\boldsymbol{\pi} = \lim_{n \to \infty} P^{(n)} = (\lim_{n \to \infty} P^{(n-1)}) P = \boldsymbol{\pi} P$$

shows for finite state space S (Lawler, 2006). These arguments can be extended to chains with countable state space S as well.

2.3.5 Diffusion processes

We now introduce an example of a Markov process where both the index set and the state space are continuous. A *diffusion process* is a continuous time Markov process where additionally every sample path $X(t)$ is continuous in t. We assume that $T = \mathbb{R}_{\geq 0}$ and $S = \mathbb{R}$ if not otherwise indicated. This section as well as the following sections are mainly based on the work by Karlin and Taylor (1981).

A central example of diffusion processes is the so-called *Wiener process*. It is also referred to as *Brownian Motion* since it models random continuous motion such as the displacement $\mathbf{X}(t)$ of a particle in a fluid at time t given that $\mathbf{X}(0) = 0$. It plays a fundamental role in the theory of stochastic processes and has many applications across different fields like economics or biology (Karlin and Taylor, 1975).

Definition 2.8. *A* Wiener process *or* Brownian motion *is a stochastic process* $\{X(t) \mid t \in \mathbb{R}_{\geq 0}\}$ *taking values in* $S = \mathbb{R}$ *such that*

(i) $X(0) = 0$,

(ii) it has stationary independent increments,

(iii) the increments $X(s + t) - X(s)$ *for* $s + t > s \geq 0$ *are normally distributed with mean 0 and variance* $\sigma^2 t$ *where the* variance parameter σ *is fixed,*

(iv) the paths are continuous, i.e., the maps $t \mapsto X(t)$ *are continuous.*

A d-dimensional Wiener process *is a vector valued stochastic process*

$$\mathbf{X}(t) = \left(X^1(t), \ldots, X^d(t)\right)^T$$

on the state space $S = \mathbb{R}^d$ *where each component is a Wiener process itself and the component Wiener processes are independent.*

If the variance parameter $\sigma^2 = 1$, the process is called *standard Brownian motion*. From now on if not otherwise indicated we only consider Wiener processes with $\sigma^2 = 1$, since any Wiener process can be scaled by $\frac{1}{\sigma}$ and therefore reduced to standard Brownian motion.

Since the considered stochastic processes are now continuous, the distribution of the process at time t can no longer be specified by a (possibly infinite) probability vector but by a probability density. The transition probability density $p(x, t|x_0)$ of a particle being at a position with x-coordinate $X(t_0 + t)$ at time $t_0 + t$ given that $X(t_0) = x_0$ can be shown to satisfy the following partial differential equation

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2},$$

a variant of the so-called *diffusion equation*. Under suitable boundary conditions it is uniquely solved by

$$p(x, t|x_0) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x - x_0)^2}{2t}\right). \tag{9}$$

This transition density satisfies the continuous version of the Chapman-Kolmogorov equation (8)

$$p(y, s + t|x) = \int_S p(z, s|x)p(y, t|z)dz.$$

Combining the transition density with a suitable initial distribution yields the probability density of the process at time t as

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right) \tag{10}$$

(Karlin and Taylor, 1975; Borghini, 2012).

A Wiener process has time-homogeneous transition probabilities, but it is not a stationary process. The unconditional density $p(x, t)$ shows that the variance and therefore the distribution of $X(t)$ depends on t.

So far, we considered the case $\mathbb{E}[X(t)] = 0$, i.e., the case where there is no drift to the process. A process that drifts away from the origin in a direction μ, that is, with $\mathbb{E}[X(t)] = \mu t$, can be defined similarly to Definition 2.8.

Definition 2.9. *Let $\tilde{\boldsymbol{X}}(t)$ be a d-dimensional Wiener process (where $d = 1$ is possible) with variance parameter (matrix) σ^2. Let $\boldsymbol{\mu} \in \mathbb{R}^d$. The stochastic process*

$$\mathbf{X}(t) = \tilde{\boldsymbol{X}}(t) + \boldsymbol{\mu} t$$

is called d-dimensional Wiener process with drift. *The constant $\boldsymbol{\mu}$ is called the* drift parameter*.*

It can be shown that such a process fulfills the conditions in Definition 2.8 with the only difference that the mean of the increments is $\boldsymbol{\mu} t$ instead of zero. The formula for the transition density (9) as well as the unconditional density (10) of one component can be extended accordingly (Karlin and Taylor, 1975; Lawler, 2006).

There are different alternative characterizations of diffusion processes. One way is to characterize a diffusion process by two conditions, namely the existence of the two limits

$$\begin{aligned} \lim_{h \downarrow 0} \frac{\mathbb{E}[X(t+h) - X(t) | X(t) = x]}{h} &=: \mu(x,t) \\ \text{and } \lim_{h \downarrow 0} \frac{\mathbb{E}[(X(t+h) - X(t))^2 | X(t) = x]}{h} &=: \sigma^2(x,t). \end{aligned}$$

They describe the mean and variance of the infinitesimal displacements in a one-dimensional diffusion process. The function μ is called *drift parameter* or *infinitesimal mean* and the function σ^2 is called *diffusion parameter* or *infinitesimal variance*. We assume that these infinitesimal parameters are continuous in x and t and that $\sigma^2(x,t) > 0$. If the process is time-homogeneous, they are both independent of the time t and we can write $\mu(x,t) = \mu(x)$ and $\sigma^2(x,t) = \sigma^2(x)$.

The concept can also be extended to multidimensional diffusion processes. Then, the drift $\boldsymbol{\mu}$ is a vector-valued function with

$$\mu_i(\mathbf{x}, t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[X_i(t+h) - X_i(t) | \mathbf{X}(t) = \mathbf{x}]$$

and σ^2 is a matrix-valued function with

$$\sigma^2_{ij}(\mathbf{x}, t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[(X_i(t+h) - X_i(t))(X_j(t+h) - X_j(t)) | \mathbf{X}(t) = \mathbf{x}]$$

such that $\sigma^2(\mathbf{x}, t)$ is positive definite.

Example 2.10. A Wiener process is a diffusion process with drift parameter $\mu(x) = 0$ and diffusion parameter $\sigma^2(x) = \sigma^2$ where σ^2 is a constant. A Wiener process with drift μ has drift parameter $\mu(x) = \mu$ and diffusion parameter $\sigma^2(x) = \sigma^2$ where μ and σ^2 are both constants.

We can also extend the concept of stationary or equilibrium distribution and limiting distribution to diffusion processes. If existent, a stationary density ψ satisfies

$$\psi(y) = \int \psi(x)p(y,t|x)dx$$

for all $t > 0$. Under suitable conditions, a stationary density ψ defines also a limiting distribution, i.e., it satisfies

$$\lim_{t\to\infty} p(y,t|x) = \psi(y).$$

2.3.6 Stochastic differential equations

A different way to characterize a diffusion process is in terms of a *stochastic differential equation*, i.e., a differential equation where at least one of the terms and therefore also the solution is a stochastic process. For example, we can reformulate the Wiener process with drift and arbitrary variance as a stochastic differential equation in terms of standard Brownian motion.

We revisit the example of the x-coordinate of the position $X(t)$ of a particle suspended in a fluid. The particle's motion is driven by two principal forces: a deterministic motion induced by the nature of the fluid and possible external forces on the system as well as a random movement caused by collisions and interactions with other particles. The latter can be described over short time durations by a standard Brownian motion $W(t)$. Then the displacement of the particle along the x-axis after a short period of time Δt is approximated by

$$\Delta X(t) := X(t+\Delta t) - X(t) \approx m(x,t)\Delta t + s(x,t)\Delta W(t),$$

where $\Delta W(t) = W(t+\Delta t) - W(t)$ and $x = X(t)$. Further, $m(x,t)$ is the instantaneous velocity of the particle, and $s(x,t) > 0$ is the instantaneous variance associated with collisions. The first part of the approximation is deterministic while the second component is random.

Assuming that $m(x,t)$ and $s(x,t)$ are sufficiently continuous deterministic functions, we can infer that $X(t)$ is a diffusion process. Computing the infinitesimal mean and variance of the process yields $\mu(x,t) = m(x,t)$ and $\sigma^2(x,t) = s^2(x,t)$. However, we cannot directly evaluate the limit of $\frac{\Delta X(t)}{\Delta t}$ as it does not converge due to $W(t)$ not being differentiable (Lawler, 2006). Therefore, it is not possible to deduce

$$\frac{\text{“}dX(t)\text{”}}{dt} = \mu(x,t) + \sigma(x,t)\,\frac{\text{“}dW(t)\text{”}}{dt} \tag{11}$$

as “$dW(t)/dt$” is not well-defined.

It can be shown that the analog of “$dW(t)/dt$” in discrete time would be a sequence of independent normal random variables with mean zero and unit variance often referred to as *white noise*. A continuous time white noise process whose values at all time points are independent is difficult to realize so we have to resort to a more abstract way of handling “$dW(t)/dt$” (Karlin and Taylor, 1981).

An extended version of standard differential calculus allows us to rewrite (11) as

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dW(t) \tag{12}$$

in differential notation. Such an extension is provided by the so-called *Ito calculus* that stretches the notion of the classical Riemann-Integral to stochastic processes by introducing the *Ito integral* – a stochastic integral. We will not go further into detail on the exact specifications of Ito calculus.

Equation (12) is the general form of a stochastic differential equation modeling a diffusion process $X(t)$ with infinitesimal drift and variance $\mu(x, t)$ and $\sigma^2(x, t)$, respectively. We can interpret the equation as follows: if $X(t)$ is for instance the process representing the x-coordinate of a particle in a fluid, it looks at time t like a Wiener process with drift $\mu(X(t), t)$ and variance $\sigma^2(X(t), t)$ (Lawler, 2006).

Rewriting Equation (12) in integral notation yields

$$X(t) = X(0) + \int_0^t \mu(X(u), u)du + \int_0^t \sigma(X(u), u)dW(u),$$

where the first integral term is a standard Lebesgue integral. Under specific assumptions on growth and smoothness, this equation admits a unique solution on a chosen time interval where the initial condition is specified. The solution also satisfies the Markov property (Gardiner, 1985). For our purposes, it will be sufficient to interpret the second integral term, i.e., the Ito integral with respect to a standard Brownian motion process $W(u)$, as a random noise term. In general, we can view such a diffusion process as a process determined by a deterministic force and a random force.

Equation (12) can also be extended to d-dimensional stochastic processes. Then we have

$$d\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t), t)dt + \sigma(\mathbf{X}(t), t)d\mathbf{W}(t),$$

where $\mathbf{X}(t)$ is a d-dimensional stochastic process, $\boldsymbol{\mu}(\mathbf{X}(t), t)$ a d-dimensional vector, $\mathbf{W}(t)$ a n-dimensional standard Brownian motion, and $\sigma^2(\mathbf{X}(t), t)$ a $d \times n$ matrix (Vatiwutipong and Phewchean, 2019).

If the considered process is time-homogeneous, equation (12) reduces to

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dW(t),$$

where $W(t)$ is a standard Wiener process.

Example 2.11. The Wiener process with drift $\mu \in \mathbb{R}$ and diffusion parameter $\sigma^2 > 0$ is defined by the stochastic differential equation

$$dX(t) = \mu dt + \sigma dW(t).$$

2.3.7 The Ornstein-Uhlenbeck process

The Wiener process can be used to model the position of a particle, so naturally its derivative should represent the velocity of the particle. However, a central result in the theory of diffusion processes is that the path of a Wiener process $W(t)$ is nowhere differentiable. This limitation of the Wiener process is overcome by the *Ornstein-Uhlenbeck process*. It directly models the particle's velocity as a function of time.

The one-dimensional Ornstein-Uhlenbeck process is defined by the stochastic differential equation

$$dX(t) = -\alpha X(t)dt + \sigma dW(t)$$

with infinitesimal drift and diffusion parameters given by $\mu(x) = -\alpha x$ with $\alpha > 0$ and $\sigma^2(x) = \sigma^2$. The intuition behind the drift parameter is that the frictional resistance of the surrounding medium, e.g. a fluid, is assumed to proportionally reduce the velocity: the farther the particle is away from the origin (or the long-term mean), the slower it becomes.

It can be shown that a one-dimensional Ornstein-Uhlenbeck process is uniquely given by

$$X(t) = X(0)e^{-\alpha t} + \sigma \int_0^t e^{-\alpha(t-s)} dW(s).$$

If $X(0)$ is deterministic or Gaussian distributed (and independent of the white noise for $t > 0$), the process is a Gaussian process with mean and variance

$$\mathbb{E}[X(t)] = \mathbb{E}[X(0)]e^{-\alpha t} \text{ and}$$
$$\mathrm{Var}(X(t)) = \mathrm{Var}(X(0))e^{-2\alpha t} + \sigma \int_0^t e^{-2\alpha(t-s)} ds = \left(\mathrm{Var}(X(0)) - \frac{\sigma^2}{2\alpha}\right) e^{-2\alpha t} + \frac{\sigma^2}{2\alpha}$$

(Gardiner, 1985). For a deterministic initial condition $X(0) = x_0$ for instance, it is a Gaussian process with

$$\mathbb{E}[X(t)] = xe^{-\alpha t} \text{ and } \mathrm{Var}(X(t)) = \left(-\frac{\sigma^2}{2\alpha}\right) e^{-2\alpha t} + \frac{\sigma^2}{2\alpha} = \frac{(1 - e^{-2\alpha t})\sigma^2}{2\alpha}.$$

For $t \to \infty$, the mean and variance approach values that are independent of t, thereby giving the limiting distribution $\mathcal{N}(0, \frac{\sigma^2}{2\alpha})$. Choosing this distribution as the initial distribution yields a stationary solution $X(t)$ of the stochastic differential equation with $X(t) \sim \mathcal{N}(0, \frac{\sigma^2}{2\alpha})$ for all $t \geq 0$ (Arnold, 1974; Gardiner, 1985). Given an appropriate initial condition, the Ornstein-Uhlenbeck process is in fact the only non-trivial Markov process that is Gaussian and stationary (Borghini, 2012). Intuitively, the process admits a stationary distribution due to the drift term being dependent on the current value at time t and thereby in a way balancing out the values around the long-term mean, in our case zero (Vatiwutipong and Phewchean, 2019).

The definition of the Ornstein-Uhlenbeck process generalizes to multiple dimensions as follows.

Definition 2.12 (Gardiner (1985)). *The* p-dimensional Ornstein-Uhlenbeck process *is defined by the stochastic differential equation*

$$d\mathbf{X}(t) = M\mathbf{X}(t)dt + D \cdot d\mathbf{W}(t), \tag{13}$$

where $\mathbf{W}(t)$ *is a* p*-dimensional standard Wiener process and* $M, D \in \mathbb{R}^{p \times p}$ *are constant matrices.*

The resulting process is a Gaussian process with mean and covariance matrix

$$\mathbb{E}[\mathbf{X}(t)] = \mathbb{E}[(\mathbf{X}(0)]e^{Mt} \text{ and}$$
$$\operatorname{Var}(\mathbf{X}(t)) = e^{Mt}\operatorname{Var}(\mathbf{X}(0))e^{Mt} + \int_0^t e^{M(t-s)}DD^T e^{M^T(t-s)}ds$$

(Gardiner, 1985).

In the one-dimensional case, we required the parameter α to be positive such that the factor $-\alpha$ in the drift function $\mu(x) = -\alpha x$ is negative. The equivalent in multiple dimensions is a restriction on the eigenvalues of M. It can be shown that if the real parts of the eigenvalues of M are negative, a stationary Gaussian process solving (13) exists. Its stationary distribution has mean $\mathbb{E}[\mathbf{X}(t)] = 0$ and the covariance matrix $\Sigma := \operatorname{Var}(\mathbf{X}(t))$ satisfies

$$M\Sigma + \Sigma M^T = -C$$

with $C := DD^T$ (Arnold, 1974). This equation is called the *continuous Lyapunov equation*. The matrix M encoding the interactions of the coordinates of $\mathbf{X}(t)$ is called *drift matrix*, while the matrix C is often referred to as *volatility matrix*.

Remark 2.13. It may seem odd that we dropped the minus sign of the drift term in Definition 2.12. A similar definition can of course be made by adding the minus sign in front of all occurrences of M and requiring all eigenvalues of M to have positive real part to ensure a stationary solution (Gardiner, 1985). We chose the notation without minus sign here to adhere to the definition proposed by Varando and Hansen (2020) and Dettling et al. (2022a).

Remark 2.14. Definition 2.12 only covers the special case where the long-term mean of the process is zero. The drift parameter $\boldsymbol{\mu}(\mathbf{X}(t)) = M\mathbf{X}(t)$ can more generally be specified as $\boldsymbol{\mu}(\mathbf{X}(t)) = M(\mathbf{X}(t) - \boldsymbol{\mu})$ where $\boldsymbol{\mu} \in \mathbb{R}^d$ is a fixed value giving the long-term mean of the process (Vatiwutipong and Phewchean, 2019). We restrict ourselves to the case with $\boldsymbol{\mu} = 0$, i.e., we assume that the observations are centered.

2.4 Introducing the Lyapunov model

After a detour in the area of stochastic processes, we take the newly learned concepts and apply them to the process we want to model. The discrete-time VAR(1) process we encountered in Section 2.2 is a Markov process. Its continuous-time analog is the multivariate Ornstein-Uhlenbeck process, being a multivariate continuous-time auto-regressive process. Thus, it is a natural next step to replace the discrete process developed from the classical directed Gaussian graphical model by the Ornstein-Uhlenbeck process. This approach was first proposed by Fitch (2019) and then further refined and extended by Varando and Hansen (2020) and Dettling et al. (2022a,b).

To incorporate the temporal perspective without having to rely on time-series data, we view a single multivariate observation $\mathbf{X} = (X_1, \dots, X_p)^T$ as a one-time snapshot of a multivariate dynamic process in equilibrium (Dettling et al., 2022a). This implicit introduction of a temporal context allows the modeling of feedback loops and the inference of information about cycles in the network (Young et al., 2019).

It is especially useful to reason about the equilibrium of a process instead of the process itself when the measurements are too slow or too imprecise and thus feedback loops cannot be unrolled into the actual dynamical process they depict. The original graph can then be interpreted as a shorthand notation of the change in variables at each step (Shalizi, 2013).

Therefore, we assume a random vector $\mathbf{X}$ to arise from the multivariate Ornstein-Uhlenbeck process in equilibrium. The covariance matrix of this equilibrium and thus of $\mathbf{X}$ satisfies the continuous Lyapunov equation (Dettling et al., 2022a). As discussed above, the Ornstein-Uhlenbeck process admits a stationary or equilibrium distribution if the eigenvalues of the drift matrix M have negative real part. We formalize this property in the following definition.

Definition 2.15. *A matrix is said to be* stable *if all its eigenvalues have a strictly negative real part. We denote by* $\mathrm{Stab}(p)$ *the set of stable* $p \times p$ *matrices. By*

$$\mathrm{Stab}(E) := \mathrm{Stab}(p) \cap \mathbb{R}^E,$$

we denote the subset of stable matrices with zero-pattern given by E*.*

Remark 2.16. In the special case of triangular matrices, the eigenvalues are precisely the diagonal entries. Since we only consider matrices with real entries, a triangular matrix in our setting lies in $\mathrm{Stab}(p)$ if and only if all its diagonal entries are negative. This provides us with a simple criterion for deciding whether a real-valued triangular matrix is stable.

It can be shown that the continuous Lyapunov equation has a positive definite solution Σ given a positive definite volatility matrix C if and only if the drift matrix M is stable (Bhaya et al., 2003). The solution Σ is unique provided that M is stable (Varando and Hansen, 2020). Consequently, we can define the graphical continuous Lyapunov model as the set of all multivariate normal distributions with mean zero whose covariance matrix solves the continuous Lyapunov equation for a fixed positive definite volatility matrix C and an arbitrary stable drift matrix M.

Definition 2.17. *Let* $G_p = (V, E)$ *be a directed graph with vertex set* $V := [p]$ *and edge set* E *that includes all self-loops* $i \to i, i \in [p]$*. The* graphical continuous Lyapunov model *of* G *given a fixed* $C \in \mathrm{PD}_p$ *is the family of all multivariate normal distributions* $\mathcal{N}(0, \Sigma)$ *with covariance matrix* Σ *in the set*

$$\mathcal{M}_{G,C} = \{\Sigma \in \mathrm{PD}_p \mid \exists\, M \in \mathrm{Stab}(E) : M\Sigma + \Sigma M^T + C = 0\}.$$

Again, we identify the model with the set $\mathcal{M}_{G,C}$ of covariance matrices of distributions belonging to the model and consequently refer to $\mathcal{M}_{G,C}$ as the Lyapunov model.

Remark 2.18. The more general definition of the Lyapunov model by Dettling et al. (2022b) uses the requirement $M \in \mathbb{R}^E$ in Definition 2.17 above. As noted by Bhaya et al. (2003), the Lyapunov equation has a positive definite solution Σ for some $C \in \mathrm{PD}_p$ if and only if the matrix M is stable. Therefore, we directly use the condition $M \in \mathrm{Stab}(E)$ as proposed by Dettling et al. (2022b).

Before giving a first example of a Lyapunov model, we briefly analyze some properties of such a covariance matrix Σ. The matrix $C = DD^T$ arising from the diffusion term of the Ornstein-Uhlenbeck process is symmetric. Thus, any solution Σ of the Lyapunov equation for a stable matrix M given a positive definite matrix C is also symmetric: since

$$0 = M\Sigma + \Sigma M^T + C \iff 0 = \left(M\Sigma + \Sigma M^T + C\right)^T = M\Sigma^T + \Sigma^T M^T + C$$

holds, the uniqueness of the solution implies $\Sigma = \Sigma^T$. Further, Σ is invariant under rescaling of M and C, as

$$(\lambda M)\Sigma + \Sigma(\lambda M)^T = -\lambda C \iff M\Sigma + \Sigma M^T = -C$$

holds for any $\lambda \neq 0$.

Example 2.19. Let $p = 3$ and consider the graph $G = (V, E)$ with $V = [p]$ depicted in Figure 4.

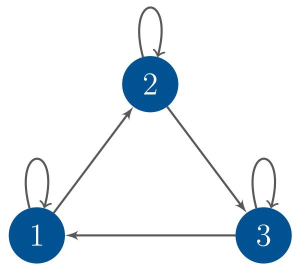


Figure 4: Graph G with self-loops on three nodes that includes a directed cycle.

Any drift matrix $M \in \mathbb{R}^E$ is of the form

$$M = \begin{pmatrix} m_{11} & 0 & m_{13} \\ m_{21} & m_{22} & 0 \\ 0 & m_{32} & m_{33} \end{pmatrix},$$

i.e., it encodes the zero pattern of the graph G. The entries of M can be viewed as the edge weights. Assuming that M is stable, the induced Lyapunov equation has a unique solution. Then, the graphical continuous Lyapunov model of the directed cycle on three nodes, given a volatility matrix C, is given by the family of multivariate normal distributions with mean zero whose covariance matrix solves the Lyapunov equation for such a matrix M. As stated in the introduction, we will consider the Lyapunov model of the directed path in the following.

3 First examples of conditional independence

In this chapter, we start by reviewing the concept of conditional independence – in particular for the multivariate normal distribution. Then, we explore as first examples the Lyapunov model of the directed path on two, three, and four nodes and investigate the conditional independence properties.

Notation. For a subset $S \subseteq [p]$ with more than one element, we write $\mathbf{X_S} := ((X_v)_{v \in S})^T$. When indexing subvectors of random vectors or submatrices of covariance matrices, we leave out set brackets and union signs. For example, let $i, j \in [p]$ and $S, Z \subseteq [p]$. For the subvector of all variables with index in $ijSZ := \{i, j\} \cup S \cup Z$ in the given order we write $\mathbf{X}_{ijSZ}$. For a submatrix of a covariance matrix Σ, we again index with the rows and columns belonging to the respective variables; for example, we write $\Sigma_{iS,jS}$ for the submatrix of rows in $iS := \{i\} \cup S$ and columns in $jS := \{j\} \cup S$ of Σ. If $S = \emptyset$, we have $\Sigma_{iS,jS} = \Sigma_{ij}$.

When considering a random vector or matrix, the expectation is taken component-wise. For a random vector $\mathbf{X} \in \mathbb{R}^p$, we write

$$\mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_p])^T$$

for its expectation and

$$\mathrm{Var}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T] = (\mathrm{Cov}(X_i, X_j))_{i,j=1,\ldots,p}$$

for its covariance matrix. Note that the covariance matrix is symmetric and positive semi-definite: for all $i, j \in [p]$, we have $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$ and for any vector $z \in \mathbb{R}^p$, we have

$$z^T \mathrm{Var}(\mathbf{X}) z = \mathbb{E}[z^T (\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T z] = \mathbb{E}[v^2] \geq 0$$

with $v := z^T (\mathbf{X} - \mathbb{E}[\mathbf{X}]) = v^T \in \mathbb{R}$.

3.1 Conditional independence

Consider a probability space $(\Omega, \mathcal{B}, P)$ with probability measure P. Two events A and B are *independent* if $P(A \cap B) = P(A)P(B)$. For $P(B) > 0$, we can define the conditional probabilities $P(A \mid B) := \frac{P(A \cap B)}{P(B)}$. Note that the set function

$$P(\,\cdot \mid B): \ \mathcal{B} \to [0,1], \ A \mapsto P(A \mid B)$$

again defines a probability measure, called the *conditional probability measure given B*. As a result, two events A and B with $P(B) > 0$ are statistically independent if and only if $P(A \mid B) = P(A)$ (Rao, 1973; Koller and Friedman, 2009).

A more common scenario is that two events are only independent given a third event. For example, if a patient has a fever, it is very likely they also have a cough. However, if we know that the patient has the flu, it is not necessary to additionally know that they also have a fever. The flu diagnosis already tells us that the patient most likely also has a cough. In mathematical terms, having a fever and having a cough are independent events given that the patient is diagnosed with the flu. Similarly as above, we say that two events A and B are *conditionally independent* given a third event C if they are independent in the conditional probability distribution given C, i.e., if for $P(C) > 0$, we have $P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$. Then again, two events A and B with $P(B \cap C) > 0$ are conditionally independent given an event C if and only if $P(A \mid B \cap C) = P(A \mid C)$ (Koller and Friedman, 2009).

These notions generalize to random variables as well, as, for example, described by Lauritzen (1996), Koller and Friedman (2009), and Edwards (2012). Let $\mathbf{X}$ and $\mathbf{Y}$ be continuous random vectors with joint density $f_{\mathbf{XY}}$ and marginal densities $f_{\mathbf{X}}, f_{\mathbf{Y}}$. Then $\mathbf{X}$ and $\mathbf{Y}$ are *(marginally) independent* if $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y})$ for all $\mathbf{x}$ and $\mathbf{y}$. Conditional densities can be defined analogously to conditional probability measures. Then, as above, the condition for independence can be reformulated in terms of the conditional density as $f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x} \mid \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})$. We reach the following definition for conditional independence of random variables.

Definition 3.1. *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be continuous random vectors with joint density $f_{\mathbf{XYZ}}$. Then $\mathbf{X}$ is* conditionally independent *of $\mathbf{Y}$ given $\mathbf{Z}$ if*

$$f_{\mathbf{XY}|\mathbf{Z}=\mathbf{z}}(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}(\mathbf{x} \mid \mathbf{z}) f_{\mathbf{Y}|\mathbf{Z}=\mathbf{z}}(\mathbf{y} \mid \mathbf{z})$$

holds for all $\mathbf{x}, \mathbf{y}, \mathbf{z}$ such that $f_{\mathbf{Z}}(\mathbf{z}) > 0$. We use the notation

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}.$$

To extend the notation to $\mathbf{Z} = \emptyset$, we write $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ if $\mathbf{X}$ and $\mathbf{Y}$ are marginally independent.

In other words, $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent given $\mathbf{Z}$ if for each value $\mathbf{z}$ of $\mathbf{Z}$, $\mathbf{X}$ and $\mathbf{Y}$ are independent in the conditional distribution given $\mathbf{Z} = \mathbf{z}$ (Højsgaard et al., 2012). Note that the density equations only have to hold almost surely with respect to the corresponding product measure.

3.2 The multivariate normal distribution

We give a short overview on the multivariate normal distribution as in (Rao, 1973, Chapter 8).

Definition 3.2. *A p-dimensional random variable $\mathbf{X} = (X_1, \dots, X_p)^T$ in $\mathbb{R}^p$ follows a* multivariate normal distribution *$\mathcal{N}_p$ if every linear function of $\mathbf{X}$ follows a univariate normal distribution.*

Remark 3.3. An equivalent definition can be given as follows: a p-dimensional random variable $\mathbf{X}$ in $\mathbb{R}^p$ follows a multivariate normal distribution $\mathcal{N}_p$ if there exists a $k \times 1$ vector of independent univariate standard normal variables such that $\mathbf{X}$ can be expressed in the form $\mathbf{X} = AY + b$, where $A \in \mathbb{R}^{k \times p}$ and $b \in \mathbb{R}^p$.

By definition, the marginal distribution of any subset of q components of $\mathbf{X}$ is again a q-variate normal distribution. Further, the components X_i are univariate normal, so $\mathbb{E}[X_i]$ and $\mathrm{Var}(X_i)$ exist for $i = 1 \dots, p$. Since $\mathrm{Var}(X_i) < \infty$ for all $i = 1 \dots, p$, due to the Cauchy-Schwarz inequality, the covariances $\mathrm{Cov}(X_i, X_j)$ exist as well. Therefore, we can define $\mu := \mathbb{E}[\mathbf{X}]$ and $\Sigma := \mathrm{Var}(\mathbf{X})$. The p-variate normal distribution is then fully specified by μ and Σ.

Definition 3.4. *Let $\mu \in \mathbb{R}^p$ and let $\Sigma \in \mathbb{R}^{p \times p}$ be a positive semi-definite matrix. Then the* multivariate normal distribution $\mathcal{N}_p(\mu, \Sigma)$ *is the p-variate normal distribution with mean vector μ and covariance matrix Σ.*

This definition includes distributions whose covariance matrices are singular and therefore have zero as eigenvalue and are consequently not positive definite. In the case that Σ is positive definite, the distribution has the usual multivariate normal density $f_{\mu,\Sigma}$. In the singular case, the distribution has no density. To investigate conditional independence in both cases, we need the following definition.

Remember that a *generalized inverse* of a matrix A is any matrix A^- satisfying

$$A = AA^-A.$$

For further details, see for example Ben-Israel and Greville (2003) and Zhang (2006).

Definition 3.5. *For a positive semi-definite matrix*

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

we define the generalized Schur complement

$$\Sigma_{11 \cdot 2} := \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}$$

of Σ_{11} where Σ_{22}^- is a generalized inverse of Σ_{22}. If Σ_{22} is invertible, then $\Sigma_{22}^- = \Sigma_{22}^{-1}$, so replacing Σ_{22}^- with the true inverse Σ_{22}^{-1} in the formula yields the Schur complement *of Σ_{11}.*

Similar definitions can be given for $\Sigma_{22 \cdot 1}$ by exchanging the indices 1 and 2 in the above formula. Note that Definition 3.5 is independent of the choice of generalized inverse Σ_{11}^- (Zhang, 2006).

Since Σ is positive semi-definite, Σ_{11} and Σ_{22} are positive semi-definite as well. It can be shown that the respective Schur complements are also positive semi-definite. The same statement holds if we replace "positive semi-definite" with "positive definite" (Zhang, 2006). If Σ is positive definite and Σ_{22} is therefore invertible, we have that $\Sigma_{22}^- = \Sigma_{22}^{-1}$.

Now, we turn our attention to the conditional independence properties of the multivariate normal distribution. We know that for positive definite covariance matrices, the normal distribution is closed under marginalization and conditioning, i.e., any marginal or conditional distribution of a normal distribution is again normal Maathuis et al. (2019). The same holds in the general case for multivariate normal distributions with positive semi-definite covariance matrices.

Proposition 3.6 (Rao (1973)). *Consider a random vector* $\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma)$ *where*

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathcal{N}_p\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

with $\mathbf{X}_1, \mu_1 \in \mathbb{R}^{d_1}$ *and* $\mathbf{X}_2, \mu_2 \in \mathbb{R}^{d_2}$*, as well as* $\Sigma_{11} \in \mathbb{R}^{d_1 \times d_1}, \Sigma_{22} \in \mathbb{R}^{d_2 \times d_2}$*, and* $\Sigma_{21} = \Sigma_{12}^T \in \mathbb{R}^{d_2 \times d_1}$ *such that* $d_1 + d_2 = p$ *and* Σ *positive semi-definite.*

Then, the marginal distribution of $\mathbf{X}_1$ *is* $\mathcal{N}_{d_1}(\mu_1, \Sigma_{11})$ *and the conditional distribution of* $\mathbf{X}_1$ *given* $\mathbf{X}_2$ *is*

$$\mathcal{N}_{d_1}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-}(\mathbf{X}_2 - \mu_2), \Sigma_{11\cdot 2}\right),$$

where $\Sigma_{11\cdot 2}$ *is the generalized Schur complement of* Σ_{11}*.*

In this general setting, we derive results for independence and conditional independence.

Corollary 3.7. *Consider the same setup as in Proposition 3.6 with positive semi-definite covariance matrix* Σ*. Then,* $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_2$ *if and only if* $\Sigma_{12} = 0$*.*

Proof. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, their covariance Σ_{12} is zero. If $\Sigma_{12} = 0$, the random variable $\mathbf{X}_1$ given $\mathbf{X}_2$ is normally distributed with mean $\mu_1 + \Sigma_{12}\Sigma_{22}^{-}(\mathbf{X}_2 - \mu_2) = \mu_1$ and covariance $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21} = \Sigma_{11}$ by Proposition 3.6. Since the mean and covariance are the same as for $\mathbf{X}_1$, we conclude that they follow the same normal distribution; therefore $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_2$. This argument is based on the reasoning in (Maathuis et al., 2019, Chapter 9). An alternative argument relying on characteristic functions is given in (Rao, 1973, Chapter 8). $\square$

Remark 3.8. Let $\mathbf{X} = (X_1, \ldots, X_p)^T \sim \mathcal{N}_p(\mu, \Sigma)$ be a random vector with positive semi-definite covariance matrix Σ and let $i, j \in [p],\ i \neq j$ be indices. It is a well-known fact that

$$X_i \perp\!\!\!\perp X_j \iff \Sigma_{ij} = 0.$$

This result can be deduced by first applying Proposition 3.6 to obtain the marginal distribution of (X_i, X_j) and then applying Corollary 3.7 to this distribution. It motivates the following Lemma.

Lemma 3.9. *Consider* $\mathbf{X} = (X_1, \ldots, X_p)^T \sim \mathcal{N}_p(\mu, \Sigma)$ *with* Σ *positive semi-definite. Let* $i, j \in [p],\ i \neq j$*, and* $S \subseteq [p] \setminus \{i, j\}$ *with* $S \neq \emptyset$*.*

(a) Then, the conditional distribution of $\mathbf{X_{ij}} = (X_i, X_j)^T$ *given* $\mathbf{X_S}$ *is normal with covariance matrix*

$$\Sigma_{ij,ij\cdot S} := \Sigma_{ij,ij} - \Sigma_{ij,S}\Sigma_{S,S}^{-}\Sigma_{S,ij} \in \mathbb{R}^{2\times 2}.$$

The conditional covariance of X_i *and* X_j *given* $\mathbf{X_S}$ *is*

$$(\Sigma_{ij,ij\cdot S})_{12} = \Sigma_{ij} - \Sigma_{i,S}\Sigma_{S,S}^{-}\Sigma_{S,j}.$$

(b) Moreover, the following statements are equivalent:

(i) $X_i \perp\!\!\!\perp X_j \mid \mathbf{X_S}$*;*

(ii) $(\Sigma_{ij,ij\cdot S})_{12} = 0$*.*

Proof. Following the notation in Proposition 3.6, we consider $\mathbf{X}_1 := \mathbf{X_{ij}}$ and $\mathbf{X}_2 := \mathbf{X_S}$ and reorder and partition the covariance matrix Σ accordingly.
(a) From Proposition 3.6, we know that the considered conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2$ is normal with covariance matrix

$$\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21} = \Sigma_{ij,ij} - \Sigma_{ij,S}\Sigma_{S,S}^{-}\Sigma_{S,ij} = \Sigma_{ij,ij\cdot S} \in \mathbb{R}^{2\times 2}$$

as defined above. The conditional covariance of X_i and X_j given $\mathbf{X_S}$ is the off-diagonal entry of the 2×2 covariance matrix $\Sigma_{11\cdot 2}$, i.e.,

$$(\Sigma_{11\cdot 2})_{12} = (\Sigma_{ij,ij\cdot S})_{12} = (\Sigma_{ij,ij})_{12} - (\Sigma_{ij,S}\Sigma_{S,S}^{-}\Sigma_{S,ij})_{12} = \Sigma_{ij} - \Sigma_{i,S}\Sigma_{S,S}^{-}\Sigma_{S,j}.$$

(b) The second statement follows directly by applying Corollary 3.7 to the conditional distribution of $(X_i, X_j) \mid \mathbf{X_S}$. □

Assume we have a p-dimensional vector $\mathbf{X}$ following a multivariate normal distribution $\mathcal{N}_p(\mu, \Sigma)$ with Σ positive semi-definite. We are, for example, interested in the statement $X_1 \perp\!\!\!\perp X_p \mid X_2, \ldots, X_{p-1}$. Then, we need to partition $\mathbf{X}$ into $\mathbf{X_{1p}} = (X_1, X_p)^T$ and $\mathbf{X_S} = (X_2, \ldots, X_{p-1})^T$ while reordering and partitioning Σ accordingly as well. The conditional covariance of X_1 and X_p given $\mathbf{X_S}$ is given by $(\Sigma_{1p,1p\cdot S})_{12} = \Sigma_{1,p} - \Sigma_{1,S}\Sigma_{S,S}^{-}\Sigma_{S,p}$ – the off-diagonal entry in the 2×2 conditional covariance matrix $\Sigma_{1p,1p\cdot S}$ of $\mathbf{X_{1p}}$ given $\mathbf{X_S}$.

If the covariance matrix Σ of a multivariate normal distribution is positive definite and therefore not singular, we can rely on a criterion that is even easier to check.

Lemma 3.10 (Maathuis et al. (2019)). *Consider $\mathbf{X} = (X_1, \ldots, X_p)^T \sim \mathcal{N}_p(\mu, \Sigma)$ with Σ positive definite. Let $i, j \in [p]$, $i \neq j$, and $S \subseteq [p] \setminus \{i, j\}$. Then the following statements are equivalent:*

(i) $X_i \perp\!\!\!\perp X_j \mid \mathbf{X_S}$;

(ii) $\det \Sigma_{iS,jS} = 0$.

Consequently, if the covariance matrix Σ in the example above is positive definite, we can simply check the condition $\det \Sigma_{1S,pS} = 0$ to determine whether $X_1 \perp\!\!\!\perp X_p \mid \mathbf{X_S}$ holds.

3.3 The directed path model

Having established the necessary theory for conditional independence in the multivariate normal distribution, we begin our investigation by considering some first examples. We are interested in the conditional independence properties of the Lyapunov model of the directed path $G_p = (V_p, E_p)$ on p nodes with self-loops as defined in the introduction. The corresponding graphical continuous Lyapunov model is given by

$$\mathcal{M}_{G_p,C_p} = \{\Sigma \in \mathrm{PD}_p \mid \exists\, M \in \mathrm{Stab}(E_p) : M\Sigma + \Sigma M^T + C_p = 0\}.$$

We fix $C_p := 2 \cdot \mathrm{I}_p$ for the remainder of this thesis. The Lyapunov model of the directed path of length p, or short "(directed) path model", is then the set of all p-variate normal distributions with mean zero and covariance matrix in $\mathcal{M}_{G_p,C_p}$. Remember that we also refer to $\mathcal{M}_{G_p,C_p}$ as the model itself. To gain intuition on the path model and conditional independence relations in the model, we consider a few examples.

The directed path on three nodes

We start by considering the directed path with self-loops on three nodes in Figure 5.

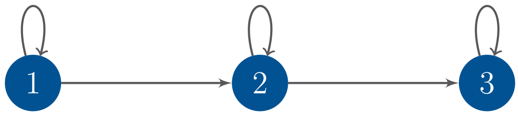


Figure 5: Directed path G with self-loops on three nodes.

The corresponding graphical continuous Lyapunov model is given by

$$\mathcal{M}_{G_3,C_3} = \{\Sigma \in \mathrm{PD}_3 \mid \exists\, M \in \mathrm{Stab}(E_3) : M\Sigma + \Sigma M^T + C_3 = 0\},$$

that is the set of all 3×3 positive definite, real matrices that fulfill the Lyapunov equation for a stable 3×3 drift matrix M and the volatility matrix C_3.

In the introduction, we postulated the conjecture that for two nodes $i, j \in V_3$ with $i < j$, and a set of nodes $S \subseteq V_3 \setminus \{i, j\}$, no conditional independence statement of the form

$$i \perp\!\!\!\perp j \mid S$$

holds in the Lyapunov model of the directed path of length 3. To prove this for $p = 3$, we want to find a counterexample, i.e., for every such statement, we want to find a distribution in the model where the statement does not hold. That means we have to check for any candidate distribution whether the condition in Lemma 3.9 is fulfilled. A convenient way to do this in the case of a positive definite covariance matrix Σ is via Lemma 3.10: we want to show that there is a $\Sigma \in \mathcal{M}_{G_3,C_3}$ such that $\det \Sigma_{iS,jS} \neq 0$. In other words, the goal is to verify that this determinant is not the zero polynomial. Thus, we aim to find a matrix $M \in \mathrm{Stab}(E_3)$ such that the resulting Σ has $\det \Sigma_{iS,jS} \neq 0$.

A matrix $M \in \mathbb{R}^{E_3}$ is of the form

$$M = \begin{pmatrix} d_1 & 0 & 0 \\ m_{21} & d_2 & 0 \\ 0 & m_{32} & d_3 \end{pmatrix}$$

with diagonal entries $d_i \in \mathbb{R}, i = 1, 2, 3$ and edge weights $m_{21}, m_{32} \in \mathbb{R}$. Then, we have $M \in \mathrm{Stab}(E_3)$ if and only if $d_1, d_2, d_3 < 0$.

We focus on the case $S \neq \emptyset$ and investigate the three possible conditional independence statements

(a) $1 \perp\!\!\!\perp 2 \mid 3$,

(b) $1 \perp\!\!\!\perp 3 \mid 2$,

(c) $2 \perp\!\!\!\perp 3 \mid 1$,

none of which we suspect to hold true in the path model on three nodes. The corresponding determinants of interest are

(a) $\det \Sigma_{13,23}$,

(b) $\det \Sigma_{12,32}$,

(c) $\det \Sigma_{21,31}$.

We consider different examples for M in the three-node case and investigate whether examples can even exist such that one of the determinants of interest vanishes.

Example 3.11. The following simple matrix

$$M = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \in \mathrm{Stab}(E_3)$$

yields the covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{4} & \frac{7}{8} & \frac{15}{8} \end{pmatrix}$$

as the solution of the Lyapunov equation. The resulting determinants are

(a) $\det \Sigma_{13,23} = \frac{1}{2} \cdot \frac{15}{8} - \frac{1}{4} \cdot \frac{7}{8} = \frac{23}{32} \neq 0$,

(b) $\det \Sigma_{12,32} = \frac{1}{4} \cdot \frac{3}{2} - \frac{1}{2} \cdot \frac{7}{8} = -\frac{1}{16} \neq 0$,

(c) $\det \Sigma_{21,31} = 1 \cdot \frac{7}{8} - \frac{1}{4} \cdot \frac{1}{2} = \frac{3}{4} \neq 0$.

None of these determinants is zero for this drift matrix M; consequently, neither can be the zero polynomial. Note that the marginal independence statements $1 \perp\!\!\!\perp 2$, $2 \perp\!\!\!\perp 3$, and $1 \perp\!\!\!\perp 3$ do also not hold in $\mathcal{M}_{G_3,C_3}$, as the corresponding entries of the covariance matrix, i.e., Σ_{12}, Σ_{23}, and Σ_{13}, are all non-zero.

From this example, we can already see that none of the three possible conditional independencies (a), (b), or (c), nor any marginal independencies hold in the path model on three nodes $\mathcal{M}_{G_3,C_3}$. However, it might still be instructive to find examples where one of the determinants of interest is zero, i.e., to find specific distributions where one of the conditional independence statements actually holds.

Example 3.12. We want to find examples where one of the determinants of interest is zero. First, we set the entries on the diagonal to a constant $d \in \mathbb{R} \setminus \{0\}$ and consider the matrix

$$M = \begin{pmatrix} d & 0 & 0 \\ m_{21} & d & 0 \\ 0 & m_{32} & d \end{pmatrix} \in \mathbb{R}^{E_3}$$

yielding

$$\Sigma = \begin{pmatrix} -\frac{1}{d} & \frac{m_{21}}{2d^2} & -\frac{m_{21}m_{32}}{4d^3} \\ \frac{m_{21}}{2d^2} & -\frac{1}{d} - \frac{m_{21}^2}{2d^3} & \frac{m_{32}}{2d^2} + \frac{3m_{21}^2 m_{32}}{8d^4} \\ -\frac{m_{21}m_{32}}{4d^3} & \frac{m_{32}}{2d^2} + \frac{3m_{21}^2 m_{32}}{8d^4} & -\frac{1}{d} - \frac{m_{32}^2}{2d^3} - \frac{3m_{21}^2 m_{32}^2}{8d^5} \end{pmatrix}$$

as the corresponding solution of the Lyapunov equation. The determinants of interest evaluate to

(a) $\det \Sigma_{13,23} = -\frac{m_{21}}{2d^3} - \frac{m_{21}m_{32}^2}{8d^5} - \frac{3m_{21}^3 m_{32}^2}{32d^7} = 0 \iff m_{21} = 0$ or $(m_{21} = \pm\frac{2\sqrt{-4d^4 - d^2 m_{32}^2}}{\sqrt{3m_{32}^2}}$ and $m_{32} \neq 0)$,

(b) $\det \Sigma_{12,32} = -\frac{m_{21}^3 m_{32}}{16d^6} = 0 \iff m_{21} = 0$ or $m_{32} = 0$,

(c) $\det \Sigma_{21,31} = -\frac{m_{32}}{2d^3} - \frac{m_{21}^2 m_{32}}{4d^5} = 0 \iff m_{32} = 0$ or $m_{21} = \pm i\sqrt{2}d$.

The only values for the entries of M resulting in zero determinants are either zero values or specific values in $\mathbb{C} \setminus \mathbb{R}$.

Thus, if we only consider matrices M with non-zero real edge weights and diagonal entries, the considered determinants are always non-zero for a matrix M with constant diagonal. Otherwise, we can, for example, choose

$$M = \begin{pmatrix} d & 0 & 0 \\ m & d & 0 \\ 0 & 0 & d \end{pmatrix} \in \mathbb{R}^{E_3}$$

with $m \in \mathbb{R} \setminus \{0\}$ yielding

$$\Sigma = \begin{pmatrix} -\frac{1}{d} & \frac{m}{2d^2} & 0 \\ \frac{m}{2d^2} & -\frac{1}{d} - \frac{m^2}{2d^3} & 0 \\ 0 & 0 & -\frac{1}{d} \end{pmatrix}$$

as the covariance matrix. Then, the first determinant is non-zero, while the other two determinants are zero. This is obvious since $1 \perp\!\!\!\perp 3$ and $2 \perp\!\!\!\perp 3$ holds in the distribution induced by Σ.

If we allow possibly distinct values $d_1, d_2, d_3 \in \mathbb{R} \setminus \{0\}$ on the diagonal, there exist cases where the Lyapunov equation has no solution.

Example 3.13. For instance, any matrix

$$M = \begin{pmatrix} -1 & 0 & 0 \\ m_{21} & 1 & 0 \\ 0 & m_{32} & -1 \end{pmatrix} \in \mathbb{R}^{E_3}$$

does not yield a solution Σ of the Lyapunov equation as M is not stable.

Example 3.14. The matrix

$$M = \begin{pmatrix} -1 & 0 & 0 \\ m_{21} & -2 & 0 \\ 0 & m_{32} & -1 \end{pmatrix} \in \mathrm{Stab}(E_3),$$

however, leads to the matrix

$$\Sigma = \begin{pmatrix} 1 & \frac{m_{21}}{3} & \frac{m_{21}m_{32}}{6} \\ \frac{m_{21}}{3} & \frac{1}{2} + \frac{{m_{21}}^2}{6} & \frac{m_{32}}{6} + \frac{m_{21}^2 m_{32}}{9} \\ \frac{m_{21}m_{32}}{6} & \frac{m_{32}}{6} + \frac{m_{21}^2 m_{32}}{9} & 1 + \frac{m_{32}^2}{6} + \frac{m_{21}^2 m_{32}^2}{9} \end{pmatrix}$$

as the solution of the Lyapunov equation. The three determinants are

(a) $\det \Sigma_{13,23} = \frac{m_{21}}{3} + \frac{m_{21}m_{32}^2}{36} + \frac{m_{21}^3 m_{32}^2}{54} = 0 \iff m_{21} = 0$ or $(m_{21} = \pm\frac{\sqrt{-3(12+m_{32}^2)}}{\sqrt{2m_{32}^2}}$ and $m_{32} \neq 0)$,

(b) $\det \Sigma_{12,32} = \frac{m_{21}m_{32}}{36} - \frac{m_{21}^3 m_{32}}{108} = 0 \iff m_{21} = 0$ or $m_{32} = 0$ or $m_{21} = \pm\sqrt{3}$,

(c) $\det \Sigma_{21,31} = \frac{m_{32}}{6} + \frac{m_{21}^2 m_{32}}{18} = 0 \iff m_{32} = 0$ or $m_{21} = \pm i\sqrt{3}$.

For determinants (a) and (c), we clearly have no real and non-zero solutions for the edge weights that would result in a zero determinant. Determinant (b) is zero, for instance, with

$$M = \begin{pmatrix} -1 & 0 & 0 \\ \sqrt{3} & -2 & 0 \\ 0 & 1 & -1 \end{pmatrix} \in \mathrm{Stab}(E_3).$$

A similar result can be reached by setting $d_3 = -3$: the matrix

$$M = \begin{pmatrix} -1 & 0 & 0 \\ \sqrt{3} & -2 & 0 \\ 0 & 1 & -3 \end{pmatrix}$$

also renders determinant (b) zero, whereas both other determinants are non-zero for any real non-zero edge weights. A different order of the diagonal entries, however, gives different results. For instance, with $d_1 = -2$, $d_2 = -1$, $d_3 = -1$, as well as $d_1 = -1$, $d_2 = -1$, $d_3 = -2$, or $d_1 = -3$, $d_2 = -2$, $d_3 = -1$, there are no real non-zero values for the subdiagonal of M that would render one of the determinants zero.

We successfully constructed examples where the second determinant (b) is zero, but we have not yet found examples of matrices M such that determinants (a) or (c) become zero without setting the edge weights m_{21} or m_{32} to zero.

Example 3.15. We consider the general case with arbitrary real non-zero diagonal entries and edge weights. The drift matrix

$$M = \begin{pmatrix} d_1 & 0 & 0 \\ m_{21} & d_2 & 0 \\ 0 & m_{32} & d_3 \end{pmatrix} \in \mathbb{R}^{E_3}$$

yields a symmetric covariance matrix

$$\Sigma = \begin{pmatrix} -\frac{1}{d_1} & \frac{m_{21}}{d_1(d_1+d_2)} & -\frac{m_{21}m_{32}}{d_1(d_1+d_2)(d_1+d_3)} \\ * & -\frac{1}{d_2} - \frac{{m_{21}}^2}{d_1 d_2(d_1+d_2)} & \frac{(d_1(d_1+d_2)(d_1+d_3)+(d_1+d_2+d_3)m_{21}^2)m_{32}}{d_1 d_2(d_1+d_2)(d_1+d_3)(d_2+d_3)} \\ * & * & -\frac{1}{d_3} - \frac{(d_1(d_1+d_2)(d_1+d_3)+(d_1+d_2+d_3)m_{21}^2)m_{32}^2}{d_1 d_2 d_3(d_1+d_2)(d_1+d_3)(d_2+d_3)} \end{pmatrix}.$$

For a unique solution Σ to exist, we require $d_1 + d_2 \neq 0$, $d_2 + d_3 \neq 0$, and $d_1 + d_3 \neq 0$. This holds, for example, for a stable matrix. The resulting determinants are

(a) $\det \Sigma_{13,23} = 0 \iff m_{21} = 0$ or $(m_{21} = \pm i\frac{\sqrt{(d_1+d_2)(d_1+d_3)}\sqrt{d_2(d_1+d_3)(d_2+d_3)+d_1 m_{32}^2}}{\sqrt{(d_1+d_2+d_3)m_{32}^2}}$ and $m_{32} \neq 0)$ or $(d_3 = -(d_1 + d_2)$ and $m_{32} = \pm i d_2)$,

(b) $\det \Sigma_{12,32} = 0 \iff m_{21} = 0$ or $m_{32} = 0$ or $m_{21} = \pm\sqrt{-d_1^2 + d_2^2}$,

(c) $\det \Sigma_{21,31} = 0 \iff m_{32} = 0$ or $m_{21} = \pm i \frac{\sqrt{d_3+d_1}(d_1+d_2)}{\sqrt{d_1+2d_2+d_3}}$.

For the first determinant to be zero, we have three conditions one of which has to be fulfilled. If we want to find an example with non-zero edge weights m_{21} and m_{32}, the second and third conditions are of particular interest. Fulfilling the third condition would imply that m_{32} is non-real, which we are not interested in. Only, meeting the second condition might lead to a drift matrix M with real and non-zero edge weights and diagonal entries.

In the second condition, however, for m_{21} to be real, we need that either all three radicands are negative or exactly one radicand is negative while the other two are positive. In all four of those cases, the condition implies that there have to be positive and negative entries on the diagonal, meaning the matrix M cannot be chosen stable to render the first determinant zero. Thus, we cannot find a matrix $\Sigma \in \mathcal{M}_{G_3,C_3}$ with non-zero edge weights such that the first determinant is zero. An example with $m_{21} = 0$ is given by the drift matrix

$$M = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \in \mathrm{Stab}(E_3)$$

yielding the matrix

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{pmatrix}$$

with $\det \Sigma_{13,23} = 0$.

For the second determinant, m_{21} is non-zero, real, and fulfills the solution condition if and only if $|d_1| < |d_2|$. To get a zero determinant, we can choose arbitrary non-zero, real diagonal entries satisfying $|d_1| < |d_2|$ and thereby specify two possibilities for m_{21}. The remaining edge weight m_{32} and diagonal entry d_3 can be set to arbitrary (non-zero) real values. We have already found such a matrix M in example Example 3.14.

The third determinant is zero if either m_{32} is zero or if we find a value for m_{21} that fulfills the condition while being non-zero and real. The entry m_{21} is real and non-zero if and only if ($d_1 + d_3 > 0$ and $d_1 + 2d_2 + d_3 < 0$) or ($d_1 + d_3 < 0$ and $d_1 + 2d_2 + d_3 > 0$). Note that this entry cannot be real for either only positive or only negative diagonal entries. It is non-zero if and only if $d_3 \neq -d_1$ and $d_2 \neq -d_1$. For instance, the drift matrix

$$M = \begin{pmatrix} -1 & 0 & 0 \\ 4 & 3 & 0 \\ 0 & 1 & -2 \end{pmatrix} \in \mathbb{R}^{E_3}$$

yields

$$\Sigma = \begin{pmatrix} 1 & -1 & -\frac{1}{3} \\ -1 & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

as the solution of the Lyapunov equation. While determinant (c) is zero in this example, the matrix M is not stable and, consequently, Σ is not in the model $\mathcal{M}_{G_3,C_3}$.

Apparently, it is not trivial to find distributions in the Lyapunov model where the determinants of interest are zero without resorting to zero edge weights on the subdiagonal. After having investigated conditional independence in the path model on three nodes, we consider a few more examples with varying numbers of nodes.

Marginal independence on the directed path

For completeness, we also consider the case $S = \emptyset$, i.e., the case of marginal independence.

Example 3.16. On the directed path of length $p = 2$, we naturally have $S = \emptyset$. Consider $\mathcal{M}_{G_2,C_2}$ and let

$$M = \begin{pmatrix} d_1 & 0 \\ m_{21} & d_2 \end{pmatrix} \in \mathbb{R}^{E_2}.$$

The Lyapunov equation is

$$0 = M\Sigma + \Sigma M^T + 2I_2 = \begin{pmatrix} 2d_1\Sigma_{11} + 2 & \Sigma_{12}(d_1 + d_2) + \Sigma_{11}m_{21} \\ \Sigma_{21}(d_1 + d_2) + \Sigma_{11}m_{21} & 2d_2\Sigma_{22} + m_{21}(\Sigma_{12} + \Sigma_{21}) + 2 \end{pmatrix}.$$

By solving the equations given by the four entries, we obtain

$$\Sigma_{11} = -\frac{1}{d_1},$$

$$\Sigma_{12} = \Sigma_{21} = \frac{m_{21}}{d_1(d_1 + d_2)},$$

and

$$\Sigma_{22} = -\frac{1}{d_2}\left(\frac{m_{21}^2}{d_1(d_1 + d_2)} + 1\right) = -\frac{1}{d_2} - \frac{m_{21}^2}{d_1 d_2(d_1 + d_2)}.$$

For a unique solution Σ to exist, we require $d_1 + d_2 \neq 0$. This is fulfilled if M is stable. We can directly see that

$$1 \perp\!\!\!\perp 2 \iff \Sigma_{12} = 0 \iff m_{21} = 0.$$

Thus, any stable 2×2 matrix M with non-zero edge weight and diagonal entries yields a valid counterexample to the statement. One such matrix is – not surprisingly –

$$M = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} \in \mathrm{Stab}(E_2).$$

To construct a distribution in the model where $1 \perp\!\!\!\perp 2$ actually holds, we have to set $m_{21} = 0$. Such a drift matrix is, for example, given by $M = -I_2$. The resulting covariance matrix is $\Sigma = I_2$.

Example 3.17. On the directed path on $p = 3$ nodes, we can as well find examples of distributions in the Lyapunov model where the marginal independencies do hold. By setting the edge weights m_{21} and m_{32} to zero and the diagonal entries of M to -1 as before, we have $M = -I_3$, yielding $\Sigma = I_3$. In the corresponding distribution, all variables are pairwise independent.

Remark 3.18. We can extend this reasoning to construct a distribution with pairwise independency of any two nodes on the directed path of length p. Consider the statement

$$i \perp\!\!\!\perp j$$

in the model $\mathcal{M}_{G_p,C_p}$. By setting all edge weights to zero, we remove any influence between the nodes. The corresponding drift matrix $M = -I_p$ is the negative of the identity matrix, reducing the Lyapunov equation to

$$-2\Sigma + 2I_p = 0.$$

In the distribution defined by the solution $\Sigma = I_p$, all nodes are pairwise independent.

Longer directed paths

By extending the first simple counterexample we found in Example 3.11, we can also construct counterexamples to conditional independence statements on longer paths.

Example 3.19. Consider the Lyapunov model $\mathcal{M}_{G_4,C_4}$ of the directed path on 4 nodes. We have

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{2} & \frac{3}{2} & \frac{7}{8} & \frac{1}{2} \\ \frac{1}{4} & \frac{7}{8} & \frac{15}{8} & \frac{19}{16} \\ \frac{1}{8} & \frac{1}{2} & \frac{19}{16} & \frac{15}{16} \end{pmatrix} \in \mathcal{M}_{G_4,C_4},$$

since it solves the Lyapunov equation given by the drift matrix

$$M = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \in \mathrm{Stab}(E_4).$$

For example, the statement

$$2 \perp\!\!\!\perp 4 \mid 1,3$$

does not hold in the model, as

$$\det \Sigma_{213,413} = -\frac{19}{256} \neq 0.$$

Comparing this covariance matrix to the covariance matrix in Example 3.11, we see that the 3×3 covariance matrix is a leading principal submatrix of this 4×4 matrix. Let us take a closer look at the corresponding Lyapunov equation of the model on four nodes. Solving it in terms of the entries of Σ, we obtain the following recursive representation

$$\Sigma = \begin{pmatrix} 1 & \frac{\Sigma_{11}}{2} & \frac{\Sigma_{12}}{2} & \frac{\Sigma_{13}}{2} \\ * & \frac{\Sigma_{12}+\Sigma_{21}}{2} & \frac{\Sigma_{22}+\Sigma_{13}}{2} & \frac{\Sigma_{23}+\Sigma_{14}}{2} \\ * & * & \frac{\Sigma_{23}+\Sigma_{32}}{2} & \frac{\Sigma_{33}+\Sigma_{24}}{2} \\ * & * & * & \frac{\Sigma_{34}+\Sigma_{43}}{2} \end{pmatrix}$$

of the symmetric covariance matrix Σ. Starting in the second row and column, every entry is the arithmetic mean of the entry on the left and the entry above. The entries in the first row and column are powers of $\frac{1}{2}$. By increasing the size of M and Σ accordingly, this pattern propagates through Σ. Thus, if we add another row and column to M with 1 on the first subdiagonal and -1 on the diagonal, the first four rows and columns of the resulting covariance matrix are the same as for the path on four nodes but with a fifth row and column added according to this pattern.

Example 3.20. Consider the directed path with self-loops on five nodes. Based on the observations above, we know that

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{2} & \frac{3}{2} & \frac{7}{8} & \frac{1}{2} & \frac{9}{32} \\ \frac{1}{4} & \frac{7}{8} & \frac{15}{8} & \frac{19}{16} & \frac{47}{64} \\ \frac{1}{8} & \frac{1}{2} & \frac{19}{16} & \frac{15}{16} & \frac{107}{128} \\ \frac{1}{16} & \frac{9}{32} & \frac{47}{64} & \frac{107}{128} & \frac{235}{128} \end{pmatrix}$$

lies in $\mathcal{M}_{G_5,C_5}$, as it is the solution of the Lyapunov equation induced by

$$M = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \in \mathrm{Stab}(E_5).$$

In this distribution, the conditional independence statement $2 \perp\!\!\!\perp 4 \mid 1,3$ does not hold either. The corresponding determinant is the same as in the previous example, where we already saw that the determinant is non-zero.

This observation provides an efficient way to solve the Lyapunov equation for this specific drift matrix M in higher dimensions. If we know the solution Σ in lower dimensions, we only need to compute the entries in the additional rows and columns. This finding is especially helpful for simulations with computer algebra systems such as Mathematica, as it significantly increases the speed of computation of Σ.

Now, we have constructed a few low-dimensional examples of distributions in the Lyapunov model of the directed path. We saw that it is possible to construct counterexamples for all possible conditional or marginal independence statements on the path on three nodes. Further, we found that it is not straightforward to construct distributions where some of the conditional independence statements hold, such that these distributions are actually in the model. These observations motivate us in the following chapter to find a general way to construct distributions in the Lyapunov model as counterexamples to conditional independence statements.

4 Conditional independence in the path model

This chapter marks the main contribution of the thesis. We propose a theorem that verifies the conjecture postulated in the introduction for conditional independence statements up to a certain number of variables. In order to prove the theorem, we develop several lemmas to reduce a general conditional independence statement to a standard form. Then, we combine these lemmas and prove the theorem.

As introduced in the previous chapter, the Lyapunov model of the directed path of length p is formally given by

$$\mathcal{M}_{G_p,C_p} = \{\Sigma \in \mathrm{PD}_p \mid \exists\ M \in \mathrm{Stab}(E_p) : M\Sigma + \Sigma M^T + C_p = 0\},$$

where we fix $C_p := 2 \cdot \mathrm{I}_p$. Remember that the matrices $M \in \mathrm{Stab}(E_p)$ are lower triangular matrices with negative diagonal entries, arbitrary entries on the first subdiagonal and zero entries everywhere else.

As stated in the introduction, we aim to show that no conditional independence relations hold in the path model. Note that when we say that a conditional independence statement does not hold in the Lyapunov model, we mean that there exists a distribution in the Lyapunov model such that the statement does not hold in this distribution. The conditional independence statement can still hold for some distributions in the model. We saw in the previous examples that it is convenient to employ the determinant condition in Lemma 3.10 to check whether the condition for conditional independence in Lemma 3.9 is fulfilled. Therefore, we reformulate the conjecture and overarching goal accordingly.

Conjecture Assume $p \in \mathbb{N}_{\geq 2}$. Let $i, j \in V_p$, $i < j$, and $S \subseteq V_p \setminus \{i, j\}$. Then, there is no conditional independence statement of the form

$$i \perp\!\!\!\perp j \mid S \tag{14}$$

that holds for all distributions in the Lyapunov model of the directed path of length p, i.e., there is always a matrix $\Sigma \in \mathcal{M}_{G_p,C_p}$ such that $\det \Sigma_{iS,jS} \neq 0$.

Goal For every such statement as (14), find a matrix $M \in \mathrm{Stab}(E_p)$ such that the resulting solution of the Lyapunov equation $\Sigma \in \mathcal{M}_{G_p,C_p}$ has

$$\det \Sigma_{iS,jS} \neq 0.$$

4.1 Approach

Proving the conjecture is a complex endeavor as the number N_p of potential conditional independence relations with $S \neq \emptyset$ is exponential in the number of nodes p. This can be shown with a short combinatorial argument: if we first pick i and j such that $i < j$ (i.e., counting $i \perp\!\!\!\perp j \mid S$ and $j \perp\!\!\!\perp i \mid S$ as one statement) and then form S by picking l nodes for every possible l from the remaining $p-2$ nodes such that $s_1 < \cdots < s_l$, we reach

$$N_p = \sum_{l=1}^{p-2} \frac{p(p-1)}{2}\binom{p-2}{l} = \frac{p(p-1)}{2}\left(\sum_{l=0}^{p-2}\binom{p-2}{l}1^l - 1\right) = \frac{p(p-1)}{2}(2^{p-2}-1)$$

as the number of potential conditional independence statements for p nodes. Therefore, we aim to find a way to extend known counterexamples as seen in the previous chapter to varying numbers of nodes and different conditional independence patterns.

The following theorem is the main result of this thesis. It states that any conditional independence statement that involves at most 100 conditioning variables occurring between i and j does not hold in the directed path model, thus proving the conjecture for a restricted set of conditional independence relations.

Theorem 4.1. *Consider the Lyapunov model on the directed path $\mathcal{M}_{G_p,C_p}$ with $p \geq 2$. Let $i, j \in V_p$ such that $i < j$ and let $S \subseteq V_p \setminus \{i, j\}$ be a set of nodes. Define*

$$Z := S \cap \{i, \dots, j\}$$

as the subset of S containing all conditioning variables that lie between i and j.
If $|Z| \leq 100$, the conditional independence statement

$$i \perp\!\!\!\perp j \mid S$$

does not hold for all distributions in the model $\mathcal{M}_{G_p,C_p}$.

Before verifying the statement rigorously, we sketch the main ideas behind the proof.

Example 4.2. Assume we want to prove that the statement

$$2 \perp\!\!\!\perp 6 \mid 1, 4$$

does not hold in the Lyapunov model on $p = 7$ nodes. We already know that the covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{4} & \frac{7}{8} & \frac{15}{8} \end{pmatrix} \text{ with drift matrix } M = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

defines a counterexample to the statement $1 \perp\!\!\!\perp 3 \mid 2$ on a three-node directed path. The idea is now to extend this example to a counterexample for statement (14) on seven nodes by adding nodes to the model accordingly while preserving the conditional independence structure of the nodes in the smaller example.

How can we achieve this? There are two main strategies we follow: to an existing counterexample we add nodes that are independent from all existing nodes and nodes that are perfectly correlated with existing nodes. Both types of added nodes do not or do only slightly change the conditional independence structure of the original example. In the above example this means adding an independent node before the first and after the last node of the example on three nodes. Additionally, we need to add a node that is perfectly correlated with node 1 and one that is perfectly correlated with node 2 in the example on three nodes. That is, nodes that behave exactly like the existing nodes and therefore do not change the conditional independence structure.

Adding nodes in this example means extending the given distribution on three nodes to a distribution on seven nodes. Note that we cannot simply add rows and columns to

$\Sigma^* \in \mathcal{M}_{G_3,C_3}$ to create a $\Sigma \in \mathcal{M}_{G_p,C_p}$ since the model $\mathcal{M}_{G_p,C_p}$ is parametrized via stable matrices M that fulfill the Lyapunov equation together with Σ. Therefore, we have to construct drift matrices M that yield suitable covariance matrices Σ as solutions to the Lyapunov equation. These covariance matrices then contain the rows and columns of a covariance matrix of a counterexample on fewer nodes as a submatrix.

If we change our perspective by taking a path of length p as a starting point for our deliberations, we can also formulate the idea as follows: we construct a counterexample on the longer path of length p in such a way that crossing out rows and corresponding columns of the covariance matrix $\Sigma \in \mathcal{M}_{G_p,C_p}$ yields a covariance matrix $\Sigma^* \in \mathcal{M}_{G_{p^*},C_{p^*}}$ that now induces a distribution of the model on the shorter path of length $p^* < p$. We formalize the idea of embedding a smaller counterexample into a larger counterexample as a projection onto a submatrix.

Definition 4.3. *Let $p \geq 2$, $\emptyset \neq K \subseteq [p]$, and $p^* := |K|$. Define the projection map*

$$\begin{aligned} \Pi_K : \mathcal{M}_{G_p,C_p} &\to \mathbb{R}^{(p^*)\times(p^*)}, \\ \Sigma = (\Sigma_{ij})_{i,j=1,\dots,p} &\mapsto (\Sigma_{ij})_{i,j\in K} = \Sigma_{K,K} \end{aligned}$$

that gives the submatrix of Σ with rows and columns in K.

If $K = [p] \setminus \{k\} = \{k\}^C$ for some $k \in [p]$, we write $\Pi_{(-k)} := \Pi_{[p]\setminus\{k\}}$ for the map that removes the k-th row and column from a matrix.

The projection $\Pi_{(-k)}$ can be used to "cross out" the k-th node of the directed path on p nodes by applying it to a suitable covariance matrix. When extending a known counterexample to a counterexample on more nodes, we want the extension to be defined in such a way that applying a suitable projection map to the newly constructed example yields the existing smaller example.

It is important to note that not for every $\Sigma \in \mathcal{M}_{G_p,C_p}$, we have $\Pi_{(-k)}(\Sigma) \in \mathcal{M}_{G_{p-1},C_{p-1}}$. Not every distribution in the Lyapunov model on the directed path on p nodes results in a new distribution in the Lyapunov model on the directed path of $p-1$ nodes when crossing out one node, as illustrated in the following example.

Example 4.4. Consider the path model on four nodes. We saw in Example 3.19 that the drift matrix $M \in \mathrm{Stab}(E_4)$ with only -1 on the diagonal and only 1 on the first subdiagonal induces a covariance matrix $\Sigma \in \mathcal{M}_{G_4,C_4}$. Crossing out the third row and column of Σ yields the 3×3 matrix

$$\Sigma^* := \Pi_{(-3)}(\Sigma) = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{8} \\ \frac{1}{2} & \frac{3}{2} & \frac{1}{2} \\ \frac{1}{8} & \frac{1}{2} & \frac{35}{16} \end{pmatrix}.$$

This matrix, however, lies not in $\mathcal{M}_{G_3,C_3}$ since there is no $M^* \in \mathrm{Stab}(E_3)$ solving the induced Lyapunov equation

$$M^*\Sigma^* + \Sigma^*(M^*)^T + C_3 = 0.$$

For instance, if we require a solution M^* to be a lower triangular matrix, the resulting unique solution of the Lyapunov equation is

$$M^* = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ -\frac{9}{323} & \frac{575}{1292} & -\frac{180}{323} \end{pmatrix},$$

which is stable but not in $\mathbb{R}^{E_3}$. Therefore, Σ^* is not in the model $\mathcal{M}_{G_3,C_3}$ of the three node sub-path.

4.2 Constructing independent nodes

Since we already found suitable counterexamples for conditional independence statements on two, three, and four nodes, it is a natural question to ask whether we can use those examples to construct suitable counterexamples on a longer path.

4.2.1 Independent nodes not in the statement

We start by illustrating a simple way to embed a counterexample on a shorter path into a new counterexample on a longer path.

Example 4.5. Let $p = 5$ in the Lyapunov model of the directed path. Does $2 \perp\!\!\!\perp 4 \mid 3$ hold in this model? The variables involved in the statement correspond to nodes on the sub-path from node 2 to 4, while the nodes 1 and 5 do not appear. This observation motivates the following construction. We have already seen that the matrix

$$M^* = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

defines a suitable counterexample for the corresponding statement $1 \perp\!\!\!\perp 3 \mid 2$ on a three-node path since the resulting covariance matrix Σ^* has $\det \Sigma^*_{12,32} \neq 0$. If we extend this drift matrix to five nodes by adding -1 on the diagonal and setting the remaining entries to zero, the resulting matrix

$$M = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & & & & 0 \\ 0 & & M^* & & 0 \\ 0 & & & & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

is again stable, and yields

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{7}{8} & 0 \\ 0 & \frac{1}{4} & \frac{7}{8} & \frac{15}{8} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & & & & 0 \\ 0 & & \Sigma^* & & 0 \\ 0 & & & & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

as the solution of the Lyapunov equation. Calculating the determinant of interest

$$\det \Sigma_{23,43} = -\det \begin{pmatrix} \Sigma_{23} & \Sigma_{24} \\ \Sigma_{33} & \Sigma_{34} \end{pmatrix} = -\det \begin{pmatrix} \Sigma^*_{12} & \Sigma^*_{13} \\ \Sigma^*_{22} & \Sigma^*_{23} \end{pmatrix} = -\frac{1}{16} \neq 0$$

then boils down to calculating the same determinant on the covariance matrix of the three-node example, which we already know to be non-zero.

The resulting covariance matrix in the example shows that by setting the edge weights between nodes 1 and 2 as well as 4 and 5 to zero, the first and the last nodes become independent from all other nodes. This is consistent with the intuition that by setting the edge weights to zero, we remove any influence of either the first node on the following nodes or the first four nodes on the last node. Consequently, in Example 4.5, we constructed a distribution on the path of length 5 by adding a new independent variable each at the beginning and end of the path, thereby extending the existing distribution on the path of length 3.

This observation gives rise to a general approach of embedding a distribution of the Lyapunov model on a shorter path into a distribution of the Lyapunov model on a longer path by adding independent variables.

Consider two nodes a and b with $a < b$ on the path of length p and let $p^* := b - a + 1 < p$. We want to extend a p^*-dimensional distribution in the Lyapunov model $\mathcal{M}_{G_{p^*}, C_{p^*}}$ on the path from a to b to a p-dimensional distribution in the Lyapunov model of the full path of length p. In some cases, searching for a counterexample to a conditional independence statement on the full path can be reduced to searching for a counterexample to the corresponding statement on a shorter path with a and b suitably defined.

We follow the same strategy as in the example by extending the drift matrix of the p^*-dimensional model with -1 entries on the diagonal and 0 entries on the first subdiagonal. To map the relevant indices of the path of length p to a shorter path of length p^*, we define a suitable bijection.

Definition 4.6. *For $a, b \in \mathbb{N}_{>0}$ with $a < b$, we define the index map*

$$\varphi_{a,b} : \{a, \dots, b\} \to \{1, \dots, b - a + 1\}, \quad x \mapsto x - a + 1.$$

Before we formalize the idea in a lemma, we give a short example.

Example 4.7. Consider the statement $2 \perp\!\!\!\perp 4 \mid 3$ on $p = 5$ nodes. Let $a = 2$ and $b = 4$ with $\varphi := \varphi_{a,b}$ and $p^* = 3$. We obtain $\varphi(2) = 1$, $\varphi(3) = 2$, and $\varphi(4) = 3$. In Example 4.5, we saw that $2 \perp\!\!\!\perp 4 \mid 3$ does not hold in the model on five nodes if $1 \perp\!\!\!\perp 3 \mid 2$ does not hold in the model on three nodes.

Lemma 4.8. *Let $p \geq 2$ and $a, b \in [p]$ such that $a < b$. Define $K := \{a, \dots, b\} \subset [p]$ and $p^* := b - a + 1$. Let $\Sigma^* \in \mathcal{M}_{G_{p^*}, C_{p^*}}$. Then, there is a matrix $\Sigma \in \mathcal{M}_{G_p, C_p}$ such that*

$$\Pi_K(\Sigma) = \Sigma^*,$$

and

$$\Pi_{K^C}(\Sigma) = I_{p-p^*},$$

and all other entries of Σ are zero.

Proof. Since for $a = 1$ and $b = p$ there is nothing to prove, we only consider the case where $K \subsetneq [p]$, so $a > 1$ or $b < p$. We define $\varphi := \varphi_{a,b}$ as in Definition 4.6.

Let $\Sigma^* \in \mathcal{M}_{G_{p^*}, C_{p^*}}$. Then, there is a stable matrix $M^* := (m^*_{kl}) \in \mathbb{R}^{p^* \times p^*}$ such that the Lyapunov equation

$$M^*\Sigma^* + \Sigma^*(M^*)^T + C_{p^*} = 0$$

is fulfilled.

First, we consider the case $1 < a$ and $b < p$. Let

$$M := (m_{kl}) = \begin{pmatrix} -I_{a-1} & 0 & 0 \\ 0 & M^* & 0 \\ 0 & 0 & -I_{p-b} \end{pmatrix} \in \mathbb{R}^{p \times p},$$

so we have

$$m_{kl} = \begin{cases} -1, & \text{if } k = l < a \text{ or } k = l > b; \\ m^*_{\varphi(k)\varphi(l)}, & \text{if } a \leq k, l \leq b; \\ 0, & \text{else.} \end{cases}$$

Thereby, we extend the matrix M^* with -1 on the diagonal and 0 on the first sub-diagonal and fill the remaining entries with zeros. Note that since M^* is stable and the added diagonal entries are negative, M is again stable.

In the following, we leave out the specific dimensions of the identity matrices for better readability. Computing the left-hand side of the Lyapunov equation by block matrix multiplication with

$$\Sigma = \begin{pmatrix} A & B & C \\ D & E & F \\ G & H & J \end{pmatrix} \in \mathbb{R}^{p \times p},$$

where $A \in \mathbb{R}^{(a-1)\times(a-1)}$, $E \in \mathbb{R}^{p^* \times p^*}$, and $J \in \mathbb{R}^{(p-b)\times(p-b)}$, now yields

$$\begin{aligned} &M\Sigma + \Sigma M^T + C_p \\ &= \begin{pmatrix} -I & 0 & 0 \\ 0 & M^* & 0 \\ 0 & 0 & -I \end{pmatrix} \begin{pmatrix} A & B & C \\ D & E & F \\ G & H & J \end{pmatrix} + \begin{pmatrix} A & B & C \\ D & E & F \\ G & H & J \end{pmatrix} \begin{pmatrix} -I & 0 & 0 \\ 0 & (M^*)^T & 0 \\ 0 & 0 & -I \end{pmatrix} + C_p \\ &= \begin{pmatrix} -2A + 2I & B(M^* - I)^T & -2C \\ (M^* - I)D & M^*E + E(M^*)T + C_{p^*} & (M^* - I)F \\ -2G & H(M^* - I)^T & -2I + 2J \end{pmatrix}. \end{aligned} \tag{15}$$

Setting (15) to zero and solving for the block partitions of Σ directly gives $A = I_{a-1}$, $J = I_{p-b}$, $C = G^T = 0$, and $E = \Sigma^*$. The remaining equations to solve are

$$\begin{aligned} &(M^* - I)D = 0 \text{ and } (M^* - I)B^T = 0 \text{ for } D, B^T \in \mathbb{R}^{p^* \times (a-1)} \text{ and} \\ &(M^* - I)F = 0 \text{ and } (M^* - I)H^T = 0 \text{ for } F, H^T \in \mathbb{R}^{p^* \times (p-b)}. \end{aligned}$$

We already see $B^T = D$ and $H^T = F$ due to the symmetry of the Lyapunov equation. The matrix M^* is a lower triangular matrix. Since the matrix is stable, all its diagonal entries are strictly negative, so $M^* - I$ has non-zero entries only on the diagonal. Therefore, the kernel of $M^* - I$ contains only zero, hence $B^T = D = 0$ and $H^T = F = 0$. Then,

$$\Sigma = \begin{pmatrix} I_{a-1} & 0 & 0 \\ 0 & \Sigma^* & 0 \\ 0 & 0 & I_{p-b} \end{pmatrix}, \text{ i.e., } \Sigma_{kl} = \begin{cases} 1, & \text{if } k = l < a \text{ or } k = l > b; \\ \Sigma^*_{\varphi(k)\varphi(l)}, & \text{if } a \leq k, l \leq b; \\ 0, & \text{else.} \end{cases}$$

Thus, $\Pi_K(\Sigma) = \Sigma_{K,K} = \Sigma^*$, $\Pi_{K^C}(\Sigma) = \Sigma_{K^C,K^C} = I_{p-p^*}$, and all other entries of Σ are zero.

Now consider the case where either $a = 1$ or $b = p$. Here, we can perform the same calculations as above by leaving out the corresponding parts of the matrices. If $a = 1$, we leave out the corresponding rows and columns of the first block of the inverse identity matrix in the definition of M and compute the left-hand side of the Lyapunov equation with

$$M = \begin{pmatrix} M^* & 0 \\ 0 & -I_{p-b} \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} E & F \\ H & J \end{pmatrix}.$$

This yields the covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma^* & 0 \\ 0 & I_{p-b} \end{pmatrix}$$

with $\Sigma_{k,l} = \Sigma^*_{\varphi(k),\varphi(l)}$ for all $k, l \leq b$, so $\Pi_K(\Sigma) = \Sigma^*$, $\Pi_{K^C}(\Sigma) = I_{p-b}$, and all other entries of Σ are zero.

If $b = p$, the same holds with

$$M = \begin{pmatrix} -I_{a-1} & 0 \\ 0 & M^* \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} A & B \\ D & E \end{pmatrix} = \begin{pmatrix} I_{a-1} & 0 \\ 0 & \Sigma^* \end{pmatrix},$$

where $\Sigma_{k,l} = \Sigma^*_{\varphi(k),\varphi(l)}$ for all $k, l \geq a$, so $\Pi_K(\Sigma) = \Sigma^*$, $\Pi_{K^C}(\Sigma) = I_{a-1}$, and all other entries of Σ are zero. □

Remark 4.9. Let $K \subsetneq [p]$ and $p^* := |K|$. Then, Lemma 4.8 says that for every matrix $\Sigma^* \in \mathcal{M}_{G_{p^*},C_{p^*}}$, we find a $\Sigma \in \mathcal{M}_{G_p,C_p}$ with $\Pi_K(\Sigma) = \Sigma^*$. This can be reformulated as

$$\mathcal{M}_{G_{p^*},C_{p^*}} \subseteq \Pi_K(\mathcal{M}_{G_p,C_p}).$$

Now we employ the lemma to show that any counterexample for the statement (14) can be extended to a counterexample for the same statement on a longer path by adding independent nodes at the start and end of the path.

Corollary 4.10. *Consider the conditional independence statement* (14) *in the model* $\mathcal{M}_{G_p,C_p}$ *with* $p \geq 2$*. Let* $a := \min(\{i\} \cup S)$ *and* $b := \max(\{j\} \cup S)$*. Define* $p^* := b - a + 1$ *and* $\varphi := \varphi_{a,b}$*. If there exists a distribution in the model* $\mathcal{M}_{G_{p^*},C_{p^*}}$ *such that*

$$\varphi(i) \perp\!\!\!\perp \varphi(j) \mid \varphi(S)$$

does not hold in this distribution, then there exists a distribution in $\mathcal{M}_{G_p,C_p}$ *for which*

$$i \perp\!\!\!\perp j \mid S$$

does not hold.

Proof. If $p = 2$, we are done. We assume from now on that $p \geq 3$.

Assume that there exists a distribution in the model $\mathcal{M}_{G_{p^*}, C_{p^*}}$ such that the conditional independence statement $\varphi(i) \perp\!\!\!\perp \varphi(j) \mid \varphi(S)$ does not hold in this distribution. In other words, the statement does not hold in the Lyapunov model of the path of length p^*. Hence, there exists $\Sigma^* \in \mathcal{M}_{G_{p^*}, C_{p^*}}$ with

$$\det \Sigma^*_{\varphi(i)\varphi(S), \varphi(j)\varphi(S)} \neq 0.$$

Let $K := \{a, \dots, b\} \subseteq [p]$. Due to Lemma 4.8, there is $\Sigma \in \mathcal{M}_{G_p, C_p}$ such that

$$\Pi_K(\Sigma) = \Sigma^*. \tag{16}$$

The submatrix $\Sigma_{iS,jS}$ selects only rows of Σ with indices in $\{i\} \cup S$ and only columns with indices in $\{j\} \cup S$. So for all entries Σ_{kl} of Σ occurring in the determinant of interest, we have $a \leq k, l \leq b$ and therefore $k, l \in K$. Thus, due to (16), the determinant evaluates to

$$\det \Sigma_{iS,jS} = \det(\Pi_K(\Sigma))_{\varphi(i)\varphi(S), \varphi(j)\varphi(S)} = \det \Sigma^*_{\varphi(i)\varphi(S), \varphi(j)\varphi(S)} \neq 0.$$

Using the criterion in Lemma 3.10, we deduce that $i \perp\!\!\!\perp j \mid S$ does not hold in the distribution defined by Σ. Therefore, we have found a distribution in the model $\mathcal{M}_{G_p, C_p}$ for which $i \perp\!\!\!\perp j \mid S$ does not hold. $\square$

4.2.2 Independent nodes in the statement

Our next goal is to be able to ignore all remaining nodes that occur before i and after j on the path – regardless of whether they are conditioning variables in S or not. To achieve this, we follow the same strategy as before and construct them as nodes that are independent of all other nodes. First, we consider a few examples to gain intuition on how this works. Then, we formulate a statement that extends Corollary 4.10 to nodes before i and after j that do not occur in the statement.

Example 4.11. Suppose we want to contradict the statement $1 \perp\!\!\!\perp 2 \mid 3$ on the directed path of length three. If this statement holds and if additionally $2 \perp\!\!\!\perp 3$ holds, the contraction axiom, one of the rules of conditional independence (see for example Pearl (2009)), implies $1 \perp\!\!\!\perp 2$. The contrapositive of this rule says that if we can find a distribution where $2 \perp\!\!\!\perp 3$ holds while $1 \not\perp\!\!\!\perp 2$, then also $1 \not\perp\!\!\!\perp 2 \mid 3$. The easiest way to achieve this is again by setting the edge weight m_{32} between the nodes 2 and 3 to zero while keeping the edge weight m_{21} between node 1 and 2 non-zero. A possible drift matrix is then given by

$$M = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

which yields the covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{3}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We see that $2 \perp\!\!\!\perp 3$ holds in this example, since $\Sigma_{32} = 0$, while $1 \not\!\perp\!\!\!\perp 2$, since $\Sigma_{21} = \frac{1}{2} \neq 0$. Due to the contraction axiom, we can conclude $1 \not\!\perp\!\!\!\perp 2 \mid 3$. The same argument works with $1 \perp\!\!\!\perp 3$ instead of $2 \perp\!\!\!\perp 3$, since $\Sigma_{31} = 0$ is also implied by setting $m_{32} = 0$. Checking the determinant of interest gives the same result. By computing

$$\det \Sigma_{13,23} = \det \begin{pmatrix} \Sigma_{12} & \Sigma_{13} \\ \Sigma_{32} & \Sigma_{33} \end{pmatrix} = \Sigma_{12}\Sigma_{33} - \Sigma_{13}\Sigma_{32} = \Sigma_{12} \cdot 1 - 0 = \Sigma_{12}$$

and specifically

$$\det \Sigma_{13,23} = \det \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{2} \neq 0,$$

we see that the determinant is non-zero if $1 \not\!\perp\!\!\!\perp 2$ and at least one of $1 \perp\!\!\!\perp 3$ and $2 \perp\!\!\!\perp 3$ hold, assuming that $\Sigma_{33} \neq 0$.

In the previous example, only one conditioning variable s occurred after j, so we reduced the search for a counterexample to the statement $i \perp\!\!\!\perp j \mid s$ to finding a counterexample for $i \perp\!\!\!\perp j$ on a shorter path and setting the remaining edge weight to zero. We discover in the following two examples that the same strategy can be applied if conditioning variables occur both before i and after j as well as in between i and j on the directed path. In this case, we set all edge weights occurring before i and after j to zero.

Example 4.12. Let $p = 4$ and consider the statement $2 \perp\!\!\!\perp 3 \mid 1, 4$ in the path model. If we want to set both edge weights between the nodes 1 and 2 as well as 3 and 4 to zero while keeping the edge weight m_{32} non-zero, one possible drift matrix is given by

$$M = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Solving the Lyapunov equation yields the covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{3}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where we can already see that $1 \perp\!\!\!\perp 2$, $1 \perp\!\!\!\perp 3$, and $1 \perp\!\!\!\perp 4$ hold as well as $4 \perp\!\!\!\perp 3$ and $4 \perp\!\!\!\perp 2$, whereas $2 \not\!\perp\!\!\!\perp 3$. Computing the determinant corresponding to the conditional independence statement results in

$$\det \Sigma_{214,314} = (-1)^2 \det \Sigma_{124,134} = \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \frac{1}{2} \neq 0.$$

The same strategy can be used if we additionally have conditioning variables s with $i < s < j$, and we only want to eliminate the conditioning variables that occur before i and after j.

Example 4.13. Let $p = 5$ and consider the conditional independence statement

$$2 \perp\!\!\!\perp 4 \mid 1, 3, 5$$

in the path model. We want to reduce the statement to the case where conditioning nodes before i and after j do not occur. In this example, it means removing the nodes 1 and 5 from the statement and therefore reducing to the statement $1 \perp\!\!\!\perp 3 \mid 2$ on three nodes, where we already know a valid counterexample such that the determinant of interest is non-zero, namely

$$M^* = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

If we then extend this matrix to five nodes by keeping the missing diagonal entries as -1 and setting the missing edge weights to zero, we have the drift matrix

$$M = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & & & & 0 \\ 0 & & M^* & & 0 \\ 0 & & & & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Solving the Lyapunov equation for Σ yields

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{7}{8} & 0 \\ 0 & \frac{1}{4} & \frac{7}{8} & \frac{15}{8} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & & & & 0 \\ 0 & & \Sigma^* & & 0 \\ 0 & & & & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

where Σ^* is the covariance matrix corresponding to M^*. Computing the determinant of interest then boils down to computing

$$\begin{aligned} \det \Sigma_{2135,4135} = (-1)^3 \det \Sigma_{1235,1345} = -\det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & \frac{3}{2} & \frac{7}{8} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = -1 \cdot 1 \cdot \det \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{3}{2} & \frac{7}{8} \end{pmatrix} \cdot 1 \\ = -\det \Sigma^*_{12,23} = \det \Sigma^*_{12,32} \neq 0, \end{aligned}$$

which we know to be non-zero in this example.

Comparing this example to Example 4.5 shows that we performed the exact same calculations – in Example 4.5 for nodes appearing before i and after j that are not conditioning nodes and now for nodes before i and after j that are conditioning nodes in the considered statements. Therefore, we can combine both operations into one step: remove all nodes before $a := i$ and after $b := j$ from the path by setting the corresponding edge weights to zero. We achieve this by applying Lemma 4.8 with the index map $\varphi_{a,b} = \varphi_{i,j}$ defined as in Definition 4.6 and the projection map Π_K with $K := \{i, \dots, j\}$ defined accordingly as in Definition 4.3.

For computing the determinant in the independence criterion, we need all rows and columns of the covariance matrix with indices in $S \cup \{i, j\}$. The corresponding conditioning nodes can occur everywhere on the path, but we only want to keep the ones between i and j in the statement. Therefore, if $S \neq \emptyset$, we partition the conditioning set

$$S = U \,\dot{\cup}\, Z \,\dot{\cup}\, W$$

in three disjoint ordered sets $U = \{u_1, \dots, u_{l_u}\}$, $Z = \{z_1, \dots, z_{l_z}\}$, and $W = \{w_1, \dots, w_{l_w}\}$ such that

$$u_1 < \dots < u_{l_u} < i < z_1 < \dots < z_{l_z} < j < w_1 < \dots < w_{l_w}.$$

We can also write $Z = S \cap K$. If $Z = \emptyset$, the conditioning variables only occur before i or after j on the path.

Corollary 4.14. *Consider the conditional independence statement (14) in the model $\mathcal{M}_{G_p,C_p}$ with $p \geq 2$. Define $p^* := j-i+1$ and $\varphi := \varphi_{i,j}$. Further, let $K := \{i, \dots, j\} \subseteq [p]$ and U, W, and Z as above. If there exists a distribution in the model $\mathcal{M}_{G_{p^*},C_{p^*}}$ such that*

$$1 \perp\!\!\!\perp p^* \mid \varphi(Z)$$

does not hold in this distribution, then there exists a distribution in $\mathcal{M}_{G_p,C_p}$ for which

$$i \perp\!\!\!\perp j \mid S$$

does not hold.

Proof. The proof is similar to Corollary 4.10; we only have to exercise caution with the computation of the determinant. If $S = \emptyset$, we have already proven the statement in Corollary 4.10. Thus, we assume that $S \neq \emptyset$.

Assume that there exists a distribution in the model $\mathcal{M}_{G_{p^*},C_{p^*}}$ such that $1 \perp\!\!\!\perp p^* \mid \varphi(Z)$ does not hold. Note that $1 = \varphi(i)$ and $p^* = \varphi(j)$, so, in other words, there is a matrix $\Sigma^* \in \mathcal{M}_{G_{p^*},C_{p^*}}$ with $\det \Sigma^*_{\varphi(i)\varphi(Z),\varphi(j)\varphi(Z)} \neq 0$. Due to Lemma 4.8, there exists $\Sigma \in \mathcal{M}_{G_p,C_p}$ such that

$$\Pi_K(\Sigma) = \Sigma^*, \tag{17}$$
$$\Pi_{K^C}(\Sigma) = I_{p-p^*}, \tag{18}$$

and all other entries of Σ are zero.

Since $\{i, j\} \cup Z \subseteq \{i, \dots, j\} = K$, equation (17) implies

$$\Sigma_{iZ,Zj} = \Sigma^*_{\varphi(i)\varphi(Z),\varphi(Z)\varphi(j)}. \tag{19}$$

Further, equation (18) implies that

$$\Sigma_{U,U} = I_{l_u} \text{ and } \Sigma_{W,W} = I_{l_w} \tag{20}$$

since $U, W \subseteq K^C$.

We need to compute the determinant of the submatrix of Σ with rows in $\{i\} \cup S$ and columns in $\{j\} \cup S$. To clarify the computation, we reorder the rows and columns and thereby the indices from smallest to largest, and partition Σ in the respective block matrices.

If both $U = \emptyset$ and $W = \emptyset$, we are already done, as this case is covered by Corollary 4.10. Therefore, we first assume $U \neq \emptyset$ and $W \neq \emptyset$. Due to (19) and (20), the resulting determinant is

$$\begin{aligned}
\det \Sigma_{iS,jS} &= \det \Sigma_{iUZW,jUZW} \\
&= (-1)^{2l_u+l_z} \det \Sigma_{UiZW,UZjW} \\
&= (-1)^{l_z} \det \left(\begin{array}{c|cc|c} \Sigma_{U,U} & \Sigma_{U,Z} & \Sigma_{U,j} & \Sigma_{U,W} \\ \hline \Sigma_{i,U} & \Sigma_{i,Z} & \Sigma_{i,j} & \Sigma_{i,W} \\ \Sigma_{Z,U} & \Sigma_{Z,Z} & \Sigma_{Z,j} & \Sigma_{Z,W} \\ \hline \Sigma_{W,U} & \Sigma_{W,Z} & \Sigma_{W,j} & \Sigma_{W,W} \end{array}\right) \\
&= (-1)^{l_z} \det \left(\begin{array}{c|cc|c} I_{l_u} & 0 & 0 & 0 \\ \hline 0 & \multicolumn{2}{c|}{\multirow{2}{*}{$\Sigma_{iZ,Zj}$}} & 0 \\ 0 & & & 0 \\ \hline 0 & 0 & 0 & I_{l_w} \end{array}\right) \\
&= (-1)^{l_z} \det \left(\begin{array}{c|cc|c} I_{l_u} & 0 & 0 & 0 \\ \hline 0 & \multicolumn{2}{c|}{\multirow{2}{*}{$\Sigma^*_{\varphi(i)\varphi(Z),\varphi(Z)\varphi(j)}$}} & 0 \\ 0 & & & 0 \\ \hline 0 & 0 & 0 & I_{l_w} \end{array}\right) \\
&= \det(I_{l_u}) \det(I_{l_w}) (-1)^{l_z} \det \Sigma^*_{\varphi(i)\varphi(Z),\varphi(Z)\varphi(j)} \\
&= \det \Sigma^*_{\varphi(i)\varphi(Z),\varphi(j)\varphi(Z)} \neq 0.
\end{aligned} \tag{21}$$

In the last equality, we switched back the order of the columns in the considered submatrix of Σ^*.

If either $U = \emptyset$ or $W = \emptyset$, the computation of the determinant can be performed similarly by leaving out all blocks with indices in either U or W, respectively, in the calculation (21) above. Similarly, if $Z = \emptyset$, we leave out all blocks with indices in Z.

Using the criterion in Lemma 3.10, we infer that $i \perp\!\!\!\perp j \mid S$ does not hold in the distribution defined by Σ. Thus, we have found a distribution in the model $\mathcal{M}_{G_p,C_p}$ for which $i \perp\!\!\!\perp j \mid S$ does not hold. $\square$

4.3 Constructing perfectly correlated nodes

Due to Corollary 4.14, all that remains is to consider statements of the form

$$1 \perp\!\!\!\perp p \mid S, \text{ where } S \subseteq \{2, \ldots, p-1\}, \tag{22}$$

in the model $\mathcal{M}_{G_p,C_p}$. Note that the set S of conditioning nodes occurring between the nodes 1 and p is allowed to be empty. Our next goal is to construct the nodes between 1 and p that do not occur in S in such a way that we can ignore them when considering the conditional independence statement (22). In other words, we want to be able to "cross out" any node between the first and last node that does not occur in the conditioning set S and then use an existing counterexample on $p-1$ nodes. If we can do this repeatedly for all such nodes not occurring in S, it suffices in the end to find counterexamples to the statement

$$1 \perp\!\!\!\perp p \mid 2, \dots, p-1 \tag{23}$$

in the model $\mathcal{M}_{G_p,C_p}$. Thus, we want to find a way to extend counterexamples for (23) to counterexamples for (22).

We start again by looking at examples.

Example 4.15. While a first intuition might be to follow the same strategy of adding independent nodes between i and j as before, it becomes clear that we need a different approach. Let $p = 4$ and consider the conditional independence statement $1 \perp\!\!\!\perp 4 \mid 2$ in the path model. We already know a counterexample to $1 \perp\!\!\!\perp 3 \mid 2$ on three nodes with drift matrix

$$M^* = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Following our previous strategy would mean extending M^* to

$$M = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

The resulting covariance matrix is then

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{3}{2} & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{3}{2} \end{pmatrix}.$$

Computing the determinant of interest yields

$$\det \Sigma_{12,42} = -\det \Sigma_{12,24} = -\det \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{3}{2} & 0 \end{pmatrix} = 0,$$

so $1 \perp\!\!\!\perp 4 \mid 2$ holds in this distribution. Taking a closer look at the covariance matrix Σ tells us why: by inserting a node that is independent of all existing nodes on the path, we get in particular $\Sigma_{14} = \Sigma_{24} = 0$. Thus, the nodes 1 and 2 are both independent of node 4. The weak union axiom (see again Pearl (2009)) then implies $1 \perp\!\!\!\perp 4 \mid 2$.

The choice of entries of M in the example does not yield a counterexample to the statement, so we have to find a different way of extending the drift matrix M^*. We start with the structure

$$M = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & ? & ? & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix},$$

where nodes 1 and 2 of the smaller example correspond to nodes 1 and 2 of the larger example. The last node, i.e., node 3 of the smaller example, corresponds to the last node, i.e., node 4 of the extended example. The edge weight and the diagonal entry of M marked with "?" belong to the node to be constructed at index 3. This node now gives rise to an additional row and column at index 3 in the 4×4 covariance matrix Σ. Since we only need the rows and columns of Σ for the computation of the determinant that correspond to the nodes in $\{i, j\} \cup S = \{1, 2, 4\}$, we would ideally like to keep these rows and columns as the respective rows and columns of Σ^* where we know said determinant to be non-zero.

How can we choose the entries in M to achieve this? Intuitively, we want the added node to behave exactly like one of its neighbors – here, for instance, node 2 – while preserving all existing dependence relations among the nodes 1, 2, and 4. Then, nodes 2 and 3 in the extended example can essentially be viewed as one node, and we are back at the smaller example. With respect to the covariance matrix Σ, this means that we want the third row and column to be duplicates of the second row and column, i.e., we want Σ to take on the form

$$\Sigma = \begin{pmatrix} \Sigma^*_{11} & \Sigma^*_{12} & \Sigma^*_{12} & \Sigma^*_{13} \\ \Sigma^*_{21} & \Sigma^*_{22} & \Sigma^*_{22} & \Sigma^*_{23} \\ \Sigma^*_{21} & \Sigma^*_{22} & \Sigma^*_{22} & \Sigma^*_{23} \\ \Sigma^*_{31} & \Sigma^*_{32} & \Sigma^*_{32} & \Sigma^*_{33} \end{pmatrix}. \tag{24}$$

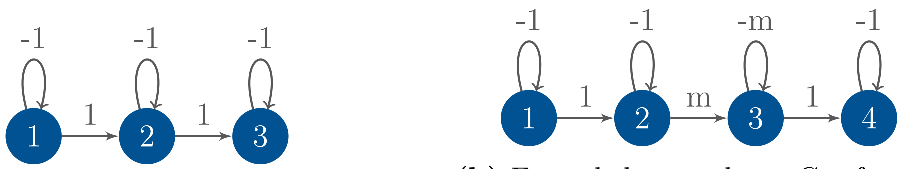
Then, we have, for example, correlation $\rho(X_2, X_3) = \frac{\Sigma^*_{22}}{\sqrt{\Sigma^*_{22}}\sqrt{\Sigma^*_{22}}} = 1$. Therefore, we aim to find an edge weight m_{32} and a diagonal entry m_{33} for M such that this Σ is the solution of the corresponding Lyapunov equation.

We start by setting $m_{32} := m$ for some $m \in \mathbb{N}_{>0}$. Remember that M is the drift matrix of an Ornstein-Uhlenbeck process in equilibrium. As presented in Section 2.3.5, the drift term of a diffusion process stems from the infinitesimal transition from one state to the next. It describes the interaction of the coordinates of the process at the transition. Therefore, the larger the entry m_{32} is in comparison to the corresponding noise term, the stronger the influence of node 2 on node 3. Thus, we are interested in drift matrices M where m is large.

It is clear that the diagonal entry m_{33} has to be negative for M to be stable. If we again make the obvious choice of $m_{33} = -1$ as diagonal entry, the matrix is stable. However, in terms of self-regulation, node 3 does not behave exactly like node 2. The proportion of the change from node 2 to 3 and the decay at node 3 is not balanced as in the case of node 2. There, the diagonal entry is the negative of the edge weight from node 1 to 2. To replicate this behavior at node 3, we set the diagonal entry $m_{33} := -m$. Thus, we define drift matrices of the form

$$M^{(m)} := \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & m & -m & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

with $m \in \mathbb{N}_{>0}$. The corresponding model is illustrated in Figure 6.



(a) The directed graph G_3 with example edge weights.

(b) Extended example on G_4 after inserting one node that is highly correlated with node 2 for large m.

Figure 6: Embedding of an example on three nodes into an example on four nodes.

Solving the resulting Lyapunov equation yields the symmetric covariance matrix

$$\Sigma^{(m)} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2}\frac{m}{1+m} & \frac{1}{4}\frac{m}{1+m} \\ * & \frac{3}{2} & \frac{m}{2}\frac{4+3m}{(1+m)^2} & \frac{m}{8}\frac{9+7m}{(1+m)^2} \\ * & * & \frac{2+4m+6m^2+3m^3}{2m(1+m)^2} & \frac{8+16m+24m^2+21m^3+7m^4}{8m(1+m)^3} \\ * & * & * & \frac{8+24m+48m^2+45m^3+15m^4}{8m(1+m)^3} \end{pmatrix}.$$

If we now let m go to infinity, we find

$$\lim_{m\to\infty} \Sigma^{(m)} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{2} & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{2} & \frac{3}{2} & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{4} & \frac{7}{8} & \frac{7}{8} & \frac{15}{8} \end{pmatrix} =: \Sigma,$$

where the third row and column are a copy of the second row and column, respectively. All other entries are the same as in Σ^*, with indices shifted accordingly. Therefore, we can construct a sequence of distributions such that the sequence of covariance matrices $\Sigma^{(m)}$ solving the Lyapunov equation converges to the matrix Σ for $m \to \infty$.

The idea of embedding a smaller example into a larger example such that the additional nodes are perfectly correlated with an existing node can be formulated via a projection that removes one node at an index k as in Definition 4.3. A corresponding index map that maps the original indices of the directed path of length p to the corresponding indices of the directed path where the k-th node has been removed can be defined as follows. It shifts all indices larger than k down by one.

Definition 4.16. *For $k \in \mathbb{N}_{>0}$, we define the index map*

$$\psi_{-k} : \mathbb{N} \setminus \{k\} \to \mathbb{N}, \quad x \mapsto \begin{cases} x & \text{if } x < k; \\ x-1 & \text{if } x > k. \end{cases}$$

Now we extend the argument to arbitrary dimensions. Given a matrix Σ that solves the Lyapunov equation, we want to construct a node at index k that is perfectly correlated with the node at index $k-1$. That means shifting all rows and columns of Σ with index starting at k one index higher and inserting a new k-th row and column at index k as the exact duplicate of the $(k-1)$-th row and column, respectively. We can approximately achieve this by setting the $(k-1)$-th edge weight $m_{k,k-1} := m$ for a large value $m \in \mathbb{N}_{>0}$, while defining the diagonal entry as $m_{kk} := -m$ and shifting the entries with higher indices up by one index. Again, we formalize this approach in a lemma.

Lemma 4.17. *Let $p \geq 3$, $k \in \{2, \dots, p-1\}$, and let $\Sigma^* \in \mathcal{M}_{G_{p-1},C_{p-1}}$. Then, there is a matrix $\Sigma \in \overline{\mathcal{M}_{G_p,C_p}}$ such that*

$$\Pi_{(-k)}(\Sigma) = \Sigma^* \tag{25}$$

and

$$\Sigma_{k,j} = \Sigma_{k-1,j} \text{ as well as } \Sigma_{j,k} = \Sigma_{j,k-1} \text{ for all } j \in [p]. \tag{26}$$

Proof. The proof is similar to the proof of Lemma 4.8 in the sense that we define an appropriate matrix M and then solve the Lyapunov equation for Σ. Assume we have $\Sigma^* \in \mathcal{M}_{G_{p-1},C_{p-1}}$. Then, there is a stable matrix $M^* := (m^*_{ij}) \in \mathbb{R}^{(p-1)\times(p-1)}$ such that the Lyapunov equation

$$M^*\Sigma^* + \Sigma^* {M^*}^T + C_{p-1} = 0 \tag{27}$$

is fulfilled.

We want to extend this model by inserting a node at index k. Let $U := \{1, \dots, k-1\}$ and $V := \{k+1, \dots, p\}$ and define $\psi := \psi_{-k}$ as in Definition 4.16. Then, we have $\psi(U) = U$ and $\psi(V) = \{k, \dots, p-1\}$. We can partition

$$\Sigma^* = \left(\begin{array}{c|c} \Sigma^*_{U,U} & \Sigma^*_{U,\psi(V)} \\ \hline \Sigma^*_{\psi(V),U} & \Sigma_{\psi(V),\psi(V)} \end{array}\right)$$

as well as

$$M^* = \left(\begin{array}{c|c} M^*_{U,U} & 0 \\ \hline M^*_{\psi(V),U} & M^*_{\psi(V),\psi(V)} \end{array}\right), \quad \text{where} \quad M^*_{\psi(V),U} = \begin{pmatrix} 0 & \cdots & 0 & m^*_{k,k-1} \\ & & & 0 \\ & 0 & & \vdots \\ & & & 0 \end{pmatrix}.$$

For better readability, we write $\Sigma^*_U := \Sigma^*_{U,U}$ and $\Sigma^*_{\psi(V)} := \Sigma^*_{\psi(V),\psi(V)}$; similarly for M^*. Considering (27) block-wise results in the following four equations

(a) $M^*_U \Sigma^*_U + \Sigma^*_U {M^*_U}^T + 2I_{k-1} = 0$,

(b) $M^*_U \Sigma^*_{U,\psi(V)} + \Sigma^*_{U,\psi(V)} {M^*_{\psi(V)}}^T + \Sigma^*_U {M^*_{\psi(V),U}}^T = 0$,

(c) $M^*_{\psi(V)} \Sigma^*_{\psi(V)} + \Sigma^*_{\psi(V)} {M^*_{\psi(V)}}^T + 2I_{p-k} + M^*_{\psi(V),U} \Sigma^*_{U,\psi(V)} + \Sigma^*_{\psi(V),U} {M^*_{\psi(V),U}}^T = 0$,

(d) $M^*_{\psi(V)} \Sigma^*_{\psi(V),U} + \Sigma^*_{\psi(V),U} {M^*_U}^T + M^*_{\psi(V),U} \Sigma^*_U = 0$

$\iff \; M^*_U {\Sigma^*_{\psi(V),U}}^T + {\Sigma^*_{\psi(V),U}}^T {M^*_{\psi(V)}}^T + \Sigma^*_U {M^*_{\psi(V),U}}^T = 0.$

Given the stable matrix M^*, these equations are uniquely fulfilled by the corresponding submatrices of Σ^*.

Now let $m \in \mathbb{N}_{>0}$ and define

$$M^{(m)} := (m_{ij}^{(m)}) = \left(\begin{array}{c|c|c} M_U^* & 0 & 0 \\ \hline 0 \quad m & -m & 0 \\ \hline 0 & \begin{array}{c} m_{k,k-1}^* \\ 0 \end{array} & M_{\psi(V)}^* \end{array}\right) \in \mathbb{R}^{p \times p},$$

i.e., we have

$$m_{ij}^{(m)} = \begin{cases} m_{ij}^*, & \text{if } i,j \leq k-1 \text{ or } i,j \geq k+1; \\ m, & \text{if } i = k,\ j = k-1; \\ -m, & \text{if } i = j = k; \\ m_{k,k-1}^*, & \text{if } i = k+1,\ j = k; \\ 0 & \text{else.} \end{cases}$$

Let

$$\Sigma^{(m)} = \left(\begin{array}{c|c|c} \Sigma_U^{(m)} & \Sigma_{U,k}^{(m)} & \Sigma_{U,V}^{(m)} \\ \hline \Sigma_{k,U}^{(m)} & \Sigma_{k,k}^{(m)} & \Sigma_{k,V}^{(m)} \\ \hline \Sigma_{V,U}^{(m)} & \Sigma_{V,k}^{(m)} & \Sigma_V^{(m)} \end{array}\right) \in \mathbb{R}^{p \times p}$$

be the solution of the Lyapunov equation

$$M^{(m)}\Sigma^{(m)} + \Sigma^{(m)}{M^{(m)}}^T + C_p = 0$$

induced by $M^{(m)}$. We use the same abbreviations of the form $\Sigma_U^{(m)} := \Sigma_{U,U}^{(m)}$ as above. Due to the symmetry of the Lyapunov equation, any unique solution is symmetric. Since $M^{(m)}$ is stable, the solution $\Sigma^{(m)}$ is unique and therefore symmetric, so we only have to solve the following six matrix equations:

(A) $M_U^*\Sigma_U^{(m)} + \Sigma_U^{(m)}{M_U^*}^T + 2I_{k-1} = 0$

This equation is solved by $\Sigma_U^{(m)} = \Sigma_U^*$ due to (a).

(B) $M_U^*\Sigma_{U,k}^{(m)} + m\Sigma_{U,k-1}^{(m)} - m\Sigma_{U,k}^{(m)} = 0$

This equation can be rewritten to

$$\left(\frac{1}{m}M_U^* - I_{k-1}\right)\Sigma_{U,k}^{(m)} = -\Sigma_{U,k-1}^{(m)}.$$

The matrix $\frac{1}{m}M_U^* - I_{k-1}$ is stable and thus invertible, so solving for $\Sigma_{U,k}^{(m)}$ yields

$$\begin{aligned}\Sigma_{U,k}^{(m)} &= -\left(\frac{1}{m}M_U^* - I_{k-1}\right)^{-1}\Sigma_{U,k-1}^{(m)}\\ &\overset{(A)}{=} -\left(\frac{1}{m}M_U^* - I_{k-1}\right)^{-1}\Sigma_{U,k-1}^{*} \xrightarrow[m\to\infty]{} -\left(0 - I_{k-1}\right)^{-1}\Sigma_{U,k-1}^{*} = \Sigma_{U,k-1}^{*},\end{aligned}$$

where we used that $k-1 \in U$ and that matrix inversion is continuous. We deduce that $\lim_{m\to\infty}\Sigma_{U,k}^{(m)} = \Sigma_{U,k-1}^{*}$.

(C) $M_U^*\Sigma_{U,V}^{(m)} + \Sigma_{U,V}^{(m)}{M_{\psi(V)}^*}^T + \Sigma_{U,k}^{(m)}(m_{k,k-1}^*, 0, \ldots, 0) = 0$

Taking the limit of the left-hand side for $m \to \infty$ yields

$$\begin{aligned}&M_U^*\Sigma_{U,V}^{(m)} + \Sigma_{U,V}^{(m)}{M_{\psi(V)}^*}^T + \Sigma_{U,k}^{(m)}(m_{k,k-1}^*, 0, \ldots, 0)\\ \xrightarrow[m\to\infty]{(B)}\ & M_U^*\left(\lim_{m\to\infty}\Sigma_{U,V}^{(m)}\right) + \left(\lim_{m\to\infty}\Sigma_{U,V}^{(m)}\right){M_{\psi(V)}^*}^T + \Sigma_{U,k-1}^{*}(m_{k,k-1}^*, 0, \ldots, 0),\end{aligned}$$

where we employed (B) in the last summand. This last summand can be rewritten further as

$$\Sigma_{U,k-1}^{*}(m_{k,k-1}^*, 0, \ldots, 0) = \Sigma_U^* {M_{\psi(V),U}^*}^T$$

by attaching a $(k-2)\times(k-1)$ zero matrix to the existing row vector and thereby extending it to ${M_{\psi(V),U}^*}^T$. Setting the limit to zero yields equation (b) which is solved by $\Sigma_{U,\psi(V)}^*$. This gives us $\lim_{m\to\infty}\Sigma_{U,V}^{(m)} = \Sigma_{U,\psi(V)}^*$.

(D) $2 - 2m\Sigma_{k,k}^{(m)} + m\left(\Sigma_{k,k-1}^{(m)} + \Sigma_{k-1,k}^{(m)}\right) = 0$

Solving for $\Sigma_{k,k}^{(m)}$ yields

$$\Sigma_{k,k}^{(m)} = \frac{1}{m} + \frac{\Sigma_{k,k-1}^{(m)} + \Sigma_{k-1,k}^{(m)}}{2} \xrightarrow[m\to\infty]{(B)} \frac{\Sigma_{k-1,k-1}^{*} + \Sigma_{k-1,k-1}^{*}}{2} = \Sigma_{k-1,k-1}^{*},$$

where we used $\Sigma_{k,k-1}^{(m)} \to \Sigma_{k-1,k-1}^{*}$ for $m \to \infty$ due to (B). We conclude that $\lim_{m\to\infty}\Sigma_{k,k}^{(m)} = \Sigma_{k-1,k-1}^{*}$.

(E) $m\Sigma_{k-1,V}^{(m)} - m\Sigma_{k,V}^{(m)} + \Sigma_{k,k}^{(m)}(m_{k,k-1}^*, 0, \ldots, 0) + \Sigma_{k,V}^{(m)}{M_{\psi(V)}^*}^T = 0$

Rewriting the equation yields

$$\begin{aligned}&\Sigma_{k,V}^{(m)}\left(-\frac{1}{m}{M_{\psi(V)}^*}^T + I_{p-k}\right) = \Sigma_{k-1,V}^{(m)} + \frac{1}{m}\Sigma_{k,k}^{(m)}(m_{k,k-1}^*, 0, \ldots, 0)\\ \iff\ & \Sigma_{k,V}^{(m)} = \left(\Sigma_{k-1,V}^{(m)} + \frac{1}{m}\Sigma_{k,k}^{(m)}(m_{k,k-1}^*, 0, \ldots, 0)\right)\left(-\frac{1}{m}{M_{\psi(V)}^*}^T + I_{p-k}\right)^{-1}\\ &\xrightarrow[m\to\infty]{(C),(D)} (\Sigma_{k-1,\psi(V)}^* + 0)(0 + I_{p-k})^{-1} = \Sigma_{k-1,\psi(V)}^*.\end{aligned}$$

Here, we used (C) together with $k-1 \in U$ in the limit. Due to (D), $\Sigma_{k,k}^{(m)}$ converges to a constant and therefore $\frac{1}{m}\Sigma_{k,k}^{(m)} \to 0$. Additionally, the inverted matrix is stable and therefore invertible, and matrix inversion is continuous. We deduce that $\lim_{m\to\infty}\Sigma_{k,V}^{(m)} = \Sigma_{k-1,\psi(V)}^*$.

(F) $M^*_{\psi(V)}\Sigma^{(m)}_V + \Sigma^{(m)}_V {M^*_{\psi(V)}}^T + 2I_{p-k} + (m^*_{k,k-1}, 0, \dots, 0)^T \Sigma^{(m)}_{k,V} + \Sigma^{(m)}_{V,k}(m^*_{k,k-1}, 0, \dots, 0) = 0$

We use that $\Sigma^{(m)}_{k,V} \to \Sigma^*_{k-1,\psi(V)}$ as well as $\Sigma^{(m)}_{V,k} \to \Sigma^*_{\psi(V),k-1}$ for $m \to \infty$ due to (E) and take the limit on both sides. This yields for the left-hand side

$$\begin{aligned} & M^*_{\psi(V)}\Sigma^{(m)}_V + \Sigma^{(m)}_V {M^*_{\psi(V)}}^T + 2I_{p-k} \\ & \qquad + (m^*_{k,k-1}, 0, \dots, 0)^T \Sigma^{(m)}_{k,V} + \Sigma^{(m)}_{V,k}(m^*_{k,k-1}, 0, \dots, 0) \\ \xrightarrow[m\to\infty]{(E)} \; & M^*_{\psi(V)} \left(\lim_{m\to\infty} \Sigma^{(m)}_V\right) + \left(\lim_{m\to\infty} \Sigma^{(m)}_V\right) {M^*_{\psi(V)}}^T + 2I_{p-k} \\ & \qquad + (m^*_{k,k-1}, 0, \dots, 0)^T \Sigma^*_{k-1,\psi(V)} + \Sigma^*_{\psi(V),k-1}(m^*_{k,k-1}, 0, \dots, 0) \\ = \; & M^*_{\psi(V)} \left(\lim_{m\to\infty} \Sigma^{(m)}_V\right) + \left(\lim_{m\to\infty} \Sigma^{(m)}_V\right) {M^*_{\psi(V)}}^T + 2I_{p-k} \\ & \qquad + M^*_{\psi(V),U}\Sigma^*_{U,\psi(V)} + \Sigma^*_{\psi(V),U}{M^*_{\psi(V),U}}^T, \end{aligned}$$

where we employed the same trick of rewriting the last two summands as for (C). Setting the limit to zero yields equation (c) that is solved by $\Sigma^*_{\psi(V)}$. We conclude $\lim_{m\to\infty} \Sigma^{(m)}_V = \Sigma^*_{\psi(V)}$.

Combining these results, we have

$$\lim_{m\to\infty} \Sigma^{(m)} = \left(\begin{array}{c|c|c} \Sigma^*_U & \Sigma^*_{U,k-1} & \Sigma^*_{U,\psi(V)} \\ \hline \Sigma^*_{k-1,U} & \Sigma^*_{k-1,k-1} & \Sigma^*_{k-1,\psi(V)} \\ \hline \Sigma^*_{\psi(V),U} & \Sigma^*_{\psi(V),k-1} & \Sigma^*_{\psi(V)} \end{array}\right) =: \Sigma \in \mathbb{R}^{p\times p}$$

with

$$\Sigma_{ij} = \begin{cases} \Sigma^*_{\psi(i),\psi(j)}, & \text{if } i, j \neq k; \\ \Sigma^*_{k-1,\psi(j)}, & \text{if } i = k,\ j \neq k; \\ \Sigma^*_{\psi(i),k-1}, & \text{if } i \neq k,\ j = k; \\ \Sigma^*_{k-1,k-1}, & \text{if } i = j = k. \end{cases}$$

Note that Σ is now a copy of Σ^* with one row and one column added at index k where we duplicated and inserted the $k-1$-th row and column of Σ^*, which are the $k-1$-th row and column of Σ. Thus, $\Sigma_{k,j} = \Sigma_{k-1,j}$ as well as $\Sigma_{j,k} = \Sigma_{j,k-1}$ for all $j \in [p]$, so (25) and (26) hold. □

The lemma implies that any covariance matrix of a distribution in the model on $p-1$ nodes can be embedded into a covariance matrix of a distribution in the closure of the model on p nodes by inserting a copy of the $(k-1)$-th row and column, respectively, at index k.

Remark 4.18. The first conclusion (25) in Lemma 4.17 can be reformulated as

$$\mathcal{M}_{G_{p-1},C_{p-1}} \subseteq \Pi_{(-k)}(\overline{\mathcal{M}_{G_p,C_p}}),$$

since for $\Sigma^* \in \mathcal{M}_{G_{p-1},C_{p-1}}$, this is equivalent to the existence of $\Sigma \in \overline{\mathcal{M}_{G_p,C_p}}$ with

$$\Sigma^* = \Pi_{(-k)}(\Sigma).$$

The limit operation is performed element-wise on the matrices, and the projection does not change the values themselves. Thus, by interchanging the limit operation and the projection we obtain

$$\mathcal{M}_{G_{p-1},C_{p-1}} \subseteq \overline{\Pi_{(-k)}(\mathcal{M}_{G_p,C_p})}.$$

Now we have everything in place to return to the statement

$$1 \perp\!\!\!\perp p \mid S \ \text{ with } \ S \subsetneq \{2,\dots,p-1\}.$$

Assume that $k \in \{2,\dots,p-1\} \setminus S$. Our goal is to extend a counterexample $\Sigma^* \in \mathcal{M}_{G_{p-1}}$ for $1 \perp\!\!\!\perp p-1 \mid \psi(S)$ to a counterexample $\Sigma \in \mathcal{M}_{G_p}$ for $1 \perp\!\!\!\perp p \mid S$. So we start with a matrix Σ^* where $\det \Sigma^*_{1\psi(S),(p-1)\psi(S)} \neq 0$. Then, we create a new row and column at index k in Σ^* via a suitable drift matrix M such that $\det \Sigma_{1S,pS} \neq 0$.

Ideally, as in (24), we would like to shift all rows and columns with index starting at k one index higher and insert a new k-th row and column at index k as the exact duplicate of the $(k-1)$-th row and column, respectively. Then, both determinants would be the same: due to $k \notin S$, the k-th row and column do not occur in the computation of the determinant while all other entries stay as in Σ^*.

We first formulate the technical aspect of the statement with respect to non-zero determinants and then formulate it in terms of conditional independence statements. The difficulty lies in the fact that any matrix we construct via Lemma 4.17 is in the closure of the path model on p nodes but not necessarily in the model. Consequently, we cannot take such a matrix itself as a counterexample for the conditional independence statement $1 \perp\!\!\!\perp p \mid S$ in $\mathcal{M}_{G_p,C_p}$.

Corollary 4.19. *Let $p \geq 3$, $S \subsetneq \{2,\dots,p-1\}$, and $k \in \{2,\dots,p-1\} \setminus S$. Define $\psi := \psi_{-k}$ as above. Then, the following holds: If there is*

$$\Sigma^* \in \mathcal{M}_{G_{p-1},C_{p-1}} \ \textit{such that} \ \det \Sigma^*_{1\psi(S),(p-1)\psi(S)} \neq 0,$$

then there is

$$\Sigma \in \mathcal{M}_{G_p,C_p} \ \textit{such that} \ \det \Sigma_{1S,pS} \neq 0.$$

Proof. We split the proof into two parts. First, we apply Lemma 4.17 to construct such a matrix with non-zero determinant of interest on the boundary of the model. In the second step, we exploit the continuity of the determinant to find a covariance matrix with non-zero determinant in the model.

Assume we have $\Sigma^* \in \mathcal{M}_{G_{p-1},C_{p-1}}$ with $\det \Sigma^*_{1\psi(S),(p-1)\psi(S)} \neq 0$. We want to show that there is

(i) $\widehat{\Sigma} \in \overline{\mathcal{M}_{G_p,C_p}}$ such that $\det \widehat{\Sigma}_{1S,pS} \neq 0$ and therefore there is

(ii) $\Sigma \in \mathcal{M}_{G_p,C_p}$ such that $\det \Sigma_{1S,pS} \neq 0$.

We first show (i). Due to Lemma 4.17, there is a sequence

$$\widehat{\Sigma}^{(m)} \in \mathcal{M}_{G_p,C_p} \text{ with } \lim_{m\to\infty} \widehat{\Sigma}^{(m)} =: \widehat{\Sigma} \in \overline{\mathcal{M}_{G_p,C_p}} \text{ such that } \Pi_{(-k)}(\widehat{\Sigma}) = \Sigma^*.$$

Since $k \notin \{1,p\} \cup S$, all entries $\widehat{\Sigma}_{ij}$ occurring in the computation of the determinant $\det \widehat{\Sigma}_{1S,pS}$ fulfill $\widehat{\Sigma}_{ij} = \Sigma^*_{\psi(i),\psi(j)}$. Computing the determinant then yields

$$\det \widehat{\Sigma}_{1S,pS} = \det \Sigma^*_{\psi(1)\psi(S),\psi(p)\psi(S)} = \det \Sigma^*_{1\psi(S),(p-1)\psi(S)} \neq 0.$$

For (ii), we observe that, due to Lemma 4.17, $\widehat{\Sigma}$ has a duplicated row and column and is therefore singular. This implies that it is not positive definite and thus $\widehat{\Sigma} \notin \mathcal{M}_{G_p,C_p}$.

Since $\widehat{\Sigma} \in \overline{\mathcal{M}_{G_p,C_p}} \setminus \mathcal{M}_{G_p,C_p}$ holds, the matrix $\widehat{\Sigma}$ is an element of the boundary of $\mathcal{M}_{G_p,C_p}$ with $\det \widehat{\Sigma}_{1S,jS} \neq 0$. Since the determinant is a continuous function on $\mathbb{R}^{p\times p}$, there is a neighborhood U around $\widehat{\Sigma}$ where this determinant is non-zero as well. Then, for any $\Sigma \in U \cap \mathcal{M}_{G_p,C_p}$, we have $\det \Sigma_{1S,jS} \neq 0$. Therefore, we can choose any matrix $\Sigma \in U \cap \mathcal{M}_{G_p,C_p}$. □

Corollary 4.20. *Consider the conditional independence statement (22) in the model $\mathcal{M}_{G_p,C_p}$ with $p \geq 3$. Let $S \subsetneq \{2,\dots,p-1\}$ and $k \in \{2,\dots,p-1\} \setminus S$. Define $\psi := \psi_{-k}$ as above. If there exists a distribution in the model $\mathcal{M}_{G_{p-1},C_{p-1}}$ such that*

$$1 \perp\!\!\!\perp p-1 \mid \psi(S)$$

does not hold in this distribution, then there exists a distribution in $\mathcal{M}_{G_p,C_p}$ for which

$$1 \perp\!\!\!\perp p \mid S$$

does not hold.

Proof. Assume that there exists a distribution in the model $\mathcal{M}_{G_{p-1},C_{p-1}}$ on $p-1$ nodes such that $1 \perp\!\!\!\perp p-1 \mid \psi(S)$ does not hold in this distribution. Then there is $\Sigma^* \in \mathcal{M}_{G_{p-1},C_{p-1}}$ with $\det \Sigma^*_{1\psi(S),(p-1)\psi(S)} \neq 0$. Due to Corollary 4.19, there is a matrix $\Sigma \in \mathcal{M}_{G_p,C_p}$ such that $\det \Sigma_{1S,pS} \neq 0$. This implies that $1 \perp\!\!\!\perp p \mid S$ does not hold in the distribution defined by Σ. Therefore, we have found a distribution in the model M_{G_p,C_p} for which $1 \perp\!\!\!\perp p \mid S$ does not hold. □

By applying this corollary consecutively, we can extend any counterexample for the conditional independence statement $1 \perp\!\!\!\perp p^* \mid 2,\dots,p^*-1$ on p^* nodes with $p^* < p$ to a counterexample for $1 \perp\!\!\!\perp p \mid S$ with $S \subsetneq \{2,\dots,p-1\}$.

4.4 Proving the theorem

We now have established the necessary tools to prove the main result. The proof combines the results of Corollary 4.14 and Corollary 4.20 with some additional direct calculations. First, we summarize the rather technical results of the chapter in the following proposition. The proposition is applicable for statements with an arbitrary number of variables occurring between the nodes i and j on paths of arbitrary length.

Proposition 4.21. *Consider the Lyapunov model on the directed path $\mathcal{M}_{G_p,C_p}$ with $p \geq 2$. Let $i, j \in V_p$ such that $i < j$ and let $S \subseteq V_p \setminus \{i, j\}$. Define*

$$Z := S \cap \{i, \ldots, j\} \quad \textit{and} \quad q := |Z| + 2$$

as the subset of S containing all conditioning variables that lie between i and j, and the number of these variables including i and j, respectively. If there exists a distribution in the model $\mathcal{M}_{G_q,C_q}$ such that

$$1 \perp\!\!\!\perp q \mid 2, \ldots, q-1 \quad (\textit{or } 1 \perp\!\!\!\perp 2 \textit{ if } q = 2)$$

does not hold in this distribution, then there exists a distribution in $\mathcal{M}_{G_p,C_p}$ for which

$$i \perp\!\!\!\perp j \mid S$$

does not hold.

Proof. If $p = 2$, we are done. Therefore, we assume $p \geq 3$.

We again define the bijective index map $\varphi := \varphi_{i,j}$ as in Definition 4.6. Let $p^* := \varphi(j) = j - i + 1$ and $Z^* := \varphi(Z) \subseteq \{2, \ldots, p^* - 1\}$. Thereby, we consider the sub-path from node i to node j as a shorter path of length p^*.

The following proof consists of two steps. First, we apply Corollary 4.20, which means adding highly correlated nodes to an existing counterexample distribution. Then, we apply Corollary 4.14 which allows us to further add independent nodes at the start and end of the path in the existing counterexample distribution.

Assume that there exists a distribution in the model $\mathcal{M}_{G_q,C_q}$ for which

$$1 \perp\!\!\!\perp q \mid 2, \ldots, q-1 \text{ (or } 1 \perp\!\!\!\perp 2 \text{ if } q = 2) \tag{28}$$

does not hold.

Step 1. We show that there is a distribution in $\mathcal{M}_{G_{p^*},C_{p^*}}$ for which $1 \perp\!\!\!\perp p^* \mid Z^*$ does not hold.

Let $K := \{2, \ldots, p^* - 1\} \setminus Z^*$, so

$$\{2, \ldots, p^* - 1\} = Z^* \mathbin{\dot\cup} K.$$

The nodes in Z^* correspond to the conditioning nodes $2, \ldots, q-1$ that are already in the statement. The elements of K are the nodes between 1 and p^* that are not in the conditioning set. We insert them consecutively at the correct indices by applying Corollary 4.20 $|K|$ times.

If $Z^* = \{2, \ldots, p^* - 1\}$ and therefore $q = p^*$ and $K = \emptyset$, we are already done. Thus, we assume $Z^* \subsetneq \{2, \ldots, p^* - 1\}$ in the following. Then, we have $q < p^*$ and $K \neq \emptyset$.

If $Z^* = \emptyset$, i.e. $q = 2$, we have $K = \{2, \ldots, p^* - 1\}$. In that case, we apply Corollary 4.20 $p^* - 2$ times. We know that there is a distribution in the model $\mathcal{M}_{G_2,C_2}$ for which

$$1 \perp\!\!\!\perp 2 \mid \emptyset$$

does not hold. We can rewrite this statement to

$$1 \perp\!\!\!\perp 2 \mid \psi_{-k}(\emptyset)$$

for any $k \in K$ using Definition 4.16 of the index map. Then, Corollary 4.20 implies that there is a distribution in the model $\mathcal{M}_{G_3,C_3}$ for which

$$1 \perp\!\!\!\perp 3 \mid \emptyset$$

does not hold. We can perform this step for every $k \in K$ until we reach the statement that there is a distribution in the model $\mathcal{M}_{G_{p^*},C_{p^*}}$ for which

$$1 \perp\!\!\!\perp p^* \mid \emptyset$$

does not hold. Note that due to the empty conditioning sets the position of the newly added nodes between the first and last node is irrelevant. Therefore, the argument works regardless of the order of the elements in K.

If $Z^* \neq \emptyset$ and $K \neq \emptyset$, the reasoning is not as straightforward, as we have to shift the conditioning nodes to their correct indices to construct the nodes in the given order. Let $k_1 := \min K$. To show by contradiction that $k_1 \in \{2, \dots, q\}$, assume $k_1 \notin \{2, \dots, q\}$. This implies $k_1 > q$, i.e., the minimal element of K is larger than q, so $\{2, \dots, q\} \subseteq Z^*$. Then, we have $|Z^*| \geq q - 1$ which is a contradiction as $|Z^*| = q - 2$. Thus, we can write

$$\{2, \dots, q-1\} = \psi_{-k_1}(\{2, \dots, q\} \setminus \{k_1\}),$$

as the map ψ_{-k_1} shifts all values larger than k_1 down by one. This allows us to rewrite (28) in the case $q > 2$ as

$$1 \perp\!\!\!\perp q \mid \psi_{-k_1}(\{2, \dots, q\} \setminus \{k_1\}).$$

With Corollary 4.20, we conclude that there exists a distribution in the model $\mathcal{M}_{G_{q+1},C_{q+1}}$ such that

$$1 \perp\!\!\!\perp q+1 \mid \{2, \dots, q\} \setminus \{k_1\} \tag{29}$$

does not hold in this distribution.

Now let $k_2 := \min(K \setminus \{k_1\})$. With the same argument as above, we show by contradiction that $k_2 \in \{2, \dots, q+1\} \setminus \{k_1\}$. Assume that $k_2 \notin \{2, \dots, q+1\} \setminus \{k_1\}$. This implies $k_2 > q+1$, so $\{2, \dots, q+1\} \setminus \{k_1\} \subseteq Z^*$. Then, we again have $|Z^*| \geq q-1$, a contradiction. Thus, we can write as above

$$\{2, \dots, q\} \setminus \{k_1\} = \psi_{-k_2}(\{2, \dots, q+1\} \setminus \{k_1, k_2\}),$$

so (29) can in turn be rewritten as

$$1 \perp\!\!\!\perp q+1 \mid \psi_{-k_2}(\{2, \dots, q+1\} \setminus \{k_1, k_2\}).$$

Again, Corollary 4.20 implies that there exists a distribution in the model $\mathcal{M}_{G_{q+2},C_{q+2}}$ for which

$$1 \perp\!\!\!\perp q+2 \mid \{2, \dots, q+1\} \setminus \{k_1, k_2\}$$

does not hold. We can continue this chain of reasoning until we conclude that there exists a distribution in the model $\mathcal{M}_{G_{q+|K|},C_{q+|K|}}$ for which

$$1 \perp\!\!\!\perp q + |K| \mid \{2, \ldots, q + |K| - 1\} \setminus K \tag{30}$$

does not hold. We have $q + |K| = 2 + |Z| + |K| = 2 + p^* - 2 = p^*$ and

$$\{2, \ldots, q + |K| - 1\} \setminus K = Z^*,$$

so (30) not holding for a distribution in the model $\mathcal{M}_{G_{p^*},C_{p^*}}$ is equivalent to

$$1 \perp\!\!\!\perp p^* \mid Z^*$$

not holding for a distribution in the model $\mathcal{M}_{G_{p^*},C_{p^*}}$, which is what we wanted to show.

Step 2. We show that there exists a distribution in the model $\mathcal{M}_{G_p,C_p}$ for which $i \perp\!\!\!\perp j \mid S$ does not hold.

This is exactly the claim of Corollary 4.14, which we have already proven. □

Now we use Proposition 4.21 to prove Theorem 4.1.

Proof of Theorem 4.1. First, we show that there is a distribution in the model $\mathcal{M}_{G_2,C_2}$ for which $1 \perp\!\!\!\perp 2$ does not hold and that for $q = 3, \ldots, 102$, there is a distribution in the model $\mathcal{M}_{G_q,C_q}$ for which $1 \perp\!\!\!\perp q \mid 2, \ldots, q-1$ does not hold. Let $2 \leq q \leq 102$ and define

$$M := \begin{pmatrix} -1 & & & \\ 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{q \times q}$$

to be the $q \times q$ matrix with -1 on the diagonal, 1 on the first subdiagonal, and 0 in all other entries. We solve the Lyapunov equation

$$M\Sigma + \Sigma M^T + C_q = 0$$

with the computer algebra system Mathematica (Wolfram Research, Inc.) employing the efficient way of computation explained in section 3.3. Computing the determinant of interest yields

$$\det \Sigma_{1\ldots q-1, 2\ldots q} \neq 0$$

for $2 \leq q \leq 102$. Thus, there exists a distribution in the model $\mathcal{M}_{G_q,C_q}$ such that the conditional independence statement $1 \perp\!\!\!\perp q \mid 2, \ldots, q-1$ (or $1 \perp\!\!\!\perp 2$ for $q = 2$) does not hold in this distribution. The code for the calculations can be found in the additional material.

Now let $q := |Z| + 2$ and assume that $|Z| \leq 100$. Then, we have $2 \leq q \leq 102$, so there exists a distribution in the model $\mathcal{M}_{G_q,C_q}$ for which

$$1 \perp\!\!\!\perp q \mid 2, \ldots, q-1 \text{ (or } 1 \perp\!\!\!\perp 2 \text{ if } q = 2)$$

does not hold. Applying Proposition 4.21 then implies that there is a distribution in the model $\mathcal{M}_{G_p,C_p}$ such that
$$i \perp\!\!\!\perp j \mid S$$
does not hold in this distribution. □

We close this chapter by returning to Example 4.2.

Example 4.22 (continuation of Example 4.2)**.** We want to show that the conditional independence statement
$$2 \perp\!\!\!\perp 6 \mid 1,4 \tag{31}$$
does not hold in the Lyapunov model $\mathcal{M}_{G_7,C_7}$ of the directed path of length $p = 7$. Using the terminology from Proposition 4.21 and Theorem 4.1, we have $S = \{1, 4\}$ and $Z = \{4\}$. Thus, we define $q := |Z| + 2 = 3$. Proposition 4.21 states that (31) does not hold (for all distributions) in $\mathcal{M}_{G_7,C_7}$ if
$$1 \perp\!\!\!\perp 3 \mid 2 \tag{32}$$
does not hold (for all distributions) in the model $\mathcal{M}_{G_3,C_3}$. We apply the two steps of the proof of Proposition 4.21 to construct the covariance matrix of a counterexample distribution.

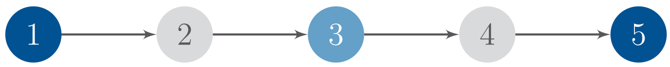
We know that the drift matrix
$$M^{**} = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$
induces a distribution in the Lyapunov model where (32) does not hold. The first step of the proof of Proposition 4.21 is the consecutive application of Corollary 4.20: we construct a node between node 1 and 2 that is highly correlated with node 1 as well as a node between node 2 and 3 that is highly correlated with node 2. The resulting distribution contradicts the statement
$$1 \perp\!\!\!\perp 5 \mid 3$$
in the model $\mathcal{M}_{G_5,C_5}$ on $p^* = 5$ nodes. In the proof of Corollary 4.20, we choose the final covariance matrix that is used as a counterexample from a neighborhood of the occurring limit. Consequently, we cannot write down the actual drift matrix that induces the counterexample. We denote this new 5×5 drift matrix with M^*.

Applying Corollary 4.14 constitutes the second step of the proof of Proposition 4.21: we construct nodes occurring before the first node and after the last node, each independent from all other nodes. We do this by adding -1 on the diagonal and 0 entries on the subdiagonal of the extended drift matrix, regardless of whether the added nodes are part of the conditioning set S. The resulting drift matrix

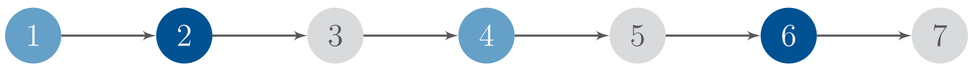
$$M = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & & & & & & 0 \\ 0 & & & & & & 0 \\ 0 & & & M^* & & & 0 \\ 0 & & & & & & 0 \\ 0 & & & & & & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

then yields a counterexample to (31). We illustrate the steps of extending a distribution in the Lyapunov model to a distribution on a longer path in Figure 7.

(a) Path with $q = 3$ nodes. We know that we can contradict statement (32) with a suitable distribution in the model.



(b) Path with $p^* = 5$ nodes. We extend the distribution to five nodes by adding two nodes before and after the conditioning node and shifting the indices accordingly. The added nodes are constructed so that they are highly correlated with existing nodes, thereby preserving the dependence structure.



(c) Path with $p = 7$ nodes. We extend the distribution to seven nodes by adding an independent conditioning node at the start and an independent node that does not appear in the statement at the end of the path. Indices are shifted accordingly. The dependence structure is preserved as well, so we constructed a counterexample for (31).

Figure 7: The two steps of proving Proposition 4.21 illustrated by an example. The conditioning nodes are marked light blue, and the newly constructed nodes are marked grey. For better readability, we do not draw the self-loops.

5 Towards a general statement

In the previous chapter, we gave a proof verifying our conjecture for a subset of conditional independence statements (14) in the path model, namely by restricting the number of conditioning variables between i and j appearing in the statement. We now consider potential strategies for expanding Theorem 4.1 to a general statement proving the conjecture while illustrating the arising challenges.

The crucial point in verifying the conjecture is the first part of the proof of Theorem 4.1, where we calculated determinants manually. The second part of the proof, where we applied Proposition 4.21, works for any number of variables in the statement if a valid counterexample exists showing that

$$1 \perp\!\!\!\perp p \mid 2, \dots, p-1 \tag{33}$$

does not hold in $\mathcal{M}_{G_p,C_p}$ for a suitable value $p \in \mathbb{N}_{\geq 2}$. Thus, finding a way to generate a counterexample to (33), i.e., a $\Sigma \in \mathcal{M}_{G_p,C_p}$ where the determinant $\det \Sigma_{1\dots p-1,2\dots p}$ is non-zero, for any $p \in \mathbb{N}_{\geq 2}$ would imply that we can prove the conjecture.

5.1 Computing determinants

We saw in the proof of Theorem 4.1 that the distribution induced by a drift matrix $M \in \mathrm{Stab}(E_p)$ with -1 on the diagonal, 1 on the first subdiagonal, and 0 everywhere else is a valid counterexample for $p = 2, \dots, 102$. Can we extend this example to arbitrary dimensions?

To determine independence, we computed the determinant of the upper right $(p-1) \times (p-1)$ submatrix of Σ that arises as a solution of the Lyapunov equation defined by M. The values of the computed determinants start to notably diverge away from zero at around $p = 50$. For instance, while the determinant of interest is 0.0025 for $p = 10$ and -0.9994 for $p = 50$, it already reaches values like 2.4637×10^9 for $p = 100$ and 6.6458×10^9 for $p = 102$. Performing the same computation for some randomly chosen larger values of p also yields non-zero determinants, indicating that the distribution induced by such a matrix M might be a valid counterexample for any p.

In these computations, we can observe that there is no recognizable pattern of the sign of the calculated determinants, in particular, the computed determinants are not exclusively positive or negative. Instead, the determinants are negative or positive in irregular frequencies, thereby oscillating around zero with increasing absolute value for larger values p.

Taking a closer look at the entries of such a covariance matrix Σ reveals that the matrix entries become smaller and smaller along its rows and columns. Consequently, the values of the entries are close together, which complicates the estimation of a bound away from zero for the determinant. One solution could be to construct the drift matrices in a way that the entries of Σ grow in such a manner along its rows and columns that determining a bound between zero and the determinant is possible.

One way to construct growing entries in Σ is by keeping the entries of the diagonal of M fixed at -1 as before and defining the entries on the subdiagonal as an ascending sequence, for example with $m_{i+1,i} = 2^i$ or $m_{i+1,i} = 10^i$. In principle, we could use Gaussian elimination on the submatrix $\Sigma_{1\ldots p-1,2\ldots p}$ to bring it in a diagonal form. Then we can determine whether its determinant is non-zero: it is non-zero if and only if all of the resulting diagonal entries are non-zero. However, a general proof is not straightforward because simulations show that such a matrix Σ still produces positive and negative determinants of the submatrix $\Sigma_{1\ldots p-1,2\ldots p}$ depending on the number of nodes p. Therefore, bounding the determinant away from zero is still a challenge.

A similar way to construct growing entries in Σ is by keeping the subdiagonal entries fixed at 1 and defining the diagonal entries as a descending sequence, for example with $m_{i,i} = 2^{-i}$. Simulations suggest that in this case, the determinant of interest stays positive in every dimension p. However, we are still left with the issue of computing a bound on the determinant asserting that the determinant always stays positive. It becomes clear that the approach of computing the determinant directly would require a tractable way of computing the determinant of the desired submatrix of Σ in any dimension p allowing us to bound the determinant away from zero. Thus, it might be more sensible to find a different approach that does not require the computation of a determinant that is not easily tractable in arbitrary dimensions.

5.2 Constructing independent conditioning nodes

We saw in Section 4.3 that we are able to add nodes between 1 and p such that they behave approximately like existing nodes in the model. In the proof of Theorem 4.1, we used this way of constructing nodes only for nodes not occurring in the conditioning set. Pursuing this idea further naturally leads to the idea of making the conditioning nodes highly correlated with each other: the conditioning nodes $2, \ldots, p-1$ then effectively behave as one node as in the path model with three nodes. Thus, we would only need to prove that (33) does not hold for $q = 3$, which we already have. At first glance, this strategy seems straightforward. However, it creates a new set of challenges.

5.2.1 The candidate matrix

We start with the three-node counterexample with drift and covariance matrices

$$M^* = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \in \mathrm{Stab}(E_3) \quad \text{and} \quad \Sigma^* = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{4} & \frac{7}{8} & \frac{15}{8} \end{pmatrix} \in \mathcal{M}_{G_3,C_3},$$

respectively. Now let $p \in \mathbb{N}_{>3}$. Can we follow a similar strategy as in Section 4.3 to construct a matrix Σ on the boundary of the model $\mathcal{M}_{G_p,C_p}$ such that the statement (33) does not hold in the distribution induced by Σ? If so, we can aim to show that there is also a matrix in the model where the statement does not hold.

Let $p \in \mathbb{N}_{\geq 3}$. The case $p = 2$ is considered in Chapter 3. Consider the matrix

$$\overline{\Sigma} := \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{2} & \cdots & \frac{3}{2} & \frac{7}{8} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{2} & \frac{3}{2} & \cdots & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{4} & \frac{7}{8} & \cdots & \frac{7}{8} & \frac{15}{8} \end{pmatrix} \in \mathbb{R}^{p \times p} \tag{34}$$

where all middle rows and columns are duplicates of each other. Following our deliberations above, we suspect that (33) does not hold in the distribution defined by this matrix.

To prove this, our first instinct might be to apply the determinant criterion in Lemma 3.10. However, the matrix $\overline{\Sigma}$ is only positive semi-definite for $p > 3$ and the submatrix $\overline{\Sigma}_{1\ldots p-1,\ldots p}$ is singular for $p > 3$. The criterion on the determinant only applies to multivariate normal distributions with positive definite covariance matrix. As a consequence, we have to resort back to the elementary condition for conditional independence in Lemma 3.9 based on the covariance matrix of the conditional distribution of $(X_1, X_p)^T | X_2, \ldots, X_{p-1}$.

Let $S := \{2, \ldots, p-1\}$. We want to show that the conditional covariance

$$(\overline{\Sigma}_{1p,1p \cdot S})_{12} = \overline{\Sigma}_{1,p} - \overline{\Sigma}_{1,S} (\overline{\Sigma}_{S,S})^{-} \overline{\Sigma}_{S,p} \tag{35}$$

of X_1 and X_p given X_S is non-zero in the distribution defined by $\overline{\Sigma}$.

As the generalized inverse $(\overline{\Sigma}_{S,S})^{-}$ in the formula (35), we choose the *Moore-Penrose inverse* denoted by $(\overline{\Sigma}_{S,S})^{+}$ since it can be easily computed via the singular value decomposition of the matrix $\overline{\Sigma}_{S,S}$.

Lemma 5.1. *We have*

$$(\overline{\Sigma}_{1p,1p \cdot S})_{12} \neq 0,$$

i.e., $1 \perp\!\!\!\perp p \mid 2, \ldots, p-1$ *does not hold for* $\mathbf{X} = (X_1, \ldots, X_p)^T \sim \mathcal{N}_p(0, \overline{\Sigma})$.

Proof. Let $\mathbf{X}$ follow the multivariate normal distribution $\mathcal{N}_p(0, \overline{\Sigma})$. If $p = 3$, the matrix $\overline{\Sigma}$ is positive definite. Then we can apply the determinant criterion where we already know that the determinant is non-zero.

Thus, we assume from now on that $p > 3$. Then, the matrix $\overline{\Sigma}$ is positive semi-definite. Following Lemma 3.9, the conditional covariance of X_1 and X_p given $\mathbf{X_S}$ is given by the Schur complement (35). Let $n := p - 2$. Hence, we have

$$\overline{\Sigma}_{1,p} = \frac{1}{4}, \quad \overline{\Sigma}_{1,S} = \frac{1}{2}(1, \ldots, 1) \in \mathbb{R}^{1 \times n}, \quad \text{and} \quad \overline{\Sigma}_{S,p} = \frac{7}{8} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{n \times 1}.$$

The only thing left to compute is the generalized inverse of

$$\overline{\Sigma}_{S,S} = \frac{3}{2} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

The Moore-Penrose inverse $(\overline{\Sigma}_{S,S})^+$ can be computed via the singular value decomposition $\overline{\Sigma}_{S,S} = U\Delta V^T$ with $U, V \in \mathbb{R}^{n\times n}$ orthonormal and $\Delta \in \mathbb{R}^{n\times n}$ diagonal. The columns of U are the eigenvectors of $\overline{\Sigma}_{S,S}\overline{\Sigma}_{S,S}{}^T$ and the columns of V are the eigenvectors of $\overline{\Sigma}_{S,S}{}^T\overline{\Sigma}_{S,S}$, while the diagonal entries of Δ are the square roots of the corresponding eigenvalues. Since $\overline{\Sigma}_{S,S}{}^T = \overline{\Sigma}_{S,S}$, we have $U = V$ and therefore only need to compute the eigenvectors and eigenvalues of $(\overline{\Sigma}_{S,S})^2$.

We have

$$(\overline{\Sigma}_{S,S})^2 = \frac{9}{4}\begin{pmatrix} n & \cdots & n \\ \vdots & \ddots & \vdots \\ n & \cdots & n \end{pmatrix} = \frac{9}{4}n\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

The eigenvalues are $\lambda_1 = \frac{9}{4}n^2$ and $\lambda_2 = \cdots = \lambda_n = 0$. A corresponding set of eigenvectors is formed by

$$\tilde{e}_1 = \begin{pmatrix} 1 \\ \\ \vdots \\ \\ 1 \end{pmatrix}, \quad \tilde{e}_2 = \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad \ldots, \quad \tilde{e}_n = \begin{pmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Since the matrix is symmetric, we know that the eigenspaces corresponding to the eigenvalues $\frac{9}{4}n^2$ and 0 are orthogonal to each other. Therefore, we normalize $\tilde{e}_1$ and then apply the Gram-Schmidt process to the eigenvectors $\tilde{e}_2, \ldots, \tilde{e}_n$ yielding an orthonormal basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$. Thus,

$$U = V = \begin{pmatrix} | & | & & | \\ e_1 & e_2 & \cdots & e_n \\ | & | & & | \end{pmatrix} \quad \text{and} \quad \Delta = \operatorname{diag}\left(\frac{3}{2}n, 0 \ldots, 0\right).$$

The Moore-Penrose inverse of $\overline{\Sigma}_{S,S}$ with

$$\Delta^+ = \operatorname{diag}\left(\frac{2}{3n}, 0, \ldots, 0\right)$$

is then

$$(\overline{\Sigma}_{S,S})^+ = V\Delta^+U^T = \frac{2}{3n}e_1 \cdot e_1{}^T = \frac{2}{3n}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Therefore,

$$(\overline{\Sigma}_{1p,1p\cdot S})_{12} = \overline{\Sigma}_{1,p} - \overline{\Sigma}_{1,S}\overline{\Sigma}_{S,S}^{-}\overline{\Sigma}_{S,p} = \frac{1}{4} - \frac{7}{16}\frac{2}{3n^2}n^2 = \frac{1}{4} - \frac{7}{24} = -\frac{1}{24} \neq 0,$$

which implies that $1 \perp\!\!\!\perp p \mid 2, \ldots, p-1$ does not hold in this distribution. $\square$

We have found the covariance matrix of a distribution where the conditional independence statement (33) does not hold. In this distribution, the nodes 2 to $p-1$ are perfectly correlated. It is obvious that the matrix $\overline{\Sigma}$ is singular and therefore not positive definite. Consequently, it cannot lie in the model $\mathcal{M}_{G_p,C_p}$. Instead, we show in the following that the matrix is an element of the closure $\overline{\mathcal{M}_{G_p,C_p}}$ of the model.

Lemma 5.2. *Let $p \geq 3$ and define $\overline{\Sigma} \in \mathbb{R}^{p\times p}$ as above. Then we have*

$$\overline{\Sigma} \in \overline{\mathcal{M}_{G_p,C_p}},$$

i.e., there is a sequence $(\Sigma^{(m)})_{m\in\mathbb{N}_{>0}}$ in $\mathcal{M}_{G_p,C_p}$ such that

$$\lim_{m\to\infty} \Sigma^{(m)} = \overline{\Sigma}.$$

Proof. We prove this fact by induction. The idea is to repeatedly apply Lemma 4.17 to a specific distribution on three nodes and thereby repeatedly duplicating the middle node. To be able to specify the matrix $\overline{\Sigma}$ in different dimensions, we indicate the dimension of the respective matrices via a subscript by writing $\overline{\Sigma}_{(3)} \in \mathbb{R}^{3\times 3}$, $\overline{\Sigma}_{(p-1)} \in \mathbb{R}^{(p-1)\times(p-1)}$, and $\overline{\Sigma}_{(p)} \in \mathbb{R}^{p\times p}$.

We start with the case $p = 3$ by considering the distribution induced by the drift matrix

$$M = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \in \mathrm{Stab}(E_3).$$

In Example 3.11, we saw that the Lyapunov equation

$$M\Sigma + \Sigma M^T + C_3 = 0$$

is solved by the covariance matrix $\Sigma = \overline{\Sigma}_{(3)}$, so $\overline{\Sigma}_{(3)} \in \mathcal{M}_{G_3,C_3} \subseteq \overline{\mathcal{M}_{G_3,C_3}}$. We can further set $\Sigma^{(m)} := \overline{\Sigma}_{(3)}$ for all $m \in \mathbb{N}_{>0}$.

For the induction step, let $p \in \mathbb{N}_{>3}$ and consider

$$\overline{\Sigma}_{(p-1)} = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{2} & \cdots & \frac{3}{2} & \frac{7}{8} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{2} & \frac{3}{2} & \cdots & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{4} & \frac{7}{8} & \cdots & \frac{7}{8} & \frac{15}{8} \end{pmatrix} \in \mathbb{R}^{(p-1)\times(p-1)}.$$

We assume that $\overline{\Sigma}_{(p-1)} \in \overline{\mathcal{M}_{G_{p-1},C_{p-1}}}$. Then, there is a sequence

$$(\Sigma^{(n)}_{(p-1)})_{n\in\mathbb{N}_{>0}} \text{ in } \mathcal{M}_{G_{p-1},C_{p-1}}$$

such that

$$\lim_{n\to\infty} \Sigma^{(n)}_{(p-1)} = \overline{\Sigma}_{(p-1)}.$$

We cannot directly apply Lemma 4.17 to $\overline{\Sigma}_{(p-1)}$, as the matrix that we want to extend has to be in the model $\mathcal{M}_{G_{p-1},C_{p-1}}$. Instead, we can apply Lemma 4.17 to the elements $\Sigma^{(n)}_{(p-1)}$ of the matrix sequence.

Due to Lemma 4.17, for every $n \in \mathbb{N}_{>0}$, there is a $p \times p$ matrix $\Gamma^{(n)} \in \overline{\mathcal{M}_{G_p,C_p}}$ such that

$$\Pi_{(-(p-1))}(\Gamma^{(n)}) = \Sigma^{(n)}_{(p-1)}$$

and

$$(\Gamma^{(n)})_{p-1,j} = (\Gamma^{(n)})_{p-2,j} \text{ as well as } (\Gamma^{(n)})_{j,p-1} = (\Gamma^{(n)})_{j,p-2} \text{ for all } j \in 1, \dots, p.$$

Thus, the matrix $\Gamma^{(n)}$ is constructed from $\Sigma^{(n)}_{(p-1)}$ by inserting a row and column at index $p-1$ such that the $(p-1)$-th row and column are a duplicate of the $(p-2)$-th row and column. This gives us a sequence of singular matrices in $\overline{\mathcal{M}_{G_p,C_p}} \setminus \mathcal{M}_{G_p,C_p}$.

Define $\Gamma := \lim_{n\to\infty} \Gamma^{(n)}$. Since $\overline{\mathcal{M}_{G_p,C_p}}$ is closed, we have $\Gamma \in \overline{\mathcal{M}_{G_p,C_p}}$. Note that the limit operation is applied element-wise to the matrices and that the projection map does not change the values of the remaining entries. Therefore, we have

$$\Pi_{(-(p-1))}(\Gamma) = \Pi_{(-(p-1))}(\lim_{n\to\infty} \Gamma^{(n)}) = \lim_{n\to\infty} \Pi_{(-(p-1))}(\Gamma^{(n)}) = \lim_{n\to\infty} \Sigma^{(n)}_{(p-1)} = \overline{\Sigma}_{(p-1)}$$

as well as

$$\Gamma_{p-1,j} = (\lim_{n\to\infty} \Gamma^{(n)})_{p-1,j} = \lim_{n\to\infty} (\Gamma^{(n)})_{p-1,j} = \lim_{n\to\infty} (\Gamma^{(n)})_{p-2,j} = (\lim_{n\to\infty} \Gamma^{(n)})_{p-2,j} = \Gamma_{p-2,j}$$

for all $j \in 1, \dots, p$, where the same holds for columns. This implies

$$\Gamma = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{2} & \cdots & \frac{3}{2} & \frac{3}{2} & \frac{7}{8} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{1}{2} & \frac{3}{2} & \cdots & \frac{3}{2} & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{2} & \frac{3}{2} & \cdots & \frac{3}{2} & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{4} & \frac{7}{8} & \cdots & \frac{7}{8} & \frac{7}{8} & \frac{15}{8} \end{pmatrix} \in \mathbb{R}^{p\times p},$$

so $\Gamma = \overline{\Sigma}_{(p)}$ and consequently $\overline{\Sigma}_{(p)} \in \overline{\mathcal{M}_{G_p,C_p}}$. We conclude that there is a sequence $(\Sigma^{(m)}_{(p)})_{m\in\mathbb{N}_{>0}}$ in $\mathcal{M}_{G_p,C_p}$ such that

$$\lim_{m\to\infty} \Sigma^{(m)}_{(p)} = \overline{\Sigma}_{(p)}.$$

□

Remark 5.3. It is important to note here that we only showed the existence of a sequence in $\mathcal{M}_{G_p,C_p}$ converging to $\overline{\Sigma}_{(p)}$, not an explicit construction of such a sequence. Consequently, we do not know what the matrices in the sequence look like. For the considerations presented in the following, the existence of such a sequence is sufficient. However, especially for simulations, it is useful to be able to explicitly construct a sequence. We claim that such a construction is possible by solving the Lyapunov equation for a sequence of drift matrices

$$M^{(m)} := \begin{pmatrix} -1 & & & & & \\ 1 & -1 & & & & \\ & m & -m & & & \\ & & \ddots & \ddots & & \\ & & & m & -m & \\ & & & & 1 & -1 \end{pmatrix} \in \mathrm{Stab}(E_p)$$

for $m \in \mathbb{N}_{>0}$. The resulting sequence of covariance matrices $\Sigma^{(m)}$ then converges to the matrix $\overline{\Sigma} \in \overline{\mathcal{M}_{G_p,C_p}}$ defined in (34). The proof is similar to the proof of Lemma 4.17. We partition the matrices $M^{(m)}$ and $\Sigma^{(m)}$ in nine block partitions and then solve, due to the symmetry of the equation, the resulting six unique equations of the block partitions by consecutively taking the limits and inserting the results from the equations solved before. We do not give a rigorous proof of this fact, as we do not necessarily require the result for the remainder of this thesis.

Combining the result of Lemma 5.1 and Lemma 5.2 shows that the independence statement

$$1 \perp\!\!\!\perp p \mid 2, \ldots, p-1 \tag{36}$$

does not hold in the closure $\overline{\mathcal{M}_{G_p,C_p}}$ of the model. All that remains is to find a distribution that is actually in the path model $\mathcal{M}_{G_p,C_p}$, such that (36) does not hold either in this distribution.

5.2.2 Continuity and the Schur complement

In the proof of Corollary 4.20, we were faced with a similar task – the only difference being that the submatrix of interest of the covariance matrix in the proof of Corollary 4.20 was non-singular so that the determinant criterion could be applied. There, we solved the boundary issue by exploiting the continuity of the determinant: if the determinant is non-zero on the model's boundary, then there is a matrix Σ in the interior and therefore in the model with non-zero determinant.

This strategy, however, is not applicable here as the submatrix $\overline{\Sigma}_{2\ldots p,1\ldots p-1}$ is singular, so we have to compute the conditional covariance directly as seen in Lemma 5.1. Now we face the challenge that, unlike the actual inverse, a generalized inverse is in general not continuous (Stewart, 1969; Ben-Israel and Greville, 2003). As a consequence, the conditional covariance that is defined as the Schur complement of a submatrix of $\overline{\Sigma}$ is in general not continuous either. Can we still prove convergence of the Schur complement in our specific example?

We keep the notation from Lemma 5.1 and Lemma 5.2 and assume that $p \in \mathbb{N}_{>3}$. We consider a sequence of positive definite covariance matrices $\Sigma^{(m)}$ in the model $\mathcal{M}_{G_p,C_p}$ converging to the positive semi-definite but not positive definite covariance matrix $\overline{\Sigma}$ in the closure $\overline{\mathcal{M}_{G_p,C_p}}$ of the model, that is

$$\Sigma^{(m)} \to \overline{\Sigma} \text{ for } m \to \infty.$$

The conditional covariance of X_1 and X_p given X_S obtained through the Schur complement is

$$\Sigma^{(m)}_{\text{cond}} := (\Sigma^{(m)}_{1p,1p\cdot S})_{12} = \Sigma^{(m)}_{1,p} - \Sigma^{(m)}_{1,S}(\Sigma^{(m)}_{S,S})^{-1}\Sigma^{(m)}_{S,p} \in \mathbb{R}$$

for every $\Sigma^{(m)}$ and

$$\overline{\Sigma}_{\text{cond}} := (\overline{\Sigma}_{1p,1p\cdot S})_{12} = \overline{\Sigma}_{1,p} - \overline{\Sigma}_{1,S}(\overline{\Sigma}_{S,S})^{+}\overline{\Sigma}_{S,p} \in \mathbb{R}$$

for $\overline{\Sigma}$. Given that $\Sigma^{(m)} \to \overline{\Sigma}$, the question is now whether

$$\Sigma^{(m)}_{\text{cond}} \to \overline{\Sigma}_{\text{cond}} \text{ for } m \to \infty.$$

For easier notation, we set $n := p-2$ and we define

$$\begin{aligned} v^{(m)} &:= \Sigma^{(m)T}_{1,S} \quad \text{and } v := \overline{\Sigma}_{1,S}^{\;T} = \frac{1}{2}1_n, \\ w^{(m)} &:= \Sigma^{(m)}_{S,p} \quad \text{and } w := \overline{\Sigma}_{S,p} = \frac{7}{8}1_n, \\ A^{(m)} &:= \Sigma^{(m)}_{S,S} \quad \text{and } A := \overline{\Sigma}_{S,S} = \frac{3}{2}1_{n\times n}. \end{aligned}$$

Then we have $v = \lim_{m\to\infty} v^{(m)}$, $w = \lim_{m\to\infty} w^{(m)}$, and $A = \lim_{m\to\infty} A^{(m)}$. We can illustrate these definitions in matrix notation as

$$\Sigma^{(m)} = \left(\begin{array}{c|c|c} * & \text{---}\quad v^{(m)T} \quad\text{---} & \Sigma^{(m)}_{1,p} \\ \hline * & A^{(m)} & \begin{array}{c} | \\ w^{(m)} \\ | \end{array} \\ \hline * & * & * \end{array}\right) \xrightarrow{m\to\infty} \left(\begin{array}{c|c|c} * & \text{---}\quad v^{T} \quad\text{---} & \overline{\Sigma}_{1,p} \\ \hline * & A & \begin{array}{c} | \\ w \\ | \end{array} \\ \hline * & * & * \end{array}\right) = \overline{\Sigma}. \tag{37}$$

The first summand of $\Sigma^{(m)}_{\text{cond}}$, i.e., $\Sigma^{(m)}_{1,p}$, is an entry of $\Sigma^{(m)}$ so we know that it is convergent. The second part of the term is the inverse of $\Sigma^{(m)}_{S,S}$ multiplied with two vectors that are submatrices of $\Sigma^{(m)}$ – in the newly introduced notation it is the product $v^{(m)T}(A^{(m)})^{-1}w^{(m)}$. The following example suggests that this product might be convergent as well.

Example 5.4. We consider the case $p = 4$, where we have

$$\overline{\Sigma} = \left(\begin{array}{c|cc|c} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ \hline \frac{1}{2} & \frac{3}{2} & \frac{3}{2} & \frac{7}{8} \\ \frac{1}{2} & \frac{3}{2} & \frac{3}{2} & \frac{7}{8} \\ \hline \frac{1}{4} & \frac{7}{8} & \frac{7}{8} & \frac{15}{8} \end{array}\right)$$

and

$$\Sigma^{(m)} = \left(\begin{array}{c|cc|c} 1 & \frac{1}{2} & \frac{1}{2}\frac{m}{1+m} & \frac{1}{4}\frac{m}{1+m} \\ \hline \frac{1}{2} & \frac{3}{2} & \frac{4m+3m^2}{2(1+m)^2} & \frac{9m+7m^2}{8(1+m)^2} \\ \frac{1}{2}\frac{m}{1+m} & \frac{4m+3m^2}{2(1+m)^2} & \frac{2+4m+6m^2+3m^3}{2m(1+m)^2} & \frac{8+16m+24m^2+21m^3+7m^4}{8m(1+m)^3} \\ \hline \frac{1}{4}\frac{m}{1+m} & \frac{9m+7m^2}{8(1+m)^2} & \frac{8+16m+24m^2+21m^3+7m^4}{8m(1+m)^3} & \frac{8+24m+48m^2+45m^3+15m^4}{8m(1+m)^3} \end{array}\right)$$

constructed via Lemma 4.17 through a sequence of drift matrices

$$M^{(m)} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & m & -m & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

This implies $S = \{2, 3\}$ and $n = 2$. We further define $v^{(m)}$, $w^{(m)}$, $A^{(m)}$, as well as the corresponding limits v, w, and A as above in (37).

Observe that for every $m \in \mathbb{N}_{>0}$, we have

$$\det A^{(m)} = \frac{6 + 24m + 48m^2 + 41m^3 + 12m^4}{4m(1+m)^4} > 0,$$

so the matrix $A^{(m)}$ is invertible with

$$(A^{(m)})^+ = (A^{(m)})^{-1} = \begin{pmatrix} \frac{2(1+m)^2(2+4m+6m^2+3m^3)}{6+24m+48m^2+41m^3+12m^4} & -\frac{2m^2(1+m)^2(4+3m)}{6+24m+48m^2+41m^3+12m^4} \\ -\frac{2m^2(1+m)^2(4+3m)}{6+24m+48m^2+41m^3+12m^4} & \frac{6m(1+m)^4}{6+24m+48m^2+41m^3+12m^4} \end{pmatrix}.$$

On the contrary, A is not invertible – due to its linearly dependent rows and columns. The Moore-Penrose inverse of A computed via the singular value decomposition as in Lemma 5.1 is

$$A^+ = \frac{2}{3} \cdot \frac{1}{2^2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{9} A.$$

Unfortunately, we have

$$\lim_{m\to\infty} (A^{(m)})^+ = \lim_{m\to\infty} (A^{(m)})^{-1} = \begin{pmatrix} \infty & -\infty \\ -\infty & \infty \end{pmatrix} \neq A^+,$$

so we cannot interchange taking the limit and forming the Moore-Penrose inverse. This example illustrates that the Moore-Penrose inverse is in general not continuous.

Remark 5.5. There exist special cases where continuity holds (Stewart, 1969; Ben-Israel and Greville, 2003). For example, under specific conditions on the error, a sequence of matrices converges in the Moore-Penrose inverse if and only if the rank of the matrices converges as well. In general terms, convergence of the Moore-Penrose inverse of $A^{(m)}$ is equivalent to the existence of $m_0 \in \mathbb{N}$ such that $\operatorname{rank}(A^{(m)}) = \operatorname{rank}(A)$ for all $m > m_0$. This is not the case in our example, as the matrices in the sequence all have full rank while $\operatorname{rank}(A) = 1$.

We turn our attention to the product $\left(v^{(m)}\right)^T A^{(m)} w^{(m)}$ that we suspect to converge to the product $v^T A w$ in this setting.

Example 5.6 (continuation of Example 5.4)**.** We make the following observation:

$$(A^{(m)})^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \left(\begin{pmatrix} 1 & 1 \end{pmatrix} (A^{(m)})^{-1}\right)^T = \begin{pmatrix} \frac{4(1+m)^4}{6+24m+48m^2+41m^3+12m^4} \\ \frac{2m(1+m)^2(3+2m)}{6+24m+48m^2+41m^3+12m^4} \end{pmatrix} \xrightarrow{m\to\infty} \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}.$$

Therefore, we have

$$v^T(A^{(m)})^{-1} \xrightarrow{m\to\infty} \frac{1}{2}\begin{pmatrix}\frac{1}{3} & \frac{1}{3}\end{pmatrix} = \begin{pmatrix}\frac{1}{6} & \frac{1}{6}\end{pmatrix}$$

and

$$(A^{(m)})^{-1}w \xrightarrow{m\to\infty} \frac{7}{8}\begin{pmatrix}\frac{1}{3} \\ \frac{1}{3}\end{pmatrix} = \begin{pmatrix}\frac{7}{24} \\ \frac{7}{24}\end{pmatrix}.$$

Computing the full bilinear form yields

$$v^T(A^{(m)})^{-1}w \to \frac{7}{24}.$$

The products $(v^{(m)})^T A^{(m)}$, $A^{(m)}w^{(m)}$, and $(v^{(m)})^T A^{(m)}w^{(m)}$ converge to these values as well. Thus, the Moore Penrose inverse of $A^{(m)}$ seems to converge when multiplied with specific vectors, for example, the vectors v and w.

Since they are symmetric, the matrices A and $A^{(m)}$ are diagonalizable. The matrices A and A^+ have rank 1, so they have exactly one non-zero eigenvalue. For both matrices, the eigenvector generating the eigenspace of the non-zero eigenvalue is

$$\begin{pmatrix}1 \\ \vdots \\ 1\end{pmatrix} \in \mathbb{R}^n.$$

Note that $v, w \in \mathbb{R}^n$ are both eigenvectors from this eigenspace. We did observe in Example 5.6 that the bilinear form

$$(v^{(m)})^T(A^{(m)})^{-1}w^{(m)} \tag{38}$$

converges and we know that $v^{(m)}$ converges to the eigenvector v and $w^{(m)}$ to the eigenvector w. The matrix-vector products $(v^{(m)})^T(A^{(m)})^{-1}$ and $(A^{(m)})^{-1}w^{(m)}$ converge as well. Do these observations with respect to eigenvectors hold in general?

Example 5.7. The convergence of the bilinear form (38) observed in Example 5.6 does not necessarily hold in a general example. Assume

$$A = \begin{pmatrix}1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0\end{pmatrix} \in \mathbb{R}^{n\times n} \quad \text{and} \quad A^{(m)} = \begin{pmatrix}1 & 0 & \cdots & 0 \\ 0 & \varepsilon^{(m)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varepsilon^{(m)}\end{pmatrix} \in \mathbb{R}^{n\times n}$$

such that $\varepsilon^{(m)} \to 0$ for $m \to \infty$. Further, let

$$v = w = \begin{pmatrix}1 \\ 0 \\ \vdots \\ 0\end{pmatrix} \in \mathbb{R}^n \quad \text{and} \quad v^{(m)} = w^{(m)} = \begin{pmatrix}1 \\ \eta^{(m)} \\ \vdots \\ \eta^{(m)}\end{pmatrix} \in \mathbb{R}^n$$

such that $\eta^{(m)} \to 0$ for $m \to \infty$. Then, we have $A^{(m)} \to A$, $v^{(m)} \to v$, and $w^{(m)} \to w$ for $m \to \infty$. The vectors v and w are eigenvectors of A (and of $A^{(m)}$). The inverse of the diagonal matrix $A^{(m)}$ is

$$(A^{(m)})^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & (\varepsilon^{(m)})^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\varepsilon^{(m)})^{-1} \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Computing the vector-matrix product yields

$$(A^{(m)})^{-1} w^{(m)} = \begin{pmatrix} 1 \\ \eta^{(m)}(\varepsilon^{(m)})^{-1} \\ \vdots \\ \eta^{(m)}(\varepsilon^{(m)})^{-1} \end{pmatrix}.$$

For example, with $\eta^{(m)} = \frac{1}{m}$ and $\varepsilon^{(m)} = \frac{1}{m^2}$, we have $\eta^{(m)}(\varepsilon^{(m)})^{-1} = m \to \infty$, so the inverse does not converge along $w^{(m)}$ for $m \to \infty$. Computing the bilinear form yields

$$(v^{(m)})^T (A^{(m)})^{-1} w^{(m)} = 1 + (n-1)(\eta^{(m)})^2 (\varepsilon^{(m)})^{-1}.$$

With, for example, $\eta^{(m)} = \frac{1}{m}$ and $\varepsilon^{(m)} = \frac{1}{m^3}$, we have $\eta^{(m)}(\varepsilon^{(m)})^{-1} = m \to \infty$. Then the bilinear form does not converge.

The example shows that it is not possible to show convergence of the conditional covariance matrix $\Sigma_{\text{cond}}^{(m)}$ based on the fact that $v = \Sigma_{1,S}$ and $w = \Sigma_{S,p}$ are eigenvectors of $A = \Sigma_{S,S}$. Instead, stronger assumptions are required.

5.3 Continuity of diagonalization

Since we already diagonalized A in order to compute its Moore-Penrose inverse, the idea to diagonalize the sequence elements $A^{(m)}$ as well presents itself naturally. Then we can investigate the convergence properties of the diagonal matrices and the change-of-basis matrices separately.

Remember that each matrix $A^{(m)}$ in the sequence is real, symmetric, and positive definite, so it can be diagonalized by an orthogonal matrix $U^{(m)} \in \mathbb{R}^{n \times n}$ such that

$$D^{(m)} := (U^{(m)})^T A^{(m)} U^{(m)}$$

is a diagonal matrix with real non-negative entries. Consequently, we have

$$A^{(m)} = U^{(m)} D^{(m)} (U^{(m)})^T.$$

Note that all eigenvalues of $A^{(m)}$ are real and positive. Since $A^{(m)}$ is invertible, we can write

$$(A^{(m)})^{-1} = U^{(m)} (D^{(m)})^{-1} (U^{(m)})^T$$

or equivalently

$$(D^{(m)})^{-1} = (U^{(m)})^T (A^{(m)})^{-1} U^{(m)}.$$

Since A is real and symmetric as well, we can perform a similar diagonalization for A with an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ such that

$$D := U^T AU$$

is a diagonal matrix where the eigenvalues of A are on the diagonal.

Without loss of generality, we assume that the diagonal entries of $D^{(m)}$ and D are in descending order. Since A has rank 1, we know that the diagonal of D consists of one non-zero entry followed by $n-1$ zero entries.

5.3.1 Eigenvalues

First, we study the convergence properties of the diagonal matrix $D^{(m)}$. Its diagonal consists of the n eigenvalues of $A^{(m)}$. These eigenvalues are real since $A^{(m)}$ is symmetric with real entries for all $m \in \mathbb{N}_{>0}$ and they are positive since $A^{(m)}$ is positive definite for all $m \in \mathbb{N}_{>0}$.

We denote the i-th largest eigenvalue of a symmetric real matrix B as $\lambda_i(B) \in \mathbb{R}$ for $i \in [n]$, so we have

$$\lambda_1(B) \geq \cdots \geq \lambda_n(B).$$

Weyl's perturbation theorem states that two hermitian matrices B and C fulfill

$$\max_i |\lambda_i(B) - \lambda_i(C)| \leq \|B - C\|$$

if the eigenvalues are ordered in a descending manner where $\|\cdot\|$ denotes the operator norm. The statement can be extended to

$$\|\mathrm{diag}(\lambda_1(B), \ldots, \lambda_n(B)) - \mathrm{diag}(\lambda_1(C), \ldots, \lambda_n(C))\| \leq \|B - C\|. \tag{39}$$

Note that this holds not only for the operator norm but any unitarily invariant norm. In other words, the eigenvalues are continuous functions on the space of hermitian matrices (Bhatia, 1997).

We can enumerate the eigenvalues of $A^{(m)}$ for every $m \in \mathbb{N}_{>0}$ in descending order

$$\lambda_1(A^{(m)}) \geq \cdots \geq \lambda_n(A^{(m)})$$

as well. A similar enumeration can be done for A. Since $A^{(m)}$ and A are symmetric real matrices, statement (39) above implies

$$\|D^{(m)} - D\| \leq \|A^{(m)} - A\|$$

for all $m \in \mathbb{N}_{>0}$. We conclude that if $A^{(m)}$ converges to A for $m \to \infty$, then $D^{(m)}$ converges to D for $m \to \infty$.

Remark 5.8. An alternative argument for symmetric real matrices can be given as follows. The eigenvalues of a matrix B are the roots of the characteristic polynomial of B. The coefficients of the characteristic polynomial are continuous functions in the entries of B and therefore in B itself. The roots of the polynomial are continuous functions in the coefficients of the characteristic polynomial and therefore also in B (Bhatia, 1997). Combining these results, we have that the set of eigenvalues of B is continuous in B (Kato, 1995). Choosing the maximum element of a finite set is also continuous, so by ordering the eigenvalues of $A^{(m)}$ for every $m \in \mathbb{N}_{>0}$ in descending order as above we can parametrize them as continuous functions in $A^{(m)}$ (Kato, 1995). Then, if $A^{(m)}$ converges to A for $m \to \infty$, each eigenvalue of $A^{(m)}$ converges to a corresponding eigenvalue of A.

We know that the matrix A has two distinct eigenvalues, $\lambda := \lambda_1(A) = \frac{3}{2}n$ with multiplicity 1 and $\lambda_2(A) = \cdots = \lambda_n(A) = 0$ with multiplicity $n-1$. Denoting the eigenvalues of $A^{(m)}$ by $\lambda_i^{(m)} := \lambda_i(A^{(m)})$ for $i \in [n]$, we have due to their ordering

$$D^{(m)} = \begin{pmatrix} \lambda_1^{(m)} & & & \\ & \lambda_2^{(m)} & & \\ & & \ddots & \\ & & & \lambda_n^{(m)} \end{pmatrix} \xrightarrow{m\to\infty} \begin{pmatrix} \frac{3}{2}n & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} = D.$$

5.3.2 Eigenvectors

As a next step, we study the convergence of eigenvectors. First, we revisit Example 5.4.

Example 5.9 (continuation of Example 5.4)**.** If we diagonalize the matrices $A^{(m)}$ and A as proposed, we have diagonal matrices

$$D^{(m)}$$
$$= \begin{pmatrix} \frac{2+7m+12m^2+6m^3+\sqrt{4+4m+m^2+64m^4+96m^5+36m^6}}{4m(1+m)^2} & 0 \\ 0 & \frac{2+7m+12m^2+6m^3-\sqrt{4+4m+m^2+64m^4+96m^5+36m^6}}{4m(1+m)^2} \end{pmatrix}$$

and

$$D = \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}.$$

We form the orthogonal matrix $U^{(m)}$ by choosing corresponding orthonormal eigenvectors of $A^{(m)}$. In this example, simulations show that the resulting change-of-basis matrix $U^{(m)}$ actually converges to an orthogonal matrix that can be chosen as the change-of-basis matrix U for diagonalizing A.

Can we in general choose the eigenvectors that form $U^{(m)}$ in a way such that $U^{(m)}$ converges to U? This task is not as straightforward as establishing the convergence of the eigenvalues. The eigenspaces behave in a way more singular than the eigenvalues. To illustrate this fact, we study an example.

Example 5.10. For $x \in \mathbb{R}$, consider the matrix function

$$B(x) = \begin{cases} \begin{pmatrix} 1 & x \\ x & 1 \end{pmatrix}, & \text{if } x \geq 0; \\ \begin{pmatrix} 1+x & 0 \\ 0 & 1-x \end{pmatrix}, & \text{if } x < 0; \end{cases}.$$

Note that we have

$$B(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

so $B(x)$ is continuous in x. The eigenvalues of $B(x)$ are given by $1+x$ and $1-x$ for $x \in \mathbb{R}$, so they are distinct for $x \neq 0$. Thus, the eigenvectors of $B(x)$ are uniquely determined up to multiplicity for $x \neq 0$. For $x > 0$, they are, for example, given by the orthonormal vectors $v_1 = (1, 0)^T$ and $v_2 = (0, 1)^T$. For $x < 0$, they are, for instance, given by the orthonormal vectors $w_1 = \frac{1}{\sqrt{2}}(1, 1)^T$ and $w_2 = \frac{1}{\sqrt{2}}(1, -1)^T$. At $x = 0$, the matrix $B(x)$ only has eigenvalue 1, so the eigenvectors can be chosen as any, possibly orthonormal, basis of $\mathbb{R}^2$. It is clear that regardless of what we choose as eigenvectors for $B(0)$, the eigenvectors for $x > 0$ and $x < 0$ stay apart. In other words, it is not possible to choose orthonormal eigenvectors of $B(x)$ continuously in x (Bhatia, 1997).

The example shows that given a matrix function $B(x)$ depending on a real parameter x, continuity in x of said function is not sufficient to ensure the existence of orthonormal eigenvectors continuous in x. There are further examples that illustrate that even if the matrix function is infinitely differentiable for real x, the eigenvectors might not be continuous (Kato, 1995). Therefore, stronger assumptions on the function $B(x)$ are needed. We recap the required notions as follows.

A complex function defined on an open set V is *holomorphic* in V if it is complex differentiable at every point of its domain V (Freitag and Busam, 2009). The function is called holomorphic at a point x_0 if it is holomorphic in a neighborhood of x_0 and it is called holomorphic on a non-open set X if it is holomorphic at every point of X. If a function is holomorphic, it can be shown that it is infinitely complex differentiable at every point of the domain (Freitag and Busam, 2009). This does not hold the other way around when restricting to a point (Conway, 1995): if a function is infinitely complex differentiable at a point, this does not imply that it is differentiable at any other point in an open neighborhood. Thus, being holomorphic is a stronger property than being infinitely (complex) differentiable. Note that the notion of *analytic* functions, i.e., functions that can locally be represented by convergent power series, is closely related as it can be shown that complex analytic functions are equivalent to holomorphic functions (Conway, 1995; Krantz and Parks, 2002). Consequently, the term analytic is often used in literature instead of holomorphic.

Given a family of symmetric operators $A(x)$ with $x \in C$ for some open set $C \in \mathbb{C}$ intersecting with the real axis, we know that for each real $x \in C$ there is an orthonormal basis of eigenvectors $\{u_i(x)\}_{i\in[n]}$. Kato (1995) shows that if $A(x)$ is holomorphic in x, the orthonormal eigenvectors can be chosen as holomorphic and therefore continuous functions in x as well for every real x. In other words, the function $A(x)$ being holomorphic and symmetric is sufficient for the existence of an orthonormal basis that depends smoothly on x.

Let $A(m) := A^{(m)}$ for $m \in \mathbb{N}_{>0}$. We define the matrix function $A(x)$ by continuously extending $A(m)$ to the positive real line by replacing the parameter m with positive real values x, i.e., we define the map

$$A: \ (0, \infty) \to \mathbb{R}^{n\times n}, \ x \mapsto A(x) := A^{(x)}$$

for $x \in \mathbb{R}_{>0}$. Note that the entries of $A(x)$ are quotients of polynomials in x. In particular, the coefficients of the polynomials in the denominator of each entry of $A(x)$ are all positive due to the recursive construction of $A^{(m)}$ by solving the Lyapunov equation. Thus, the polynomials in the denominators do not have roots in $\mathbb{R}_{>0}$.

Polynomials are holomorphic functions in $\mathbb{C}$. Quotients of holomorphic functions away from the roots of the denominator are holomorphic as well (Conway, 1995). Therefore, $A(x)$ is holomorphic on the complement $C \subseteq \mathbb{C}$ of the roots of the denominators of all matrix entries of $A(x)$. The set C is open and it contains $\mathbb{R}_{>0}$, so we obtain that $A(x)$ is holomorphic on $\mathbb{R}_{>0}$.

Thus, we can conclude that there is an orthonormal basis $\{u_i(x)\}_{i\in[n]}$ that depends continuously on x and consist of eigenvectors of $A(x)$. We choose these vectors as the columns of the change-of-basis matrix $U^{(m)}$ by setting $u_i^{(m)} := u_i(m)$ ordered such that they correspond to the correct eigenvalues. Then, we can write

$$U^{(m)} := \begin{pmatrix} | & | & & | \\ u_1^{(m)} & u_2^{(m)} & \dots & u_n^{(m)} \\ | & | & & | \end{pmatrix}.$$

As $A^{(m)}$ goes to A for $m \to \infty$, the holomorphically chosen eigenvectors $u_i^{(m)}$ of $A^{(m)}$ converge to eigenvectors u_i of A, so we can choose

$$U := \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_n \\ | & | & & | \end{pmatrix}.$$

5.4 Challenges

We recapitulate the progress we made so far. Remember that we want to show that

$$(v^{(m)})^T \underbrace{U^{(m)}(D^{(m)})^{-1}(U^{(m)})^T}_{(A^m)^{-1}} w^{(m)} \xrightarrow{m\to\infty} v^T \underbrace{UD^+U^T}_{A^+} w, \tag{40}$$

in other words that the Moore-Penrose inverse of $A^{(m)}$ converges in this bilinear form multiplied with $v^{(m)}$ and $w^{(m)}$ to the Moore-Penrose inverse of A multiplied with v and w. This would imply convergence of the conditional covariances $\Sigma_{\text{cond}}^{(m)} \to \overline{\Sigma}_{\text{cond}}$ for $m \to \infty$.

We collect everything we know about the convergence of the variables in the statement:

- The invertible diagonal matrix $D^{(m)}$ converges to the singular diagonal matrix D (see Section 5.3.1).
- The orthogonal matrix $U^{(m)}$ converges to the orthogonal matrix U (both defined in Section 5.3.2).
- The vectors $v^{(m)}$ and $w^{(m)}$ converge to the vectors v and w, respectively (as they are subvectors of $\Sigma^{(m)}$ and Σ).

The Moore-Penrose inverse of the diagonal matrix $D^{(m)}$ does not converge by itself due to some eigenvalues converging to zero, so we need to consider the product as a whole. Further note that v and w are in the eigenspace belonging to the non-zero eigenvalue λ of A. Therefore, they are orthogonal to all columns of U except for the first column. Thus, we have $u_i^T v = 0$ for $2 \leq i \leq n$.

Both vector-matrix products

$$(v^{(m)})^T U^{(m)} \to v^T U \quad \text{and} \quad (U^{(m)})^T w^{(m)} \to U^T w$$

converge since their factors converge themselves. Then, it is enough to prove the convergence of either $((v^{(m)})^T U^{(m)})(D^{(m)})^{-1}$ or $(D^{(m)})^{-1}(U^{(m)} w^{(m)})$ in (40), as the respective other factors already converge. Thus, we aim to prove that

$$((v^{(m)})^T U^{(m)})(D^{(m)})^{-1} \to (v^T U) D^+$$

for $m \to \infty$.

Writing out the definitions yields

$$((v^{(m)})^T U^{(m)})(D^{(m)})^{-1} = \begin{pmatrix} (u_1^{(m)})^T v^{(m)} \\ (u_2^{(m)})^T v^{(m)} \\ \vdots \\ (u_n^{(m)})^T v^{(m)} \end{pmatrix}^T (D^{(m)})^{-1} = \begin{pmatrix} \frac{(u_1^{(m)})^T v^{(m)}}{\lambda_1^{(m)}} \\ \frac{(u_2^{(m)})^T v^{(m)}}{\lambda_2^{(m)}} \\ \vdots \\ \frac{(u_n^{(m)})^T v^{(m)}}{\lambda_n^{(m)}} \end{pmatrix}^T$$

and

$$(v^T U) D^+ = \begin{pmatrix} \frac{(u_1)^T v}{\lambda} \\ 0 \\ \vdots \\ 0 \end{pmatrix}^T.$$

Since one eigenvalue of $A^{(m)}$ converges to $\lambda \neq 0$ and $n-1$ eigenvalues converge to 0, we need to distinguish two cases. We have to show that the quotients

$$q_i^{(m)} := \frac{(u_i^{(m)})^T v^{(m)}}{\lambda_i^{(m)}}, \quad i \in [n],$$

converge to $\frac{(u_1)^T v}{\lambda}$ for $i = 1$ and to 0 for $2 \leq i \leq n$ when $m \to \infty$. The case $i = 1$ is straightforward since $\lambda_1^{(m)}$ converges to $\lambda \neq 0$. The numerator $(u_1^{(m)})^T v^{(m)}$ converges to $u_1^T v$, so $q_1^{(m)}$ converges as well for $m \to \infty$.

The case $2 \leq i \leq n$ is not as straightforward, as the eigenvalue $\lambda_i^{(m)}$ converges to zero. The numerator $(u_i^{(m)})^T v^{(m)}$ converges to $u_i^T v = 0$ as well, so we have to show that $(u_i^{(m)})^T v^{(m)}$ goes to zero faster than the eigenvalue itself. If we can prove that

$$q_i^{(m)} \xrightarrow{m \to \infty} 0 \tag{41}$$

for $2 \leq i \leq n$, then we can conclude that the conditional covariance $\Sigma_{\text{cond}}^{(m)}$ converges to $\overline{\Sigma}_{\text{cond}}$ for $m \to \infty$. Since we already know that $\overline{\Sigma}_{\text{cond}} \neq 0$, we can then find a $\hat{\Sigma} \in \mathcal{M}_{G_p,C_p}$ in a neighborhood of $\overline{\Sigma}_{\text{cond}}$ such that the conditional independence is non-zero in this distribution as well. Proving (41) lies beyond the scope of this thesis. We conclude the chapter by formulating the challenges that remain on the way to a proof of (41), and we present some initial ideas on how to address these challenges.

Rate of convergence. Let $2 \leq i \leq n$. We want to show that the $q_i^{(m)}$ converge to zero or, in other words, that the denominators of the $q_i^{(m)}$ converge more slowly to zero than the numerators of the $q_i^{(m)}$. That means we have to relate the convergence of the numerators $\left(u_i^{(m)}\right)^T v^{(m)}$ to the convergence of the eigenvalues $\lambda_i^{(m)}$. This marks the main challenge on the way to a proof.

As a first idea, we can infer

$$\begin{aligned}
|(u_i^{(m)})^T v^{(m)} - 0| &= |(u_i^{(m)})^T v^{(m)} - u_i^T v| \\
&= |(u_i^{(m)})^T v^{(m)} - (u_i^{(m)})^T v + (u_i^{(m)})^T v - u_i^T v| \\
&= |(u_i^{(m)})^T (v^{(m)} - v) + (u_i^{(m)} - u_i)^T v| \\
&\leq |(u_i^{(m)})^T (v^{(m)} - v)| + |(u_i^{(m)} - u_i)^T v| \\
&\leq \|u_i^{(m)}\|_2 \|v^{(m)} - v\|_2 + \|u_i^{(m)} - u_i\|_2 \|v\|_2 \\
&= \|v^{(m)} - v\|_2 + \frac{\sqrt{n}}{2} \|u_i^{(m)} - u_i\|_2 \xrightarrow{m \to \infty} 0,
\end{aligned}$$

where we use the triangle inequality, the Cauchy-Schwarz inequality, and eventually the convergence of $v^{(m)}$ and $u_i^{(m)}$. This result allows us to trace the convergence of the product back to the convergence of its factors $v^{(m)}$ and $u_i^{(m)}$. We do not, however, have any information on the rate of convergence of the continuously chosen orthonormal eigenvectors $u_i^{(m)}$ as well as the eigenvalues $\lambda^{(m)}$ that go to zero.

Tractable formula. The deliberations in Section 5.3.2 only verified the existence, not the construction of the continuously chosen orthonormal eigenvectors. Therefore, we do not have any formula for the entries of the eigenvectors $u_i^{(m)}$ dependent on $\Sigma^{(m)}$ that we could use to determine the rate of convergence. Using the explicit construction of $\Sigma^{(m)}$ described in Remark 5.3 would not solve this issue, as the eigenvectors are then still chosen as described in Section 5.3.2.

The eigenvalues are obtained by solving the characteristic polynomial of $A^{(m)}$. Due to the construction of $\Sigma^{(m)}$ via Lemma 5.2, we do not have an explicit formula for the entries of $A^{(m)}$ as well. Even if we had such a formula, for example, obtained by following the strategy described in Remark 5.3, the eigenvalues cannot be formulated as a function of the entries of $A^{(m)}$ in a general way that might be helpful to determine the rate of convergence. As seen in Example 5.6, the eigenvalues are already rather complicated for $n = 2$.

To summarize, it is not immediately helpful to resort to the direct construction of $\Sigma^{(m)}$ described in Remark 5.3. However, an explicit construction might allow the direct computation of the error $E^{(m)} := A^{(m)} - A$. Simulations suggest that the error $E^{(m)}$ goes to zero with $\frac{1}{m^2}$. This information, in turn, could be beneficial when considering the convergence rate of the eigenvectors. We conclude that these considerations could be a starting point for future work.

6 Conclusion and outlook

In this thesis, we investigated the conditional independence properties of directed paths in the recently proposed graphical continuous Lyapunov model. The model parametrizes the covariance matrices of its distributions as solutions of the continuous Lyapunov equation via appropriate drift and volatility matrices. The Lyapunov equation is induced by the Ornstein-Uhlenbeck process, a multivariate diffusion process that models, for example, the movement of particles in a fluid. By assuming a random vector to arise from the Ornstein-Uhlenbeck process in equilibrium, a temporal perspective is incorporated in the model that allows the modeling of cycles and thus feedback loops.

After laying out the theoretical foundation for the Lyapunov model as well as the concept of conditional independence in the multivariate normal distribution, we presented first examples of the Lyapunov model on the directed path of lengths 2, 3, and 4. In particular, for the path of length 3, we constructed a counterexample to all possible conditional independence statements in the model, i.e., a distribution in the model where none of these statements hold. Furthermore, we found that at least for some of the conditional independence statements, it is not easy to construct example distributions where the statements do hold.

Our main contribution is twofold. First, we devised a way to reduce finding a counterexample to any statement $i \perp\!\!\!\perp j \mid S$ on the path of length p to finding a counterexample to the statement $1 \perp\!\!\!\perp q \mid 2, \dots, q-1$ for a suitably chosen $q < p$. Then, we used this result to verify our conjecture of no conditional independence in the path model for a specific subset of statements $i \perp\!\!\!\perp j \mid S$, namely for all such statements where less than 100 of the nodes between i and j are conditioning variables.

We achieved the first step, given a suitable counterexample for a corresponding statement on a shorter path, by constructing independent nodes at the beginning and end of the path and highly correlated nodes in between. The independent nodes were constructed by extending the drift matrices with -1 on the diagonal and zero everywhere else. We consecutively constructed highly correlated nodes by letting the newly inserted diagonal entry of the drift matrix go to $-\infty$ and the corresponding subdiagonal entry to ∞. One caveat was that the resulting covariance matrix is singular and therefore not positive definite. Thus, a positive definite counterexample had to be chosen in a suitable neighborhood of this singular matrix. The second result was verified by an additional calculation with the computer algebra system Mathematica.

Lastly, we discussed the challenges arising on the way to proving that $1 \perp\!\!\!\perp p \mid 2, \dots, p-1$ does not hold in the model $\mathcal{M}_{G_p,C_p}$. First, we illustrated the difficulty of a straightforward computation of the determinant of interest. Then, we introduced an approach to construct the nodes $2, \dots, p-1$ as highly correlated nodes via a sequence of matrices. Due to the singularity of the involved matrices, the conditional covariance has to be computed explicitly. The formula contains the generalized inverse of a singular submatrix of the covariance matrix Σ. The main challenge here is that the generalized inverse is in general not continuous. However, in this specific application, simulations suggested that the product of the generalized inverse with two subvectors of Σ converges, given that the sequence of the matrices converges. By diagonalizing the matrices, we were able to determine a specific term for each of the eigenvalues going to zero, that would need to go to zero as well for the full product to converge.

One immediate potential goal for future work is proving that $1 \perp\!\!\!\perp p \mid 2, \ldots, p-1$ does not hold in the Lyapunov model of the directed path of length p. One way to achieve this might be to follow the strategy laid out in Chapter 5 and verify the convergence of the term that we specified there. Another approach might be to explicitly compute the error in the sequences and argue via the convergence rate of the matrix and its eigenvectors. This result together with Proposition 4.21 would fully prove the conjecture that no conditional independencies hold in the Lyapunov model of the directed path.

Recall that our overarching conjecture is that two nodes i and j, that are connected by a trek in the graph are never conditionally independent in the Lyapunov model given any set of conditioning nodes. Thus, a natural next step is to consider treks themselves, as one is illustrated in Figure 8. We presume that many of the ideas and strategies for extending counterexamples on shorter paths to longer paths we employed in this thesis can be extended for treks, albeit possibly in a modified version.

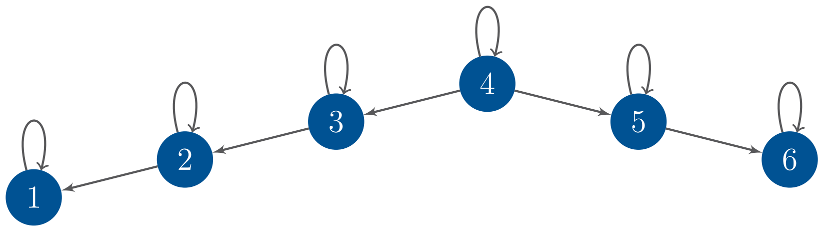


Figure 8: Example of a trek on $p = 6$ nodes.

Proving the conjecture for all treks then might allow us to verify that the only conditional independencies that hold in the Lyapunov model of an arbitrary directed graph are marginal independencies of variables that are not connected by a trek in the graph, and the conditional independencies implied by them.

References

Ludwig Arnold. *Stochastic differential equations: theory and applications.* John Wiley & Sons, New York-London-Sydney, 1974.

Adi Ben-Israel and Thomas N. E. Greville. *Generalized inverses: theory and applications*, volume 15 of *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC.* Springer Science & Business Media, second edition, 2003.

Rajendra Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics.* Springer New York, 1997.

Amit Bhaya, Eugenius Kaszkurewicz, and R. Santos. Characterizations of classes of stable matrices. *Linear Algebra Appl.*, 374:159–174, 2003.

Nicolas Borghini. Topics in nonequilibrium physics, 2012.

Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge University Press, Cambridge, 2004.

John B. Conway. *Functions of one complex variable. II*, volume 159 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1995.

Philipp Dettling, Mathias Drton, and Mladen Kolar. On the lasso for graphical continuous Lyapunov models. *arXiv preprint arXiv:2208.13572*, 2022a.

Philipp Dettling, Roser Homs, Carlos Améndola, Mathias Drton, and Niels Richard Hansen. Identifiability in continuous Lyapunov models. *arXiv preprint arXiv:2209.03835*, 2022b.

Mathias Drton. Algebraic problems in structural equation modeling. In *The 50th anniversary of Gröbner bases*, volume 77 of *Adv. Stud. Pure Math.*, pages 35–86. Math. Soc. Japan, Tokyo, 2018.

Mathias Drton and Marloes Maathuis. Structure learning in graphical modeling. *Annual Review of Statistics and Its Application*, 4:365–393, 2017.

David Edwards. *Introduction to graphical modelling.* Springer Science & Business Media, 2012.

Katherine Fitch. Learning directed graphical models from Gaussian data. *arXiv preprint arXiv:1906.08050*, 2019.

Eberhard Freitag and Rolf Busam. *Complex analysis.* Universitext. Springer-Verlag, Berlin, second edition, 2009.

Crispin W. Gardiner. *Handbook of stochastic methods*, volume 13 of *Springer Series in Synergetics.* Springer-Verlag, Berlin, second edition, 1985.

Søren Højsgaard, David Edwards, and Steffen Lauritzen. *Graphical models with R.* Springer, New York, 2012.

Antti Hyttinen, Frederick Eberhardt, and Patrik O. Hoyer. Learning linear cyclic causal models with latent variables. *The Journal of Machine Learning Research*, 13(1):3387–3439, 2012.

Samuel Karlin and Howard M. Taylor. *A first course in stochastic processes.* Academic Press, New York-London, second edition, 1975.

Samuel Karlin and Howard M. Taylor. *A second course in stochastic processes.* Academic Press, New York-London, 1981.

Tosio Kato. *Perturbation theory for linear operators.* Classics in Mathematics. Springer, Berlin, 1995.

Achim Klenke. *Probability theory: a comprehensive course.* Universitext. Springer, London, second edition, 2014.

Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques.* Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2009.

Steven G. Krantz and Harold R. Parks. *A primer of real analytic functions.* Birkhäuser Advanced Texts: Basel Textbooks. Birkhäuser Boston, Inc., Boston, MA, second edition, 2002.

Steffen L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series.* The Clarendon Press, Oxford University Press, New York, 1996. Oxford Science Publications.

Gregory F. Lawler. *Introduction to stochastic processes.* Chapman & Hall/CRC, Boca Raton, FL, second edition, 2006.

Marloes Maathuis, Mathias Drton, Steffen Lauritzen, and Martin Wainwright, editors. *Handbook of graphical models.* Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL, 2019.

Emanuel Parzen. *Stochastic processes*, volume 24 of *Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

Judea Pearl. *Causality.* Cambridge University Press, Cambridge, second edition, 2009.

C. Radhakrishna Rao. *Linear statistical inference and its applications.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-London-Sydney, second edition, 1973.

Cosma Shalizi. *Advanced data analysis from an elementary point of view.* Citeseer, 2013.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, prediction, and search.* Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, second edition, 2000.

G. W. Stewart. On the continuity of the generalized inverse. *SIAM J. Appl. Math.*, 17: 33–45, 1969.

Gherardo Varando and Niels Richard Hansen. Graphical continuous Lyapunov models. In *Conference on Uncertainty in Artificial Intelligence*, pages 989–998. PMLR, 2020.

Pat Vatiwutipong and Nattakorn Phewchean. Alternative way to derive the distribution of the multivariate Ornstein-Uhlenbeck process. *Advances in Difference Equations*, 2019(1):1–7, 2019.

Wolfram Research, Inc. Mathematica, Version 13.3. URL https://www.wolfram.com/mathematica. Champaign, IL, 2023.

William Chad Young, Ka Yee Yeung, and Adrian E. Raftery. Identifying dynamical time series model parameters from equilibrium samples, with application to gene regulatory networks. *Stat. Model.*, 19(4):444–465, 2019.

Fuzhen Zhang. *The Schur complement and its applications*. Springer New York, NY, 2006.