Cecilie Olesen Recke

Identifiability in Statistical Models via Algebraic Geometry

PHD THESIS

this thesis has been submitted to the phd school of
the faculty of science, University of Copenhagen

Department of Mathematical Sciences
University of Copenhagen

March 2026

Cecilie Olesen Recke
cor@math.ku.dk
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
2100 Copenhagen
Denmark

Thesis title:	Identifiability in Statistical Models via Algebraic Geometry
Supervisor:	Professor Niels Richard Hansen University of Copenhagen
Assessment Committee:	Professor Carsten Wiuf (chair) University of Copenhagen
	Associate Professor Liam Solus KTH Royal Institute of Technology
	Associate Professor Kaie Kubjas Aalto University
Date of Submission:	March 31, 2026
Date of Defense:	June 8th, 2026
ISBN:	978-87-7629-223-2

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen on 31 March 2026. It was supported by the Novo Nordisk Foundation (research grant NNF20OC0062897).

Acknowledgments

First and foremost, I would like to thank Niels for believing in the idea of me doing first a master's thesis and then a PhD in this cross field of algebraic geometry and statistics, I have really enjoyed our collaboration in this area. I would also like to thank Niels for his seemingly endless enthusiasm to discuss all areas of mathematics and his advice and guidance especially by helping me not undersell my accomplishments.

I am also very grateful to Elisenda for introducing me to the wonderful world of applied algebraic geometry during my master's and for continuing to invite me along to all your activities in the applied algebra group.

I want to thank Mathias and the entire group in Munich for hosting me for four months. I had a great time and enjoyed all the great discussions and positive environment at the office. In particular, I would like to thank Sarah for being both my friend and collaborator and showing me around Munich on the weekends, I look forward to returning the favor when you come to Copenhagen soon. I am also grateful to Sarah for providing feedback on the introduction to this thesis.

I am very grateful to both Kaie and Elina for proposing the interesting projects for the apprenticeship week and the Oxford workshop, respectively which have resulted in two of the projects featured in this thesis. I would also like to thank all of my co-authors: Alex, Elina, Jane, Janike, Jiayi, Nataliia, Niels, May, Sarah and Thomas.

Doing a PhD at the Department of Mathematical Sciences has been a great experience, most of all because of all the wonderful and interesting people there, many of which I am happy to call my friends. I am truly grateful to all the wonderful people at the department. Thank you for always making the offices on the third floor a great place to be: Alex C., Alex M., Anton, Christian, Christine, Ella, Emma, Francesco, Frederik, Gabriel, Harry, Jeff, Jin, Matt, Myrto, Nena, Pedja, Phillip, Rasmus, Silvan, Tessa. A special thank you to the original chocolate club: Margherita and Ulises.

I am also fortunate to be surrounded by great friends not necessarily at the department who have brought me much joy the past three years. An extra thank you to my fellow PhD students Jonathan, Marius, Oscar and Tim for always being up for ice cream and talking about the maze that is academia.

Lastly, I would like to thank my entire family. This PhD would not have been possible without all of your endless encouragement throughout my life. I appreciate your willingness to listen to me talk endlessly about all of my specialized interests (both mathematical and non-mathematical) and cheering me on, even if you didn't understand what I was talking about.

Cecilie Olesen Recke
March, 2026

Abstract

If we are only given partial information about a probability distribution how much about said distribution or its generating process can we recover? This broad question of identifiability will be posed and answered in different ways throughout this thesis using tools from algebraic geometry. The first chapter serves as an introduction to the different types of identifiability in the thesis and how algebraic geometry can be used to answer such questions.

The second chapter introduces the graphical discrete Lyapunov models which arise as the steady-state distributions of first-order vector autoregressive models. It is a model of cross-sectional observations from a dynamical system. A directed graph encodes the sparsity pattern of the parameter matrix. Under the assumption of non-Gaussian error terms the tensor equations for the higher-order cumulants are derived, the so-called discrete Lyapunov equations. In this framework we prove generic identifiability of the parameter matrix for any directed acyclic graph with all self-loops and no isolated nodes and local identifiability for all directed graphs with all self-loops and no isolated nodes.

The third chapter introduces a different model of cross-sectional observations from a dynamical system, the continuous Lyapunov model. They arise as the steady-state distributions of a stochastic differential equation with a linear drift matrix whose sparsity pattern is encoded by a directed graph. We derive the higher-order cumulant tensor equations for the steady-state, a generalization of the continuous covariance Lyapunov equation. We prove generic identifiability of the drift matrix for any connected graph with all self-loops. We propose a new semiparametric estimator of the drift matrix and derive its asymptotic distribution by utilizing the identifiability result.

The fourth chapter is on-going work, where we propose an observation noise on top of the steady-state distributions in the previous chapter. First we discuss how to give the models a causal interpretation and then show how the identifiability result from Chapter 3 translates to this set-up when the observation noise is Poisson.

The fifth chapter considers a different type of identifiability than the parameter identifiability discussed in the previous chapters. Instead we consider the completion problem for log-linear models. So we assume that we only observe some coordinates of a discrete probability distribution and discuss when it can completed to a full probability distribution in a given log-linear model. In the case where there are finitely many completions we show that there is either a one or two completions to the log-linear model.

Sammenfatning

Hvis vi kun har adgang til delvis information om en sandsynlighedsfordeling, hvor meget information om denne fordeling eller dens underliggende genererende proces kan vi så genfinde? Dette brede identifikationsspørgsmål vil blive stillet og besvaret på forskellige måder i denne afhandling ved brug af redskaber fra algebraisk geometri. Det første kapitel fungerer som en introduktion til de forskellige typer af identificerbarhed som studeres i denne afhandling og hvordan algebraisk geometri kan bruges til at besvare sådanne spørgsmål.

I andet kapitel introducerer vi den grafiske diskrete Lyapunov model, der opstår som stationære fordelinger af førsteordens vektor-autoregressive modeller. Det er en model af tværsnitsobservationer fra et dynamisk system. En orienteret graf angiver mønstret af nulindgange i parametermatricen. Under antagelse af, at fejlledene ikke er normalfordelte, udleder vi tensorligninger, der beskriver højere ordens kumulanterne af fordelingerne, de såkaldte diskrete Lyapunov ligninger. Ved hjælp af disse ligninger beviser vi generisk identificerbarhed af parametermatricen for en vilkårlig orienteret acyklisk graf med alle løkker og ingen isolerede knuder og lokal identificerbarhed for alle orienterede grafer med alle løkker og ingen isolerede knuder.

Tredje kapitel introducerer en anden model af tværsnitsobservationer fra et dynamisk system, den kontinuerte Lyapunov model. Den opstår som stationære fordelinger af en stokastisk differential ligning med lineært driftled, hvis mønster af nulindgange bliver angivet af en orienteret graf. Vi udleder tensorligninger, der beskriver højere ordens kumulanterne af fordelingerne. De er en generalisering af den kontinuerte Lyapunov ligning for kovariansen. Vi beviser generisk identificerbarhed af driftmatricen for an vilkårlig sammenhængende graf med alle løkker. Vi foreslår en ny semiparametrisk estimator af driftmatricen og udleder dens asymptotiske fordeling ved brug af identificerbarhedsresultatet.

Det fjerde kapitel er igangværende arbejde, hvor vi foreslår observationsstøj ovenpå de stationære sandsynlighedsfordelinger i forrige kapitel. Først diskuterer vi, hvordan disse modeller kan gives en kausal fortolkning og viser derefter, hvordan identifikationsresultatet fra kapitel 3 kan overføres til modellen med Poisson støj.

Det femte kapitel betragter et andet type identifikationsproblem end de tidligere kapitler. Vi betragter i stedet fuldstændiggørelsesproblemet for log-lineære modeller. Vi antager, at vi kun observerer enkelte koordinater fra en diskret sandsynlighedsfordeling og undersøger, hvornår sådan en partiel observation kan fuldstændiggøres til en fuld sandsynlighedsfordeling i en given log-lineær model. Hvis der er et endeligt antal fuldstændiggørelser, viser vi, at der er en eller to fuldstændiggørelser til den givne log-lineære model.

Contributions and Structure

Chapter 1 is an introduction which will provide an overview of the concepts which will be present throughout the thesis. In it we will introduce and discuss what identifiability is, the idea behind it and how algebraic geometry tools can be used to determine it. The following four chapters each represent a paper. For the three papers already available online, there are minor changes to formatting, typo corrections and references have been updated if a preprint has been published.

Chapter 2 defines the graphical discrete Lyapunov models and proves generic and local identifiability results using higher-order cumulants. It corresponds to the following paper:

[**IDLM**] [Recke et al., 2026]. C. O. Recke, S. Lumpp, N. Kushnerchuk, J. Oldekop, J. Li, J. I. Coons, and E. Robeva. Identifiability in graphical discrete Lyapunov models. *arXiv: 2601.21818*, 2026.

Chapter 3 defines the continuous Lyapunov model and proves generic identifiability up to scaling using the second-order and one higher-order cumulant. It also introduces the singular value estimator and derives its asymptotic distribution by using the identifiability result. Everything except the last Appendix 3.C corresponds to the paper:

[**ICLM**] [Recke and Hansen, 2026]. C. O. Recke and N. R. Hansen. Identifiability and estimation in continuous Lyapunov models. *arXiv: 2603.17142*, 2026.

Chapter 4 gives the model from the previous chapter a causal interpretation and introduces an observation noise model. It corresponds to the work in progress:

[**LOSD**] [Markham et al., 2026+]. A. Markham, C. O. Recke, and N. R. Hansen. Causal inference with latent operator-selfdecomposable distributions. *Working paper*, 2026+.

Chapter 5 studies the problem of completing a partial observation of a discrete probability distribution to a log-linear model and corresponds to the following paper:

[**CLLM**] [Cai et al., 2024]. M. Cai, C. O. Recke, and T. Yahl. Completions to discrete probability distributions in log-linear models. *Algebraic Statistics*, 15(2):225–247, Dec. 2024. ISSN 2693-2997. doi: 10.2140/astat.2024.15.225.

Contents

1 Introduction

A common theme throughout all projects in this thesis is the idea of identifiability if we think of it more broadly as: if we observe *some* information about a probability distribution how much information about the entire distribution or its generating process can we recover? We study a version of this question for different statistical models, both discrete and continuous. The second unifying idea is using tools from algebraic geometry to answer such questions. We first provide an example to illustrate the two different identifiability problems explored in the thesis. We then focus on parameter identifiability and use this as the main example of how to use tools from algebraic geometry to answer questions in statistics. Lastly, we introduce and compare the different Lyapunov models which are the focus points of three out of four chapters in the remainder of the thesis.

1.1 A Tale of Two Identification Problems

We consider the independence model for two discrete random variables and how it can yield two different identification problems see e.g. Drton et al. [2009], Sullivant [2018] for background.

Example 1.1.1. Let X_1 and X_2 be two discrete random variables with state spaces $[n_i] = \{1, \dots, n_i\}$ for $n_i \in \mathbb{N}$ and $i = 1, 2$. Then X_1 and X_2 are independent if their joint probability factors into the marginals,

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2) \tag{1}$$

for all $x_1 \in [n_1]$ and $x_2 \in [n_2]$. A consequence of this is that the matrix with entries equal to the joint probability distribution $p_{ij} = P(X_1 = i, X_2 = j)$ for $i = 1, \dots, n_1$ and $j = 1, \dots n_j$ has rank equal to one if and only if it comes from the independence model [Drton et al., 2009, Proposition 1.1.2]. We let $\Delta_{k-1} = \{(p_1, \dots, p_k) \mid p_i \geq 0, \sum_{i=1}^{k} p_i = 1\}$ denote the probability simplex, the set of all possible distributions for a variable with k possible outcomes. Then the definition of the discrete independence model can also be written as exactly the set of matrices $p = (p_{ij})$ which have rank 1

$$\mathcal{M}_{X_1 \perp\!\!\!\perp X_2} = \{p \in \Delta_{n_1 n_2 - 1} \mid \text{rank(p)} = 1\}.$$

The above is an implicit description of the independence model, but it can also be parametrized using (1) as

$$\phi : \Delta_{n_1 - 1} \times \Delta_{n_2 - 1} \to \Delta_{n_1 n_2 - 1}, \quad (\alpha, \beta) \mapsto \alpha\beta^T,$$

with α and β representing the marginal probability distributions, so $\text{im}(\phi) = \mathcal{M}_{X_1 \perp\!\!\!\perp X_2}$. A natural question to ask now is: if we only know the joint probabilities $p \in \mathcal{M}_{X_1 \perp\!\!\!\perp X_2}$ can we recover the marginal probabilities, α and β? In this case, the answer is yes. It is easily seen that we can sum the rows or columns of p accordingly to obtain the marginals

$$\alpha_i = \sum_{j=1}^{n_2} p_{ij}, \quad \beta_j = \sum_{i=1}^{n_1} p_{ij}.$$

Another way to phrase the question above is whether the map ϕ from the set of parameters (the marginals) to the statistical model (the joint distribution) is injective. Injectivity of the map parameterizing a statistical model is what we generally mean by *parameter identifiability*, see Definitions 1.2.1 and 1.2.5.

A different, and as it will turn out much harder, problem we could ask for the discrete independence model is the completion problem. That is whether if we know some of the entries of the joint distribution $p = (p_{ij})$ but not all of them can we recover the rest? For the independence model this is treated by Kubjas and Rosen [2017] by considering completions of partially observed matrices to rank one matrices in the probability simplex. Let S denote the set of observed indices of p and let

$$\pi_S : \mathcal{M}_{X_1 \perp\!\!\!\perp X_2} \to \mathbb{R}^{|S|}$$

denote the projection from the independence model to the coordinates indexed by S. A *completion* of a partially observed matrix of probabilities $p_S \in \mathbb{R}^{|S|}$ is any element in the fiber of p_S,

$$\pi_S^{-1}(p_S) = \{p \in \mathcal{M}_{X_1 \perp\!\!\!\perp X_2} \mid \pi_S^{-1}(p_S) = p\}.$$

The following three questions are the primary questions to ask about completions. For what patterns of observed entries S does a completion exist? And what additional semi-algebraic constraints need to be satisfied for p_S to be in the *completable region*, $\pi_S(\mathcal{M}_{X_1 \perp\!\!\!\perp X_2})$? For example all the coordinates obviously have to be between 0 and 1 and no row or column can sum to more than 1, so this will result in semi-algebraic constraints. Lastly, given that a completion exists, is it unique? This is a question of injectivity of π_S by characterizing the size of the fibers of a partially observed matrix $p_S \in \mathbb{R}^{|S|}$,

In Kubjas and Rosen [2017] these questions are answered for the independence model. Firstly, they associate the sparsity pattern S with a bipartite graph and show that if all the observed entries are positive then the matrix is uniquely completable to the independence model if and only if the graph is connected. They also provide an algorithm that runs in polynomial time to decide completability of an observation p_S also allowing for zero entries by providing equality and inequality constraints that have to be satisfied.

They also consider the size of the fibers of π_S and provide a characterization of the size of the completable region divided into cases of when it is a single point, two points or providing the dimension of the semi-algebraic set when there are infinitely many completions. Thus, even though we refrain from writing all the exact conditions and algorithms from Kubjas and Rosen [2017] it is clear that for the independence model the completion problem is much more involved than the parameter identifiability problem.

Example 1.1.1 showcases two different areas of identification theory: parameter identifiability and completion of discrete probability distributions. In this thesis we will explore these topics in the context of the followings papers. The paper [**CLLM**] explores the completion problem for a generalization of the independence model, namely log-linear models, which will be defined in Chapter 5. The papers [**ICLM**] and [**IDLM**] prove results about parameter identifiability for the models introduced in Section 1.4.

The completion problem as stated here is only well-defined in the case of discrete probability distributions whereas parameter identifiability can be asked for any parametrized statistical model. For this reason we will now focus the discussion on parameter identifiability. However, since both cases are essentially a question of injectivity of certain maps, several of the ideas, for example genericity and finite-to-one, can just as well be applied to the map in the completion question. The main difference is that for the completion problem the first problem is determining for which subsets of the domain the fibers are non-empty, which is typically a question answered directly by definition in the case of parameter identifiability.

1.2 Parameter Identifiability

In one of the early papers explicitly studying (parameter) identification, Koopmans and Reiersøl [1950], they write

> "The study of identifiability is undertaken in order to explore the limitations of statistical inference."

Koopmans and Reiersøl write this to motivate the study of identifiability and explain its relevance for statistical inference. The idea is that if a statistical model is not identifiable, that is that we cannot even identify the parameters of interest when we have access to the true probability distribution, it would be unrealistic to expect to be able to accurately estimate the parameters given data. In this way identifiability is a crucial first question to answer about a statistical model before trying to estimate the parameters.

Given Example 1.1.1 it might seem like parameter identifiability will always be a simple question to answer. As we shall see, this will *not* be the case. As with the completion problem, identifiability might only be feasible for a subset of the parameter set and maybe not at all. See for example [Sullivant, 2018, Ch. 16], [Bollen, 1989, Ch. 4] for a discussion.

In this section we explore and discuss different notions of identifiability. First we briefly introduce the modern definition of global identifiability, Definition 1.2.1. By giving a historical context and considering some of the first models to prompt the study of identifiability we explore examples wherein only weaker notions of identifiability can be obtained. Ultimately, this leads to several definitions of weaker forms of identifiability (Definition 1.2.5).

The strongest notion of parameter identifiability, as satisfied by the independence model in Example 1.1.1, is that the map parameterizing the statistical model is injective on the entire parameter space.

To formally define this we let a statistical model, $\mathcal{M}$, be a a family of probability measures on a measure space, or simply just a set of probability distributions. A *parametrized statistical model*, $\mathcal{M}_\nu$ is defined as the image of a map $\nu : \Theta \to \mathcal{M}_\nu$ mapping the parameter space Θ to a set of probability distributions

$$\mathcal{M}_\nu = \{\nu(\theta) \mid \theta \in \Theta\}.$$

Definition 1.2.1. *A parametrized statistical model $\mathcal{M}_\nu$ is said to be* (globally) identifiable *if the map $\nu : \Theta \to \mathcal{M}_\nu$ is injective.*

In the majority of this thesis we will identify the parameters from the covariance or some higher-order cumulants of the distributions. In this setting, the set $\mathcal{M}_\nu$ will not directly be a set of distributions but a set of cumulants satisfying a set of equations which have to be satisfied for the set of distributions. Of course, identifiability from the cumulants will imply identifiability from the corresponding distributions.

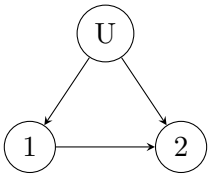
Historically, the first uses of the word identification to describe Definition 1.2.1 are given in the works of Koopmans and Reiersøl around 1950 [Koopmans, 1949; Koopmans and Reiersøl, 1950; Reiersøl, 1950]. What is common across all three papers is that they study a system of (typically linear) equations in a set of stochastic variables, where some are assumed observed and some are assumed unobserved. In this way they study identifiability of what we would now mainly classify as linear structural equation models both with independent and non-independent errors.

Furthermore, in all three papers they give a brief survey of the history of the study of identifiability up until this point. They note that most of the study so far has been carried out with different fields of application in mind. Some notable and early contributions, albeit not a complete overview of the history, include the method of path coefficients by Wright [1934] essentially studying identification in linear structural equation models using treks and Thomson [1919] who demonstrates the lack of identifiability in some factor analysis models by numerical examples.

The identifiability of all of these somewhat similar models, namely linear structural equation models, factor analysis and principal components analysis have been the subject of much study and is still an active area of research. See for example Brito and Pearl [2002]; Foygel et al. [2012b]; Barber et al. [2022] for linear structural equation models, Sturma et al. [2025] for factor analysis models and Comon [1994]; Mesters and Zwiernik [2024] for components analysis models.

To illustrate different types of identifiability we will consider a linear structural equation model on two observed variables with latent confounding, see Figure 1 and Example 1.2.2. This is a restructuring of Example 1.2 in Koopmans and Reiersøl [1950] to the standard way of writing a linear structural equation model where all observed variables can be written as a linear combination of the other observed variables and error terms. Koopmans and Reiersøl [1950] instead assume a linear relationship between unobserved true variables and add a layer of measurement noise. It is possible to move between the two set-ups by redefining what is seen as errors.

In a linear structural equation model the linear effects are encoded by a directed graph, so a directed edge $i \to j$ corresponds to there being a linear effect, λ_{ij}, of the parent X_i



(a) Directed graph with a latent confounder U.

(b) A mixed graph representing non-zero correlation between the errors.

Figure 1: Two different graphical representations of confounding of two variables.

on the child X_j, see Example 1.2.2. See for example Bollen [1989] and Drton [2018] for background on linear structural equation models.

Example 1.2.2. We consider the linear structural equation model on two observed variables X_1 and X_2 and a single unobserved (or latent) variable U. The graph in Figure 1a corresponds to the following system of linear equations

$$U = \varepsilon_u, \qquad X_1 = \lambda_{u1}U + \varepsilon_1, \qquad X_2 = \lambda_{12}X_1 + \lambda_{u2}U + \varepsilon_2. \tag{2}$$

where the assumption is that $\varepsilon_u, \varepsilon_1$ and ε_2 are all independent and Gaussian with mean 0. The parameter of interest to identify is λ_{12}, the effect of X_1 on X_2.

If the latent variable, U, is not of specific interest the graph in Figure 1a is often projected to the mixed graph in Figure 1b. To do this we redefine the errors associated to X_1 and X_2 by $\epsilon_i = \varepsilon_i + \lambda_{ui}\varepsilon_u$ for $i = 1, 2$ [Drton, 2018]. This results in the errors being correlated, which is exactly what a bidirected edge between two variables encodes in a mixed graph. In this set-up equation (2) will instead be represented as

$$X_1 = \epsilon_1, \qquad X_2 = \lambda_{12}X_1 + \epsilon_2. \tag{3}$$

However, in this representation the correlation between the errors could result from a more complicated latent structure as the latent structure is not specified. In this parametrization we explicitly have four parameters, λ_{12} and the unknown covariance matrix of ϵ_1 and ϵ_2.

Irrespective of the parametrization, the observed variables X_1 and X_2 follow a Gaussian distribution with mean 0. Thus, being able to identify the parameters from the distribution of (X_1, X_2) corresponds to identifying them from the covariance matrix of X_1 and X_2:

$$\begin{aligned} \text{var}(X_1) &= \text{var}(\epsilon_1) \\ \text{cov}(X_1, X_2) &= \lambda_{12}\text{var}(\epsilon_1) + \text{cov}(\epsilon_1, \epsilon_2) \\ \text{var}(X_2) &= \lambda_{12}^2\text{var}(\epsilon_1) + \text{var}(\epsilon_2). \end{aligned}$$

We only have three equations and four unknown parameters and therefore we conclude that the model is not identifiable, as per Definition 1.2.5 (iv) introduced below. Furthermore, even just the parameter of interest, λ_{12}, is non-identifiable from the covariance.

Example 1.2.2 illustrates one of the simplest ways for parameter identifiability to fail, that there are more unknown parameters to identify than equations to identify them from. Small variations of the assumptions made in Example 1.2.2 can yield identifiability.

Example 1.2.3 (Continuation of Example 1.2.2)**.** We consider two different modifications to the assumptions in Example 1.2.2 which result in the parameter of interest λ_{12} being identifiable.

If we assumed there is no confounding, i.e., no U variable or mixed edge, so $\epsilon_1 \perp\!\!\!\perp \epsilon_2$, we are just in a classic linear regression set-up and we can recover the coefficient as

$$\lambda_{12} = \frac{\text{cov}(X_1, X_2)}{\text{var}(X_1)}.$$

The only condition is that $\text{var}(X_1)$ is non-zero, which is just an assumption of the distribution being non-degenerate, i.e., non-constant. We can use this to recover the variance of the errors ϵ as well. Thus, the model is (globally) identifiable.

Another option, also explored in Koopmans and Reiersøl [1950], where there is still allowed to be confounding is to introduce some non-Gaussianity. Specifically, what they do in Koopmans and Reiersøl [1950] is to assume that ε_u and ε_2 are Gaussian but that ε_1 is non-Gaussian. This results in both X_1 and X_2 being non-Gaussian. In this case it is possible to utilize a higher-order cumulant of X to identify λ_{12}. Cumulants are the higher-order version of the covariance, see McCullagh [2018] for background on cumulants. The Gaussian distribution is the only distribution that has all cumulants of order three and above equal to zero. Non-Gaussianity will turn out to be a useful assumption to obtain identifiability where it is not possible from just the covariance, see Remark 1.2.6 and the three projects [**ICLM**], [**IDLM**] and [**LOSD**].

Since ε_1 is non-Gaussian at least one cumulant of order at least three is non-zero. For notational purposes, we assume that the third-order cumulant κ_3 is non-zero. Then

$$\kappa_3(X_1, X_1, X_2) = \kappa_3(\varepsilon_1 + \lambda_{u1}\varepsilon_u, \varepsilon_1 + \lambda_{u1}\varepsilon_u, \lambda_{12}(\varepsilon_1 + \lambda_{u1}\varepsilon_u) + \varepsilon_2) = \lambda_{12}\kappa_3(\varepsilon_1, \varepsilon_1, \varepsilon_1),$$

by linearity of cumulants and that the third-order cumulant of two independent stochastic variables or any combination of only Gaussian stochastic variables is 0. Similarly, by the linearity and that ε_u is Gaussian we obtain $\kappa_3(X_1, X_1, X_1) = \kappa_3(\varepsilon_1, \varepsilon_1, \varepsilon_1)$. Thus, we can recover the parameter of interest

$$\lambda_{12} = \frac{\kappa_3(X_1, X_1, X_2)}{\kappa_3(X_1, X_1, X_1)}, \tag{4}$$

and then recover the covariance of ϵ, but not the latent edges and covariances.

Example 1.2.4. We consider a modification of Example 1.2.2 where we have one more observed variable X_0 with an incoming edge to X_1 with no latent confounding with the other variables, see Figure 2. This is known as the instrumental variable set-up discussed at length in the literature [Bowden et al., 1990]. It corresponds to changing the equation for X_1 in (2) and (3) to

$$X_1 = \lambda_{01}X_0 + \lambda_{u1}U + \varepsilon_1, \qquad X_1 = \lambda_{01}X_0 + \epsilon_1,$$

(a) Directed graph representing an instrumental variable and latent confounder U.

(b) A mixed graph representing an instrumental variable model.

Figure 2: Two different graphical representations of an instrumental variable model.

respectively and adding the additional variable $X_0 = \varepsilon_0$ (or $X_0 = \epsilon_0$ in (3)).

By assuming we have an extra observed variable, we obtain 6 distinct entries in the covariance matrix of the observed variables, X_0, X_1 and X_2. If we consider the mixed graph we also have exactly six unknowns in the form of two edges between observed nodes and the variance of each ϵ_i and $\text{cov}(\epsilon_1, \epsilon_2)$. Thus, we are no longer in the case of Example 1.2.2 where there are more unknowns than parameters. Furthermore, if we write out the expression for the covariance matrix of (X_0, X_1, X_2) we obtain the solution

$$\lambda_{01} = \frac{\text{cov}(\text{X}_0, \text{X}_1)}{\text{var}(X_0)}, \qquad \lambda_{12} = \frac{\text{cov}(X_0, X_2)}{\text{cov}(X_0, X_1)} \tag{5}$$

for the parameters of interest. Thereafter, we can find the covariance of the errors ϵ as in the previous example. However, there is one subtlety here compared to Example 1.2.3 and Example 1.1.1. The solution for λ_{12} only exists if λ_{01} is non-zero. Heuristically, this makes sense because if λ_{01} was zero the model would be the same as Example 1.2.2 where λ_{12} was not identifiable. This model is therefore not globally identifiable, but what we will call generically identifiable and furthermore rationally identifiable (see Definition 1.2.5), where we only have identifiability for a specific subset of the set of parameters.

With these different examples in mind we introduce the following definition of different types of identifiability for parametrized statistical models. They are ordered by decreasing strength starting with globally identifiable and ending with non-identifiable. This is Definition 2.2.5 from [**IDLM**] with the addition of point (ii) and the removal of the assumption of the parametrization being rational.

Definition 1.2.5. *Let $f : \Theta \to \mathcal{M}_f$ be a map from a finite dimensional parameter space $\Theta \subseteq \mathbb{R}^k$ to a statistical model such that $im(f) = \mathcal{M}_f$. The model is said to be*

(i) globally identifiable if $f^{-1}(f(\theta)) = \theta$ for all $\theta \in \Theta$;

(ii) rationally identifiable if $f^{-1}(f(\theta)) = \theta$ for all $\theta \in \Theta \setminus A$ where A is a proper algebraic subset of Θ with f and f^{-1} both rational maps;

(iii) generically identifiable if $f^{-1}(f(\theta)) = \theta$ for all $\theta \in \Theta \setminus A$ where A is a proper algebraic subset of Θ;

(iv) locally identifiable if $|f^{-1}(f(\theta))| < \infty$ for all $\theta \in \Theta \setminus A$ where A is a proper algebraic subset of Θ;

(v) non-identifiable if $|f^{-1}(f(\theta))| = \infty$ *for generic* θ*.*

Note that for the second, third and fourth notion of identifiability the Zariski closure $\overline{\Theta}$ *of* Θ *is assumed to be an irreducible variety.*

Note that different notions of identifiability have different names in the literature and some of the adjectives are also given different meanings. Definitions *(i), (ii), (iv)* and *(v)* are almost always given by the definitions above. Locally identifiable is also sometimes known as *finite-to-one* or *algebraically* identifiable when f is a rational map. Even though it seems misleading (given the usefulness of generic identifiability) non-identifiable is sometimes used to denote anything which is not globally identifiable.

It is mainly the notion of *generic* identifiability which is used in many different ways. It is always a notion of the parametrization being injective on the entire parameter set except some exceptional subset A. In some instances A is simply assumed to be a null set with respect to the Lebesgue measure. An algebraic subset (the vanishing set of a set of polynomials) is a specific instance of a null set, which is more well-behaved than general null sets. For example an algebraic subset cannot be dense in the ambient space with respect to the standard Euclidean topology, which a general null set does not exclude. Furthermore, in practice it is typically possible to prove the (stronger and more insightful) definition of generic identifiability provided here and often even rational identifiability, see Section 1.3 and the papers [**ICLM**] and [**IDLM**] where we provide proofs of rational and local identifiability. As will be apparent global identifiability is not possible in the models studied in these papers.

Furthermore for *(v)* the notion of 'generic' is purposefully left more vague here than in *(iii)* to just specify that *most* of the parameter space is not-identifiable as would be the case if *(v)* is true for all parameters except a null set.

The reason for assuming that the Zariski closure of the parameter set is irreducible is to ensure that the algebraic subset A is of strictly lower dimension than the parameter set Θ. In practice a lot of parameter sets are open subsets of $\mathbb{R}^k$ which is an instance of a parameter set with irreducible Zariski closure.

Remark 1.2.6 (A Note on Structural Identifiability). Especially, in the context of graphical models, like linear structural equation models, identifiability is sometimes taken to mean what we will call *structural identifiability*. That is if we have a finite set of graphs $G_1, \ldots, G_p$ each defining a linear structural equation model on k observed variables is it possible to distinguish between the models by the observed distributions? This is important in order to be able to learn the directed graph from observed data as is typically the goal in causal discovery [Peters et al., 2017].

For example in the well-known case of two observed variables X_1 and X_2 with a directed edge between them (so Figure 1a without the confounding) the distribution of (X_1, X_2) is the same no matter if the effect goes from X_1 to X_2 or the other way if they follow a joint Gaussian distribution. Thus, they are not structurally identifiable.

However, if the errors are assumed non-Gaussian the distributions for the two different graphs can actually be distinguished. This is a consequence of the Darmois-Skitovich theorem, see Comon and Jutten [2010]. It can also be seen by finding an equation which

has to be true in the moments (or cumulants) for one graph but which is not true in the other graph, see for example Améndola et al. [2023]. In Shimizu et al. [2006] they prove structural identifiability for all directed acyclic graphs with no latent confounding (based on the result of Comon [1994] for independent component analysis) and utilized it for causal discovery, naming the models LiNGAM (Linear Non-Gaussian Acyclic Model).

In [**IDLM**] Section 2.6 we show some first results towards structural identifiability for a certain class of discrete Lyapunov models.

1.3 Why Algebra in Statistics?

The idea of using algebraic geometry tools to answer statistical problems has seen rapid development in the last 30 years, see Sullivant [2018]; Drton et al. [2009]; Casanellas et al. [2019] for an overview of different research directions. Algebraic geometric methods are mostly of relevance if the statistical model can at least in part be parametrized by a rational map. Examples of this includes various discrete probability distributions such as log-linear models as we study in [**CLLM**] and a wide collection of graphical models including linear structural equation models [Drton, 2018] and graphical Lyapunov models for example studied in [**ICLM**] and [**IDLM**] in this thesis.

Once a rational parametrization is given, computational algebra can in principal be utilized to obtain polynomial equality and inequality constraints which must be satisfied in the model. Such polynomials can for example be used to obtain results about conditional independence implications, parameter and structural identifiability and sampling from conditional distributions.

As identifiability is a main focus point of this thesis, we exemplify the application of algebraic geometry to solve statistical problems by showing how to use it to study parameter and structural identifiability of rationally parametrized statistical models. Parameter identifiability using computational algebraic geometry was first discussed in García-Puente et al. [2010], we base the following on that as well as Foygel et al. [2012b] and Barber et al. [2022]. A general background on polynomial ideals and affine varieties used below can for example be found in Cox et al. [2015].

Let $\Theta \subseteq \mathbb{R}^k$ and $f : \Theta \to \mathcal{M}_f \subseteq \mathbb{R}^d$ denote a rational parametrization of a statistical model $\mathcal{M}_f$ given by

$$f(\theta) = \left(\frac{f_1(\theta)}{g_1(\theta)}, \dots, \frac{f_d(\theta)}{g_d(\theta)} \right), \tag{6}$$

where we adopt the convention that $\theta = (\theta_1, \dots, \theta_k)$. We let the coordinates in $\mathbb{R}^d$ be denoted by $p_1, \dots, p_d$.

In order to study (rational) parameter identifiability we need to find polynomials relating the coordinates of the probability distribution, p, and the parameters, θ. Using algebraic geometry we can find what is called the vanishing ideal of the graph of f, $\mathcal{I}_f$ (equation (8)). It is the largest set of polynomials $h(\theta, p) \in \mathbb{R}[\theta, p]$ that vanish on the graph of f,

$$G(f) = \{(\theta, f(\theta)) \mid \theta \in \Theta\} \subseteq \mathbb{R}^k \times \mathbb{R}^d.$$

Or equivalently, the vanishing set of $\mathcal{I}_f$ is the smallest affine variety containing the graph of f. Thus, if we cannot use the polynomials in $\mathcal{I}_f$ to obtain a rational expression for θ_i in terms of $p_1, \ldots, p_d$, θ_i is *not* rationally identifiable in the model $\mathcal{M}_f$, see Theorem 1.3.1 below for the precise statement.

To define the vanishing ideal of the graph, we first define the ideal

$$\mathcal{J}_f = \langle g_1(\theta)p_1 - f_1(\theta), \ldots, g_d(\theta)p_d - f_d(\theta), 1 - (g_1 \cdots g_d)y \rangle \tag{7}$$

in the polynomial ring $\mathbb{R}[\theta, p, y]$. It is the smallest ideal containing the polynomials in the brackets. The first d polynomials correspond to $p_i = f_i(\theta)/g_i(\theta)$, so exactly that $(\theta, p) \in G(f)$. The variable y is an extra variable in order to ensure that the denominators are never zero. If we knew that was never the case (for example if f was a polynomial) we could remove the last polynomial in the generating set as it would be trivial and $\mathcal{J}_f$ would already be the vanishing ideal of the graph of f. The vanishing ideal of the graph is simply the elimination ideal of $\mathcal{J}_f$ where the variable y has been eliminated,

$$\mathcal{I}_f = \mathcal{J}_f \cap \mathbb{R}[\theta_1, \ldots, \theta_k, p_1, \ldots, p_d], \tag{8}$$

so it is all the polynomials in $\mathcal{J}_f$ only involving θ and p.

Similarly, to be able to obtain results about structural identifiability we can use the polynomials which characterize the image of f, so if these polynomials are different for two models they are structurally identifiable from each other. Thus, we need to find the vanishing ideal of the image of f, $\mathcal{I}(\mathcal{M}_f)$. It is the largest set of polynomials $h(p) \in \mathbb{R}[p]$ which vanish on the image of f. The vanishing ideal of the image can be obtained from (7) by eliminating both θ and y,

$$\mathcal{I}(\mathrm{im}(f)) = \mathcal{I}(\mathcal{M}_f) = \mathcal{J}_f \cap \mathbb{R}[p_1, \ldots, p_d], \tag{9}$$

as also specified by the implicitization theorem [Cox et al., 2015, p. 138].

The notion of the vanishing ideal of the image, $\mathcal{I}(\mathcal{M}_f)$ is also needed for the precise statement of how to show identifiability for each parameter θ_i as given below, as it is used to formally write that a polynomial $a(p)$ does not vanish on the model $\mathcal{M}_f$.

Theorem 1.3.1. *Let $\Theta \subseteq \mathbb{R}^k$ be a subset such that its Zariski closure $\overline{\Theta}$ is irreducible and let $f : \Theta \to \mathbb{R}^d$ define a rationally parametrized statistical model by (6). Define the ideal $\mathcal{I}_f$ as in equation (8) then for $i \in \{1 \ldots, k\}$, θ_i is rationally identifiable if and only if $\mathcal{I}_f$ contains a polynomial of the form $a(p)\theta_i + b(p)$ with $a \notin \mathcal{I}(\mathcal{M}_f)$.*

The proof of Theorem 1.3.1 follows by essentially the same proof as in Foygel et al. [2012a] for linear structural equation models by applying the same arguments to general rationally parametrized models.

As is explained in García-Puente et al. [2010] Theorem 1.3.1 can easily be generalized to show local identifiability by existence of a polynomial of the form $a(p)\theta_i^l + b(p)$ instead. Of course in this case θ_i might still be generically identifiable if only one of the roots of the polynomial belongs to Θ. Another option to prove local identifiability is by proving that the Jacobian of f has full rank [Sullivant, 2018, Prop. 16.1.7]). This is done in Section 2.5 of [**IDLM**].

Theorem 1.3.1 can be made into an algorithm to determine rational identifiability by obtaining a Gröbner basis of $\mathcal{I}_f$, see Algorithm 1. The proof of correctness of Algorithm 1 follows by essentially the same proof as in Foygel et al. [2012a] for linear structural equation models. The algorithm could start using either $\mathcal{J}_f$ or $\mathcal{I}_f$. It is essentially equivalent, since the elimination ideal $\mathcal{I}_f$ is obtained from $\mathcal{J}_f$ by finding a Gröbner basis with respect to a monomial order where y is greater than the remaining indeterminates. Thus, in practice it can be simpler to do it in one step to only find one Gröbner basis.

Algorithm 1. Algebraic method to check rational identifiability

1 Choose a monomial ordering on the variables θ and p such that for all indices i, j we have $y > \theta_i > p_j$
2 Let $\mathcal{J}_f$ be as given in equation (8) and compute a reduced Gröbner basis T of $\mathcal{J}_f$ with respect to the monomial order from step 1.
3 The model $\mathcal{M}_f$ is rationally identifiable if and only if for each $i = 1, \ldots, k$ the basis T contains an element whose leading term is a monomial in p times θ_i.

Example 1.3.2 (Continuation of Example 1.2.4)**.** We consider the instrumental variable model from Example 1.2.4. Since it is a Gaussian model the model, $\mathcal{M}_f$ is a subset of covariance matrices, so let Σ denote the covariance matrix of (X_0, X_1, X_2). Then the entries of Σ will correspond to the p variables in the definitions above.

Foygel et al. [2012a] present a modified version of Algorithm 1 specialized to linear structural equation models by replacing $\mathcal{J}_f$ with an ideal which includes less equations but aside from that the algorithm is the same. Using this modified algorithm we find that the Gröbner basis of $\mathcal{I}_f$ contains the following two polynomials,

$$\lambda_{12}\Sigma_{01} - \Sigma_{02}, \qquad \lambda_{01}\Sigma_{00} - \Sigma_{01},$$

when picking a lexicographic ordering with $\lambda_{01} > \lambda_{12}$. Clearly, 3) in Algorithm 1 is satisfied so we rediscover that the model is rationally identifiable. Furthermore, since all elements of $\mathcal{I}_f$ vanish on the model we can solve for λ_{12} and λ_{01} in the equations above by setting them equal to zero. In this way we recover exactly the identifying formulas obtained in equation (5).

The Gröbner basis calculation can be carried out using Buchberger's algorithm [Cox et al., 2015, p. 91], however even for small (in terms of dimensions k and d) problems it can be challenging in terms of running times and and can for example depend a lot on the specific monomial order. Because of this it is typically worth it if the algorithm can be tailored to the specific models classes (to reduce the generating set of $\mathcal{J}_f$) [Barber et al., 2022; Foygel et al., 2012a]). It is also still an active area of research to optimize algorithms for computing Gröbner bases, see for example Cummings and Hollering [2026]. Implementations of Gröbner bases algorithms exist in most computer algebra programming languages such as Macaulay2, Oscar, SageMath, Maple and Mathematica.

In practice it is often most useful to use Gröbner bases calculations to explore identifiability for low dimensional models (i.e. a graph on a small number of vertices) to

obtain an idea about what could potentially be proven for arbitrary dimensions. This is how computational algebra was used in both [**ICLM**] and [**IDLM**]. Furthermore, in [**IDLM**] a Jacobian rank computation for all small graphs (≤ 5 nodes) carried out in Maple is used as the base case of the induction proof of Theorem 2.5.6.

In this section the discussion has only focused on the (easier) task of finding polynomial equalities which are true in the model. In practice it is also very relevant to try to find semi-algebraic descriptions which contain both equality and inequality polynomial constraints. However, this is a much more difficult task. In Boege and Solus [2024] they show in the case of a (globally) rationally identifiable model how is it is possible to exploit this to obtain both equality and inequality constraints in the statistical model. In [**CLLM**] we provide semi-algebraic descriptions of the completable region in specific examples.

1.4 Discrete and Continuous Lyapunov Models

In the three projects [**ICLM**], [**LOSD**] and [**IDLM**] we study the so-called Lyapunov models with [**ICLM**] studying the continuous Lyapunov model, [**LOSD**] the continuous one with a measurement noise model, and [**IDLM**] studying the discrete Lyapunov model. As implied by their similar sounding names, the two models share certain characteristics but deviate in other aspects. In this section we provide a short introduction to both models with focus on how the they relate to each other. A summary of the two models is provided by Table 1.

The main unifying idea of the two models is that both the discrete and continuous Lyapunov models arise as distributions of the steady-state solution of a stochastic dynamical system. The difference between them is the type of stochastic dynamical system. Namely, for the continuous model it is the steady-state of a stochastic differential equation with linear drift, so a continuous time system and for the discrete model it is the steady-state of a first-order vector autoregressive model, so a discrete time system. The definition of both dynamical systems is provided in Table 1. Both models are then considered to be models of cross-sectional data from the dynamical systems once they have stabilized.

Historically, the covariance equations of the steady-state distributions first arose from the study of stability theory for continuous-time and discrete-time linear systems, where they can be used to verify whether such a system is stable [Gajic and Qureshi, 1995]. Stability theory was first studied in Russian mathematician Alexander Mikhailovitch Lyapunov's thesis in 1892 (first full English translation is from 1992, Lyapunov [1992]) whom these types of equations are named after. The equations are used in many other aspects of mathematics for example control theory, differential equations and signal processing [Gajic and Qureshi, 1995]. However, a key difference is that in the classical linear stability set-up they study the opposite problem than the one relevant for statistical inference. Namely, they assume that the parameter matrix, A or M, respectively is known and seek to answer if the covariance equations are solvable for a positive definite Σ. The answer to this is positive when the parameter matrices satisfy two different notions of

	Continuous	**Discrete**
Parameters	$M \in \mathbb{R}^E$ Hurwitz Stable and $(Z_t)_{t\geq 0}$ a Lévy process with $\mathcal{C}_n = \text{cum}_n(Z_1)$	$A \in \mathbb{R}^E$ Schur Stable and ε_t i.i.d random variables with $\Omega^{(n)} = \text{cum}_n(\varepsilon_1)$
Dynamics	$dX_t = MX_t dt + dZ_t$	$X_1 = \varepsilon_1$, $X_t = AX_{t-1} + \varepsilon_t \quad \text{for } t > 1$
Covariance	$M\Sigma + \Sigma M^T + \mathcal{C}_2 = 0$	$\Sigma = A\,\Sigma\,A^T + \Omega^{(2)}$
Cumulants	$\mathcal{K}_n \times_1 M + \ldots + \mathcal{K}_n \times_n M + \mathcal{C}_n = 0$	$\mathcal{K}_n = \mathcal{K}_n \times_1 A \cdots \times_n A + \Omega^{(n)}$
Trek rule	Yes (Prop. 3.B.11)	Yes (equi-treks, Prop. 2.3.2)
Papers	[**ICLM**], [**LOSD**]	[**IDLM**]

Table 1: Comparison of discrete and continuous Lyapunov Models for a directed graph $G = ([d], E)$ on d nodes.

stability. For the discrete equation A needs to be Schur stable, i.e., have all eigenvalues with absolute value less than one and for the continuous equation M needs to be Hurwitz stable, i.e., have the real part of all eigenvalues be negative [Gajic and Qureshi, 1995]. In statistics we seek to answer the opposite problem: if we observe the covariance (and potentially higher-order cumulants) can we recover the parameter matrix, A or M, respectively?

Even though the covariance equations of the steady-state distributions have been well-known for a long time [Gajic and Qureshi, 1995] they have only recently been considered to define a graphical model both in the continuous case [Fitch, 2020; Varando and Hansen, 2020] and in the discrete case [Young et al., 2019]. In all previous work for the continuous and discrete Lyapunov models only the covariance equations have been used. In Varando and Hansen [2020]; Fitch [2020]; Dettling et al. [2024] they study estimation of M, in Boege et al. [2025] they study conditional independence and in Dettling et al. [2023]; Améndola et al. [2025] they study identifiability in the continuous Lyapunov models. In Young et al. [2019] they study both estimation and identifiability in the discrete Lyapunov model and Liu [2025] show results about structural identifiability.

The graphical models are defined by assuming that the parameter matrix has its sparsity pattern encoded by a directed graph. Let $G = ([d], E)$ be a directed graph on d nodes. We define the set of $d \times d$ matrices with sparsity pattern according to G by

$$\mathbb{R}^E = \{B \in \mathbb{R}^{d\times d} \mid B_{ij} = 0 \text{ if } j \to i \notin E\},$$

as it is defined in [**ICLM**] and [**IDLM**]. For both models we assume that the parameter matrix is in this set and satisfies the appropriate notion of stability. The suitable notion of stability ensures that the steady-state distributions exist.

As opposed to the previous literature on graphical Lyapunov models, we study identifiability and estimation of M and A, respectively not only terms of the covariance

but also the higher-order cumulants of the steady-state distribution, which are derived in [**ICLM**] and [**IDLM**], respectively. Thus, we treat the matrices M and A as the parameters of interest and the cumulants $\mathcal{C}_n$ and $\Omega^{(n)}$ as nuisance parameters. For this reason we will sometimes refer to $\mathcal{C}_n$ and $\Omega^{(n)}$ as the error cumulant tensors.

By the definition of mode products ($\times_i$, see definitions in [**IDLM**] or [**ICLM**]), we see that the covariance equations are exactly the cumulant equations for $n = 2$ as is to be expected. In order for the higher-order cumulants to be non-zero the steady-state distributions need to be non-Gaussian. For the continuous Lyapunov model this implies that the Lévy process, $(Z_t)_{t\geq 0}$, should *not* be a Brownian motion. For the discrete Lyapunov model the errors ε_t should be non-Gaussian.

As can be seen in Table 1 the cumulant equations are not the same for the two models even though a first-order vector autoregressive model can be seen as a discretization of a stochastic differential equation with linear drift. One of the main differences is that the continuous cumulant equations are linear in the entries of M whereas the discrete equations are in general polynomial of order equal to the order of the cumulant in the entries of A.

However, as explained in Chapter 3 in Gajic and Qureshi [1995] it is possible for a given $(A, \Omega^{(2)})$ solving the second-order discrete Lyapunov equation for Σ to find a pair of parameters $(M, \mathcal{C}_2)$ solving the second-order continuous Lyapunov equation for the same Σ. We provide one example but Gajic and Qureshi [1995] actually provide two different ways of doing this. If Σ solves the second-order discrete Lyapunov equation for $(A, \Omega^{(2)})$ then the pair

$$M_A = (A - I)(A + I)^{-1}, \qquad (\mathcal{C}_2)_A = 2(A + I)^{-T}\Omega^{(2)}(A + I)^{-1}$$

will solve the second-order continuous Lyapunov equation for Σ. However, it does not appear possible to find a transformation such that the same M_A could be used for both the covariance and a higher-order cumulant. Even though such a transformation is possible at least in the second-order case it does *not* provide an obvious way to yield information about parameter identifiability in one model based on the other one. This is namely because the sparsity patterns of M_A and $(\mathcal{C}_2)_A$ will in general be different than those of A and $\Omega^{(2)}$. Generally, M_A will have more non-zero entries than A, and $(\mathcal{C}_2)_A$ will generally not be diagonal even if $\Omega^{(2)}$ is assumed diagonal.

The identifiability results for either type of Lyapunov model in this thesis or elsewhere either rely on a known error covariance matrix [Dettling et al., 2023; Young et al., 2019]; or a known higher-order cumulant error tensor ([**ICLM**] Proposition 3.B.12); or on using more than one order of cumulant equation with diagonal error cumulants ([**ICLM**] and [**IDLM**]). Therefore, the lack of preservation of diagonal error cumulants or of known $\Omega^{(2)}$ implying known $(\mathcal{C}_2)_A$ appears to make it impossible to transfer the results between the two models in this way. However, it would be interesting if this correspondence between the equations could be used for something, potentially to transfer results about structural identifiability.

In [**ICLM**] the main identifiability result is rational identifiability of $(M, \mathcal{C}_2, \mathcal{C}_r)$ up to a joint scaling factor from the second-order and r-th-order continuous Lyapunov equation

when the error cumulants are assumed diagonal and the defining graph G is connected and contains all self-loops. In [**IDLM**] the first main identifiability result is rational identifiability of $(A, \Omega^{(2)}, \Omega^{(3)}, \Omega^{(4)})$ from the second-, third- and fourth-order cumulants when the error cumulants are assumed diagonal and the graph is a directed acyclic graph with all self-loops and no isolated nodes. The second main identifiability result is local identifiability of $(A, \Omega^{(2)}, \Omega^{(3)})$ from the second- and third-order cumulants when the error cumulants are assumed diagonal and G is any graph with all self-loops and no isolated nodes.

Thus, the similarities between the identification results are that they both assume some order of the error cumulants to be diagonal, that the graph must contain all self-loops and the use of a combination of the covariance with higher-order cumulants. Another common factor is that the proofs in both papers rely heavily on the combinatorial description of the entries of the cumulants given by their respective trek rules. The main difference is that for the continuous Lyapunov model we only have identifiability up to common scaling factor. For this model this is the best possible result when we assume that the the error of the cumulants are unknown. This is inherent to the system because we can never determine the speed at which the solution developed. This also explains why the graph has to be connected since otherwise there would be a separate scaling factor for each connected component. This is not a problem for the discrete Lyapunov model because the underlying dynamical system is discrete.

Both [**ICLM**] and [**IDLM**] focus on identification theory within the models, with [**ICLM**] also including a semiparametric estimator of M based on said identifiability results. However, the study of both models are motivated by potential application to observing cross-sectional data from a dynamical system as for example seen in single cell biology [Wang et al., 2023]. Furthermore, both models can be given a causal interpretation, where the entries of the parameter matrices would correspond to direct causal effects. An advantage of both models in this context is that because of the underlying temporal nature both models naturally allow for feedback loops, encoded as cycles in the graphs. In a classical causal graphical model, linear structural equations [Drton, 2018], cycles are both mathematically and interpretationally more challenging to include.

The ongoing project [**LOSD**] focuses exactly on including a causal interpretation for the continuous Lyapunov model. In [**LOSD**] we further introduce a measurement noise model similar to Lorch et al. [2026] to make the setting more realistic.

2 Identifiability in Graphical Discrete Lyapunov Models

Cecilie Olesen Recke, Sarah Lumpp, Nataliia Kushnerchuk, Janike Oldekop, Jiayi Li, Jane Ivy Coons, and Elina Robeva

Abstract

In this paper, we study discrete Lyapunov models, which consist of steady-state distributions of first-order vector autoregressive models. The parameter matrix of such a model encodes a directed graph whose vertices correspond to the components of the random vector. This combinatorial framework naturally allows for cycles in the graph structure. We focus on the fundamental problem of identifying the entries of the parameter matrix. In contrast to the classical setting, we assume non-Gaussian error terms, which allows us to use the higher-order cumulants of the model. In this setup, we show generic identifiability for directed acyclic graphs with self-loops at each vertex and show how to express the parameters as a rational function of the cumulants. Furthermore, we establish local identifiability for *all* directed graphs containing self-loops at each vertex and no isolated vertices. Finally, we provide first results on the defining equations of the models, showing model equivalence for certain graphs and paving the way towards structure learning.

2.1 Introduction

Vector autoregressive models are widely used in control theory, the health sciences [van der Krieke et al., 2017], and econometrics [Sims, 1980]. Furthermore, they are applied to model gene regulatory networks in computational biology [Michailidis and d'Alché Buc, 2013; Rajapakse and Mundra, 2011]. In particular, the first-order vector autoregressive model, or VAR(1) model, is a common modeling choice. Its dynamics are given by

$$\begin{aligned} X_1 &= \varepsilon_1, \\ X_t &= AX_{t-1} + \varepsilon_t \quad \text{for } t > 1, \end{aligned} \tag{2.1.1}$$

where X_t is a p-dimensional random vector, $A \in \mathbb{R}^{p\times p}$ is a parameter matrix encoding the interactions among the components of X_t, and ε_t is a p-dimensional random vector

modeling independent errors with mean zero and diagonal covariance matrix D. Under certain conditions on the eigenvalues of A, the system (2.1.1) has a unique *stationary solution*. Our primary object of interest is the steady-state distribution of the stationary solution. This is also called the equilibrium, limiting or simply stationary distribution in the literature. The covariance matrix Σ of this steady-state distribution is parametrized by the *discrete Lyapunov equation*

$$A\Sigma A^T + D = \Sigma. \tag{2.1.2}$$

If the errors are Gaussian, then the steady-state distribution is also Gaussian and equation (2.1.2) completely characterizes this distribution by its covariance. In the case of non-Gaussian errors, which we study here, the steady-state distribution can further be characterized by its *higher-order cumulants* which satisfy higher-order analogues of equation (2.1.2). These equations are described in Section 2.2.1.

The VAR(1) model allows a natural graphical representation. Let $G = (V, E)$ be a directed graph whose vertices V correspond to the p coordinates of the random process X_t, and the edges $E \subseteq V \times V$ represent the sparsity pattern of the parameter matrix A. In other words, $a_{ji} = 0$ if $i \to j \notin E$. Note that this definition allows for self-loops at each node. To simplify indexing later on, we usually number the vertices as $V = \{0\} \cup [p-1]$. Parallel to the continuous setting [Recke and Hansen, 2026], we refer to the set of centered distributions satisfying the discrete Lyapunov equations up to order n as the *nth order discrete Lyapunov model of the directed graph* G (see Definition 2.2.3).

Unlike structural causal models in which the random vector X does not evolve with time [Peters et al., 2017], VAR(1) models naturally allow for *cycles*, reflecting feedback loops in many real-world processes (e.g., stability of control systems [Passino et al., 2002], gene regulatory networks [Karlebach and Shamir, 2008; Young et al., 2019], and macro-economic dynamics [Kozlov and Kozlova, 2009]). In particular, it is common to allow the existence of self-loops where some of the variables depend on their own past, capturing the temporal and cyclic nature of the model. An interpretation of cycles in linear structural equation models is discussed in Hyttinen et al. [2012], where it corresponds to the error in (2.1.1) not depending on time. The authors of Hyttinen et al. [2012] discuss the limits of this deterministic view point and propose the set-up in the current paper as a possible extension.

Since our primary interest lies in the steady-state distribution of the time series, each of our samples corresponds to a single snapshot of cross-sectional data once the process has stabilized. This perspective is critical in fields such as computational biology, where measuring the same system at multiple time points may be infeasible due to the destructive nature of RNA-sequencing technologies [Young et al., 2019]. Even though only one time point is observed per sample, the underlying dynamic structure is preserved by virtue of the model's stationarity. Furthermore, by keeping the interpretation of the data as arising from the steady-state, there is a natural interpretation of cycles.

Variations of this approach were recently established in Young et al. [2019] and Varando and Hansen [2020]. While Young et al. [2019] focuses on the Gaussian stationary VAR(1) model, Varando and Hansen [2020] proposes the equivalent in a continuous setting – the graphical continuous Lyapunov model. This model parametrizes

stationary distributions of diffusion processes, such as the Ornstein-Uhlenbeck process, via the continuous Lyapunov equation. Recent results include advances in parameter identifiability in the Gaussian and the non-Gaussian cases [Dettling et al., 2023; Recke and Hansen, 2026], as well as new findings on structure identifiability and model equivalence [Améndola et al., 2025]. Further work has addressed statistical estimation and structure learning [Varando and Hansen, 2020; Dettling et al., 2024]. In more general diffusion models with potentially nonlinear drift, recent contributions characterize the model's conditional independence structure [Boege et al., 2025] and propose intervention-based learning procedures [Lorch et al., 2024].

While continuous Lyapunov models continue to be an active area of research, analogous results for discrete-time models are limited. In the case of continuous Lyapunov models with unknown errors, the strongest possible identifiability result can only ever be up to a scaling of the parameter matrix [Dettling et al., 2023]. However, as we will show, this is not the case for discrete Lyapunov models. To reliably *estimate the graph* G from data in the discrete setting, it is crucial to first provide theoretical guarantees for *identifiability* of the model parameters if the graph is known. This is precisely the question we address in the present work. To do so, we use tools from algebraic geometry, graph theory, and algebraic statistics.

In the case of Gaussian error terms, the resulting stationary distribution is Gaussian as well, implying that all cumulants of order higher than two are zero. For general graphs, however, the parameters are not identifiable from only the first and second-order moments [Comon and Jutten, 2010] – for instance, flipping the sign of the parameter matrix A already results in the same covariance matrix. Prior work by Young et al. [2019] mitigates these issues by assuming that the error covariance D is known as well as imposing a particular sparsity pattern on A and assuming the signs of all the diagonal elements of A are known. In this specific setting, they are able to show identifiability of the parameters in A from the covariance matrix as well as consistency and asymptotic efficiency of the maximum likelihood estimator of A. Recent work by Liu [2025] presents first results on structure identifiability from the covariance matrix. In particular, they consider model distinguishability based on the Jacobian matroid of the parametrization map. They provide sufficient conditions for generic structure distinguishability based on model dimension and graphical structures in special cases.

In this work, in contrast, we consider the non-Gaussian setting which allows us to leverage higher-order cumulants as well. We show in Theorem 2.5.6 and following corollaries that the entries of the parameter matrix A together with the cumulants of the error terms are locally identifiable through an algebraic characterization of the second-, third-, and fourth-order cumulants for all graphs with all self-loops without an isolated node. We also show in Theorem 2.4.3 that when the sparsity pattern of A specifies a DAG where all self-loops are present, the model parameters are generically (rationally) identifiable from second-, third- and fourth-order cumulants. Furthermore, we investigate model equivalence and the polynomial equations satisfied by the cumulants for certain families of graphs.

The remainder of the paper is organized as follows. In Section 2.2, we derive the discrete Lyapunov equations for higher-order cumulants and use them to define higher-

order discrete Lyapunov models. In Section 2.3, we show how to parametrize the model using a trek rule and how this trek rule can be restricted to base treks without self-loops in a specific setting. Section 2.4 addresses the question of parameter identifiability when the underlying graph is a DAG (with self-loops). In this case, we prove that the matrix A and the cumulants of the error are generically identifiable from the second-, third-, and fourth-order cumulants of the stationary distribution of X_t. Section 2.5 extends these results by proving local identifiability for arbitrary directed graphs that may contain cycles while considering the Jacobian of only the second- and third-order cumulants. In Section 2.6, we provide initial results on model equivalence as well as equations in the ideal of the model.

2.2 Preliminaries

In this section, we provide preliminary results and definitions for discrete Lyapunov models. In Section 2.2.1, we first derive the tensor equations for the cumulants of the steady-state of a VAR(1) process; these equations are referred to as the *discrete Lyapunov equations*. In Section 2.2.2, we define the nth-order discrete Lyapunov model for a given directed graph G. Finally, in Section 2.2.3, we introduce different types of identifiability and provide initial remarks on the types of identifiability that can be feasibly established for discrete Lyapunov models.

2.2.1 Higher-Order Cumulants and the Discrete Lyapunov Equations

A steady-state distribution of a VAR(1) process as described in (2.1.1) exists in the general setting if A is a Schur stable matrix, which is a matrix where all eigenvalues have absolute values strictly less than 1, and that the ε_t are independent with mean 0 and have the same covariance. Under these condition, Example 8.4.1 of Brockwell and Davis [Brockwell and Davis, 2016] (combined with Example 2.2.1) shows that there is a unique stationary solution to (2.1.1), given by

$$X_t \;=\; \sum_{j=0}^{\infty} A^j \, \varepsilon_{t-j}. \tag{2.2.3}$$

Intuitively, requiring $|\lambda_i(A)| < 1$ ensures that the powers A^j decay as $j \to \infty$, making the above series converge absolutely. Hence, for large t, the distribution of X_t stabilizes to a unique steady-state distribution, also sometimes called its limiting, equilibrium or stationary distribution.

We aim to characterize steady-state distributions of VAR(1) processes by their higher-order cumulants. Therefore, we will assume that the errors, ε_t, are independent and identically distributed so that all their higher-order cumulants are identical. We derive equations that describe the cumulants of the steady-state based on the parameter matrix A and the corresponding cumulant tensors of the errors ε_t. Denote by $\Omega^{(n)} = \mathrm{cum}_n(\varepsilon_t)$ the nth-order cumulant tensor of the noise. Note that $\Omega^{(2)} = D$ as introduced in (2.1.1). As shown in Young et al. [2019], the second-order cumulant of a stationary

process X_t, i.e., $\mathrm{Var}(X_t)$, is described by the discrete Lyapunov equation in (2.1.2) or see remark after Proposition 2.2.1. We show that the higher-order cumulants of the steady-state distribution of X_t follow higher-order tensor equivalents of the discrete Lyapunov equation.

In the following, we will denote the second-, third-, and fourth-order cumulants by $S = T_2$, $T = T_3$, and $R = T_4$, respectively. A notion of taking the product between a tensor and a matrix is needed to write the higher-order discrete Lyapunov equations, this is called the k-mode product. The definition of the k-mode product of a tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with a matrix $A \in \mathbb{R}^{J \times I_k}$, denoted $T \times_k A$, is a $I_1 \times \cdots \times I_{k-1} \times J \times I_{k+1} \times \cdots \times I_N$ tensor with element-wise entries

$$(T \times_n A)_{i_1 \cdots i_{k-1} j i_{k+1} \cdots i_N} = \sum_{i_k=1}^{I_k} T_{i_1 \cdots i_{k-1} i_k i_{k+1} \cdots i_N} A_{j i_k}.$$

When performing the multiplication along every mode at once (as in equation (2.2.4)), this is also called the Tucker product.

Proposition 2.2.1 (nth-Order Cumulants)**.** *Let $A \in \mathbb{R}^{d \times d}$ be a Schur stable matrix, meaning it has all eigenvalues $|\lambda_i(A)| < 1$. Suppose ε_t has finite nth order cumulant $\Omega^{(n)}$. Then the steady-state distribution of the VAR(1) model (2.1.1) has finite nth cumulant T_n, and*

$$T_n = \sum_{i=0}^{\infty} \Omega^{(n)} \times_1 A^i \times_2 \cdots \times_n A^i, \tag{2.2.4}$$

where $\times_k$ denotes the k-mode product. Two other equivalent equations for T_n are the recursive formula

$$T_n \;=\; T_n \;\times_1 A \;\times_2 A \;\cdots\; \times_n A \;+\; \Omega^{(n)}, \tag{2.2.5}$$

and the vectorized product

$$\mathrm{vec}(T_n) = (I - \underbrace{A \otimes \cdots \otimes A}_{n \text{ times}})^{-1} \mathrm{vec}(\Omega^{(n)}). \tag{2.2.6}$$

Proof. From (2.1.1), we can apply the properties of how cumulants behave under linear and affine transformations (see, e.g., [McCullagh, 2018, Section 2.4]) and repeatedly use the linear relation $X_{t+1} = AX_t + \varepsilon_{t+1}$. For nth-order cumulants, one obtains the recursive identity

$$\mathrm{cum}_n(X_{t+1}) \;=\; \big(\mathrm{cum}_n(X_t)\big) \;\times_1 A \;\times_2 A \;\cdots\; \times_n A \;+\; \Omega^{(n)}. \tag{2.2.7}$$

Under stationarity (and by (2.2.3)), X_t converges in distribution to the limit X_∞ whose nth order cumulant T_n satisfies

$$T_n \;=\; \lim_{t \to \infty} \mathrm{cum}_n(X_t).$$

Unraveling the recursion yields the infinite series (2.2.4). Convergence follows because A is Schur stable; so $A^i \to 0$ as $i \to \infty$ sufficiently fast to guarantee that the infinite sum is well-defined.

The vectorized equation (2.2.6) arises by vectorizing (2.2.4) and using properties of the k-mode product and Kronecker product.

Equation (2.2.5) also follows from equation (2.2.7) since at stationarity the cumulant cannot change from time step to time step, so the cumulant of the steady-state distribution has to satisfy (2.2.5). Alternatively, [Xu and Wang, 2022, Corollary 3.2] yields that, when A is Schur stable, (2.2.4) is the unique solution to the recursive equation (2.2.5) implying that the two equations are equivalent. □

The recursive equation (2.2.5) is the generalization of what has classically been known as the (second-order) discrete Lyapunov equation, (2.1.2). Since they are all equivalent, we will refer to all the equations of the form (2.2.4), (2.2.5) and (2.2.6) as *discrete Lyapunov equations*, to differentiate between them the order of the cumulant should be specified.

Remark 2.2.2 (Gaussian special case). If in addition ε_t is Gaussian, one recovers the standard result: in the limit, X_t follows a multivariate Gaussian distribution with mean $\mathbf{0}$ and covariance matrix Σ satisfying the *discrete Lyapunov equation*

$$\Sigma \;=\; A\,\Sigma\,A^T \;+\; D,$$

where D is the noise covariance. Moreover,

$$\Sigma \;=\; \sum_{j=0}^{\infty} A^j\,D\,(A^T)^j \;=\; (I - A \otimes A)^{-1}\,\mathrm{vec}(D),$$

highlighting that the requirement $|\lambda_i(A)| < 1$ ensures invertibility of $(I - A \otimes A)$.

2.2.2 Defining the Model via Cumulant Equations

A directed graph G is a pair (V, E), where V is the set of vertices, which we usually denote as $V = \{0\} \cup [p-1]$, and $E \subseteq V \times V$ is the set of edges, which are ordered pairs (i, j), often denoted as $i \to j$. We denote by $\mathrm{pa}(i)$ the set of parents of a node $i \in V$, i.e., $\mathrm{pa}(i) = \{j \in V : j \to i \in E\}$.

Given a directed graph $G = (V, E)$, we define the (graphical) *discrete Lyapunov model* of order n as the set of all cumulant tensors satisfying the appropriate discrete Lyapunov equations, where A is a stable matrix with sparsity pattern according to the edges of G, that is $a_{ji} = 0$ if $i \to j \notin E$. We denote this set by $\mathbb{R}^E$. Furthermore, as explained in Section 2.2.1, the matrix A needs to be Schur stable in order for the steady-state distribution to exist, so we will further assume that $A \in \mathbb{R}^E_{\mathrm{stab}}$, the Schur stable matrices which have sparsity pattern according to G. In the following let $\mathrm{Sym}_m(\mathbb{R}^l)$ denote the set of symmetric tensor of order m on $\mathbb{R}^l$.

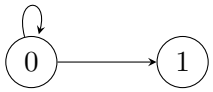


Figure 2.2.1: A directed graph on two nodes with a single self-loop at the source.

Definition 2.2.3. *Let $G = (V, E)$ be a directed graph. The* nth-order discrete Lyapunov cumulant model *of G is the set*

$$\mathcal{M}_G^{\leq n} = \{(T_2, \ldots, T_n) \ : T_m = \sum_{i=0}^{\infty} \Omega^{(m)} \times_1 A^i \times_2 \cdots \times_m A^i \text{ for all } 2 \leq m \leq n, \\ A \in \mathbb{R}^E_{\text{stab}},\ \Omega^{(m)} \in \text{Sym}_m(\mathbb{R}^{|V|}) \text{ diagonal for all } 2 \leq m \leq n\}.$$

The nth-order cumulant ideal *of G is the ideal $\mathcal{I}^{\leq n}(G)$ of polynomials in the entries $T_2, \ldots, T_n$ that vanish when $(T_2, \ldots, T_n) \in \mathcal{M}_G^{\leq n}$.*

Notice that $\mathcal{I}^{\leq n}(G)$ is a homogeneous ideal. Usually, we will focus on the case where $n = 3$ or 4, corresponding to Lyapunov models of order at most four.

Example 2.2.4. Consider the third-order discrete Lyapunov cumulant model of the graph on two nodes in Fig. 2.2.1. The graph gives rise to the following A matrix, and the error cumulants are assumed to be diagonal:

$$A = \begin{pmatrix} a_{00} & 0 \\ a_{10} & 0 \end{pmatrix}, \qquad \Omega^{(2)} = \begin{pmatrix} \omega_{00} & 0 \\ 0 & \omega_{11} \end{pmatrix}, \qquad \Omega^{(3)} = \text{diag}(\omega_{000}, \omega_{111}).$$

The second- and third-order cumulants are given by

$$S = ASA^\top + \Omega^{(2)}, \qquad T = T \times_1 A \times_2 A \times_3 A + \Omega^{(3)}.$$

Writing these in entries yields

$$\begin{aligned} &s_{00} = a_{00}^2 s_{00} + \omega_{00}, \quad &&s_{01} = a_{00}a_{10}s_{00}, \quad &&s_{11} = a_{10}^2 s_{00} + \omega_{11}, \\ &t_{000} = a_{00}^3 t_{000} + \omega_{000}, \quad &&t_{001} = a_{00}^2 a_{10} t_{000}, \quad &&t_{011} = a_{00}a_{10}^2 t_{000}, \quad &&t_{111} = a_{10}^3 t_{000} + \omega_{111}. \end{aligned}$$

Eliminating the parameters $a_{00}, a_{10}, \omega_{000}$ leads to the polynomial relation, $s_{01}^3 t_{000}^2 - s_{00}^3 t_{001} t_{011} = 0$, which generates the third-order cumulant ideal

$$\mathcal{I}^{\leq 3}(G) = \left\langle s_{01}^3 t_{000}^2 - s_{00}^3 t_{001} t_{011} \right\rangle.$$

2.2.3 Identifiability

In this paper we study identifiability of the parameters in the discrete Lyapunov model for a known graph G. Identifiability is about determining injectivity of the map from the set of parameters of interest, A, and the noise parameters, $\Omega^{(n)}$, to the model $\mathcal{M}_G^{\leq n}$. In this section we introduce different notions of identifiability. In Section 2.4 we prove

generic identifiability of all DAGs (with self-loops) from second-, third- and some fourth-order cumulants, and in Section 2.5 we prove local identifiability for most directed graphs from only the second- and third-order cumulants.

If we let the error cumulants $\Omega^{(n)}$ be represented by their non-zero diagonal entries, then the parameters $(A, \Omega^{(2)}, \Omega^{(3)}, \Omega^{(4)})$ of the fourth-order discrete Lyapunov model lie in the following set:

$$\Theta_{G,\leq 4} = \mathbb{R}^E_{\text{stab}} \times \mathbb{R}^p_+ \times (\mathbb{R} \setminus \{0\})^p \times \mathbb{R}^p_+.$$

The question of identifiability for the fourth-order discrete Lyapunov models is then whether the (rational) map

$$\phi_{G,\leq 4} : \Theta_{G,\leq 4} \to PD_p \times \text{Sym}^3(\mathbb{R}^p) \times \text{Sym}^4(\mathbb{R}^p), \quad (A, \Omega^{(2)}, \Omega^{(3)}, \Omega^{(4)}) \mapsto (S, T, R)$$

is injective, where (S, T, R) are the solutions to the corresponding second-, third- and fourth-order discrete Lyapunov equations. The identifiability results will initially be stated as identifying A since it is clear by considering (2.2.5) that if A can be identified from the cumulants (S, T, R) then we can also identify the error cumulants.

We provide definitions of several different types of identifiability for a parametrized statistical model, see for example [Sullivant, 2018, Chapter 16] for an introduction. By a statistical model we mean a set of probability measures or a representation of such a set, e.g. via cumulants up to a certain order.

Definition 2.2.5. *Let $f : \Theta \to \mathcal{M}_f$ be a rational map from a finite dimensional parameter space $\Theta \subseteq \mathbb{R}^k$ to a statistical model such that $im(f) = \mathcal{M}_f$. The model is said to be*

(i) globally identifiable if $f^{-1}(f(\theta)) = \theta$ for all $\theta \in \Theta$;

(ii) generically identifiable if $f^{-1}(f(\theta)) = \theta$ for all $\theta \in \Theta \setminus A$ where A is a proper algebraic subset of Θ;

(iii) locally identifiable if $|f^{-1}(f(\theta))| < \infty$ for all $\theta \in \Theta \setminus A$ where A is a proper algebraic subset of Θ;

(iv) non-identifiable if $|f^{-1}(f(\theta))| = \infty$ for generic θ.

Note that for the second and third notion of identifiability the Zariski closure $\overline{\Theta}$ of Θ is assumed to be an irreducible variety.

In some literature, generic (and local) identifiability is defined as identifiability away from a null set. Definition 2.2.5 (ii) is more specific than this since an algebraic set is also always a null set. Requiring the exceptional set to be algebraic has several justifications; general null sets can behave pathologically (for example, they may be dense in the ambient space) in ways algebraic sets cannot and in the following theorems it is the stronger version (Definition 2.2.5 (ii)) that we are able to prove in all cases. Local identifiability is also known as finite-to-one identifiability. In the case where $\mathcal{M}_f$

is generically identifiable and $f^{-1}(p)$ is a rational function of the entries of p for almost all $p \in \mathcal{M}_f$, we say that $\mathcal{M}_f$ is generically *rationally* identifiable.

We seek to answer the question of identifiability for the discrete Lyapunov models (primarily up to fourth-order) so we will apply the definitions to the map $\phi_{G,\leq 4}$. The parameter set $\Theta_{G,\leq 4}$ is an open subset (in the standard Euclidean topology) so its Zariski closure is therefore irreducible and Definition 2.2.5 (ii) and (iii) can be applied. If the discrete Lyapunov model of a certain order corresponding to a graph G is identifiable it is sometimes denote this as the graph G being identifiable.

Note that the discrete Lyapunov model with unknown errors can never be globally identifiable no matter the order of the cumulants. This is because the subset of the parameter space corresponding to isolating a node in the graph (or isolating all of them by taking A to be diagonal) yields an under determined system (see Remark 2.4.4). In the case where A is diagonal, the higher order cumulants in the model are diagonal as well (see the treks rules in the following section) with entries only depending on the corresponding diagonal entries of A and the respective error cumulants. This is why the results of Section 2.4 and Section 2.5 focus on generic and local identifiability.

Furthermore, if one were to consider only the second-order discrete Lyapunov model with unknown diagonal errors, it is not possible to prove generic identifiability without imposing sign conditions on several entries of A. As discussed by Young et al. [2019], this can be immediately seen for the second-order discrete Lyapunov model by observing that flipping the signs of a parameter matrix $A \in \mathbb{R}^E_{\text{stab}}$ to $-A \in \mathbb{R}^E_{\text{stab}}$ yields the same covariance matrix. This is why the results of Section 2.4 and Section 2.5 use third- and fourth-order cumulants as well as the covariance.

2.3 Trek Parametrization

In this section we consider the individual entries of the cumulants T_n and the expression for each such entry in terms of the parameters A and $\Omega^{(n)}$ induced by the formulas in equation (2.2.4). In Section 2.3.1 we give a combinatorial way, called the trek rule, to read off these expressions directly from the graph. In Section 2.3.2 we show how in special cases one can simplify the expressions for the entries of T_n. Finally, in Section 2.3.3 we show how missing treks in the graph imply marginal independence between some of the random variables.

2.3.1 Trek Rules

Trek rules are commonly used to express the entries of the moments and cumulants in linear structural equation models [Sullivant et al., 2010; Robeva and Seby, 2021; Améndola et al., 2023] and continuous Lyapunov models [Boege et al., 2025; Hansen, 2025]. We prove that similar rules hold for the expression of the entries of the cumulants T_n in our discrete Lyapunov models, with the only difference that the paths of the treks need to have the same length. We call such treks *equitreks*. They are a specialization of the treks introduced in Sullivant et al. [2010] and Robeva and Seby [2021] for linear structural equation models.

Definition 2.3.1. *Let $G = (V, E)$ be a directed graph. A k-equitrek between k vertices $i_1, \dots, i_k$ is an ordered set of directed paths $(P_1, \dots, P_k)$ of the same length l from a common source $v \in V$ to sinks $i_1, \dots, i_k$, respectively. We say that such a k-equitrek has length l. We let $\mathcal{T}(i_1, \dots, i_k)$ denote the set of all k-equitreks between the nodes $i_1, \dots, i_k$.*

Rewriting the formula for T_n obtained in Proposition 2.2.1 yields the following trek rule with the proof given in Appendix 2.A. The cumulant $\Omega^{(n)}$ of the noise vector ε_t is a diagonal tensor, so we only use its p diagonal entries denoted by $w_1^{(n)}, \dots, w_p^{(n)}$ to simplify the notation.

Proposition 2.3.2. *Let $G = (V, E)$ be a directed graph on p nodes, where $V = \{0\} \cup [p-1]$, and let the cumulant T_n satisfy the nth order discrete Lyapunov equation (2.2.4). Then T_n satisfies the trek rule*

$$(T_n)_{i_1,\dots i_n} = \sum_{\tau=(\tau_1,\dots,\tau_n)\in\mathcal{T}(i_1,\dots,i_n)} w^{(n)}_{top(\tau)} a^{\tau_1} \cdots a^{\tau_n},$$

where a^P denotes the path monomial associated to the path P given as $a^P = \prod_{\alpha\to\beta\in P} a_{\beta\alpha}$.

Remark 2.3.3. If the error cumulant $\Omega^{(n)}$ is not assumed to be diagonal, the top of a trek can be considered to be an n-fold blunt (or multidirected) edge corresponding to a non-zero entry in $\Omega^{(n)}$. In this way each of the paths could potentially start at different vertices. It is clear that the proof of Proposition 2.3.2 generalizes to this case. See, for example, Varando and Hansen [2020]; Foygel et al. [2012b] for a trek rule of this nature in the covariance case for linear structural equation models and continuous Lyapunov models, respectively.

The entries of the second- and third-order cumulants in a discrete Lyapunov model are then explicitly given by the following equations.

Corollary 2.3.4. *Let $G = ([p], E)$ be a directed graph. For $(S, T, \dots) \in \mathcal{M}_G^{\leq n}$ with $n \geq 3$, we obtain the following trek rules:*

$$s_{ij} = \sum_{\tau=(\tau_1,\tau_2)\in\mathcal{T}(i,j)} w^{(2)}_{top(\tau)} a^{\tau_1} a^{\tau_2}, \qquad t_{ijk} = \sum_{\tau=(\tau_1,\tau_2,\tau_3)\in\mathcal{T}(i,j,k)} w^{(3)}_{top(\tau)} a^{\tau_1} a^{\tau_2} a^{\tau_3}.$$

The following example illustrates the trek rule.

Example 2.3.5. Let G be the same graph on two nodes as in Example 2.2.4, shown in Fig. 2.2.1. Then, by using the formulas from Corollary 2.3.4, we get

$$s_{00} = \sum_{t=0}^{\infty} w_0^{(2)} a_{00}^{2t} = \frac{w_{00}^{(2)}}{1 - a_{00}^2}, \quad s_{01} = s_{10} = \sum_{t=0}^{\infty} w_0^{(2)} a_{00} a_{10} a_{00}^{2t} = \frac{w_0^{(2)} a_{00} a_{10}}{1 - a_{00}^2}, \quad s_{11} = w_1^{(2)}.$$

Analogously,

$$t_{000} = \frac{w_0^{(3)}}{1 - a_{00}^3}, \qquad t_{001} = \frac{w_0^{(3)} a_{00}^2 a_{10}}{1 - a_{00}^3}, \qquad t_{011} = \frac{w_0^{(3)} a_{00} a_{10}^2}{1 - a_{00}^3}, \qquad t_{111} = w_1^{(3)}.$$

An intuitive explanation of our trek rule is as follows. If we think about the VAR(1) process, in order to calculate the cumulant for two or more variables from the same time point, we would need to trace all possible tuples of paths that start at a common source back in time and end at the variables of interest. Since the variables are at the same time point, the paths will have the same length, which means they form an equitrek. This is also why if there are no cycles (for example self-loops) there are very few equitreks in the graph and many entries of the cumulants will be zero.

Remark 2.3.6. Liu [2025] introduce *maximal classes* as the set of all nodes reachable from a source strongly connected component. The concept of treks offers a complementary viewpoint by interpreting these classes as sets of nodes connected by treks where all top nodes belong to the same source strongly connected component. Our trek rule allows us to reinterpret their structural identifiability results in the language of treks, as they relate the notion of maximal classes to entries of Σ being non-zero.

2.3.2 A Restricted Trek Rule for DAGs

Since the infinite sums in the general trek rule may be hard to compute, we derive a trek rule for DAGs in a restriction of the model where we assume that the entries on the diagonal of the parameter matrix A are constant, i.e., that $a_{ii} = t$ for all $i \in V$. This allows us to obtain a finite sum representation. Similar restrictions have been considered by Boege et al. [2025] and Hansen [2025] in the continuous Lyapunov model based on second-order cumulants.

To formulate the trek rule in a concise way, we define the following notation. Let $\mathcal{T}^*(i,j)$ be the set of *base treks* between i and j, that is, the set of treks that visit each node on the trek only once on each path of the trek (i.e., no self-loops and cycles). Note that we do not require these treks to be equitreks. For a base trek $\tau = (\tau_i, \tau_j) \in \mathcal{T}^*(i,j)$, let $d(\tau_i)$ be the edge-length of the path τ_i, i.e., the distance measured in number of edges between the top node, $\mathrm{top}(\tau)$, and the leaf i. Note that τ does not contain self-loops, so the path monomial a^{τ_i} only contains off-diagonal entries of A.

The idea is to collect all terms that account for the edges of a base trek between i and j (a monomial in the off-diagonal entries of A) as well as collect the terms that encode the different ways of adding self-loops to this base trek to obtain all possible equitreks between i and j with this underlying base trek (a rational function in t). This rational function in t is the product of a power of t balancing out the potentially different lengths of both legs of a base trek and of a rational function whose numerator is a polynomial $p_{x,y}(t)$ and whose denominator is a power of $(1-t^2)$ accounting for all combinations in which arbitrary numbers of self-loops can be added along the trek while still obtaining an equitrek. We derive the following finite sum representation for the entries of the covariance matrix S.

Proposition 2.3.7. *Let G be a DAG with constant self-loop parameter t and consider the corresponding discrete Lyapunov model. Then the entries of the covariance matrices*

S in the model are parametrized by the following base trek rule

$$s_{ij} = \sum_{\tau=(\tau_i,\tau_j)\in\mathcal{T}^*(i,j)} C(d(\tau_i), d(\tau_j); t) a^{\tau_i} a^{\tau_j} \omega^{(2)}_{\mathrm{top}(\tau)}, \tag{2.3.8}$$

where for distances $x, y \geq 0$ along the trek, the rational function $C(x, y; t)$ is defined as

$$C(x, y; t) = t^{|x-y|} \frac{p_{x,y}(t)}{(1-t^2)^{x+y+1}},$$

and the polynomial $p_{x,y}(t)$ in the numerator for distances $x, y \geq 0$ is given by

$$p_{x,y}(t) = \sum_{l=0}^{\min(x,y)} t^{2l} \binom{\max(x,y)}{\min(x,y)-l} \binom{\min(x,y)}{l} = \sum_{l=0}^{\min(x,y)} t^{2(l-\min(x,y))} \binom{x}{l} \binom{y}{l}.$$

Note that for example when $x \leq y$, we can simplify

$$p_{x,y}(t) = \sum_{l=0}^{x} t^{2l} \binom{y}{x-l} \binom{x}{l}.$$

The proof can be found in Appendix 2.A. This result is parallel to the restricted trek rule for DAGs in the continuous Lyapunov model proposed by Boege et al. [2025]. However, while in the continuous setting the trek monomial is weighted by a binomial coefficient, here it is weighted by a sum of binomial coefficients together with a rational term accounting for the equitrek parametrization of the model. The proof of Proposition 2.3.7 is based on induction by employing the recursive formula for the entries of S and can be found in Section 2.A.2. The restricted trek rule simplifies significantly in the case of a tree or a directed path as demonstrated in the following examples.

Example 2.3.8. In the case of a (poly)tree, two nodes i and j have at most one common source in the graph. If i and j have a common source, there is a base trek $\tau = (\tau_i, \tau_j)$ with this source node as its top node. Every base trek between i and j is contained in the trek τ and has a common ancestor k of i and j as its top node. We denote this base trek by $\tau^k = (\tau_i^k, \tau_j^k)$. Consequently, the set of base treks between i and j can directly be parametrized by the common ancestors of i and j. Further let $z = \max(\mathrm{An}(i) \cap \mathrm{An}(j))$ be the common ancestor closest to i and j where $\mathrm{An}(i)$ denotes the ancestors of a node i including i itself. Then the trek rule simplifies further to

$$s_{ij} = t^{|d(\tau_i^z)-d(\tau_j^z)|} \sum_{k\in\mathrm{An}(i)\cap\mathrm{An}(j)} \frac{p_{d(\tau_i^k),d(\tau_j^k)}(t)}{(1-t^2)^{d(\tau_i^k)+d(\tau_j^k)+1}} a^{\tau_i^k} a^{\tau_j^k} w_k^{(2)}.$$

Example 2.3.9. The case of a directed path yields an even simpler formula. Assuming a topological order, the distance between two nodes is given by the difference of their indices. The trek rule is then

$$s_{ij} = t^{|i-j|} \sum_{k=0}^{\min(i,j)} \frac{p_{i-k,j-k}(t)}{(1-t^2)^{(i-k)+(j-k)+1}} a^{\tau_i^k} a^{\tau_j^k} w_k^{(2)}.$$

This illustrates the role of each term in the formula: the first factor $t^{|i-j|}$ adds the minimum number of missing self-loops to turn a base trek into an equitrek. The index k ranges over every possible top node of a trek between i and j. The denominator accounts for the infinite number of additional self-loops that can be added while maintaining equal length on both sides. The numerator $p_{i-k,j-k}(t)$ accounts for all possible ways the self-loops can be positioned along the considered trek with top node k. For illustration, we compute some of these entries for the directed path on four nodes. The appearing polynomials $p_{x,y}(t)$ can be ordered into a symmetric square matrix for $x, y \in \{0,1,2,3\}$ as follows:

$$[p_{x,y}(t)]_{x,y=0,\dots,3} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1+t^2 & 2+t^2 & 3+t^2 \\ 1 & 2+t^2 & 1+2\cdot 2t^2+t^4 & 3+2\cdot 3t^2+t^4 \\ 1 & 3+t^2 & 3+2\cdot 3t^2+t^4 & 1+3\cdot 3t^2+3\cdot 3t^4+t^6 \end{bmatrix}.$$

Observe that the coefficients of t^0 correspond to Pascal's triangle (as they are given by $\binom{x}{y}$ for $x \geq y$) while the other coefficients are scaled and shifted versions thereof. Then we can for example write out

$$s_{23} = \frac{t(3+6t^2+t^4)}{(1-t^2)^6} a_{10}^2 a_{21}^2 a_{32} w_0^{(2)} + \frac{t(2+t^2)}{(1-t^2)^4} a_{21}^2 a_{32} w_1^{(2)} + \frac{t\cdot 1}{(1-t^2)^2} a_{32} w_2^{(2)}.$$

The restricted trek rule for DAGs can be extended to higher-order cumulants as

$$(T_n)_{i_1,\dots,i_n} = \sum_{\tau=(\tau_1,\dots,\tau_n)\in\mathcal{T}^*(i_1,\dots,i_n)} C(d(\tau_1),\dots,d(\tau_n);t)\, a^{\tau_1}\cdots a^{\tau_n} w^{(n)}_{top(\tau)}$$

$$\text{where} \quad C(x_1,\dots,x_n;t) = t^{n\cdot\max(x_1,\dots,x_n)-\sum_{l=1}^n x_l} \frac{p_{x_1,\dots,x_n}(t)}{(1-t^n)^{\sum_{l=1}^n x_l+1}}.$$

While the same induction strategy as for proving Proposition 2.3.7 can be applied, indexing the terms and defining p become much more involved. We conjecture that $p_{x,y}(t)$ generalizes to higher n as

$$p_{x_1,\dots,x_n}(t) = \sum_{l=0}^{\sum_{i=1}^n x_i-\max(x_1,\dots,x_n)} t^{n(\sum_{i=1}^n x_i-\max(x_1,\dots,x_n)-l)} \sum_{k=0}^{l} (-1)^{l-k} \binom{x_1+\cdots+x_1+1}{l-k} \prod_{i=1}^n \binom{x_i+k}{k}.$$

Intuition on this based on generating functions of known integer sequences is given in Section 2.A.3.

2.3.3 Marginal Independence

Similar to Varando and Hansen [2020] and Boege et al. [2025] in the continuous setting, we can derive marginal independencies from missing treks in the graph. The trek parametrization allows us to infer the marginal independence properties of the distributions in the discrete Lyapunov model based on results connecting higher-order cumulants and independence. Consider a multivariate distribution that is uniquely determined by

its moments. If all mixed higher-order cumulants of X_i and X_j of such a distribution are zero, then X_i and X_j are marginally independent [McCullagh, 2018]. Combining this result with the parametrization of the cumulants via equitreks, we obtain the following result on marginal independence in the discrete Lyapunov model. Note that this statement can also be extended to a partition of the indices $I \cup J = V$ where there is no equitrek between the two sets of nodes.

Lemma 2.3.10. *Let G be a directed graph and consider a distribution in the discrete Lyapunov model $\mathcal{M}_G$ that is uniquely determined by its moments. If there is no equitrek between i and j in G, then all mixed higher-order cumulants of X_i and X_j are zero. In particular, $X_i \perp\!\!\!\perp X_j$ holds in the considered distribution.*

Proof. The trek rule allows us to write the $i \cdots ij \cdots j$-th entry of any higher order cumulant as a sum over all equitreks between i and j with the corresponding number of leafs. If there is no equitrek between i and j, then all corresponding mixed higher-order cumulants are zero. As detailed above, this implies $X_i \perp\!\!\!\perp X_j$ for all such distributions. □

This is a more general statement than [Varando and Hansen, 2020, Cor. 2.3], as it includes non-Gaussian distributions. However, as in the continuous time setting, we can choose to restrict the error distributions of the underlying VAR(1) processes to be Gaussian. Then all cumulants of order higher than two are zero, so it is enough to study the covariance structure. To emphasize this, we use the notation from Remark 2.2.2.

Corollary 2.3.11. *Let G be a directed graph and $\Sigma \in \mathcal{M}_G^{\leq 2}$ restricted to Gaussian distributions. If there is no equitrek between i and j in G, then*

$$\Sigma_{ij} = 0.$$

In particular, $X_i \perp\!\!\!\perp X_j$ holds in all distributions in $\mathcal{M}_G^{\leq 2}$.

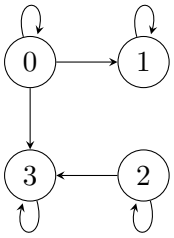
Let $\hat{G} = (V, \hat{E})$ be the bidirected *equitrek graph* derived from G defined by including a bidirected edge between i and j in $\hat{E}$ if there is an equitrek between i and j in G. The same argument as in Boege et al. [2025] allows us to describe the implied conditional independencies that hold for all distributions in the model in the Gaussian case by means of a graphical separation in $\hat{G}$ as follows.

Lemma 2.3.12. *Let G be a directed graph with corresponding equitrek graph $\hat{G}$, and let $I, J, K \subseteq V$ be disjoint index sets. Then*

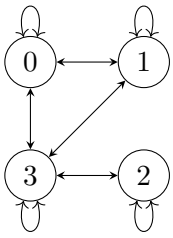
$$V \setminus (I \cup J \cup K) \text{ separates } I \text{ and } J \text{ in } \hat{G} \quad \Rightarrow \quad X_I \perp\!\!\!\perp X_J \mid X_K$$

in the second-order discrete Lyapunov model $\mathcal{M}_G^{\leq 2}$ restricted to Gaussian distributions.

Proof. See the argument in Boege et al. [2025] based on the connected set Markov property for bidirected graphs [Drton and Richardson, 2008]. □



(a) Directed graph G on $p = 4$ nodes including all self-loops.



(b) Bidirected equitrek graph $\hat{G}$.

Figure 2.3.2: Directed graph with corresponding equitrek graph.

Similar to the continuous case, this statement is a special case of a setup with more general Markov processes. A general result in the case of discrete-time Markov processes is given by the second statement of [Niemiro and Rajkowski, 2023, Theorem 3.4], where it is shown that a certain graphical separation in the original graph implies a conditional independence statement holding for the corresponding stationary distributions. As conjectured in Niemiro and Rajkowski [2023], this graphical separation might also be necessary for conditional independence, in the sense that if the separation does not hold, there exists a stationary distribution of a discrete-time Markov process with sparsity pattern given by the graph G where the corresponding conditional independence does not hold. Similar to the approach for the continuous case in Boege et al. [2025], it might be possible to construct these counterexample distributions directly in the discrete Lyapunov model as shown in the following example.

Example 2.3.13. Consider the second-order discrete Lyapunov model $\mathcal{M}_G^{\leq 2}$ restricted to Gaussian distributions for the graph on four nodes with all self-loops depicted in Fig. 2.3.2. Since there is no (equi)trek between $\{0, 1\}$ and $\{2\}$ in G, we conclude from Corollary 2.3.11 that

$$X_{\{0,1\}} \perp\!\!\!\perp X_2$$

holds for all distributions in $\mathcal{M}_G^{\leq 2}$. Based on the resulting separations in the corresponding equitrek graph $\hat{G}$, Lemma 2.3.12 further implies that

$$X_1 \perp\!\!\!\perp X_2 \mid X_0 \text{ and } X_0 \perp\!\!\!\perp X_2 \mid X_1$$

hold in all distributions in the model. To show that no other conditional independencies hold for all distributions in the model, we need to construct a distribution in the model where, for example,

$$X_1 \perp\!\!\!\perp X_3 \mid X_0$$

does not hold. This is equivalent to the corresponding determinant

$$\det \Sigma_{\{1,0\}\times\{3,0\}} = \frac{a_{10}a_{30}(\omega_0^{(2)})^2}{(1-a_{00}^2)(1-a_{00}a_{11})(1-a_{00}a_{33})(1-a_{11}a_{33})}$$

being non-zero. It is clear that for almost all choices of the parameters of A and D, this is fulfilled. Since we require the matrix A to be stable, an easy choice to construct such a distribution is setting the edge weights to 1, the self-loop parameters to $\frac{1}{2}$, and the error covariances to 1. Then we have $\det \Sigma_{\{1,0\}\times\{3,0\}} = \frac{256}{81}$, so $X_1 \not\perp\!\!\!\perp X_3 \mid X_0$ in the corresponding distribution. Even though this approach works in small dimensions, a general way of constructing counterexample distributions to arbitrary conditional independence statements is required to prove the conjecture.

Remark 2.3.14. Note that if all self-loops are present (or the graph is a DAG and at least all source nodes of the graph have a self-loop), the equitrek graph coincides with the trek graph known from the continuous setting. However, if some of the self-loops at the source nodes are missing, the equitrek graph is a sub-graph of the trek graph, thereby encoding additional independence statements.

2.4 Generic Identifiability

In this section we study identifiability of the parameters in the discrete Lyapunov model for a known graph G. Recall that the parameters $(A, \Omega^{(2)}, \Omega^{(3)}, \Omega^{(4)})$ of the fourth-order discrete Lyapunov model lie in the parameter space

$$\Theta_{G,\leq 4} = \mathbb{R}^E_{\text{stab}} \times \mathbb{R}^p_+ \times (\mathbb{R} \setminus \{0\})^p \times \mathbb{R}^p_+.$$

The question of identifiability for fourth-order discrete Lyapunov models is whether the (rational) map

$$\phi_{G,\leq 4} : \Theta_{G,\leq 4} \to PD_p \times \text{Sym}^3(\mathbb{R}^p) \times \text{Sym}^4(\mathbb{R}^p), \quad (A, \Omega^{(2)}, \Omega^{(3)}, \Omega^{(4)}) \mapsto (S, T, R)$$

is injective, where (S, T, R) denote the solutions to the corresponding second-, third- and fourth-order discrete Lyapunov equations. All of the identifiability results will initially be stated as identifying A since it is clear by considering (2.2.5) that if A can be identified from the cumulants (S, T, R), then we can also identify the error cumulants.

In Section 2.4.1 we establish generic (rational) identifiability for DAGs which have a self-loop at each vertex. In Section 2.4.2 we prove generic (rational) identifiability for polytrees which have self-loops at each source node. Finally, in Section 2.4.3 we present examples of some non-identifiable graphs.

2.4.1 Identifiability for DAGs With All Self-Loops

We first study generic identifiability for directed graphs G that have a self-loop at each vertex and are otherwise acyclic. We refer to such graphs as *DAGs with all self-loops*. Considering the definition of the VAR(1) process, the assumption of all the vertices having a self-loop is a reasonable one to make as the it corresponds to the value of the variable at time t affecting its value at time $t + 1$. This is why we initially consider this case.

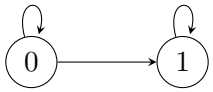


Figure 2.4.3: A graph on two vertices with two self-loops.

Example 2.4.1. Consider the graph G in Fig. 2.4.3 on two nodes, which serves as the base case for the inductive proof of generic identifiability for DAGs with all self-loops in Theorem 2.4.3.

By employing the trek rule, we obtain the following equations for the second-, third- and fourth-order cumulants, denoted by s, t, and r with respective subscripts, and the edge weight parameters a_{00}, a_{10} and a_{11}:

$$\frac{s_{01}}{s_{00}} = \frac{a_{00}a_{10}}{1 - a_{00}a_{11}}, \qquad \frac{t_{001}}{t_{000}} = \frac{a_{00}^2 a_{10}}{1 - a_{00}^2 a_{11}}, \qquad \frac{r_{0001}}{r_{0000}} = \frac{a_{00}^3 a_{10}}{1 - a_{00}^3 a_{11}}.$$

These equations can be solved uniquely for a_{00}, a_{10} and a_{11}, yielding

$$\begin{aligned}
a_{00} &= \frac{-r_{0001}(s_{00}t_{001} - s_{01}t_{000})}{s_{01}(r_{0000}t_{001} - r_{0001}t_{000})}, \\
a_{10} &= \frac{-s_{01}^2 t_{001}(r_{0000}s_{01}t_{001} + r_{0001}s_{00}t_{001} - 2r_{0001}s_{01}t_{000})(r_{0000}t_{001} - r_{0001}t_{000})}{r_{0001}^2(s_{00}t_{001} - s_{01}t_{000})^3}, \\
a_{11} &= \frac{(r_{0000}t_{001} - r_{0001}t_{000})s_{01}^2(r_{0000}s_{00}t_{001}^2 - r_{0001}s_{01}t_{000}^2)}{(s_{00}t_{001} - s_{01}t_{000})^3 r_{0001}^2}.
\end{aligned}$$

By employing the trek-rule once more, we see that the denominators above can never be zero if we assume the following: A is Schur stable, $a_{00} \neq 0, a_{10} \neq 0$, and $\omega_0^{(2)} \neq 0, \omega_0^{(3)} \neq 0, \omega_0^{(4)} \neq 0$. Thus, we conclude that G is generically identifiable from the second-, third- and fourth-order cumulants.

Alternatively, it is possible to replace the fourth-order equation by another third-order equation involving t_{011}/t_{000}. In this case, there are two solutions, implying that G is locally identifiable from the second- and third-order discrete Lyapunov equations alone. From experiments it furthermore seems to be the case that only one of these solutions yields a Schur stable matrix A. If this could be proven, then A would also be generically identifiable from just the second-and third-order cumulants.

Remark 2.4.2. In the proofs of this section, we encounter the situation where we have several (potentially) different expressions for the same parameter a_{ij} in terms of the cumulants of the model. However, all these equations are directly derived from the model definition and we further only obtain a single solution for a_{ij} each time. Consequently, they all have to agree since otherwise, the corresponding original equation would not have been true in the model.

Theorem 2.4.3. *Let $G = (V, E)$ be a DAG with $p \geq 2$ nodes that has no isolated nodes and a non-zero self-loop at each node. Consider the corresponding discrete Lyapunov model with edge weight matrix A. If G is known (i.e., the zero pattern of A is known), then A is generically identifiable.*

The proof can be found in Appendix 2.B. Once we have identified the matrix A, we can easily identify all cumulants of the noise, for instance by using the recursive equation (2.2.5).

Remark 2.4.4. A graph which has an isolated node with a self-loop is neither generically nor locally identifiable using any number of cumulants since then the isolated node would itself have to be generically/locally identifiable. This, however, is not possible since there will always be one more parameter than available equations.

Corollary 2.4.5. *Let $G = (V, E)$ be a DAG with $p \geq 2$ nodes that has no isolated nodes and a non-zero self-loop at each node. Consider the corresponding discrete Lyapunov model with edge weight matrix A. If G is known (i.e., the zero pattern of A is known), then A, $\Omega^{(2)}$, $\Omega^{(3)}$, and $\Omega^{(4)}$ are generically identifiable.*

2.4.2 Other Types of DAGs

We now consider graphs which might not have all self-loops present but are otherwise still acyclic. For a variable to not have a self-loop implies that its value at time t does not affect it at time $t + 1$, this could for example be the case of the time jumps are big compared to the speed of the process.

Example 2.4.6. Consider the graph on two nodes that has a directed edge from the source, 0, to the other node, 1, but only the source has a self-loop, as shown in Fig. 2.2.1. By employing the trek rule, as well as directly considering the discrete Lyapunov equations, we obtain the following equations for the second- and third-order cumulants:

$$\begin{aligned} s_{00} &= w_0^{(2)} \frac{1}{1 - a_{00}^2}, & t_{000} &= w_0^{(3)} \frac{1}{1 - a_{00}^3}, \\ s_{01} &= a_{00} a_{10} s_{00}, & t_{001} &= a_{00}^2 a_{10} t_{000}, \\ s_{11} &= a_{10}^2 s_{00} + w_1^{(2)}, & t_{111} &= a_{10}^3 t_{000} + w_1^{(3)}. \end{aligned}$$

The third-order cumulant entry t_{011} follows a similar pattern, but is not required for the following computation. Combining these equations yields

$$\begin{aligned} a_{00} &= \frac{s_{00} t_{001}}{s_{01} t_{000}}, & a_{10} &= \frac{s_{01}}{s_{00} a_{00}} = \frac{s_{01}^2 t_{000}}{s_{00}^2 t_{001}}, \\ w_0^{(2)} &= s_{00}(1 - a_{00}^2) = \frac{s_{00} s_{01}^2 t_{000}^2 - s_{00}^3 t_{001}^2}{s_{01}^2 t_{000}^2}, & w_1^{(2)} &= s_{11} - a_{10}^2 s_{00} = \frac{s_{00}^3 s_{11} t_{001}^2 - s_{01}^4 t_{000}^2}{s_{00}^4 t_{001}^2}, \\ w_0^{(3)} &= t_{000}(1 - a_{00}^3) = \frac{s_{00} s_{01}^3 t_{000}^3 - s_{00}^4 t_{001}^3}{s_{01}^3 t_{000}^3}, & w_1^{(3)} &= t_{111} - a_{10}^3 t_{000} = \frac{t_{001}^3 t_{111} s_{00}^6 - s_{01}^6 t_{000}^4}{s_{00}^6 t_{001}^3}. \end{aligned}$$

This example serves as the base case for the following theorem.

Theorem 2.4.7. *Let $G = (V, E)$ be a polytree with $p \geq 2$ nodes that has no isolated nodes and non-zero self-loops at all sources. Consider the corresponding discrete Lyapunov model with edge weight matrix A. If G is known (i.e., the zero pattern of A is known), then A is generically identifiable.*

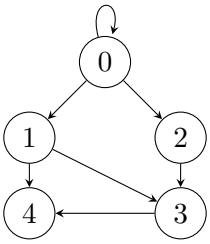


Figure 2.4.4: A generically identifiable graph.

The proof can be found in Appendix 2.B. As in Theorem 2.4.3, it is easy to identify the cumulants of the noise once A is identified.

Corollary 2.4.8. *Let $G = (V, E)$ be a polytree with $p \geq 2$ nodes that has no isolated nodes and has a non-zero self-loop at all the sources. Consider the corresponding discrete Lyapunov model with edge weight matrix A. If G is known (i.e. the zero pattern of A is known), then A, $\Omega^{(2)}$, $\Omega^{(3)}$, and $\Omega^{(4)}$ are generically identifiable.*

Remark 2.4.9. The motivation behind Theorem 2.4.7 was to explore identifiability results in less restrictive cases than those in Theorem 2.4.3. One natural generalization is to consider graphs where some self-loops are missing. Note, that sources must still have self-loops, see Example 2.4.11. However, merely requiring this is not sufficient to guarantee identifiability, see Example 2.4.15.

In the proofs of Theorem 2.4.7 and Theorem 2.4.3, we rely on the following: in the induction step, for each vertex, we are able to find linearly independent equations that identify the incoming edges. Having all the self-loops in a graph, or assuming the graph is a tree, ensures this property. However, there are alternative ways to formulate this condition.

For example, the following statement could be proved using the same technique: Let $G = (V, E)$ be a DAG with $p \geq 2$ nodes, no isolated nodes, and self-loops at all sources. Further, for each non-source node i with parents $\mathrm{pa}(i) = \{v_1, \ldots, v_d\}$, suppose there exist d nodes $\{u_1, \ldots, u_d\}$ that appear before i in the topological order, such that u_i is an ancestor of v_i (possibly equal if v_i has a self-loop), and u_i is not an ancestor of v_j for $i \neq j$, and $u_i \neq v_j$ for $j \neq i$. Consider the corresponding discrete Lyapunov model with edge weight matrix A. If G is known, then A is generically identifiable. In the proof, one would need to consider equations for $s_{iu_1}, \ldots, s_{i,u_d}$ to identify the edges $a_{iv_1}, \ldots, a_{iv_d}$.

Finally, note that there exist graphs not satisfying the conditions above that are still identifiable, see Example 2.4.10. Thus, a natural direction for future research is to completely characterize the class of generically identifiable graphs.

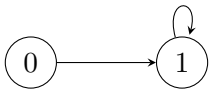
Example 2.4.10. The graph in Fig. 2.4.4 does not satisfy the conditions from Theorem 2.4.3, Theorem 2.4.7, or Remark 2.4.9. Nevertheless, the parameters can still be generically identified using $s_{04}, t_{004}, s_{03}, s_{43}$.

2.4.3 Nonidentifiable Graphs

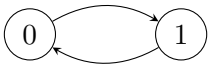
In this subsection, we present examples of small graphs that are not identifiable.

2.4.3.1 Examples on Two Nodes

We consider graphs on two nodes with various edge configurations, including graphs in which not all self-loops are present (in particular, the self-loop at the source) and graphs that contain a cycle, as shown in Fig. 2.4.5.



(a) A directed graph on two vertices with a single self-loop at the sink.



(b) A cyclic directed graph on two vertices with no self-loops.

Figure 2.4.5: Non-identifiable directed graphs on two nodes.

Example 2.4.11. We consider the graph in Fig. 2.4.5a. The unknown parameters are a_{10}, a_{11}, and $w_0^{(2)}$, $w_1^{(2)}, \ldots, w_0^{(n)}, w_1^{(n)}$, when considering cumulants up to order n. We obtain the following equations:

$$\begin{array}{lll} s_{00} = w_0^{(2)}, & t_{000} = w_0^{(3)}, & q_{0\ldots00} = w_0^{(n)}, \\ & t_{001} = 0, & q_{0\ldots01} = 0, \\ s_{01} = 0, & & \cdots \quad \vdots \\ & t_{011} = 0, & q_{01\ldots1} = 0, \\ s_{11} = \dfrac{w_0^{(2)} a_{10}^2 + w_1^{(2)}}{1 - a_{11}^2}, & t_{111} = \dfrac{w_0^{(3)} a_{10}^3 + w_1^{(3)}}{1 - a_{11}^3}, & q_{11\ldots1} = \dfrac{w_0^{(n)} a_{10}^n + w_1^{(n)}}{1 - a_{11}^n}, \end{array}$$

where the last equations describe the nth order cumulants. Note that there are no equitreks between 0 and 1, so the respective cumulant entries are zero. Consequently, there are $2(n-1)$ equations and $2(n-1)+2$ unknowns. Therefore, the parameters cannot be identified using these equations for any n. The same issue arises if the graph has one edge and no self-loops.

Example 2.4.12. We consider the cyclic graph with two edges in Fig. 2.4.5b. The unknown parameters are a_{10}, a_{01}, and $w_0^{(2)}, w_1^{(2)}, \ldots, w_0^{(n)}, w_1^{(n)}$, when considering cumulants up to order n. We obtain the following equations:

$$\begin{array}{llll} s_{00} = \dfrac{w_0^{(2)} + w_1^{(2)} a_{01}^2}{1 - a_{10}^2 a_{01}^2}, & t_{000} = \dfrac{w_0^{(3)} + w_1^{(3)} a_{01}^3}{1 - a_{10}^3 a_{01}^3}, & \cdots & q_{0\ldots0} = \dfrac{w_0^{(n)} + w_1^{(n)} a_{01}^n}{1 - a_{10}^n a_{01}^n}, \\ s_{11} = \dfrac{w_0^{(2)} a_{10}^2 + w_1^{(2)}}{1 - a_{10}^2 a_{01}^2}, & t_{111} = \dfrac{w_0^{(3)} a_{10}^3 + w_1^{(3)}}{1 - a_{10}^3 a_{01}^3}, & \cdots & q_{1\ldots1} = \dfrac{w_0^{(n)} a_{10}^n + w_1^{(n)}}{1 - a_{10}^n a_{01}^n}, \end{array}$$

where the last equations describe the nth order cumulants. Again, there are no equitreks between 0 and 1, so the respective cumulant entries are zero. Consequently, there are again $2(n-1)$ equations and $2(n-1)+2$ unknowns, making this graph non-identifiable as well.

Remark 2.4.13. These examples illustrate a more general observation: The number of nth order cumulants of a p-dimensional random vector is $\binom{p+n-1}{n}$, so in the nth order discrete Lyapunov model, we have at most $\sum_{i=2}^{n}\binom{i+p-1}{i} = \binom{p+n}{n} - (p+1)$ equations. Thus, if the model is generically identifiable, we have that $|E|+p(n-1) \leq \binom{p+n}{n}-(p+1)$. This bound can be further tightened by subtracting the number of zero cumulants due to missing i-equitreks for $i = 2, \ldots, n$ in the graph from the right-hand side (see, e.g., Dettling et al. [2023]). In the examples above, this yields $|E| + 2(n-1) \leq 2(n-1)$. If there are no or only few equitreks missing, however, this bound does not present a notable restriction.

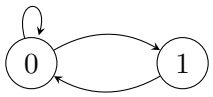


Figure 2.4.6: A cyclic graph on two vertices with one self-loop.

Example 2.4.14. We consider the cyclic graph with three edges in Fig. 2.4.6. The unknown parameters are a_{00}, a_{10}, a_{01}, and $w_0^{(2)}, w_1^{(2)}, \ldots, w_0^{(n)}, w_1^{(n)}$, when considering the cumulants up to order n. For this edge configuration, the parameters are locally identifiable from second- and third-order cumulants.

2.4.3.2 Diamond-type Graphs

There are also examples of non-identifiable graphs that do not arise from a lack of equations compared to the number of parameters.

Example 2.4.15. We consider the graph on four vertices shown in Fig. 2.4.7.

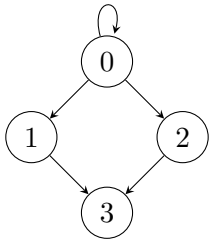


Figure 2.4.7: A diamond-type graph.

Since the subgraph induced by $0, 1, 2$ satisfies the conditions of Theorem 2.4.7, the parameters a_{00}, a_{10} and a_{20} are identifiable from the second-, third- and fourth-order

cumulants (see Example 2.4.6). We also have

$$s_{03} = w_0^{(2)} \frac{a_{10}a_{31} + a_{20}a_{32}}{1 - a_{00}^2}.$$

Hence, the quantity $y = a_{10}a_{31} + a_{20}a_{32}$ is identifiable from these cumulants. Note that this is a linear combination of the two unknown values a_{31} and a_{32} with known coefficients. However, we claim that a_{31} and a_{32} are not identifiable from any cumulants, regardless of how high the order is allowed to be.

Let $q_{0^{k_0}1^{k_1}2^{k_2}3^{k_3}}$ denote the ith order cumulant with k_j copies of j in its subscript for $j = 0, \ldots, 3$, where $i = k_0 + \cdots + k_3$. Note that a_{31} or a_{32} appears in the trek parametrization of this cumulant if and only if $k_3 \neq 0$. In this case, for the off-diagonal cumulants (i.e., those with $k_3 \neq i$), we have

$$q_{0^{k_0}1^{k_1}2^{k_2}3^{k_3}} = w_0^{(i)} \frac{a_{00}^{2k_0}(a_{00}a_{10})^{k_1}(a_{00}a_{20})^{k_2}y^{k_3}}{1 - a_{00}^i}.$$

All ith order cumulants with $k_3 = 0$ are rational functions only of a_{00}, a_{10} and a_{20}.

Let A be any parameter matrix consistent with the prescribed cumulants of all orders, and let $\hat{A}$ be any parameter matrix which satisfies

$$\hat{a}_{00} = a_{00}, \qquad \hat{a}_{10} = a_{20}, \qquad \hat{a}_{20} = a_{20}, \qquad a_{10}\hat{a}_{31} + a_{20}\hat{a}_{32} = a_{10}a_{31} + a_{20}a_{32}.$$

Then $\hat{A}$ has the same prescribed cumulants as A. Therefore, there is a line of parameters that give the same cumulants, yielding this model non-identifiable. Note that in the context of the proof of Theorem 2.4.3, this property results in the product of submatrices of the A and S in (2.B.2) having rank 1.

2.5 Local Identifiability Using the Jacobian

If the Jacobian of the parametrization has full rank, then the parametrization is locally identifiable [Sullivant, 2018, Section 16.1]. In this section, we derive an expression for the Jacobian of both the second- and third-order cumulants of the discrete Lyapunov model. We show that, for any graph where all connected components have at least size three and all self-loops are present, the joint second- and third-order Jacobian has full rank. This proves that all discrete Lyapunov models for such a graph are locally identifiable. By further including the Jacobian of the fourth-order cumulant we can prove that any graph with all self-loops and no isolated nodes are locally identifiable.

As was seen in Example 2.4.1, the graph on two nodes with both self-loops and an edge from 0 to 2 is generically identifiable from second-, third-, and fourth-order cumulants. However, it is already two-to-one identifiable (so locally identifiable) from the second- and third-order cumulants alone. It is this phenomenon which we will generalize in the section.

Fix a directed graph G on p vertices labeled from 0 to $p-1$. Let $\phi_{G,\leq 3}$ be the map which sends a tuple of parameters $(A, \omega^{(2)}, \omega^{(3)})$ to the corresponding second- and

third-order cumulants (S, T) under the discrete Lyapunov model specified by G, so the third-order version of $\phi_{G,\leq 4}$ from Section 2.4. Let $\mathrm{Jac}^G(S,T)$ denote the Jacobian of $\varphi_{G,\leq 3}$.

We derive the following expression for the column of the Jacobian corresponding to the parameter $a_{\beta\alpha}$ using the Kronecker parametrization, equation (2.2.6). By applying standard rules for differentiating a matrix inverse, which can be found for example in [Petersen et al., 2008, Section 2.2], we obtain

$$\begin{aligned}\mathrm{vec}\left(\frac{\partial S}{\partial a_{\beta\alpha}}\right) &= -(I - A\otimes A)^{-1}\frac{\partial(I - A\otimes A)}{\partial a_{\beta\alpha}}(I - A\otimes A)^{-1}\mathrm{vec}(\Omega^{(2)}) \\ &= (I - A\otimes A)^{-1}(E_{\beta\alpha}\otimes A + A\otimes E_{\beta\alpha})(I - A\otimes A)^{-1}\mathrm{vec}(\Omega^{(2)}), \end{aligned} \tag{2.5.9}$$

where $E_{\beta\alpha}$ is the matrix that is 1 at the (β,α)th position and 0 elsewhere. Note that the product of the last two matrices in (2.5.9) is exactly $\mathrm{vec}(S)$ as given by the Kronecker product parametrization.

The column corresponding to the parameter $\omega_i^{(2)}$ is

$$\mathrm{vec}\left(\frac{\partial S}{\partial \omega_i^{(2)}}\right) = (I - A\otimes A)^{-1}\frac{\partial \mathrm{vec}(\Omega^{(2)})}{\partial \omega_i^{(2)}} = (I - A\otimes A)^{-1}\mathrm{vec}(E_{ii}).$$

Similarly, we derive the following expression for the Jacobian of the third-order cumulant with respect to the entries of A:

$$\begin{aligned}\mathrm{vec}\left(\frac{\partial T}{\partial a_{\beta\alpha}}\right) &= -(I - A\otimes A\otimes A)^{-1}\frac{\partial(I - A\otimes A\otimes A)}{\partial a_{\beta\alpha}}(I - A\otimes A\otimes A)^{-1}\mathrm{vec}(\Omega^{(3)}) \\ &= (I - A\otimes A\otimes A)^{-1}(E_{\beta\alpha}\otimes A\otimes A + A\otimes E_{\beta\alpha}\otimes A \\ &\quad + A\otimes A\otimes E_{\beta\alpha})(I - A\otimes A\otimes A)^{-1}\mathrm{vec}(\Omega^{(3)}) \end{aligned} \tag{2.5.10}$$

and with respect to the $\omega_i^{(3)}$'s

$$\mathrm{vec}\left(\frac{\partial T}{\partial \omega_i^{(3)}}\right) = (I - A\otimes A\otimes A)^{-1}\frac{\partial \mathrm{vec}(\Omega^{(3)})}{\partial \omega_i^{(3)}} = (I - A\otimes A\otimes A)^{-1}\mathrm{vec}(E_{iii}).$$

From the above derivative computations, we observe that $J(S,T)$ can be written as a product of block matrices:

$$\mathrm{Jac}^G(S,T) = \begin{pmatrix}(I - A\otimes A)^{-1} & 0 \\ 0 & (I - A\otimes A\otimes A)^{-1}\end{pmatrix}\begin{pmatrix}J_2^G(S,a) & J_2^G(S,\omega^{(2)}) & 0 \\ J_3^G(T,a) & 0 & J_3^G(T,\omega^{(3)})\end{pmatrix}. \tag{2.5.11}$$

We define the blocks of this second matrix as follows. The matrix $J_2^G(S,a)$ is a $\binom{p+1}{2}\times \#E$ matrix whose $\beta\alpha$-th column (i.e., the column corresponding to the edge $\alpha\to\beta$) is

$$(E_{\beta\alpha}\otimes A + A\otimes E_{\beta\alpha})(I - A\otimes A)^{-1}\mathrm{vec}(\Omega^{(2)}).$$

Similarly, $J_3^G(S,A)$ is the $\binom{p+2}{3} \times \#E$ matrix whose $\beta\alpha$-th column is

$$(E_{\beta\alpha} \otimes A \otimes A + A \otimes E_{\beta\alpha} \otimes A + A \otimes A \otimes E_{\beta\alpha})(I - A \otimes A \otimes A)^{-1}\mathrm{vec}(\Omega^{(3)}).$$

The matrix $J_2^G(S,\omega^{(2)})$ is the $\binom{p+1}{2} \times p$ matrix whose i-th column is $\mathrm{vec}(E_{ii})$. Finally, $J_3^G(S,\omega^{(3)})$ is the $\binom{p+2}{3} \times p$ matrix whose i-th column is $\mathrm{vec}(E_{iii})$.

Remark 2.5.1. If we want the image of φ_G to contain cumulants of order higher than 3 as well, analogous formulas can be derived for the rows of the Jacobian corresponding to these higher-order cumulants.

Definition 2.5.2. *Let G be a directed graph and let $A \in \mathbb{R}^E$. The* modified Jacobian *of the second- and third-order cumulants of the discrete Lyapunov model, denoted by $\mathcal{J}^G(S,T)$, is defined by*

$$\mathcal{J}^G(S,T) = \begin{pmatrix} J_2^G(S,a) & J_2^G(S,\omega^{(2)}) & 0 \\ J_3^G(T,a) & 0 & J_3^G(T,\omega^{(3)}) \end{pmatrix},$$

as described in (2.5.11).

Since the original Jacobian is the modified Jacobian multiplied by an invertible block matrix from the left, their ranks are equal. Since the entries of the modified Jacobian will turn out to have a simpler description (see Proposition 2.5.3 and Proposition 2.5.4), we will determine its rank instead.

We notice that the last $2p$ columns of the modified Jacobian are in fact unit vectors, so by reordering the rows of the modified Jacobian its rank decomposes as

$$\mathrm{rank}(\mathcal{J}^G(S,T)) = \mathrm{rank}\begin{pmatrix} J_2^G(S,a)_{\mathrm{off}} & 0 & 0 \\ J_3^G(T,a)_{\mathrm{off}} & 0 & 0 \\ J_2^G(S,a)_{\mathrm{diag}} & I_p & 0 \\ J_3^G(T,a)_{\mathrm{diag}} & 0 & I_p \end{pmatrix} = 2p + \mathrm{rank}\begin{pmatrix} J_2^G(S,a)_{\mathrm{off}} \\ J_3^G(T,a)_{\mathrm{off}} \end{pmatrix}, \quad (2.5.12)$$

where we denote the submatrix whose rows correspond to diagonal entries of the cumulants with 'diag' and the submatrix whose rows correspond to off-diagonal entries of the cumulants with 'off'. Thus,

$$\mathcal{J}_{\mathrm{off}}^G(S,T) = \begin{pmatrix} J_2^G(S,a)_{\mathrm{off}} \\ J_3^G(T,a)_{\mathrm{off}} \end{pmatrix}, \quad (2.5.13)$$

has full rank if and only if the modified Jacobian has full rank. We will refer to this matrix as the off-diagonal modified Jacobian. In order to be able to show that the modified Jacobian has full rank we first show the following two results describing its entries, with proofs provided in Appendix 2.C.

Proposition 2.5.3. *The entries of $J_2^G(S,a)$ are given by*

$$J_2^G(S,a)_{(ij),\alpha\to\beta} = \delta_j(\beta)\sum_{l=0}^{p-1} a_{il}s_{l\alpha} + \delta_i(\beta)\sum_{k=0}^{p-1} a_{jk}s_{k\alpha}.$$

Furthermore, the description can be given by employing treks in G:

$$J_2^G(S,a)_{(ij),\alpha\to\beta} = \delta_j(\beta) \sum_{\tau\in\mathcal{T}_{(1,0)}(i,\alpha)} m_\tau + \delta_i(\beta) \sum_{\tau\in\mathcal{T}_{(1,0)}(j,\alpha)} m_\tau,$$

where $\mathcal{T}_{(1,0)}(i,j)$ denotes the set of treks between i and j where the path going to i is one longer than the path going to j, and m_τ is the trek monomial for a trek τ.

Proposition 2.5.4. *The entries of $J_3^G(T,a)$ are given by*

$$J_3^G(T,a)_{(ijk),\alpha\to\beta} = \delta_i(\beta)\sum_{m=0}^{p-1}\sum_{n=0}^{p-1} a_{jm}a_{kn}t_{\alpha mn} + \delta_j(\beta)\sum_{l=0}^{p-1}\sum_{n=0}^{p-1} a_{il}a_{kn}t_{l\alpha n} + \delta_k(\beta)\sum_{l=0}^{p-1}\sum_{m=0}^{p-1} a_{il}a_{jm}t_{lm\alpha}.$$

Furthermore, the description can be given by employing treks in G:

$$J_3^G(T,a)_{(ijk),\alpha\to\beta} = \delta_i(\beta) \sum_{\tau\in\mathcal{T}_{(1,1,0)}(j,k,\alpha)} m_\tau + \delta_j(\beta) \sum_{\tau\in\mathcal{T}_{(1,1,0)}(i,k,\alpha)} m_\tau + \delta_k(\beta) \sum_{\tau\in\mathcal{T}_{(1,1,0)}(i,j,\alpha)} m_\tau,$$

where $\mathcal{T}_{(1,1,0)}(i,j,k)$ denotes the set of all treks between i,j and k where the paths going to i and j are one longer than the path going to k, and m_τ is the trek monomial for a trek τ.

Remark 2.5.5. Similar formulas for the modified Jacobian can be derived analogously for higher-order cumulants.

We are now ready to state the main result of this section.

Theorem 2.5.6. *Let $G=(V,E)$ be any connected directed graph with all self-loops on $p \geq 3$ nodes, then $\mathcal{J}_{\text{off}}^G(S,T)$ has full rank ($= |E|$) generically. Therefore, the corresponding discrete Lyapunov model is locally identifiable from its second- and third-order cumulants.*

The proof of Theorem 2.5.6 (and its modifications) proceed by using the local identifiability of models associated to subgraphs of G to recursively prove that the model associated to G itself is locally identifiable. In particular, we consider subgraphs obtained by edge-disconnecting G.

Definition 2.5.7. *Let $G=(V,E)$ be a connected graph. A subset $F \subset E$* edge-disconnects *G if the graph $(V, E\setminus F)$ is disconnected. The set $\Xi(G)$ is the set of all graphs of the form $(V, E\setminus F)$ with exactly two connected components.*

Specifically, the proof of Theorem 2.5.6 proceeds by induction on the number of nodes, where the induction step relies on being able to edge-disconnect the graph into two connected components of size at least three. Therefore, the base case includes all connected graphs on three, four or five nodes and all graphs on at least six nodes where it is not possible to disconnect the graph into two connected components of at least size three.

Therefore, to understand the base, we must characterize the connected graphs where no matter how the graph is edge-disconnected a connected component of size one or

two is always created. Note that the self-loops do not impact this property, so they will be omitted in the following discussion. Furthermore, this property is completely characterized by the skeleton of the graph and therefore the description will be given for undirected graphs.

It is possible to describe the base cases by considering the intersections of paths in the graph. For example, let G be a graph such that every graph in $\Xi(G)$ has an isolated node. Then every pair of paths in G of vertex length two intersect at a single vertex. In other words, there is a vertex v that is in every edge, so the skeleton of G is a star, as pictured in Fig. 2.5.8. Similarly, consider the graphs such that every graph in $\Xi(G)$ has a connected component with exactly two vertices. In this case, every pair of paths of vertex length three must intersect in at least one vertex. In Lemma 2.C.1, we show that if G has six or more vertices, these graphs are the so-called *generalized 2 stars*.

Definition 2.5.8. *Let $G = (V, E)$ be a connected undirected graph with $|V| \geq 6$. Then G is a* generalized two star *if all paths of vertex length 3 intersect in the same node. We call this node the* center *of the graph.*

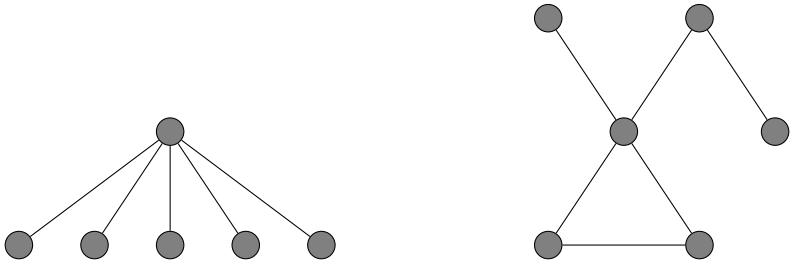


Figure 2.5.8: Examples of a star (left) and a generalized two star (right).

Due to the block structure of the Jacobian, the statement and proof of Theorem 2.5.6 can be adapted to directed graphs with several connected components. Furthermore, connected components of size two are allowed as long as they are not the complete graph on two nodes, since they are also locally identifiable from the second- and third-order cumulants, see Example 2.4.1 or the calculations performed in Maple[1].

Corollary 2.5.9. *Let $G = (V, E)$ be any directed graph with all self-loops. If all connected components of G have at least size two and the components of size two are not the two-cyle then G is generically locally identifiable from the second- and third-order cumulants.*

The restrictions on the size of the connected components of the graph are motivated by non-identifiability of the one-node graph (with a self-loop) and the two-cycle on two nodes (with two self-loops) from just the second- and third-order cumulants. A graph

[1] https://github.com/cecilie2424/Local-Identifiability-in-Non-Gaussian-Discrete-Lyapunov-Models.

which has an isolated node with a self-loop is not locally identifiable using any number of cumulants as discussed in Remark 2.4.4. The two-cycle can only be locally identified by also employing some fourth-order cumulants.

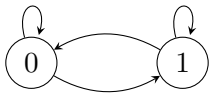


Figure 2.5.9: The complete graph on two nodes.

Example 2.5.10. For the two-cycle on two nodes, the second- and third-order cumulants are not enough to obtain full rank of the modified Jacobian, since there will be one more parameter than there are equations. However, it is shown computationally in Maple[1] that, by adding the two diagonal fourth-order cumulants, the two $\omega^{(4)}$'s and one off-diagonal fourth-order cumulant (so in total using three fourth-order cumulants) we can prove that the modified Jacobian of both the second-, third- and the three fourth-order cumulants will have full rank. Therefore, the actual Jacobian will also have full rank.

From Example 2.5.10 it follows that the proof of Theorem 2.5.6 could be modified to also allow for the use of fourth-order cumulants which would then yield local identifiability of discrete Lyapunov models for any graph which has all self-loops and no isolated nodes. Thus, it includes the few graphs not covered by Corollary 2.5.9, i.e. the graphs which have a connected component which is the complete graph on two nodes. However, all other graphs were already included in Corollary 2.5.9. We make this distinction of when the fourth-order cumulants are needed because in practice it is difficult to estimate higher-order cumulants accurately, and the difficulty grows with the order of the cumulant. Therefore, it is desirable to obtain the identifiability result from the lowest possible order of cumulants

Corollary 2.5.11. *Let $G = (V, E)$ be any directed graph with all self-loops on $p \geq 2$ nodes with no isolated nodes, then $\mathcal{J}^G_{\text{off}}(S, T)$ has full rank ($= |E|$) generically. It follows that the corresponding discrete Lyapunov model is locally identifiable from its second-, third-, and fourth-order cumulants.*

Proof of Theorem 2.5.6. The proof proceeds by induction on the number of nodes. The base case is any graph on three, four, or five nodes, and all graphs with $|V| \geq 6$ whose skeleton is a star or generalized two star, as explained in the preceding paragraph. The base cases are covered by Lemmas 2.C.2 and 2.C.4.

Let $G = (V, E)$ be a graph on $p \geq 6$ nodes, where $V = \{0\} \cup [p-1]$. If G is a star or generalized two star, then G is covered by the base case. Therefore, we now assume this is not the case. This implies that G can be edge-disconnected into exactly two connected components where each component has at least three vertices. Let $G' = (V, E')$ denote such an edge-disconnecting, and let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be

the two connected components of G'. Further, let $|V_i| = n_i$ for $i = 1, 2$, and write the vertices as $V_i = \{g_1^i, \dots, g_{n_i}^i\}$. We now let the sparsity pattern of A be given by G'.

Let $\mathcal{J}_{\text{off}}^{G_i}(S, T)$ denote the columns of the modified Jacobian corresponding to G_i. The induction hypothesis then implies that $\mathcal{J}_{\text{off}}^{G_1}(S, T)$ and $\mathcal{J}_{\text{off}}^{G_2}(S, T)$ both have full rank generically.

We aim to show that $\mathcal{J}_{\text{off}}^{G}(S, T)$ has generically full rank. To simplify the notation, we construct a matrix X that contains $\mathcal{J}_{\text{off}}^{G}(S, T)$ as a submatrix. The only difference is that X will (potentially) include columns corresponding to edges that are missing in G.

The rows and columns of X will be partitioned into four groups. The first two groups correspond to submatrices of $\mathcal{J}_{\text{off}}^{G_1}(S, T)$ and $\mathcal{J}_{\text{off}}^{G_2}(S, T)$, respectively. The rows are chosen so that the submatrix has full rank, which is possible by the induction hypothesis. Specifically, the rows are

$$R_{G_1} = \{(i, j) \mid i, j \in V_1, i \neq j\} \cup \{(i, j, k)\ i, j, k \in V_1, i \neq j \neq k\},$$
$$R_{G_2} = \{(i, j) \mid i, j \in V_2, i \neq j\} \cup \{(i, j, k)\ i, j, k \in V_2, i \neq j \neq k\},$$

and the columns correspond to the edges in each connected component

$$C_{G_1} = \{\alpha \to \beta \in E_1\} \qquad \text{and} \qquad C_{G_2} = \{\alpha \to \beta \in E_2\},$$

respectively. We denote these two matrices by $\mathcal{J}^{G_1}$ and $\mathcal{J}^{G_2}$. The remaining blocks correspond to potential edges between the connected components. For $j = 1, \dots, n_2$, consider the rows and columns

$$R_{G_1 \to g_j^2} = \{(g_k^1 g_k^1 g_j^2) \mid k = 1, \dots, n_1\}, \qquad C_{G_1 \to g_j^2} = \{(g_k^1 \to g_j^2) \mid k = 1, \dots, n_1\}.$$

Conversely, for $j = 1, \dots, n_1$, consider the rows and columns

$$R_{G_2 \to g_j^1} = \{(g_k^2 g_k^2 g_j^1) \mid k = 1, \dots, n_2\}, \qquad C_{G_2 \to g_j^1} = \{(g_k^2 \to g_j^1) \mid k = 1, \dots, n_2\}.$$

With this ordering of rows and columns, the matrix X has the following block structure:

$$\begin{array}{c} \\ R_{G_1} \\ R_{G_2} \\ R_{G_1 \to g_1^2} \\ \vdots \\ R_{G_1 \to g_{n_2}^2} \\ R_{G_2 \to g_2^1} \\ \vdots \\ R_{G_2 \to g_{n_1}^1} \end{array} \begin{array}{c} \begin{array}{cccccccc} C_{G_1} & C_{G_2} & C_{G_1 \to g_1^2} & \dots & C_{G_1 \to g_{n_2}^2} & C_{G_2 \to g_1^1} & \dots & C_{G_2 \to g_{n_1}^1} \end{array} \\ \begin{pmatrix} \mathcal{J}^{G_1} & & & & & & & \\ & \mathcal{J}^{G_2} & & & & & & \\ & & M_{G_1 \to g_1^2} & & & 0 & & \\ & & & \ddots & & & & \\ & & & & M_{G_1 \to g_{n_2}^2} & & & \\ & & 0 & & & M_{G_2 \to g_2^1} & & \\ & & & & & & \ddots & \\ & & & & & & & M_{G_2 \to g_{n_1}^1} \end{pmatrix} \end{array}.$$

We see that X is block diagonal by using Propositions 2.5.3 and 2.5.4 to infer the zero entries as follows. Generally, the zero entries arise either due to the sink of an edge not

matching the indices of the considered row or due to the missing treks between the two connected components.

Case 1: Zeros due to sinks not matching row indices.

First, in order for there to be a potential non-zero entry at position $(lm), \alpha \to \beta$ or $(klm), \alpha \to \beta$ in this matrix, β needs to equal either l or m, or k, l or m, respectively. This accounts for the zeros between rows R_{G_1} and columns with sink in G_2, namely columns in C_{G_2} and $C_{G_1 \to g_j^2}$. The same argument holds for rows R_{G_2} and columns in C_{G_1} and $C_{G_2 \to g_j^1}$.

Now consider the columns $C_{G_1 \to g_j^2}$ for $j = 1, \ldots, n_2$. For fixed j, entries in $C_{G_1 \to g_j^2}$ can only be non-zero if they correspond to $R_{G_1 \to g_j^2}$, indicated as blocks $M_{G_1 \to g_j^2}$ in the matrix. For all $i \neq j$, however, the entries of $C_{G_1 \to g_j^2}$ in $R_{G_1 \to g_i^2}$ are zero, since $g_j^2 \notin V_1 \cup \{g_i^2\}$. The same argument holds, if we flip the indices 1 and 2.

Case 2: Zeros due to missing treks.

We are now left with the cases where the sink β of an edge $\alpha \to \beta$ might match one of the indices (lm) or (klm). Here we use the existence (or non-existence) of treks to deduce the zero pattern. Specifically, we use the characterization of the entries of $J_2^G(S)$ and $J_3^G(T)$ based on treks where the length of the legs differ by one. If there is no match between β and (lm) or (klm), the first case applies.

Assume without loss of generality that $\beta = m$. For the corresponding entry in $J_2^G(S)$ to be non-zero there needs to exist a base trek between l and α (we can use self-loops to obtain treks where the length of one of the legs is off by one). Since the two connected components are disconnected in G', this is only possible if l and α are in the same connected component. Similarly, $J_3^G(T)$ can only be non-zero if, again without loss of generality, $\beta = m$ and there exists a base trek between k, l, and α. Thus, k, l, and α all need to belong to the same connected component.

Consequently, the remaining entries in the columns $C_{G_2 \to g_j^1}$ for each $j = 1, \ldots, n_1$ are zero in the rows R_{G_1}, $R_{G_1 \to g_1^2}, \ldots, R_{G_1 \to g_{n_2}^2}$, since $\alpha \in V_2$ is never in the same connected component as $k, l \in V_1$. The same argument holds for each $j = 1, \ldots, n_2$ for the remaining entries of the columns $C_{G_1 \to g_j^2}$ in the rows R_{G_2} and $R_{G_2 \to g_1^1}, \ldots, R_{G_2 \to g_{n_1}^1}$.

The remaining entries to consider are in the columns in C_{G_1} and C_{G_2}. In the columns in C_{G_1}, again the entries in $R_{G_2 \to g_1^1}, \ldots, R_{G_2 \to g_{n_1}^1}$ are zero, since $\alpha \in V_1$ and $k, l \in V_2$. Similarly, in the columns in C_{G_2}, the entries in $R_{G_1 \to g_1^2}, \ldots, R_{G_1 \to g_{n_2}^2}$ are zero, since $\alpha \in V_2$ and $k, l \in V_1$. Further, for the columns C_{G_1} and the rows $R_{G_1 \to g_j^2}$, the sink $\beta \in V_1$ will always coincide with at least one index of the rows. However, since one of the indices in (klm) will always be g_j^2, there would have to be a trek between β, g_j^2 and another element in G_1 for the entry to be non-zero, which is not the case. For all other indices, the first case again applies. The argument is analogous for the columns in C_{G_2}.

To show that X generically has full rank, it suffices to show that each block on the diagonal generically has full rank. The matrices $\mathcal{J}^{G_1}$ and $\mathcal{J}^{G_2}$ generically have full rank by the induction hypothesis. Employing Lemma 2.C.6, we further obtain that that each of the M matrices generically has full rank as well. Since the generic choices ensuring the M matrices have generically full rank do not depend on G, the submatrix $\mathcal{J}_{\text{off}}^G(S, T)$

will also have full rank. □

2.6 Defining Equations

In this section, we explore the vanishing ideal of the discrete Lyapunov model for specific families of graphs. First, in Section 2.6.1, we consider the case where the graph G is a polytree with a single source and only one self-loop. In this setting, the vanishing ideal is toric. In Section 2.6.2, we present several determinantal results which give rise to polynomials in the vanishing ideal for other types of graphs. In Section 2.6.3 we explore a different approach for obtaining polynomials in the vanishing ideal.

2.6.1 Polytrees With One Source

In this subsection, we focus on the case of G being a polytree with one source node (i.e., a directed tree) and a single self-loop at the source. For such a graph, the vanishing ideal $\mathcal{I}^{\leq n}(G)$ is toric for any $n \geq 2$. The model admits a monomial parametrization given by the *shortest* equitreks. In a directred tree with a single self-loop at the source, there is a unique shortest equitrek between any set of nodes.

Let $G = (V, E)$ be such a polytree, where $V = \{0\} \cup [p-1]$ and 0 is the unique source node. For the nth order model, we introduce new parameters $v_i^{(2)}, v_i^{(3)}, \ldots, v_i^{(n)}$ for each $i \in V$, such that

$$v_i^{(2)} = w_0^{(2)} \frac{\left(a^{\lambda_i}\right)^2}{1 - a_{00}^2}, \qquad v_i^{(3)} = w_0^{(3)} \frac{\left(a^{\lambda_i}\right)^3}{1 - a_{00}^3}, \qquad \ldots \qquad v_i^{(n)} = w_0^{(n)} \frac{\left(a^{\lambda_i}\right)^n}{1 - a_{00}^n},$$

where λ_i is the unique shortest path from the source to vertex i, and a^{λ_i} is the corresponding path polynomial. This invertible change of parameters allows us to derive the *shortest equitrek rule*.

For vertices $i_1, \ldots, i_l$, let $\tau(i_1, \ldots, i_l)$ denote the unique shortest l-equitrek between $i_1, \ldots, i_l$ consisting of paths $\tau_1, \ldots, \tau_l$. Let $top(\tau)$ denote the top vertex of the trek τ. Using the new v parameters, we can write the second- and third-order cumulants as

$$s_{i_1,i_2} = v^{(2)}_{top(\tau(i_1,i_2))} a^{\tau_1} a^{\tau_2}, \qquad t_{i_1,i_2,i_3} = v^{(3)}_{top(\tau(i_1,i_2,i_3))} a^{\tau_1} a^{\tau_2} a^{\tau_3}.$$

More generally,

$$(T_n)_{i_1,\ldots,i_n} = v^{(n)}_{top(\tau(i_1,\ldots,i_n))} a^{\tau_1} \ldots a^{\tau_n}.$$

Notice that the new parametrization is monomial, so the Zariski closure of its image is a toric variety [Sullivant, 2018]. We can determine the associated parametrization matrix P. Its rows are indexed by the parameters and its columns are indexed by the cumulants of the model.

Example 2.6.1. Let G be the graph in Fig. 2.6.10.

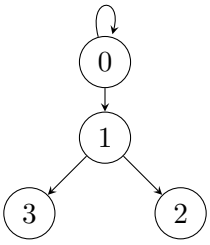


Figure 2.6.10: A polytree on four vertices with a single self-loop at the source.

For the third-order cumulant model, the parameters are $v_0^{(2)}, \ldots, v_3^{(2)}$, $v_0^{(3)}, \ldots, v_3^{(3)}$ and $a_{00}, \ldots, a_{31}$. The associated parametrization matrix P is given by

$$\begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 2 & 2 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 2 & 4 & 4 & 1 & 3 & 3 & 2 & 2 & 2 & 0 & 2 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 2 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 0 & 3 & 3 & 3 & 3 & 3 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 2 & 1 & 0 & 0 & 1 & 0 & 2 & 1 & 0 & 0 & 2 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 2 & 0 & 0 & 1 & 0 & 1 & 2 & 0 & 1 & 2 & 0
\end{pmatrix}.$$

The associated toric ideal lives in a polynomial ring with 30 variables s_{ij} and t_{ijk} for $i, j, k \in \{0, 1, 2, 3\}$, and has degree 1374. Its Gröbner basis contains 640 generators.

Even though the ideal itself is intractable, we can still recover certain information about the graph from the toric ideal.

Definition 2.6.2. *Define the* level *of a vertex v in a directed tree to be the length of the shortest path from the source node to v. For each $\ell \geq 0$, define $V_\ell(G)$ to be the set of all vertices in G at level ℓ.*

The following result shows how the levels of vertices can be recovered from the vanishing ideal.

Proposition 2.6.3. *Let $G = (V, E)$ be a directed tree with a single self-loop at the source, where $V = \{0\} \cup [p-1]$, and let $\mathcal{I}^{\leq n}(G)$ be the vanishing ideal of the nth order cumulant model associated to G for $n \geq 3$. Then,*

1. *For $i \in V$, $s_{ij}^3 t_{iii}^2 - s_{ii}^3 t_{iij} t_{ijj} \in \mathcal{I}^{\leq n}(G)$ for all $j \in V$ if and only if i is the source node.*

2. *Let 0 be the source node. Then, for $i, j \in V$, $s_{0i} t_{00j} - s_{0j} t_{00i} \in \mathcal{I}^{\leq n}(G)$ if and only if i and j are on the same level.*

3. *For $i, j \in V$ on different levels, $s_{ij}t_{00j} - s_{0j}t_{0ij} \in \mathcal{I}^{\leq n}(G)$ if and only if the level of j is greater than level of i (i.e., j is further from the source node than i).*

The proof can be found in Appendix 2.D.

Example 2.6.4. For the polytree in Fig. 2.6.11, the polynomial $s_{02}s_{34}t_{224} - s_{03}s_{22}t_{244}$ lies in the vanishing ideal. Note that this polynomial does not satisfy any of the conditions from Proposition 2.6.3. Here, vertex 2 is the top of the equitrek between vertices 3 and 4.

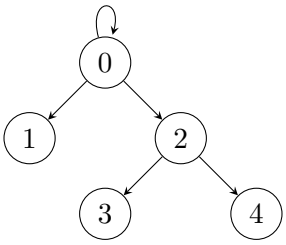


Figure 2.6.11: A polytree on five vertices with a single self-loop at the source.

The extra polynomial found in Example 2.6.4, corresponding to a trek which does not have a top node equal to the source, is an instance of a more general result, as described in Proposition 2.6.5 below.

Proposition 2.6.5. *Let $G = (V, E)$ be a directed tree with a single self-loop at the source, where $V = \{0\} \cup [p-1]$, and let $I^{\leq n}(G)$ be the vanishing ideal of the kth order cumulant model associated to G for $n \geq 3$. Let $i, j, l \in V$. Then,*

$$s_{0l}s_{ij}t_{llj} - s_{0i}s_{ll}t_{ljj} \in \mathcal{I}^{\leq n}(G)$$

if and only if l is the top of the shortest equitrek between i and j.

Consider two directed trees G and H with self-loops at their respective sources and the same number of nodes. It follows from Proposition 2.6.3 and Proposition 2.6.5 that if the levels of the graphs differ, or if there are two vertices for which the tops of the shortest equitreks are different, then the ideals are different. It turns out that these are also necessary conditions for the ideals to be different.

Theorem 2.6.6. *Two directed trees G and H on the same vertex set with self-loops at their respective sources define the same ideal if and only if*

(i) for all $\ell \geq 0$, the vertices at level ℓ in G are the same as the vertices at level ℓ in H, i.e. $V_\ell(G) = V_\ell(H)$, and

(ii) for each pair of vertices, the tops of the shortest equitreks between them are the same in G and H.

This result will follow from Lemma 2.6.7 and Lemma 2.6.8, both of which are proven in Appendix 2.D. The main idea is that if the graphs G and H satisfy the conditions of Theorem 2.6.6, then their corresponding parametrization matrices are row-equivalent, which implies that the vanishing ideal coincide.

Lemma 2.6.7. *Let $G = (V, E)$ be a directed tree with $V = \{0\} \cup [p-1]$, where 0 is the source with a self-loop. Let i and j be two vertices from the same level, each having at most one outgoing edge. Suppose there is no vertex i' such that the shortest equitrek between i and i' or between j and i' is shorter than the equitrek between i and j. Let H be the graph constructed by swapping i and j in G. Then, for any order n of cumulants, the vanishing ideals coincide: $\mathcal{I}^{\leq n}(H) = \mathcal{I}^{\leq n}(G)$.*

Lemma 2.6.8. *Let G and H be directed trees with self-loops at their respective sources, and with the same number of vertices p, labeled from 0 to $p-1$. Then H can be constructed from G using the swapping operations from Lemma 2.6.7 if and only if the following conditions hold:*

1. *For all $\ell \geq 0$, the vertices at level ℓ in G and H coincide, i.e., $V_\ell(G) = V_\ell(H)$.*
2. *For each pair of vertices i and j, the top vertices of the shortest equitreks are the same in G and in H. (Equivalently, the first p rows of the base trek parametrization matrix are the same.)*

Lemma 2.6.8 immediately implies Theorem 2.6.6 and constructively shows that any two graphs satisfying the conditions in the theorem have row-equivalent parametrization matrices.

2.6.2 Defining Polynomials via the Vanishing of Determinants

In this section, we consider the problem of finding elements of the vanishing ideal of the discrete Lyapunov model in cases where the defining graph is not a polytree with a single source. We do this by showing that certain graphical structures imply the vanishing of certain determinants. We further show that this is not a complete characterization and it remains a topic for future research.

We now show that certain matrices have to drop rank under certain constraints on the parents and ancestors of a set of vertices. The proofs are given in Appendix 2.D. To state these results, we define the set of parents of a subset $U = \{v_1, \ldots, v_k\} \subset V$ by

$$\mathrm{pa}(U) = \bigcup_{i=1}^{k} \mathrm{pa}(v_i).$$

Proposition 2.6.9. *Let $G = (V, E)$ be a graph on p vertices, and let $U = \{v_1, \ldots, v_k\}$ be a subset of V. Denote by S' the $(p-k) \times k$ submatrix of S, formed by columns $v_1, \ldots, v_k$ and not containing any diagonal entries. Then,*

$$rk(S') \leq |pa(U)|.$$

The same result holds for all slices of the tensor T. Here, for a fixed $i \in V$, a slice T_i is a $p \times p$ matrix T_{ijk} for $0 \leq j, k \leq p-1$.

Furthermore, the statement of Proposition 2.6.9 holds for the matrix formed by gluing S and all slices of T below each other, forming a $p \times (p + p^2)$ matrix.

Proposition 2.6.10. *Let $G = (V, E)$ be a graph on p vertices, and let $U = \{v_1, \dots, v_k\}$ be a subset of V. Form a $p(p+1) \times p$ matrix by stacking S, and T_i, $i \in \{0\} \cup [p-1]$. Denote by Q the submatrix formed by the columns $v_1, \dots, v_k$ and not containing any diagonal entries s_{ii} or t_{iii}. Then,*

$$rk(Q) \leq |pa(U)|.$$

Note that this is a generalization of the previous proposition since S' is a submatrix of Q. It follows that if we have a subset of vertices whose total number of parents is small, we can obtain polynomials in the vanishing ideal of the model by taking determinants of submatrices of the matrix Q.

(a) Vertices 1 and 2 have one parent.

(b) Vertices 1 and 2 have one grandparent.

Figure 2.6.12: Graphs with vanishing determinants.

Example 2.6.11. Let G be the graph in Fig. 2.6.12a, and let $U = \{1, 2\}$. Note that $|pa(U)| = 1$. Consider the matrix Q as defined in Proposition 2.6.10:

$$Q^t = \begin{pmatrix} s_{01} & t_{001} & t_{011} & t_{012} & t_{011} & t_{112} & t_{012} & t_{112} \\ s_{02} & t_{002} & t_{012} & t_{022} & t_{012} & t_{122} & t_{022} & t_{012} \end{pmatrix}.$$

Observe that some columns of the matrix above repeat due to the symmetries of the cumulant tensor. By the proposition, all 2×2 minors of Q must vanish. However, these polynomials alone do not generate the whole ideal $\mathcal{I}_G^{\leq 3}$.

Proposition 2.6.12. *Let $G = (V, E)$ be a graph on p vertices, and let $U = \{v_1, \dots, v_k\}$ be a subset of V. Let $an_2(U)$ be the set of all grandparents of U, and $sib(U)$ to be the set of siblings:*

$$an_2(U) = \bigcup_{w \in pa(U)} pa(w), \quad sib(U) = \{w \in V : pa(w) \in pa(U)\}.$$

Denote by S' the $(p - |sib(U)|) \times k$ submatrix of S, formed by taking columns $v_1, \dots, v_k$ and removing all rows containing entries s_{ij} with $i, j \in sib(U)$. Then,

$$rk(S') \leq |an_2(U)|.$$

The same statement holds for all slices of the tensor T as well. Moreover, the statement holds for a matrix, formed by gluing S and all slices of T below each other (to form a $p \times (p + p^2)$ matrix).

Example 2.6.13. Let G be the graph in Fig. 2.6.12b, and let $U = \{1, 2\}$. Here, $|an_2(U)| = 1$, and $|sib(U)| = 2$. For the slice T_3, consider the submatrix T_3' introduced in Proposition 2.6.12:

$$T_3' = \begin{pmatrix} t_{013} & t_{023} \\ t_{113} & t_{123} \\ t_{123} & t_{223} \\ t_{133} & t_{233} \end{pmatrix}.$$

In this case, the matrix T_3' has the same number of rows as T_3, as none of them need to be deleted. By Proposition 2.6.12, all 2×2 minors of T_3' vanish.

Finally, we show that the problem of computing the vanishing ideal of the model can be simplified by removing entries of the cumulants indexed only by a sink. Here we call a vertex v a *sink* if it has no outgoing edges directed to other vertices.

Proposition 2.6.14. *Let $G = (V, E)$ be a directed graph, and consider the third-order vanishing ideal $\mathcal{I}_{\widehat{G}}^{\leq 3} \subseteq \mathbb{R}[s_{ij}, t_{ijk} \mid i, j, k \in V]$ of the corresponding discrete Lyapunov model. If v is a sink in G, then there exists a generating set of $\mathcal{I}_{\widehat{G}}^{\leq 3}$ such that the variables s_{vv} and t_{vvv} do not appear in any generator.*

Proof. Suppose that $f = \sum_{t=0}^{k} s_{vv}^t f_t$ belongs to a generating set, and assume that no terms of f_t can be divided by s_{vv}. Using the trek rule, f can be written as an element of $\mathbb{R}(a_{ij}, \omega_i^{(2)}, \omega_i^{(3)},\ i, j \in V)$. Now assign arbitrary values to a_{ij} and $\omega_i^{(3)}$ for all $i, j \in V$ and to $\omega_i^{(2)}$ for all $i \in V \setminus \{v\}$ from the parameter space Θ. Denote the resulting polynomial as $g \in \mathbb{R}[\omega_v^{(2)}]$. Note that g has degree k. Since $f \in \mathcal{I}_{\widehat{G}}^{\leq 3}$, it must vanish for all assignments of the variables $a_{ij}, \omega_i^{(2)}, \omega_i^{(3)}$. Thus, $g(\omega_v^{(2)}) = 0$ for all $\omega_v^{(2)} \in \mathbb{R}$. It follows that $g = 0$, and hence all its coefficients are 0. The coefficients are evaluations of f_t or $\frac{f_t}{(1-a_{vv})^t}$, depending on whether there is a self-loop at v. Thus, $f_t \in \mathcal{I}_{\widehat{G}}^{\leq 3}$, and we can add them to the generating set instead of f. We can perform a similar process to get generators not involving t_{vvv}. □

Corollary 2.6.15. *The elimination of s_{vv} and t_{vvv} does not change the vanishing ideal corresponding to G in case v is a sink of G.*

2.6.3 Defining Polynomials via Birational Implicitization

In Boege and Solus [2024], the authors develop a strategy for efficiently computing the polynomials and semialgebraic constraints that define a statistical model under strict assumptions on its parametrization. In particular, their approach requires that the parametrization be *ambirational*.

Definition 2.6.16. *Let $\alpha : \mathbb{R}^m \to \mathbb{R}^n$ be a rational function, $\Theta \subset \mathbb{R}^m$ a parameter space, and $\mathcal{M} = \alpha(\Theta)$ a parametric statistical model. If α admits a rational inverse $\beta : \mathbb{R}^n \dashrightarrow \mathbb{R}^m$, then $\mathcal{M}$ is* ambirational.

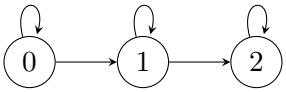


Figure 2.6.13: A directed path with all self-loops

In particular, this means that Θ and $\mathcal{M}$ are birationally equivalent, and the rational functions realizing this equivalence are also isomorphisms of their respective ambient spaces. In this setting, one can apply [Boege and Solus, 2024, Theorem 3.10] to use the polynomial constraints on the parameter space Θ to derive polynomial constraints on $\mathcal{M}$. We have shown in Theorem 2.4.3 that when G is a DAG with all self-loops, the map from the parameter space consisting of the matrix A and cumulants K_2, K_3 and K_4 to the space of 2nd, 3rd and 4th order cumulants admits a rational inverse. We have not shown that this map is ambirational; indeed, Theorem 2.4.3 shows that for a DAG G with all the self loops, for generic cumulants in $\mathcal{M}_G^{\leq 4}$ the parameters A and $\Omega^{(m)}$ can be found as a rational function of the cumulants. This does not rule out that possibility that the map from parameter space to the cumulants is somewhere many-to-one; in general, these models are generically identifiable but not globally identifiable. Nevertheless, we can attempt to apply the methods of Boege and Solus [2024] to compute candidates for defining polynomials of the model without the guarantee that the resulting polynomials will actually belong to the vanishing ideal.

Consider the directed path G with all self-loops pictured in Fig. 2.6.13. Given generic second-, third- and fourth-order cumulants in $\mathcal{M}_G^{\leq 4}$, Theorem 2.4.3 allows us to solve for the adjacency matrix A as a rational function of these cumulants. This also gives an expression for a_{20} in terms of the cumulants; but since A is a weighted adjacency matrix of G, the $(2,0)$ entry of A should be zero. If the map which sends $\mathcal{M}_G^{\leq 4}$ to the matrix A were ambirational, then by [Boege and Solus, 2024, Theorem 3.10], the numerator of a_{20} as a function of the second-, third- and fourth-order cumulants belongs to the vanishing ideal $\mathcal{I}^{\leq 4}(G)$. Using Macaulay2, we compute that this numerator has two irreducible factors, $a_{20} = f_1 f_2$. If we assume that $a_{20} \in \mathcal{I}^{\leq 4}(G)$, then since this ideal is prime, one of f_1 or f_2 must belong to $\mathcal{I}^{\leq 4}(G)$. By generating random cumulants in $\mathcal{M}_G^{\leq 4}$ and substituting them into these polynomials, we find that the one that vanishes on these random cumulants is the following polynomial, f_1:

$$
\begin{aligned}
&s_{00}s_{01}^3s_{02}s_{11}t_{001}^4t_{002}r_{0000}^3\\
&+s_{00}s_{01}^3s_{02}s_{12}t_{000}t_{001}^4r_{0000}^2r_{0001}\\
&-s_{01}^4s_{02}s_{11}t_{000}^2t_{001}^2t_{002}r_{0000}^2r_{0001}\\
&-2s_{00}s_{01}^3s_{02}s_{11}t_{000}t_{001}^3t_{002}r_{0000}^2r_{0001}\\
&+s_{00}^2s_{01}^3s_{12}t_{001}^4t_{002}r_{0000}^2r_{0001}\\
&+s_{00}s_{01}^3s_{02}s_{12}t_{000}^2t_{001}^3r_{0000}r_{0001}^2\\
&+s_{00}^3s_{01}s_{02}s_{12}t_{001}^5r_{0000}r_{0001}^2\\
&-2s_{01}^5s_{02}t_{000}^2t_{001}^2t_{002}r_{0000}r_{0001}^2\\
&+2s_{00}s_{01}^4s_{02}t_{000}t_{001}^3t_{002}r_{0000}r_{0001}^2\\
&-s_{01}^4s_{02}^2t_{000}^3t_{001}^2r_{0001}^3\\
&-4s_{00}^2s_{01}^2s_{02}s_{12}t_{000}^2t_{001}^3r_{0001}^3\\
&-s_{01}^4s_{02}s_{11}t_{000}^4t_{002}r_{0001}^3\\
&-2s_{00}s_{01}^4s_{12}t_{000}^3t_{001}t_{002}r_{0001}^3\\
&-3s_{00}^3s_{01}^2s_{12}t_{000}t_{001}^3t_{002}r_{0001}^3\\
&-s_{00}s_{01}^4s_{12}t_{001}^4t_{002}r_{0000}^3\\
&+s_{00}s_{01}^3s_{02}^2t_{001}^5r_{0000}^2r_{0001}\\
&+s_{01}^5s_{12}t_{000}^2t_{001}^2t_{002}r_{0000}^2r_{0001}\\
&+s_{00}s_{01}^4s_{12}t_{000}t_{001}^3t_{002}r_{0000}^2r_{0001}\\
&-s_{01}^4s_{02}s_{12}t_{000}^3t_{001}^2r_{0000}r_{0001}^2\\
&-2\,s_{00}s_{01}^3s_{02}^2t_{000}t_{001}^4r_{0000}r_{0001}^2\\
&+2s_{01}^4s_{02}s_{11}t_{000}^3t_{001}t_{002}r_{0000}r_{0001}^2\\
&+s_{00}s_{01}^3s_{02}s_{11}t_{000}^2t_{001}^2t_{002}r_{0000}r_{0001}^2\\
&+s_{00}^2s_{01}^3s_{12}t_{000}t_{001}^3t_{002}r_{0000}r_{0001}^2\\
&+2s_{00}s_{01}^3s_{02}s_{12}t_{000}^2t_{001}^2r_{0001}^3\\
&+3s_{00}^3s_{01}s_{02}s_{12}t_{000}t_{001}^4r_{0001}^3\\
&+s_{01}^5s_{12}t_{000}^4t_{002}r_{0001}^3\\
&-s_{00}s_{01}^4s_{02}t_{000}^2t_{001}^2t_{002}r_{0001}^3\\
&+s_{00}^4s_{01}s_{12}t_{001}^4t_{002}r_{0001}^3\\
&-s_{01}^4s_{02}^2t_{000}t_{001}^4r_{0000}^2r_{0001}\\
&-s_{00}^2s_{01}^2s_{02}s_{12}t_{001}^5r_{0000}^2r_{0001}\\
&+s_{01}^5s_{02}t_{000}t_{001}^3t_{002}r_{0000}^2r_{0001}\\
&-s_{00}s_{01}^4s_{02}t_{001}^4t_{002}r_{0000}^2r_{0001}\\
&+2s_{01}^4s_{02}^2t_{000}^2t_{001}^3r_{0000}r_{0001}^2\\
&-s_{00}^2s_{01}^2s_{02}s_{12}t_{000}t_{001}^4r_{0000}r_{0001}^2\\
&-s_{01}^5s_{12}t_{000}^3t_{001}t_{002}r_{0000}r_{0001}^2\\
&-s_{00}s_{01}^4s_{12}t_{000}^2t_{001}^2t_{002}r_{0000}r_{0001}^2\\
&-s_{00}^3s_{01}^2s_{12}t_{001}^4t_{002}r_{0000}r_{0001}^2\\
&+s_{00}s_{01}^3s_{02}^2t_{000}^2t_{001}^3r_{0001}^3\\
&-s_{00}^4s_{02}s_{12}t_{001}^5r_{0001}^3\\
&+s_{01}^5s_{02}t_{000}^3t_{001}t_{002}r_{0001}^3\\
&+4s_{00}^2s_{01}^3s_{12}t_{000}^2t_{001}^2t_{002}r_{0001}^3
\end{aligned}
$$

The fact that f_1 vanishes on cumulants in the model generated from random values of the entries of A is strong evidence that f_1 belongs to the vanishing ideal of the model. Indeed, if f_1 were not in $\mathcal{I}^{\leq 4}(G)$, then f_1 would be nonzero on almost all cumulants in $\mathcal{M}_G^{\leq 4}$. We leave both a statistical interpretation of this polynomial and a proof that it belongs to $\mathcal{I}^{\leq 4}(G)$ as directions for future research. This computation also leads us to the question: under what circumstances can the requirement of ambirationality in Boege and Solus [2024] be relaxed?

2.7 Discussion

In this paper, we presented a first study of discrete Lyapunov models with non-Gaussian errors. The non-Gaussianity gives rise to non-Gaussian equilibrium distributions, allowing us to consider not only its covariance matrix but also its higher-order cumulants. We showed that the entries of the cumulants of the equilibrium distribution can be expressed combinatorially via *equitreks* in the graph.

This combinatorial interpretation allowed us to derive several parameter identifiability results. In particular, we provided generic identifiability from second-, third-, and fourth-order cumulants results for DAGs with self-loops at each node (Section 2.4). Furthermore, we showed local identifiability for all directed graphs containing a self-loop at each node and no isolated noted (Section 2.5). Finally, we described some of the equations characterizing the implicit description of the model in Section 2.6.

This work is only the first study of such models, and numerous questions remain open. Most notably, these models can be used in causal discovery, and, therefore, one of the main problems is to design algorithms for learning the graph from samples of the equilibrium distribution. One of the ways this could be approached is by discovering more equations that vanish on the model for each graph, and then testing such equations on the sample cumulants obtained from data. We here have only discovered some of the

equations that vanish on the model for some special graphs, and it is still an open problem to characterize the full vanishing ideal of the model.

Funding. This project originated at the Workshop for Women in Algebraic Statistics, held at St John's College, Oxford from July 8 to July 18, 2024. The workshop was supported by St John's College, Oxford, the L'Oreal-UNESCO For Women in Science UK and Ireland Rising Talent Award in Mathematics and Computer Science (awarded to Jane Coons), the Heilbronn Institute for Mathematical Research, and the UKRI/EPSRC Additional Funding Programme for the Mathematical Sciences.

Elina Robeva was also supported by a Canada CIFAR AI Chair. Cecilie Olesen Recke was also supported by Novo Nordisk Foundation Grant NNF20OC0062897. Sarah Lumpp was also supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 883818) as well as the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Research, Technology and Space.

Appendix for 'Identifiability in Graphical Discrete Lyapunov Models'

2.A. Proofs for Section 2.3

2.A.1 Proof of Proposition 2.3.2

Proof. Spelling out the l-th mode products from Proposition 2.2.1, we obtain

$$\begin{aligned}(T_n)_{i_1\dots i_n} &= \sum_{l=0}^{\infty} \sum_{1\le r_1,\dots,r_n\le p} (\Omega^{(n)})_{r_1,\dots,r_n} (A^l)_{i_1 r_1}\cdots(A^l)_{i_n r_n} \\ &= \sum_{l=0}^{\infty}\sum_{r=1}^{p} w_r^{(n)} (A^l)_{i_1 r}\cdots(A^l)_{i_n r} = \sum_{r=1}^{p} w_r^{(n)} \sum_{l=0}^{\infty} (A^l)_{i_1 r}\cdots(A^l)_{i_n r},\end{aligned}$$

where in the second equality we used that $\Omega^{(n)}$ is diagonal. Further note that

$$(A^l)_{ij} = \sum_{P\in P_l(j,i)} a^P,$$

where $P_l(j,i)$ denotes the set of all paths from j to i of length l. Substituting this representation yields

$$\begin{aligned}(T_n)_{i_1\dots i_n} &= \sum_{r=1}^{p} w_r^{(n)} \sum_{l=0}^{\infty} \left(\sum_{P_1\in P_l(r,i_1)} a_1^P\right)\cdots\left(\sum_{P_n\in P_l(r,i_n)} a_n^P\right) \\ &= \sum_{\tau=(\tau_1,\dots,\tau_n)\in\mathcal{T}(i_1,\dots,i_n)} w_{top(\tau)}^{(n)} a^{\tau_1}\dots a^{\tau_n}.\end{aligned}$$

□

2.A.2 Proof of Proposition 2.3.7

The proof of Proposition 2.3.7 is based on induction by employing the recursive formula for the entries of S. Induction then basically means adding an edge on one or both sides of the considered base trek. Thus, the main challenge is to correctly account for the added possibilities of adding self-loops along the the trek, which is expressed in the following recursive properties of the rational coefficients $C(x,y;t)$ and the polynomials $p_{x,y}(t)$.

Lemma 2.A.1. *Let $0\le x\le y$. The polynomial p and the rational function C satisfy the following recursive relations:*

(i) $p_{x+1,y+1}(t) = t^2 p_{x,y+1}(t) + (1+(t^2-1)\delta_{xy})p_{x+1,y}(t) + (1-t^2)p_{x,y}(t)$;

(ii) $C(x+1,y+1;t) = \frac{1}{1-t^2}\left(t(C(x,y+1;t)+C(x+1,y;t)) + C(x,y;t)\right)$.

Proof. (i) We prove this using a similar strategy as in the proof of the Chu-Vandermonde identity by looking at coefficients in the expansion of the binomial formula.

First observe that $p_{x,y}(t)$ is equal to the coefficient of s^x in the polynomial $(1+s)^y(1+st^2)^x$ by writing out the appropriate binomial expansion. Thus, if we can prove the stronger claim of a corresponding recursive equality for these polynomials, it follows in particular that the coefficients of s^x are equal. This strategy works only when $x < y$, so the proof is now split into the two cases $x < y$ and $x = y$.

Case 1: For $x < y$, the recursive formula simplifies to

$$p_{x+1,y+1}(t) = t^2 p_{x,y+1}(t) + p_{x+1,y}(t) + (1-t^2)p_{x,y}(t).$$

By the explained correspondence, $p_{x,y+1}(t)$ is the coefficient of s^x in $(1+s)^{y+1}(1+st^2)^x$, $p_{x+1,y}(t)$ is the coefficient of s^{x+1} in $(1+s)^y(1+st^2)^{x+1}$, $p_{x,y}(t)$ is the coefficient of s^x in $(1+s)^y(1+st^2)^x$, and $p_{x+1,y+1}(t)$ is the coefficient of s^{x+1} in $(1+s)^{y+1}(1+st^2)^{x+1}$. Therefore, it suffices to prove the polynomial identity

$$(1+s)^{y+1}(1+st^2)^{x+1} = st^2(1+s)^{y+1}(1+st^2)^x + (1+s)^y(1+st^2)^{x+1} + s(1-t^2)(1+s)^y(1+st^2)^x.$$

Rewriting the right-hand side yields

$$\begin{aligned}
&st^2(1+s)^{y+1}(1+st^2)^x + (1+s)^y(1+st^2)^{x+1} + s(1-t^2)(1+s)^y(1+st^2)^x \\
&= (1+s)^y(1+st^2)^x(st^2(1+s) + (1+st^2) + s(1-t^2)) \\
&= (1+s)^y(1+st^2)^x(1+s+st^2+s^2t^2) = (1+s)^{y+1}(1+st^2)^{x+1},
\end{aligned}$$

which proves the required polynomial equality, concluding this case.

Case 2: For $x = y$, we write out the definition of p directly to prove the result. Since $p_{x,y}$ is symmetric in x and y, the recursive formula to prove becomes

$$p_{x+1,x+1}(t) = 2t^2 p_{x,x+1}(t) + (1-t^2)p_{x,x}(t).$$

Using the definition to write out the right-hand side yields

$$\begin{aligned}
&2t^2 p_{x,x+1}(t) + (1-t^2)p_{x,x}(t) \\
&= \sum_{l=0}^{x} t^{2(l+1)} \left(2\binom{x+1}{x-l}\binom{x}{l} - \binom{x}{l}^2 \right) + \sum_{l=0}^{x} t^{2l}\binom{x}{l}^2 \\
&= \sum_{k=1}^{x+1} t^{2k} \left(2\binom{x+1}{x-(k-1)}\binom{x}{k-1} - \binom{x}{k-1}^2 \right) + \sum_{l=0}^{x} t^{2l}\binom{x}{l}^2 \\
&= \sum_{k=1}^{x} t^{2k} \left(2\binom{x+1}{x-(k-1)}\binom{x}{k-1} - \binom{x}{k-1}^2 + \binom{x}{k}^2 \right) + t^{2(x+1)} + 1,
\end{aligned}$$

by shifting the index of the first sum. Note that the last two terms outside of the sum correspond to the extreme terms in $p_{x+1,x+1}(t)$. In order to prove the recursive identity,

the only thing left to compute is

$$\begin{aligned}
&2\binom{x+1}{k}\binom{x}{k-1} - \binom{x}{k-1}^2 + \binom{x}{k}^2 \\
&= \binom{x+1}{k}^2 \frac{2k}{x+1} - \left[\binom{x+1}{k}\frac{k}{x+1}\right]^2 + \left[\binom{x+1}{k}\frac{x+1-k}{x+1}\right]^2 \\
&= \binom{x+1}{k}^2 \left[\frac{2k(x+1)}{(x+1)^2} - \frac{k^2}{(x+1)^2} + \frac{(x+1-k)^2}{(x+1)^2}\right] \\
&= \binom{x+1}{k}^2
\end{aligned}$$

to obtain the formula for $p_{x+1,x+1}(t)$.

(*ii*) Writing out the definitions on the right-hand side, we obtain

$$\begin{aligned}
&\frac{1}{1-t^2}\left(t(C(x,y+1;t) + C(x+1,y;t)) + C(x,y;t)\right) \\
&= t^{|x-(y+1)|+1}\frac{p_{x,y+1}(t)}{(1-t^2)^{x+y+3}} + t^{|(x+1)-y|+1}\frac{p_{x+1,y}(t)}{(1-t^2)^{x+y+3}} + t^{|x-y|}\frac{(1-t^2)p_{x,y}(t)}{(1-t^2)^{x+y+3}} \\
&= \frac{t^{|(x+1)-(y+1)|}}{(1-t^2)^{(x+1)+(y+1)+1}}\left(t^2 p_{x,y+1}(t) + (1+(t^2-1)\delta_{xy})p_{x+1,y}(t) + (1-t^2)p_{x,y}(t)\right) \\
&= t^{|(x+1)-(y+1)|}\frac{p_{x+1,y+1}(t)}{(1-t^2)^{(x+1)+(y+1)+1}} = C(x+1,y+1;t).
\end{aligned}$$

In the second equality, we use that $x \leq y$, so we have

$$x < y+1, \text{ so } |x-(y+1)|+1 = y+1-x+1 = |x-y|+2.$$

If we consider $x+1$, the power of t depends on whether $x = y$ or not. In the case $x < y$, we have

$$x+1 \leq y, \text{ so } |(x+1)-y|+1 = y-(x+1)+1 = |x-y|.$$

In the case $x = y$, however, we have,

$$x+1 > y, \text{ so } |(x+1)-y|+1 = x-y+2 = |x-y|+2,$$

inducing the factor $(1+(t^2-1)\delta_{xy})$ to account for both cases in the equation. The result follows from the recursive property of the polynomial $p_{x,y}(t)$ shown in (*i*). □

Proof of Proposition 2.3.7. Writing out the recursive formula for s_{ij}, we obtain

$$s_{ij} = \sum_{\substack{k\in\mathrm{pa}(i),\\ l\in\mathrm{pa}(j)}} a_{ik}a_{jl}s_{kl} + \delta_{ij}\omega_i^{(2)} = t^2 s_{ij} + \sum_{\substack{k\in\mathrm{pa}(i),\\ l\in\mathrm{pa}(j),\\ (k,l)\neq(i,j)}} a_{ik}a_{jl}s_{kl} + \delta_{ij}\omega_i^{(2)}.$$

Solving for s_{ij} yields

$$\begin{aligned} s_{ij} &= \sum_{\substack{k\in\mathrm{pa}(i),\\ l\in\mathrm{pa}(j),\\ (k,l)\neq(i,j)}} \frac{1}{1-t^2} a_{ik}a_{jl}s_{kl} + \delta_{ij}\frac{1}{1-t^2}\omega_i^{(2)} \\ &= \sum_{l\in\mathrm{pa}(j)\setminus\{j\}} \frac{1}{1-t^2} t a_{jl}s_{il} + \sum_{k\in\mathrm{pa}(i)\setminus\{i\}} \frac{1}{1-t^2} t a_{ik}s_{kj} \\ &\quad + \sum_{\substack{k\in\mathrm{pa}(i)\setminus\{i\},\\ l\in\mathrm{pa}(j)\setminus\{j\}}} \frac{1}{1-t^2} a_{ik}a_{jl}s_{kl} + \delta_{ij}\frac{1}{1-t^2}\omega_i^{(2)}. \end{aligned}$$

This formula motivates an inductive proof similar to the proof of [Boege et al., 2025, Prop. 4.3]. We assume that (2.3.8) holds for s_{il}, s_{kj}, and s_{kl} as the induction hypothesis. Fix a topological ordering of the nodes and the induced lexicographic ordering of the entries s_{ij} of S, where $s_{kl} < s_{ij}$ if and only if $l < j$, or $l = j$ and $k < i$.

The smallest S-entries are given by s_{ii}, where i is a source node of G. In this case, we have

$$s_{ii} = \frac{1}{1-t^2}\omega_i^{(2)} = C(0,0;t)\omega_i^{(2)},$$

which corresponds to the sum over the empty trek in (2.3.8). Note that if i and j are two distinct source nodes, we have $s_{ij} = 0$ corresponding to the claim in (2.3.8) as well.

Now consider a non-empty base trek (i.e., with no cycles and self-loops) between two nodes i and l, denoted by $\tilde{\tau} \in \mathcal{T}^*(i,l)$. Let k be the unique parent of i on this trek. Then the trek $\tilde{\tau}$ can be decomposed uniquely into the remaining trek τ between k and l and the edge $k \to i$. Hence, we have a bijection between $\mathcal{T}^*(i,l)$ and $\bigcup_{k\in\mathrm{pa}(i)\setminus\{i\}} \mathcal{T}^*(k,l)$. Moreover, $d(\tilde{\tau}_i) = d(\tau_k) + 1$ and $a^{\tilde{\tau}_i} = a^{\tau_k}\cdot a_{ik}$. A similar decomposition can be applied to the path to the second leaf of the treak or to both at the same time. Applying the

induction hypothesis together with these observations and Lemma 2.A.1, we obtain

$$\begin{aligned}
s_{ij} &= \sum_{l\in \mathrm{pa}(j)\setminus\{j\}} \sum_{\tilde\tau=(\tilde\tau_i,\tilde\tau_l)\in\mathcal{T}^*(i,l)} \frac{1}{1-t^2} tC(d(\tilde\tau_i),d(\tilde\tau_l);t) a^{\tilde\tau_i} a^{\tilde\tau_l} a_{jl} w^{(2)}_{\mathrm{top}(\tilde\tau)} \\
&\quad + \sum_{k\in \mathrm{pa}(i)\setminus\{i\}} \sum_{\tilde\tau=(\tilde\tau_k,\tilde\tau_j)\in\mathcal{T}^*(k,j)} \frac{1}{1-t^2} tC(d(\tilde\tau_k),d(\tilde\tau_j);t) a^{\tilde\tau_k} a_{ik} a^{\tilde\tau_j} w^{(2)}_{\mathrm{top}(\tilde\tau)} \\
&\quad + \sum_{\substack{k\in \mathrm{pa}(i)\setminus\{i\},\\ l\in \mathrm{pa}(j)\setminus\{j\}}} \sum_{\tau=(\tau_k,\tau_l)\in\mathcal{T}^*(k,l)} \frac{1}{1-t^2} C(d(\tau_k),d(\tau_l);t) a^{\tau_k} a_{ik} a^{\tau_l} a_{jl} w^{(2)}_{\mathrm{top}(\tau)} + \delta_{ij}\frac{1}{1-t^2}\omega_i^{(2)} \\
&= \sum_{\substack{k\in \mathrm{pa}(i)\setminus\{i\},\\ l\in \mathrm{pa}(j)\setminus\{j\}}} \sum_{\tau=(\tau_k,\tau_l)\in\mathcal{T}^*(k,l)} \frac{1}{1-t^2} [t(C(d(\tau_k)+1,d(\tau_l);t) + C(d(\tau_k),d(\tau_l)+1;t)) \\
&\quad + C(d(\tau_k),d(\tau_l);t)] \cdot a^{\tau_k} a_{ik} a^{\tau_l} a_{jl} w^{(2)}_{\mathrm{top}(\tau)} + \delta_{ij}\frac{1}{1-t^2}\omega_i^{(2)} \\
&= \sum_{\substack{k\in \mathrm{pa}(i)\setminus\{i\},\\ l\in \mathrm{pa}(j)\setminus\{j\}}} \sum_{\tau=(\tau_k,\tau_l)\in\mathcal{T}^*(k,l)} C(d(\tau_k)+1,d(\tau_l)+1;t) a^{\tau_k} a_{ik} a^{\tau_l} a_{jl} w^{(2)}_{\mathrm{top}(\tau)} + \delta_{ij}\frac{1}{1-t^2}\omega_i^{(2)} \\
&= \sum_{\tau=(\tau_i,\tau_j)\in\mathcal{T}^*(i,j)} C(d(\tau_i),d(\tau_j);t) a^{\tau_i} a^{\tau_j} w^{(2)}_{\mathrm{top}(\tau)}.
\end{aligned}$$

In the case $i = j$, the last term $\frac{1}{1-t^2}\omega_i^{(2)}$ accounts for the empty trek that does not involve true parents of i. □

Remark 2.A.2. The proof of the restricted trek rule for third-order cumulant entries is analogous to the argument above. However, it requires substantially more bookkeeping of indices during the induction, as well as proving a more involved recursion for the appearing polynomials $p_{x,y,z}(t)$.

2.A.3 Restricted trek rule for higher-order cumulants

The general form of an entry of the nth order cumulant in the discrete Lyapunov model of a DAG is given as

$$(T_n)_{i_1,\dots,i_n} = \sum_{\tau=(\tau_1,\dots,\tau_n)\in\mathcal{T}^*(i_1,\dots,i_n)} C(d(\tau_1),\dots,d(\tau_n);t) a^{\tau_1}\cdots a^{\tau_n} w^{(n)}_{top(\tau)},$$

where C is the rational function in the edge lengths of all n components of the considered base trek as well as the self-loop parameter t given by

$$C(x_1,\dots,x_n;t) = t^{n\cdot\max(x_1,\dots,x_n)-\sum_{l=1}^n x_l} \frac{p_{x_1,\dots,x_n}(t)}{(1-t^n)^{\sum_{l=1}^n x_l+1}}.$$

For instance, in the case $n = 3$, the rational function is given as

$$C(x,y,z;t) = t^{3\cdot\max(x,y,z)-(x+y+z)} \frac{p_{x,y,z}(t)}{(1-t^3)^{x+y+z+1}}.$$

While it is clear how to compute the powers of t and $(1-t^3)$, the polynomial $p_{x,y,z}(t)$ is much less tractable.

In the case $n = 2$, the polynomial $p_{x,y}(t)$ seems simple, however it contains a more intricate structure that allows us to extend the definition to higher n. First, observe that in the case $x = y$, i.e., a base trek with equal length paths, the formula collapses to

$$p_{x,x}(t) = \sum_{l=0}^{x} t^{2l} \binom{x}{l}^2. \tag{2.A.1}$$

The coefficients appearing in $p_{x,x}(t)$ are the squared entries of the x-th row of Pascal's triangle. The resulting integer sequence is known as OEIS A008459. Combinatorially, this sequence admits several interpretations; for example, it counts the lattice paths from $(0,0)$ to (x,x) with steps $(1,0)$ and $(0,1)$, having l right turns.

In the case $n = 3$, the coefficients of the polynomial $p_{x,x,x}(t)$ are given by a known integer sequence as well, namely OEIS A181544. This sequence can be defined via the row generating function of row x:

$$B(x) = \left(\sum_{i \geq 0} \binom{x+i}{i}^3 s^i \right) (1-s)^{3x+1} = \sum_{l \geq 0} \left(\sum_{k=0}^{l} (-1)^k \binom{3x+1}{k} \binom{x+l-k}{l-k}^3 \right) s^l.$$

This representation allows us to infer the closed formula for each coefficient in that row, yielding

$$p_{x,x,x}(t) = \sum_{l=0}^{2x} t^{3(2x-l)} \sum_{k=0}^{l} (-1)^{l-k} \binom{3x+1}{l-k} \binom{x+k}{k}^3.$$

Interestingly, a similar formula can be derived from the row generating function for $n = 2$, namely

$$p_{x,x}(t) = \sum_{l=0}^{x} t^{2(x-l)} \sum_{k=0}^{l} (-1)^{l-k} \binom{2x+1}{l-k} \binom{x+k}{k}^2.$$

It can be shown that the coefficients of this polynomial coincide with the squared binomial coefficients appearing in (2.A.1). Generalizing this to the product of two different binomial coefficients reveals the underlying combinatorial pattern of the coefficients in $p_{x,y}(t)$ as

$$\binom{x}{l}\binom{y}{l} = \sum_{k=0}^{l} (-1)^{l-k} \binom{x+y+1}{l-k} \binom{x+k}{k} \binom{y+k}{k}$$

holds (by rewriting the binomial coefficients in the sum as a part independent of i and a part dependent on i that can further be rewritten as an application of the hypergeometric function ${}_3F_2$ – which can be rewritten again in terms of a fraction of binomial coefficients). When generalizing this pattern to the case $n = 3$, we obtain

$$p_{x,y,z}(t) = \sum_{l=0}^{x+y+z-\max(x,y,z)} t^{3(x+y+z-\max(x,y,z)-l)} \sum_{k=0}^{l} (-1)^{l-k} \binom{x+y+z+1}{l-k} \binom{x+k}{k} \binom{y+k}{k} \binom{z+k}{k}.$$

Consequently, we conjecture that this formula generalizes to higher n as

$$p_{x_1,\dots,x_n}(t) = \sum_{l=0}^{\sum_{i=1}^n x_i - \max(x_1,\dots,x_n)} t^{n(\sum_{i=1}^n x_i - \max(x_1,\dots,x_n)-l)} \sum_{k=0}^{l} (-1)^{l-k} \binom{x_1+\cdots+x_1+1}{l-k} \prod_{i=1}^{n} \binom{x_i+k}{k}.$$

2.B. Proofs for Section 2.4

2.B.1 Proof of Theorem 2.4.3

Proof. Let the node set V of G be topologically ordered as $0 \leq \cdots \leq p-1$. The proof of generic identifiability proceeds by induction. At each step j, we show that we can generically identify all parameters a corresponding to edges going into node j, i.e., a_{ji} for $i \leq j$.

Since G is a DAG, the first node 0 in the topological order is a source. If we consider the subgraph of G consisting of 0 and its first child i according to the topological order, we are in the case of Example 2.4.1 and can generically identify a_{00}.

We now proceed with the induction step. Assume we have generically identified all edges going into the nodes up until $j-1$, i.e., all a_{kl} such that $k \leq j-1$ and $l \leq j-1$. Now consider node j. There are two cases, either it is or is not a source. If it is a source proceed similarly to the base case to identify a_{jj}.

Otherwise, j is a non-source node. In that case, we obtain the following equations by using the recursive formulas for S and T, when $i \neq j$. Without loss of generality, we assume that the first source in the topological order on G that is an ancestor of j is node 0. Then,

$$s_{ij} = \sum_{k \in \mathrm{pa}(i), l \in \mathrm{pa}(j)} a_{ik} a_{jl} s_{kl}, \qquad t_{00j} = \sum_{m \in \mathrm{pa}(j)} a_{00}^2 a_{jm} t_{00m}.$$

Let $\{i_1, \dots, i_d\}$ denote the parents of j not including j itself in topological order. We collect the equations for $s_{i_1 j}, \dots, s_{i_d j}$ together with the equation for t_{00j} as follows:

$$\begin{bmatrix} s_{i_1 j} \\ \vdots \\ s_{i_d j} \\ t_{00j} \end{bmatrix} = \begin{bmatrix} A_{\{i_1,\dots,i_d\}\times(\{0\}\cup[j-1])} & 0 \\ 0 & a_{00}^2 \end{bmatrix} \begin{bmatrix} S_{(\{0\}\cup[j-1])\times\{i_1,\dots,i_d,j\}} \\ t_{00i_1} \quad \cdots \quad t_{00i_d} \quad t_{00j} \end{bmatrix} \cdot \begin{bmatrix} a_{ji_1} \\ \vdots \\ a_{ji_d} \\ a_{jj} \end{bmatrix}$$

$$= \begin{bmatrix} A_{\{i_1,\dots,i_d\}\times(\{0\}\cup[j-1])} \cdot S_{(\{0\}\cup[j-1])\times\{i_1,\dots,i_d\}} & A_{\{i_1,\dots,i_d\}\times(\{0\}\cup[j-1])} S_{(\{0\}\cup[j-1])\times j} \\ a_{00}^2 t_{00i_1} \quad \cdots \quad a_{00}^2 t_{00i_d} & a_{00}^2 t_{00j} \end{bmatrix} \cdot \begin{bmatrix} a_{ji_1} \\ \vdots \\ a_{ji_d} \\ a_{jj} \end{bmatrix}.$$

We denote the matrix above by P^j with rows indexed by $\{i_1, \dots, i_d, 0\}$ and columns indexed by $\{i_1, \dots, i_d, j\}$ in the indicated order. Our goal is to show that P^j is generically invertible. This is equivalent to the upper left block $P^j_{\{i_1,\dots,i_d\}\times\{i_1,\dots,i_d\}}$ as well as its Schur complement in P^j being generically invertible.

To show this, we pick a lower triangular matrix A with non-zero diagonal entries such that there is only one path from 0 to j with non-zero edge weights. All other edge weights that are not self-loops are set to zero. Without loss of generality, assume that

this path to j passes through the first parent i_1. We denote the unique nodes on the path by $(0, k_1, \ldots, k_r, i_1, j)$. As a consequence, we have that $t_{00i_1} \neq 0$ and $t_{00j} \neq 0$, while $t_{00i_2}, \ldots, t_{00i_d} = 0$. Furthermore, we have in particular that $s_{0j}, s_{k_1 j}, \ldots, s_{k_r j}, s_{i_1 j} \neq 0$, while $s_{lj} = 0$ for all nodes l that are not on that path.

Computing the upper left block of P^j, we obtain

$$P^j_{\{i_1,\ldots,i_d\}\times\{i_1,\ldots,i_d\}} = \begin{bmatrix} \cdots & 0 & a_{i_1k_r} & a_{i_1i_1} & 0 & \cdots & \cdots & & \\ \cdots & 0 & 0 & 0 & a_{i_2i_2} & 0 & \cdots & & \\ \ddots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & & \\ \cdots & 0 & 0 & 0 & \cdots & 0 & a_{i_di_d} & 0 & \cdots \end{bmatrix} \begin{bmatrix} s_{0i_1} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ s_{k_{r-1}i_1} & \vdots & & & \vdots \\ s_{k_ri_1} & \vdots & & & \vdots \\ s_{i_1i_1} & 0 & \cdots & \cdots & 0 \\ 0 & s_{i_2i_2} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & s_{i_di_d} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} a_{i_1k_r}s_{k_ri_1} + a_{i_1i_1}s_{i_1i_1} & & & \\ & a_{i_2i_2}s_{i_2i_2} & & \\ & & \ddots & \\ & & & a_{i_di_d}s_{i_di_d} \end{bmatrix},$$

so it is generically invertible by taking the inverse of the diagonal elements. Note that we depicted the rows and columns indexed by k_r, $i_1, \ldots, i_d$ to be adjacent in the matrix, however they do not have to be consecutive nodes in the topological order. In that case, there would be additional zero rows and columns in the respective matrices.

Its Schur complement in P^j is of the form

$$P^j_{0j\cdot\{i_1,\ldots,i_d\}} = a^2_{00}t_{00j} - \begin{bmatrix} a^2_{00}t_{00i_1} & \ldots & a^2_{00}t_{00i_d} & a^2_{00}t_{00j} \end{bmatrix} \cdot \left(P^j_{\{i_1,\ldots,i_d\}\times\{i_1,\ldots,i_d\}}\right)^{-1} \cdot P^j_{\{i_1,\ldots,i_d\}\times j}.$$

We further have

$$P^j_{\{i_1,\ldots,i_d\}\times j} = A_{\{i_1,\ldots,i_d\}\times(\{0\}\cup[j-1])} \cdot S_{(\{0\}\cup[j-1])\times j}$$

$$= A_{\{i_1,\ldots,i_d\}\times(\{0\}\cup[j-1])} \cdot \begin{bmatrix} * \\ \vdots \\ * \\ s_{k_{r-1}j} \\ s_{k_rj} \\ s_{i_1j} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{i_1k_r}s_{k_rj} + a_{i_1i_1}s_{i_1j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Combining these expressions yields

$$P^j_{0j\cdot\{i_1,\ldots,i_d\}} = a^2_{00}t_{00j} - a^2_{00}t_{00i_1}\frac{a_{i_1k_r}s_{k_rj} + a_{i_1i_1}s_{i_1j}}{a_{i_1k_r}s_{k_ri_1} + a_{i_1i_1}s_{i_1i_1}}.$$

Thus, we now have to choose the values of the non-zero edge weights in A such that

$$t_{00j}\left(a_{i_1k_r}s_{k_ri_1} + a_{i_1i_1}s_{i_1i_1}\right) - t_{00i_1}\left(a_{i_1k_r}s_{k_rj} + a_{i_1i_1}s_{i_1j}\right) \neq 0. \tag{2.B.1}$$

Note that

$$t_{00j} = \frac{a_{00}^2 a_{ji_1}}{1 - a_{00}^2 a_{jj}} t_{00i_1},$$

so the expression above simplifies to

$$a_{00}^2 a_{i_1j}\left(a_{i_1k_r}s_{k_ri_1} + a_{i_1i_1}s_{i_1i_1}\right) - \left(1 - a_{00}^2 a_{jj}\right)\left(a_{i_1k_r}s_{k_rj} + a_{i_1i_1}s_{i_1j}\right) \neq 0.$$

Observe that not all terms in the above expression depend on $\omega_{i_1}^{(2)}$, as there are, for example, no treks between k_r and either i_1 or j starting at i_1. The only terms that depend on $\omega_{i_1}^{(2)}$ are $s_{i_1i_1}$ and s_{i_1j}, and we have

$$\begin{aligned}
s_{i_1i_1} &= \cdots + \omega_{i_1}^{(2)} \frac{1}{1 - a_{i_1i_1}^2}, \\
s_{i_1j} &= \frac{1}{1 - a_{i_1i_1}a_{jj}}\left(a_{i_1k_r}a_{ji_1}s_{k_ri_1} + a_{i_1k_r}a_{jj}s_{k_rj} + a_{i_1i_1}a_{ji_1}s_{i_1i_1}\right) \\
&= \cdots + \omega_{i_1}^{(2)} \frac{a_{i_1i_1}a_{ji_1}}{(1 - a_{i_1i_1}^2)(1 - a_{i_1i_1}a_{jj})}
\end{aligned}$$

due to the structure of the chosen A. Consequently, the coefficient of $\omega_{i_1}^{(2)}$ in the numerator of the Schur complement (2.B.1) is

$$\begin{aligned}
&a_{00}^2 a_{i_1j}a_{i_1i_1}\frac{1}{1 - a_{i_1i_1}^2} - (1 - a_{00}^2 a_{jj})a_{i_1i_1}\frac{a_{i_1i_1}a_{ji_1}}{(1 - a_{i_1i_1}^2)(1 - a_{i_1i_1}a_{jj})} \\
&= a_{i_1j}a_{i_1i_1}\frac{a_{00}^2(1 - a_{i_1i_1}a_{jj}) - (1 - a_{00}^2 a_{jj})a_{i_1i_1}}{(1 - a_{i_1i_1}^2)(1 - a_{i_1i_1}a_{jj})} \\
&= a_{i_1j}a_{i_1i_1}\frac{a_{00}^2 - a_{i_1i_1}}{(1 - a_{i_1i_1}^2)(1 - a_{i_1i_1}a_{jj})}.
\end{aligned}$$

Finally, if we set all non-zero entries of A to $\frac{1}{2}$, for instance, the coefficient evaluates to $-\frac{1}{9} \neq 0$. □

2.B.2 Proof of Theorem 2.4.7

Proof. Let the node set V of G be topologically ordered as $0 \leq \ldots \leq p-1$. The proof proceeds by induction.

The base case of the induction is to identify the self-loop of the first source. We can identify the self-loop of the source by Example 2.4.1 or Example 2.4.6, depending on whether the child has a self-loop.

We now proceed with the induction step. Consider node j. If it is a source node, repeat the base case. Otherwise, let $i_1, \ldots, i_d$ be the parents of j. There are two cases, depending on whether j has a self-loop or not.

Assume first that j does not have a self-loop. In this case, we need to identify d incoming edges at this step. For each parent i_k, let e_k denote one of it's ancestors that is a source. Notice that all $e_1, \ldots, e_d$ are distinct; otherwise, G would not be a polytree. Then, by collecting the recursive formulas for $s_{e_1 j}, \ldots s_{e_d j}$, we have

$$\begin{bmatrix} s_{e_1 j} \\ \vdots \\ s_{e_d j} \end{bmatrix} = \left[A_{\{e_1,\ldots,e_d\}\times(\{0\}\cup[j-1])}\right] \left[S_{(\{0\}\cup[j-1])\times\{i_1,\ldots,i_d\}}\right] \cdot \begin{bmatrix} a_{ji_1} \\ \vdots \\ a_{ji_d} \end{bmatrix}. \tag{2.B.2}$$

Notice that, since $e_1, \ldots, e_d$ are sources, they have no incoming edges except for the self-loops. Consequently, each row e_l has only one non-zero entry in column e_l. Therefore, the product of the first two matrices above can be written as the product of two square matrices:

$$\left[A_{\{e_1,\ldots,e_d\}\times(\{0\}\cup[j-1])}\right] \left[S_{(\{0\}\cup[j-1])\times\{i_1,\ldots,i_d\}}\right] = \left[A_{\{e_1,\ldots,e_d\}\times\{e_1,\ldots,e_d\}}\right] \left[S_{\{e_1,\ldots,e_d\}\times\{i_1,\ldots,i_d\}}\right].$$

The product is diagonal, since $s_{e_l i_k} = 0$ unless $l = k$. Hence, we can write

$$\begin{bmatrix} s_{e_1 j} \\ \vdots \\ s_{e_d j} \end{bmatrix} = \begin{bmatrix} a_{e_1 e_1} s_{e_1 i_1} & & & \\ & a_{e_2 e_2} s_{e_2 i_2} & & \\ & & \ddots & \\ & & & a_{e_d e_d} s_{e_d i_d} \end{bmatrix} \cdot \begin{bmatrix} a_{ji_1} \\ \vdots \\ a_{ji_d} \end{bmatrix}.$$

Now, let A satisfy $a_{kk} \neq 0$ for all k and $a_{ji_1} \neq 0$. Then, by the equitrek rule, we have $s_{e_k i_k} \neq 0$ for all $1 \leq k \leq d$, since there exists an equitrek between e_k and i_k for each k. For this choice of parameters, the matrix is invertible and therefore, the parameters are generically identifiable.

Now assume that j has a self-loop. Without loss of generality, assume $0 = e_1$. It follows that $t_{00 i_1} \neq 0$, while $t_{00 i_k} = 0$ for all $2 \leq k \leq d$. Consider the system

$$\begin{bmatrix} s_{0j} \\ \vdots \\ s_{e_d j} \\ t_{00 j} \end{bmatrix} = \begin{bmatrix} A_{\{0,e_2,\ldots,e_d\}\times(\{0\}\cup[j-1])} & 0 \\ 0 & a_{00}^2 \end{bmatrix} \begin{bmatrix} S_{(\{0\}\cup[j-1])\times\{i_1,\ldots,i_d,j\}} \\ t_{00 i_1} \quad 0 \cdots \quad 0 \quad t_{00 j} \end{bmatrix} \cdot \begin{bmatrix} a_{ji_1} \\ \vdots \\ a_{ji_d} \\ a_{jj} \end{bmatrix}$$

$$= \begin{bmatrix} a_{00} s_{0 i_1} & 0 & \ldots & 0 & a_{00} s_{0j} \\ 0 & a_{e_2 e_2} s_{e_2 i_2} & \ldots & 0 & a_{e_2 e_2} s_{e_2 j} \\ & \vdots & & & \\ 0 & \ldots & 0 & a_{e_d e_d} s_{e_d i_d} & a_{e_d e_d} s_{e_d j} \\ a_{00}^2 t_{00 i_1} & 0 & \ldots & 0 & a_{00}^2 t_{00 j} \end{bmatrix} \cdot \begin{bmatrix} a_{ji_1} \\ \vdots \\ a_{ji_d} \\ a_{jj} \end{bmatrix}.$$

As before, let A satisfy $a_{kk} \neq 0, 1$ for all k and $a_{ji_1} \neq 0$. Then, by the equitrek rule, we have $s_{e_k i_k}, t_{e_k e_k i_k} \neq 0$ for all $1 \leq k \leq d$ and $s_{0j}, t_{00j} \neq 0$, since there is an equitrek between e_k and i_k for each k.

Hence, the upper-left $d \times d$ block of this matrix is invertible for this choice of A. To prove that the entire matrix is invertible, it suffices to check that the Schur complement

with respect to this block, which is a 1×1 matrix, is nonzero under these conditions. The Schur complement is

$$a_{00}^2 t_{00j} - \frac{a_{00}^2 t_{00i_1} s_{0j}}{s_{0i_1}}.$$

For the sake of contradiction, suppose that the Schur complement is zero. Clearing denominators and dividing through by a_{00}, which is nonzero by assumption, yields

$$s_{0i_1} t_{00j} - s_{0j} t_{00i_1} = 0. \tag{2.B.3}$$

Since G is a polytree, the only path from 0 to j passes through i_1. Therefore, we have

$$s_{0j} = \frac{a_{00} a_{ji_1} s_{0i_1}}{1 - a_{00} a_{jj}} \qquad \text{and} \qquad t_{00j} = \frac{a_{00}^2 a_{ji_1} t_{00i_1}}{1 - a_{00}^2 a_{jj}}.$$

Plugging these into (2.B.3), we obtain

$$s_{0i_1} t_{00i_1} \left(\frac{a_{00}^2 a_{ji_1}}{1 - a_{00}^2 a_{jj}} - \frac{a_{00} a_{ji_1}}{1 - a_{00} a_{jj}} \right) = 0.$$

Clearing denominators and dividing through by $s_{0i_1} t_{00i_1}$, which is assumed to be nonzero, gives

$$0 = (1 - a_{00}a_{jj}) a_{00}^2 a_{ji_1} - (1 - a_{00}^2 a_{jj}) a_{00} a_{ji_1} = a_{00}^2 a_{ji_1} - a_{00} a_{ji_1} = a_{ji_1} a_{00} (a_{00} - 1).$$

This contradicts our assumption that $a_{ji_1} \neq 0$ and $a_{00} \neq 0, 1$. Hence, the Schur complement is nonzero, and the matrix is invertible, as required. □

2.C. Proofs for Section 2.5

We provide the proof for the entries of the modified Jacobians, described in Proposition 2.5.3 and Proposition 2.5.4.

Proof of Proposition 2.5.3. By equation (2.5.9) and the fact that $(I - A \otimes A)^{-1} \mathrm{vec}(\Omega^{(2)}) = \mathrm{vec}(S)$, the entry of the Jacobian indexed by $(ij), (\alpha \to \beta)$ is given by the following sum:

$$\begin{aligned} J_2^G(S, a)_{(ij),\alpha\to\beta} &= \sum_{l=0}^{p-1} \sum_{k=0}^{p-1} (A \otimes E_{\beta\alpha} + E_{\beta\alpha} \otimes A)_{ip+j,lp+k} \cdot s_{lk} \\ &= \sum_{l=0}^{p-1} \sum_{k=0}^{p-1} s_{lk} \left(a_{il} \delta_{(j,k)}(\beta, \alpha) + \delta_{(i,l)}(\beta, \alpha) a_{jk} \right) \\ &= \delta_j(\beta) \sum_{l=0}^{p-1} a_{il} s_{l\alpha} + \delta_i(\beta) \sum_{k=0}^{p-1} a_{jk} s_{k\alpha}, \end{aligned}$$

as stated in the statement of the proposition. To get to the trek description, note that $s_{l\alpha} = \sum_{\tau \in \mathcal{T}(l,\alpha)} m_\tau$, where the sum is over all equitreks between l and α as given by the trek rule (Proposition 2.3.2). Multiplying this by a_{il} corresponds to adding the edge $l \to i$ to the trek and thereby transforming the equitrek between l and α to a trek between i and α where the path going to i is one longer than the path going to α. □

Proof of Proposition 2.5.4. By equation (2.5.10) and the fact that $(I - A \otimes A \otimes A)^{-1}\text{vec}(\Omega^{(3)}) = \text{vec}(T)$, the entry of the Jacobian indexed by $(ijk), (\alpha \to \beta)$ is given by the following sum:

$$
\begin{aligned}
&J_3^G(T,a)_{(ijk),\alpha\to\beta} \\
&= \sum_{l=0}^{p-1}\sum_{m=0}^{p-1}\sum_{n=0}^{p-1}((E_{\beta\alpha}\otimes A)\otimes A + (A\otimes E_{\beta\alpha})\otimes A + (A\otimes A)\otimes E_{\beta\alpha})_{(ip+j)p+k,(lp+m)p+n}t_{lmn} \\
&= \sum_{l=0}^{p-1}\sum_{m=0}^{p-1}\sum_{n=0}^{p-1}t_{lmn}\left((E_{\beta\alpha}\otimes A)_{ip+j,lp+m}a_{kn} + (A\otimes E_{\beta\alpha})_{ip+j,lp+m}a_{kn} + (A\otimes A)_{ip+j,lp+m}\delta_{(k,n)}(\beta,\alpha)\right) \\
&= \sum_{l=0}^{p-1}\sum_{m=0}^{p-1}\sum_{n=0}^{p-1}t_{lmn}\left(\delta_{(i,l)}(\beta,\alpha)a_{jm}a_{kn} + a_{il}\delta_{(j,m)}(\beta,\alpha)a_{kn} + a_{il}a_{jm}\delta_{(k,n)}(\beta,\alpha)\right) \\
&= \delta_i(\beta)\sum_{m=0}^{p-1}\sum_{n=0}^{p-1}a_{jm}a_{kn}t_{\alpha mn} + \delta_j(\beta)\sum_{l=0}^{p-1}\sum_{n=0}^{p-1}a_{il}a_{kn}t_{l\alpha n} + \delta_k(\beta)\sum_{l=0}^{p-1}\sum_{m=0}^{p-1}a_{il}a_{jm}t_{lm\alpha},
\end{aligned}
$$

which matches the formula stated in the proposition. Arriving at the trek description is entirely analogous to the covariance case (Proposition 2.5.3). □

Below we provide the lemmas required for the main proof of local identifiability using the Jacobian matrix. Lemma 2.C.1 provides a sufficient condition for a graph to be a generalized two star. The base cases of Theorem 2.5.6 are covered by Lemma 2.C.2 and Lemma 2.C.4.

Lemma 2.C.1. *Let $G = (V, E)$ be a connected undirected graph on at least six nodes where every pair of paths of vertex length 3 intersect by at least one vertex. Then G is a generalized two star.*

Proof. Assume that there are at least three different vertex paths of length 3; otherwise the statement is trivial.

Let n be the length of the longest path in G with no repeated vertices. The proof proceeds by checking possible values for n, which are $n = 3, 4$, or 5. If n was larger, this path could be split into two non-intersecting paths of vertex length at least three.

If $n = 3$, any two paths must share the middle vertex. Indeed, if they shared only the beginning or end vertex, this would result in a path of vertex length 4, which is a contradiction. Thus, all distinct vertex paths must share the same center vertex. In this case, the graph is a star.

If $n = 4$, consider a path of vertex length 4, $0-1-2-3$ in the graph. Then any other vertex in the graph cannot be connected to 0 or 3 in this path, since otherwise there would be a path of vertex length 5. Thus, any remaining vertices in the graph have to be connected to vertex 1 or 2. If there is a remaining vertex 4 that is connected to 1, there cannot be another vertex 5 connected to 2, since otherwise $0-1-4$ and $3-2-5$ would be two paths of vertex length 3 which do not intersect. Furthermore, the vertex 4 could not be connected to both 1 and 2 because then there would be a path of vertex length 5. Thus, any additional vertices in the graph have to be connected to 1 or 2 but only one of them, making this one the center of the graph. The center is unique because in order for the graph to have at least six vertices there must be at least two nodes aside from

the given vertex path of length four. Without loss of generality, we let 1 be the center. If 4 is connected to 1 then arguing as before any additional vertex can be connected to 1 or a single vertex can be connected to 4, but at most one since otherwise two vertex disjoint paths of vertex length three are created.

If $n = 5$, consider a path of vertex length 5, $0 - 1 - 2 - 3 - 4$. Arguing as in the $n = 4$ case, it is only possible for any remaining vertices to be connected to the graph by vertex 2 (so it is possible as above to add a single vertex or a path of two (potentially connected) vertices going out from 2) since otherwise a path of length 6 is created or two vertex disjoint paths of length 3 are created. Thus, any two paths of vertex length three in this graph must intersect at node 2. □

The calculations for the following Lemma were carried out in Maple, with code available at: https://github.com/cecilie2424/Local-Identifiability-in-Non-Gaussian-Discrete-Lyapunov-Models.

Lemma 2.C.2. *Let $G = (V, E)$ be any directed graph on three, four or five nodes with all self-loops. Then, the matrix $\mathcal{J}_{\text{off}}^{G}(S, T)$ has full rank generically, so G is locally identifiable.*

Proof. We consider all connected polytrees on 3, 4 and 5 nodes. We show that when A has the sparsity pattern of such a graph, the off-diagonal modified Jacobian for the complete graph has full column rank ($= p^2$). This shows that for any graph G' on p nodes which has a given polytree on p nodes as a subgraph, its off diagonal modified Jacobian, $\mathcal{J}_{\text{off}}^{G'}(S, T)$, will have full column rank. Therefore, G' is locally identifiable. To see this, remove the columns from the complete graph which are not in G'. Since the off-diagonal modified Jacobian of the entire graph had full column rank ($= p^2$), it will just drop by 1 in column rank when removing a column until we are only left with columns corresponding to edges in G', which then has full column rank equal to the number of edges in E'. The chosen A is valid for any graph which has the specific polytree as a subgraph. In this way we are guaranteed to check any connected graph if we check every polytree since any connected graph has at least one polytree as a subgraph.

In order to ensure that all possible polytrees are checked (some possibly more than once), we consider all undirected unlabeled trees on three, four and five nodes. There are one, two and three respectively (consider Fig. 2.C.1 without labeling). Then all unlabeled polytrees on three, four and fives nodes will be included in the polytrees created by choosing all possible combinations of directions of the edges with skeleton equal to the undirected trees. This creates $4, 16$ and 48 different A's to check to make sure that all possible directed trees are checked at least once. We choose a labeling, see Fig. 2.C.1, to be able to pick specific A's. The matrices are created and checked in the maple code. They are all chosen such that they have $a_{ii} = 1/2,\ i = 1, \dots, p$ on the diagonal and the off-diagonal entries corresponding to existing edges in the given polytree set to one. See Example 2.C.5 to see such an A in the case of a graph on two nodes. □

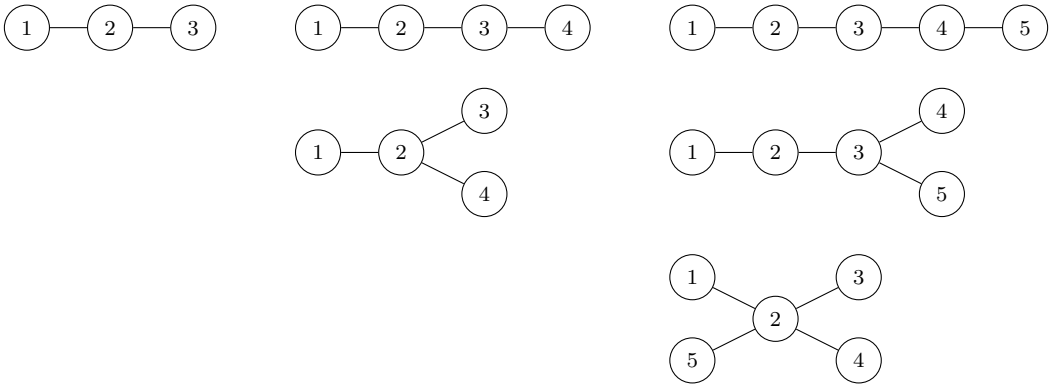


Figure 2.C.1: The different structures for undirected trees (with labels) on three, four and five nodes.

Lemma 2.C.3. *Let $G = (\{0\} \cup [p-1], E)$ be a directed graph on at least three nodes whose undirected skeleton is a generalized two star. Furthermore, assume all nodes not equal to the center either have an edge from it to the center node or is connected (direction irrelevant) to another node that has an edge from it to the center. Then $\mathcal{J}_{\text{off}}^{G}(S, T)$ generically has full rank, and the graph is locally identifiable.*

Proof. The proof proceeds by picking a matrix A and showing that the modified Jacobian has full rank for this choice. We choose a zero pattern for A such that all outgoing edges from 0 are set to 0. This implies that the only treks in the graph are between 0 and any other node $i \in [p]$, between a node i and itself, and lastly in a generalized two star there can exist an edge between a pair of non-center nodes i, j as long as their only other edges (other than the self-loops) are edges to the center node. Therefore, there can also exist pairs $i, j \in [p]$, $i, j \neq 0$, such that there is a trek between i and j, but no treks between i (resp. j) with any other node not equal to $i, j, 0$.

Assume that there are k such pairs of nodes, and assume, without loss of generality, that they are the first $1, \ldots, 2k$ nodes such that i is paired with $i+1$ for i an odd number less than $2k$. We refer to these as the *pair nodes*. The remaining $m = p - 1 - (2k)$ non-center nodes are not connected to any other nodes except the center; we refer to these as the *pure star nodes*.

We make the following partition of the rows and columns. For each odd integer i between 1 and $2k - 1$, let E_i denote the set of edges involving at least one node from the ith pair, that is,

$$E_i = \{x \to y \in E \mid x, y \in \{0, i, i+1\}\} \setminus \{0 \to 0\}.$$

For every pair except the first, we select a subset of rows

$$\begin{aligned} R_i = &\{(0i), (0(i+1)), (i(i+1)), \\ &(00i), (00(i+1)), (0ii), (0i(i+1)), (0(i+1)(i+1)), (ii(i+1)), (i(i+1)(i+1))\} \end{aligned}$$

of cardinality $|E_i|$ such that the modified Jacobian on three nodes corresponding to the three node graph on $0, i$ and $i+1$ has full rank, as guaranteed to exist by Lemma 2.C.2. The corresponding column set is

$$C_i = \{x \to y \in E_i\},$$

for i taking odd values from 3 to $2k-1$.

For the first pair, we include the edge $0 \to 0$ in its columns and therefore need to include one additional row. Thus, it will be a subset of

$$R_1 = \{(01), (02), (12), (001), (002), (011), (012), (022), (112), (122)\}$$

of cardinality $|E_i|+1$ such that the modified Jacobian on three nodes corresponding to the three node graph on $0, 1$ and 2 has full rank, as guaranteed to exist by Lemma 2.C.2. The corresponding column set is

$$C_1 = \{x \to y \in E_1\} \cup \{0 \to 0\}.$$

For the pure star part, namely the nodes $(p-1)-(2k), \ldots, p-1$, we define a row and column set for each node j as follows. We choose any subset of

$$R_j = \{(0j), (00j), (0jj)\},$$

of cardinality equal to the number of edges involving j (at least two and possibly three). For the column set, we pick all the edges involving j. By assumption, this set always includes the self-loop and the incoming edge to 0, and may additionaly include the outgoing edge from 0. Thus,

$$C_j = \{j \to j, j \to 0\} \cup (E \cap \{0 \to j\}).$$

If there is at least one pair, then the specified submatrix of the modified Jacobian will be block triangular:

$$\begin{array}{c} \\ R_{p-1} \\ \vdots \\ R_{2k+1} \\ R_{2k-1} \\ R_{2k-3} \\ \vdots \\ R_1 \end{array} \begin{array}{c} \begin{array}{ccccccc} C_{p-1} & \cdots & C_{2k+1} & C_{2k-1} & C_{2k-3} & \cdots & C_1 \end{array} \\ \begin{pmatrix} J^S_{R_{p-1},C_{p-1}} & & & & & & \\ & \ddots & & & & & \\ & & J^S_{R_{2k+1},C_{2k+1}} & & & & \\ & & & J^P_{R_{2k-1},C_{2k-1}} & & & \\ & & & & J^P_{R_{2k-3},C_{2k-3}} & & \\ & 0 & & & & \ddots & \\ & & & & & & J^P_{R_1,C_1} \end{pmatrix} \end{array}$$

The zero-pattern follows from Proposition 2.5.3 and Proposition 2.5.4 as in the proof of Theorem 2.5.6. Moreover, above the diagonal, nonzero entries can occur only in column C_1. By Lemma 2.C.2, all the pair matrices $J^P_{R_i,C_i}$ for i all the odd numbers from

1 to $2k-1$ have full rank. For all the pure star matrices $J^S_{R_i,C_i}$, it follows by direct computation in Example 2.C.5 that they have full rank, since each corresponds to a specific submatrix (with at most three columns) of the joint Jacobian for a connected graph on two nodes with both self-loops. This concludes the proof in the case where there is at least one pair.

If there are no pairs, so that the graph has a star skeleton, we need to place the $0 \to 0$ column in one of the column sets corresponding to a single node, and correspondingly add a row. We choose the nodes set, which now corresponds to node 1, since there were no pairs. We select a subset of

$$R_1 = \{(01), (001), (011), (122)\}$$

of cardinality equal to the number of edges involving node 1 (at least two and possibly three) plus one, where at least one of the rows is (122). The corresponding column set is

$$C_1 = \{1 \to 1, 1 \to 0, 0 \to 0\} \cup (E \cap \{0 \to j\}).$$

In this case, the specified submatrix of the modified Jacobian will be block triangular and takes the form

$$\begin{array}{c} \\ R_{p-1} \\ \vdots \\ R_2 \\ R_1 \end{array} \begin{array}{c} \begin{array}{cccc} C_{p-1} & \dots & C_2 & C_1 \end{array} \\ \begin{pmatrix} J^S_{R_{p-1},C_{p-1}} & & & \\ & \ddots & & \\ 0 & & J^S_{R_2,C_2} & \\ & & & J_{R_1,C_1} \end{pmatrix} \end{array}$$

Again, the zero-pattern follows from Proposition 2.5.3 and Proposition 2.5.4. All the matrices on the block diagonal except for J_{R_1,C_1} were already shown to have full rank in the previous case. For J_{R_1,C_1}, notice that it is block triangular of the form

$$\begin{array}{c} \\ (01) \\ (001) \\ (011) \\ (122) \end{array} \begin{array}{c} \begin{array}{cccc} 0 \to 0 & 1 \to 0 & 1 \to 1 & 0 \to 1 \end{array} \\ \begin{pmatrix} & \mathcal{J}^{G'}_{\text{off}}(S,T) & & \\ & & & \\ 0 & & & J^G_3(T,a)_{(122),(0\to 1)} \end{pmatrix} \end{array},$$

where G' is the graph on nodes 0 and 1 with both self-loops and the edge $1 \to 0$. This matrix has generically full rank because $\mathcal{J}^{G'}_{\text{off}}(S,T)$ has full rank, as G' is locally identifiable by Example 2.C.5, and $J^G_3(T,a)_{(122),(0\to 1)}$ is generically non-zero in the graph G due to existence of equitreks between 0 and 2. □

Lemma 2.C.4. *Let $G = (V, E)$ be a directed graph on $p \geq 3$ nodes whose undirected skeleton is a generalized two star. Then $\mathcal{J}^G_{\text{off}}(S,T)$ generically has full rank, and the graph is locally identifiable.*

Proof. The proof proceeds by induction. The base case contains generalized two stars (including the stars as a special case), which are covered by Lemma 2.C.3, as well as generalized two stars on at most five nodes, which are handled by Lemma 2.C.2.

Now, let $p \geq 6$ and let $G = (\{0\} \cup [p-1], E)$ be a generalized two star on p nodes, with 0 denoting the center. If G is a star or a generalized two star such that all non-center nodes either have an edge from it to the center or is connected (direction irrelevant) to another node that has an edge from it to the center, then it is covered by Lemma 2.C.3. Therefore, assume that G is a graph not covered by the base case. This implies that one of the five two- or three-node subgraphs in Fig. 2.C.2 is a subgraph of G; denote this subgraph by H. Moreover, if we remove the nodes not equal to the center, the remaining graph would have at least three nodes. By the induction hypothesis, if the subgraph H is removed (except 0 and its self-loop), this remaining graph G' is locally identifiable, since it has $p-1$ or $p-2$ nodes.

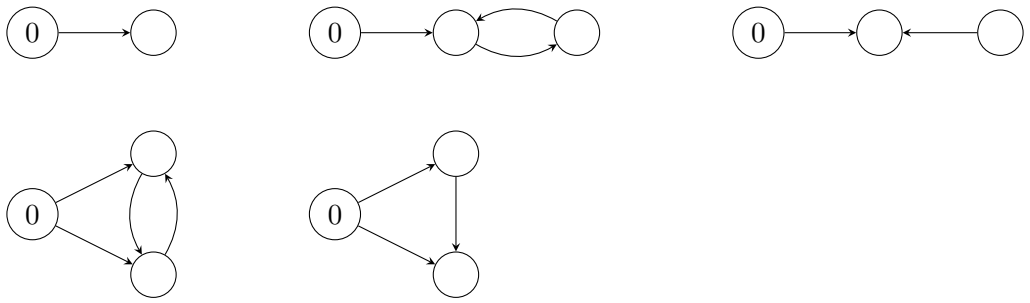


Figure 2.C.2: Possible types of extensions from the center in the induction step, shown without self-loops.

Since all edges involving 0 in the graphs from Figure 2.C.2 are outgoing, adding any of these five subgraphs to G' (to obtain G) will not change any of the treks among the nodes in G', and therefore does not affect the corresponding cumulants, Jacobian or modified Jacobian. Consequently, by the induction hypothesis, all edges in the subgraph G' are indeed locally identifiable in G.

The final step is to conclude that the remaining edges are locally identifiable. There are at least two and at most six such edges, depending on the case from Fig. 2.C.2 (including self-loops). Consider the Jacobian of the second- and third-order cumulants involving only the center node 0 and the potential one or two extra nodes, denoted α and β. As before, it suffices to consider the non-diagonal cumulant rows of the modified Jacobian. The relevant cumulants are indexed by

$$R = \{(0\alpha), (0\beta), (\alpha\beta), (00\alpha), (00\beta), (0\alpha\alpha), (0\alpha\beta), (0\beta\beta), (\alpha\alpha\beta)(\alpha\beta\beta)\},$$

or, in the case where only one node is added, only the ones involving α and 0. In all cases, there are more cumulants than edges added. Furthermore, these cumulants can be viewed as equations only in the new edges (those between 0, α and β), with the remaining variables being either other cumulants or known a-parameters which are

already identified. Here, we highlight that all the cumulants not involving combinations of 0, α and β were not affected by adding α and β. Thus, to estabilsh local identifiability of the entire graph, it suffices to prove local identifiability of the edges between 0, α and β, assuming the edges in G' are already known.

Therefore, we consider the submatrix of the modified Jacobian indexed by rows in R and columns corresponding to edges between 0, α and β (excluding $0 \to 0$), denoted $\mathcal{J}_{\text{off}}^{G\setminus G'}$. We wish to prove that this submatrix generically has full rank. Let us choose A generically such that all edges not involving only 0, α and β are set to zero. Then the only treks will be among 0, α and β, and $\mathcal{J}_{\text{off}}^{G\setminus G'}$ will be exactly equal to the submatrix of the modified Jacobian of H (on two or three nodes), $\mathcal{J}_{\text{off}}^{H}$, with the column $0 \to 0$ removed. Since H is either a DAG or a graph on three nodes, $\mathcal{J}_{\text{off}}^{H}$ has full column rank by Lemma 2.C.2 and Theorem 2.4.3. It will also have full column rank if the $0 \to 0$ column is removed. Therefore, $\mathcal{J}_{\text{off}}^{G\setminus G'}$ generically has full column rank and the edges among 0, α and β (except $0 \to 0$) can be locally identified, assuming all edges in G' had already been locally identified, as guaranteed by the induction hypothesis. This completes the induction step. □

Example 2.C.5. Let $G = (\{0, 1\}, E)$ be either the complete graph on two nodes or the complete graph without the edge $1 \to 0$. Then any 3×3 submatrix of the $3 \times |E|$ matrix $\mathcal{J}_{\text{off}}^{G}(S, T)$ generically has rank 3. We see this by letting

$$A = \begin{pmatrix} 1/2 & 0 \\ 1 & 1/2 \end{pmatrix},$$

and computing the rank for all 3×3 submatrices of $\mathcal{J}_{\text{off}}^{G}(S, T)$ in Maple[2].

Lemma 2.C.6. *The matrix $M_{G_i \to g_k^j}$ with $i \neq j$, $i, j \in \{1, 2\}$, which appears in the proof of Theorem 2.5.6 as a submatrix of the modified Jacobian, generically has full rank.*

Proof. Let A be diagonal. Then $M_{G_i \to g_k^j}$ is also diagonal with generically non-zero diagonal entries: By Proposition 2.5.4, $M_{G_i \to g_k^j}$ is an $n_i \times n_i$ matrix with rows indexed by $R_{G_i \to g_k^j}$ and columns indexed by $C_{G_i \to g_k^j}$, where the (l, m)-th entry is given by

$$\left(M_{G_i \to g_k^j}\right)_{l,m} = \left(J_3^G(T, a)\right)_{(g_l^i g_l^i g_k^j),(g_m^i \to g_k^j)} = \sum_{x \in \text{pa}(g_l^i)} \sum_{y \in \text{pa}(g_l^i)} a_{g_l^i x} a_{g_l^i y} t_{g_m^i xy}.$$

Thus, the only way for the above to be non-zero for diagonal A is exactly when $g_l^i = x = y = g_m^i$, since A being diagonal implies that T is diagonal by the trek rule. Consequently, $M_{G_i \to g_k^j}$ is indeed diagonal, with diagonal entries $a_{g_l^i g_l^i}^2 t_{g_l^i g_l^i g_l^i}$ for $l = 1, \ldots, n_i$. For A diagonal with non-zero diagonal entries (and $\Omega^{(3)}$ non-zero diagonal), these diagonal entries are non-zero by the trek rule. Therefore, $M_{G_i \to g_k^j}$ generically has full rank. □

[2] https://github.com/cecilie2424/Local-Identifiability-in-Non-Gaussian-Discrete-Lyapunov-Models.

2.D. Proofs for Section 2.6

2.D.1 Proofs for Section 2.6.1

Proof of Proposition 2.6.3. For a vertex $i \in V$, let $|\tau_i|$ denote the length of the shortest path from 0 to i, and let a^{τ_i} denote the corresponding path monomial. In the following, we assume that $v_i^{(2)}$, $v_i^{(3)}$, and a_{ij} do not vanish for all $i, j \in V$, since the equations hold generically.

1. By the monomial parametrization,
$$s_{ij} = v^{(2)}_{top(\tau(i,j))} a^{\tau_1} a^{\tau_2}, \qquad s_{ii} = v_i^{(2)}.$$
Since each a_{ij} is nonzero, the polynomials $s_{ij}^3 t_{iii}^2 - s_{ii}^3 t_{iij} t_{ijj}$ vanish for every j if and only if for every j, $v^{(2)}_{top(\tau(i,j))} = v_i^{(2)}$, which is equivalent to i being the source.

2. By the monomial parametrization,
$$s_{0i} = v_0^{(2)} a_{00}^{|\tau_i|} a^{\tau_i}, \qquad t_{00i} = v_0^{(3)} a_{00}^{2|\tau_i|} a^{\tau_i}.$$
Then the polynomial becomes
$$v_0^{(2)} a_{00}^{|\tau_i|} a^{\tau_i} \cdot v_0^{(3)} a_{00}^{2|\tau_j|} a^{\tau_j} - v_0^{(2)} a_{00}^{|\tau_j|} a^{\tau_j} \cdot v_0^{(3)} a_{00}^{2|\tau_i|} a^{\tau_i} = 0.$$
Since each a_{ij}, $v_i^{(2)}$ and $v_i^{(3)}$ is nonzero, this holds if and only if $a_{00}^{|\tau_i|} - a_{00}^{|\tau_j|} = 0$, which implies that $|\tau_i| = |\tau_j|$.

3. Since i and j are on different levels, all equitreks between them must start at the source. Therefore,
$$s_{ij} = v_0^{(2)} a^{\tau_i} a^{\tau_j} a_{00}^{||\tau_i| - |\tau_j||}.$$
The polynomial becomes
$$v^{(2)} a^{\tau_i} a^{\tau_j} a_{00}^{||\tau_i| - |\tau_j||} \cdot v_0^{(3)} a_{00}^{2|\tau_j|} a^{\tau_j} - v_0^{(2)} a_{00}^{|\tau_j|} a^{\tau_j} \cdot v_0^{(3)} a^{\tau_i} a^{\tau_j} a_{00}^{\max(|\tau_i|,|\tau_j|) + ||\tau_i| - |\tau_j||} = 0$$
The nonvanishing of the parameters again implies that $a_{00}^{|\tau_j|} - a_{00}^{\max(|\tau_i|,|\tau_j|)} = 0$. So we conclude that $|\tau_j| = \max(|\tau_i|, |\tau_j|)$. □

Proof of Proposition 2.6.5. $\Leftarrow$: Suppose l is the top of the shortest equitrek between i and j, and let τ_i and τ_j denote the paths from l to i and j, respectively. By the monomial parametrization, we have
$$s_{ij} = s_{ll} a^{\tau_i} a^{\tau_j}, \qquad s_{0i} = s_{0l} a_{00}^{|\tau_i|} a^{\tau_i}, \qquad t_{ljj} = t_{lj} \frac{a^{\tau_j}}{a_{00}^{|\tau_i|}}.$$
As a result,
$$s_{0l} s_{ij} t_{llj} - s_{0i} s_{ll} t_{ljj} = s_{0l} s_{ll} a^{\tau_i} a^{\tau_j} t_{llj} - s_{0l} a_{00}^{|\tau_i|} a^{\tau_i} s_{ll} t_{llj} \frac{a^{\tau_j}}{a_{00}^{|\tau_i|}} = 0.$$

$\Rightarrow$: When writing the parametrization of variables, note that $s_{ll} = v_l^{(2)}$. For the polynomial to vanish, this parameter $v_l^{(2)}$ must appear in the first term with degree one. The variables in the first terms are s_{0l}, t_{llj} and s_{ij}, so the parameter $v_l^{(2)}$ might only appear from the parametrization of s_{ij}. That implies that l is the top of the shortest equitrek between i and j. □

Proof of Lemma 2.6.7. Let k denote the top vertex of the shortest equitrek between i and j. Let A and B be the weighted adjacency matrices of G and H, respectively. Let $p(i)$ and $p(j)$ denote the parents of i and j in G, and let $c(i)$ and $c(j)$ denote the children of i and j in G, if they exist. The matrices A and B differ in eight entries. After swapping the edges, the entries $a_{ip(i)}, a_{jp(j)}, a_{c(i)i}$ and $a_{c(j)j}$ become zero, while the new edges $b_{jp(i)}, b_{ip(j)}, b_{c(j)i}, b_{c(i)j}$ appear in B. In the special case where $p(i) = p(j)$, only four elements change, since $a_{ip(i)} = b_{ip(j)}$ and $a_{jp(j)} = b_{jp(i)}$; however, this case is already covered by the general one.

Figure 2.D.1: Swapping operation for vertices i and j.

Let P_G and P_H denote the $(p + |E(G)|) \times \frac{p(p-1)}{2}$ shortest equitrek parametrization matrices of the second-order cumulant models corresponding to G and H, respectively. The first p rows are labeled by $v_i^{(2)}$ for $i \in \{0\} \cup [p-1]$, and the remaining rows are labeled by edges. We will show that these matrices are row equivalent. A similar reasoning works for higher-order cumulant models.

Notice that the first p rows of P_G and P_H coincide, as the top vertices of all shortest equitreks remain unchanged after the swapping operation. Indeed, consider two vertices $u, v \in V(G)$.

- If u and v have different levels, then the top vertex of their shortest equitrek is 0. Since the swap does not alter vertex levels, u and v still lie at different levels in H, and their top vertex remains 0.
- If u and v have the same level and are located above i and j, then the top of their shortest equitrek also does not change after the swap, since the equitrek between them does not use any of the changed edges.
- Suppose u and v lie on the same level and are located below i and j. If the shortest equitrek in G passes through both i and j, then its top vertex must be k, which remains unchanged after the swap. Both paths of the trek can not pass through either i or j, since these vertices have at most one child; such a trek would therefore

not be the shortest. One of the paths might pass through i or j, but then there exists a corresponding trek in H, which uses the swapped edges.

- The last case is when u and v lie on the same level as i and j. If they do not coincide with neither i nor j, then the shortest equitrek between them in G does not use any of the edges affected by the swap, and hence they have the same equitrek in H. If the pair u, v coincides with i, j, then the top of the shortest equitrek remains unchanged after swapping. Finally, suppose without loss of generality that $u = i$ and $v \neq j$. Let k' denote the top of the trek between i and v in G. By the statement, k' lies above k, and the path from k' to i passes through k. A corresponding shortest equitrek between i and v in H can then be constructed as follows: the path from k' to v remains unchanged, while the path from k' to i follows the path from k' to k, then to $p(j)$, and finally uses the swapped edge to reach i.

Some rows of P_H coincide with some rows of P_G. We now identify all such equal rows. First, observe that $(P_G)_{a_{00}} = (P_H)_{b_{00}}$. For a column s_{uv}, the corresponding entry in this row corresponds to the number of self-loops used in a trek between u and v, which is determined by the levels of the vertices.

A row corresponding to a_{uv} in G is equal to a row corresponding to b_{uv} in H if the edge is not adjacent to either i or j. This follows since for each column index s_{xy}, the entry $(P_G)_{a_{uv},s_{xy}}$ equals the number of times the edge a_{uv} is used in the shortest equitrek between x and y.

Finally, by the same reasoning, the following rows also coincide:

$$(P_H)_{b_{c(i)j}} = (P_G)_{a_{c(i)i}}, \qquad (P_H)_{b_{c(j)i}} = (P_G)_{a_{c(j)j}}.$$

The only remaining rows that change correspond to edges leading to i and j from their parents. We claim that

$$(P_H)_{b_{ip(j)}} = (P_G)_{a_{ip(i)}} - (P_G)_{a_{c(i)i}} + (P_G)_{a_{c(j)j}}.$$

We verify this equality column by column. Fix some column corresponding to s_{uv}.

- First, suppose that $(P_H)_{b_{ip(j)},s_{uv}} = 0$. Then none of the paths of the equitrek between u and v pass through i in H. This implies that none of the paths of the equitrek pass through $c(j)$ in H, and therefore none pass through $c(j)$ in G. Hence, $(P_G)_{a_{c(j)j},s_{uv}} = 0$. Moreover, $(P_G)_{a_{ip(i)},s_{uv}} = (P_G)_{a_{c(i)i},s_{uv}}$ unless either u or v coincides with i. But if u or v were i, then the corresponding entry $(P_H)_{b_{ip(j)},s_{uv}}$ would be nonzero, which contradicts the assumption. Therefore, the right-hand side also evaluates to zero.

- Next, suppose that $(P_H)_{b_{ip(j)},s_{uv}} = 1$. This means that one of the paths of the trek between u and v crosses i in H. This happens if, for example, the path of the trek stops there, so one of u and v coincides with i. In this case, $(P_G)_{a_{c(j)j},s_{uv}} = 0$, since none of the treks go through $c(j)$. Moreover, $(P_G)_{a_{ip(i)},s_{uv}} = (P_G)_{a_{c(i)i},s_{uv}} + 1$, so the claimed equality holds in this case.

In the other case, neither u nor v coincide with i. Then one of the paths must pass through $c(j)$ in H. This implies that in G, one of the paths also passes though $c(j)$, so $(P_G)_{a_{c(j)j},s_{uv}} = 1$. On the other hand, the entries of the rows $(P_G)_{a_{ip(i)}}$ and $(P_G)_{a_{c(i)i}}$ are equal.

- Finally, suppose that $(P_H)_{b_{ip(j)},s_{uv}} = 2$. This means that both paths of the equitrek pass through i in H. By the same reasoning as above, equality holds.

We have verified that the claimed equality holds for all columns of $(P_H)_{b_{ip(j)}}$. By an analogous argument, the remaining changed row satisfies

$$(P_H)_{b_{jp(i)}} = (P_G)_{a_{jp(j)}} - (P_G)_{a_{c(j)j}} + (P_G)_{a_{c(i)i}}.$$

This shows that P_G and P_H are row equivalent, and the explicit row equivalence has been identified.

Now suppose that P_G and P_H are the parametrization matrices for cumulants up to order k. These matrices have $kp + |E(G)|$ rows. Again, the first kp rows are equal, since the top vertices of the equitreks remain unchanged after the edge swap. For the rows corresponding to edges, we need to show that the same relations hold as in the second-order case described above. This can be checked column by column in a similar way. Indeed, in the new columns corresponding to cumulants of order three and higher, the only difference is that we consider equitreks between i-tuples of vertices for $3 \leq i \leq k$, and the argument above still holds. □

Proof of Lemma 2.6.8. $\Leftarrow$: The swapping operation does not change the equitreks and the levels, as shown in the previous lemma.
$\Rightarrow$: We proceed by induction on the number of levels. The base case covers graphs with two levels. Such graphs define equivalent models if they have the same source and the same number of vertices.

Assume the statement holds for graphs with $k-1$ levels. Let G and H be graphs with p vertices and k levels, satisfying the conditions from the statement. Let G' and H' be the graphs obtained from G and H, respectively, by removing all vertices in the last level. Then G' and H' also satisfy the conditions, and by the induction hypothesis, H' can be obtained from G' by a sequence of swapping operations.

We now explain why the same swapping operations can be applied to H. Observe that the vertices at level $k-1$ of G and H that have more than one child coincide, since every vertex with at least two children is the top of an equitrek. Moreover, if u and v are vertices at the last level of G whose equitrek has top w, then the equitrek between $pa(u)$ and $pa(v)$ has also top w. Hence, if swapping vertices in H' preserves the tops of equitreks, then the same swapping operations preserve equitrek tops in H as well.

Applying the same swapping operations to H, the resulting graph agrees with G on all levels except possibly the last one, using the induction hypothesis. Finally, additional swaps within the last level can be performed to make the two graph identical. This completes the inductive step. □

2.D.2 Proofs for Section 2.6.2

Proof of Proposition 2.6.9. The matrix S' has size $(p-k)\times k$, hence $rk(S') \leq \min(k, p-k)$. Assume that $|pa(U)| < \min(k, p-k)$. We prove the desired rank inequality by decomposing S' as a product of two matrices of appropriate sizes. Let s_{iv_j} be an entry of S'. Then $i \in V \setminus U$. For any $u \in pa(U)$ such that $u \notin pa(v_j)$, we have $a_{v_j u} = 0$. Since s_{iv_j} is not a diagonal entry, we may write

$$s_{iv_j} = \sum_{u\in pa(U)} \sum_{w\in pa(i)} a_{v_j u} a_{iw} s_{wu} = \sum_{u\in pa(U)} a_{v_j u} \sum_{w\in pa(i)} a_{iw} s_{wu} = \sum_{u\in pa(U)} a_{v_j u} K_{ui},$$

where $K_{ui} \coloneqq \sum_{w\in pa(i)} a_{iw} s_{wu}$. The quantities K_{ui} form a matrix K of size $|pa(U)| \times (p-k)$. Let A' be the $k \times |pa(U)|$ submatrix of A whose rows correspond to $v_1, \ldots, v_k$ and whose columns correspond to vertices in $pa(U)$. Then

$$S' = (A'K)^T,$$

and therefore $rk(S') \leq |pa(U)|$.

We now argue analogously for the Tensor T. Fix i and consider the slice T_i. Let T_i' be the submatrix of T_i formed by the columns $v_1, \ldots, v_k$ and excluding diagonal entries t_{jjj}. If $i \in U$, then T_i' has size $(p-1)\times k$; otherwise, it has size $p \times k$. Assume that $|an_2(U)|$ is smaller than both dimensions of T_i'.

Let t_{ijv_l} be an entry of T_i'. Then

$$t_{ijv_l} = \sum_{u\in pa(U)} \sum_{w\in pa(i)} \sum_{x\in pa(j)} a_{v_j u} a_{iw} a_{jx} t_{uwx} = \sum_{u\in pa(U)} a_{v_j u} K_{ui},$$

where $K_{ui} \coloneqq \sum_{w\in pa(i)} \sum_{x\in pa(j)} a_{iw} a_{jx} t_{uwx}$. Again, the quantities K_{ui} form a matrix K with $|pa(U)|$ rows. Let A' be the $k \times |pa(U)|$ submatrix of A, with rows indexed by $v_1, \ldots, v_j$ and columns indexed by $pa(U)$. Then

$$T_i' = (A'K)^T,$$

and consequently $rk(T_i') \leq |pa(U)|$. □

Proof of Proposition 2.6.10. By Proposition 2.6.9, we can decompose S' and each slice into a product of two matrices, A' and K. Note that the first factor A' is the same in all such decompositions. Using these factorizations, we construct a decomposition of Q by taking A' and stacking the corresponding matrices K in the appropriate order. □

Proof of Proposition 2.6.12. The matrix S' has size $(p - |sib(U)|) \times k$, hence $rk(S') \leq \min(k, p - |sib(U)|)$. Assume that $|an_2(U)| < \min(k, p - |sib(U)|)$. We establish the desired rank inequality by decomposing S' as a product of two matrices of appropriate

sizes. Let s_{iv_j} be an entry of S'. Then $i \in V \setminus sib(U)$. We compute

$$\begin{aligned}
s_{iv_j} &= \sum_{u \in pa(U)} \sum_{w \in pa(i)} a_{v_j u} a_{iw} s_{wu} \\
&= \sum_{u \in pa(U)} \sum_{w \in pa(i)} a_{v_j u} a_{iw} \left(\sum_{g \in pa(u)} \sum_{x \in pa(w)} a_{ug} a_{wx} s_{gx} \right) \\
&= \sum_{u \in pa(U)} \sum_{w \in pa(i)} a_{v_j u} a_{iw} \left(\sum_{g \in an_2(U)} \sum_{x \in pa(w)} a_{ug} a_{wx} s_{gx} \right) \\
&= \sum_{u \in pa(U)} \sum_{w \in pa(i)} \sum_{g \in an_2(U)} \sum_{x \in pa(w)} a_{v_j u} a_{iw} a_{ug} a_{wx} s_{gx} \\
&= \sum_{g \in an_2(U)} \left(\sum_{u \in pa(U)} a_{v_j u} a_{ug} \right) \left(\sum_{w \in pa(i)} \sum_{x \in pa(w)} a_{iw} a_{wx} s_{gx} \right) \\
&= \sum_{g \in an_2(U)} B_{v_j g} K_{gi}.
\end{aligned}$$

Here we define

$$B_{v_j g} \coloneqq \sum_{u \in pa(U)} a_{v_j u} a_{ug}, \qquad K_{gi} \coloneqq \sum_{w \in pa(i)} \sum_{x \in pa(w)} a_{iw} a_{wx} s_{gx}.$$

These quantities form matrices B of size $k \times |an_2(U)|$ and K of size $|an_2(U) \times p - |sib(U)|$. Consequently,

$$S' = (BK)^T,$$

and therefore $rk(S') \leq |an_2(U)|$.

Now consider a slice T_i of the tensor T. Let T_i' be the submatrix of T_i formed by the columns $v_1, \ldots, v_k$ and obtained by removing all rows containing entries t_{ijl} with $i, j, l \in sib(U)$. For an entry t_{ilv_j} of T_i', we have

$$\begin{aligned}
t_{ilv_j} &= \sum_{u \in pa(U)} \sum_{w \in pa(i)} \sum_{x \in pa(l)} a_{v_j u} a_{iw} a_{lx} t_{wux} \\
&= \sum_{u \in pa(U)} \sum_{w \in pa(i)} \sum_{x \in pa(l)} a_{v_j u} a_{iw} a_{lx} \left(\sum_{g \in an_2(U)} \sum_{y \in pa(w)} \sum_{z \in \mathrm{pa}(x)} a_{ug} a_{wy} a_{xz} t_{qyz} \right) \\
&= \sum_{g \in an_2(U)} B_{v_j g} K_{gi},
\end{aligned}$$

where

$$B_{v_j g} \coloneqq \sum_{u \in pa(U)} a_{v_j u} a_{ug}, \qquad K_{gi} \coloneqq \sum_{w \in pa(i)} \sum_{x \in pa(l)} \sum_{y \in pa(w)} \sum_{z \in pa(x)} a_{iw} a_{lx} a_{wy} a_{xz} t_{qyz}.$$

As before, these quantities form matrices B of size $k \times |an_2(U)|$ and K of compatible dimensions, and we obtain

$$T'_i = (BK)^T,$$

which implies $rk(T'_i) \leq |an_2(U)|$.

Finally, observe that both S' and each slices T'_i admit a decomposition as a product of two matrices B and K, where the first factor is the same in all cases. Using these decompositions, we construct a decomposition of Q by fixing B and stacking the corresponding matrices K in the appropriate order. □

3 Identifiability and Estimation in Continuous Lyapunov Models

Cecilie Olesen Recke, Niels Richard Hansen

Abstract

Cross-sectional observations from a dynamical system can be modeled via steady-state distributions of Markov processes. The major challenge is then to determine whether the process parameters can be identified and estimated from the steady-state distributions. We study this problem for continuous Lyapunov models that arise as steady-state distributions of the solution to a multivariate stochastic differential equation, whose linear drift matrix is parametrized by a directed graph. We derive equations for the cumulant tensors of any order for this distribution, which generalize the well-known covariance Lyapunov equation. Under a non-Gaussianity assumption we prove generic identifiability of the drift matrix for any connected graph using the equations for the higher-order cumulants. Based on the identifiability result, we propose a new semiparametric estimator of the drift matrix, and we derive its asymptotic distribution. A simulation study demonstrates the asymptotic validity of the estimator but shows that it is only accurate for relatively large sample sizes, illustrating the hardness of the unconstrained estimation problem.

Some key words: Cross-Sectional Observations; Graphical Modeling; Higher-Order Cumulants; Lyapunov Equations; Parameter Identifiability; Steady-State Distributions

3.1 Introduction

3.1.1 Background

The problem of estimating the parameters of a dynamical model from cross-sectional data has recently received considerable attention [Varando and Hansen, 2020; Lorch et al., 2024; Guan et al., 2024; Bleile et al., 2026]. An important application is learning biological mechanisms from single cell data [Wang et al., 2023; Rohbeck et al., 2024; Lorch et al., 2026], but this is challenging with access to only temporal snapshots rather than dynamic trajectories. One largely outstanding question is to what extent the parameters are identifiable from cross-sectional distributions. Partial results were given

by Dettling et al. [2023] for a Gaussian Markov process, and they showed, in particular, that the drift parameters of this process are only identifiable under strong assumptions.

We consider a similar but more general class of Markov processes and give the first positive identifiability result regarding the drift parameters when the process is non-Gaussian. This relies on a thorough treatment of the algebraic structures of higher-order cumulants determined by the model. We use our results to propose a new semiparametric estimator of the drift parameters and investigate its properties, asymptotically and in simulations.

3.1.2 Main contributions

We consider the d-dimensional steady-state distribution of the stationary Markov process that solves the stochastic differential equation (SDE)

$$dX_t = MX_t dt + dZ_t, \tag{3.1.1}$$

where M is a stable $d \times d$ matrix and $Z = (Z_t)_{t \geq 0}$ is a Lévy process. Such a distribution is known as an M-selfdecomposable distribution [Sato and Yamazato, 1984; Masuda, 2004], and we regard it as a model of a single cross-sectional observation from the stationary process. The entries of the matrix M are the drift parameters of the process, and following Varando and Hansen [2020] we associate with any directed graph G on nodes $\{1, \ldots, d\}$ a sparsity pattern of M: only entries corresponding to a directed edge in G can be non-zero, see Figure 3.2.1. For a given graph G, our focus is on identification and estimation of M from cross-sectional observations.

We first derive, in Proposition 3.2.2, the equations that determine all cumulants of the steady-state distribution from the matrix M and the cumulants $\mathcal{C}_2, \mathcal{C}_3, \ldots$ of Z_1. Our main results are then Theorem 3.3.1 and Corollary 3.3.2. For those we effectively assume that the Lévy process has independent coordinates and is not a Brownian motion. If the graph G is also connected then $(M, \mathcal{C}_2, \mathcal{C}_r)$ is *generically* identifiable up to a common scaling factor from the covariance matrix and the r-th-order cumulant tensor. The possible exception set where identifiability fails is a proper algebraic subset of the parameter set. We refer to Theorem 3.3.1 and Corollary 3.3.2 for the details, but note that identifiability up to a scaling factor is the best possible result without additional assumptions, such as assuming the cumulants of Z_1 known.

To prove Theorem 3.3.1 we derive trek rules of independent interest, collected in Appendix 3.B.2, that express how the k-th-order cumulants depend on M and $\mathcal{C}_k$. This allows us to show that the equations from Proposition 3.2.2 can generically be solved uniquely for the parameters in terms of the cumulants. These equations naturally suggest an estimator of M, up to a scaling factor. We show consistency and asymptotic normality of this estimator in Theorem 3.4.2, with the most important contribution being the explicit formula for the asymptotic covariance matrix. The Julia package `https://github.com/nielsrhansen/SteadyStateStatistics.jl` includes an implementation. A simulation study shows that while the estimator behaves asymptotically according to the theory, it has a notable finite sample bias and the sample size needs to be fairly large for accurate estimation.

3.1.3 Relations to existing literature

It is well known that the covariance matrix Σ for an M-selfdecomposable distribution solves the continuous Lyapunov equation

$$M\Sigma + \Sigma M^T + \mathcal{C}_2 = 0, \tag{3.1.2}$$

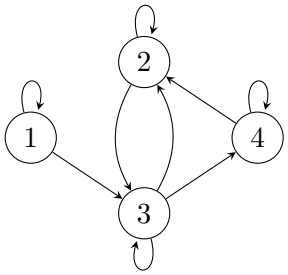
where $\mathcal{C}_2$ is the covariance matrix of Z_1. A proof is given by Jacobsen [1993] assuming the Lévy process is a Brownian motion, and similar arguments apply in general once the SDE (3.1.1) is known to have a unique stationary solution. In fact, it also follows rather easily from Theorem 4.1 by Sato and Yamazato [1984] combined with their equations (2.15) and (2.18). Our Proposition 3.2.2 generalizes this result and shows that the k-th-order cumulant solves the *k-th-order continuous Lyapunov equation*. Though this equation has been studied previously [Xu and Wang, 2022], its connection to the cumulants of M-selfdecomposable distributions is new.

Estimation of M from cross-sectional observations, assuming a sparse graph and a diagonal $\mathcal{C}_2$, was investigated by Fitch [2020], Varando and Hansen [2020] and Dettling et al. [2024] using lasso-type estimators. They only considered the covariance equation (3.1.2), effectively assuming observations to be Gaussian. Dettling et al. [2023] showed global identifiability from the second order equation (3.1.2) when $\mathcal{C}_2$ is known and the graph is simple, but they also showed by example that there exist sparse graphs where M is not even generically identifiable from (3.1.2). Still considering only the second order Lyapunov equation (3.1.2), van Seeventer and Salehkaleybar [2026] derived results on identification of just the signs of entries in M.

Young et al. [2019] and Recke et al. [2026] derived related identifiability results for discrete Lyapunov models where the cumulant equations arise from the steady-state distribution of a vector autoregressive model. Only Recke et al. [2026] considered the non-Gaussian setting.

Lorch et al. [2024] and Bleile et al. [2026] considered a setup similar to ours with cross-sectional observations from a stationary diffusion, that is, a stationary solution to an SDE driven by Brownian motion. They proposed and investigated different estimators of drift and diffusion parameters. Guan et al. [2024] also investigated estimation via cross-sectional observations from a diffusion, but they relied on observations from multiple time points from a process out of its steady-state to get identification. While these developments are closely related to the problem we consider, they are not directly applicable, because the solution to the SDE (3.1.1) is not a diffusion unless the Lévy process is a Brownian motion.

Our contributions are more inspired by the literature on identifiability of the mixing matrix in independent component analysis [Comon, 1994], and its applications to linear non-Gaussian acyclic models [Shimizu et al., 2006]. Of particular relevance are the recent contributions by Mesters and Zwiernik [2024] on identifiability and estimation of the mixing matrix using higher-order moments and cumulants. While their overall strategy is similar to ours, the algebraic techniques we use are quite different as the fundamental equations we study are different.



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

Figure 3.2.1: Example of a directed graph $G = ([4], E)$ and corresponding zero-pattern of M in $\mathbb{R}^E$

3.2 The Continuous Lyapunov Model

In this section we define the semiparametric *continuous Lyapunov model*. Then we present the cumulant defining equations that we will use to show our main identifiability results for the parameters of interest, and which form the basis of our estimator.

The model consists of d-dimensional steady-state distributions of solutions to the SDE (3.1.1), parametrized by the drift matrix M and the distribution of the Lévy process $(Z_t)_{t\geq 0}$. There is a unique such steady-state distribution when M is a stable matrix, that is, all its eigenvalues have negative real part, and when the Lévy process satisfies the weak integrability condition

$$\mathbb{E}(\log(1 + \|Z_1\|)) < \infty. \tag{3.2.3}$$

This follows from Theorem 4.1 by Sato and Yamazato [1984]. The steady-state distribution is known as an *M-selfdecomposable distribution*.

The drift matrix is the parameter of primary interest, and we will allow for it to have a particular sparsity structure according to a directed graph, see Figure 3.2.1. Formally, with $[d] = \{1, \ldots, d\}$ and with $G = ([d], E)$ a directed graph with node set $[d]$ and edge set E, we define the set

$$\mathbb{R}^E = \{M \in \mathbb{R}^{d\times d} \mid M_{ij} = 0 \text{ if } j \to i \notin E\}.$$

We denote by $\mathbb{R}^E_{\text{stab}}$ the subset of $\mathbb{R}^E$ where M is also stable.

Definition 3.2.1. *Let $G = ([d], E)$ be a directed graph. The continuous Lyapunov model is the set of probability distributions $\mathcal{P}_G$ on $\mathbb{R}^d$ that are steady-state distributions of solutions to the SDE (3.1.1) for an $M \in \mathbb{R}^E_{\text{stab}}$ and a Lévy process $(Z_t)_{t\geq 0}$ that satisfies the integrability condition (3.2.3).*

We will parametrize the distribution of the Lévy process by the distribution of Z_1. This infinitely divisible distribution, which we denote Q_1, uniquely determines the distribution of the entire process. We denote by $\mathcal{Q}$ the set of infinitely divisible distributions Q_1 on $\mathbb{R}^d$ satisfying the integrability condition

$$\int \log(1 + \|z\|))Q_1(\mathrm{d}z) < \infty. \tag{3.2.4}$$

For any graph G, the model $\mathcal{P}_G$ is therefore a set of M-selfdecomposable distributions parametrized by the finite dimensional parameter space $\mathbb{R}^E_{\text{stab}}$ and the infinite dimensional space of infinitely divisible distributions $\mathcal{Q}$. We denote this parametrization by

$$\psi : \mathbb{R}^E_{\text{stab}} \times \mathcal{Q} \to \mathcal{P}_G.$$

Provided that a distribution in $\mathcal{P}_G$ has finite k-th-order moment, we can derive a simple generalization of the second-order Lyapunov equation (3.1.2) that determines the k-th-order cumulant of the distribution. See McCullagh [2018] for background on cumulants.

In order to write these higher-order cumulant equations we use the notion of the n-mode product, also known as the Tucker product, between a tensor and a matrix. The n-mode product of a tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with a matrix $M \in \mathbb{R}^{J \times I_n}$, denoted $T \times_n M$, is an $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$ tensor with coordinate-wise entries

$$(T \times_n M)_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} T_{i_1 \cdots i_{n-1} i_n i_{n+1} \cdots i_N} M_{j i_n}.$$

Proposition 3.2.2. *Let M be a $d \times d$ stable matrix and let $Z = (Z_t)_{t \geq 0}$ denote a d-dimensional Lévy process with finite k-th-order moment. Then the corresponding M-selfdecomposable distribution has finite k-th-order moment, and the k-th-order cumulant tensor $\mathcal{K}$ solves the equation*

$$\mathcal{K} \times_1 M + \ldots + \mathcal{K} \times_k M + \mathcal{C}_k = 0 \tag{3.2.5}$$

where $\mathcal{C}_k = \text{cum}_k(Z_1)$ is the k-th-order cumulant tensor of Z_1. Equation (3.2.5) will be denoted the k-th-order continuous Lyapunov equation.

The k-th-order continuous Lyapunov equation is linear in both the parameters $(M, \mathcal{C}_k)$ and the k-th-order cumulant $\mathcal{K}$. Furthermore, given a stable M and any k-th-order tensor $\mathcal{C}_k$, there exists a unique k-th-order tensor $\mathcal{K}$ solving (3.2.5). This follows from Corollary 3.1 by Xu and Wang [2022]. Appendix 3.A.1 includes the proof of Proposition 3.2.2 and additional details related to equation (3.2.5), such as the statement of Corollary 3.1 by Xu and Wang [2022] and an alternative proof of existence and uniqueness of the solution to equation (3.2.5) that also applies to non-stable M.

For our identification results, we will focus on the combination of the covariance equation and one additional r-th-order continuous Lyapunov equation. For estimation it is in practice most relevant to consider $r = 3$ or $r = 4$, or a combination to possibly achieve a more efficient estimator.

In order to show our identification results, we consider a subset of distributions in $\mathcal{P}_G$ where the r-th-order cumulants exist, and where the second- and r-th-order cumulants of the Lévy process are diagonal with non-zero entries on the diagonal entries.

Definition 3.2.3. *Let $G = ([d], E)$ be a directed graph on d nodes and let $r \geq 3$ be an integer. Denote by $\mathcal{P}_G^{2,r} \subseteq \mathcal{P}_G$ the subset of the continuous Lyapunov model for which the Lévy process has finite r-th-order moment and with cumulant tensors $\mathcal{C}_2$ and $\mathcal{C}_r$ being diagonal with non-zero diagonal entries.*

We let $\mathcal{Q}^{2,r}$ denote the subset of $\mathcal{Q}$ consisting of distributions with finite r-th-order moment and with the cumulant tensors $\mathcal{C}_2$ and $\mathcal{C}_r$ being diagonal with non-zero diagonal entries. Then

$$\mathcal{P}_G^{2,r} = \psi(\mathbb{R}^E_{\text{stab}} \times \mathcal{Q}^{2,r}),$$

that is, the submodel $\mathcal{P}_G^{2,r}$ is parametrized by $\mathbb{R}^E_{\text{stab}} \times \mathcal{Q}^{2,r}$ via the parametrization ψ. If the Lévy process has *independent* coordinates, that is, if Q_1 is a product measure, any cumulant tensor is diagonal, provided it exists. Since we only need that $\mathcal{C}_2$ and $\mathcal{C}_r$ are diagonal, we avoid making the stronger independence assumption in the definition of $\mathcal{P}_G^{2,r}$.

We also assume that the diagonal entries of $\mathcal{C}_2$ and $\mathcal{C}_r$ are non-zero. For $\mathcal{C}_2$ this assumption is a non-degeneracy condition, while for $\mathcal{C}_r$ this is the specific non-Gaussianity assumption we need to establish identifiability.

In addition to the abstract parameter and model spaces we will also introduce the finite-dimensional spaces of cumulants resulting from the model spaces. Let $\text{Sym}^k(\mathbb{R}^d)$ denote the set of k-th-order $d \times d \times \ldots \times d$ symmetric tensors, and let $\text{Sym}^k(\mathbb{R}^d)^*$ denote the subset of tensors whose diagonal entries are non-zero. For $k = 2$ we let $\text{PD}_d \subseteq \text{Sym}^2(\mathbb{R}^d)^*$ denote the subset of symmetric $d \times d$-matrices that are also positive definite.

Definition 3.2.4. *Let $G = ([d], E)$ be a directed graph with d nodes and let $r \geq 3$ be an integer. The second- and r-th-order Lyapunov cumulant model is the set of second- and r-th-order tensors in the set*

$$\mathcal{M}_G^{2,r} = \left\{ (\Sigma, \mathcal{K}) \in (\text{PD}_p, \text{Sym}^r(\mathbb{R}^d)) \;\middle|\; \begin{array}{l} M\Sigma + \Sigma M^T + \mathcal{C}_2 = 0, \\ \mathcal{K} \times_1 M + \cdots + \mathcal{K} \times_r M + \mathcal{C}_r = 0, \\ M \in \mathbb{R}^E_{\text{stab}}, \text{ and } \mathcal{C}_2 \in \text{PD}_d, \mathcal{C}_r \in \text{Sym}^r(\mathbb{R}^d)^* \text{ diagonal} \end{array} \right\}.$$

Let $\text{cum}_k(P) \in \text{Sym}^k(\mathbb{R}^d)$ denote the k-th-order cumulant for a probability distribution P with finite k-th-order moment. Then for $(M, Q_1) \in \mathbb{R}^E_{\text{stab}} \times \mathcal{Q}$ with Q_1 having finite k-th-order moment,

$$\text{cum}_k(\psi(M, Q_1)) = \mathcal{K}$$

with $\mathcal{K}$ solving (3.2.5) for $\mathcal{C}_k = \text{cum}_k(Q_1)$. Letting additionally $\text{cum}_{2,r}(P) = (\text{cum}_2(P), \text{cum}_r(P))$, we see that

$$\text{cum}_{2,r} : \mathcal{P}_G^{2,r} \to \mathcal{M}_G^{2,r}$$

maps the distribution P in the model $\mathcal{P}_G^{2,r}$ to its corresponding pair $(\Sigma, \mathcal{K})$ of second- and r-th-order cumulants in $\mathcal{M}_G^{2,r}$. See the commutative diagram in Figure 3.3.4 for an overview of these maps and sets. While $\text{cum}_{2,r}$ may not be surjective for all r, we show that it is for r odd, and for r even its image has full dimension, see Proposition 3.B.13 in Appendix 3.B.4 in combination with the commutative diagram in Figure 3.3.4.

3.3 Identifiability

3.3.1 Overview

Usually, the parameter M would be said to be identified from the probability measure $P \in \mathcal{P}_G$ if it were uniquely determined by P. Section 3.3.2 shows that this is not possible and that we can only hope to identify M up to scaling in general. We give necessary conditions, including that G is connected, for identification up to a global scaling factor.

Our main Theorem 3.3.1 shows that the conditions are generically essentially sufficient as well and describes the precise injectivity properties of the parameterization of the cumulants via the Lyapunov equations. The conclusion is rephrased in Corollary 3.3.2 as $(M, \mathcal{C}_2, \mathcal{C}_r)$ being generically identifiably, up to a joint scaling of these parameters, from a combination of the covariance matrix and one r-th-order cumulant tensor.

The proof of Theorem 3.3.1 exploits the linearity of the continuous Lyapunov equations, which allows us to transfer the question of identifiability to rank questions for certain matrices. We describe the necessary reorganization of the Lyapunov equations in full detail in Section 3.3.4. Section 3.4 on estimation also relies heavily on the reorganization. The non-trivial, but technical, arguments about rank properties are given in Appendix 3.A.2. When G is not connected, Section 3.3.5 shows that identification up to a global scaling factor fails, and that identification is only possible up to a scaling factor for each connectivity component.

3.3.2 Limitations to identifiability

For the parametrized model $\mathcal{P}_G$, the standard notion of identifiability is simply the question of whether the map $\psi : \mathbb{R}^E_{\text{stab}} \times \mathcal{Q} \to \mathcal{P}_G$ is injective. To see that ψ is not injective, we write (3.1.1) as a stochastic integral equation and do a change-of-variable in the integration to get

$$X_{ct} = X_0 + \int_0^t (cM) X_{cs} \mathrm{d}s + Z_{ct}$$

for any $c > 0$. This shows that $\psi(M, Q_1) = \psi(cM, Q_c)$ for all $c > 0$, where Q_c is the distribution of Z_c, and ψ is thus not injective. The best possible identification result would therefore be that $\psi^{-1}(\psi(M, Q_1)) = \{(cM, Q_c) \mid c > 0\}$, that is, that the parameter (M, Q_1) is identifiable up to a common scaling factor.

From the Lévy-Khintchine representation of Q_c it likewise follows that $\operatorname{cum}_k(Z_c) = c\mathcal{C}_k$ for any k. We see that the lack of identifiability is also reflected in the cumulant equations (3.2.5): if $(M, \mathcal{C}_k)$ is a solution with M stable for given $\mathcal{K}$ then $(cM, c\mathcal{C}_k)$ is a solution with cM stable for all $c > 0$. This particular lack of identifiability is intuitively reasonable, since we cannot identify the absolute speed of the stationary solution of (3.1.1) from observations at a single time point.

Since M is our parameter of interest, we are primarily interested in whether there exists a map

$$P = \psi(M, Q_1) \mapsto \{cM \mid c > 0\} \tag{3.3.6}$$

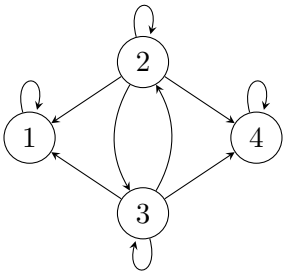


Figure 3.3.2: Directed graph, G, such that the drift matrix, M, is *not* identifiable up to scaling from the Gaussian model $\psi(M, \mathcal{N}(0, \mathcal{C}_2))$ for any diagonal $\mathcal{C}_2$ [Dettling et al., 2023].

from $\mathcal{P}_G$ to the equivalence classes in $\mathbb{R}^E_{\text{stab}}$ of matrices identical up to scaling. Such a map cannot exists in general either, because if G has at least two connectivity components, we can rescale the parameters for each component independently, see also Corollary 3.3.5. Additionally, even if G is connected, but $Q_1 = \mathcal{N}(0, \mathcal{C}_2)$ is a Gaussian distribution with diagonal covariance matrix $\mathcal{C}_2$, then M cannot always be identified up to a scaling from $P = \psi(M, Q_1)$. Indeed, Dettling et al. [2023] showed this by their Example 8.7 for the graph in Figure 3.3.2 for any choice of diagonal $\mathcal{C}_2$.

Our goal is to find a submodel of $\mathcal{P}_G$ such that we can establish the existence of the map (3.3.6) defined on this submodel, provided that G is connected. A candidate for such a submodel is $\mathcal{P}_G^{2,r}$ from Definition 3.2.3, which rules out all Gaussian distributions. It turns out that we need to further remove an exception set from $\mathcal{P}_G^{2,r}$, but we can argue that this set is, in a certain sense, negligible. The precise way to phrase this is via the finite-dimensional parametrization of the second- and r-th-order Lyapunov cumulant model $\mathcal{M}_G^{2,r}$ in terms of $(M, \mathcal{C}_2, \mathcal{C}_r)$ of which the negligible set will be an algebraic subset.

3.3.3 Main results

To simplify notation, we identify *diagonal* $\mathcal{C}_2$ and $\mathcal{C}_r$ tensors with their diagonals as elements in $\mathbb{R}^d_+$ and $(\mathbb{R} \setminus \{0\})^d$, respectively, and we introduce the parameter set

$$\Theta_G = \mathbb{R}^E_{\text{stab}} \times \mathbb{R}^d_+ \times (\mathbb{R} \setminus \{0\})^d.$$

The set Θ_G is an open set in standard Euclidean topology. Elements in Θ_G are denoted $\theta = (M, \mathcal{C}_2, \mathcal{C}_r)$. The question of identifiability is then settled in terms of the rational parametrization

$$\begin{aligned} \phi_{G,r} : \Theta_G &\to \mathcal{M}_G^{2,r} \\ (M, \mathcal{C}_2, \mathcal{C}_r) &\mapsto (\Sigma, \mathcal{K}), \end{aligned}$$

where Σ and $\mathcal{K}$ are the unique solutions to the corresponding second- and r-th-order Lyapunov equation, as guaranteed to exist by Proposition 3.A.2. By Definition 3.2.4, $\phi_{G,r}$ is surjective.

We can now state our main identifiability result pertaining to the injectivity properties of $\phi_{G,r}$.

Theorem 3.3.1. *Let $d \geq 2$, $r \geq 3$ be integers and let $G = ([d], E)$ be any connected graph with all self-loops present. Then there exists a proper algebraic subset $\mathcal{N}_G \subset \Theta_G$ such that for $\theta \in \Theta_G \setminus \mathcal{N}_G$,*

$$\phi_{G,r}^{-1}(\phi_{G,r}(\theta)) = \{c\theta \mid c > 0\}. \tag{3.3.7}$$

There are different notions in the literature of weak forms of identifiability. See Chapter 16 by Sullivant [2018] for an overview. A result such as Theorem 3.3.1 can be phrased as θ being *generically* identifiable up to scaling. The term *generic* means that the possible exception set, where θ cannot be identified, is negligible in some sense. In Theorem 3.3.1 the negligible set $\mathcal{N}_G$ is a proper algebraic subset, which means that it is a proper subset and the vanishing set of a collection of polynomials, i.e., a variety. This makes $\mathcal{N}_G$ a Lebesgue null set in the ambient Euclidean space, but more importantly, since Θ_G is open, $\mathcal{N}_G$ is a strictly lower-dimensional set, it is closed and $\Theta_G \backslash \mathcal{N}_G$ is open and dense in Θ_G. For these reasons, (3.3.7) is regarded as holding for most points in Θ_G, and it is reasonable to regard the exception set $\mathcal{N}_G$ as negligible. Using the conclusion of Theorem 3.3.1 as the definition of what we mean by generic identifiability, we get the following corollary.

Corollary 3.3.2. *Let $d \geq 2$, $r \geq 3$ be integers and let $G = ([d], E)$ be any connected graph with all self-loops present. Then the drift matrix M and the diagonal and non-zero cumulants $\mathcal{C}_2$ and $\mathcal{C}_r$ are jointly generically identifiable from the cumulant tensors $(\Sigma, \mathcal{K}) \in \mathcal{M}_G^{2,r}$ up to a common scaling.*

Remark 3..3. Theorem 3.3.1 and Corollary 3.3.2 make two assumptions about the graph; that it is connected and that it contains all self-loops. Section 3.3.5 argues that the connectedness assumption is necessary. If the graph is acyclic, the drift matrix can also only be stable if it contains all self-loops, and our proof relies on the existence of an acyclic subgraph corresponding to a stable M. The assumption that the graph contains all self-loops is harmless from a practical perspective as it is almost always sensible to assume that a variable affects itself. However, it is of theoretical interest to determine if the assumption can be relaxed. We have by computation found examples of cyclic graphs for which Theorem 3.3.1 remains true for $r = 3$ even though some self-loops are removed, see Figure 3.3.3 for two examples on three nodes.

Theorem 3.3.1 and Corollary 3.3.2 do not explicitly answer to what extent the map (3.3.6) exists. To do so we introduce the map

$$\begin{aligned} \pi_{2,r} :& \mathbb{R}^E_{\text{stab}} \times \mathcal{Q}^{2,r} \to \Theta_G \\ & (M, Q_1) \mapsto (M, \text{cum}_2(Q_1), \text{cum}_r(Q_1)), \end{aligned}$$

which makes the diagram in Figure 3.3.4 commute. Under the assumptions of Theorem 3.3.1, the map (3.3.6) can then be defined on $\psi(\mathbb{R}^E_{\text{stab}} \times \mathcal{Q}^{2,r} \backslash \pi_{2,r}^{-1}(\mathcal{N}_G))$ by composing

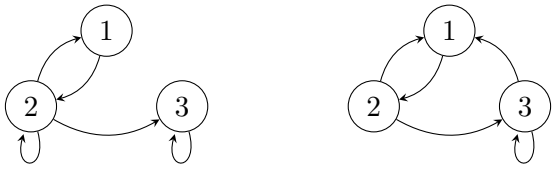


Figure 3.3.3: Two directed graphs on three nodes with one and two self-loops missing, respectively, for which M is still identifiable in the sense of Theorem 3.3.1 for $r = 3$.

$$\begin{array}{ccc} \mathbb{R}^E_{\text{stab}} \times \mathcal{Q}^{2,r} & \xrightarrow{\psi} & \mathcal{P}^{2,r}_G \\ \downarrow{\scriptstyle \pi_{2,r}} & & \downarrow{\scriptstyle \text{cum}_{2,r}} \\ \Theta_G & \xrightarrow{\phi_{G,r}} & \mathcal{M}^{2,r}_G \end{array} \qquad \begin{array}{ccc} (M, Q_1) & \xrightarrow{\psi} & P \\ \downarrow{\scriptstyle \pi_{2,r}} & & \downarrow{\scriptstyle \text{cum}_{2,r}} \\ (M, \mathcal{C}_2, \mathcal{C}_r) & \xrightarrow{\phi_{G,r}} & (\Sigma, \mathcal{K}) \end{array}$$

Figure 3.3.4: Relations between model spaces and parametrizations as a commutative diagram.

$\phi^{-1}_{G,r} \circ \text{cum}_{2,r}$ with a projection onto the first coordinate. The extent to which $\pi^{-1}_{2,r}(\mathcal{N}_G)$ is negligible in $\mathbb{R}^E_{\text{stab}} \times \mathcal{Q}^{2,r}$ is not entirely clear. We do, however, show Proposition 3.B.13 in Appendix 3.B.4 stating that $\pi_{2,r}$ is surjective if r is odd, and the image of $\pi_{2,r}$ is, at least, of full dimension if r is even. Importantly, the image of $\pi_{2,r}$ in Θ_G is therefore not a lower dimensional set like $\mathcal{N}_G$, which justifies that generic identifiability in the sense of Theorem 3.3.1 and Corollary 3.3.2 is of relevance.

3.3.4 Outline of the proof of Theorem 3.3.1 via reorganization of the Lyapunov equations

In this section we reorganize the second- and r-th-order cumulant equations from Proposition 3.2.2 so that they appear as a homogeneous system of linear equations in the vectorization of the parameters M, $\mathcal{C}_2$ and $\mathcal{C}_r$. For the vectorization we do not assume that $\mathcal{C}_2$ and $\mathcal{C}_r$ are diagonal. Since the (non-zero) parameter vector is in the kernel of the matrix that defines the linear equations, the conclusion of Theorem 3.3.1, equation (3.3.7), is true whenever this kernel is one-dimensional. This reduces our question about identifiability to a question about the rank of a matrix. To illustrate the idea behind the general proof, Example 3.3.4 gives the proof in the case of the complete graph on two nodes and with $r = 3$.

In order to vectorize the cumulant equations we use a so called symmetric, or unique, vectorization operator. That is, it is a vectorization of only the unique entries of the symmetric covariance matrix and the symmetric r-th-order cumulant tensor. We denote the unique vectorization operator by vec_u and the ordinary vectorization operator by vec.

Since the k-th order Lyapunov equation is linear in M, its unique vectorization can

be written as

$$A_k(\mathcal{K})\text{vec}(M) + \text{vec}_u(\mathcal{C}_k) = 0 \tag{3.3.8}$$

for some matrix $A_k(\mathcal{K})$ depending on the k-th-order cumulant $\mathcal{K}$. To understand how $A_k(\mathcal{K})$ is organized, we first consider the case $k = 2$, where the equation is

$$A_2(\Sigma)\text{vec}(M) + \text{vec}_u(\mathcal{C}_2) = 0 \tag{3.3.9}$$

with $A_2(\Sigma)$ a $(d(d+1)/2) \times d^2$ matrix. To write down the entries of this matrix, we index the second-order Lyapunov equation and the unique vectorization of $\mathcal{C}_2$ by $\{(i_1 i_2) \mid i_1 \leq i_2\}$. We index the columns of A_2 by the d^2 *potential* edges $\alpha \to \beta$ of a graph, corresponding to all possible entries of M. The expressions for the entries are then

$$A_2(\Sigma)_{(i_1 i_2),(\alpha\to\beta)} = \begin{cases} 0 & \text{if } \beta \neq i_1, i_2 \\ 2\Sigma_{i_1 i_2} & \text{if } \beta = i_1 = i_2 \\ \Sigma_{i_2\alpha} & \text{if } \beta = i_1 \neq i_2 \\ \Sigma_{i_1\alpha} & \text{if } \beta = i_2 \neq i_1, \end{cases} \tag{3.3.10}$$

see also equation (4.3) by Dettling et al. [2023].

Similarly, the k-th-order Lyapunov equation and unique vectorization of $\mathcal{C}_k$ are indexed by $\{(i_1 \dots i_k) \mid i_1 \leq \cdots \leq i_k\}$. The matrix $A_k(\mathcal{K})$ has dimensions

$$\binom{d + (k-1)}{k} \times d^2.$$

In general, the entry $A_k(\mathcal{K})_{(\beta i_2 \dots i_k),(\alpha\to\beta)}$ is an integer multiple of $\mathcal{K}_{\alpha i_2 \dots i_k}$, and likewise for β equal to one of the other indices. The remaining entries are zero. To determine the integer, one needs to count how many of the mode products in the equation will contribute with the same term, which will depend on how many of the i-indices are equal to β. Thus, again written in the case $\beta = i_1$, the general expression is given by

$$A_k(\mathcal{K})_{(i_1 \dots i_k),(\alpha\to\beta)} = N((i_1 \dots i_k), \beta) \cdot \mathcal{K}_{\alpha i_2 \dots i_k}, \tag{3.3.11}$$

where $N((i_1 \dots i_k), \beta)$ is the number of row indices $i_1 \leq \dots \leq i_k$ equal to β.

Writing this out, we obtain a complete description of $A_3(\mathcal{K})$ by

$$A_3(\mathcal{K})_{(i_1 i_2 i_3),(\alpha\to\beta)} = \begin{cases} 0 & \beta \neq i_1, i_2, i_3 \\ 3\mathcal{K}_{i_1 i_2 i_3} & \beta = i_1 = i_2 = i_3 \\ 2\mathcal{K}_{i_1 \alpha i_3} & \beta = i_1 = i_2 \neq i_3 \\ 2\mathcal{K}_{\alpha i_2 i_3} & \beta = i_2 = i_3 \neq i_1 \\ \mathcal{K}_{i_1 i_2 \alpha} & \beta = i_3, i_2 \neq i_3 \\ \mathcal{K}_{\alpha i_2 i_3} & \beta = i_1, i_1 \neq i_2 \\ \mathcal{K}_{i_1 \alpha i_3} & \beta = i_2, i_2 \neq i_3, i_2 \neq i_1. \end{cases} \tag{3.3.12}$$

Again, letting $r \geq 3$ be a fixed integer, we can combine the second-order vectorized equation, (3.3.9) and (3.3.8) for $k = r$ to obtain the following linear system, still allowing $\mathcal{C}_2$ and $\mathcal{C}_r$ to be non-diagonal,

$$\begin{pmatrix} A_2(\Sigma) & I_{d(d+1)/2} & 0 \\ A_r(\mathcal{K}) & 0 & I_{\binom{d+(r-1)}{r}} \end{pmatrix} \begin{pmatrix} \operatorname{vec}(M) \\ \operatorname{vec}_u(\mathcal{C}_2) \\ \operatorname{vec}_u(\mathcal{C}_r) \end{pmatrix} = 0. \tag{3.3.13}$$

The question of identifiability is now reduced to considering the rank of the matrix above. Without any restrictions on the cumulants $\mathcal{C}_2$ and $\mathcal{C}_r$ there are as many unique entries in these as there are unique equations. Therefore, it is impossible to establish identifiability if we allow M, $\mathcal{C}_2$ and $\mathcal{C}_r$ to be arbitrary. This would not change by adding additional equations for other higher-order cumulants, since we would always be adding the same number of equations as unknown parameters. Therefore, certain constraints are necessary on the entries of M, $\mathcal{C}_2$ or $\mathcal{C}_r$.

For $\mathcal{C}_2$ and $\mathcal{C}_r$ diagonal, (3.3.13) reduces to

$$\begin{pmatrix} A_2(\Sigma)_{\text{off}} & 0 & 0 \\ A_r(\mathcal{K})_{\text{off}} & 0 & 0 \\ A_2(\Sigma)_{\text{diag}} & I_d & 0 \\ A_r(\mathcal{K})_{\text{diag}} & 0 & I_d \end{pmatrix} \begin{pmatrix} \operatorname{vec}(M) \\ \operatorname{diag}(\mathcal{C}_2) \\ \operatorname{diag}(\mathcal{C}_r) \end{pmatrix} = 0, \tag{3.3.14}$$

where $A_2(\Sigma)_{\text{off}}$ and $A_r(\mathcal{K})_{\text{off}}$ denote all the rows indexed by the off-diagonal tensor entries of $A_2(\Sigma)$ and $A_r(\mathcal{K})$, respectively. That is, all the rows not indexed by (ii) or $(i \dots i)$. Similarly, $A_2(\Sigma)_{\text{diag}}$ and $A_r(\mathcal{K})_{\text{diag}}$ denote the rows indexed by the diagonal entries. In the proof of Theorem 3.3.1 we establish that the upper left block of the matrix in (3.3.14),

$$A_{\text{off}} = \begin{pmatrix} A_2(\Sigma)_{\text{off}} \\ A_r(\mathcal{K})_{\text{off}} \end{pmatrix}, \tag{3.3.15}$$

generically has rank $d^2 - 1$. This shows the generic identifiability of M up to scaling, and from the remaining equations in (3.3.14), the diagonal cumulants $\mathcal{C}_2$ and $\mathcal{C}_r$ are also identified up to scaling.

To prove that A_{off} generically has rank $d^2 - 1$ it is enough to exhibit one choice of $\theta = (M, \mathcal{C}_2, \mathcal{C}_3) \in \Theta_G$ such that A_{off} has rank $d^2 - 1$. This is because the entries of the cumulants, Σ and $\mathcal{K}$, can be written as rational functions in the parameters M, $\mathcal{C}_2$ and $\mathcal{C}_r$, since the continuous Lyapunov equations are also linear in the respective cumulants, see Proposition 3.A.1 in Appendix 3.A.1. Alternatively, this can also be seen using the general trek rule in Appendix 3.B.2. Thus, any determinant of a submatrix of A will also be a rational function in M, $\mathcal{C}_2$ and $\mathcal{C}_r$. Therefore, the determinant of a submatrix of A being non-zero corresponds to a polynomial in M, $\mathcal{C}_2$ and $\mathcal{C}_r$ being non-zero, and to show that a polynomial is not the zero polynomial it is enough to show that it is non-zero in a single point.

Example 3.3.4. To illustrate by example how we prove Theorem 3.3.1, we consider the smallest non-simple graph, the complete graph on $d = 2$ nodes, see Figure 3.3.5, and

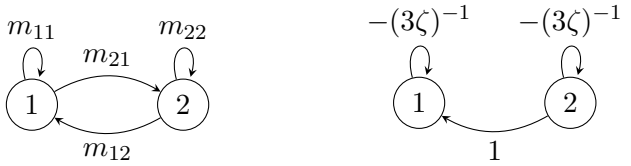


Figure 3.3.5: The complete graph on two nodes and four edges corresponding to the parameters in M (left). Example 3.3.4 shows identifiability of M up to scaling by computing the rank of A_{off} for a particular choice of parameters (right).

prove that (3.3.15) for $r = 3$ generically has rank $d^2 - 1 = 3$. For diagonal $\mathcal{C}_2$ and $\mathcal{C}_3$, the equation system in equation (3.3.13) can be written as

$$\begin{array}{c} \\ (11) \\ (12) \\ (22) \\ (111) \\ (112) \\ (122) \\ (222) \end{array} \begin{array}{c} \begin{array}{cccc} 1\to 1 & 1\to 2 & 2\to 1 & 2\to 2 \end{array} \\ \left(\begin{array}{cccc|cccc} 2\Sigma_{11} & 0 & 2\Sigma_{12} & 0 & 1 & 0 & 0 & 0 \\ \Sigma_{12} & \Sigma_{11} & \Sigma_{22} & \Sigma_{12} & 0 & 0 & 0 & 0 \\ 0 & 2\Sigma_{12} & 0 & 2\Sigma_{22} & 0 & 1 & 0 & 0 \\ 3\mathcal{K}_{111} & 0 & 3\mathcal{K}_{112} & 0 & 0 & 0 & 1 & 0 \\ 2\mathcal{K}_{112} & \mathcal{K}_{111} & 2\mathcal{K}_{122} & \mathcal{K}_{112} & 0 & 0 & 0 & 0 \\ \mathcal{K}_{122} & 2\mathcal{K}_{112} & \mathcal{K}_{222} & 2\mathcal{K}_{122} & 0 & 0 & 0 & 0 \\ 0 & 3\mathcal{K}_{122} & 0 & 3\mathcal{K}_{222} & 0 & 0 & 0 & 1 \end{array}\right) \end{array} \begin{pmatrix} M_{11} \\ M_{21} \\ M_{12} \\ M_{22} \\ (\mathcal{C}_2)_{11} \\ (\mathcal{C}_2)_{22} \\ (\mathcal{C}_3)_{111} \\ (\mathcal{C}_3)_{222} \end{pmatrix} = 0.$$

Because of the sparsity structure exhibited by the $\mathcal{C}$ columns, the matrix has full rank minus one if the submatrix with row indices corresponding to three off-diagonal entries of the tensors has rank three. These are rows two, five and six. Removing the four $\mathcal{C}$ columns as well, we arrive at the 3×4 matrix

$$\begin{array}{c} \\ (12) \\ (112) \\ (122) \end{array} \begin{array}{c} \begin{array}{cccc} 1\to 1 & 1\to 2 & 2\to 1 & 2\to 2 \end{array} \\ \begin{pmatrix} \Sigma_{12} & \Sigma_{11} & \Sigma_{22} & \Sigma_{12} \\ 2\mathcal{K}_{112} & \mathcal{K}_{111} & 2\mathcal{K}_{122} & \mathcal{K}_{112} \\ \mathcal{K}_{122} & 2\mathcal{K}_{112} & \mathcal{K}_{222} & 2\mathcal{K}_{122} \end{pmatrix} \end{array} = A_{\text{off}},$$

which we need to show has rank 3 generically.

As argued right before the example, to prove this generic rank property we only need to exhibit one choice of parameters where the matrix has rank 3. We use the specialized trek rule in Corollary 3.A.4 and let

$$M = \begin{pmatrix} -1/(3\zeta) & 1 \\ 0 & -1/(3\zeta) \end{pmatrix},$$

see Figure 3.3.5. We take $\mathcal{C}_2$ and $\mathcal{C}_3$ to be the identity matrix and tensor, respectively. By Corollary 3.A.4 we obtain the following expression for the entries of Σ and $\mathcal{K}$

$$\Sigma_{11} = \frac{3}{2}\zeta \qquad\qquad \mathcal{K}_{111} = \zeta \qquad\qquad \mathcal{K}_{122} = 2\zeta^3$$
$$\Sigma_{12} = \left(\frac{3}{2}\right)^2 \zeta^2 \qquad \Sigma_{22} = \frac{3}{2}\zeta + 2\left(\frac{3}{2}\right)^3 \zeta^3 \qquad \mathcal{K}_{112} = \zeta^2 \qquad \mathcal{K}_{222} = \zeta + 6\zeta^4.$$

Plugging these into the matrix A_{off}, its determinant becomes $3/2\zeta^7 + 3/4\zeta^4$. This is not the zero polynomial, so generically the determinant does not vanish. Thus, the parameters are generically identifiable up to a joint scaling.

In this small example, the argument would work by picking just one value of ζ, for example, $\zeta = 1$. However, in the full proof the number of terms in the general ζ-polynomial grows rapidly, and the signs of some of the terms will be opposite. For that reason, the strategy for proving that the determinant is non-zero in general is to prove that the coefficient of the lowest degree term in ζ is non-zero.

3.3.5 Identification when G is not connected

Theorem 3.3.1 gives rise to two natural questions: is the conclusion true if the graph is not connected; and is the exception set $\mathcal{N}_G$ actually non-empty? We show the following corollary that answers both questions with the proof provided in Appendix 3.A.2.

Corollary 3.3.5. *Let $d \geq 2$, $r \geq 3$ be integers and let $G_0 = ([d], E_0)$ be a graph with m connected components and all self-loops present. Then for $(\Sigma, \mathcal{K}) \in \mathcal{M}_{G_0}^{2,r}$,*

$$A_{\text{off}} = \begin{pmatrix} A_2(\Sigma)_{\text{off}} \\ A_r(\mathcal{K})_{\text{off}} \end{pmatrix} \tag{3.3.16}$$

has rank at most $d^2 - m$, and generically in Θ_{G_0} it has rank $d^2 - m$.

If a graph G_0 has $m \geq 2$ connected components, we can, of course, apply Theorem 3.3.1 for each connected component separately, and conclude that the parameters are generically identifiable up to component-dependent scaling factors. Corollary 3.3.5 shows that this is the best we can hope for, since the rank deficiency of A_{off} is at least the number of connectivity components m. This result is intuitively reasonable since each connectivity component corresponds to a subsystem of the d-dimensional stochastic process, and these subsystems do not interact. Therefore, the processes can evolve at different speeds independently of each other, and we cannot determine how fast any of them move around from cross-sectional observations alone.

It also follows from Corollary 3.3.5 that if G is connected then $\mathcal{N}_G \neq \emptyset$. To see this, suppose that

$$\phi_{G,r}(\theta) = (\Sigma, \mathcal{K}) \in \mathcal{M}_{G_0}^{2,r} \subseteq \mathcal{M}_G^{2,r}$$

for G_0 a subgraph of G having at least two connectivity components. Then the corresponding A_{off} matrix has rank at most $d^2 - 2$ by Corollary 3.3.5. This implies that $\phi_{G,r}^{-1}(\phi_{G,r}(\theta))$ is at least two-dimensional, and we conclude that $\phi_{G,r}^{-1}(\mathcal{M}_{G_0}^{2,r}) \subseteq \mathcal{N}_G$. Since $\phi_{G,r}^{-1}$ is surjective, $\mathcal{N}_G$ is non-empty. This shows that θ can never be globally identified up to scaling. It remains an open question whether $\mathcal{N}_G$ contains parameters not in $\phi_{G,r}^{-1}(\mathcal{M}_{G_0}^{2,r})$ for some disconnected subgraph G_0. That is, whether there exist exceptional choices of parameters where identifiability up to scaling fails without disconnecting the graph.

3.4 Estimation

Our identification results are based on the linear equation (3.3.14). In this section we turn such an identification equation into an estimator of the drift parameter M. To this end, suppose A is any $b \times d^2$ matrix, depending linearly on the cumulants, such that

$$\ker(A) = \operatorname{span}(\operatorname{vec}(M)). \tag{3.4.17}$$

By Theorem 3.3.1 and its proof we know that (3.4.17) holds for $\theta = (M, \mathcal{C}_2, \mathcal{C}_r) \in \Theta_G \backslash \mathcal{N}_G$ when

$$A = \begin{pmatrix} A_2(\Sigma)_{\text{off}} \\ A_r(\mathcal{K})_{\text{off}} \end{pmatrix} \tag{3.4.18}$$

for $(\Sigma, \mathcal{K}) = \phi_{G,r}(\theta)$. This particular A-matrix has

$$b = \frac{d(d-1)}{2} + \binom{d + (r-1)}{r} - d \geq d^2 - 1$$

rows.

For a given connected graph G, it follows from the proof of Theorem 3.3.1 that (3.4.17) holds if we choose A as a certain submatrix of (3.4.18) containing only $d^2 - 1$ of these rows. On the other hand, we can also include additional rows from $A_k(\mathcal{K})_{\text{off}}$ for $k \notin \{2, r\}$ without violating (3.4.17). The choice of which rows to include involves a tradeoff between computational complexity and statistical efficiency. For estimation purposes with a known G we would usually also drop all columns of A that correspond to entries of M that are zero according to G. This would reduce A to a $b \times |E|$ matrix. To keep the notation simple, we will here only treat estimation of the entire matrix M corresponding to the complete graph, with obvious modifications for any choice of subgraph. We will, however, allow for any choice of A satisfying (3.4.17).

We will from hereon assume that M is normalized, that is, $\|M\|_F = 1$ where $\|\cdot\|_F$ denotes the Frobenius norm. For any specific choice of A satisfying (3.4.17) we define an estimator of M, which we show is consistent and asymptotically normal as the sample size tends to infinity. In practice we may choose A given by (3.4.18) with $r = 3$ or $r = 4$, or possibly a combination. Including more rows from higher order cumulants is computationally more costly but could potentially make the estimator more efficient. We report on this tradeoff in Section 3.5.

Definition 3.4.1. *Suppose $\hat{A}$ is an estimator of the $b \times d^2$ matrix A with a unique smallest singular value. Then the least singular value estimator of M is defined by*

$$\operatorname{vec}(\hat{M}) = v_{\min}(\hat{A}), \tag{3.4.19}$$

where $v_{\min}(\hat{A})$ denotes the right singular vector for $\hat{A}$ corresponding to its smallest singular value and such that $\operatorname{tr}(\hat{M}) < 0$.

It is natural to think of $\operatorname{span}(v_{\min}(\hat{A}))$ as an estimator of $\ker(A)$, see Figure 3.4.6. Since the right singular vector $v_{\min}(\hat{A})$ is only determined up to a sign, we fix its sign

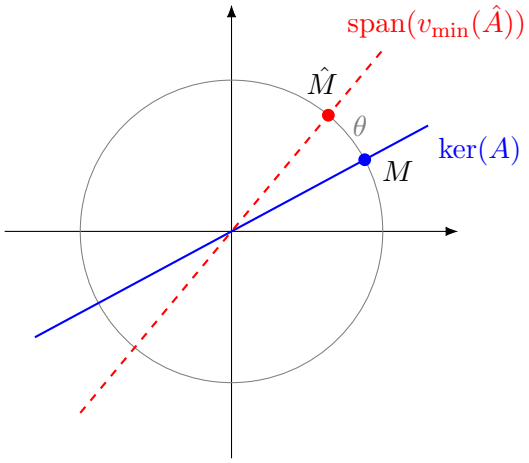


Figure 3.4.6: With $v_{\min}(\hat{A})$ a right singular vector of $\hat{A}$ corresponding to its smallest singular value, the estimator $\hat{M}$ is a unit vector in span($v_{\min}(\hat{A})$).

in Definition 3.4.1 by requiring that $\mathrm{tr}(\hat{M}) < 0$. This is a sensible convention since M is stable with $\mathrm{tr}(M) < 0$, but it does not guarantee that $\hat{M}$ is stable.

We may also observe that

$$\mathrm{vec}(\hat{M}) = \underset{v:\|v\|=1}{\arg\min} \|\hat{A}v\|^2 \tag{3.4.20}$$

with the same convention as in Definition 3.4.1 for fixing the sign of the minimizer. Thus, $v_{\min}(\hat{A})$ is also an eigenvector of $\hat{A}^T\hat{A}$ corresponding to its smallest eigenvalue. Since the eigenvalues of a matrix are roots of the characteristic polynomial, the estimator is well-defined generically since the matrix $\hat{A}^T\hat{A}$ will generically have a unique smallest eigenvalue. However, least singular value estimation can result in a non-stable $\hat{M}$ since stability is not imposed as a constraint in the minimization problem (3.4.20).

All the estimators of A that we consider are linear maps of the form

$$\mathrm{vec}(\hat{A}) = \mathcal{A}\hat{\kappa} \tag{3.4.21}$$

where $\mathcal{A}$ is a fixed matrix with $b \cdot d^2$ rows, and $\hat{\kappa}$ denotes a unique vectorization of the empirical cumulants entering into A. The matrix $\mathcal{A}$ has integer entries and is generally extremely sparse.

Whenever the M-selfdecomposable distribution has finite $2k$-th-order moment,

$$\sqrt{n}(\hat{\kappa}_n - \kappa) \xrightarrow{D} \mathcal{N}(0, \Omega) \tag{3.4.22}$$

for some covariance matrix Ω by the central limit theorem. Therefore, we obtain the following theorem giving the asymptotics of the estimator proven in Appendix 3.B.1.

Theorem 3.4.2. *Let $X_1, \ldots, X_n$ be an i.i.d. sample from a distribution in $\mathcal{P}_G$ with finite k-th-order moment, and suppose that A, whose entries are cumulants of order at most*

k, satisfies (3.4.17). Let $\hat{A}_n$ denote the plug-in estimator of A based on the empirical cumulants and let $\hat{M}_n$ be the least singular value estimator given by (3.4.19) in terms of $\hat{A}_n$. Then

$$\hat{A}_n \xrightarrow{P} A, \qquad \hat{M}_n \xrightarrow{P} M \tag{3.4.23}$$

for $n \to \infty$. If, additionally, the distribution has finite $2k$-th-order moment,

$$\sqrt{n}(\mathrm{vec}(\hat{A}_n - A)) \xrightarrow{D} \mathcal{N}\left(0, \mathcal{A}\Omega\mathcal{A}^T\right)$$
$$\sqrt{n}(\mathrm{vec}(\hat{M}_n - M)) \xrightarrow{D} \mathcal{N}\left(0, \left(\mathrm{vec}(M)^T \otimes A^+\right)\mathcal{A}\Omega\mathcal{A}^T\left(\mathrm{vec}(M) \otimes (A^+)^T\right)\right)$$

for $n \to \infty$, where A^+ denotes the Moore-Penrose inverse of A.

To use these results in practice, we need to estimate the asymptotic covariance matrix, and for this we can exploit the explicit formula above to modularize the estimation. The matrix Ω can be estimated separately based on the empirical cumulants only.

The following corollary, proved in Appendix 3.B.1, provides a compact way of representing the asymptotic covariance matrix, which avoids explicit usage of the large matrices $\mathrm{vec}(M)^T \otimes A^+$ and $\mathcal{A}$. The representation uses the matrices $B_k(M)$, which are defined by equation (3.A.3) in Appendix 3.A.1 and enter in the unique vectorizations of the Lyapunov equations when written as linear equations in the cumulants instead of in M.

Corollary 3.4.3. *If A is given by equation (3.4.18), the asymptotic covariance matrix of $\mathrm{vec}(\hat{M}_n)$ can be written as*

$$(A^+ B_{2,r}(M))\Omega(A^+ B_{2,r}(M))^T,$$

where $B_{2,r}(M)$ is the block diagonal matrix with $B_2(M)$ and $B_r(M)$ on the diagonal.

If not all rows in (3.4.18) are included in A, the matrix $B_{2,r}(M)$ should simply be replaced by the submatrix of $B_{2,r}(M)$ only indexed by the same rows as the chosen A.

3.5 Numerical Experiments

The Julia package `https://github.com/nielsrhansen/SteadyStateStatistics.jl` implements simulation from M-selfdecomposable distributions and computation of the least singular value estimator of M, based on the multivariate cumulant estimators in the Cumulants package [Domino et al., 2018].

To illustrate properties of the estimator, we ran a simulation study with the Lévy process having independent and identically distributed coordinates, each of which is a compound Poisson process with beta distributed jumps. We parametrize the Lévy process by the three parameters

$$\begin{aligned} \lambda > 0 &: \text{the rate parameter of the Poisson process} \\ \mu \in (0,1) &: \text{the mean parameter of the beta jump distribution} \\ \nu > 0 &: \text{the size parameter of the beta jump distribution.} \end{aligned}$$

The stable M-matrix is parametrized by the two parameters $\gamma \in \mathbb{R}$ and $\rho \in (-1/(d-1), 1)$. The parameter ρ is a correlation parameter, which means that with Σ the covariance matrix that solves the Lyapunov equation (3.1.2), then

$$\rho = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

for all $i \neq j$. The parameter γ controls the asymmetry of M with $\gamma = 0$ corresponding to the symmetric $M = -c\Sigma$ for some $c > 0$. The details of this parametrization are given in Appendix 3.B.5, which also describes how observations are sampled.

We show results for the settings given by the grid of the following simulation parameters:

$$\begin{aligned} d &: 3, 6, 12 \\ n &: 1000, 2000, 4000, 8000 \\ \gamma &: 5.0, 10.0, 15.0 \\ \rho &: 0.2, 0.8 \end{aligned}$$

and with $\lambda = 0.5$, $\mu = 0.8$ and $\nu = 1.0$. For each setting, $N = 100$ replications were done, and for each replication we computed the estimate $\hat{M}^{(l)}$, which by construction is standardized to have Frobenius norm one, that is, $\|\hat{M}^{(l)}\|_F^2 = \sum_{i,j}(\hat{M}_{ij}^{(l)})^2 = 1$. The matrix M was also standardized to have Frobenius norm one, and the mean squared error was computed as

$$\text{MSE} = \frac{1}{N}\sum_{l=1}^{N} \|\hat{M}^{(l)} - M\|_F^2.$$

Due to the standardization, $\|\hat{M}^{(l)} - M\|_F^2 = 2(1 - \cos(\theta^{(l)}))$, where $\theta^{(l)}$ is the angle between the subspaces spanned by $\hat{M}^{(l)}$ and M, see Figure 3.4.6. The MSE of the standardized estimator is thus equal to the mean of twice the cosine distance between $\hat{M}^{(l)}$ and M.

The MSE can be decomposed into a total variance term and a total squared bias term,

$$\text{MSE} = \frac{1}{N}\sum_{l=1}^{N} \|\hat{M}^{(l)} - \overline{M}\|_F^2 + \|\overline{M} - M\|_F^2.$$

According to Theorem 3.4.2, the $\sqrt{n}$-scaled total bias, $\sqrt{n}\|\overline{M} - M\|_F$, should vanish with n tending to infinity, and the $\sqrt{n}$-scaled root mean squared error should tend to the square root of the asymptotic total variance, that is, the square root of the trace of the asymptotic covariance matrix given by Theorem 3.4.2 or Corollary 3.4.3.

Figure 3.5.7 shows the scaled root mean squared error as well as the scaled total bias as a function of the sample size for the different dimensions and different values of the simulation parameters. The estimator is the least singular value estimator based on all second- and third-order cumulants. We see that for several of the settings there is a notable finite sample bias, in particular when the correlation is set to the high value

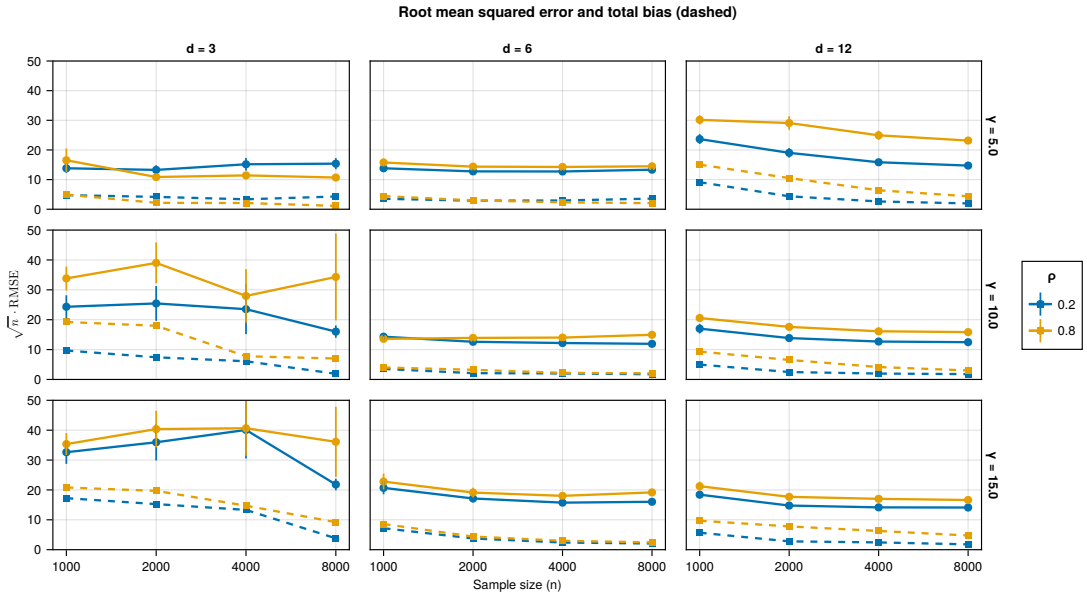


Figure 3.5.7: Estimation errors for the least singular value estimator based on all second- and third-order cumulants. Circles connected with full lines show the $\sqrt{n}$-scaled root mean squared error, while the squares connected with dashed lines show the $\sqrt{n}$-scaled total bias.

$\rho = 0.8$, but for large enough sample sizes the bias disappears and the scaled root mean squared error stabilizes.

Figure 3.5.8 compares the scaled root mean squared error to the square root of the asymptotic total variance obtained by Corollary 3.4.3. For all settings presented here, the asymptotic total variance eventually concurs with the mean squared error when the sample size is large enough. This is aligned with the observation that the total bias vanishes.

We ran simulations for other parameter values (data not shown). For $d = 24$ or $\gamma = 0$, the estimation problem is harder, and even larger sample sizes are required for accurate estimation. Other choices of parameters for the compound Poisson process were also tried, e.g., $\lambda = 0.1, 2.0$, $\mu = 0.2$ and $\nu = 0.1, 10$. While the results differed quantitatively from the results shown, the qualitative differences were minor.

In addition to the estimator based on second- and third-order cumulants only, we implemented and tested the estimator that additionally included the fourth-order cumulants. Figure 3.B.2 in Appendix 3.B.5 shows the results for this least singular value estimator based on all second-, third- and fourth-order cumulants. For $\rho = 0.8$, inclusion of fourth-order cumulants generally increased the estimation error. For $\rho = 0.2$, the effect of including the fourth-order cumulants was mixed and depended on d and γ.

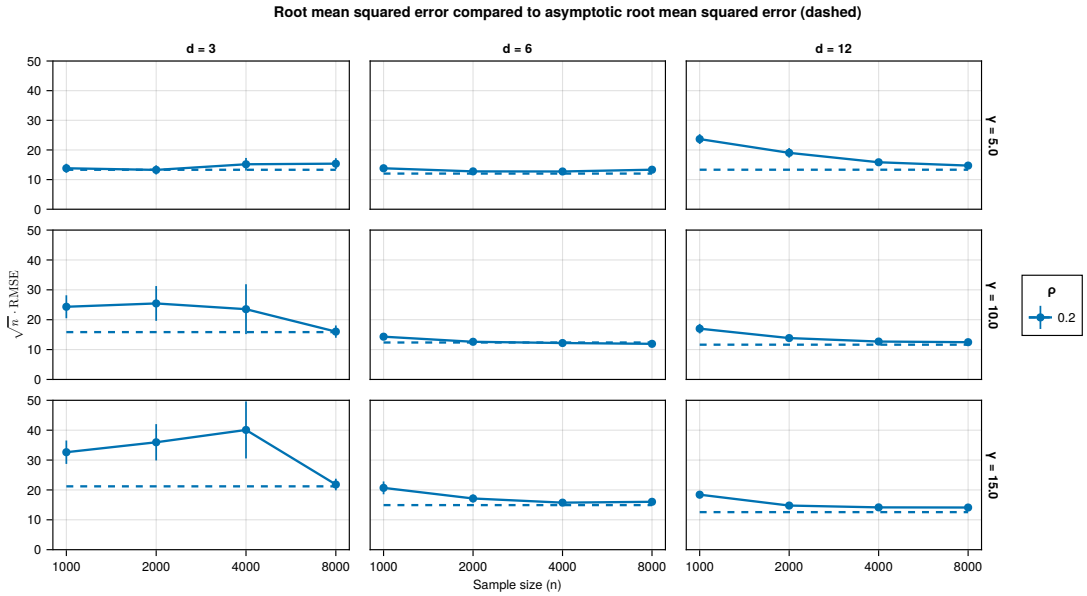


Figure 3.5.8: Root mean squared error scaled with $\sqrt{n}$ compared to the square root of the total asymptotic variance (dashed).

3.6 Discussion

Our main identifiability result, Theorem 3.3.1, leaves two minor questions open. First, it remains unknown to what extent the assumption that the graph contains all self-loops can be relaxed. Our proof strategy relies on the self-loops to ensure the existence of a sufficiently simple and stable M that allows for explicit computations, see also Remark 3..3. Second, it remains unknown whether the exceptional set, where identification fails, corresponds only to parameter values that disconnect the graph. If we are willing to make additional assumptions, stronger results can also be obtained. If we, for example, additionally assume that the r-th-order cumulant of the Lévy process, for some $r \geq 3$, is known and diagonal, Proposition 3.B.12 in Appendix 3.B.3 shows that the drift matrix M is, in fact, generically identifiable for *any* directed graph with all self-loops present.

There are also natural generalizations and extensions of our results. One direction of future research is to allow the cumulants of the Lévy process to be non-diagonal, as encoded by blunt edges in Appendix 3.B.2. Following the arguments by Varando and Hansen [2020], the marginalization of an M-selfdecomposable distribution is a distribution with cumulants solving continuous Lyapunov equations for a new M and a new $\mathcal{C}$, whose sparsity patterns are obtained by projecting the directed graph to a graph with directed and blunt edges. This generalization would be a way of expanding the model class to be able to handle latent variables. However, the marginalized M is not guaranteed to be stable, and the marginalized $\mathcal{C}_2$ is, for example, not guaranteed to be positive definite [Varando and Hansen, 2020]. The proof of Theorem 3.3.1 could, nev-

ertheless, easily be modified to allow for specific r-th-order cumulants to be non-zero, e.g., those whose indices do not correspond to the rows used to prove the generic rank of A_{off}. Another direction of future research is to consider structural identifiability instead. The results by Améndola et al. [2025] on structural identifiability are based on the second order continuous Lyapunov equation only, and by expanding their approach to take higher order cumulants into account, we could likely answer some of the open questions by Améndola et al. [2025].

The present paper focuses on the identification theory and does not consider specific applications, but the theory is motivated by the possible applications to cross-sectional observations from single cell biology. The M-selfdecomposable distributions can, as steady-state distributions of solutions to stochastic differential equations, be given natural causal interpretations [Sokol and Hansen, 2014; Varando and Hansen, 2020], which was exploited by, for example, Wang et al. [2023] to learn gene regulatory networks from temporal snapshots. The network is then encoded by the sparsity pattern of M, and if the model has a causal interpretation, an edge represents a direct causal effect. An alternative to M-selfdecomposable distributions is classical linear structural equations [Drton, 2018], but feedback loops present both mathematical and interpretational challenges for such models. The implicit temporal nature of M-selfdecomposable distributions allows feedback loops to be naturally incorporated into the model, and the interpretation and properties of the model do not depend on whether there are feedback loops or not. However, a realistic application to single cell data requires a measurement noise model, such as the zero-inflated Poisson model by Lorch et al. [2026], and this is beyond the scope of the current paper.

Finally, as demonstrated by our simulation study, even when the drift matrix is identifiable in the non-Gaussian case from the higher-order cumulants, estimation is hard for small sample sizes. This is also seen in the general literature that uses non-Gaussianity and higher-order cumulants for identification and estimation [Shimizu et al., 2006; Schkoda et al., 2024; Tramontano et al., 2025], and Wang and Drton [2023] explicitly comments on the need for large sample sizes. An ongoing research project explores how additional assumptions, e.g., sparsity of M or homogeneity of the Lévy process across dimensions, can improve finite sample estimation accuracy.

Acknowledgment

We would like to thank Mathias Drton and Anton Rask Lundborg for helpful discussions. This work was supported by a research grant (NNF20OC0062897) from Novo Nordisk Fonden. Proposition 3.2.2, that the cumulants of the steady-state distribution satisfy the higher-order continuous Lyapunov equations, was first proved by Jeffrey Adams and included in his PhD thesis [Adams, 2024]. The proof given here was obtained independently.

Appendix for 'Identifiability and Estimation in Continuous Lyapunov Models'

3.A. Appendix – proofs of main results

3.A.1 The Lyapunov equations

Proof of Proposition 3.2.2. When M is a stable matrix and Z is a Lévy process satisfying (3.2.3), the stochastic integral

$$X = \int_0^\infty e^{sM} \mathrm{d}Z_s \tag{3.A.1}$$

is well defined, and the distribution of X is the unique steady-state distribution in $\mathcal{P}_G$ determined by M and the Lévy process. This was known by Sato and Yamazato [1984], and stated indirectly in their introduction with reference to an abstract by Wolfe [1982]. It is a direct consequence of Theorem 5.2(i) by Sato and Yamazato [1983].

Using the integral representation (3.A.1) and multilinearity of the cumulant operator,

$$\begin{aligned}
\mathcal{K} &= \mathrm{cum}_k(X) \\
&= \int_0^\infty \cdots \int_0^\infty \mathrm{cum}_k(\mathrm{d}Z_{s_1}, \ldots, \mathrm{d}Z_{s_k}) \times_1 e^{s_1 M} \times_2 e^{s_2 M} \cdots \times_k e^{s_k M} \\
&= \int_0^\infty \mathcal{C}_k \times_1 e^{sM} \times_2 e^{sM} \cdots \times_k e^{sM} \mathrm{d}s
\end{aligned}$$

where we have used that for a Lévy process, the k-th-order cumulant measure is "diagonal" and equals

$$\mathrm{cum}_k(\mathrm{d}Z_{s_1}, \ldots, \mathrm{d}Z_{s_k}) = \mathcal{C}_k \delta_{s_1,\ldots,s_k} H^1(ds_1, \ldots, ds_k)$$

with H^1 the 1-dimensional Hausdorff measure. That $\mathcal{K}$ solves (3.2.5) follows from Proposition 3.A.2. □

For the second order Lyapunov equation it is well-known that it can be vectorized as

$$(I \otimes M + M \otimes I)\mathrm{vec}(\Sigma) + \mathrm{vec}(\mathcal{C}_k) = 0,$$

which shows that it has a unique solution for Σ if no pair of eigenvalues of M add up to zero. This can be generalized to the continuous Lyapunov equation of any order.

Proposition 3.A.1. *The vectorization of the k-th-order continuous Lyapunov equation, (3.2.5), can be written as*

$$0 = \mathrm{vec}(\mathcal{C}_k) + \left(\sum_{i=1}^{k} (\underbrace{I_d \otimes \cdots \otimes I_d}_{k-i \text{ times}} \otimes M \otimes \underbrace{I_d \otimes \cdots \otimes I_d}_{i-1 \text{ times}}) \right) \mathrm{vec}(\mathcal{K}). \tag{3.A.2}$$

As a consequence of this, the k-th-order continuous Lyapunov equation is uniquely solvable whenever no k eigenvalues (the same eigenvalue can be picked multiple times) add up to zero.

Proof. By rules of vectorizations and Kronecker products, see Chapter 2 by Magnus and Neudecker [2019], we can write

$$\mathrm{vec}(\mathcal{K} \times_i M) = (\underbrace{I_d \otimes \cdots \otimes I_d}_{k-i \text{ times}} \otimes M \otimes \underbrace{I_d \otimes \cdots \otimes I_d}_{i-1 \text{ times}})\mathrm{vec}(\mathcal{K}),$$

for $\mathcal{K} \in \mathrm{Sym}^k(\mathbb{R}^d)$ and $M \in \mathbb{R}^{d\times d}$. This yields equation (3.A.2).

Furthermore, since the eigenvalues of $A \otimes B$ are equal to products of the eigenvalues of A and B, it follows that no matter the place of the single M, the eigenvalues of any tensor product, $I_d \otimes \cdots \otimes I_d \otimes M \otimes I_d \otimes \cdots \otimes I_d$, is just equal to the eigenvalues of M (each with multiplicity d). Furthermore, since all the matrices of this form, $I_d \otimes \cdots \otimes I_d \otimes M \otimes I_d \otimes \cdots \otimes I_d$, commute with each other (since all the M's are placed in different modes), the eigenvalues of the sum of them in equation (3.A.2) is the sum of the their eigenvalues. Thus, any eigenvalue of the sum in (3.A.2) is a sum of k eigenvalues of M. And if no k eigenvalues of M add up to zero, the sum is a matrix of full rank, and the k-th-order continuous Lyapunov equation is uniquely solvable for $\mathcal{K}$ given M and $\mathcal{C}_k$. □

Equation (3.A.2) is the standard vectorization of the Lyapunov equation, and we could instead do the unique vectorization of both the equation and of $\mathcal{K}$,

$$0 = \mathrm{vec}_u(\mathcal{C}_k) + B_k(M)\mathrm{vec}_u(\mathcal{K}). \tag{3.A.3}$$

The rows of $B_k(M)$ can be obtained by adding the columns of the matrix in equation (3.A.2) corresponding to the same cumulant, i.e., for an index set $(i_1 \dots i_k)$ taking all columns corresponding to an index of $\mathcal{K}$ which is a permutation of $(i_1 \dots i_k)$. To also have it be the unique vectorization of the equation we only keep the rows indexed by $(i_1 \dots i_k)$ such that $i_1 \leq \cdots \leq i_k$ as was done for the unique vectorization isolating $\mathrm{vec}(M)$ instead in equation (3.3.8).

When M is stable, there is an explicit formula for the solution of the continuous Lyapunov equation.

Proposition 3.A.2 (Corollary 3.1, Xu and Wang [2022])**.** *Let M be a $d\times d$ stable matrix, then the k-th-order continuous Lyapunov equation (3.2.5) has the unique solution*

$$\mathcal{K} = \int_0^\infty \mathcal{C}_k \times_1 e^{Mt} \times_2 e^{Mt} \cdots \times_k e^{Mt} \mathrm{d}t. \tag{3.A.4}$$

Equation (3.A.4) is the key to deriving the so-called trek formulas for the entries of the cumulants of the continuous Lyapunov equations. The general trek rules provide a graphical description of the entries of the cumulants that can allow one to determine zeros in the cumulants just from the absence of certain treks, corresponding to the absence of several paths, in the graph. Additionally, a specialized version of the trek formulas is a key ingredient in the proof of the main identifiability result, Theorem 3.3.1. The details and proofs concerning treks and the trek representations for cumulants are provided in Appendix 3.B.2.

3.A.2 Proof of the main result and additional lemmas and corollaries

The proof of Theorem 3.3.1 uses a specific restricted trek rule for the second- and r-th-order cumulants, very similar to the restricted trek rule given in Proposition 4.3 by Boege et al. [2025]. Such restricted trek rules can be seen as special cases of a more general trek rule first published by Hansen [2025] in the covariance case. We provide the extension of the general trek rule from the covariance case to cumulants of any order in Appendix 3.B.2, where a definition of treks is also given. For the purpose of proving Theorem 3.3.1, we give here a simple proof of the restricted trek rule for directed acyclic graphs.

Lemma 3.A.3. *Let $G = ([d], E)$ be any directed acyclic graph (DAG) with all self-loops and with the nodes ordered in a topological order. Let $r \geq 3$ be an integer and consider the second- and r-th-order Lyapunov model with $\mathcal{C}_2$ and $\mathcal{C}_r$ equal to the identity matrix and identity tensor, respectively, and $M \in \mathbb{R}^E$ such that $M_{ii} = -1/(r\zeta)$. Then we have the following specialized trek rules for the second- and r-th-order cumulants*

$$\Sigma_{i_1 i_2} = \sum_{\tau \in \mathcal{T}(i_1, i_2)} \zeta^{\ell_{i_1} + \ell_{i_2} + 1} \left(\frac{r}{2}\right)^{\ell_{i_1} + \ell_{i_2} + 1} \binom{\ell_{i_1} + \ell_{i_2}}{\ell_{i_1}} \omega(M, \tau)$$

$$\mathcal{K}_{i_1 \dots i_r} = \sum_{\tau \in \mathcal{T}(i_1, \dots, i_r)} \zeta^{\sum_{j=1}^r \ell_{i_j} + 1} \frac{(\sum_{j=1}^r \ell_{i_j})!}{\prod_{j=1}^r (\ell_{i_j}!)} \omega(M, \tau),$$

where $\mathcal{T}(i_1, i_2)$ and $\mathcal{T}(i_1, \dots, i_r)$ denote all treks without self-loops between i_1 and i_2 and $i_1, \dots, i_r$, respectively, and $\omega(M, \tau)$ denotes the trek monomial. That is, $\omega(M, \tau) = \prod_{\alpha \to \beta \in \tau} M_{\beta\alpha}$ since $\mathcal{C}_2$ and $\mathcal{C}_r$ are the identity.

Proof. For notational convenience the proof is given for the case $r = 3$. It is, however, easily seen from this proof how the general case follows by keeping track of r mode products instead of 3, and replacing 3 by r everywhere.

Let $M_{\text{DAG}} = M - (-1/(3\zeta) I_{d \times d})$ denote the strictly lower triangular part of M which corresponds to all the edges in G which are not self-loops, by subtracting a scaling of the $d \times d$ identity matrix, $I_{d \times d}$. We decompose M as

$$M = M_{\text{DAG}} - \frac{1}{3\zeta} I_{d \times d}.$$

Since M_{DAG} is lower-triangular it has spectral radius 0, which is especially less than 1. We can use this decomposition of M instead of the $(\Lambda - I_{d \times d})$ used in Proposition 3.B.11, or in Proposition 2.4 by Hansen [2025] for the covariance case, and do exactly the same proof with the identity instead of $\mathcal{C}$ and M_{DAG} instead of Λ and the exponent of e will be $-2/(3\zeta)$ and $-1/\zeta$, respectively. When letting $I_{d \times d \times d}$ denote the identity $d \times d \times d$ tensor, writing out the derivation of the third-order trek formula in this special

case becomes

$$\begin{aligned}
\mathcal{K} &= \int_0^\infty I_{d\times d\times d} \times_1 e^{Mt} \times_2 e^{Mt} \times_3 e^{Mt} \mathrm{d}t \\
&= \int_0^\infty I_{d\times d\times d} \times_1 e^{(M_{\mathrm{DAG}} - 1/(3\zeta)I_{d\times d})t} \times_2 e^{(M_{\mathrm{DAG}} - 1/(3\zeta)I_{d\times d})t} \times_3 e^{(M_{\mathrm{DAG}} - 1/(3\zeta)I_{d\times d})t} \mathrm{d}t \\
&= \int_0^\infty e^{-\frac{1}{\zeta}t} I_{d\times d\times d} \times_1 e^{M_{\mathrm{DAG}}t} \times_2 e^{M_{\mathrm{DAG}}t} \times_3 e^{M_{\mathrm{DAG}}t} \mathrm{d}t \\
&= \int_0^\infty e^{-\frac{1}{\zeta}t} I_{d\times d\times d} \times_1 \sum_{n=0}^\infty \frac{t^n}{n!} M_{\mathrm{DAG}}^n \times_2 \sum_{m=0}^\infty \frac{t^m}{m!} M_{\mathrm{DAG}}^m \times_3 \sum_{k=0}^\infty \frac{t^k}{k!} M_{\mathrm{DAG}}^k \mathrm{d}t \\
&= \sum_{n=0}^\infty \sum_{m=0}^\infty \sum_{k=0}^\infty \int_0^\infty \frac{t^{n+m+k} e^{-\frac{1}{\zeta}t}}{n!m!k!} I_{d\times d\times d} \times_1 M_{\mathrm{DAG}}^n \times_2 M_{\mathrm{DAG}}^m \times_3 M_{\mathrm{DAG}}^k \mathrm{d}t \\
&= \sum_{n=0}^\infty \sum_{m=0}^\infty \sum_{k=0}^\infty \frac{\zeta^{n+m+k+1}\Gamma(n+m+k+1)}{n!m!k!} I_{d\times d\times d} \times_1 M_{\mathrm{DAG}}^n \times_2 M_{\mathrm{DAG}}^m \times_3 M_{\mathrm{DAG}}^k.
\end{aligned}$$

The (i,j,k) entry of the tensor $I_{d\times d\times d} \times_1 M_{\mathrm{DAG}}^n \times_2 M_{\mathrm{DAG}}^m \times_3 M_{\mathrm{DAG}}^k$ corresponds to the sum of all treks in the acyclic part between i, j and k with a given top node with the paths going to each of them having lengths n, m, k, respectively. To see this we write out the (i,j,k) entry

$$\left(I_{d\times d\times d} \times_1 M_{\mathrm{DAG}}^n \times_2 M_{\mathrm{DAG}}^m \times_3 M_{\mathrm{DAG}}^k\right)_{ijk} = \sum_{\alpha=1}^d (M_{\mathrm{DAG}}^n)_{i\alpha} \cdot (M_{\mathrm{DAG}}^m)_{j\alpha} \cdot (M_{\mathrm{DAG}}^k)_{k\alpha},$$

and note that $(M_{\mathrm{DAG}}^n)_{i\alpha}$ is exactly the sum of all directed walk polynomials of length n from α to i in the directed acyclic part of G since M_{DAG} is the adjacency matrix (with weights) of the directed acyclic part of G. See also the proof of Proposition 3.B.7.

Because the graph is assumed to be a DAG, the sum above becomes finite, and for the (i,j,k)-th entry we therefore obtain exactly the finite sum over all treks between i, j and k as given the statement of the proposition by inserting the above into the sum.

The proof is completely analogous for the covariance just with one term less and the exponent of the exponential being $(-2/(3\zeta))t$ instead of $(-1/\zeta)t$. □

The specific trek rule above gives a finite sum representation of the entries of the cumulants, as opposed to the potential infinite sums in the general trek rules in Appendix 3.B.2. The proof of Theorem 3.3.1 uses the special case given in the following corollary.

Corollary 3.A.4. *If we further assume that $M_{\beta\alpha} = 1$ for $\alpha \to \beta \in E$ with $\beta \neq \alpha$, then*

$$\begin{aligned}
\Sigma_{i_1 i_2} &= \sum_{\tau \in \mathcal{T}(i_1, i_2)} \zeta^{\ell_{i_1} + \ell_{i_2} + 1} \left(\frac{r}{2}\right)^{\ell_{i_1} + \ell_{i_2} + 1} \binom{\ell_{i_1} + \ell_{i_2}}{\ell_{i_1}} \\
\mathcal{K}_{i_1 \dots i_r} &= \sum_{\tau \in \mathcal{T}(i_1, \dots, i_r)} \zeta^{\sum_{j=1}^r \ell_{i_j} + 1} \frac{(\sum_{j=1}^r \ell_{i_j})!}{\prod_{j=1}^r (\ell_{i_j}!)}.
\end{aligned}$$

In addition to the specific trek rule, the proof of Theorem 3.3.1 also relies on the following sum identity.

Lemma 3.A.5. *Let d, q be non-negative integers such that $q \leq d-1$ and $r \geq 3$ an integer, then*

$$\sum_{i=0}^{d-1} \left(\frac{r}{2}\right)^{d-1-i} (-1)^i c(d,q,i) = \left(\frac{r}{2}-1\right)^{d-1} (-1)^q, \tag{3.A.5}$$

where the coefficients are

$$c(d,q,i) = \sum_{j=0}^{i} (r-1)^j \binom{q}{j} \binom{d-1-q}{i-j}.$$

Proof. We first prove the following recursive identity for the coefficients

$$c(d,q,i) = c(d,q-1,i) + (r-2)c(d-1,q-1,i-1).$$

To prove this, note that $c(d,q,i)$ is the coefficient in front of x^i if you expand the following polynomial in x using the binomial theorem, $(1+(r-1)x)^q(1+x)^{d-1-q}$. Thus, $c(d,q-1,i)$ is the coefficient in front of x^i in $(1+(r-1)x)^{q-1}(1+x)^{d-1-(q-1)}$ and $c(d-1,q-1,i-1)$ is the coefficient in front of x^{i-1} in $(1+(r-1)x)^{q-1}(1+x)^{d-2-(q-1)}$. Hence, $c(d,q-1,i)+c(d-1,q-1,i-1)$ will be the coefficient in front of x^i in

$$\begin{aligned}
&(1+(r-1)x)^{q-1}(1+x)^{d-1-(q-1)} + (r-2)x(1+(r-1)x)^{q-1}(1+x)^{d-2-(q-1)} \\
&= (1+(r-1)x)^{q-1}(1+x)^{d-2-(q-1)}(1+x+(r-2)x) \\
&= (1+(r-1)x)^{q}(1+x)^{d-1-q}.
\end{aligned}$$

Thus, the recursion formula is true.

We now obtain the result about the summation by induction on (d,q), where we induct over this set by what the sum $d+q$ is.

The base case is $d=2$ and $q=0$. The proof is the same for any d, so we just prove the formula for any d and $q=0$. When q is 0 we obtain

$$c(d,0,i) = \sum_{j=0}^{i} (r-1)^j \binom{0}{j} \binom{d-1}{i-j} = \binom{d-1}{i},$$

since the only non-zero contribution to the sum above is when $j=0$. Inserting this into the left hand side of equation (3.A.5) we can prove the base case by application of the binomial formula

$$\sum_{i=0}^{d-1} \left(\frac{r}{2}\right)^{d-1-i} (-1)^i c(d,0,i) = \sum_{i=0}^{d-1} \left(\frac{r}{2}\right)^{d-1-i} (-1)^i \binom{d-1}{i} = \left(\frac{r}{2}-1\right)^{d-1}.$$

We now consider the induction step, so consider the lefthand side in equation (3.A.5) and apply the recursion formula to obtain

$$\begin{aligned}
&\sum_{i=0}^{d-1}\left(\frac{r}{2}\right)^{d-1-i}(-1)^i c(d,q,i)\\
&=\sum_{i=0}^{d-1}\left(\frac{r}{2}\right)^{d-1-i}(-1)^i c(d,q-1,i)+(r-2)\sum_{i=0}^{d-1}\left(\frac{r}{2}\right)^{d-1-i}(-1)^i c(d-1,q-1,i-1).
\end{aligned}$$

By the induction hypothesis the left sum is equal to $(r/2-1)^{d-1}(-1)^{q-1}$. The rightmost sum can be written as

$$\begin{aligned}
&\sum_{i=0}^{d-1}\left(\frac{r}{2}\right)^{d-1-i}(-1)^i c(d-1,q-1,i-1)\\
&=-\sum_{i=0}^{d-1}\left(\frac{r}{2}\right)^{(d-2)-(i-1)}(-1)^{i-1} c(d-1,q-1,i-1)\\
&=-\sum_{i=-1}^{d-2}\left(\frac{r}{2}\right)^{(d-2)-i}(-1)^i c(d-1,q-1,i)\\
&=-\sum_{i=0}^{d-2}\left(\frac{r}{2}\right)^{(d-2)-i}(-1)^i c(d-1,q-1,i)-\left(\frac{r}{2}\right)^{d-1}c(d-1,q-1,-1)\\
&=-\left(\frac{r}{2}-1\right)^{d-2}(-1)^{q-1},
\end{aligned}$$

by using the induction hypothesis and that $c(d-1,q-1,-1)=0$. Inserting the expressions for the two sums we obtain

$$\begin{aligned}
\sum_{i=0}^{d-1}\left(\frac{r}{2}\right)^{d-1-i}(-1)^i c(d,q,i)&=\left(\frac{r}{2}-1\right)^{d-1}(-1)^{q-1}-(r-2)\left(\frac{r}{2}-1\right)^{d-2}(-1)^{q-1}\\
&=\left(\frac{r}{2}-1\right)^{d-1}(-1)^{q}\left(-1+\frac{r-2}{r/2-1}\right)\\
&=\left(\frac{r}{2}-1\right)^{d-1}(-1)^{q}\left(\frac{-r/2+1+r-2}{r/2-1}\right)\\
&=\left(\frac{r}{2}-1\right)^{d-1}(-1)^{q},
\end{aligned}$$

as we needed to show. □

In the proof of Theorem 3.3.1 we will only need the general expression for the rows of $A_r(\mathcal{K})$ when precisely two distinct subscripts appear, namely the rows $(i_1 i_2)$ and $(i_1 \dots i_1 i_2)$ (and $(i_1 i_2 \dots i_2)$), when $i_1 \neq i_2$. The explicit expression for these rows is

$$A_r(\mathcal{K})_{(i_1\dots i_1 i_2),(\alpha\to\beta)} = \begin{cases} 0 & \text{if} \beta \neq i_1, i_2 \\ (r-1)\mathcal{K}_{i_1\dots i_1 i_2 \alpha} & \text{if} \beta = i_1 \\ \mathcal{K}_{i_1\dots i_1 \alpha} & \text{if} \beta = i_2. \end{cases} \tag{3.A.6}$$

However, as discussed in Section 3.3.4, the general expressions can easily be obtained as well.

Proof of Theorem 3.3.1. Since we only consider diagonal $\mathcal{C}_2$ and $\mathcal{C}_r$, the linear system corresponding to the second and r-th-order Lyapunov equations from equation (3.3.13) reduces to

$$\begin{pmatrix} A_2(\Sigma)_{\text{off}} & 0 & 0 \\ A_r(\mathcal{K})_{\text{off}} & 0 & 0 \\ A_2(\Sigma)_{\text{diag}} & I_d & 0 \\ A_r(\mathcal{K})_{\text{diag}} & 0 & I_d \end{pmatrix} \begin{pmatrix} \text{vec}(M) \\ \text{diag}(\mathcal{C}_2) \\ \text{diag}(\mathcal{C}_r) \end{pmatrix} = 0, \tag{3.A.7}$$

where $A_2(\Sigma)_{\text{off}}$ and $A_r(\mathcal{K})_{\text{off}}$ denote all the rows indexed by the off-diagonal tensor entries of $A_2(\Sigma)$ and $A_r(\mathcal{K})$, respectively. That is all the rows not indexed by (ii) or $(i\dots i)$, and $A_2(\Sigma)_{\text{diag}}$ and $A_r(\mathcal{K})_{\text{diag}}$ denote the rows indexed by the diagonal entries.

Thus, the question of identifiability, i.e., whether this matrix in equation (3.A.7) has full rank minus 1, reduces to the question of whether the matrix

$$\begin{pmatrix} A_2(\Sigma)_{\text{off}} \\ A_r(\mathcal{K})_{\text{off}} \end{pmatrix}$$

has rank d^2-1. However, this matrix is in general not square, so we have to consider the determinant of a $(d^2-1)\times(d^2-1)$ submatrix.

Any connected graph will have at least one polytree as a subgraph without the self-loops, which is a connected directed acyclic graph with exactly $d-1$ edges. Pick any such subpolytree and let $G' = ([d], E')$ be the subgraph of G where we have the edges in the polytree as well as all the self-loops. Thus, G' will be a connected directed acyclic graph (disregarding the self-loops) with $d-1$ non-self-loop edges and all self-loops and be a subgraph of G. Firstly, we can without loss of generality assume that the nodes of G' are ordered according to a topological order, $\leq$. We order any possible edges on d nodes according to a lexicographic ordering with the ordering on each coordinate the topological order, thus $(i\to j)\leq(i'\to j')$ if $i\leq i'$ or $i=i'$ and $j\leq j'$. According to this order the first element is $1\to 1$, this is the column we will omit from the $(d^2-1)\times(d^2-1)$ submatrix we consider.

We pick the following rows. We pick all of $A_2(\Sigma)_{\text{off}}$ and of $A_r(\mathcal{K})_{\text{off}}$ we pick the rows of the form $(ij\dots j)$ with $i<j$ and $(i\dots ij)$ such that $i\to j\in E'$. By counting, these three groups each contribute with $d(d-1)/2$, $d(d-1)/2$ and $d-1$ rows, respectively, so in total this is exactly d^2-1 rows. We will denote this submatrix by A.

From now on when referring to rows of A they will be (ij), $(ij\dots j)$ or $(i\dots ij)$, which will implicitly inform whether it is a row coming from the second- or r-th-order Lyapunov equation. The rest of the proof consists of showing that A has full rank generically.

As discussed in Section 3.3.4 we can prove that A has full rank generically, that is, outside of a proper algebraic subset, by exhibiting a choice of parameters $(M, \mathcal{C}_2, \mathcal{C}_3)$ where A has full rank, since the entries of A can be written as rational functions in M, $\mathcal{C}_2$ and $\mathcal{C}_r$. Let $\mathcal{C}_2$ and $\mathcal{C}_r$ be the identity matrix and identity $d \times \cdots \times d$ tensor, respectively and let $M \in \mathbb{R}^{E'}$ with diagonal $M_{ii} = -1/(r\zeta)$ and $M_{\beta\alpha} = 1$ for $\alpha \to \beta \in E'$ when $\alpha \neq \beta$. Since we picked E' such that G' (without the self-loops) is a polytree, this implies that we can use the polynomial expressions in ζ for the entries of Σ and $\mathcal{K}$ given by Corollary 3.A.4.

Thus, $\det(A)$ will be a polynomial in ζ. The remaining idea of the proof is to show that this polynomial is not the zero polynomial. This is done by using Leibniz' formula for determinants to achieve an expression for the lowest degree non-vanishing term of $\det(A)$,

$$\det(A) = \sum_{\pi \in S_{d^2-1}} \operatorname{sgn}(\pi) \prod_{i=0}^{d^2-1} A_{i\pi(i)}. \tag{3.A.8}$$

From Corollary 3.A.4 we see that a given non-zero entry of A will have lowest degree in ζ equal to the total length of the shortest trek in G' between the nodes indexing the entry of Σ or $\mathcal{K}$ plus 1. Thus, only entries of A equal to the diagonal entries Σ_{ii} and $\mathcal{K}_{i\ldots i}$ will have low $\deg(\Sigma_{ii}) = $ low $\deg(\mathcal{K}_{i\ldots i}) = 1$, the lowest possible lowest degree. Similarly, the only way to obtain lowest degree equal to 2 is the case where $i \neq j$ and $i \to j \in E'$, then low $\deg(\Sigma_{ij}) = $ low $\deg(\mathcal{K}_{i\ldots ij}) = 2$, and low $\deg(\mathcal{K}_{ij\ldots j}) = r$. For two distinct nodes with no edge between them, the lowest degree of Σ_{ij} or $\mathcal{K}_{i\ldots ij}$ will be at least three.

Equations (3.3.10) and (3.A.6) yield that in row (ij) there is a diagonal entry in the columns $(i \to j)$ and $(j \to i)$, and in row $(i \ldots ij)$ there is a diagonal entry in column $(i \to j)$ and similarly for $(ij \ldots j)$ it will only be in column $(j \to i)$. Thus, there will never be a diagonal entry in the self-loop columns in A since we do not have the (ii) or $(i \ldots i)$ rows.

Now the claim is that it is possible to pick a permutation (there will be several) such that all non self-loop columns correspond to a diagonal entry and all self-loop columns will correspond to a Σ_{ij} or $\mathcal{K}_{i\ldots ij}$ with an edge $i \to j \in E'$. By the discussion of possible lowest degrees for entries of Σ and $\mathcal{K}$ this is the term of lowest possible degree of the determinant of A by Leibniz' formula (3.A.8). We now characterize precisely which permutations $\pi \in S_{d^2-1}$ obtain this degree.

We subdivide the rows and columns into the following four sets

$$\begin{aligned}
\mathbf{R}_1 &= \{(ij) \mid i < j,\ i \to j \in E'\} & \mathbf{C}_1 &= \{(i \to j) \mid i \to j \in E'\} \\
\mathbf{R}_2 &= \{(i \ldots ij) \mid i < j,\ i \to j \in E'\} & \mathbf{C}_2 &= \{(i \to i) \mid i \in [d] \setminus \{1\}\} \\
\mathbf{R}_3 &= \{(ij) \mid i < j,\ i \to j \notin E'\} & \mathbf{C}_3 &= \{(i \to j) \mid i < j,\ i \to j \notin E'\} \\
\mathbf{R}_4 &= \{(ij \ldots j) \mid i < j\} & \mathbf{C}_4 &= \{(j \to i) \mid i < j\}.
\end{aligned}$$

For the columns $\mathbf{C}_3$ the only row where it is possible to obtain a diagonal entry for $(i \to j) \in \mathbf{C}_3$ is exactly the matching row $(ij) \in R_3$. Thus, any permutation $\pi \in S_{d^2-1}$ has to fix these rows to their corresponding columns.

Furthermore, for the rows $(ij\dots j) \in \mathbf{R}_4$ the only possible way to obtain either $\mathcal{K}_{i\dots i}$ or $\mathcal{K}_{i\dots ij}$ is in columns $(j \to i)$ or $(i \to j)$, respectively. However, in column $i \to j$ it will always be possible to pick π to obtain a diagonal entry from one of the other rows instead. Therefore, any permutation π also has to fix the rows $\mathbf{R}_4$ to be mapped to the set of columns going against the topological order, $\mathbf{C}_4$.

By reordering the rows and columns (and only changing the determinant up to a sign) we can assume π is just the identity permutation on $\mathbf{R}_3 \cup \mathbf{R}_4$ to $\mathbf{C}_3 \cup \mathbf{C}_4$. By considering Leibniz' formula they will always contribute with a factor of

$$\left(\frac{r}{2}\right)^{d(d-1)/2-(d-1)} \zeta^{d(d-1)-(d-1)}$$

to the lowest degree term of the determinant.

We are left to consider the $2(d-1)\times 2(d-1)$ submatrix $A_{(\mathbf{R}_1\cup\mathbf{R}_2),(\mathbf{C}_1\cup\mathbf{C}_2)}$ with columns indexed by the edges in the polytree, C_1 and the self-loops except the one for the first source node, $\mathbf{C}_2$. Recalling the lexicographic ordering of the edges in E', we will now order each of the sets of rows individually according to this as well and order the self-loops $\mathbf{C}_2$ *not* according to the topological order of the nodes but according to when a node first appears in the lexicographically ordered edges $E' = \{(e_{11} \to e_{12}, \dots, e_{(d-1)1} \to e_{(d-1)2}\}$, denote this $\{(i_2 \to i_2), \dots (i_d \to i_d)\}$. This is not guaranteed to coincide with the topological order, but it is possible, for example if G' is a path. Thus, it is guaranteed that $i_2 = e_{12}$ and that i_j will be at least $e_{(j-1)1}$ or $e_{(j-1)2}$. So the matrix will be indexed like this:

	$(e_{11} \to e_{12})$	$\cdots$	$(e_{(d-1)1} \to e_{(d-1)2})$	$(i_2 \to i_2)$	$\cdots$	$(i_d \to i_d)$
$(e_{11}e_{12})$	$*$			$*$		
$\vdots$						
$(e_{(d-1)1}e_{(d-1)2})$						
$(e_{11}\dots e_{11}e_{12})$						
$\vdots$						
$(e_{(d-1)1}\dots e_{(d-1)1}e_{(d-1)2})$						

To obtain the lowest possible degree, the first $d-1$ columns should all contribute with a diagonal entry in the product, and the last $d-1$ columns should contribute with a Σ_{ij} or $\mathcal{K}_{i\dots ij}$, potentially multiplied by $(r-1)$, with $i \to j \in E'$. For example, for the first row, the permutation has to map it to one of the columns marked with a '$*$' above .

We number the rows and columns $1, \dots, 2(d-1)$ in the order explained above. For the first $d-1$ columns to correspond to diagonal entries, the permutation π has to satisfy

$$\pi(i) = i \text{ if } i = 1, \dots, d-1 \qquad \text{or} \qquad \pi(i) = i-(d-1) \text{ if } i = d, \dots, 2(d-1),$$

for all $i = 1, \dots, 2(d-1)$. In the first case we pick the possible diagonal entry in $A_2(\Sigma)$ and in the second case in $A_r(\mathcal{K})$ and there are always exactly these two choices. They will contribute with $(r/2)\zeta$ and ζ to the lowest degree term, respectively.

Then for the $(d-1)$ rows not picked out to have a diagonal entry, it has to be matched with a self-loop column instead. This is done consecutively, so first we consider either row 1, $(e_{11}e_{12})$, or row $1+(d-1)$, $(e_{11}\dots e_{11}e_{12})$, and whichever one was not picked for the first column $e_{11} \to e_{12}$ has to be assigned column $e_{12} \to e_{12}$, since $e_{11} \to e_{11}$ was the column removed from the matrix, and so on. Depending on which i was picked above the opposite happens now,

$$\pi(i) = i + (d-1) \text{ if } \; i = 1, \dots, d-1 \qquad \text{or} \qquad \pi(i) = i \text{ if } \; i = d, \dots, 2(d-1),$$

for all $i = 1..2(d-1)$. This will contribute with a factor $(r/2)^2\zeta^2$ or ζ^2 (or $(r-1)\zeta^2$), respectively.

If there is only one source in G' each row with indices equal to those of edge $e_{i1} \to e_{i2} \in E'$, $(e_{i1}e_{i2})$ or $(e_{i1}\dots e_{i1}e_{i2})$ will be matched with the self-loop corresponding to the child, e_{i2}, when paired with a self-loop column because $e_{11} \to e_{11}$ is the only column not in the matrix.

If there are multiple sources, then for all descendants of the first source e_{11}, the rows corresponding to the edge $e_{i1} \to e_{i2} \in E'$, $(e_{i1}e_{i2})$ or $(e_{i1}\dots e_{i1}e_{i2})$ will again always be matched with the child when matched with a self-loop column. However, for the nodes which are not descendants of the first source (so descendants of a different source and not also a descendant of the first source) the rows corresponding to the edge $e_{i1} \to e_{i2} \in E'$, $(e_{i1}e_{i2})$ or $(e_{i1}\dots e_{i1}e_{i2})$ are matched with the self-loop of the parent e_{i1}, when matched up with a self-loop column.

The paths starting at different sources in the graph than e_{11} all have to converge with the path starting at e_{11} because G' is a tree. So the parent node is picked until you hit the spot where this path converges with the path which started at e_{11}. After the convergence point you then again always have to match a row with the child's self-loop as stated above. Such a convergence point must exist because E' is a tree. However, the main point is that once it is decided which rows will be mapped to the first $d-1$ columns by π everything else about π is decided by going down the remaining rows in order.

By the ordering of the rows and columns it further follows that π will be a product of 2-cycles. Since for each $i = 1, \dots, d-1$ either both

$$\pi(i) = i \qquad \text{and} \qquad \pi(i + (d-1)) = i + (d-1)$$

or they result in a two cycle where the two are flipped

$$\pi(i) = i + (d-1) \qquad \text{and} \qquad \pi(i + (d-1)) = i.$$

Thus, if k is the number of the first $(d-1)$ first rows that are mapped to the first $(d-1)$ columns, then π will be a product of $(d-1)-k$ two-cycles. Since we do not care about the sign of the determinant, the main point is that the permutations with even k will always have the same sign and all the permutations with odd k will always have the same sign, since a 2-cycle has sign -1.

One problem remains before the formula for the lowest degree term of the determinant of A, up to its sign, can be written. In the equation (3.A.6) for $A_r(\mathcal{K})$ there can be a

factor $r-1$ in front of $\mathcal{K}_{i\dots ij}$. This will happen exactly when in row $(e_{i1}\dots e_{i1}e_{i2})$ we pick the self-loop corresponding to the parent, i.e., $e_{i1}\to e_{i1}$. Let $q=0,\dots,d-2$ denote the number of non-self-loop edges on the paths starting at a different source node than the first source node, e_{11}, until the convergence points with the path starting at e_{11}. Then for each permutation π such that we are in the first case above where (ij) (so $A_2(\Sigma)$) contributes a diagonal entry and i or j is not a descendant of the first source node (so it is a descendant of another source before the convergence point) there will be a factor $r-1$ contributed by $A_r(\mathcal{K})$ from the $(i\dots ij)$ row. So q is the largest possible number of rows, which can contribute a $(r-1)$ factor.

Thus, we can obtain the following expression for the lowest degree term using Leibniz rule by going through the permutations by how many indices i where the first $(d-1)$ rows are mapped to the first $(d-1)$ columns. In other words, how many of the $A_2(\Sigma)$ rows are picked to contribute the diagonal entries. Thus, the sum over permutations can be written as a sum over i:

$$\text{low deg term}(\det(A_{(\mathbf{R}_1\cup\mathbf{R}_2),(\mathbf{C}_1\cup\mathbf{C}_2)})) = \zeta^{2(d-1)}\zeta^{(d-1)}\sum_{i=0}^{d-1}\left(\frac{r}{2}\right)^{i}\left(\left(\frac{r}{2}\right)^{2}\right)^{d-1-i}(-1)^i c(d,q,i),$$

where

$$c(d,q,i)=\sum_{j=0}^{i}(r-1)^j\binom{q}{j}\binom{d-1-q}{i-j}.$$

For a given number i of $A_2(\Sigma)$-rows picked for diagonal entries, the sum $c(d,q,i)$ counts the added contribution by potentially having to pick a factor $(r-1)$, depending on the graph, in the r-th-order rows. The summation index j counts the number of columns picked where an $(r-1)$ is included, and the binomial coefficient product counts the number of ways, potentially zero, that the columns can be picked such that exactly j columns with a factor $(r-1)$ are picked. By Lemma 3.A.5

$$\begin{aligned}\text{low deg term}(\det(A_{(\mathbf{R}_1\cup\mathbf{R}_2),(\mathbf{C}_1\cup\mathbf{C}_2)})) &= \zeta^{2(d-1)}\zeta^{(d-1)}\left(\frac{r}{2}\right)^{d-1}\sum_{i=0}^{d-1}\left(\frac{r}{2}\right)^{d-1-i}(-1)^i c(d,q,i)\\ &= \zeta^{2(d-1)}\zeta^{(d-1)}\left(\frac{r}{2}\right)^{d-1}\left(\frac{r}{2}-1\right)^{d-1}(-1)^q.\end{aligned}$$

Thus, the coefficient in front of the smallest degree term in the determinant of all of A is not zero, it is

$$\left(\frac{r}{2}\right)^{d-1}\left(\frac{r}{2}-1\right)^{d-1}(-1)^q\left(\frac{r}{2}\right)^{d(d-1)/2-(d-1)}.$$

The determinant of A is therefore not the zero polynomial in ζ, and there will exist a ζ-value such that the corresponding M, $\mathcal{C}_2$ and $\mathcal{C}_r$ result in the determinant of A being non-zero. □

Remark 3.A.6. In the proof we show that the rank of A is d^2-1 by removing the column corresponding to the self-loop of the first node in the topological ordering and showing

this matrix has rank $d^2 - 1$. However, the proof will work by removing any of the self-loop columns. The only thing this will change is how to determine q, which counts the columns where it is possible to obtain a factor $(r-1)$.

Remark 3.A.7. From the matrix A we could also remove columns corresponding to the zero entries of M and only solve for the non-zero entries. This would not change the proof in any substantial way. It would only make the set of rows R_3 and R_4, and corresponding column sets $\mathbf{C_3}$ and $\mathbf{C_4}$, smaller, and their fixed contribution to the determinant later in the proof would be smaller as well. But no arguments change or depend on how many of the rows or columns in these sets are included as long as the only deleted rows and columns correspond to zero entries in M given by the graph.

Proof of Corollary 3.3.5. Let $S_1, \dots, S_m$ be the connected components of G. Let $\mathbf{C}_{S_i}$ and $\mathbf{R}_{S_i}$ denote all columns and rows, respectively where all indices belong to the connected component S_i. Let $\mathbf{C}_{\text{mix}}$ and $\mathbf{R}_{\text{mix}}$ denote the rows and columns with indices in more than one component. By definition of A_2 and A_r, see equations (3.3.10) and (3.3.11), we see that A_{off} exhibits the following block structure if we organize the rows and columns by connected components

$$A_{\text{off}} = \begin{array}{c} \\ \mathbf{R}_{S_1} \\ \vdots \\ \mathbf{R}_{S_m} \\ \mathbf{R}_{\text{mix}} \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{C}_{S_1} & \cdots & \mathbf{C}_{S_m} & \mathbf{C}_{\text{mix}} \end{array} \\ \begin{pmatrix} A_{\text{off}}^{S_1} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & A_{\text{off}}^{S_m} & 0 \\ 0 & \cdots & 0 & A_{\text{off}}^{\text{mix}} \end{pmatrix} \end{array}.$$

The zero-pattern is explained as follows: if we consider $\alpha \to \beta \in S_i$ then $A_2(\Sigma)_{(kj),\alpha\to\beta} = 0$ if $k \notin S_i$ or $j \notin S_i$. Similarly, for the r-th-order Lyapunov equation rows. Explicitly, by equation (3.3.11) we have that $A_r(\mathcal{K})_{i_1,\dots,i_r,\alpha\to\beta}$ is given as some integer multiple of $\mathcal{K}_{\alpha i_2,\dots,i_r}$ given that $\beta = i_1$, and similarly if β was equal to one of the other indices. Thus, an entry can only be non-zero if β overlaps with $(i_1, \dots i_r)$, and there is a trek between α and the remaining i's, so α and the remaining i's need to belong to the same connected component.

By the block diagonal structure

$$\operatorname{rank}(A_{\text{off}}) = \sum_{i=1}^{m} \operatorname{rank}\left(A_{\text{off}}^{S_i}\right) + \operatorname{rank}\left(A_{\text{off}}^{\text{mix}}\right). \tag{3.A.9}$$

Each block, $A_{\text{off}}^{S_i}$ can have at most rank $|S_i| - 1$ showing the first part of the Corollary, $\operatorname{rank}(A_{\text{off}}) \leq d - m$.

By Theorem 3.3.1, $\operatorname{rank}(A_{\text{off}}^{S_i}) = |S_i| - 1$ generically. This would be enough for the m-dimensional identifiability of the graph with m connected components. However, it is not enough in terms of understanding the exceptional set for the connected graphs.

To show that $A_{\text{off}}^{\text{mix}}$ has full rank we notice that the square submatrix only indexed by the r-th-order off-diagonal rows of the form $(i \dots ij)$ and all columns in $\mathbf{C}_{\text{mix}}$ has full rank. This is seen by noting that this matrix exhibits a block diagonal structure.

For each connected component S_i and each $s \in S_i$ group the columns $t \to s$ and rows $(t \dots ts)$ for $t \in [d] \setminus S_i$. The argument for the zero-pattern is completely analogous to that for A_{off} by either using no overlap in the sink and indices of the rows or when there is a matching a lack of trek between the source and the remaining indices.

Let $s \in S_i$ for a given connected component and number the elements in $t_1, \dots t_l \in [d] \setminus S_i$, then the block on the diagonal corresponding to s has the form

$$(D_s)_{(t_i \dots t_i s),(t_j \to s)} = \mathcal{K}_{t_i \dots t_i t_j}.$$

Thus, if we pick M to be diagonal, each D_s will become diagonal with $\mathcal{K}_{t_1 \dots t_1}, \dots, \mathcal{K}_{t_l \dots t_l}$ showing that each D_s generically has full rank. Thus, the entire square matrix generically has full rank, as does $A_{\text{off}}^{\text{mix}}$. Combing this we obtain by equation (3.A.9)

$$\operatorname{rank}(A_{\text{off}}) = d^2 - m,$$

so the rank drops by exactly the number of connected components generically. □

3.B. Appendix – additional proofs and supplementary material

3.B.1 Proofs about the estimator

Proof of Theorem 3.4.2. By the law of large numbers, the empirical moments up to order k converge in probability to the true moments. Since the cumulants can be obtained by a continuous, and smooth, transformation of the moments, see, e.g., Section 2.4.3 by McCullagh [2018] for an explicit formula we can apply the continuous mapping theorem, Theorem 2.3 by Vaart [1998], to obtain that the empirical cumulants also converge in probability to the true cumulants. As the matrix A is assumed to have entries equal to cumulants, up to scaling by a real number, it follows by another use of the continuous mapping theorem that

$$\hat{A}_n \xrightarrow{P} A \text{ for } n \to \infty.$$

By definition, $\nu_{\min}(\hat{M})$ can be found as the unit eigenvector of $\hat{A}^T\hat{A}$ corresponding to the smallest eigenvalue of $\hat{A}^T\hat{A}$ picked such that $\operatorname{tr}(\hat{M}) < 0$. Let G denote this function. Then G is continuous and smooth in an open neighborhood of the true A. This holds since G is the composition of two maps, which are smooth in a neighborhood of A. Clearly, the map $X \mapsto X^TX$ is smooth. The unit eigenvector map for a simple eigenvalue of a real symmetric matrix, X_0, is also smooth in a neighborhood of X_0, see Theorem 1 by Magnus [1985]. To apply this result for A^TA we note that we have assumed $\ker(A) = \operatorname{span}(\operatorname{vec}(M))$, so the smallest eigenvalue of A^TA is 0, and since the kernel is one-dimensional it is a simple eigenvalue. Thus, we can apply the continuous mapping theorem to conclude that

$$\hat{M}_n \xrightarrow{P} M \text{ for } n \to \infty.$$

If the distribution, additionally, has finite $2k$-th-order moment, asymptotic normality follow by an application of the delta method. By the central limit theorem, the empirical

moments up to order k will be asymptotically normally distributed. Thus, since the cumulants can be obtained by a differentiable function of the moments it follows by the delta method that the empirical cumulants are also asymptotically normally distributed, as written in equation (3.4.22). As $A = \mathcal{A}\kappa$ for $\mathcal{A}$ a fixed matrix, we obtain

$$\sqrt{n}(\text{vec}(\hat{A}_n - A)) \xrightarrow{D} \mathcal{N}\left(0, \mathcal{A}\Omega\mathcal{A}^T\right)$$

for $n \to \infty$ by use of the delta method.

As argued above M and $\hat{M}$ are obtained by applying the function G to A and $\hat{A}$, respectively, and since G is smooth in a neighborhood of the true A, it follows by the delta method that

$$\sqrt{n}(\text{vec}(\hat{M}_n - M)) \xrightarrow{D} \mathcal{N}\left(0, DG(A)\mathcal{A}\Omega\mathcal{A}^T(DG(A))^T\right)$$

for $n \to \infty$, where $DG(A)$ denotes the Jacobian evaluated in A. By Lemma 3.B.1, we have

$$DG(A) = -\text{vec}(M)^T \otimes A^+,$$

and we obtain the asymptotic covariance matrix specified in the theorem. □

We derive the Jacobian of a vectorized version of a map of matrices. This is done using notation and conventions from Chapter 6 by Magnus and Neudecker [2019].

Lemma 3.B.1. *Let $G : \mathbb{R}^{b\times d^2} \to \mathbb{R}^{d^2}$ denote the function mapping a matrix X to $v_{\min}(X)$, the right unit singular vector of the smallest singular value of X. Consider a matrix $A \in \mathbb{R}^{b\times d^2}$ whose smallest singular value is simple and equal to zero with right singular vector denoted $\text{vec}(M)$. Then in a neighborhood of A, G is a smooth function with Jacobian evaluated in A given by*

$$DG(A) = \frac{\partial \text{vec}(G)}{\partial \text{vec}(X)^T}\bigg|_{X=A} = -\text{vec}(M)^T \otimes A^+, \tag{3.B.1}$$

where A^+ denotes the Moore-Penrose inverse.

Proof. We first note that the smallest singular value and its corresponding right singular vector of a matrix X can be found as the corresponding smallest eigenvalue and eigenvector of X^TX. Thus, G can be viewed as the composition of maps $\text{eig}_{\min} \circ F$, where F maps a matrix X to X^TX and $\text{eig}_{\min}$ maps a matrix to the unit eigenvector of its smallest eigenvalue. We compute the Jacobian of each of these maps. Theorem 1 by Magnus [1985] gives that the Jacobian of $\text{eig}_{\min}$ at A^TA is given as

$$D\text{eig}_{\min}(A^TA) = \frac{\partial \text{eig}_{\min}}{\partial(\text{vec}(X))^T}\bigg|_{X=A^TA} = -\text{vec}(M)^T \otimes (A^TA)^+,$$

since the smallest eigenvalue is zero with eigenvector $\text{vec}(M)$.

By letting $\varepsilon > 0$ and Y any matrix of the same dimension as X we can write

$$\begin{aligned}\mathrm{vec}(F(X+\varepsilon Y)) &= \mathrm{vec}\left((X+\varepsilon Y)^T(X+\varepsilon Y)\right)\\ &= \mathrm{vec}(X^TX) + \varepsilon\mathrm{vec}\left(Y^TX + X^TY\right) + \varepsilon^2\mathrm{vec}(Y^TY)\\ &= \mathrm{vec}(X^TX) + \varepsilon((X^T\otimes I)K + I\otimes X^T)\mathrm{vec}(Y) + \varepsilon^2\mathrm{vec}(Y^TY),\end{aligned}$$

by using rules of vectorization and Kronecker products, see Chapter 2 by Magnus and Neudecker [2019], and where K denotes the commutation matrix that switches the vectorization of a matrix with the vectorization of its transpose. From this we can conclude that

$$DF(A) = \frac{\partial F}{\partial(\mathrm{vec}(X))^T}\bigg|_{X=A} = (A^T\otimes I)K + I\otimes A^T.$$

Thus, by the chain rule we obtain the following expression for the Jacobian of G at A

$$\begin{aligned}DG(A) = \frac{\partial\mathrm{vec}(G)}{\partial\mathrm{vec}(X)^T}\bigg|_{X=A} &= D\mathrm{eig}_{\min}(A^TA)DF(A)\\ &= -\mathrm{vec}(M)^T\otimes(A^TA)^+((A^T\otimes I)K + I\otimes A^T)\\ &= -(\mathrm{vec}(M)^TA^T\otimes(A^TA)^+)K - \mathrm{vec}(M)^T\otimes(A^TA)^+A^T\\ &= -\mathrm{vec}(M)^T\otimes(A^TA)^+A^T = -\mathrm{vec}(M)^T\otimes A^+,\end{aligned}$$

by using that $A\mathrm{vec}(M) = 0$. □

Proof of Corollary 3.4.3. We rewrite

$$\left(\mathrm{vec}(M)^T\otimes A^+\right)\mathcal{A}$$

in a way to avoid forming the large matrices $\mathrm{vec}(M)^T\otimes A^+$ and $\mathcal{A}$. The columns of $\mathcal{A}$ are indexed by the unique vectorizations of the second- and r-th-order cumulants $\Sigma_{11},\ldots,\Sigma_{dd},\mathcal{K}_{1\cdots1},\ldots,\mathcal{K}_{d\cdots d}$. We let the second- and r-th-order cumulants jointly be given by κ and let j index this set. If we consider column j of $\mathcal{A}$, $\mathcal{A}_{*,j}$ as being written as the vectorization of a $D\times d^2$ matrix. Then this matrix can be seen to be A evaluated at the point where every cumulant except j is equal to zero and $j=1$. If we let $\kappa_j = 1$ denote the vector of second- and r-th-order cumulants where the j-th is set to one and the remaining to zero, we obtain that

$$\mathcal{A}_{*,j} = \mathrm{vec}(A(\kappa_j = 1)).$$

Then

$$\left(\mathrm{vec}(M)^T\otimes A^+\right)\mathcal{A}_{*,j} = A^+A(\kappa_j=1)\mathrm{vec}(M),$$

and therefore,

$$\left(\mathrm{vec}(M)^T\otimes A^+\right)\mathcal{A} = A^+\left[A(\kappa_1=1)\mathrm{vec}(M)\cdots A(\kappa_q=1)\mathrm{vec}(M)\right].$$

Each column $A(\kappa_j = 1)\text{vec}(M)$ is by definition of A the off-diagonal Lyapunov equation, of order 2 and r combined, evaluated at M, and the cumulant κ_j for which the j-th-order cumulant is one and the rest are zero. Thus, this matrix is just equal to the block diagonal matrix

$$[A(\kappa_1 = 1)\text{vec}(M) \cdots A(\kappa_q = 1)\text{vec}(M)] = B_{2,r}(M) = \begin{pmatrix} B_2(M) & 0 \\ 0 & B_r(M) \end{pmatrix},$$

with $B_k(M)$ defined by equation (3.A.3), so we obtain that

$$\left(\text{vec}(M)^T \otimes A^+\right) \mathcal{A} = A^+ B_{2,r}(M),$$

which finishes the proof. □

3.B.2 General Lyapunov models and trek formulas

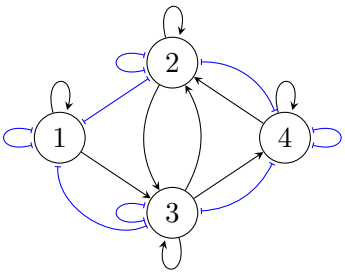
For our main results we assume that the continuous Lyapunov model has diagonal cumulants of the driving Lévy process. To encode a more general sparsity pattern of the cumulant we introduce graphs with blunt edges, as was done in covariance case by Varando and Hansen [2020]. In this appendix we will provide this definition for any order cumulant of a continuous Lyapunov model and derive general trek formulas in this setup.

Definition 3.B.2. *Let $G = ([d], E, B)$ be a k-th-order mixed graph with E denoting the set of directed edges ($\to$) and B denoting the k-fold blunt edges ($\mapsto$). Let*

$$\mathbb{R}^E = \{M \in \mathbb{R}^{d\times d} \mid M_{ij} = 0 \text{ if } j \to i \notin E\}$$
$$\text{Sym}^k(\mathbb{R}^d)^B = \{C \in \text{Sym}^k(\mathbb{R}^d) \mid C_{i_1,\dots,i_k} = 0 \text{ if } (i,\dots,i_k) \notin B\}.$$

If $(M, \mathcal{C}_k) \in \mathbb{R}^{d\times d} \times \text{Sym}^k(\mathbb{R}^d)$ satisfy that $M \in \mathbb{R}^E$ and $\mathcal{C}_k \in \text{Sym}^k(\mathbb{R}^d)^B$ we will say that they are compatible with the mixed graph G.

In Figure 3.B.1 there is an example of a second-order mixed graph and corresponding zero-pattern of $(M, \mathcal{C}_2)$. The directed edges are directed and drawn in black and the blunt edges are drawn in blue and undirected, so the order the nodes are written in is irrelevant.



$$M = \begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & * & * \\ * & * & * & 0 \\ 0 & 0 & * & * \end{pmatrix} \qquad \mathcal{C}_2 = \begin{pmatrix} * & * & * & 0 \\ * & * & 0 & * \\ * & 0 & * & * \\ 0 & * & * & * \end{pmatrix}$$

Figure 3.B.1: Example of a second-order mixed graph $G = ([4], E, B)$ and corresponding zero-pattern of pair $(M, \mathcal{C}_2)$ compatible with G.

Since the steady-state solution to equation (3.1.1) only exists if the drift matrix M is stable, we let $\mathbb{R}^E_{\text{stab}}$ denote the subset of $\mathbb{R}^E$ that consists of stable matrices as in the main paper. Furthermore, some cumulants, e.g., the covariance, satisfy positive definiteness, so when such further positivity constraints are applicable, this should also be included in the model definitions. As in the main paper, we will always assume that the directed part of the graph $G = ([d], E)$ contains all self-loops.

We can define the continuous Lyapunov model of a certain order k by the set of distributions in $\mathcal{P}_G$, where G represents only the directed part of the mixed graph, and where the Lévy process has finite k-th-order moment and the k-th-order cumulant of the Lévy process is compatible with the blunt edges in G, i.e., $\mathcal{C}_k \in \text{Sym}^k(\mathbb{R}^d)^B$. As in the main paper, we will also consider the finite-dimensional set of cumulants of order k, which can arise in this way, as solutions to the k-th-order continuous Lyapunov equation with $(M, \mathcal{C}_k)$ corresponding to a mixed graph.

Definition 3.B.3. *Let $G = ([d], E, B)$ be a k-th-order mixed graph on d nodes. Then the continuous k-th-order Lyapunov cumulant model is the set of k-th-order tensors in the set*

$$\mathcal{M}^k_G = \{\mathcal{K} \in \text{Sym}^k(\mathbb{R}^d) \mid \mathcal{K} \times_1 M + \cdots \mathcal{K} \times_k M + \mathcal{C}_k = 0 \text{ for } M \in \mathbb{R}^E_{\text{stab}} \text{ and } \mathcal{C}_k \in \text{Sym}^k(\mathbb{R}^d)^B\}.$$

In the following we will derive two different general trek rules for the k-th-order Lyapunov equation.

When considering the integral equation (3.A.4) for the solution to the k-th-order Lyapunov equation there are at least two different options of how to derive a trek formula from it. It depends upon whether you are willing to make an assumption on the spectral radius of M or not. The first option is to consider

$$\mathcal{K}(s) = \int_0^s \mathcal{C}_k \times_1 e^{Mt} \times_2 e^{Mt} \cdots \times_k e^{Mt} \mathrm{d}t, \tag{3.B.2}$$

which clearly satisfies $\lim_{s\to\infty} \mathcal{K}(s) = \mathcal{K}$, and then obtain a trek formula expression for $\mathcal{K}(s)$ as is done in the case $k = 2$ by Varando and Hansen [2020]. The other option requires an assumption which allows one to exchange integration and sums, as is done for $k = 2$ by Hansen [2025]. Before we proceed with either, we will first give a general definition of a k-trek.

Definition 3.B.4. *Let $G = ([d], E, B)$ be a k-th-order mixed graph. A k-trek τ between the nodes $i_1, \ldots, i_k$ consists of the following: a k-fold blunt edge $((j_1)_1, \ldots, (j_k)_1)$ in the graph G and k directed walks in G*

$$\begin{aligned} &\pi_1 : (j_1)_1 \to (j_1)_2 \cdots \to i_1 \\ &\vdots \\ &\pi_k : (j_n)_1 \to (j_n)_2 \cdots \to i_k. \end{aligned}$$

Let $\ell_{i_m}(\tau)$ denote the length of the walk going to i_m and $\mathcal{T}(i_1, \ldots, i_k)$ denote the set-up of all treks between the nodes $i_1, \ldots, i_k$.

Note that each of the walks is allowed to have length 0. If the blunt part of the graph only contains the self-loop blunt edges (corresponding to the $\mathcal{C}$ being diagonal) we can also define the trek from only a directed graph, because in this case the top nodes have to be the same in all the paths constituting the trek.

Example 3.B.5. We can compare the 2-treks in the graph in Figure 3.2.1 and Figure 3.B.1, where the directed part of the graph is the same, but the blunt edges are different. In Figure 3.2.1 there are not drawn, which corresponds there only being the self-loop blunt edges and $\mathcal{C}_2$ diagonal. We only consider the simple 2-treks meaning the ones without directed loops, since otherwise there are infinitely many but they all have to arise from adding loops to one of the simple treks. We provide two examples of pairs of nodes and the treks between them. If the top of the trek is a self-loop blunt edge $v \mapsto v$ we do by convention not write the blunt edge but just the given node in the middle of the trek as $\leftarrow v \rightarrow$.

Treks between 1 and 2: In Figure 3.2.1 there are none, in Figure 3.B.1 there are two simple treks

$$1 \mapsto 2, \;\; 1 \mapsto 3 \rightarrow 2.$$

Treks between 2 and 4: In Figure 3.2.1 we have

$$2 \leftarrow 4, \;\; 2 \rightarrow 3 \rightarrow 4, \;\; 2 \leftarrow 3 \rightarrow 4, \;\; 2 \leftarrow 3 \leftarrow 1 \rightarrow 3 \rightarrow 4.$$

In Figure 3.B.1 there are all the same as in Figure 3.2.1 as well as all the following

$$\begin{aligned}
&2 \mapsto 4, \\
&2 \mapsto 1 \rightarrow 3 \rightarrow 4, 2 \leftarrow 3 \leftarrow 1 \mapsto 2 \rightarrow 3 \rightarrow 4, \\
&2 \leftarrow 3 \mapsto 1 \rightarrow 3 \rightarrow 4, 2 \leftarrow 4 \leftarrow 3 \leftarrow 1 \mapsto 3 \rightarrow 4, 2 \leftarrow 3 \mapsto 1 \rightarrow 3 \rightarrow 4, \\
&2 \leftarrow 4 \leftarrow 3 \mapsto 1 \rightarrow 3 \rightarrow 4 \\
&2 \leftarrow 3 \mapsto 4, 2 \leftarrow 4 \leftarrow 3 \mapsto 4, 2 \leftarrow 4 \mapsto 3 \rightarrow 4.
\end{aligned}$$

In order to prove the trek formulas, we provide the definition of a trek polynomial.

Definition 3.B.6. *Let $G = ([d], E, B)$ be a k-th-order mixed graph and let $(M, \mathcal{C})$ be a pair of a $d \times d$ stable matrix and k-th-order tensor that is compatible with G then for τ a k-trek with the blunt edge $((j_1)_1, \ldots, (j_n)_1)$ as its center, we define*

$$\omega(M, C, \tau) = \mathcal{C}_{(j_1)_1, \ldots, (j_k)_1} \prod_{m=1}^{k} \prod_{\alpha \rightarrow \beta \in \pi_m} M_{\beta\alpha}.$$

This is called the trek polynomial corresponding to the trek τ.

We will prove two different trek rules. In both cases, we write the proof for $k = 3$ for notational convenience, it is easy to see from the proof how it generalizes to general k and we will write this result as well.

Proposition 3.B.7. *Let $G = ([d], E, B)$ be a third-order mixed graph and let $(M, \mathcal{C})$ be a pair of a $d \times d$ stable matrix and $\mathcal{C}$ third-order tensor that is compatible with G then for $\mathcal{K}(s)$ as defined by equation (3.B.2) (for $k = 3$) we obtain*

$$(\mathcal{K}(s))_{i_1 i_2 i_3} = \sum_{\tau \in \mathcal{T}(i_1,i_2,i_3)} \frac{s^{\ell_{i_1}(\tau)+\ell_{i_2}(\tau)+\ell_{i_3}(\tau)+1}}{\ell_{i_1}(\tau)!\ell_{i_2}(\tau)!\ell_{i_3}(\tau)!(\ell_{i_1}(\tau)+\ell_{i_2}(\tau)+\ell_{i_3}(\tau)+1)}\omega(M, \mathcal{C}, \tau).$$

Proof. First we consider the general expression for $\mathcal{K}(s)$. We obtain the following by using the series expansion for the matrix exponential and using that we can exchange the sum and the integral since M is stable so the infinite sum is convergent.

$$\begin{aligned}
\mathcal{K}(s) &= \int_0^s \mathcal{C} \times_1 e^{Mt} \times_2 e^{Mt} \times_3 e^{Mt} \mathrm{d}t \\
&= \int_0^s \mathcal{C} \times_1 \sum_{n=0}^{\infty} \frac{t^n}{n!} M^n \times_2 \sum_{m=0}^{\infty} \frac{t^m}{m!} M^m \times_3 \sum_{k=0}^{\infty} \frac{t^k}{k!} M^k \mathrm{d}t \\
&= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{k=0}^{\infty} \int_0^s \mathcal{C} \times_1 \frac{t^n}{n!} M^n \times_2 \frac{t^m}{m!} M^m \times_3 \frac{t^k}{k!} M^k \mathrm{d}t \\
&= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{k=0}^{\infty} \int_0^s \frac{t^{n+m+k}}{n!m!k!} \mathcal{C} \times_1 M^n \times_2 M^m \times_3 M^k \mathrm{d}t \\
&= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{k=0}^{\infty} \frac{s^{n+m+k+1}}{n!m!k!(n+m+k+1)} \mathcal{C} \times_1 M^n \times_2 M^m \times_3 M^k
\end{aligned}$$

To derive an expression for the $(i_1 i_2 i_3)$-th entry we need an expression for the $(i_1 i_2 i_3)$-th entry of $\mathcal{C} \times_1 M^n \times_2 M^m \times_3 M^k$. By definition

$$(\mathcal{C} \times_1 M^n \times_2 M^m \times_3 M^k)_{i_1 i_2 i_3} = \sum_{\alpha=1}^{d}\sum_{\beta=1}^{d}\sum_{\gamma=1}^{d} \mathcal{C}_{\alpha\beta\gamma} \cdot (M^n)_{i_1\alpha} \cdot (M^m)_{i_2\beta} \cdot (M^k)_{i_3\gamma}.$$

Since M is an adjacency matrix $(M^n)_{i_1\alpha}$ is the sum of all walk polynomials of directed walks in G from α to i_1 of length n. Thus, the sum above will be the sum of the trek polynomials over all treks between i_1, i_2 and i_3 where the three walks going to i_1, i_2 and i_3 have lengths n, m and k, respectively. Combining this with the expression above where we sum over all n, m and k we obtain

$$(\mathcal{K}(s))_{i_1 i_2 i_3} = \sum_{\tau \in \mathcal{T}(i_1,i_2,i_3)} \frac{s^{\ell_{i_1}(\tau)+\ell_{i_2}(\tau)+\ell_{i_3}(\tau)+1}}{\ell_{i_1}(\tau)!\ell_{i_2}(\tau)!\ell_{i_3}(\tau)!(\ell_{i_1}(\tau)+\ell_{i_2}(\tau)+\ell_{i_3}(\tau)+1)}\omega(M, \mathcal{C}, \tau),$$

as we wanted to prove. □

Proposition 3.B.8. *Let $G = ([d], E, B)$ be a k-th-order mixed graph and let $(M, \mathcal{C})$ be a pair of a $d \times d$ stable matrix and $\mathcal{C}$ k-th-order tensor that is compatible with G then*

for $\mathcal{K}(s)$ as defined by equation (3.B.2) we obtain

$$(\mathcal{K}(s))_{i_1\dots i_k} = \sum_{\tau\in\mathcal{T}(i_1,i_2,\dots,i_n)} \frac{s^{\sum_{j=1}^k \ell_{i_j}(\tau)+1}}{\prod_{j=1}^k \ell_{i_j}(\tau)!\left(\sum_{j=1}^k \ell_{i_j}(\tau)+1\right)}\omega(M,\mathcal{C},\tau).$$

Remark 3.B.9. The primary usefulness of this result is that the k-th-order cumulant $(\mathcal{K}(s))_{i_1\dots i_k} = 0$ if there is no k-trek between $i_1,\dots,i_k$ and by taking the limit this also applies to $\mathcal{K}$.

If we make additional assumptions on M we will be allowed to interchange the sum and limit operation to obtain a trek formula for the cumulant directly as is done for $k = 2$ by Hansen [2025]. For any M we can decompose it by subtracting a scaled version of the identity matrix, $I_{d\times d}$,

$$M = \Lambda - s\cdot I_{d\times d}. \tag{3.B.3}$$

One way to choose s would be to choose it such that $s\cdot I_{d\times d}$ equals one or more of the diagonal elements of M. Then there is one less non-zero entry in Λ than in M, which might make the trek formulas simpler, because there will be fewer treks. Another choice would be $s = 1$. Then as long as $M_{ii} \neq -1$, Λ and M would have the same zero-patterns and therefore the same treks. We will let $\omega(\Lambda,\mathcal{C},\tau)$ denote the trek polynomial for Λ independently of whether it will be the same as for M. In the case of $s = 1$, this yields the following trek rule.

Proposition 3.B.10. *Let $G = ([d],E,B)$ be a third-order mixed graph and let $(M,\mathcal{C})$ be a pair of a $d\times d$ stable matrix and a k-th-order tensor that is compatible with G, and let Λ be defined by equation (3.B.3). If Λ has spectral radius strictly less than 1, then for $\mathcal{K}\in\mathcal{M}_G^3$*

$$(\mathcal{K})_{i_1i_2i_3} = \sum_{\tau\in\mathcal{T}(i_1,i_2,i_3)} \omega(\Lambda,\mathcal{C},\tau)\frac{3^{-\ell_{i_1}(\tau)-\ell_{i_2}(\tau)-\ell_{i_3}(\tau)-1}\Gamma(\ell_{i_1}(\tau)+\ell_{i_2}(\tau)+\ell_{i_3}(\tau)+1)}{\ell_{i_1}(\tau)!\ell_{i_2}(\tau)!\ell_{i_3}(\tau)!}.$$

Proof. We obtain the following using the definition of Λ, the matrix exponential and the gamma distribution, and that we can interchange the sum and integral because of the

spectral radius assumption.

$$\begin{aligned}
\mathcal{K} &= \int_0^\infty \mathcal{C} \times_1 e^{Mt} \times_2 e^{Mt} \times_3 e^{Mt} \mathrm{d}t \\
&= \int_0^\infty \mathcal{C} \times_1 e^{(\Lambda - I_{d\times d})t} \times_2 e^{(\Lambda - I_{d\times d})t} \times_3 e^{(\Lambda - I_{d\times d})t} \mathrm{d}t \\
&= \int_0^\infty e^{-3t} \cdot \mathcal{C} \times_1 e^{\Lambda t} \times_2 e^{\Lambda t} \times_3 e^{\Lambda t} \mathrm{d}t \\
&= \int_0^\infty e^{-3t} \cdot \mathcal{C} \times_1 \sum_{n=0}^\infty \frac{t^n}{n!}\Lambda^n \times_2 \sum_{m=0}^\infty \frac{t^m}{m!}\Lambda^m \times_3 \sum_{k=0}^\infty \frac{t^k}{k!}\Lambda^k \mathrm{d}t \\
&= \sum_{n=0}^\infty \sum_{m=0}^\infty \sum_{k=0}^\infty \int_0^\infty \frac{t^{n+m+k}e^{-3t}}{n!m!k!} \mathcal{C} \times_1 \Lambda^n \times_2 \Lambda^m \times_3 \Lambda^k \mathrm{d}t \\
&= \sum_{n=0}^\infty \sum_{m=0}^\infty \sum_{k=0}^\infty \frac{3^{-n-m-k-1}\Gamma(n+m+k+1)}{n!m!k!} \mathcal{C} \times_1 \Lambda^n \times_2 \Lambda^m \times_3 \Lambda^k.
\end{aligned}$$

We take the $(i_1 i_2 i_3)$-th entry and use the same derivation as in the proof of Proposition 3.B.7 to obtain an expression for $\mathcal{C} \times_1 \Lambda^n \times_2 \Lambda^m \times_3 \Lambda^k$ by 3-treks to obtain that

$$(\mathcal{K})_{i_1 i_2 i_3} = \sum_{\tau \in \mathcal{T}(i_1, i_2, i_3)} \omega(\Lambda, \mathcal{C}, \tau) \frac{3^{-\ell_{i_1}(\tau) - \ell_{i_2}(\tau) - \ell_{i_3}(\tau) - 1} \Gamma(\ell_{i_1}(\tau) + \ell_{i_2}(\tau) + \ell_{i_3}(\tau) + 1)}{\ell_{i_1}(\tau)! \ell_{i_2}(\tau)! \ell_{i_3}(\tau)!},$$

as we wanted to prove. □

The above calculation can easily be carried out for a general decomposition of M as given in equation (3.B.3) to obtain a trek rule that might be more useful in specific cases. An example of this is Lemma 3.A.3. The trickiest part in terms of using this trek rule is to ensure that Λ has spectral radius strictly less than 1, which is something that can be more easily ensured when M is lower-triangular, i.e., the graph is a DAG with self-loops, as in Lemma 3.A.3. Below is the generalization for higher-order k.

Proposition 3.B.11. *Let $G = ([d], E, B)$ be a k-th-order mixed graph and let (M, C) be a pair of a $d \times d$ stable matrix and a k-th-order tensor that is compatible with G, and let Λ be defined by equation (3.B.3). If Λ has spectral radius strictly less than 1, then for $\mathcal{K} \in \mathcal{M}_G^k$*

$$\mathcal{K}_{i_1 \dots i_k} = \sum_{\tau \in \mathcal{T}(i_1, i_2, \dots, i_k)} \omega(\Lambda, \mathcal{C}, \tau) \frac{k^{-\sum_{j=1}^k \ell_{i_j}(\tau) - 1} \Gamma\left(\sum_{j=1}^k \ell_{i_j}(\tau) + 1\right)}{\prod_{j=1}^k \ell_{i_j}(\tau)!}.$$

3.B.3 Additional identifiability results

In this section we provide an additional identifiability result for continuous Lyapunov models. Contrary to Theorem 3.3.1, we assume that a given cumulant of order at least three of the Lévy process is known, as explored by Dettling et al. [2023] in the covariance case. In this case we show generic identifiability of M for any graph with all self-loops.

Proposition 3.B.12. *Let $G = ([d], E)$ be any directed graph with all self-loops and $r \geq 3$ an integer then the r-th-order continuous Lyapunov model $\mathcal{M}_G^r$ where $\mathcal{C}_r$ is assumed known and diagonal with non-zero diagonal is generically identifiable.*

Proof. To show generic identifiability, we exhibit a choice of M in $\mathbb{R}^E_{\text{stab}}$ such that a suitable submatrix of $A_r(\mathcal{K})$ as defined by equation (3.3.8) has rank $|E|$, so that we can solve the system in equation (3.3.8) for the non-zero entries of M.

We let M be any stable diagonal matrix (so it has negative values on the diagonal). This choice of M will be in $\mathbb{R}^E_{\text{stab}}$ for any directed graph with all self-loops, which we have assumed the graph to have. When M is diagonal (with all diagonal elements non-zero) it follows by the general trek-rule that exactly $\mathcal{K}_{i\dots i} \neq 0$ for all i but all other entries of $\mathcal{K}$ are zero since $\mathcal{C}_r$ is assumed diagonal.

Consider the submatrix A of $A_r(\mathcal{K})$ with the columns indexed by $i \to j \in E$ and rows indexed by $(i \dots ij)$ for $i \to j \in E$. This especially includes the diagonal rows $(i \dots i)$ for all $i \in [d]$).

By equation (3.A.6) we see that see that $A_r(\mathcal{K})_{(i\dots ij),(i\to j)} = \mathcal{K}_{i\dots i}$ for $i \neq j$ and $i \to j \in E$ and that $A_r(\mathcal{K})_{(i\dots ij),(\alpha\to\beta)}$ is never equal to a diagonal entry of $\mathcal{K}$ if $(\alpha \to \beta) \neq (i \to j)$ by equation (3.A.6).

Furthermore, $A_r(\mathcal{K})_{(i\dots i),i\to i} = r\mathcal{K}_{i\dots i}$ and for any other column $A_r(\mathcal{K})_{(i\dots i),(\alpha\to\beta)}$ is never equal to a diagonal entry of $\mathcal{K}$.

Thus, we conclude that the submatrix A has diagonal entries of $\mathcal{K}$ (upto multiplication by r) on its diagonal and all off-diagonal entries equal to off-diagonal entries of $\mathcal{K}$. Thus, by picking M to be diagonal with negative diagonal the determinant of A is just a product of diagonal entries of $\mathcal{K}$

$$\det(A) = \prod_{i\to j\in E, i\neq j} \mathcal{K}_{i\dots i} \prod_{i\in[d]} r\mathcal{K}_{i\dots i},$$

which are all non-zero by choice of M and the diagonal of $\mathcal{C}_r$ being non-zero . Therefore, $\det(A) \neq 0$ for this choice of M, which completes the proof. □

3.B.4 Auxiliary results when the Lévy process is a compound Poisson process

Since Z_1 is infinitely divisible, its characteristic function is of the form $\mathbb{E}(e^{iz^T Z_1}) = e^{\Psi(z)}$ where Ψ is the Lévy exponent. Its k-th-order cumulants can then be expressed as

$$\mathcal{C}_k = (-i)^n D^k \Psi(0) \tag{3.B.4}$$

provided that Ψ is k times differentiable in 0. If the Lévy process is a compound Poisson process, its Lévy exponent is of the form

$$\Psi(z) = \boldsymbol{\lambda} \int e^{iz^T u} P(\mathrm{d}u),$$

where $\boldsymbol{\lambda} > 0$ is the rate parameter of the homogeneous Poisson process, and the probability measure P is the jump distribution. Using (3.B.4) we find that for a compound Poisson process, the k-th-order cumulant for $k \geq 2$ equals

$$\mathcal{C}_k = \boldsymbol{\lambda} \int \underbrace{u \otimes u \otimes \ldots \otimes u}_{k \text{ factors}} P(\mathrm{d}u). \tag{3.B.5}$$

Suppose next that the coordinates of Z are, in fact, independent univariate compound Poisson processes. The i-th coordinate, denoted $Z^{(i)}$, has Lévy measure $\lambda_i P_i$, where $\lambda_i > 0$ and P_i is the univariate probability distribution of the jumps of $Z^{(i)}$. Then Z is a compound Poisson process with $\boldsymbol{\lambda} = \lambda_1 + \ldots + \lambda_p$ and

$$P = \sum_{i=1}^{d} \frac{\lambda_i}{\boldsymbol{\lambda}} \tilde{P}_i,$$

where the probability measure $\tilde{P}_i$ on $\mathbb{R}^d$ is given by

$$\int f(u) \tilde{P}_i(\mathrm{d}u) = \int f(0, \ldots, 0, u_i, 0, \ldots, 0) P_i(\mathrm{d}u_i).$$

From this it follows that $\mathcal{C}_k$ is diagonal with

$$\operatorname{diag}(\mathcal{C}_k) = \left(\lambda_1 \int u^k \; P_1(\mathrm{d}u), \ldots, \lambda_p \int u^k \; P_p(\mathrm{d}u) \right). \tag{3.B.6}$$

Note that for a compound Poisson process with independent coordinates, it is with probability one always only one coordinate that jumps at a time.

We can use specific compound Poisson processes to prove the following about the surjectivity of $\pi_{2,r}$.

Proposition 3.B.13. *Let r be an integer such that $r \geq 3$.*

(i) If r is odd the map $\pi_{2,r}$ is surjective.

(ii) if r is even the map $\pi_{2,r}$ is not surjective but does map onto a fully dimensional subset of Θ_G.

Proof. The proof is constructive and only focuses on the part of the map $Q_1 \mapsto (\mathcal{C}_2, \mathcal{C}_r)$, since $\pi_{2,r}$ is the identity on the M coordinate.

Furthermore, since $\mathcal{C}_2$ and $\mathcal{C}_r$) are assumed diagonal, we can construct the example by considering a Lévy process, and corresponding infinitely divisible distribution, with independent coordinates. For this reason, it is enough to construct a univariate Lévy process with a given covariance c_2 and r-th-order cumulant c_r.

We let the Lévy process be a compound Poisson process. By equation (3.B.6) the cumulants of the compound Poisson can be found as the rate parameter multiplied by the raw moments of the jump distribution, P,

$$\mathcal{C}_r = \lambda \int u^k \; P(\mathrm{d}u).$$

Thus, (i) and (ii) follow by (i) and (ii) in Lemma 3.B.14. □

Lemma 3.B.14. *Let r be an integer such that $r \geq 3$ and $(c_2, c_r) \in \mathbb{R}^2$ with $c_2 > 0$ and $c_r \neq 0$.*

(i) If r is odd there exists a probability distribution P such that $\mathbb{E}[X^2] = c_2$ and $\mathbb{E}[X^r] = c_r$ for $X \sim P$.

(ii) if r is even and $|c_r| \geq c_2^{r/2}$ there exists a probability distribution P such that $\mathbb{E}[X^2] = c_2$ and $\mathbb{E}[X^r] = c_r$ for $X \sim P$.

Proof. The proof is constructive. We provide an example of 2-point distribution P such that $\mathbb{E}[X^2] = c_2$ and $\mathbb{E}[X^r] = c_r$ if $X \sim P$, with the extra condition $|c_r| \geq c_2^{r/2}$ if r is even. Let

$$P(X = a) = p, \qquad P(X = b) = 1 - p$$

for $a, b \in \mathbb{R}$ and $p \in [0, 1]$. Then the condition $\mathbb{E}[X^2] = c_2$ and $\mathbb{E}[X^r] = c_r$ is equivalent to being able to solve the following system of two polynomial equations

$$c_2 = a^2 p + b^2(1 - p), \qquad c_r = a^r p + b^r(1 - p) \tag{3.B.7}$$

for a, b and p.

If $|c_r| \geq c_2^{r/2}$ then

$$b = 0, \quad a = \operatorname{sign}(c_r) \cdot \left|\frac{c_r}{c_2}\right|^{1/(r-2)}, \quad p = c_2 \left|\frac{c_2}{c_r}\right|^{2/(r-2)}$$

is a solution to the polynomial equations (3.B.7), and the assumption $|c_r| \geq c_2^{r/2}$ guarantees that $p \in (0, 1]$. If $|c_r| = c_2^{r/2}$ we have $p = 1$, and the distribution becomes a 1-point distribution. This proves (ii) and acts as the first case of (i).

If $|c_r| < c_2^{r/2}$ and r is odd then

$$b = -\sqrt{c_2}, \quad a = \sqrt{c_2}, \quad p = \frac{1}{2}\left(1 + \frac{c_r}{c_2^{r/2}}\right)$$

is a solution to the polynomial equations (3.B.7). The assumption that $|c_r| < c_2^{r/2}$ guarantees that $p \in (0, 1)$. If r were even, $a^r = b^r$, and it follows that a, b and p is only a solution for odd r. This concludes the proof of (i). □

Remark 3.B.15. The inequality for even r in (ii) is necessary in the sense that it can be seen by Jensen's inequality that $\mathbb{E}[X^r] \geq \mathbb{E}[X^2]^{r/2}$ for any probability distribution when r is even.

3.B.5 Details regarding the numerical experiments and additional results

Simulations from the steady-state distribution are obtained by using the representation (3.A.1). Supposing that M can be diagonalized as $M = QDQ^{-1}$, where $D = \operatorname{diag}(\delta_1, \ldots, \delta_d)$ is a diagonal matrix of complex eigenvalues, we get that

$$X = \int_0^\infty e^{sM} \mathrm{d}Z_s = \int_0^\infty Q e^{sD} Q^{-1} \mathrm{d}Z_s = Q \begin{pmatrix} \int_0^\infty e^{\delta_1 s} \mathrm{d}\tilde{Z}_s^{(1)} \\ \vdots \\ \int_0^\infty e^{\delta_d s} \mathrm{d}\tilde{Z}_s^{(d)} \end{pmatrix} \tag{3.B.8}$$

where $\tilde{Z}_s^{(i)} = (Q^{-1} Z_s)_i$.

Our numerical experiments use simulations from M-selfdecomposable distributions given in terms of:

1. A Lévy process Z with independent coordinates, each being a compound Poisson process with beta distributed jumps.
2. A stable M-matrix parametrized by a correlation parameter ρ and a parameter γ controlling the asymmetry of M.

The i-th compound Poisson process is

$$Z_t^{(i)} = \sum_{j=1}^{N_t^{(i)}} J_j^{(i)},$$

where $N^{(i)}$ is a homogeneous Poisson process with rate $\lambda_i > 0$ and $J_1^{(i)}, J_2^{(i)}, \ldots$ are i.i.d. beta distributed jumps. In this case, each univariate stochastic integral in (3.B.8) is an infinite sum over jumps of $\tilde{Z}_s^{(i)}$,

$$\int_0^\infty e^{\delta_i s} \mathrm{d}\tilde{Z}_s^{(i)} = \sum_{j=1}^{\infty} \tilde{J}_j^{(i)} e^{\delta_i s_j}, \tag{3.B.9}$$

with the exponential factor decaying rapidly due to stability of M. Simulations of X are implemented via truncation of the infinite series (3.B.9).

The i-th Poisson process has rate $\lambda_i > 0$, and all d independent compound Poisson processes have beta distributed jumps sharing the parameters $\nu > 0$ and $\mu \in (0, 1)$, known as the mean and size parameters, respectively, of a beta distribution. The shape parameters are given as $(\alpha, \beta) = (\mu\nu, (1-\mu)\nu)$ in terms of these parameters.

Since the k-th raw moment of a beta distribution is

$$\int u^k P(\mathrm{d}u) = \prod_{r=0}^{k-1} \frac{\mu\nu + r}{v + r},$$

the diagonal k-th-order cumulants are by (3.B.6) given as $\mathcal{C}_{1,i} = \mathbb{E}(Z_1^{(i)}) = \lambda\mu$ and

$$\mathcal{C}_{k,i\ldots i} = \lambda \prod_{r=0}^{k-1} \frac{\mu\nu + r}{\nu + r} = \frac{\mu\nu + k - 1}{\nu + k - 1} \mathcal{C}_{k-1,i\ldots i}$$

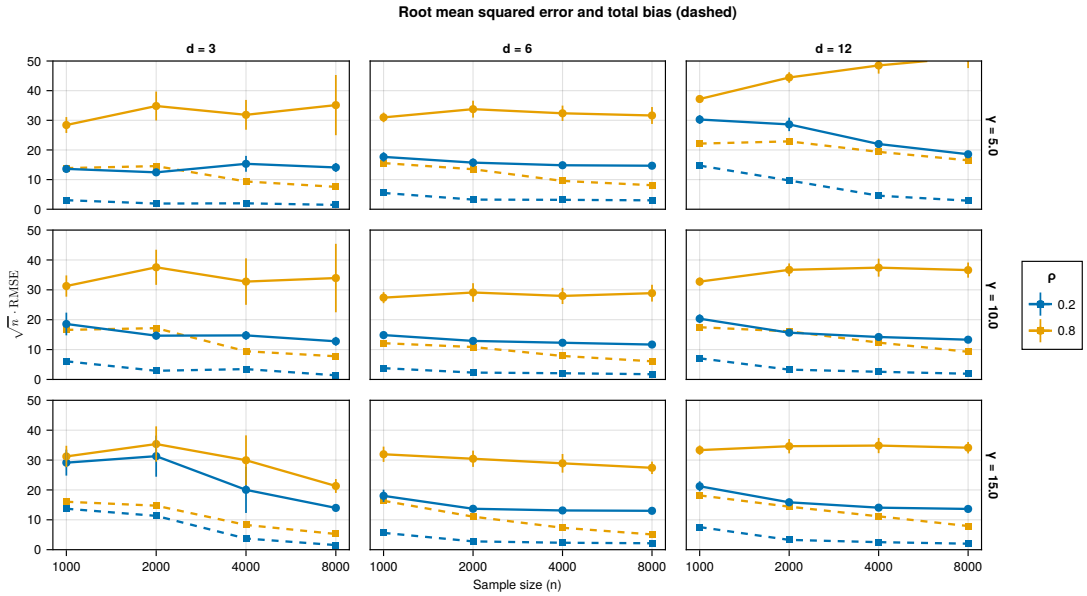


Figure 3.B.2: Estimation errors for the least singular value estimator based on all second-, third- and fourth-order cumulants. Circles connected with full lines show the $\sqrt{n}$-scaled root mean squared error, while the squares connected with dashed lines show the $\sqrt{n}$-scaled total bias.

for $k \geq 2$ and $i = 1, \ldots, d$. The $d \times d$ matrix M is parametrized as

$$M = \left(\gamma E^{\text{skew}} - dI\right)(I - \eta E) = \begin{pmatrix} -d & \gamma & \ldots & \gamma \\ -\gamma & -d & \ldots & \gamma \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma & -\gamma & \ldots & -d \end{pmatrix} \begin{pmatrix} 1-\eta & -\eta & \ldots & -\eta \\ -\eta & 1-\eta & \ldots & -\eta \\ \vdots & \vdots & \ddots & \vdots \\ -\eta & -\eta & \ldots & 1-\eta \end{pmatrix}.$$

where I is the identity matrix, E is the matrix with $E_{ij} = 1$, and E^{skew} is the skew symmetric matrix with $E^{\text{skew}}_{ij} = 1$ for $i < j$. The parameters are $\gamma \in \mathbb{R}$ and $\eta < 1/d$. If $\lambda_1 = \ldots = \lambda_d = \lambda$, then with

$$c = \frac{\lambda\mu(\mu\nu + 1)}{2d(\nu + 1)}$$

we have $\mathcal{C}_{2,ii} = 2dc$, and it is straightforward to see that the positive definite matrix $\Sigma = c\,(I - \eta E)^{-1}$ solves the second-nd order Lyapunov equation for the choice of M above. By the Sherman-Morrison formula

$$\Sigma = c\,(I - \eta E)^{-1} = c\left(I + \frac{\eta}{1 - d\eta} E\right).$$

The variance of all coordinates is

$$\sigma^2 = c\left(1 + \frac{\eta}{1 - d\eta}\right) = c\frac{1 + (1-d)\eta}{1 - d\eta},$$

and the correlation between any two coordinates is

$$\rho = \frac{\eta}{1 + (1-d)\eta} = \frac{1}{\eta^{-1} + (1-d)}.$$

The correlation is thus entirely determined by η, it increases monotonically with η, it is positive for $\eta \in (0, 1/d)$, tending to 1 for $\eta \to 1/d$, it is 0 for $\eta = 0$, and it is negative for $\eta < 0$, tending to $-1/(d-1)$ for $\eta \to -\infty$. The parameter γ does not affect the covariance matrix but determines the asymmetry of M. For the numerical experiments, M is parametrized by the correlation parameter ρ via

$$\eta = \frac{\rho}{1 + \rho(d-1)}$$

for $\rho \in (-1/(d-1), 1)$.

3.C. Linear Shrinkage estimator

As an alternative to the singular value estimator from Definition 3.4.1 one could also a imagine an estimator where $A\text{vec}(M)$ is minimized under a linear constraint instead. To this end let $w \in \mathbb{R}^{d^2}$ denote a vector which is linearly independent of the rows of A. Then the matrix defined by

$$A_w = \begin{pmatrix} A \\ w^T \end{pmatrix}, \tag{3.C.1}$$

will have full rank ($= d^2$) for the true $(\Sigma, \mathcal{K}) \in \mathcal{M}_G^{2,r}$. Thus, A_w determines a unique element $\text{vec}(M_w) \in \ker(A)$ characterized by being the unique solution to

$$A_w \text{vec}(M) = e_{b+1},$$

where $e_{b+1} \in \mathbb{R}^{b+1}$ is the vector which is 0 for all the first b entries and with the last entry equal to 1. So $\text{vec}(M_w)$ is the unique element in the kernel of A satisfying the linear constraint $w^T \text{vec}(M_w) = 1$. Throughout we assume w is picked such that M_w is stable. Since A_w has full rank, we can write

$$\text{vec}(M_w) = (A_w^T A_w)^{-1} A_w^T e_{b+1}.$$

So a sensible estimator of $\text{vec}(M_w)$ would be

$$\text{vec}(\hat{M}_w) = (\hat{A}_w^T \hat{A}_w)^{-1} \hat{A}_w^T e_{b+1}. \tag{3.C.2}$$

In practice when we estimate A by its sample version $\hat{A}$ it will typically already have full rank, so by adding a linear constraint, the sample version $\hat{A}_w$ yields an overdetermined system. Thus, the estimator defined by equation (3.C.2) will not satisfy the constraint $w^T \text{vec}(\hat{M}_w) = 1$ exactly but instead shrink towards it. It will only satisfy the constraint exactly in the limit.

We explore how the asymptotics of $\hat{M}_w$ and the singular value estimator $\hat{M}$ compare. To this end we note that A_w is given as an affine map in the cumulants of the form

$$\text{vec}(\hat{A}_w) = \mathcal{A}_0\hat{\kappa} + \text{vec}(0, \dots, 0, w), \tag{3.C.3}$$

where $\mathcal{A}_0$ is given by the block matrix with $\mathcal{A}$ as defined in equation (3.4.21) above an additional p^2 rows of zeros corresponding to the last row of A_w which doesn't depend on the cumulants. We let $\hat{\kappa}$ denote the unique vectorizations of the empirical cumulants. By essentially the same proof as that of Theorem 3.4.2 we obtain the following asymptotic result about the estimator $\hat{M}_w$.

Theorem 3.C.1. *Let $X_1, \dots, X_n$ be an i.i.d. sample from a distribution in $\mathcal{P}_G$ with finite k-th order moments, and suppose that A_w, is a matrix whose entries are linearly dependent on the cumulants with full rank of the form (3.C.1). Let $(\hat{A}_w)_n$ denote the plug-in estimator of A_w based on the empirical cumulants and let $\text{vec}(\hat{M}_w)$ be the estimator defined by equation (3.C.2). Then*

$$(\hat{A}_w)_n \xrightarrow{P} A_w, \qquad (\hat{M}_w)_n \xrightarrow{P} M_w \tag{3.C.4}$$

for $n \to \infty$. If, additionally, the distribution has finite $2k$-th order moments,

$$\sqrt{n}(\text{vec}((\hat{A}_w)_n - A_w)) \xrightarrow{D} \mathcal{N}\left(0, \mathcal{A}_0\Omega\mathcal{A}_0^T\right)$$
$$\sqrt{n}(\text{vec}((\hat{M}_w)_n - M_w)) \xrightarrow{D} \mathcal{N}\left(0, \left(\text{vec}(M_w)^T \otimes A_w^+\right)\mathcal{A}_0\Omega\mathcal{A}_0^T\left(\text{vec}(M_w) \otimes (A_w^+)^T\right)\right)$$

for $n \to \infty$, where A_w^+ denotes the Moore-Penrose inverse of A_w, M_w is the element in $\ker(A)$ such that $w^T\text{vec}(M_w) = 1$ and Ω is defined by equation (3.4.22).

Proof. The proof proceeds exactly as the proof of Theorem 3.4.2 by using the continuous mapping theorem and the delta method with Lemma 3.C.4 at the end of the section providing the needed derivative. □

As was the case for Theorem 3.4.2 we can obtain a corollary where the asymptotic covariance is rewritten slightly.

Corollary 3.C.2. *If A is defined by equation (3.4.18) and A_w by equation (3.C.1) the asymptotic covariance of $\text{vec}(\hat{M}_w)$ can also be written as*

$$\|M_w\|^2\left[\left(\left(I - \text{vec}(M_w)w^T\right)A^+\right)B_{2,r}(M)\right]\Omega\left[\left(\left(I - \text{vec}(M_w)w^T\right)A^+\right)B_{2,r}(M)\right]^T,$$

where $B_{2,r}(M)$ is the block diagonal matrix with $B_2(M)$ and $B_r(M)$ on the diagonal, M denotes the stable matrix of Frobenius norm one such that $\text{vec}(M) \in \ker(A)$, and $\|\cdot\|$ denotes the Frobenius norm of a matrix.

If $w = \text{vec}(M)$ then the asymptotic covariance of $\text{vec}(\hat{M}_w)$ simplifies to

$$(A^+B_{2,r}(M))\Omega(A^+B_{2,r}(M))^T.$$

Proof. The first part of the proof is almost identical to the proof of Corollary 3.4.3 if we replace M by M_w, A by A_w and $\mathcal{A}$ by $\mathcal{A}_0$. As in that proof, we let the second- and r-th-order cumulants jointly be given by κ and let j index this set. Doing the same argument as in Corollary 3.4.3 we can write

$$(\mathcal{A}_0)_{*,j} = \mathrm{vec}(A_0(\kappa_j = 1)),$$

where A_0 is defined as A_w with $w = 0$. We then obtain

$$\left(\mathrm{vec}(M_w)^T \otimes A_w^+\right) \mathcal{A}_0 = A_w^+ \left[A_0(\kappa_1 = 1)\mathrm{vec}(M_w) \cdots A_0(\kappa_q = 1)\mathrm{vec}(M_w)\right].$$

Each column $A_0(\kappa_j = 1)\mathrm{vec}(M_w)$ is by definition the off-diagonal Lyapunov equation of order 2 and r combined with an additional zero entry at the end, evaluated at M_w where $\kappa_j = 1$ and the rest of the cumulants are equal to zero. Thus, this matrix is just equal to the block diagonal matrix

$$\left[A_0(\kappa_1 = 1)\mathrm{vec}(M_w) \cdots A_0(\kappa_q = 1)\mathrm{vec}(M_w)\right] = B_{2,r,0}(M_w) = \begin{pmatrix} B_2(M_w) & 0 \\ 0 & B_r(M_w) \\ 0 & 0 \end{pmatrix},$$

So we obtain that

$$\left(\mathrm{vec}(M_w)^T \otimes A_w^+\right) \mathcal{A}_0 = A_w^+ B_{2,r,0}(M_w). \tag{3.C.5}$$

Let $\mathrm{vec}(M)$ denote the element in $\ker(A)$ which has norm 1 with M stable, then we can write $M_w = \|M_w\| M$. Since $B_{2,r,0}$ is linear in M we can write $B_{2,r,0}(M_w) = \|M_w\| B_{2,r,0}(M)$.

To obtain the expression given in the Corollary we need an expression for the Moore-Penrose inverse of A_w in terms of A and w. Since it is given by the block structure,

$$A_w = \begin{pmatrix} A \\ w^T \end{pmatrix},$$

we obtain by the Greville formula [Greville, 1960] for the Moore-Penrose inverse that

$$A_w^+ = \begin{pmatrix} A^+ - \frac{c^T w^T A^+}{\|c\|^2} & \frac{c^T}{\|c\|^2} \end{pmatrix},$$

where $c = w^T(I - AA^+)$ under the assumption that w^T is not in the rowspace of A, which is a consequence of the assumption that w is such that A_w has full rank.

Because of the zero row in $B_{2,r,0}(M)$ the product of A_w^+ and $B_{2,r,0}(M)$ is simply the the left-most block in A_w^+ multiplied by $B_{2,r}(M)$, so we obtain that equation (3.C.5) can be written as

$$A_w^+ B_{2,r,0}(M_w) = \|M_w\| \left(A^+ - \frac{c^T w^T A^+}{\|c\|^2}\right) B_{2,r}(M), \tag{3.C.6}$$

by also using $B_{2,r,0}(M_w) = \|M_w\| B_{2,r,0}(M)$.

To obtain the formula given in the corollary we observe that $(I - AA^+)$ is the orthogonal projection onto the kernel of A. Therefore, if we let $\text{vec}(M)$ denote the element in the kernel of A which has norm 1 with M stable, $(I - AA^+) = \text{vec}(M)\text{vec}(M)^T$, so

$$c = w^T \text{vec}(M)\text{vec}(M)^T,$$

and

$$\|c\|^2 = (w^T \text{vec}(M))^2.$$

Thus, if we use that $M_w = \|M_w\| M$ and $w^T M_w = 1$ we obtain that $c^T/(\|c\|^2) = \text{vec}(M_w)$. Inserting this into equation (3.C.6) yields the expression given in the corollary.

In the case $w = \text{vec}(M)$, the constraint $w^T \text{vec}(M) = 1$ results in $M = M_w$. Furthermore, for w equal to any element in $\ker(A)$ (so a multiple of $\text{vec}(M)$) the product $w^T A^+$ is 0. To see this we first rewrite it as $(A^+)^T w$. By definition of the Moore-Penrose inverse the kernel of $(A^+)^T$ is equal to the kernel of A, so if $w \in \ker(A)$, $(A^+)^T w = 0$. Inserting this and $M = M_w$ into the expression for the covariance we obtain the special case for $w = \text{vec}(M)$. □

Thus, if we could pick $w = \text{vec}(M)$ the asymptotic covariance of $\text{vec}(\hat{M}_w)$ and the singular value estimator, $\text{vec}(\hat{M})$, would be the same. However, $\text{vec}(\hat{M}_w)$ will generally not have norm one like the singular value estimator, $\text{vec}(\hat{M})$. With the current definition of $\text{vec}(\hat{M}_w)$ the asymptotic covariance scales with the norm of $\text{vec}(M_w)$ which is only 1 if we pick $w = \text{vec}(M)$. Thus, to be able to make a more sensible comparison of the asymptotic covariances of two estimators we normalize the linear shrinkage estimator. Define the following estimator

$$\text{vec}(\tilde{M}_w) = \frac{\text{vec}(\hat{M}_w)}{\|\hat{M}_w\|}, \tag{3.C.7}$$

as $\text{vec}(\hat{M}_w)$ rescaled by its norm. Thus, $\text{vec}(\tilde{M}_w)$ will always have norm equal to one just as the singular value estimator does. Furthermore, by the continuous mapping theorem it will converge in probability to $\text{vec}(M)$ which is the element in the kernel of A which is stable and of norm one. That it converges to the stable matrix (and not the negation of it) is ensured by picking the correct sign of w. It converges to the same value as the singular value estimator, so it is sensible to compare their asymptotic covariances. We show in Theorem 3.C.3 that $\text{vec}(\tilde{M}_w)$ and $\text{vec}(\hat{M})$ actually have the same asymptotic covariance for any w such that A_w has full rank and M_w is stable.

Theorem 3.C.3. *Let $w \in \mathbb{R}^{d^2}$ be any vector such that A_w has full rank and $w^T \text{vec}(M_w) = 1$ for M_w stable and $M_w \in \ker(A)$. Let $\text{vec}(\tilde{M}_w)$ be defined by equation (3.C.7) and $\text{vec}(\hat{M})$ be the singular value estimator defined by Definition 3.4.1. Then the two estimators $\text{vec}(\hat{M})$ and $\text{vec}(\tilde{M}_w)$ have the same asymptotic covariance.*

Proof. We will show that the asymptotic covariance of $\text{vec}(\tilde{M}_w)$ is equal to the expression given for the asymptotic covariance of $\text{vec}(\hat{M})$ in Corollary 3.4.3.

Let $g : \mathbb{R}^{d^2} \to \mathbb{R}^{d^2}$ denote the map that normalizes a vector, $g(x) = x/\|x\|$. Then $\text{vec}(\tilde{M}_w) = g(\text{vec}(\hat{M}_w))$. Thus, we can obtain the asymptotic distribution of $\tilde{M}_w$ by

application of the delta method and using the expression for the asymptotic covariance of $\mathrm{vec}(\hat{M}_w)$ provided in Corollary 3.C.2.

The Jacobian of g is $Dg(x) = (I - xx^T/\|x\|^2)/\|x\|$. Thus, the asymptotic covariance of $\tilde{M}_w$ is

$$Dg\left(\mathrm{vec}(M_w)\right) \Psi Dg\left(\mathrm{vec}(M_w)\right)^T, \tag{3.C.8}$$

for Ψ equal to the covariance of $\hat{M}_w$ given in Corollary 3.C.2.

Thus, the asymptotic covariances of the normed linear shrinkage estimator and the singular value estimator are equal if

$$Dg\left(\mathrm{vec}(M_w)\right) \|M_w\| \left(\left(I - \mathrm{vec}(M_w)w^T\right) A^+\right) = A^+, \tag{3.C.9}$$

by comparison of Corollary 3.4.3 and Corollary 3.C.2.

Inserting the expression for Dg and observing that the Frobenius norm of a matrix is the same as the 2-norm of the vectorization of that matrix we obtain that the left hand side equals

$$\begin{aligned}
&\left(I - \frac{\mathrm{vec}(M_w)\mathrm{vec}(M_w)^T}{\|\mathrm{vec}(M_w)\|^2}\right)\left(\left(I - \mathrm{vec}(M_w)w^T\right) A^+\right) \\
&= \left(I - \frac{\mathrm{vec}(M_w)\mathrm{vec}(M_w)^T}{\|\mathrm{vec}(M_w)\|^2}\right) A^+ - \left(I - \frac{\mathrm{vec}(M_w)\mathrm{vec}(M_w)^T}{\|\mathrm{vec}(M_w)\|^2}\right) \mathrm{vec}(M_w)w^T A^+
\end{aligned}$$

The second term vanishes since

$$\left(I - \frac{\mathrm{vec}(M_w)\mathrm{vec}(M_w)^T}{\|\mathrm{vec}(M_w)\|^2}\right)$$

is equal to the projection onto the orthogonal complement of $\mathrm{vec}(M_w)$ Therefore, when it is multiplied onto $\mathrm{vec}(M_w)$ it is equal to zero.

To see that the first term is just equal to A^+ we use the same argument as in the last part of the proof of Corollary 3.C.2, which yields that $\mathrm{vec}(M_w)^T A^+ = (A^+)^T \mathrm{vec}(M_w) = 0$ since the kernel of $(A^+)^T$ is equal to the kernel of A and $\mathrm{vec}(M_w) \in \ker(A)$ by definition. Thus, we have shown that (3.C.9) is true, which completes the proof. □

Lemma 3.C.4. *Let $G : \mathbb{R}^{n\times m} \to \mathbb{R}^m$ denote the map $G(X) = (X^TX)^{-1}X^Tb$ for $b \in \mathbb{R}^n$. If $A \in \mathbb{R}^{n\times m}$ is a matrix of full column rank such that $Av = b$ for $v \in \mathbb{R}^m$, then the Jacobian of G, evaluated at A is given by*

$$DG(A) = \left.\frac{\partial \mathrm{vec}(G)}{\partial \mathrm{vec}(X)^T}\right|_{X=A} = -v^T \otimes (A^TA)^{-1}A^T = -v^T \otimes A^+.$$

Proof. We calculate the derivative of G as in Magnus and Neudecker [2019] and then vectorize at the end to obtain the Jacobian. If we let $H = X^TX$ we obtain

$$dG = d(H^{-1})X^Tb + H^{-1}d(X^T)b, \tag{3.C.10}$$

by the product rule. We find

$$d(H^{-1}) = -H^{-1}(dH)H^{-1}, \qquad d(H) = d(X^T)X + X^Td(X),$$

and inserting this into equation (3.C.10) we obtain the differential of G

$$\begin{aligned} dG &= -H^{-1}(d(X^T)X + X^T d(X))H^{-1}X^T b + H^{-1}d(X^T)b \\ &= H^{-1}d(X^T)(b - XG(X)) - H^{-1}X^T d(X)G(X). \end{aligned}$$

Vectorizing the equation and using rules of vectorization of products, see Chapter 2 in Magnus and Neudecker [2019] we obtain

$$\begin{aligned} \operatorname{vec}(dG) &= \left((b - XG(X))^T \otimes H^{-1}\right) \operatorname{vec}(d(X^T)) - \left(G(X)^T \otimes H^{-1}X^T\right) \operatorname{vec}(d(X)) \\ &= \left(\left((b - XG(X))^T \otimes H^{-1}\right) K - \left(G(X)^T \otimes H^{-1}X^T\right)\right) \operatorname{vec}(d(X)), \end{aligned}$$

where K denotes the commutation matrix changing the vectorization of a matrix to the vectorization of its transpose. Thus, the Jacobian of $\operatorname{vec}(G)$ with respect to $\operatorname{vec}(X)$ is the matrix multiplied onto $\operatorname{vec}(d(X))$. Evaluating at $X = A$ we obtain

$$DG(A) = \left(\left((b - Av)^T \otimes (A^T A)^{-1}\right) K - \left(v^T \otimes (A^T A)^{-1}A^T\right)\right) = -\left(v^T \otimes (A^T A)^{-1}A^T\right),$$

since A has full rank and $G(A) = v$ and $Av = b$. □

4 Causal Inference with Latent Operator Selfdecomposable Distributions

Alex Markham, Cecilie Olesen Recke, Niels Richard Hansen

Abstract

This manuscript represents work in progress. The goal of the final version is to describe a complete workflow for learning *latent operator-selfdecomposable distributions* (latent OSDs) from single cell biology data. The OSD distributions are multivariate distributions, which can be interpreted as cross-sectional steady-state distributions from a stochastic dynamical system. The final workflow will be able to integrate observational and interventional data. The current version of the paper outlines the theoretical foundation. That is, it describes the OSD distributions, their causal semantics, the observation noise model, and the cumulant equations that will be central for the learning algorithms. Leveraging recent identification results for OSD distributions, we show a generic identification result for latent OSD distributions when the observation noise is Poisson.

4.1 Introduction

Cross-sectional observations from dynamical systems with feedback loops is a challenge for conventional structural causal models. A prime example is single cell data, where the cell is destroyed during measurement, making it impossible to follow the dynamics of any individual cell. Data thus consists of mere snapshots of the cell constituents determined by the cell's intricate regulatory feedback mechanisms. The ambition is, nevertheless, to reverse engineer the gene regulatory network from the cross-sectional distributions. Sethuraman et al. [2023] presented a method for learning such a network, that allows for feedback loops, in terms of conventional structural causal models with cycles.

Contrary to the literature based on structural causal models, recent approaches by Wang et al. [2023]; Rohbeck et al. [2024]; Guan et al. [2024, 2025]; Lorch et al. [2026] are based on dynamic process models of chemical reactions in the cell. A related approach to model single cell constituents via dynamical marginal models was given by Gorin et al. [2022]. It is characteristic of these approaches that they do not posit a structural equation of the cross-sectionally observed variables but rather of the underlying dynamical process. Such a model specification then implicitly defines models of cross-sectional observations.

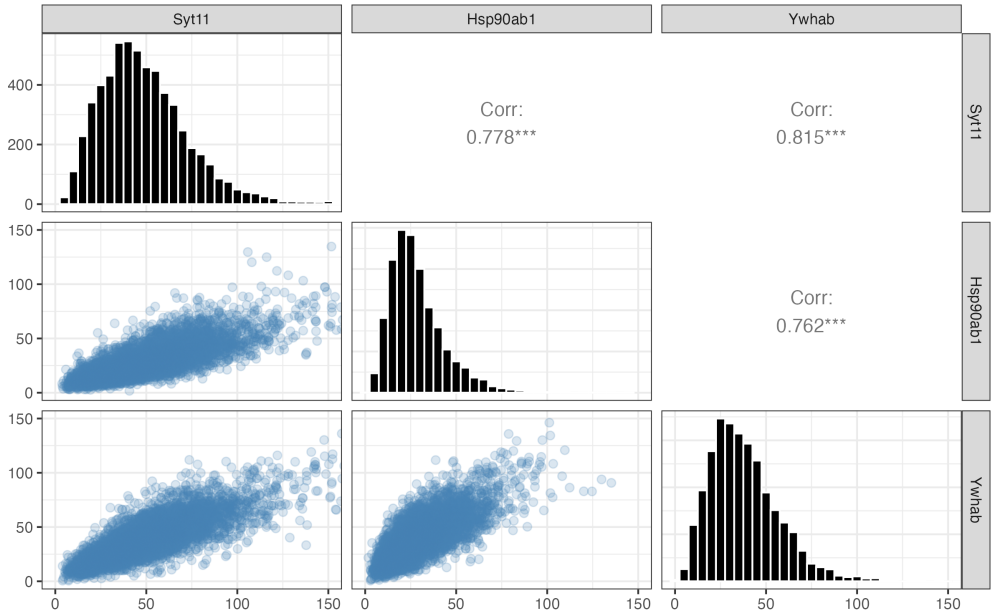


Figure 4.1.1: Scatter plots of single cell count data for three highly correlated genes and $n = 5892$ cells. The integer data points are jittered to avoid overplotting.

The ultimate purpose of learning a gene regulatory network model is to be able to reason about effects of interventions, e.g., gene knockout. We argue that the causal semantics of conventional structural causal models is inappropriate for this purpose, and we introduce an alternative causal semantics specifically for cross-sectional distributions of dynamical systems in steady-state. We do this in terms of the class of distributions known as *operator-selfdecomposable* distributions that arise as steady-state distributions of linear stochastic differential equations driven by Lévy process noise.

The models should be able to realistically replicate single cell data. Figure 4.1.1 shows an example of such data in terms of scatter plots and corresponding correlations from $n = 5892$ single cell gene expression measurements for a subset of $d = 3$ genes. The data is characteristically non-Gaussian, and the methods we propose exploit this and use information beyond correlations for inference. Additionally, single cell data are typically counts, and thus integer valued, and the measurement process itself adds noise to the observations.

Our proposal is to model single cell data via *latent* operator-selfdecomposable distributions, where the causal semantics is among the latent variables via a stochastic process model, while the model for the actual observations includes an observation noise model. The causal semantics of the latent distributions allows us to derive a collection of coherent distributions that can model both observational data and interventional data, where, e.g., one or more genes have been knocked out.

4.2 Lévy processes and OSD distributions

The objective is to build causal models of d-dimensional cross-sectional observations from a dynamical system in steady-state. For a model to be causal, it needs to capture both the observational steady-state distribution and the interventional steady-state distributions, and thus we need to specify families of multivariate distributions that enable us to do precisely that.

The *operator-selfdecomposable* (OSD) distributions on $\mathbb{R}^d$, see Definition 4.2.1, will serve this purpose. The OSD distributions form a subset of the set of multivariate infinitely divisible distributions, and they can be be equipped with a causal semantics. They include well-known elliptical distributions, such as the Gaussian and t-distributions, but they also include a range of other distributions with properties quite different from elliptical distributions.

We define the OSD distributions constructively via a stochastic integral representation. The representation leads directly to a generative model of observations from an OSD distribution. The causal semantics of OSD distributions is, however, defined by their implicit representation as steady-state distributions for a class of Markov processes, see Proposition 4.2.2, which allows us to define interventional distributions.

The definition of OSD distributions may appear technical at first, but in Section 4.2.4 we argue how they can be seen as temporal extensions of distributions defined via linear structural equations. For the case of independent errors (IE), we additionally argue how the stochastic integral representation of resulting IE-OSD distributions is closely related to Independent Component Analysis (ICA).

4.2.1 OSD distributions

The operator-selfdecomposable distributions are defined via a stochastic integral w.r.t. a Lévy process. A continuous time stochastic process $Z = (Z_t)_{t\geq 0}$ with $Z_t \in \mathbb{R}^d$ is a Lévy process if: $Z_0 = 0$; if the increments of Z are independent and stationary; and if Z is continuous in probability [Sato, 2013]. Throughout, Z denotes a Lévy process satisfying the mild integrability condition

$$\mathbb{E}(\log(1 + \|Z_1\|)) < \infty. \tag{4.2.1}$$

Recall that a matrix M is a *stable matrix* if all eigenvalues of M have strictly negative real part.

Definition 4.2.1 (OSD distributions)**.** *Let M be a $d \times d$ stable matrix and let $Z = (Z_t)_{t\geq 0}$ denote a d-dimensional Lévy process satisfying (4.2.1). The distribution of*

$$X = \int_0^\infty e^{sM} \mathrm{d}Z_s \tag{4.2.2}$$

is called M-selfdecomposable. A distribution is operator-selfdecomposable, or OSD, if it is M-selfdecomposable for some stable M and some Lévy process Z.

The Lévy process $Z_t = (Z_t^{(1)}, \ldots, Z_t^{(d)})$ in the representation (4.2.2) is a d-dimensional process, which we refer to as the *driving* Lévy process or the *error* process. Its i-th coordinate, $Z^{(i)}$, is itself a Lévy process. In our general definition we do not assume that these marginal Lévy processes are independent, but if they are, the resulting distribution is referred to as being *independent-error*-OSD or IE-OSD.

Stability of M combined with the integrability condition (4.2.1) guarantees that the integral (4.2.2) is well defined [Zbigniew and David, 1993, Section 3.6, specifically identities (3.6.17) and (3.6.18)]. The stochastic integral (4.2.2) is an integral of a deterministic, but matrix-valued, integrand e^{sM} with respect to the vector-valued stochastic process Z_s. It can be defined in several equivalent ways. By partial integration we can express X as the standard integral

$$X = -M \int_0^\infty e^{sM} Z_s \mathrm{d}s \tag{4.2.3}$$

of the vector-valued integrand $e^{sM} Z_s$ [Zbigniew and David, 1993, formula (3.6.3)], and we could use (4.2.3) as an equivalent definition of M-selfdecomposable distributions.

While (4.2.2) and (4.2.3) are useful generative representations of an OSD distribution – for simulations as well as for theoretical derivations – they do not reveal why OSD distributions are models of cross-sectional data from dynamical systems. The following proposition shows how these distributions arise naturally as steady-state distributions of Markov processes.

Proposition 4.2.2 (Sato and Yamazato [1984, Theorem 4.1])**.** *Let M be a $d \times d$ stable matrix and let $Z = (Z_t)_{t \geq 0}$ denote a d-dimensional Lévy process satisfying (4.2.1). The M-selfdecomposable distribution given by (4.2.2) is the unique steady-state distribution of the stationary Markov process solving the stochastic differential equation (SDE)*

$$\mathrm{d}X_t = MX_t \mathrm{d}t + \mathrm{d}Z_t. \tag{4.2.4}$$

4.2.2 Interventional distributions

Our main motivation for studying OSD distributions is their interesting causal semantics. Following Lauritzen and Richardson [2002]; Sokol and Hansen [2014], the SDE can be given a causal interpretation[1] with interventions defined by substitution. That is, we define the intervention on a coordinate by substituting that coordinate with a fixed value in the SDE. Interventions may, of course, be defined for any subset of coordinates similarly, and we can also define more general interventions where coordinates are not just fixed [Sokol and Hansen, 2014].

Here we consider intervening on a block of coordinates, so we write $Z_t = (Z_t^{(1)}, Z_t^{(2)})$ and $X_t = (X_t^{(1)}, X_t^{(2)})$, and we let

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

[1] This is a *structural* interpretation; the structure of the SDE is assumed invariant to interventions, and the SDE is interpreted as an infinitesimal structural causal model [Sokol and Hansen, 2014].

denote the corresponding block partition of M. Intervening by fixing $X_t^{(1)} = x^{(1)}$ for all $t \geq 0$, as defined by Lauritzen and Richardson [2002]; Sokol and Hansen [2014], gives the SDE

$$\begin{aligned} \mathrm{d}X_t^{(2)} &= (M_{22}X_t^{(2)} + M_{21}x^{(1)})\mathrm{d}t + \mathrm{d}Z_t^{(2)} \\ &= M_{22}X_t^{(2)}\mathrm{d}t + \mathrm{d}(Z_t^{(2)} + M_{21}x^{(1)}t). \end{aligned} \tag{4.2.5}$$

A process $X_t^{(2)}$ solving (4.2.5) is generally different from the second block of a process $X_t = (X_t^{(1)}, X_t^{(2)})$ solving (4.2.4). To make this explicit, we could choose to annotate the variable $X_t^{(2)}$ by, e.g., $\mathrm{do}(X^{(1)} = x^{(1)})$. This becomes a cumbersome notation, and we prefer to push intervention-annotations to the probability measure, formally by replacing (4.2.5) with the SDE

$$\mathrm{d}X_t^{(2)} = M_{22}X_t^{(2)}\mathrm{d}t + \mathrm{d}\tilde{Z}_t^{(2)}, \tag{4.2.6}$$

where $\tilde{Z}^{(2)}$ is a Lévy process defined on a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mathbb{P}^{\mathrm{do}(X^{(1)}=x^{(1)})})$ with $\tilde{Z}^{(2)} \overset{\mathcal{D}}{=} (Z_t^{(2)} + M_{21}x^{(1)}t)_{t\geq 0}$. See Appendix 4.B for elaborations on the notation. We show that the interventional distributions obtained by fixing coordinates are operator-selfdecomposable.

Proposition 4.2.3. *The process $\tilde{Z}^{(2)} \overset{\mathcal{D}}{=} (Z_t^{(2)} + M_{21}x^{(1)}t)_{t\geq 0}$ is a Lévy process fulfilling (4.2.1), and if M_{22} is stable there exists a unique stationary solution to the SDE (4.2.6). Moreover,*

$$X^{(2)} = \int_0^\infty e^{sM_{22}}\mathrm{d}\tilde{Z}_s^{(2)} \tag{4.2.7}$$

has an M_{22}-selfdecomposable distribution, which is the steady-state distribution of the solution to (4.2.6).

The proof of Proposition 4.2.3 is given in Appendix 4.C. Since $\int_0^\infty e^{-sI_{11}}\mathrm{d}s = I_{11}$ for I_{11} the identity matrix corresponding to the block $x^{(1)}$, it also follows that when M_{22} is stable,

$$\begin{pmatrix} x^{(1)} \\ X^{(2)} \end{pmatrix} = \int_0^\infty e^{s\tilde{M}}\mathrm{d}\tilde{Z}_s, \quad \tilde{M} = \begin{pmatrix} -I_{11} & 0 \\ 0 & M_{22} \end{pmatrix},$$

where $\tilde{Z}_s = (x^{(1)}s, \tilde{Z}_s^{(2)})$. Thus, when M_{22} is stable, Proposition 4.2.3 implies that the fixing-intervention defined by (4.2.6) results in a stationary solution to the SDE, and the distributions of $X^{(2)}$ as well as $(x^{(1)}, X^{(2)})$ are OSD. If M_{22} is not stable, there is no stationary solution to (4.2.6) unless $\tilde{Z}_s^{(2)}$ lies in the subspace corresponding to the stable eigenvalues of M_{22}. In the above sense, we can consider the class of OSD distributions to be closed under fixing-interventions whenever fixing defines a steady-state distribution.

To better understand the nature of the interventional distributions, observe that (4.2.7) implies

$$X^{(2)} \overset{\mathcal{D}}{=} \underbrace{\int_0^\infty e^{sM_{22}}\mathrm{d}Z_s^{(2)}}_{X^{(2),0}} + \underbrace{\left(\int_0^\infty e^{sM_{22}}\mathrm{d}s\right)}_{-(M_{22})^{-1}} M_{21}x^{(1)} = X^{(2),0} - (M_{22})^{-1}M_{21}x^{(1)}. \tag{4.2.8}$$

With $P^{\mathrm{do}(X^{(1)}=x^{(1)})} = X^{(2)}(\mathbb{P}^{\mathrm{do}(X^{(1)}=x^{(1)})})$ denoting the interventional distribution of $X^{(2)}$, it follows from (4.2.8) that $X^{(2),0}$ has distribution $P^{\mathrm{do}(X^{(1)}=0)}$, and that the map

$$x^{(1)} \mapsto P^{\mathrm{do}(X^{(1)}=x^{(1)})}$$

is an affine location transformation of $P^{\mathrm{do}(X^{(1)}=0)}$. If Z_t has finite expectation, we also have

$$\mathbb{E}^{\mathrm{do}(X^{(1)}=x^{(1)})}(X^{(2)}) = \int X^{(2)} \mathrm{d}\mathbb{P}^{\mathrm{do}(X^{(1)}=x^{(1)})} = \mathbb{E}\left(X^{(2),0}\right) - (M_{22})^{-1} M_{21} x^{(1)}. \quad (4.2.9)$$

The map $x^{(1)} \mapsto \mathbb{E}^{\mathrm{do}(X^{(1)}=x^{(1)})}(X^{(2)})$ is therefore an affine map, whose parameters are linked to the observational distribution of X via the matrix M and the distribution of Z. We elaborate on this map in Section 4.5, where we derive equations for all cumulants of the interventional distributions.

4.2.3 Graphical representations

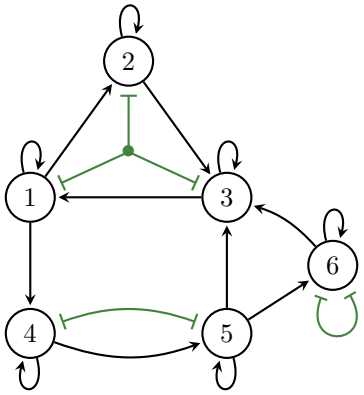
Given the causal interpretation of OSD distributions by fixed-interventions provided above, we introduce the notion of a directed graph encoding the sparsity pattern of M as a summary of the causal effects in the model. Let $G = ([d], E)$ denote a directed graph on d nodes where $[d] = \{1, \ldots, d\}$ and the set of directed edges is E. We will say that a matrix M has sparsity pattern according to the directed graph G if $M_{ji} = 0$ if $i \to j \notin E$ and denote the set of matrices satisfying this as $\mathbb{R}^E$. As only a stable M defines an OSD distribution, we will denote the subset of stable matrices with sparsity pattern encoded by G as $\mathbb{R}^E_{\mathrm{stab}}$. We will typically assume that the graph G contains all self-loops, corresponding to assuming that all variables affect themselves. See Figure 4.2.2 for an example of a directed graph and the corresponding sparsity pattern of M.

We will also represent the dependency structure between the coordinates of the driving Lévy process by a type of graph. We will draw a hyperedge between a set of k coordinates $i_1, \ldots, i_k$ if they are allowed to be non-independent. Thus, the hyperedges will represent sets (of varying sizes) of coordinates of the Lévy process, which are allowed to be dependent on each other, so two coordinates that do not belong to the same hyperedge are independent. Thus, if all the coordinates are independent there is a hyperedge for every node that only contains that node.

The hypergraph in Figure 4.2.2 contains three hyperedges $\{1, 2, 3\}, \{4, 5\}$ and $\{6\}$, which are drawn using blunt edges in green. Therefore, for this graph the first three coordinates of the Lévy process can be dependent and are independent of the last three. Similarly, the fourth and fifth coordinates are allowed to be dependent and the sixth coordinate is assumed independent of all the other coordinates. We combine the directed graph and the hypergraph, as shown in Figure 4.2.2, into a single mixed graph $\mathcal{G} = ([d], E, H)$ where E is the set of directed edges and H is the set of hyperedges.

4.2.4 Connections to linear structural causal models and ICA

Equation (4.2.4) is the fundamental causal model that implicitly defines an OSD and its causal semantics, similarly to how a conventional linear structural causal model defines



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

Figure 4.2.2: A mixed graph $\mathcal{G} = ([6], E, H)$ with the directed edges E drawn in black and the hyperedges H drawn in green. On the right is the sparsity pattern of M encoded by the directed graph on the left.

the causal semantics through the equation $X = BX + \varepsilon$. The solution of this linear equation is

$$X = (I - B)^{-1}\varepsilon, \tag{4.2.10}$$

where $(I-B)^{-1}$ is a linear operator that transforms the noise term ε into the observation vector X. The mixing matrix, $(I - B)^{-1}$, is parametrized by the coefficient matrix B, whose entries can be interpreted as direct causal effects via the causal interpretation of the linear structural causal model.

The explicit representation (4.2.2) of the steady-state distribution of the solution to the SDE (4.2.4) is similar to the representation of the solution to the linear structural causal model given by (4.2.10). It represents the steady-state OSD distribution as a mixing of the noise components in the Lévy process that drives the SDE. The particular mixing operator is still linear and reflects the time-dynamic nature of the noise; the exponential matrix factor e^{sM} represents time-dependent mixing weights – and due to stability of M this gives a built-in discounting of more distant values of Z. The parametrization of this mixing operator is in terms of the matrix M, whose entries can be interpreted as direct causal effects via the causal interpretation of the SDE.

The increments of the Lévy process are by definition independent, but we do not assume that the coordinates of the process are independent. Similarly, the noise vector ε can have dependent coordinates in general. However, if ε has independent coordinates, (4.2.10) is an ICA representation of X as a linear mixing of *independent components*. Similarly, if the coordinates of the driving Lévy process in (4.2.2) are independent, the IE-OSD distribution of X, as given by (4.2.2), is a linear mixture of independent components.

4.2.5 The landscape of multivariate distributions

The operator-selfdecomposable distributions have mostly been studied in probability theory. Appendix 4.A covers a selection of well-known facts about these distributions

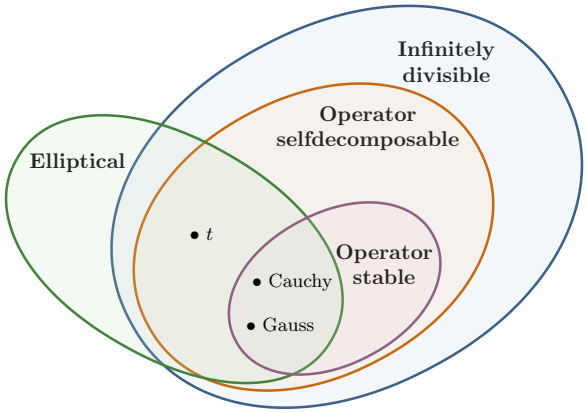


Figure 4.2.3: The OSD distributions form a subset of the infinitely divisible distributions. It includes, but is larger than, the selfdecomposable distributions. The subset of operator stable distributions includes all stable distributions, such as the well known Cauchy and Gaussian distributions. These are also both examples of elliptical distributions, and there is some overlap between the OSD distributions and the elliptical distributions, including the multivariate t-distributions.

and how they are related to other classes of multivariate distributions. The illustration in Figure 4.2.3 provides a summary.

Many classical multivariate distributions, such as Gaussian and t-distributions, are OSD, and so are multivariate generalized hyperbolic distributions [Masuda, 2004, Section 5]. The multivariate stable distributions [Sato and Yamazato, 1984, Example 4.1] are also OSD. However, since OSD distributions are infinitely divisible, there are also many distributions that are not OSD, e.g., non-degenerate distributions with compact support. The OSD distributions are, nevertheless, a lot more flexible than the few classical distributions mentioned here, and their full power as either descriptive or causal models is not well understood.

We want to make one particular comparison to the elliptical distributions, see also Appendix 4.A.4. The Gaussian and t-distributions are elliptical, and one of the properties of elliptical distributions is that they are closed under marginalization and conditioning. Additionally, the (observational) conditional expectation $x^{(1)} \mapsto \mathbb{E}(X^{(2)} \mid X^{(1)} = x^{(1)})$ is affine in $x^{(1)}$. This is, however, not the case for general OSD distributions. The fact that *interventional* expectations are affine, as demonstrated by (4.2.9), cannot be seen directly from their observational conditional distributions. This illustrates that simple regression models using observational data can misrepresent the causal semantics of the model in a very fundamental way.

4.3 Latent OSD models and Poisson observation noise

To more accurately represent real data such as integer counts, we introduce an observation noise model on top of an OSD model. If $Y \in \mathbb{N}_0^d$ denotes the d-dimensional observation vector of counts, our *latent OSD model* makes the following three assump-

tions.

1. $X \in \mathbb{R}^d$ is a random variable with an OSD distribution.
2. The coordinates of Y are independent given X, that is, $Y_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp Y_d \mid X$.
3. The conditional distribution of $Y_i \mid X$ depends only on X_i.

We refer to the distribution of $Y_i \mid X_i$ as the *observation noise model*. As a first example of an observation noise model, which will result in integer observations, we introduce a Poisson observation noise. In this model, each coordinate follows a Poisson distribution conditioned on an underlying OSD distribution. The conditional mean of each coordinate will be the corresponding coordinate of the OSD distribution.

Definition 4.3.1 (Poisson observation noise)**.** *Let M be a $d \times d$ stable matrix and let $Z = (Z_t)_{t \geq 0}$ denote a d-dimensional Lévy process satisfying (4.2.1) and let X follow the d-dimensional M-selfdecomposable distribution given by (4.2.2) such that all coordinates of X are non-negative almost surely. The distribution of $Y \in \mathbb{N}_0^d$ satisfying*

$$Y_i \mid X = x \sim \text{Pois}(x_i), \qquad Y_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp Y_d \mid X, \tag{4.3.11}$$

for all $i = 1, \ldots d$ is denoted the Poisson observation noise model of an M-selfdecomposable distribution.

The M-selfdecomposable distribution is assumed to have non-negative values for all coordinates almost surely for the definition of a Poisson observation noise model to be well-defined. A more general model would allow for

$$Y_i \mid X = x \sim \text{Pois}(\mu(x_i))$$

for a mean value function $\mu : \mathbb{R} \to [0, \infty)$. Any affine μ can be absorbed into the latent OSD distribution, and would thus be non-identifiable. However, nonlinear functions such as the ReLU, $\mu(x) = x_+$, the exponential or the softplus could be practically relevant.

We will, however, derive identification results and estimation procedures based on the cumulants of the observations, and we can only push the cumulant computations through the Poisson noise model when the latent variables enter linearly in the mean. Therefore, an important question to consider is whether there even exists a stable M such that the coordinates of the corresponding M-selfdecomposable distribution are non-negative almost surely.

An example of such matrices, is the set of all stable Metzler matrices. A Metzler matrix is a matrix M that has unconstrained diagonal and all off-diagonal entries greater than or equal to zero, $M_{ij} \geq 0$ for $i \neq j$. If M is a Metzler matrix, the matrix exponential e^{sM} is non-negative for all $t \geq 0$. Thus, if the driving Lévy process is a non-decreasing process, the M-selfdecomposable distribution defined by equation (4.2.2) for such an M and Z have all coordinates non-negative almost surely. A necessary condition for a Metzler matrix M to also be stable is that all elements on the diagonal of M are

negative. However, this is not a sufficient condition. Enforcing that M is also diagonally dominant is enough to ensure that it is both stable and Metzler.

Clearly, Metzler matrices come with the restriction that all feedback loops among the variables have to be positive feedback loops. However, the main reason for introducing Metzler matrices is not that we imagine this is the only sensible example, but instead to introduce a class of matrices for which the conditions of Definition 4.3.1 are satisfied. The definition will also be satisfied for other stable M's. Furthermore, one could also imagine M's where all the coordinates of X are not necessarily non-negative almost surely, but where they are with very high probability. If such an M was generating the latent process, one could still imagine that the Poisson observation noise model is sensible in practice.

In practice, gene regulatory networks appear to have more zeros than if it were to follow a Poisson distribution. For this reason it has been suggested in the literature to instead assume a measurement noise model corresponding to a zero-inflated Poisson [Lorch et al., 2026]. If there are only slightly more zeros we imagine that Definition 4.3.1 could still be able to approximate such a distribution in practice. The reason for introducing the Poisson observation noise model instead of the zero-inflated Poisson is that we are able to prove identifiability from the cumulants for the Poisson observation model, see Section 4.4. It would be interesting to study whether identifiability can be extended to the zero-inflated Poisson noise instead.

Figure 4.3.4 presents scatterplots for count data simulated from a latent OSD model with Poisson observation noise (see Appendix 4.D for the details behind the simulation). From this it is clear that such a distribution is highly non-Gaussian. Furthermore, if we compare this to Figure 4.1.1 which shows a similar scatterplot but for real single cell count data, these two plots do not look completely dissimilar. We take this is a positive indication that latent OSD models with Poisson observation noise could be a good choice to model such single cell count data.

4.4 Cumulants and identifiability

Recke and Hansen [2026] derived the general continuous Lyapunov tensor equation for k-th-order cumulants of OSD distributions, which for $k = 2$ gives the continuous Lyapunov equation, well known for Gaussian distributions. For Gaussian distributions, all higher-order cumulants are zero, but for non-Gaussian OSD distributions they provide constraints, which can be used to prove identifiability in OSD distributions, as done by Recke and Hansen [2026].

We first present the higher-order Lyapunov equations. Then we show how the cumulants of the underlying OSD distributions and the Poisson observation model are related. This allows us to transfer identifiability results from OSD distributions to latent OSD distributions with Poisson observation noise.

We recall the notion of an r-mode product between a tensor and a matrix. The r-mode product of a tensor $K \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_k}$ with a matrix $A \in \mathbb{R}^{J \times I_k}$, denoted $K \times_r A$, is a

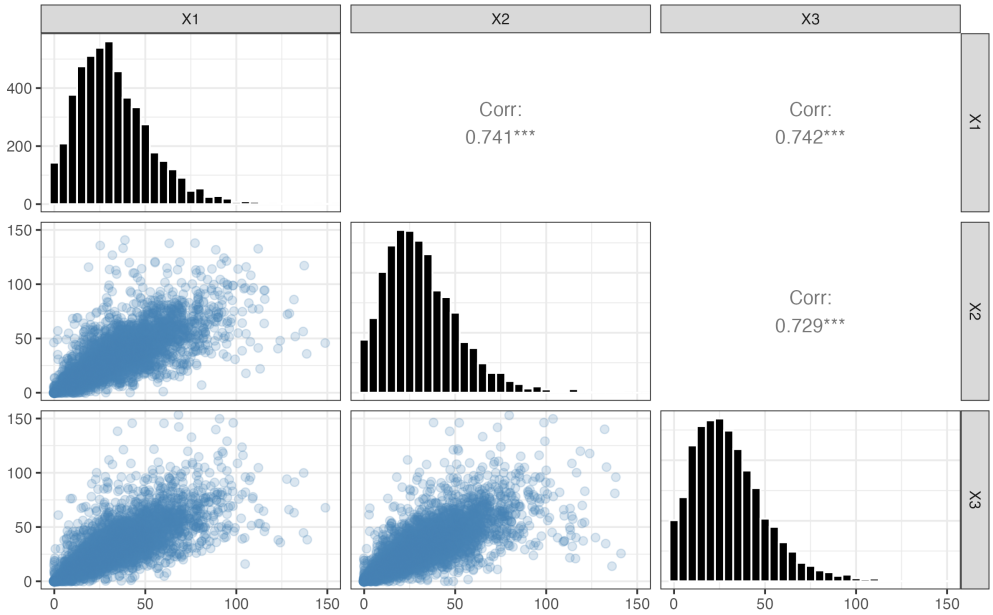


Figure 4.3.4: Scatter plots of count data from the Latent OSD model with Poisson observation noise.

$I_1 \times \cdots \times I_{r-1} \times J \times I_{r+1} \times \cdots \times I_k$ tensor with elementwise entries

$$(K \times_r A)_{i_1 \cdots i_{r-1} j i_{r+1} \cdots i_k} = \sum_{i=1}^{I_r} K_{i_1 \cdots i_{r-1} i i_{r+1} \cdots i_k} A_{ji}. \tag{4.4.12}$$

Proposition 4.4.1 (Proposition 1, Recke and Hansen [2026]). *Let M be a $d \times d$ stable matrix and let $Z = (Z_t)_{t \geq 0}$ denote a Lévy process with finite k-th moment. Then the corresponding M-selfdecomposable distribution given by (4.2.2) has finite k-th moment, and the k-th order cumulant tensor $\mathcal{K} = \mathrm{cum}_k(X)$ solves the equation*

$$\mathcal{K} \times_1 M + \ldots + \mathcal{K} \times_k M + \mathcal{C}_k = 0 \tag{4.4.13}$$

where $\mathcal{C}_k = \mathrm{cum}_k(Z_1)$ is the k-th-order cumulant tensor of Z_1.

A proof of Proposition 4.4.1 is given by Recke and Hansen [2026], but we provide an alternative derivation of (4.4.13) in Appendix 4.A.2.

Since independent variables have joint cumulants equal to 0, the allowed sparsity patterns of the cumulants of the driving Lévy process are determined by the hyperedges, H, in the mixed graph. If there is a hyperedge between $i_1, \ldots, i_k$, all cumulants using any subset of $i_1, \ldots, i_k$ of the driving Lévy process would be allowed to be non-zero.

This is not exactly the same notion as the blunt edges introduced by Recke and Hansen [2026, Definition B1], where a blunt edge of order $(i_1, \ldots, i_k)$ between k nodes only denotes that the specific k-th-order cumulant $(\mathcal{C}_k)_{i_1,\ldots,i_k}$ of the driving Lévy is allowed to be non-zero.

We are able to prove that the cumulants of the latent OSD model with Poisson observation noise can be obtained as linear combinations of the cumulants of the underlying OSD distribution and vice versa. The proof is in Appendix 4.C. The result does not depend on the latent distribution being OSD – it is only a consequence of the Poisson observation noise as defined by equation (4.3.11). Thus, it would also be possible to study the Poisson observation noise with a different latent distribution, and then it would still be possible to transfer results about identifiability from the cumulants of the latent process to identifiability from observations with Poisson noise.

Proposition 4.4.2. *Let X have an M-selfdecomposable distribution given by (4.2.2) and suppose the Lévy process has finite k-th moment. If Y follows the Poisson observation noise model as defined in Definition 4.3.1, then Y also has finite k-th moment. Furthermore, the k-th-order cumulant of X (resp. Y) can be obtained as a linear transformation of the up to k-th-order cumulant of Y (resp. X).*

A direct consequence of Proposition 4.4.2 is that any result about identifiability for M-selfdecomposable distributions obtained via its cumulants can be transferred to identifiability results about the Poisson observation noise model from its cumulants. As explained by Recke and Hansen [2026], we can never identify the parameter M up to more than a scaling, for an M-selfdecomposable distribution, unless we assume that the cumulants of the driving Lévy process are known. Heuristically, this is reasonable as we should not be able to learn the speed at which the system evolves from observations at a single time point. This can easily be seen from the k-th-order cumulant equation. If $(M, \mathcal{C}_k)$ is a solution of the k-th-order Lyapunov equation for $\mathcal{K}$ then $(cM, c\,\mathcal{C}_k)$ is also a solution with cM stable for $c > 0$. By Proposition 4.4.2 the Poisson observation model will therefore also, at most, allow us to identify M up to scaling.

By Theorem 1 and Corollary 1 by Recke and Hansen [2026] we see that we also obtain the best case scenario for the Poisson observation noise model of generic identifiability of M and diagonal error cumulants $(\mathcal{C}_2, \mathcal{C}_r)$ up to a joint scaling from the up to r-th-order cumulants. The notion of generic will be taken to mean; for all parameters in the parameter set, except a smaller dimensional set defined by the vanishing of a set of polynomials, see Recke and Hansen [2026] for more details and an explanation of why this assumption is necessary.

Corollary 4.4.3. *Let $d \geq 2$ and $r \geq 3$ be integers and $G = ([d], E)$ a connected graph with all self-loops. Let $M \in \mathbb{R}^E_{\text{stab}}$ define a M-selfdecomposable distribution with the driving Lévy process having diagonal cumulants, $\mathcal{C}_2$ and $\mathcal{C}_r$ of order 2 and r, respectively. Then $(M, \mathcal{C}_2, \mathcal{C}_r)$ is generically identifiable up to a joint scaling by a positive real number from the first r cumulants of the Poisson observation noise model, Definition 4.3.1.*

The assumptions about the cumulants $\mathcal{C}_2$ and $\mathcal{C}_r$ being diagonal is, for example, satisfied if all the coordinates of the driving Lévy process are independent. Thus, in the mixed graph representation, a way to satisfy the conditions of the corollary would be that $\mathcal{G} = ([d], E, H)$, where $([d], E)$ represents a connected directed graph with all self-loops, and the set of hyperedges would be the set of all singletons $H = \{\{1\}, \ldots, \{d\}\}$.

If the directed graph is not connected, we only have identifiability up to separate scaling for each connected component [Recke and Hansen, 2026, Corollary 2]. Intuitively, this is sensible, since if the graph is disconnected, each separate subsystem corresponding to a connected component will not interact with any other subsystem, and the process could therefore have developed at different speeds in each of them.

4.5 Interventional Cumulants

By combining Propositions 4.2.3 and 4.4.1 we can show that all cumulants of the steady-state interventional distributions also satisfy a continuous Lyapunov equation.

Proposition 4.5.1. *If M_{22} is a stable matrix, the steady-state interventional distribution of $X^{(2)}$ is M_{22}-selfdecomposable, and if $Z_1^{(2)}$ has finite k-th order moment then the k-th order cumulant tensor $\mathcal{K}^{(2)}$ of $X^{(2)}$ having distribution $P^{\mathrm{do}(X^{(1)}=x^{(1)})}$ satisfies the equation*

$$\mathcal{K}^{(2)} \times_1 M_{22} + \ldots + \mathcal{K}^{(2)} \times_k M_{22} + \mathcal{C}_k^{(2)} = 0 \tag{4.5.14}$$

where $\mathcal{C}_k^{(2)} = \mathrm{cum}_k(\tilde{Z}_1^{(2)})$ and $\tilde{Z}_t^{(2)} \overset{\mathcal{D}}{=} Z_t^{(2)} + M_{21}x^{(1)}t$. Specifically, if $Z_1^{(2)}$ has finite second moment then

$$\mathbb{E}^{\mathrm{do}(X^{(1)}=x^{(1)})}(X^{(2)}) = -(M_{22})^{-1}(\mathbb{E}(Z_1^{(2)}) + M_{21}x^{(1)}) \tag{4.5.15}$$

$$\mathrm{Var}^{\mathrm{do}(X^{(1)}=x^{(1)})}(X^{(2)}) = \Sigma^{(2)} = \int_0^\infty e^{sM_{22}}\mathcal{C}_2^{(2)}e^{sM_{22}^T}\mathrm{d}s \tag{4.5.16}$$

where $\mathcal{C}_2^{(2)} = \mathrm{Var}(Z_1^{(2)})$.

Remark 4..2. The Lyapunov equation for the interventional covariance matrix $\Sigma^{(2)}$ can also be written as

$$M_{22}\Sigma^{(2)} + \Sigma^{(2)}M_{22}^T + \mathcal{C}_2^{(2)} = 0,$$

and the integral representation (4.5.16) is thus solving this equation.

Note that if M_{22} is *not* stable, the intervened SDE (4.2.5) will generally not have a steady-state distribution and the interventional distribution is undefined.

When M_{22} is stable, the formulas (4.5.15) and (4.5.16) are superficially similar to the formulas for the (observational) conditional mean and covariance in the multivariate Gaussian distribution. The interventional mean is, e.g., an affine function of $x^{(1)}$ and the interventional covariance is independent of $x^{(1)}$. There are, however, important differences. In the non-Gaussian case, the first two interventional moments (4.5.15) and (4.5.16) do *not* characterize the entire interventional distribution. Additionally, even in the Gaussian case, the interventional mean and covariance only coincide with the observational conditional mean and covariance in special cases, see Proposition 5 by Lauritzen and Richardson [2002].

Appendix for 'Causal inference with latent operator-selfdecomposable distributions'

4.A. Operator-selfdecomposable distributions

This appendix includes historical and probabilistic facts about *operator-selfdecomposable* (OSD) distributions. We first recall some classical terminology. A random variable $X \in \mathbb{R}^d$ is said to be *decomposable* if $X \stackrel{\mathcal{D}}{=} X' + X''$ for independent X' and X''. If it is possible to take $X' = \rho X$ in this additive decomposition for any $\rho \in (0, 1)$, the distribution of X is said to be *selfdecomposable*. These distributions appear in classical probability theory as limits of centered and rescaled sums of independent random variables.

OSD distributions can be defined similarly with the number ρ replaced by an operator (a matrix) of the form e^{tM} for $t \geq 0$. The condition $\log(\rho) < 0$ translates into M being a stable matrix, that is, all its eigenvalues have strictly negative real part. We then say that X has an M-selfdecomposable distribution if there for all $t > 0$ exists a random variable X_t independent of X such that

$$X \stackrel{\mathcal{D}}{=} e^{tM} X + X_t. \tag{4.A.1}$$

If we take $M = -I$ with I the $d \times d$ identity matrix, (4.A.1) reduces to

$$X \stackrel{\mathcal{D}}{=} e^{-t} X + X_t.$$

for all $t \geq 0$, and we see that all seldecomposable distributions are OSD.

The class of OSD distributions were first studied by Urbanik [1972] under the name "Lévy's probability measures". Independently, Jurek [1982] and Wolfe [1982] showed that the distributional property (4.A.1) is equivalent to the stochastic integral representation (4.2.2) for some Lévy process Z. Sato and Yamazato [1984] gave several different characterizations of OSD distributions, including their representation as steady-state distributions of a Markov process as stated in Proposition 4.2.2. The book by Zbigniew and David [1993] collects many of the known probabilistic results about OSD distributions.

4.A.1 OSD distributions and infinite divisibility

The property of being (operator) selfdecomposable is closely linked to the property of being infinitely divisible. For once because the stochastic integral representation (4.2.2) is in terms of a Lévy process Z with the distribution of Z_1 being infinitely divisible.

By the Lévy-Khintchine representation of the infinitely divisible distribution of Z_1, its characteristic function is given as $\varphi_{Z_1}(z) = \mathbb{E}(e^{iz^T Z_1}) = e^{\Psi(z)}$ with the Lévy exponent

$$\begin{aligned}\Psi(z) &= iz^T\gamma - \frac{1}{2} z^T A z + \int e^{iz^T u} - 1 - iz^T u 1_D(u)\ \nu(\mathrm{d}u) \\ &= iz^T\gamma - \frac{1}{2} z^T A z + \int g(z,u)\ \nu(\mathrm{d}u).\end{aligned}$$

Here, $g(z,u) = e^{iz^Tu} - 1 - iz^Tu1_D(u)$, and $D = \{u \in \mathbb{R}^d \mid \|u\| \leq 1\}$ denotes the unit ball.

Theorem 4.1 by Sato and Yamazato [1984] gives that if X is given by the stochastic integral (4.2.2), then the characteristic function of X has the form

$$\varphi_X(z) = \exp\left(\int_0^\infty \Psi(e^{sM^T}z)\mathrm{d}s\right). \tag{4.A.2}$$

We can write out the exponent as

$$\begin{aligned}\Psi_X(z) &= iz^T \underbrace{\int_0^\infty e^{sM}\mathrm{d}s\gamma}_{\tilde{\gamma}} - \frac{1}{2}z^T \underbrace{\int_0^\infty e^{sM}Ae^{sM^T}\mathrm{d}s}_{\tilde{A}} z + \int\int_0^\infty g(z, e^{sM}u)\mathrm{d}s\nu(\mathrm{d}u) \\ &= iz^T\tilde{\gamma} - \frac{1}{2}z^T\tilde{A}z + \int g(z,u)\tilde{\nu}(\mathrm{d}u),\end{aligned}$$

where $\tilde{\nu}$ is the Lévy measure given by

$$\tilde{\nu}(B) = \int\int_0^\infty 1_B(e^{sM}u)\mathrm{d}s\nu(\mathrm{d}u). \tag{4.A.3}$$

This shows that OSD distributions are infinitely divisible, see Figure 4.2.3.

4.A.2 Cumulants of infinitely divisible distributions

If X is infinitely divisible with characteristic function $\varphi_X(z) = \mathbb{E}(e^{iz^TX}) = e^{\Psi_X(z)}$ and thus Lévy exponent Ψ_X, the cumulants can be defined as

$$\mathcal{K}_k = (-i)^k D^k \Psi_X(0) \tag{4.A.4}$$

provided that Ψ_X is k times differentiable in 0. Using the Lévy-Khintchine representation, we get the following formulas, given appropriate integrability conditions w.r.t. the Lévy measure, for the cumulants:

$$\begin{aligned}\mathcal{K}_1 &= \mathbb{E}(X) = \gamma + \int u1_{D^c}(u)\nu(\mathrm{d}u) \\ \mathcal{K}_2 &= \Sigma = A + \int u \otimes u\ \nu(\mathrm{d}u) \\ \mathcal{K}_3 &= \int u \otimes u \otimes u\ \nu(\mathrm{d}u) \\ &\vdots \quad \vdots \qquad\quad \vdots \\ \mathcal{K}_k &= \int \underbrace{u \otimes u \otimes \ldots \otimes u}_{k \text{ factors}} \nu(\mathrm{d}u).\end{aligned}$$

Using the representation of the Lévy measure $\tilde{\nu}$ given by (4.A.3) for an OSD distribution and the mode product notation defined by (4.4.12), we find that the k-th order cumulant tensor for OSD distributions for $k \geq 3$ has the integral representation

$$\begin{aligned}\mathcal{K}_k &= \int_0^\infty \int (e^{sM}u) \otimes (e^{sM}u) \otimes \ldots \otimes (e^{sM}u)\nu(\mathrm{d}u)\mathrm{d}s \\ &= \int_0^\infty \int u \otimes u \otimes \ldots \otimes u\nu(\mathrm{d}u) \times_1 e^{sM} \times_2 e^{sM} \ldots \times_k e^{sM}\mathrm{d}s \\ &= \int_0^\infty \mathcal{C}_k \times_1 e^{sM} \times_2 e^{sM} \ldots \times_k e^{sM}\mathrm{d}s,\end{aligned}$$

provided that the k-th order cumulant, $\mathcal{C}_k$, of Z_1 is finite.

For $k = 2$, the covariance matrix of X is given as

$$\begin{aligned}\Sigma &= \int_0^\infty e^{sM}Ae^{sM^T} + e^{sM}\int u \otimes u\ \nu(\mathrm{d}u)e^{sM^T}\,\mathrm{d}s \\ &= \int_0^\infty e^{sM}\underbrace{\left(A + \int u \otimes u\ \nu(\mathrm{d}u)\right)}_{=\mathcal{C}_2} e^{sM^T}\,\mathrm{d}s = \int_0^\infty e^{sM}\mathcal{C}_2 e^{sM^T}\,\mathrm{d}s,\end{aligned}$$

where $\mathcal{C}_2$ is the covariance matrix of Z_1. Finally, for $k = 1$,

$$\begin{aligned}\mathbb{E}(X) &= \int_0^\infty e^{sM}\gamma + e^{sM}\int u1_{D^c}\nu(\mathrm{d}u)\mathrm{d} \\ &= \int_0^\infty e^{sM}\mathrm{d}s\mathbb{E}(Z_1) = -M^{-1}\mathbb{E}(Z_1).\end{aligned}$$

With reference to Corollary 3.1 by Xu and Wang [2022], the above integral representation of $\mathcal{K}_k$ shows that it solves equation (4.4.13). Thus, the above derivation provides an alternative proof of Proposition 4.4.1.

4.A.3 Operator-stable distributions

The set of operator-stable distributions is a subset of the OSD distributions with multiple different characterizations. We give one in this appendix that is easy to state in the context of the present paper.

Suppose that X has an infinitely divisible distribution and satisfies (4.2.1), then $X = X_1$ for a Lévy process $(X_t)_{t\geq 0}$. Corollary 4.4.4 by Zbigniew and David [1993] then gives that the distribution of X is M-stable if and only if

$$X \overset{\mathcal{D}}{=} \int_0^\infty e^{sM}\mathrm{d}X_s + x \tag{4.A.5}$$

for some $x \in \mathbb{R}^d$. An infinitely divisible distribution fulfilling (4.2.1) is then called *operator-stable* if it is M-stable for some M. It follows from (4.A.5) that all operator-stable distributions are OSD. The $(-I)$-stable distributions are what is classically known as the set of stable distributions [Zbigniew and David, 1993, Section 4.14].

Two well known examples of (operator) stable distributions are the Gaussian and Cauchy distributions. The Lévy exponent for the Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ is the quadratic function

$$\Psi_{\text{Gauss}}(z) = iz^T\mu - \frac{1}{2}z^T\Sigma z.$$

The Lévy exponent for the Cauchy distribution with location parameter μ and scale matrix Σ is

$$\Psi_{\text{Cauchy}}(z) = iz^T\mu - \sqrt{z^T\Sigma z}.$$

Neither of these are of particular interest for this paper. For the Gaussian distributions, all cumulants beyond the covariance matrix are zero, while the Lévy exponent for the Cauchy distributions is non-differentiable in 0, reflecting that none of the cumulants exist. In fact, except for the Gaussian case, the operator-stable distributions have a limited number of finite moments [Zbigniew and David, 1993, Section 4.12].

4.A.4 Elliptical distributions

Some OSD distributions are also elliptical, which gives them additional properties. These distributions are, for example, straightforward to simulate from.

The distribution of X is elliptical if its characteristic function is of the form

$$\phi_X(z) = e^{iz^T\mu}\psi(z^T\Sigma z), \qquad z \in \mathbb{R}^d,$$

for a location parameter $\mu \in \mathbb{R}^d$, a positive semi-definite matrix Σ and a scalar function $\psi : \mathbb{R} \to \mathbb{C}$. The Gaussian and Cauchy distributions are seen to be elliptical directly from this definition. The definition is equivalent to X having the representation

$$X = \mu + AV$$

where A is $d \times k$, $\Sigma = AA^T$ (with rank $k \leq d$) and $V \in \mathbb{R}^k$ has a *spherical* distribution with characteristic function

$$\phi_V(z) = \psi(\|z\|_2^2), \qquad z \in \mathbb{R}^k.$$

Considering in the following $k = d$, V is spherical if and only if it has the representation

$$V = RU$$

with R and U independent, with U uniformly distributed on the unit sphere S^{d-1}, and $R \geq 0$ having some distribution . See Theorem 1 by Cambanis et al. [1981]. The Gaussian distribution corresponds to $\psi(t) = e^{-t/2}$, and R^2 having a $\chi^2(p)$-distribution. The multivariate t-distribution with $\nu > 0$ degrees of freedom is elliptical as well, and it admits the representation $\mu + RAU$ with R^2/p being $F(p, \nu)$-distributed. The expression for the characteristic function is a bit complicated.

Masuda [2004] showed in their Proposition 5.2 that the symmetric generalized hyperbolic distributions are OSD, and it follows from their characteristic function that they are also elliptical. The multivariate t-distribution is a special case.

Marginal and conditional distributions of an elliptical distribution are again elliptical, and if the conditional distribution has finite mean,

$$\mathbb{E}(X^{(2)} \mid X^{(1)} = x^{(1)}) = \xi^{(2)} + \Sigma_{21}(\Sigma_{11})^{-1}(x^{(1)} - \xi^{(1)})$$

just as for the Gaussian distribution, see Corollary 5 by Cambanis et al. [1981]. Note that even if Σ is the covariance matrix of X and if the conditional distribution has finite variance, the conditional covariance matrix *does not* generally coincide with the conditional covariance in the Gaussian distribution.

4.B. Interventional notation

This section elaborates on the choice of notation for the interventional distributions.

The driving Lévy process $Z = (Z_t)_{t\geq 0}$ is formally defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The solution of (4.2.4) can be written explicitly as

$$X_t = e^{tM}X_0 + \int_0^t e^{(t-s)M}\mathrm{d}Z_s. \tag{4.B.1}$$

By stability of M, $e^{tM}X_0 \to 0$ for $t \to \infty$, and since Z is a Lévy process, specifically that it has independent and stationary increments,

$$\int_0^t e^{(t-s)M}\mathrm{d}Z_s \overset{\mathcal{D}}{=} \int_0^t e^{sM}\mathrm{d}Z_s \to X = \int_0^\infty e^{sM}\mathrm{d}Z_s$$

for $t \to \infty$. Therefore, we have that $X_t \overset{\mathcal{D}}{\to} X$. It is fairly straightforward to show that the distribution of X is a unique steady-state distribution. Thus, we can represent the stochastic process as well as its steady-state distribution in terms of transformations of the Lévy process and thus as random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

The interventional distributions defined in Section 4.2.2 are defined via manipulation of the stochastic differential equation. If M_{22} in (4.2.5) is stable, the solution of (4.2.5) has a limiting steady-state distribution that could be represented as

$$X^{(2),\mathrm{do}(X^{(1)}=x^{(1)})} = \int_0^\infty e^{sM_{22}}(\mathrm{d}Z_s^{(2)} + M_{21}x^{(1)}\mathrm{d}s).$$

This is likewise a transformation of the Lévy process, though depending only on $Z^{(2)}$ (and the value $x^{(1)}$), and the interventional variable is again defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This notation is a *potential outcome notation* defined via manipulations of the SDE when regarded as a *structural causal model*. In this notation we distinguish different interventional distributions via the annotations on the different random variables, which are all defined on a common probability space.

In a stochastic process context, annotation of all the random variables becomes heavy notationally. For this reason we prefer to use the same random variable to denote the (steady-state) outcome under multiple different interventions, and to let the context

determine the considered interventional distribution. To achieve this formally, we modify and annotate the probability space according to the intervention. Thus we let $\tilde{Z}$ denote a Lévy process defined on a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mathbb{P}^{\mathrm{do}(X^{(1)}=x^{(1)})})$ such that $\tilde{Z} \overset{\mathcal{D}}{=} (Z_t^{(2)} + M_{21}x^{(1)}t)_{t\geq 0}$. In terms of $\tilde{Z}$ we define, if M_{22} is stable, the random variable

$$X^{(2)} = \int_0^\infty e^{sM_{22}} \mathrm{d}\tilde{Z}_s,$$

whose OSD distribution is the unique steady-state distribution of the solution to the intervened SDE (4.2.5). One standard construction of $\mathbb{P}^{\mathrm{do}(X^{(1)}=x^{(1)})}$ is as the distribution of $(Z_t^{(2)} + M_{21}x^{(1)}t)_{t\geq 0}$ on its sample space, but its construction is immaterial as long as such a construction exists. This notation is purely *distributional* and akin to notation used with *causal graphical models*.

Distributionally, the notations are equivalent. We write, for example,

$$\mathbb{P}^{\mathrm{do}(X^{(1)}=x^{(1)})}(X^{(2)} \in A)$$

for the probability that $X^{(2)}$ falls in A when we fix $X_t^{(1)} = x^{(1)}$ for all $t \geq 0$. In the potential outcome notation, we would denote this probability as

$$\mathbb{P}(X^{(2),\mathrm{do}(X^{(1)}=x^{(1)})} \in A).$$

Likewise,

$$\mathbb{E}^{\mathrm{do}(X^{(1)}=x^{(1)})}(X^{(2)})$$

is the expectation under the fixing-intervention, while in the potential outcome notation this would be

$$\mathbb{E}(X^{(2),\mathrm{do}(X^{(1)}=x^{(1)})}).$$

In either case, the interventional distribution, $P^{(2),\mathrm{do}(X^{(1)}=x^{(1)})}$, is given as a push-forward measure

$$P^{(2),\mathrm{do}(X^{(1)}=x^{(1)})} = X^{(2),\mathrm{do}(X^{(1)}=x^{(1)})}(\mathbb{P}) = X^{(2)}(\mathbb{P}^{\mathrm{do}(X^{(1)}=x^{(1)})})$$

in slightly different ways depending on which notation we use.

4.C. Proofs

Proof of Proposition 4.2.3. Since the deterministic process $t \mapsto M_{21}x^{(1)}t$ is a Lévy process, independent of $Z^{(2)}$, the sum $Z_t^{(2)} + M_{21}x^{(1)}t$ is again a Lévy process. We then show that $\tilde{Z}^{(2)}$ fulfills (4.2.1). By the triangle inequality

$$\|\tilde{Z}_1^{(2)}\| \overset{\mathcal{D}}{=} \|Z_1^{(2)} + M_{21}x^{(1)}\| \leq \|Z_1^{(2)}\| + \|M_{21}x^{(1)}\|,$$

and using monotonicity and sub-additivity of $x \mapsto \log(1+x)$, we get that

$$\begin{aligned}\mathbb{E}(\log(1 + \|\tilde{Z}_1^{(2)}\|)) &\leq \mathbb{E}(\log(1 + \|Z_1^{(2)}\|)) + \log(1 + \|M_{21}x^{(1)}\|) \\ &\leq \mathbb{E}(\log(1 + \|Z_1\|)) + \log(1 + \|M_{21}x^{(1)}\|) < \infty.\end{aligned}$$

When M_{22} is stable, the conclusion of the proposition is now a direct consequence of Definition 4.2.1 and Proposition 4.2.2. □

Proof of Proposition 4.4.2. The result is essentially just a direct application of the law of total cumulance [Brillinger, 1969]. To exemplify how it will work for higher-order cumulants we obtain first that the mean values of X_i and Y_j are equal

$$\mathbb{E}[Y_i] = \mathbb{E}[Y_i \mid X] = \mathbb{E}[X_i].$$

Furthermore, by the law of total variance (the law of total cumulance for two variables), the covariance of Y_i and Y_j is

$$\operatorname{cov}(Y_i, Y_j) = \mathbb{E}[\operatorname{cov}(Y_i, Y_j \mid X)] + \operatorname{cov}(\mathbb{E}[Y_i \mid X], \mathbb{E}[Y_j \mid X]) = \begin{cases} \mathbb{E}(X_i) + \operatorname{var}(X_i) & \text{if } i = j \\ \operatorname{cov}(X_i, X_j) & \text{if } i \neq j \end{cases},$$

since if Y_i and Y_j are distinct they are independent given X according to the Poisson observation model.

If we let $\kappa(Y_1, \ldots, Y_n)$ denote the joint cumulant of n random variables, then the law of total cumulance for the k-th-order cumulants of k coordinates of Y (potentially with repetitions) is

$$\kappa(Y_{i_1}, \ldots, Y_{i_k}) = \sum_{\pi} \kappa(\kappa(Y_{i_j} \ : i_j \in B \mid X) : B \in \pi), \tag{4.C.1}$$

where the sum runs over all partitions π of $\{1, \ldots, k\}$ and $B \in \pi$ means that the term for each π is a product of $\kappa(Y_i \ : i \in B \mid X)$ where B runs over all blocks of the given partition π.

For any order cumulant, $\kappa(Y_{i_1}, \ldots, Y_{i_l} \mid X) = 0$ whenever at least two of the indices $i_1, \ldots, i_l$ are different, because the Poisson observation model implies that two distinct coordinates of Y are conditionally independent given X. Using this and that all cumulants in the Poisson distribution are equal to the mean, the law of total cumulance can be simplified as

$$\kappa(Y_{i_1}, \ldots, Y_{i_k}) = \sum_{\pi_i} \kappa(X_{i_{\pi_i(1)}}, \ldots, X_{i_{\pi_i(b)}}), \tag{4.C.2}$$

where the sum is over partitions π_i of $i_1, \ldots, i_k$ where no block B_j in π_i contains two indices that are different, $\pi_i(j)$ denotes the index to which all elements in the j-th block, B_j, are equal, and b is the number of blocks in partition the π_i. The sum is never empty because it will always contain the partition where each block only contains a single element.

Thus, the k-th-order cumulant of Y can be written as a linear combination of the up to k-th-order cumulant of X. The equations for the cumulants of Y can easily be solved for the cumulants of X, yielding that the cumulants of X are also given as a linear combination of the cumulants of Y.

□

The proof contains the explicit formulas for going back and forth in the mean and covariance case. For the sake of implementation we write out equation (4.C.2) for the case $k = 3$ as well

$$\kappa_3(Y_{i_1}, Y_{i_2}, Y_{i_3}) = \begin{cases} \mathbb{E}(X_{i_1}) + 3\text{var}(X_{i_1}) + \kappa_3(X_{i_1}, X_{i_1}, X_{i_1}) & \text{if } i_1 = i_2 = i_3 \\ \text{cov}(X_{i_1}, X_{i_3}) + \kappa_3(X_{i_1}, X_{i_1}, X_{i_3}) & \text{if } i_1 = i_2 \neq i_3 \\ \kappa_3(X_{i_1,}, X_{i_2}, X_{i_3}) & \text{if } i_1 \neq i_2, i_2 \neq i_3, i_1 \neq i_3, \end{cases}.$$

where κ_j denotes the j-th-order cumulant. Since the cumulants are symmetric the second case above covers any combination of two of the indices being equal but not equal to the last one.

Proof of Proposition 4.5.1. The first part of the proof follows directly from a combination of Proposition 4.2.3 and Proposition 4.4.1.

When M_{22} is stable, the steady-state distribution of the solution to the interventional SDE (4.2.5) gives that the interventional distribution of $X^{(2)}$ is represented by

$$\begin{aligned} X^{(2)} &\overset{\mathcal{D}}{=} \int_0^\infty e^{sM_{22}} \mathrm{d}(Z_s^{(2)} + M_{21}x^{(1)}s) && (4.C.3) \\ &= \int_0^\infty e^{sM_{22}} \mathrm{d}Z_s^{(2)} + \int_0^\infty e^{sM_{22}} \mathrm{d}s M_{21}x^{(1)} \\ &= \int_0^\infty e^{sM_{22}} \mathrm{d}Z_s^{(2)} - (M_{22})^{-1}M_{21}x^{(1)}, && (4.C.4) \end{aligned}$$

where we have used that when M_{22} is stable,

$$\int_0^\infty e^{sM_{22}} \mathrm{d}s = (M_{22})^{-1} e^{sM_{22}} \Big|_0^\infty = -(M_{22})^{-1}.$$

When $Z^{(2)}$ has finite first moment, the expectation of (4.C.4) gives

$$\begin{aligned} \mathbb{E}^{\text{do}(X^{(1)}=x^{(1)})}(X^{(2)}) &= \int_0^\infty e^{sM_{22}} a^{(2)} \mathrm{d}s - (M_{22})^{-1}M_{21}x^{(1)} \\ &= -(M_{22})^{-1}(a^{(2)} + M_{21}x^{(1)}). \end{aligned}$$

If $Z^{(2)}$ has finite second moment, the matrix

$$\Sigma^{(2)} = \Sigma^{(2)} = \int_0^\infty e^{sM_{22}} \mathcal{C}_2^{(2)} e^{sM_{22}^T} \mathrm{d}s$$

is seen to solve Lyapunov equation for the 2-nd order cumulant, that is, the covariance matrix, whence $\text{Var}^{\text{do}(X^{(1)}=x^{(1)})}(X^{(2)}) = \Sigma^{(2)}$. □

4.D. Numerical details

4.D.1 Data example

The scatter plots in Figure 4.1.1 are of data from https://doi.org/10.5281/zenodo.7262328, made available as supporting data for the paper by Gorin et al. [2022]. The specific data used is from the file `allen_F08_Glutamatergic.loom` in the `loom` folder. The files were read into R with the loomR package. We used only the count data available in this file for the spliced genes. The file contains single cell count data for 80 genes, selected by Gorin et al. [2022], from 5892 cells.

For the plot in Figure 4.1.1 we selected the three genes for which the count data had the largest Pearson correlation.

4.D.2 Simulation of OSD distributions

Observations from an OSD distribution can be simulated via the stochastic integral representation (4.2.2). Suppose that M can be diagonalized as $M = QDQ^{-1}$, where $D = \operatorname{diag}(\delta_1, \ldots, \delta_d)$ is a diagonal matrix of complex eigenvalues, then

$$X = \int_0^\infty e^{sM} \mathrm{d}Z_s = \int_0^\infty Q e^{sD} Q^{-1} \mathrm{d}Z_s = Q \begin{pmatrix} \int_0^\infty e^{\delta_1 s} \mathrm{d}\tilde{Z}_s^{(1)} \\ \vdots \\ \int_0^\infty e^{\delta_d s} \mathrm{d}\tilde{Z}_s^{(d)} \end{pmatrix} \tag{4.D.1}$$

where $\tilde{Z}_s^{(i)} = (Q^{-1} Z_s)_i$. Note that for $s_1 < s_2$,

$$\tilde{Z}_{s_2} - \tilde{Z}_{s_1} = Q^{-1}(Z_{s_2} - Z_{s_1}).$$

If we can simulate the increments $\Delta_k Z = Z_{s_k} - Z_{s_{k-1}}$ for a grid $s_0 = 0 < s_1 < \ldots < s_n$, we can compute the approximations

$$\int_0^\infty e^{\delta_i s} \mathrm{d}\tilde{Z}_s^{(i)} \approx \sum_{k=1}^n e^{\delta_i s_k} (Q^{-1} \Delta_k Z)_i$$

for $i = 1, \ldots, d$. We make the following remarks about simulations based on the above approximation.

- If the Lévy process Z is a pure jump process, we can choose the grid to be the jump times. Then $\Delta_k Z$ is the k-th jump.
- We do not need to simulate the entire sample path of Z, only the increments $\Delta_k Z$, which are independent. For some Lévy processes, the increments have a well known distribution that is easy to simulate from.
- For the IE-OSD distributions, where the coordinates of the Lévy process are independent, we can simulate each coordinate $(\Delta_k Z)_1, \ldots, (\Delta_k Z)_d$ independently.

The simulated data for $d = 3$ presented in Figure 4.3.4 were generated with the following settings: The three coordinates of the Lévy process Z were independent and identically distributed compound Poisson processes. Their rate parameters were 5. The jump distributions were Gamma distributions with shape parameter 0.1 and scale parameter 10. The matrix M was symmetric and chosen so that all three coordinates of the OSD distributed had correlation 0.8. Given the observation x_i for the i-th coordinate, the count data were generated as

$$Y_i \mid X_i = x_i \sim \text{Pois}((3 \cdot x_i - 10)_+).$$

5 Completions to discrete probability distributions in log-linear models

May Cai, Cecilie Olesen Recke, Thomas Yahl

Abstract

Completion problems, of recovering a point from a set of observed coordinates, are abundant in applications to image reconstruction, phylogenetics, and data science. We consider a completion problem coming from algebraic statistics: to describe the completions of a point to a probability distribution lying in a given log-linear model. When there are finitely many completions, we show that these points either have a unique completion or two completions to the log-linear model depending on the set of observed coordinates.

Keywords: Log-linear model, toric variety, completion, algebraic moment map, semialgebraic set, algebraic boundary
MSC2020 subject classification: Primary: 62R01, Secondary: 14M25, 14P10,

5.1 Introduction

We consider the problem of recovering a probability distribution from partial information. This may occur as an imperfect sampling method may prevent one from observing or distinguishing certain outcomes and thus, it may be that probabilities are only known for certain outcomes. With a priori knowledge that the probability distribution belongs to a specified statistical model, the known probabilities may be used to compute the probability of each outcome. In this case, we say the original probability distribution may be recovered, or completed.

A problem of this form is known as a *probability completion problem*. We describe the general setting more precisely. We consider discrete probability distributions with outcome states $[n] = \{1, \ldots, n\}$ for a fixed positive integer n. Such a probability distribution may be represented by a tuple $(p_1, \ldots, p_n) \in \mathbb{R}^n$, whose i-th coordinate p_i is the probability of outcome $i \in [n]$. The probability simplex $\Delta_{n-1} \subseteq \mathbb{R}^n$ is the set of these discrete probability distributions and a *statistical model* $\mathcal{M} \subseteq \Delta_{n-1}$ is a subset of the probability simplex. A fixed subset of the states $E \subseteq [n]$ will index the probabilities which are to be known or observed, and we consider the coordinate projection $\pi_E : \mathcal{M} \to \mathbb{R}^E$ to the coordinates indexed by E. A point $p_E \in \mathbb{R}^E$ is called a *partial*

observation and a probability distribution in the fiber $p \in \pi_E^{-1}(p_E)$ is a *completion* of the partial observation p_E to the model $\mathcal{M}$. One looks to describe the fiber $\pi_E^{-1}(p_E)$, which is the set of completions of the partial observation p_E to $\mathcal{M}$. In addition to enumerating the completions of a partial observation, one may look to explicitly describe the *completable region* $\pi_E(\mathcal{M})$—the set of partial observations which may be completed to $\mathcal{M}$.

Probability completion problems were first considered in Kahle et al. [2017]; Kubjas and Rosen [2017], where they study completions to the independence model of two or more random variables. As the independence model is the intersection of the space of rank one tensors and the probability simplex, this may be regarded as a type of low rank tensor completion problem as in [Singer and Cucuringu, 2010; Király et al., 2012, 2015; Bernstein et al., 2020]. More generally, probability completion problems may be understood as problems in compressed sensing as described in Breiding et al. [2023].

We demonstrate an example of a probability completion problem: the Hardy-Weinberg curve used in genetics is a statistical model whose probability distributions represent the probabilities of passing on certain traits from parents to their offspring. In the case of a trait with a dominant gene X and a recessive gene Y, there are three genotypes that can be passed on—these are the homozygous combinations XX and YY, and the heterozygous combination XY. Using the variables x and y for the probabilities of passing on the homozygous combinations XX and YY respectively, and the variable z for the probability of passing on the heterozygous combination XY, the Hardy-Weinberg model $\mathcal{M}$ of possible probability distributions is defined by the equations $z^2 - 4xy = 0$ and $x + y + z = 1$, where all coordinates are non-negative. This curve is depicted in Figure 5.1.1.

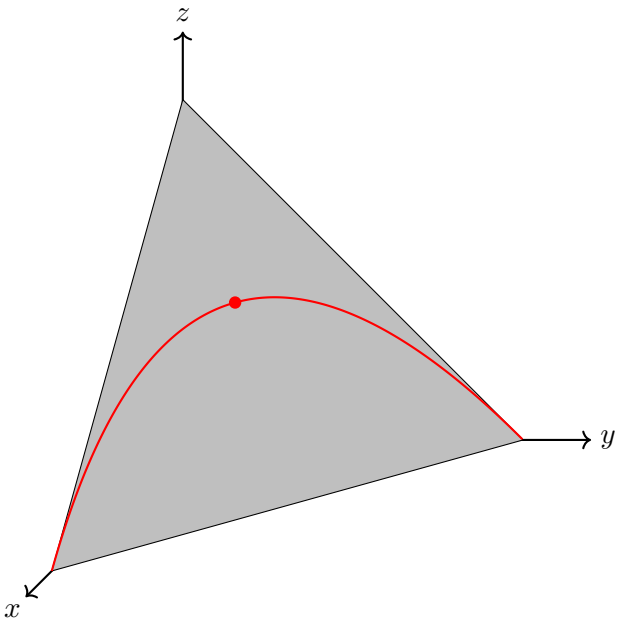


Figure 5.1.1: The Hardy-Weinberg curve

The projections onto either the first coordinate or the second coordinate are injective and the image in both cases is the interval $[0, 1]$. That is, a point on the Hardy-Weinberg curve is uniquely completable from its first coordinate or its second coordinate. Equivalently, if the probability of either homozygous combination XX or YY being passed on is known, all probabilities can be determined. However, the projection onto the third coordinate is a 2-to-1 mapping for $0 \leq z < 1/2$ and the image of the projection is $[0, 1/2]$. Thus, from a known z coordinate less than $1/2$, there are two completions to a point of the Hardy-Weinberg curve. Equivalently, given that the probability of the heterozygous combination XY being passed on is known and less than $1/2$ there are two possible probabilities for the homozygous combinations XX and YY to be passed on.

We are concerned with probability completion problems where the statistical model $\mathcal{M}$ is a log-linear model, the restriction of a toric variety to the probability simplex. The class of log-linear models encompasses many well-studied discrete models such as discrete graphical models, hierarchical models, and staged-tree models [Hoşten and Sullivant, 2002; Hoşten and Sullivant, 2007; Görgen et al., 2022]. Log-linear models are also useful as their Markov bases are understood and may be used in sampling algorithms as described in Diaconis and Sturmfels [1998]; Sullivant [2018].

Our main contributions are in showing that the behaviour demonstrated for the Hardy-Weinberg curve is typical. Precisely, we prove that for a log-linear model $\mathcal{M}$ and suitable set of observed outcomes $E \subseteq [n]$, there is either a unique completion or there are two completions for every point with non-zero coordinates in the completable region. This highlights a remarkable property of our probability completion problems–the number of solutions to the semi-algebraic problem is bounded by two, whereas the number of com-

plex solutions may be arbitrarily large. We use the relationship between toric varieties and polyhedral geometry to identify precisely when there is one completion and when there are two completions. These results are given in Theorem 5.4.13 and Theorem 5.4.16.

In addition, we give a description of the boundary and interior of the completable region in Theorem 5.4.11. We use this result to provide an algorithm for computing the defining equations for the boundary of the completable region. These defining equations often allow one to compute an explicit semialgebraic description of the completable region, as discussed in Section 5. We illustrate these results with several examples, some of which are relevant for applications.

We begin by giving background on toric varieties and log-linear models in Section 5.2. In Section 5.3, we describe the completion problem to a given toric variety and include necessary results for future sections. We present our completion results to a given log-linear model in Section 5.4. Last, in Section 5.5, we give a procedure for describing the algebraic boundary of the completable region, as well as provide explicit semialgebraic descriptions for certain models.

Acknowledgements

The authors would like to thank Kaie Kubjas for her mentorship and advice throughout the research process. We are also grateful for Serkan Hoşten, Kaie, and Bernd Sturmfels for organizing the Varieties from Statistics Apprenticeship Week at the Institute for Mathematical and Statistical Innovation (IMSI). Furthermore we would like to thank all the other participants of the Algebraic Statistics and Our Changing World program for their feedback on early drafts of the paper. Part of this research was performed while the authors were visiting the Institute for Mathematical and Statistical Innovation (IMSI), which is supported by the National Science Foundation (Grant No. DMS-1929348). This research was partially supported by NSF Grant No. DMS-1855726. Cecilie Olesen Recke was supported by Novo Nordisk Foundation Grant NNF20OC0062897.

5.2 Toric Varieties

We discuss the necessary background on toric varieties for the completion problem. Given a vector of non-negative integers $v = (v_1, \dots, v_k) \in \mathbb{Z}^k_{\geq 0}$ and variables $\theta = (\theta_1, \dots, \theta_k)$, a *monomial* is an expression of the form $\theta^v = \theta_1^{v_1} \cdots \theta_k^{v_k}$. The vector v is the exponent vector of the monomial θ^v.

Let $A \in \mathbb{Z}^{k \times n}_{\geq 0}$ be an integer matrix with columns $a_1, \dots, a_n \in \mathbb{Z}^k_{\geq 0}$. We make the assumption that the *column sums of A are equal* to a positive integer $N > 0$. This is the case for many meaningful statistical models in applications, such as discrete graphical models, hierarchical models, and staged-tree models [Hoşten and Sullivant, 2002; Hoşten and Sullivant, 2007; Görgen et al., 2022]. We define a map $\phi^A : \mathbb{C}^k \to \mathbb{C}^n$ by

$$\phi^A(\theta) = (\theta^{a_1}, \dots, \theta^{a_n}).$$

The coordinate functions of ϕ^A are monomials, and the exponent vector of each monomial is a column of A. Thus, the map ϕ^A is homogeneous in that $\phi^A(\lambda p) = \lambda^N \phi^A(p)$ and its image is a cone.

The *toric variety* X_A associated to the integer matrix $A \in \mathbb{Z}_{\geq 0}^{k\times n}$ is the Zariski closure of the image $X_A = \overline{\operatorname{im} \phi^A}$. The toric variety X_A is an irreducible variety of dimension $\dim X_A = \operatorname{rank} A$. We remark that in the more general language of Cox et al. [2011], the set $\phi^A((\mathbb{C}^\times)^k)$ is a dense torus which acts on the toric variety X_A by coordinate-wise multiplication. Precisely, if $p \in X_A$ lies in the toric variety and $\theta \in (\mathbb{C}^\times)^k$, then the coordinate-wise product $p\phi^A(\theta) \in X_A$ also lies in the toric variety. This is the origin of the term "toric variety"—the variety X_A contains a dense open set which is isomorphic to an algebraic torus and whose action on itself extends to the variety X_A.

The *toric ideal* I_A is the defining ideal of a toric variety X_A,

$$I_A = I(X_A) = \{f \in \mathbb{C}[x_1, \ldots, x_n] : f(x) = 0 \text{ for all } x \in X_A\}.$$

The toric ideal I_A is a prime ideal generated by pure binomials and generators for this ideal may be computed explicitly from the matrix A, as described in [Sullivant, 2018, Proposition 6.2.4] restated here for convenience.

Proposition 5.2.1. *If $A \in \mathbb{Z}_{\geq 0}^{k\times n}$ is an integer matrix, then*

$$I_A = \langle p^u - p^v \mid u, v \in \mathbb{N}^n \text{ and } Au = Av \rangle.$$

Further, since the column sums of A are equal, the ideal I_A is a homogeneous ideal.

There are software such as `4ti2` [4ti2 team] and `Macaulay2` [Grayson and Stillman] that contain methods used to effectively compute Gröbner bases for toric ideals. This allows one to make several computations with toric ideals, such as determining containment of points in a toric variety and computing elimination ideals.

We write $\mathbb{R}_{\geq 0}^n$ and $\mathbb{R}_{>0}^n$ for the set of points in $\mathbb{R}^n$ with non-negative coordinates and positive coordinates respectively. The *probability simplex*

$$\Delta_{n-1} = \{(p_1, \ldots, p_n) \in \mathbb{R}_{\geq 0}^n : \sum_{i=1}^n p_i = 1\}$$

is the set of probability distributions in $\mathbb{R}^n$. Our statistical models of interest are intersections of toric varieties with the probability simplex.

Definition 5.2.2. *The log-linear model defined by an integer matrix $A \in \mathbb{Z}_{\geq 0}^{k\times n}$ is the set*

$$\mathcal{M}_A = X_A \cap \Delta_{n-1}.$$

The set $\mathcal{M}_A^{>0}$ is the set of points in the log-linear model $\mathcal{M}_A$ with non-zero coordinates.

Our definition of log-linear model differs from that found in [Sullivant, 2018, Chapter 6]. Indeed, we allow for probability distributions having zero coordinates, lying in the boundary of the probability simplex. The name "log-linear" originates as for those points $p \in \mathcal{M}_A^{>0}$ with non-zero coordinates, the coordinate-wise logarithm $\log(p)$ lies in the linear space $\operatorname{im} A^T$. Many familiar discrete probability models are in fact log-linear models, such as the independence model, undirected graphical models, and hierarchical models. We describe the topology of the statistical model $\mathcal{M}_A$ via the real structure of the toric variety X_A.

Given a matrix $A \in \mathbb{Z}_{\geq 0}^{k \times n}$, the *non-negative toric variety* $X_A^{\geq 0} = X_A \cap \mathbb{R}_{\geq 0}^n$ is the set of points in X_A with non-negative real coordinates. Similarly, we write $X_A^{>0} = X_A \cap \mathbb{R}_{>0}^n$ for the set of points in X_A with positive real coordinates. If $A \in \mathbb{Z}_{\geq 0}^{k \times n}$ is such that $(1, \ldots, 1) \in \operatorname{im} A^T$, then the log-linear model $\mathcal{M}_A$ may be considered the projectivization of the non-negative toric variety $X_A^{\geq 0}$. Indeed, since I_A is homogeneous, $X_A^{\geq 0}$ is a cone and we may scale each nonzero point so that the sum of the coordinates is equal to one. The algebraic moment map provides a way to understand the topology of the projectivization of the non-negative toric variety and hence, of our log-linear model $\mathcal{M}_A$.

The algebraic moment map is defined on the projectivization of a non-negative toric variety in Fulton [1993]; Sottile [2003]. With the identification of the projectivization of the non-negative toric variety with the log-linear model $\mathcal{M}_A$, the algebraic moment map is defined as follows.

Definition 5.2.3. *The algebraic moment map $\mu_A : \mathcal{M}_A \to \mathbb{R}^n$ is defined by $\mu_A(p) = Ap$.*

For $p \in \mathcal{M}_A$, the image $\mu_A(p)$ is a convex combination of the columns of A. Hence, if P_A denotes the polytope which is the convex hull of the columns of A, then the image of the algebraic moment map is contained in P_A. In fact, the algebraic moment map is a homeomorphism of $\mathcal{M}_A$ with P_A as seen in [Sottile, 2003, Theorem 8.5]. Further, this map restricts to a homeomorphism of $\mathcal{M}_A^{>0}$ with the interior of the polytope $\operatorname{int}(P_A)$ as described in [Fulton, 1993, Chapter 4.2].

Theorem 5.2.4 (Fulton [1993]; Sottile [2003])**.** *If $A \in \mathbb{Z}_{\geq 0}^{k \times n}$ is such that $(1, \ldots, 1) \in \operatorname{im} A^T$, then the algebraic moment map $\mu_A : \mathcal{M}_A \to P_A$ is a homeomorphism. Further, the restriction $\mu_A : \mathcal{M}_A^{>0} \to \operatorname{int}(P_A)$ is a homeomorphism.*

Thus, the topology of the log-linear model $\mathcal{M}_A$ is equivalent to that of the polytope P_A, and similarly for the topology of $\mathcal{M}_A^{>0}$ and $\operatorname{int}(P_A)$. Hence, both $\mathcal{M}_A$ and $\mathcal{M}_A^{>0}$ are contractible spaces and in particular, they are connected. Further, the set $\mathcal{M}_A^{>0}$, consisting of points of the log-linear model with non-zero coordinates, is the interior of the log-linear model $\mathcal{M}_A$ and is dense in $\mathcal{M}_A$. We will make use of this connection in Section 4.

5.3 Completion to the Toric Variety

We first consider the completion problem to a toric variety. Fix an integer matrix $A \in \mathbb{Z}_{\geq 0}^{k \times n}$ and the corresponding toric variety X_A. Write $[n] = \{1, \ldots, n\}$ for the set

indexing the coordinates of $\mathbb{C}^n$. A subset $E \subseteq [n]$ determines a coordinate projection $\pi_E : \mathbb{C}^n \to \mathbb{C}^E$ to those coordinates indexed by E. We say a *partial observation* is a point $p_E \in \mathbb{C}^E$, and a *completion* of a partial observation p_E to the toric variety X_A is a point $p \in X_A$ that projects to p_E, $\pi_E(p) = p_E$.

Our goal in this section is to determine the completable region $\pi_E(X_A)$, of partial observations which can be completed to the toric variety X_A. We accomplish this by first describing the image of the monomial map ϕ^A and determining when a partial observation p_E is completable to the image $\operatorname{im}\phi^A$.

5.3.1 The Image of the Monomial Map

For a point $p = (p_1, \ldots, p_n) \in \mathbb{C}^n$, the *support* of p is the set

$$\operatorname{supp}(p) = \{i \in [n] : p_i \neq 0\},$$

consisting of the indices of non-zero coordinates of p. The following definition, taken from Geiger et al. [2006], provides a necessary condition for a point to lie in the image $\operatorname{im}\phi^A$.

Definition 5.3.1 (Geiger et al. [2006])**.** *Let $A \in \mathbb{Z}_{\geq 0}^{k\times n}$ be a matrix with column vectors $a_1, \ldots, a_n \in \mathbb{Z}^k$. A point $p \in \mathbb{C}^n$ is A-feasible if for every $j \in [n] \setminus \operatorname{supp}(p)$, the support $\operatorname{supp}(a_j)$ is not contained in the union $\bigcup_{l\in\operatorname{supp}(p)} \operatorname{supp}(a_l)$.*

We note that A-feasibility is equivalent to the notion of zero-consistency in [Kahle et al., 2017, Definition 2.1]. For points lying in the toric variety $p \in X_A$, A-feasibility is both necessary and sufficient for p to lie in the image $\operatorname{im}\phi^A$. This was proved in [Geiger et al., 2006, Theorem 3.1] for the non-negative toric variety, and we extend their result for the complex toric variety.

Proposition 5.3.2. *If $A \in \mathbb{Z}_{\geq 0}^{k\times n}$ is an integer matrix, then the image of the map ϕ^A is given by the set*

$$\operatorname{im}\phi^A = \{p \in X_A : p \text{ is } A\text{-feasible}\}.$$

Proof. Write $a_1, \ldots, a_n$ for the columns of A and a_{ij} for the (i, j)-entry of A. If $p \in \operatorname{im}\phi^A$, then $p \in \overline{\operatorname{im}\phi^A} = X_A$. Similarly, if $p = \phi^A(\theta)$ then for $j \in [n] \setminus \operatorname{supp}(p)$, there is an $i \in [k]$ such that $a_{ij} > 0$ and $\theta_i = 0$ so that $i \in \operatorname{supp}(a_j)$. However, $i \notin \operatorname{supp}(a_l)$ for each $l \in \operatorname{supp}(p)$ since otherwise $p_l = 0$. Therefore, p is A-feasible. Thus, the inclusion $\operatorname{im}\phi^A \subseteq \{p \in X_A : p \text{ is } A\text{-feasible}\}$ holds.

For the reverse inclusion, let $p \in X_A$ be A-feasible. Without loss of generality we may assume that p only has non-zero entries. Indeed, since p is A-feasible, we may restrict ϕ^A to the coordinate subspace of $\mathbb{C}^k$ such that $\theta_i = 0$ for $i \in [n] \setminus \bigcup_{l\in\operatorname{supp}(p)} \operatorname{supp}(a_l)$. We compute a preimage of this restriction by discarding all zero entries of p, and reinserting them in the appropriate places afterwards.

Let $p \in X_A$ have non-zero coordinates. From Proposition 5.2.1, for any $u \in \mathbb{Z}^n$ such that $Au = 0$, we must have that $p^u = 1$. By taking the logarithm and writing

$\log(p)$ for the coordinate-wise logarithm of p, one finds that $u^T \log(p) = 0$. Thus, $\log(p)$ annihilates the kernel of A, or equivalently, lies in the image of A^T. Writing $\log(p) = A^T v$ for some vector $v \in \mathbb{R}^k$ and applying coordinate-wise exponentiation, one finds that $p = \phi^A(e^v) \in \operatorname{im} \phi^A$. □

If A_E denotes the submatrix of A whose columns are indexed by E, then there is an equality $\phi^{A_E} = \pi_E \circ \phi^A$. Thus, there is an equality of the images $\pi_E(\operatorname{im} \phi^A) = \operatorname{im} \phi^{A_E}$, and the defining ideal of the image $\pi_E(\operatorname{im} \phi^A)$ is the toric ideal given by

$$I(\pi_E(\operatorname{im} \phi^A)) = I(\operatorname{im} \phi^{A_E}) = I(X_{A_E}) = I_{A_E}.$$

This ideal may be computed via Proposition 5.2.1, or as an elimination ideal of I_A. Indeed, I_{A_E} is obtained from I_A by eliminating the variables x_i for $i \in [n] \setminus E$. We obtain the following corollary of Proposition 5.3.2 describing when a partial observation is completable to the image $\operatorname{im} \phi^A$.

Corollary 5.3.3. *Let $p_E \in \mathbb{C}^E$ be a partial observation.*

1. *p_E is completable to a point of the image $\operatorname{im} \phi^A$ if and only if $p_E \in X_{A_E}$ and p_E is A_E-feasible.*
2. *If $p_E \in \mathbb{R}^E_{\geq 0}$ has non-negative real coordinates, then p_E is completable to $\operatorname{im} \phi^A$ if and only if there is a completion $p \in \operatorname{im} \phi^A$ with non-negative real coordinates.*

Proof. The first portion follows from the equality $\operatorname{im} \phi^{A_E} = \pi_E(\operatorname{im} \phi^A)$. A partial observation $p_E \in \mathbb{C}^E$ is completable to the image $\operatorname{im} \phi^A$ if and only if it lies in the image $\operatorname{im} \phi^{A_E}$. By Proposition 5.3.2, this is exactly when $p_E \in X_{A_E}$ and p_E is A_E-feasible.

For the second portion, if $p_E \in \mathbb{R}^E_{\geq 0}$ is a partial observation which has non-negative real coordinates and is completable to a point $p \in \operatorname{im} \phi^A$, then the point $|p| \in \operatorname{im} \phi^A$ obtained by taking the coordinate-wise absolute value of p is a completion of p_E with non-negative real coordinates. □

5.3.2 Completing to the Toric Variety

We now determine the set of partial observations which are completable to the toric variety X_A. We utilize the polyhedral structure of the polytope P_A, which is the convex hull of the columns of the matrix A. We note that as the column sums of A are equal to $N > 0$, the polytope P_A is contained in the hyperplane determined by the equation $\sum_{i=1}^n p_i = N$ and has dimension $\operatorname{rank} A - 1$.

Definition 5.3.4. *A facial set of A is a subset $F \subseteq [n]$ such that there is a vector $v \in \mathbb{R}^k$ satisfying $v^T a_i = 0$ for $i \in F$ and $v^T a_i > 0$ for $i \in [n] \setminus F$. The vector v is a inner normal vector for the face F.*

Additionally, a *facet* is a proper face which is not contained in any strictly larger proper face. Geometrically, a facial set indexes the columns of A that lie in a face of the polytope P_A. Further, an inner normal vector v of a facial set F is an inner normal

vector of a face of P_A. More on the relationship between polytopes and their faces can be found in Ewald [1996].

The following result from [Geiger et al., 2006, Lemma A.2] classifies the support of a point in X_A via the facial sets of A.

Lemma 5.3.5 (Geiger et al. [2006])**.** *Let $A \in \mathbb{Z}_{\geq 0}^{k\times n}$ be a matrix whose columns are the vectors $a_1, \ldots, a_n \in \mathbb{R}^k$. If $p \in X_A$, then* $\operatorname{supp}(p)$ *is a facial set of A.*

If $F \subseteq [n]$ is a facial set of $A \in \mathbb{Z}_{\geq 0}^{k\times n}$, we define the *characteristic vector* χ_F by $(\chi_F)_i = 1$ if $i \in F$ and $(\chi_F)_i = 0$ if $i \in [n] \setminus F$. By [Geiger et al., 2006, A.2], $\chi_F \in X_A$. Thus, for every facial set F of A, there is a point of X_A with support F.

Lemma 5.3.5 and Corollary 5.3.3 allow us to determine when a partial observation $p_E \in X_{A_E}$ is completable to the toric variety X_A. As the subset $E \subseteq [n]$ indexes the coordinates of $\mathbb{C}^E$, we may write $\operatorname{supp}(\pi_E(p)) = \operatorname{supp}(p) \cap E$ for a point $p \in \mathbb{C}^n$.

Theorem 5.3.6. *Let $p_E \in \mathbb{C}^E$ be a partial observation.*

1. *$p_E \in \mathbb{C}^E$ is completable to a point $p \in X_A$ if and only if $p_E \in X_{A_E}$ and there is a facial set F of A such that* $\operatorname{supp}(p_E) = F \cap E$.
2. *If $p_E \in \mathbb{R}_{\geq 0}^E$ has non-negative real coordinates, then p_E is completable to $X_A^{\geq 0}$ if and only if p_E is completable to X_A.*

Proof. If $p_E = \pi_E(p) \in \pi_E(X_A)$ for some $p \in X_A$, then by Lemma 5.3.5 $\operatorname{supp}(p)$ is a facial set of A and $\operatorname{supp}(p_E) = \operatorname{supp}(p) \cap E$.

Conversely, if $p_E \in X_{A_E}$ and F is a facial set of A such that $\operatorname{supp}(p_E) = E \cap F$, then p_E is $A_{E\cap F}$-feasible since p_E has non-zero coordinates $(p_E)_i$ for every $i \in E \cap F = \operatorname{supp}(p_E)$. By Corollary 5.3.3, this implies there exists $\theta \in \mathbb{C}^k$ such that the i-th coordinates $(\phi^A(\theta))_i = (p_E)_i$ are equal for every $i \in E \cap F$. Consider the coordinate-wise product $\phi^A(\theta)\chi_F \in X_A$. If $i \in E \cap F$, then the i-th coordinates $(\phi^A(\theta)\chi_F)_i = (p_E)_i$ are equal. Similarly, if $i \in E \setminus F$, then $(\phi^A(\theta)\chi_F)_i = (p_E)_i = 0$. Thus, $\phi^A(\theta)\chi_F \in X_A$ is a completion of p_E to X_A.

For the second portion, if $p_E \in \mathbb{R}_{\geq 0}^E$ is a partial observation with completion $p \in X_A$ to X_A, then the coordinate-wise absolute value $|p| \in X_A^{\geq 0}$ is a completion to $X_A^{\geq 0}$. □

Given a partial observation $p_E \in \mathbb{C}^E$, Lemma 5.3.5 and Theorem 5.3.6 allow us to quickly determine properties of completions to X_A by studying the facial sets of A. If F_1 and F_2 are facial sets, then their intersection $F_1 \cap F_2$ is a facial set. In particular, for any set $E \subseteq [n]$, there is a minimal facial set F containing E, which is the intersection of all facial sets containing E.

Corollary 5.3.7. *Let $A \in \mathbb{Z}_{\geq 0}^{k\times n}$ be an integer matrix, $E \subseteq [n]$ be a subset of the coordinates, $p_E \in \mathbb{C}^E$ be a partial observation, and $p \in X_A$ be a completion of p_E. If $F \subseteq [n]$ is the smallest facial set containing* $\operatorname{supp}(p_E)$*, then $F \subseteq \operatorname{supp}(p)$.*

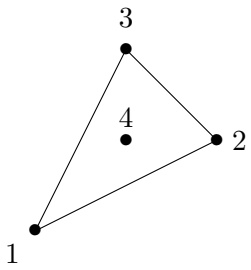


Figure 5.3.2: An example polytope P_A. The point corresponding to the 4-th column of A lies in the relative interior of the polytope.

Example 5.3.8. Consider an integer matrix $A \in \mathbb{Z}_{>0}^{k\times 4}$ such that the polytope P_A is as in Figure 5.3.2. If $E = \{4\}$ then the smallest face containing E is the whole polytope P_A. Thus, the smallest facial set containing E is $F = \{1, 2, 3, 4\}$. By Corollary 5.3.7, for any non-zero partial observation $p_E \in \mathbb{C}^E$, any completion to X_A must have all non-zero coordinates. If $p_E = 0$, then any completion has support given by one of the proper facial sets, $\emptyset$, $\{1, 2\}$, $\{1, 3\}$, or $\{2, 3\}$.

5.4 Completion to the Log-Linear Model

Let $E \subseteq [n]$ be a subset indexing some of the coordinates of $\mathbb{R}^n$ and $\pi_E : \mathbb{R}^n \to \mathbb{R}^E$ be the corresponding coordinate projection. Given an integer matrix $A \in \mathbb{Z}_{\geq 0}^{k\times n}$, we consider the completions of a partial observation $p_E \in \mathbb{R}^E$ to the log-linear model $\mathcal{M}_A$. Our analysis of the problem relies on our ability to understand completions to the non-negative toric variety $X_A^{\geq 0}$ as described in the previous section.

We provide a description of the interior and the boundary of the *completable region* $\pi_E(\mathcal{M}_A)$, which consists of partial observations which can be completed to a point in $\mathcal{M}_A$: see Theorem 5.4.11. This is accomplished in Section 5.4.1 by analyzing the singular locus of the projection $\pi_E : \mathcal{M}_A^{>0} \to \mathbb{R}^E$, which is the locus of points where the differential $(d\pi_E)_p$ drops rank. For a point $p \in \mathcal{M}_A^{>0}$ that does not lie in the singular locus, the image $\pi_E(p)$ lies in the interior of the completable region $\pi_E(\mathcal{M}_A)$. Thus, the problem of determining the interior and the boundary of the completable region rests on understanding the image of the singular locus and the image of the boundary of the log-linear model $\partial\mathcal{M}_A = \mathcal{M}_A \setminus \mathcal{M}_A^{>0}$.

In addition, we enumerate the completions for a partial observation with non-zero coordinates lying in the completable region. It will be shown that when there are finitely many completions, a partial observation can have either one or two completions to the log-linear model $\mathcal{M}_A$ depending on the subset $E \subseteq [n]$ chosen and whether the partial observation lies in the interior or the boundary of the completable region—see Theorem 5.4.13 and Theorem 5.4.16. Our results determine the number of completions of a partial observation to $\mathcal{M}_A$, but produce no general algorithm for computing these completions.

We work under the mild assumption that our subset $E \subseteq [n]$ satisfies $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1 = \dim \mathcal{M}_A$. The assumption that $|E| = \operatorname{rank} A_E \leq \dim \mathcal{M}_A$ guarantees that

the image $\pi_E : \mathcal{M}_A \to \mathbb{R}^E$ is full-dimensional—see Corollary 5.4.7. Further, we assume that $|E| = \dim \mathcal{M}_A$ so that one expects only finitely many completions to the log-linear model $\mathcal{M}_A$ for a general partial observation. If $|E| < \dim \mathcal{M}_A$, then any partial observation lying in the completable region has infinitely many completions to $\mathcal{M}_A$ and we leave it as an open problem to describe the variety of completions in this case.

5.4.1 The Singular Locus of a Coordinate Projection

We note that for a map of varieties $\pi : X \to Y$, there is a maximal rank of the differential $d\pi_p : T_pX \to T_pY$ for $p \in X$. Further, there is a Zariski open set (dense, open, path-connected set whose complement is a subvariety) $U \subseteq X$ for which this maximal rank is attained. The complement of this open set is the subvariety of X where the rank of the differential drops.

Definition 5.4.1. *The singular locus of a map of varieties $\pi : X \to Y$ is the subvariety X consisting of points $p \in X$ such that the differential $d\pi_p : T_pX \to T_pY$ has rank less than the maximal rank.*

We begin by computing the tangent space at a point of $\mathcal{M}_A^{>0}$ considered as an open subset of a variety. Let $A \in \mathbb{Z}_{\geq 0}^{k \times n}$ have entries a_{ij} for $1 \leq i \leq k$ and $1 \leq j \leq n$. By differentiating the monomial map $\phi^A : \mathbb{R}_{>0}^k \to \mathbb{R}_{>0}^n$ at $\theta \in \mathbb{R}_{>0}^k$, we obtain the map $d\phi_\theta^A : \mathbb{R}^k \to \mathbb{R}^n$ on tangent spaces defined by

$$(d\phi_\theta^A)_{ij} = a_{ji} \frac{1}{\theta_j} \phi_i^A(\theta).$$

By writing $p = \phi^A(\theta)$, the image of the differential may be written as

$$\operatorname{im} d\phi_\theta^A = \{(p_1 v_1, \ldots, p_n v_n) \in \mathbb{R}^n : v \in \operatorname{im} A^T\}.$$

Since this image has dimension $\operatorname{rank} A = \dim X_A^{>0}$, it follows that the image $\operatorname{im} d\phi_\theta^A$ coincides with the tangent space of the image $X_A^{>0}$ at p. Thus, $X_A^{>0}$ is smooth at each point and the tangent space is given by

$$T_pX_A^{>0} = \operatorname{im} d\phi_\theta^A = \{(p_1 v_1, \ldots, p_n v_n) \in \mathbb{R}^n : v \in \operatorname{im} A^T\}.$$

Write $H \subseteq \mathbb{R}^n$ for the hyperplane defined by the equation $\sum_i x_i = 1$ so that $T_pH = \{x \in \mathbb{R}^n : \sum_i x_i = 0\}$. Since $(1, \ldots, 1) \in \operatorname{im} A^T$, it follows that $(p_1, \ldots, p_n) \in T_pX_A^{>0}$ so that $T_pX_A^{>0} + T_pH = \mathbb{R}^n$—that is, $X_A^{>0}$ intersects H transversally at each point. Thus, we may regard $\mathcal{M}_A^{>0} = X_A^{>0} \cap H$ as a smooth manifold of dimension $\dim \mathcal{M}_A^{>0} = \dim X_A^{>0} - 1 = \operatorname{rank} A - 1$. Further, we write its tangent space as

$$T_p\mathcal{M}_A^{>0} = \{(p_1 v_1, \ldots, p_n v_n) \in \mathbb{R}^n : v \in \operatorname{im} A^T,\ v^T p = 0\}.$$

We now fix a subset $E \subseteq [n]$ such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$ and consider the corresponding coordinate projection $\pi_E : \mathbb{R}^n \to \mathbb{R}^E$ and its restriction to $\mathcal{M}_A^{>0}$. We show that its differential $(d\pi_E|_{\mathcal{M}_A^{>0}})_p$ is an isomorphism for most points $p \in \mathcal{M}_A^{>0}$, and give an

explicit description of the set of points where this differential is not an isomorphism, or equivalently, of the singular locus of π_E. The following proposition will aid in identifying whether the differential $(d\pi_E|_{\mathcal{M}_A^{>0}})_p$ is an isomorphism. As a linear map, we identify the differential $(d\pi_E)_p$ with the map π_E itself.

Proposition 5.4.2. *Let $E \subseteq [n]$ is such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$. Then $\dim(\operatorname{im} A^T \cap \ker \pi_E) = 1$. That is, every vector in $\operatorname{im} A^T \cap \ker \pi_E$ is a scalar multiple of any non-zero vector $\nu \in \operatorname{im} A^T \cap \ker \pi_E$.*

Proof. The map $\pi_E : (\operatorname{im} A^T + \ker \pi_E) \to \operatorname{im} A_E^T$ gives an isomorphism of the space $(\operatorname{im} A^T + \ker \pi_E)/\ker \pi_E$ with $\operatorname{im} A_E^T$ so that

$$\begin{aligned}\dim(\operatorname{im} A^T + \ker \pi_E) &= \dim \operatorname{im} A_E^T + \dim \ker \pi_E \\ &= \operatorname{rank} A_E + n - |E| \\ &= n.\end{aligned}$$

Thus, we may compute $\dim(\operatorname{im} A^T \cap \ker \pi_E) = \dim \operatorname{im} A^T + \dim \ker \pi_E - n = 1$. □

Corollary 5.4.3. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$ and $\nu \in \operatorname{im} A^T \cap \ker \pi_E$ be a non-zero vector. For $p \in \mathcal{M}_A^{>0}$, the differential $(d\pi_E|_{\mathcal{M}_A^{>0}})_p : T_p\mathcal{M}_A^{>0} \to \mathbb{R}^E$ is an isomorphism if and only if $\nu^T p \neq 0$.*

Proof. Since $\dim T_p\mathcal{M}_A^{>0}$ and $|E|$ are equal to $\operatorname{rank} A - 1$, the differential is an isomorphism if and only if it is injective. We use the description of the tangent space $T_p\mathcal{M}_A^{>0}$ found above,

$$T_p\mathcal{M}_A^{>0} = \{(p_1v_1, \ldots, p_nv_n) \in \mathbb{R}^n : v \in \operatorname{im} A^T,\ v^T p = 0\}.$$

As $p \in \mathcal{M}_A^{>0}$, a non-zero tangent vector $(p_1v_1, \ldots, p_nv_n) \in T_p\mathcal{M}_A^{>0}$ lies in $\ker(d\pi_E|_{\mathcal{M}_A^{>0}})_p$ if and only if $v \in \operatorname{im} A^T \cap \ker \pi_E$ and $v^T p = 0$. Equivalently, the kernel $\ker(d\pi_E|_{\mathcal{M}_A^{>0}})_p$ is trivial if and only if $\nu^T p \neq 0$. □

Corollary 5.4.3 effectively describes the singular locus of π_E as the set of points $p \in \mathcal{M}_A^{>0}$ such that $\nu^T p = 0$ for a non-zero vector $\nu \in \operatorname{im} A^T \cap \ker \pi_E$. We demonstrate that this locus is a proper subset of $\mathcal{M}_A^{>0}$. Recall that the algebraic moment map $\mu_A : \mathcal{M}_A \to P_A$ is a homeomorphism as described in Theorem 5.2.4.

Proposition 5.4.4. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$ and let $\nu \in \operatorname{im} A^T \cap \ker \pi_E$ be a non-zero vector. A point $p \in \mathcal{M}_A$ satisfies $\nu^T p = 0$ if and only if $\mu_A(p) \in \operatorname{im} A_E \cap P_A$.*

Proof. Note that $q \in \operatorname{im} A_E$ if and only if $v^T q = 0$ for all $v \in \ker A_E^T$—that is, q lies in the span of the columns of A indexed by E exactly when every hyperplane equation which vanishes on the columns of A indexed by E also vanish on q. Thus, for $p \in \mathcal{M}_A$, one has $\mu_A(p) \in \operatorname{im} A_E \cap P_A$ if and only if $v^T\mu_A(p) = (A^T v)^T p = 0$ for all $v \in \ker A_E^T$. However, the equality $A^T(\ker A_E^T) = \operatorname{im} A^T \cap \ker \pi_E$ holds so that the result follows. □

The intersection $\operatorname{im} A_E \cap P_A$ is the set of points in P_A spanned by the columns of A indexed by E. Since $\operatorname{rank} A_E = \operatorname{rank} A - 1$, this is a proper subset of P_A and the locus of points $p \in \mathcal{M}_A$ such that $\nu^T p = 0$ is a proper subset of $\mathcal{M}_A$. Thus, the maximal rank of the differential $(d\pi_E|_{\mathcal{M}_A^{>0}})_p$ is $\dim \mathcal{M}_A^{>0} = |E| = \operatorname{rank} A - 1$ and this rank drops exactly on the locus of points in $\mathcal{M}_A^{>0}$ lying on the hyperplane defined by $\nu^T p = 0$. Combining these results, we've proved the following.

Corollary 5.4.5. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$ and $\nu \in \operatorname{im} A^T \cap \ker \pi_E$ any non-zero vector. The singular locus of the projection $\pi_E|_{\mathcal{M}_A}$ is the set of points*

$$B_{A,E} = \{p \in \mathcal{M}_A^{>0} : \nu^T p = 0\}.$$

We note that with this terminology, Proposition 5.4.4 states that the image of the algebraic moment map applied to the singular locus $B_{A,E}$ is given by $\mu_A(B_{A,E}) = \operatorname{im} A_E \cap P_A$. That is, the image $\mu_A(B_{A,E})$ is the points of P_A that are spanned by the columns of A indexed by E. In addition, the equality $\operatorname{im} A^T \cap \ker \pi_E = A^T(\ker A_E^T)$ gives an effective method of computing a vector ν. This is illustrated in Example 5.4.6.

Example 5.4.6. Let

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 0 & 4 & 0 & 1 & 2 \\ 0 & 0 & 4 & 1 & 1 \end{pmatrix} \in \mathbb{Z}_{\geq 0}^{3\times 5}$$

and $E = \{4, 5\}$. The polytope P_A and the image $\mu_A(B_{A,E})$ are illustrated in Figure 5.4.3. Since $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1 = 2$, the result of Proposition 5.4.4 applies. The kernel $\ker A_E^T$ is generated by the vector $\omega = (1, 1, -3)$, so we may let $\nu = \frac{1}{4} A^T \omega = (1, 1, -3, 0, 0) \in A^T(\ker A_E^T)$. Thus, the singular locus of the projection $\pi_E : \mathcal{M}_A^{>0} \to \mathbb{R}^E$ is given by the hyperplane section

$$B_{A,E} = \{p \in \mathcal{M}_A^{>0} : p_1 + p_2 - 3p_3 = 0\}.$$

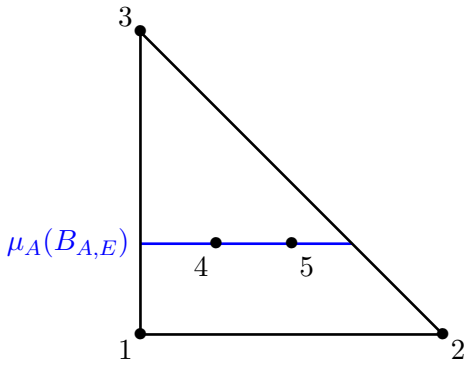


Figure 5.4.3: The polytope P_A and the image $\mu_A(B_{A,E})$ for Example 5.4.6.

We end by showing that a point $p \in \mathcal{M}_A^{>0} \setminus B_{A,E}$ not in the singular locus maps to the interior of the completable region.

Corollary 5.4.7. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$ and let $p \in \mathcal{M}_A^{>0} \setminus B_{A,E}$. There is an open subset $U \subseteq \mathbb{R}^E$ such that $\pi_E(p) \in U$ and $U \subseteq \pi_E(\mathcal{M}_A)$. In particular, the completable image is full-dimensional and $\pi_E(p) \in \operatorname{int}(\pi_E(\mathcal{M}_A))$ lies in the interior of the completable region.*

Proof. For any point $p \in \mathcal{M}_A^{>0} \setminus B_{A,E}$, the differential $(d\pi_E|_{\mathcal{M}_A^{>0}})_p : T_p\mathcal{M}_A^{>0} \to \mathbb{R}^E$ is an isomorphism. By the implicit function theorem, π_E restricts to a diffeomorphism of an open neighborhood of $p \in \mathcal{M}_A^{>0}$ to an open neighborhood of $\pi_E(p)$. □

5.4.2 The Interior and Boundary of the Completable Region

Corollary 5.4.7 shows that for $p \in \mathcal{M}_A^{>0} \setminus B_{A,E}$, the image $\pi_E(p)$ lies in the interior of the completable region $\pi_E(\mathcal{M}_A)$. We will show that the boundary of the completable region consists of the image of the boundary of the model $\pi_E(\partial\mathcal{M}_A)$ and the image of the singular locus $\pi_E(B_{A,E})$, which completely classifies the interior and the boundary of the completable region. We accomplish this by better understanding the completions of a partial observation to the non-negative toric variety $X_A^{\geq 0}$.

Proposition 5.4.8. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$ and let $\nu \in \operatorname{im} A^T \cap \ker \pi_E$ be a non-zero vector. If $p_E \in \mathbb{R}^E_{>0}$ is a partial observation with non-zero coordinates, then for any completions $p, q \in X_A^{\geq 0}$ of p_E, there exists $\alpha \in \mathbb{R}$ such that*

$$p_i = q_i e^{\alpha\nu_i} \quad \text{for all} \quad i \in \operatorname{supp}(p) \cap \operatorname{supp}(q).$$

Proof. We note that for any point $p \in X_A$, by projecting the coordinates indexed by $\operatorname{supp}(p)$, there is a completion to $\phi^A(\mathbb{R}^k_{>0})$ which agrees with p on its support. Thus, we may write $p = \chi_{\operatorname{supp}(p)}\phi^A(\theta)$ for some $\theta \in \mathbb{R}^k_{>0}$. Thus, we write $p = \chi_{F_1}\phi^A(\theta_1)$ and $q = \chi_{F_2}\phi^A(\theta_2)$ where $F_1 = \operatorname{supp}(p)$ and $F_2 = \operatorname{supp}(q)$ are facial sets that contain E, and $\theta_1, \theta_2 \in \mathbb{R}^k_{>0}$. Then by taking the coordinate-wise logarithm, we have

$$0 = \log(p_E) - \log(q_E) = A_E^T(\log(\theta_1) - \log(\theta_2)).$$

Therefore $\log(\theta_1) - \log(\theta_2) \in \ker A_E^T$ so that $\log(p) - \log(q) = \alpha\nu \in A^T(\ker A_E^T)$ for some $\alpha \in \mathbb{R}$. Exponentiating, we have that for $\phi^A(\theta_1) = \phi^A(\theta_2)e^{\alpha\nu}$. Then for $i \in F_1 \cap F_2$, we have that $p_i = q_i e^{\alpha\nu_i}$. □

We now leverage this proposition to show that the boundary of the completable region $\partial\pi_E(\mathcal{M}_A)$ is given by the image of the boundary of the model $\pi_E(\partial\mathcal{M}_A)$ and the image of the singular locus $\pi_E(B_{A,E})$. Equivalently, we show no point of the boundary $\partial\mathcal{M}_A$ or the singular locus $B_{A,E}$ map into the interior $\operatorname{int}(\pi_E(\mathcal{M}_A))$. We start by showing a point $p \in \partial\mathcal{M}_A$ is such that either $\pi_E(p)$ has some coordinate equal to zero or for any $\epsilon > 0$, $(1+\epsilon)\pi_E(p)$ is not completable the log-linear model $\mathcal{M}_A$.

Lemma 5.4.9. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$. The boundary of the model $\partial\mathcal{M}_A$ maps by π_E into the boundary of the completable region $\partial\pi_E(\mathcal{M}_A)$. That is, there is an inclusion $\pi_E(\partial\mathcal{M}_A) \subseteq \partial\pi_E(\mathcal{M}_A)$.*

Proof. Let $p \in \mathcal{M}_A$ and $p_E = \pi_E(p)$. If $p \in \partial\mathcal{M}_A$, then some coordinate of p is equal to zero and $\operatorname{supp}(p) \subseteq [n]$ is a proper facial set. There are two cases to consider. If E is not contained in $\operatorname{supp}(p)$, then $p_E \in \mathbb{R}^E$ contains some zero coordinate and any open ball around p_E contains points with negative coordinates. Those points with negative coordinates have no completion to $\mathcal{M}_A$ so that $p_E \in \partial\pi_E(\mathcal{M}_A)$.

Assume that $E \subseteq \operatorname{supp}(p)$. Since $\operatorname{rank} A_E = \operatorname{rank} A - 1$ and $\operatorname{supp}(p)$ is a proper facial set, it follows that $\operatorname{supp}(p)$ is a facet and $\operatorname{supp}(p)$ is the smallest facial set containing E. Without loss of generality, we may let $\omega \in \ker A_E^T$ be an inner normal vector of the facet $\operatorname{supp}(p)$ so that $\nu = A^T\omega$ satisfies $\nu_i = 0$ for all $i \in \operatorname{supp}(p)$ and $\nu_i > 0$ for $i \in [n] \setminus \operatorname{supp}(p)$.

Let $\epsilon > 0$ and $q_E = (1+\epsilon)p_E \in \mathbb{R}^E$. Then $(1+\epsilon)p \in X_A^{\geq 0}$ is a completion of q_E and by Proposition 5.4.8, any other completion $q \in X_A^{\geq 0}$ of q_E satisfies

$$q_i = (1+\epsilon)p_i e^{\alpha\nu_i}$$

for all $i \in \operatorname{supp}(p) \cap \operatorname{supp}(q)$. Since $E \subseteq \operatorname{supp}(q)$ and $\operatorname{supp}(p)$ is the smallest facial set containing E, we have that $\operatorname{supp}(p) \subseteq \operatorname{supp}(q)$. Further, since $\nu_i = 0$ for all $i \in \operatorname{supp}(p)$ and $q_i = (1+\epsilon)p_i$ for $i \in \operatorname{supp}(p)$. We compute

$$\sum_i q_i \geq \sum_{i \in \operatorname{supp}(p)} q_i = (1+\epsilon) \sum_{i \in \operatorname{supp}(p)} p_i = 1 + \epsilon > 1.$$

Any open set around p_E then contains points which are not completable to $\mathcal{M}_A$ so that $p_E \in \partial\pi_E(\mathcal{M}_A)$. □

Lemma 5.4.9 extends to the singular locus $B_{A,E}$ as well. We show that if $p \in B_{A,E}$, then the coordinates of $\pi_E(p)$ cannot be increased while remaining completable to the log-linear model $\mathcal{M}_A$. Thus, the image $\pi_E(B_{A,E})$ is contained in the boundary of the completable region as well.

Lemma 5.4.10. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$. The singular locus $B_{A,E}$ maps by π_E into the boundary of the completable region $\partial\pi_E(\mathcal{M}_A)$. That is, there is an inclusion $\pi_E(B_{A,E}) \subseteq \partial\pi_E(\mathcal{M}_A)$.*

Proof. We may assume E is not contained in a proper facial set, since otherwise $B_{A,E}$ is empty as can be seen by Proposition 5.4.4. Thus, by Corollary 5.3.7, if $p_E \in \mathbb{R}^E$ has non-zero coordinates, any completion $p \in X_A^{\geq 0}$ has non-zero coordinates. Let $p \in B_{A,E}$ and $p_E = \pi_E(p)$. If $\epsilon > 0$ and $q_E = (1+\epsilon)p_E$ as above, then $(1+\epsilon)p \in X_A^{\geq 0}$ is a completion. For any other completion $q \in X_A^{\geq 0}$, there exists $\alpha \in \mathbb{R}$ such that $q_i = (1+\epsilon)p_i e^{\alpha\nu_i}$ for all $i \in [n]$ since p and q have all non-zero coordinates. Since $\nu^T p = 0$, one computes

$$\sum_i q_i = (1+\epsilon)\sum_i p_i e^{\alpha\nu_i} \geq (1+\epsilon)\sum_i p_i(1+\alpha\nu_i) = 1 + \epsilon > 1.$$

Again one finds that any open set around p_E contains points which are not completable to $\mathcal{M}_A$ so that $p_E \in \partial\pi_E(\mathcal{M}_A)$. □

We combine Lemma 5.4.9 and Lemma 5.4.10 to describe the boundary and the interior of the completable region.

Theorem 5.4.11. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$. The boundary of the completable region is equal to the image by π_E of the union of the boundary of the model $\partial\mathcal{M}_A$ and the singular locus $B_{A,E}$. Precisely,*

$$\partial\pi_E(\mathcal{M}_A) = \pi_E(\partial\mathcal{M}_A) \cup \pi_E(B_{A,E}).$$

Equivalently, the interior of the completable region is equal to the image

$$\operatorname{int}(\pi_E(\mathcal{M}_A)) = \pi_E(\mathcal{M}_A^{>0} \setminus B_{A,E}).$$

Proof. For $p \in \mathcal{M}_A^{>0} \setminus B_{A,E}$, the differential $(d\pi_E)_p$ is an isomorphism so that $\pi(p) \in \operatorname{int}(\pi(\mathcal{M}_A))$. By taking the relative complement with $\pi_E(\mathcal{M}_A)$, the inclusion $\partial\pi_E(\mathcal{M}_A) \subseteq \pi_E(\partial\mathcal{M}_A) \cup \pi_E(B_{A,E})$ follows. The reverse inclusion is an immediate consequence of Lemma 5.4.9 and Lemma 5.4.10. □

We illustrate Theorem 5.4.11 in Example 5.4.12.

Example 5.4.12. Let

$$A = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \in \mathbb{Z}_{\geq 0}^{3\times 4}$$

and $E = \{1, 4\}$. The polytope P_A, the image $\mu_A(B_{A,E})$, and the completable region $\pi_E(\mathcal{M}_A)$ are illustrated in Figure 5.4.4. Using coordinates x, y, z, w for $\mathbb{R}^4$, the log-linear model is given by

$$\mathcal{M}_A = \{(x, y, z, w) \in \mathbb{R}^4_{\geq 0} : xw - yz = x + y + z + w - 1 = 0\}.$$

As $(0, 1, -1) \in \ker A_E^T$, we may take $\nu = (0, 1, -1, 0)$ so that

$$B_{A,E} = \{(x, y, z, w) \in \mathcal{M}_A^{>0} : y - z = 0\}.$$

As the subset E is not contained in a proper facial set, the boundary $\partial\mathcal{M}_A$, of points where some coordinate is equal to zero, maps to the coordinate axes of $\mathbb{R}^E_{\geq 0}$. Contrary to this, the singular locus $B_{A,E}$ maps to the curved boundary of Figure 5.4.4. The defining equation for this curve can be obtained via elimination. Indeed, the defining ideal of the singular locus is the prime ideal

$$I(B_{A,E}) = \langle xw - yz, x + y + z + w - 1, y - z \rangle.$$

Using `Macaulay2` [Grayson and Stillman] to eliminate the variables y and z, we find that the eliminant of $I(B_{A,E})$ is generated by the single polynomial

$$f(x,w) = x^2 - 2xw - w^2 - 2x - 2w + 1.$$

Thus, $f(x,w) = 0$ is the defining equation of the Zariski closure of the projection $\overline{\pi_E(B_{A,E})}$.

We note that the image $\pi_E(B_{A,E})$ is a semialgebraic set and the zero set $\{(x,w) \in \mathbb{R}^2 : f(x,w) = 0\}$ is its Zariski closure. In general, it is difficult to obtain a semialgebraic description of the image $\pi_E(B_{A,E})$ or of the completable region $\pi_E(\mathcal{M}_A)$.

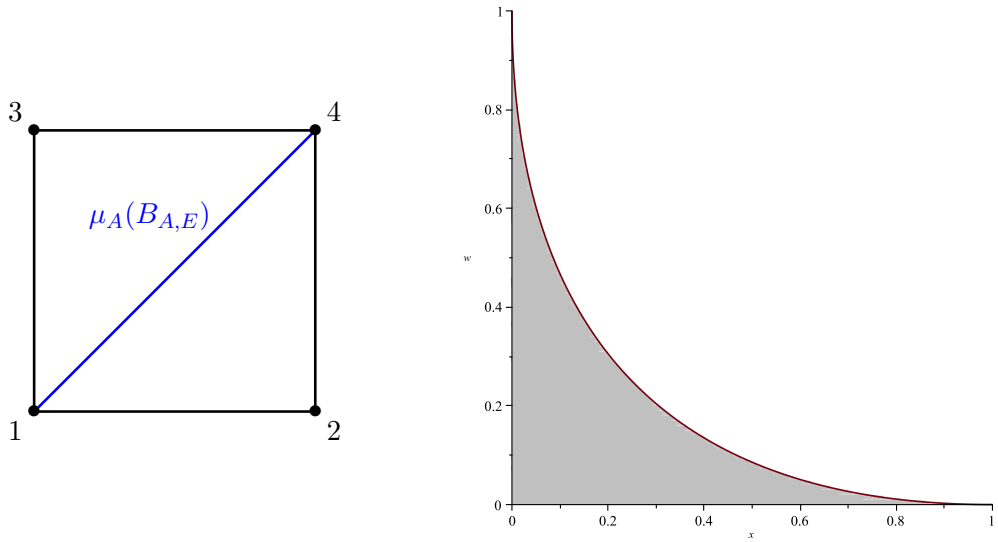


Figure 5.4.4: (Left) The polytope P_A and image $\mu_A(B_{A,E})$ for Example 5.4.12. (Right) The completable region $\pi_E(\mathcal{M}_A)$ for Example 5.4.12.

5.4.3 Enumerating Completions to the Log-Linear Model

Given a matrix $A \in \mathbb{Z}_{\geq 0}^{k \times n}$ and a subset $E \subseteq [n]$ such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$, we discuss the possible number of completions of a partial observation with non-zero coordinates in the completable region $p_E \in \pi_E(\mathcal{M}_A)$ to the log-linear model $\mathcal{M}_A$. By elementary methods, we conclude that such a partial observation has either 0, 1, or 2 completions to $\mathcal{M}_A$ depending on whether the subset $E \subseteq [n]$ lies in a proper facial set and whether p_E lies in the boundary or the interior of the completable region.

We first consider the case that a partial observation with non-zero coordinates lies in the boundary of the completable region $p_E \in \partial\pi_E(\mathcal{M}_A)$.

Theorem 5.4.13. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$. If $p_E \in \partial\pi_E(\mathcal{M}_A)$ is a partial observation with non-zero coordinates lying in the boundary of the completable region, then p_E has a unique completion to the log-linear model $\mathcal{M}_A$.*

1. *If E is contained in a proper facial set, then this completion lies in the boundary of the model $\partial\mathcal{M}_A$.*
2. *If E is not contained in a proper facial set, then this completion lies in the singular locus $B_{A,E}$.*

Proof. Let $p_E \in \partial\pi_E(\mathcal{M}_A)$ be a partial observation with non-zero coordinates in the boundary of the completable region. By Theorem 5.4.11, we may express the boundary of the completable region as

$$\partial\pi_E(\mathcal{M}_A) = \pi_E(\partial\mathcal{M}_A) \cup \pi_E(B_{A,E}).$$

If E is contained in a proper facial set, then $B_{A,E}$ is empty by Proposition 5.4.4 so that p_E has a completion $p \in \partial\mathcal{M}_A$. Similarly, if E is not contained in a proper facial set, then any completion of p_E has all non-zero coordinates so that p_E necessarily has a completion $p \in B_{A,E}$. We separate the proof into these cases:

If E is contained in a proper facial set, then p_E has a completion $p \in \partial\mathcal{M}_A$. Similar to Lemma 5.4.9, since $\operatorname{rank} A_E = \operatorname{rank} A - 1$ and $\operatorname{supp}(p)$ is a proper facial set, it follows that $\operatorname{supp}(p)$ is a facet and the smallest face containing E. We may let $\omega \in \ker A_E^T$ be an inner normal vector of the facial set $\operatorname{supp}(p)$ so that $\nu = A^T\omega$ satisfies $\nu_i = 0$ for $i \in \operatorname{supp}(p)$ and $\nu_i > 0$ for $i \in [n] \setminus \operatorname{supp}(p)$.

If $q \in \mathcal{M}_A$ is any other completion of p_E, then by Proposition 5.4.8 there exists $\alpha \in \mathbb{R}$ such that

$$q_i = p_i e^{\alpha\nu_i}$$

for $i \in \operatorname{supp}(p) \cap \operatorname{supp}(q)$. Since $E \subseteq \operatorname{supp}(q)$ and $\operatorname{supp}(p)$ is the smallest facial set containing E, we have $\operatorname{supp}(p) \subseteq \operatorname{supp}(q)$. Further, since $\nu_i = 0$ for $i \in \operatorname{supp}(p)$, we have that $q_i = p_i$ for $i \in \operatorname{supp}(p)$. Since $p, q \in \mathcal{M}_A$, we see

$$0 = \sum_{i\in[n]} (q_i - p_i) = \sum_{i\in[n]\setminus\operatorname{supp}(p)} q_i.$$

Since $q_i \geq 0$ for each i, this implies that $q_i = 0$ for $i \in [n] \setminus \operatorname{supp}(p)$ so that $q = p$.

If E is not contained in a proper facial set, then p_E has a completion $p \in B_{A,E}$. If $q \in \mathcal{M}_A$ is any other completion, then q also has non-zero coordinates. Let $\nu \in \operatorname{im} A^T \cap \ker \pi_E$ be a non-zero vector. By Proposition 5.4.8, there exists $\alpha \in \mathbb{R}$ such that

$$q_i = p_i e^{\alpha\nu_i}$$

for all $i \in [n]$. Recall that since $p \in B_{A,E}$ we have $\nu^T p = 0$ by Corollary 5.4.5. Observe that

$$\sum_{i\in[n]} p_i(e^{\alpha\nu_i} - \alpha\nu_i - 1) = \sum_{i\in[n]} (q_i - p_i) - \alpha\nu^T p = 0.$$

As each term of this summand is non-negative, this implies each summand is equal to zero. Since p has non-zero coordinates and ν is a non-zero vector, there is some index i such that $p_i \neq 0$ and $\nu_i \neq 0$, but $p_i(e^{\alpha\nu_i} - \alpha\nu_i - 1) = 0$. Therefore it must be that $\alpha = 0$ and $q = p$.

□

We turn our attention now to partial observations lying in the interior of the completable region $p_E \in \operatorname{int}(\pi_E(\mathcal{M}_A))$. Completions in this case may not be unique, however it can be seen that there are at most two.

Proposition 5.4.14. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$ and let $\nu \in \operatorname{im} A^T \cap \ker \pi_E$ be a non-zero vector. If $p_E \in \operatorname{int}(\pi_E(\mathcal{M}_A))$ is a partial observation in the interior of the completable region, then any two distinct completions $p, q \in \mathcal{M}_A$ are such that $\nu^T p$ and $\nu^T q$ are non-zero and have opposite signs.*

Proof. By Theorem 5.4.11, p_E has a completion $p \in \mathcal{M}_A^{>0} \setminus B_{A,E}$ and any other completion also lies in $\mathcal{M}_A^{>0} \setminus B_{A,E}$. By Proposition 5.4.8, if $p, q \in \mathcal{M}_A^{>0} \setminus B_{A,E}$ are any two completions, there exists $\alpha \in \mathbb{R}$ such that

$$q_i = p_i e^{\alpha\nu_i}$$

for all $i \in [n]$. Without loss of generality by interchanging p and q, we may assume that $\alpha > 0$. Observe

$$0 = \sum_i (q_i - p_i) = \sum_i p_i(e^{\alpha\nu_i} - 1) \geq \alpha \sum_i \nu_i p_i = \alpha\nu^T p.$$

Since $\alpha > 0$ and $\nu^T p \neq 0$, this implies that $\nu^T p < 0$. By the same process reversing the roles of p and q and by replacing α by $-\alpha$, one finds that $\nu^T q > 0$. □

Corollary 5.4.15. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$, and let $p_E \in \operatorname{int}(\pi_E(\mathcal{M}_A))$ be a partial observation in the interior of the completable region. Then p_E has at most two completions to $\mathcal{M}_A$.*

The number of completions to $\mathcal{M}_A$ for a partial observation in the interior of the completable region may vary. Indeed, the number of completions is depends on whether the set $E \subseteq [n]$ is contained in a proper facial set.

Theorem 5.4.16. *Let $E \subseteq [n]$ be such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$ and let $p_E \in \operatorname{int}(\pi_E(\mathcal{M}_A))$ be a partial observation lying in the interior of the completable region.*

1. *If E is contained in a proper facial set, then p_E has a unique completion to $\mathcal{M}_A$.*
2. *If E is not contained in a proper facial set, then p_E has two distinct completions to $\mathcal{M}_A$.*

Proof. 1. If $E \subseteq [n]$ is contained in a proper facial set F, we may let $\omega \in \ker A_E^T$ be an inner normal vector for F. Then $\nu = A^T\omega$ is such that $\nu_i = 0$ for $i \in F$ and $\nu_i > 0$ for $i \in [n] \setminus F$. Thus, $\nu^T p \geq 0$ for all $p \in \mathcal{M}_A$. From Proposition 5.4.14, it follows that p_E must have a unique completion to $\mathcal{M}_A$.

2. Assume $E \subseteq [n]$ is not contained in a proper facial set. If $p_E \in \text{int}(\pi_E(\mathcal{M}_A))$ is a partial observation with non-zero coordinates in the interior of the completable region, then by Theorem 5.4.11, p_E has a completion $p \in \mathcal{M}_A^{>0} \setminus B_{A,E}$. By Corollary 5.3.7 and Proposition 5.4.8, any completion $q \in X_A^{\geq 0}$ must have all non-zero coordinates and there must exist $\alpha \in \mathbb{R}$ satisfying

$$q_i = p_i e^{\alpha \nu_i}$$

for all $i \in [n]$. Setting $x = e^\alpha$, the completion $q \in X_A^{\geq 0}$ lies in the log-linear model $\mathcal{M}_A$ exactly when x is a root of the (Laurent) polynomial

$$f(x) = \sum_i p_i x^{\nu_i} - 1.$$

Conversely, any positive root of this polynomial yields a completion of p_E lying in the log-linear model $\mathcal{M}_A$. From Corollary 5.4.15, the polynomial f has at most two positive roots. We use continuity arguments to show that f always has two positive roots.

Note that since E is not contained in a proper facial set, the vector ν must have positive and negative coordinates corresponding to terms of f with positive and negative exponents. As these terms of f have positive coefficients, $f(x)$ can then be made arbitrarily large for $x > 1$ sufficiently large and for $0 < x < 1$ sufficiently small. Note that $x = 1$ is a root of f and $f'(1) = \nu^T p$. Thus, if $\nu^T p > 0$, then for $0 < x < 1$ sufficiently close to $x = 1$, $f(x)$ is negative. Thus $f(x)$ has a root in the open interval $(0, 1)$. Similarly, if $\nu^T p < 0$, then for $x > 1$ sufficiently close to $x = 1$, $f(x)$ is negative and $f(x)$ has a root in the open interval $(1, \infty)$. Thus, f always has two distinct positive roots yielding two completions of p_E. □

Remark 5..17. The upper bound of two completions to $\mathcal{M}_\mathcal{A}$ can be seen by applying Descartes' rule of signs to the polynomial $f(x)$ in the argument above. Since $f(x)$ is a Laurent polynomial with only one negative coefficient and we assume there is at least one positive zero, then there are two sign changes and hence, two positive zeros. However, Descartes' rule of signs does not guarantee that these zeros are distinct.

Example 5.4.18 demonstrates how the proof of Theorem 5.4.16 may be used to recover all completions of a partial observation $p_E \in \text{int}(\pi_E(\mathcal{M}_A))$ from a single completion.

Example 5.4.18. As in Example 5.4.12, let

$$A = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \in \mathbb{Z}_{\geq 0}^{3 \times 4}$$

and $E = \{1, 4\}$. Recall that the log-linear model is given by

$$\mathcal{M}_A = \{(x, y, z, w) \in \mathbb{R}^4_{\geq 0} : xw - yz = x + y + z + w - 1 = 0\},$$

and the singular locus is the hyperplane section of $\mathcal{M}_A^{>0}$ defined by

$$B_{A,E} = \{(x, y, z, w) \in \mathcal{M}_A^{>0} : y - z = 0\}.$$

Consider the partial observation $(\frac{1}{6}, \frac{1}{3}) \in \mathbb{R}^E_{\geq 0}$ and a completion $p = (\frac{1}{6}, \frac{1}{3}, \frac{1}{6}, \frac{1}{3}) \in \mathcal{M}_A^{>0} \setminus B_{A,E}$. Since $\nu^T p = \frac{1}{6} > 0$, there is another completion $q \in \mathbb{M}_A^{>0}$ such that $\nu^T q < 0$. This completion can be obtained by computing positive roots of the Laurent polynomial

$$f(x) = \sum_i p_i x^{\nu_i} - 1 = \frac{1}{3}x - \frac{1}{2} + \frac{1}{6}x^{-1}.$$

One finds that $x = \frac{1}{2}$ is a root of $f(x)$ so that

$$q = p\left(\frac{1}{2}\right)^{(0,1,-1,0)} = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{3}, \frac{1}{3}\right) \in \mathcal{M}_A^{>0} \setminus B_{A,E}$$

is the other completion. Theorem 5.4.16 implies that these are the only completions of the partial observation p_E to the log-linear model $\mathcal{M}_A$.

5.5 Describing the Completable Region

Given a subset $E \subseteq [n]$ and an integer matrix $A \in \mathbb{Z}^{k \times n}_{\geq 0}$, the completable region $\pi_E(\mathcal{M}_A)$ is a full-dimensional semialgebraic set—it is defined by polynomial inequalities. A *semialgebraic description* of the completable region $\pi_E(\mathcal{M}_A)$ is such a description of the completable region by polynomial inequalities. Obtaining a semialgebraic description of the completable region is difficult in general. However, in this section we will provide examples where obtaining the complete semialgebraic description of the completable region is possible by elementary methods. From Theorem 5.4.11, the algebraic boundary of the completable region $\pi_E(\mathcal{M}_A)$ is the union of the Zariski closure of the image of the boundary $\pi_E(\partial\mathcal{M}_A)$ and the Zariski closure of the image of the singular locus $\pi_E(B_{A,E})$. In both cases, the defining ideal can be computed explicitly via elimination as long as the defining ideal of the boundary $\partial\mathcal{M}_A$ and the defining ideal of the singular locus $B_{A,E}$ are known. But first we consider the simpler problem of computing defining equations for the boundary of the completable region as in Example 5.4.12.

Definition 5.5.1. *The algebraic boundary of the completable region $\pi_E(\mathcal{M}_A)$ is the Zariski closure of the Euclidean boundary $\partial\pi_E(\mathcal{M}_A)$.*

The defining ideal of the boundary $\partial\mathcal{M}_A$ is generated by the product of the coordinates $\prod_i x_i$, the hyperplane $\sum_i x_i - 1$, and I_A. However, the defining ideal of $B_{A,E}$ my be difficult to obtain. For instance, given $\nu \in \operatorname{im} A^T \cap \ker \pi_E$, the hyperplane section defined

by $\nu^T p = 0$ may intersect the Zariski closure of the log-linear model $\mathcal{M}_A$ in more than one component. However, from Proposition 5.4.4, at most one of these components intersects the model $\mathcal{M}_A$. Techniques from real algebraic geometry such as the Positivestellensatz and its variants produce methods of choosing this component in general. See Lombardi et al. [2020]; Goharshady et al. [2022] for more information on these techniques and their use in applications.

We provide pseudo-code for computing the defining ideal of the algebraic boundary. We note the reliance on several methods which we consider as black-boxes: one for computing the toric ideal I_A, one for producing the minimal primes of an ideal, one for determining whether an ideal has a positive zero, and one for computing elimination ideals.

Pseudo–Code (Defining Ideal of the Algebraic Boundary)**.**

Input	*A: An integer matrix in $\mathbb{Z}_{\geq 0}^{k\times n}$*
	E: A subset such that $\lvert E\rvert = \operatorname{rank} A_E = \operatorname{rank} A - 1$
Output	*I: The defining ideal of the algebraic boundary.*

1. *Find a non-zero vector $\nu \in A^T(\ker A_E^T)$.*
2. *Compute the minimal primes of the ideal $J = I_A + \langle \sum_i x_i - 1, \nu^T x\rangle$.*
3. *Compute $I(B_{A,E})$ as the minimal prime of J whose corresponding variety intersects the log-linear model $\mathcal{M}_A$.*
4. *Set I_1 as the ideal obtained by eliminating the variables x_i for $i \in [n] \setminus E$ from the defining ideal of the singular locus $I(B_{A,E})$.*
5. *Compute $I(\partial\mathcal{M}_A)$ as the ideal $I_A + \langle \sum_i x_i - 1, \prod_i x_i\rangle$.*
6. *Set I_2 as the ideal obtained by eliminating the variables x_i for $i \in [n] \setminus E$ from the defining ideal of the boundary of the model $I(\partial\mathcal{M}_A)$.*
7. *Return I as the radical of the ideal $I_1 I_2$.*

We note that in general, inequalities obtained from the defining equations of the algebraic boundary do not give a semialgebraic description of the completable region. This is demonstrated in Example 5.5.2.

Example 5.5.2. Let

$$A = \begin{pmatrix} 3 & 4 & 0 & 3 & 2 \\ 0 & 1 & 2 & 1 & 1 \\ 2 & 0 & 3 & 1 & 2 \end{pmatrix} \in \mathbb{Z}_{\geq 0}^{3\times 5}$$

and $E = \{4, 5\}$. We use coordinates x, y, z, u, v for $\mathbb{R}^5_{\geq 0}$. The polytope P_A is a triangle containing two interior points as illustrated in Figure 5.5.5 and the defining ideal I_A is generated by three binomials,

$$I_A = \langle u^2 - yv, v^3 - xzu, uv^2 - xyz \rangle.$$

We first consider the portion of the algebraic boundary which is the image of the singular locus. The kernel $\ker A_E^T$ is generated by the single vector $\omega = (1, -4, 1)$ so that we may write $\nu = (5, 0, -5, 0, 0)$ and

$$B_{A,E} = \{(x, y, z, u, v) \in \mathcal{M}_A^{>0} : x - z = 0\}.$$

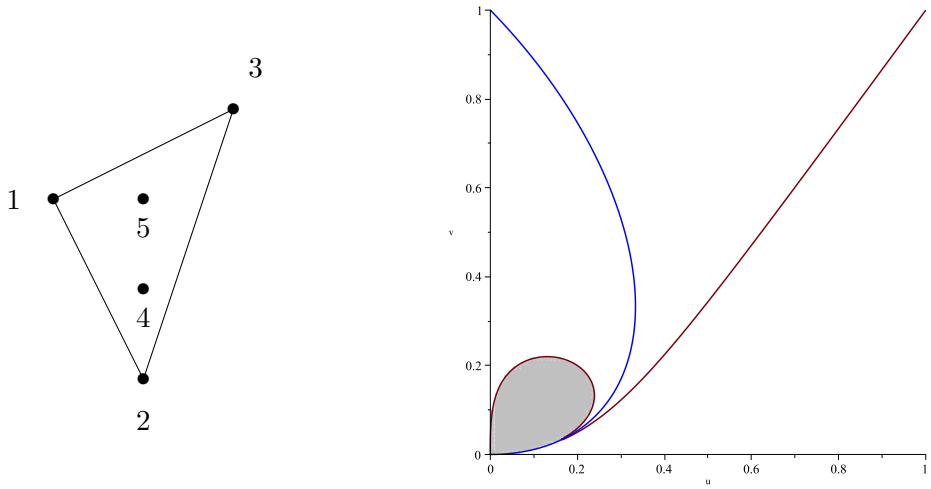


Figure 5.5.5: (Left) The polytope P_A for Example 5.5.2. (Right) The completable region and curves defined by $f(u, v) = 0$ and $g(u, v) = 0$.

As the ideal

$$I(B_{A,E}) = \langle u^2 - yv, v^3 - xzu, uv^2 - xyz, x + y + z + u + v - 1, x - z \rangle,$$

is prime, it is the defining ideal of the singular locus. Using `Macaulay2` to eliminate the variables x, y, and z from the ideal $I(B_{A,E})$, the eliminant is generated by the single polynomial

$$f(u, v) = u^5 + 2u^4v + 3u^3v^2 + 2u^2v^3 + uv^4 - 4v^5 - 2u^3v - 2u^2v^2 - 2uv^3 + uv^2.$$

Thus, the algebraic boundary of the completable region $\pi_E(\mathcal{M}_A)$ is defined by the single polynomial equation $f(u, v) = 0$.

Note if $p \in \partial\mathcal{M}_A$, then some coordinate is equal to zero so that $\operatorname{supp}(p)$ is a proper facial set. Since the only facial set intersecting E is the set $[5] = \{1, 2, 3, 4, 5\}$, it follows that $\operatorname{supp}(p) \cap E = \emptyset$ and $\pi_E(p) = (0, 0)$. That is, $\pi_E(\partial\mathcal{M}_A) = \{(0, 0)\}$. Thus, the algebraic boundary is defined by the single equation $f(u, v) = 0$.

Contrary to this, the completable region is not described by either inequality $f(u,v) \geq 0$ or $f(u,v) \leq 0$. Indeed, the completable region is illustrated in Figure 5.5.5. The red curve is the algebraic boundary defined by the equation $f(u,v) = 0$, but only the shaded area is the completable region.

To obtain a semialgebraic description of the completable region, consider a partial observation $(u,v) \in \pi_E(\mathcal{M}_A)$ with non-zero coordinates and a completion $(x,y,z,u,v) \in \mathcal{M}_A$. Multiplying through the equation $x+y+z+u+v = 1$ by zuv and using the generators of the toric ideal I_A, one finds that z satisfies the quadratic equation

$$uvz^2 + (u^3 + u^2v + uv^2 - uv)z + v^4 = 0.$$

Conversely, using the generators of I_A, any solution to this quadratic yields a completion of (u,v) to the log-linear model $\mathcal{M}_A$. Thus, this quadratic has all non-negative solutions and the sum of these solutions

$$\frac{-1}{uv}\left(u^3 + u^2v + uv^2 - uv\right) \geq 0$$

must also be non-negative. Since u and v are non-zero, this implies that

$$g(u,v) = -u^2 - uv - v^2 + v \geq 0.$$

The blue curve in Figure 5.5.5 is defined by the equation $g(u,v) = 0$ and together the inequalities $f(u,v) \geq 0$ and $g(u,v) \geq 0$ provide a semialgebraic description of the completable region. Precisely,

$$\pi_E(\mathcal{M}_A) = \{(u,v) \in \mathbb{R}^2_{\geq 0} : f(u,v) \geq 0,\ g(u,v) \geq 0\}.$$

As in Example 5.5.2, it is often the case that the ideal $I_A + \langle \sum_i x_i - 1, \nu^T x \rangle$ is prime, in which case it is equal to the defining ideal of the singular locus $I(B_{A,E})$. The following example demonstrates a situation in which the boundary of the log-linear model $\mathcal{M}_A$ maps onto the boundary of the completable region.

Example 5.5.3. A hierarchical model is a log-linear model determined by the data of a simplicial complex with positive integer weights given at each of the vertices. Each vertex corresponds to a random variable with the weight of the vertex being the number of outcome states of the variable. The faces of the simplicial structure encode dependencies among these variables. We consider the hierarchical model associated to the length two segment with each vertex weight equal to two. This corresponds to three binary random variables X_1, X_2, and X_3 with the probability of their outcomes related by certain relations determined by the dependencies.

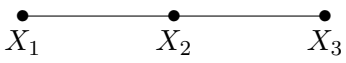


Figure 5.5.6: The length two line segment

The model has coordinates p_{ijk} given by the probability that $X_1 = i$, $X_2 = j$, and $X_3 = k$. As described in Chapter 9 of Sullivant [2018], a hierarchical model is a log-linear model and so can be described by an integer matrix $A \in \mathbb{Z}_{\geq 0}^{k \times n}$. For the hierarchical model associated to Figure 5.5.6 and all weights equal to two, the matrix A is given by

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix},$$

where the columns of A index the coordinates p_{ijk} in lexicographical order. One finds that $\operatorname{rank} A = 6$ and the toric ideal I_A is generated by the following binomials

$$I_A = \langle p_{112}p_{211} - p_{111}p_{212}, p_{122}p_{211} - p_{121}p_{222} \rangle.$$

We demonstrate two subsets $E \subseteq [n]$ for which a semialgebraic description of the completable region $\pi_E(\mathcal{M}_A)$ can be easily computed and show how a semialgebraic description can be obtained in general for this model.

Let $E = \{1, 2, 3, 4, 5\} \subseteq [8]$ so that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1 = 5$. In other words, we will observe the coordinates p_{111}, p_{112}, p_{121}, p_{122}, and p_{211}. `Macaulay2` can be used to verify that E is contained in the proper facial set $\{1, 2, 3, 4, 5, 6\}$. Thus, the singular locus $B_{A,E}$ is empty and the algebraic boundary can be obtained by eliminating p_{212}, p_{221}, and p_{222} from the ideal $I_A + \langle \sum_{i,j,k} p_{ijk} - 1, \prod_{i,j,k} p_{ijk} \rangle$. We find that the algebraic boundary is given by the single polynomial

$$f(p_{111}, p_{112}, p_{121}, p_{122}, p_{211}) = -p_{111}^2 - p_{111}p_{121} - p_{111}p_{122} - p_{111}p_{211} - p_{112}p_{211} + p_{111}.$$

Further, a semialgebraic description of the completable region is given by

$$\pi_E(\mathcal{M}_A) = \{p_E \in \mathbb{R}_{\geq 0}^5 : f(p_{111}, p_{112}, p_{121}, p_{122}, p_{211}) \geq 0\}.$$

Indeed, given a partial observation $p_E \in \pi_E(\mathcal{M}_A)$ with non-zero coordinates, the remaining coordinates can be computed in rational functions of the observed coordinates p_{111}, p_{112}, p_{121}, p_{122}, and p_{211} as

$$\begin{aligned} p_{212} &= \frac{p_{112}p_{211}}{p_{111}} \\ p_{221} &= \frac{p_{121} f(p_{111}, p_{112}, p_{121}, p_{122}, p_{211})}{p_{111}(p_{121} + p_{122})} \\ p_{222} &= \frac{p_{122} f(p_{111}, p_{112}, p_{121}, p_{122}, p_{211})}{p_{111}(p_{121} + p_{122})}. \end{aligned}$$

As our partial observation has non-zero coordinates, this completion lies in the log-linear model $\mathcal{M}_A$ exactly when $f(p_{111}, p_{112}, p_{121}, p_{122}, p_{211}) \geq 0$.

We note that for any subset $E \subseteq [n]$ which is contained in a proper facial set, the above analysis can be carried out to obtain a semialgebraic description of the completable region. Indeed, for the hierarchical model $\mathcal{M}_A$, if $E \subseteq [n]$ is contained in a proper facial set, then the coordinates p_i for $i \in [n] \setminus E$ can be explicitly computed in rational functions in the coordinates p_i for $i \in E$ via the generators of the toric ideal I_A and the relation $\sum_i p_i = 1$.

For a subset $E \subseteq [n]$ such that $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$ that is not contained in a proper facial set, then by choosing a coordinate p_{ijk} not indexed by E, the remaining coordinates are rational functions in p_{ijk} and the coordinates indexed by E. Indeed, by observing the generators of the toric ideal I_A, each of these rational functions has the form $f(p_E)p_{ijk}$ or $f(p_E)p_{ijk}^{-1}$ where $f(p_E)$ is a (Laurent) monomial in the coordinates indexed by E. Subsituting these values into the relation $\sum_{ijk} p_{ijk} = 1$, one finds that p_{ijk} satisfies a quadratic relation. As in Example 5.5.2, this quadratic has all non-negative roots for $p_E \in \pi_E(\mathcal{M}_A)$ and one obtains an additional polynomial inequality that the partial observation p_E must satisfy. Since the remaining coordinates are given as rational functions with positive coefficients in p_{ijk} and the coordinates indexed by E, any solution of this quadratic corresponds to a completion in $\mathcal{M}_A$ of p_E. Thus, one obtains a semialgebraic description of the completable region.

In either of the cases above, given a subset $E \subseteq [n]$ satisfying $|E| = \operatorname{rank} A_E = \operatorname{rank} A - 1$, one can obtain a semialgebraic description of the completable region $\pi_E(\mathcal{M}_A)$ for this hierarchical model.

Bibliography

4ti2 team. 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at https://4ti2.github.io.

J. Adams. Causal inference and causal discovery with latent variables, 2024. PhD thesis. University of Copenhagen.

C. Améndola, M. Drton, A. Grosdos, R. Homs, and E. Robeva. Third-order moment varieties of linear non-gaussian graphical models. *Information and Inference: A Journal of the IMA*, 12(3):1405–1436, 2023. ISSN 2049-8772. doi: 10.1093/imaiai/iaad007.

C. Améndola, T. Boege, B. Hollering, and P. Misra. Structural identifiability of graphical continuous Lyapunov models. *arXiv:2510.04985*, 2025.

R. F. Barber, M. Drton, N. Sturma, and L. Weihs. Half-trek criterion for identifiability of latent variable models. *The Annals of Statistics*, 50(6):3174 – 3196, 2022. doi: 10.1214/22-AOS2221.

D. I. Bernstein, G. Blekherman, and R. Sinn. Typical and generic ranks in matrix completion. *Linear Algebra and its Applications*, 585:71–104, 2020.

F. Bleile, S. Lumpp, and M. Drton. Efficient learning of stationary diffusions with Stein-type discrepancies. *arXiv:2601.16597*, 2026.

T. Boege and L. Solus. Real birational impicitization for statistical models. *arXiv:2410.23102*, 2024.

T. Boege, M. Drton, B. Hollering, S. Lumpp, P. Misra, and D. Schkoda. Conditional independence in stationary distributions of diffusions. *Stochastic Processes and their Applications*, 184:104604, 2025. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2025.104604.

K. A. Bollen. *Introduction Structural Equations with Latent Variables*. John Wiley & Sons, Ltd, 1989. ISBN 9781118619179.

R. J. Bowden, R. J. Bowden, and D. A. Turkington. *Instrumental variables*. Cambridge university press, 1990.

P. Breiding, F. Gesmundo, M. Michałek, and N. Vannieuwenhoven. Algebraic compressed sensing. *Applied and Computational Harmonic Analysis*, 65:374–406, 2023.

D. R. Brillinger. The calculation of cumulants via conditioning. *Annals of the Institute of Statistical Mathematics*, 21(1):215–218, 1969.

C. Brito and J. Pearl. A new identification condition for recursive models with correlated errors. *Structural Equation Modeling*, 9(4):459–474, 2002.

P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer Texts in Statistics. Springer International Publishing, 2016. ISBN 9783319298528.

M. Cai, C. O. Recke, and T. Yahl. Completions to discrete probability distributions in log-linear models. *Algebraic Statistics*, 15(2):225–247, Dec. 2024. ISSN 2693-2997. doi: 10.2140/astat.2024.15.225.

S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385, 1981.

M. Casanellas, S. Petrović, and C. Uhler. Algebraic statistics in practice: Applications to networks. *Annual Review of Statistics and Its Application*, 2019.

P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3): 287–314, 1994. ISSN 0165-1684. Higher Order Statistics.

P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010. ISBN 9780080884943.

D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics. Springer International Publishing, 2015. ISBN 9783319167206.

D. A. Cox, J. B. Little, and H. K. Schenck. *Toric varieties*, volume 124 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2011. ISBN 978-0-8218-4819-7. doi: 10.1090/gsm/124.

J. Cummings and B. Hollering. Computing implicitizations of multi-graded polynomial maps. *Journal of Symbolic Computation*, 132:102459, 2026. ISSN 0747-7171. doi: https://doi.org/10.1016/j.jsc.2025.102459.

P. Dettling, R. Homs, C. Améndola, M. Drton, and N. R. Hansen. Identifiability in continuous Lyapunov models. *SIAM Journal on Matrix Analysis and Applications*, 44 (4):1799–1821, 2023.

P. Dettling, M. Drton, and M. Kolar. On the lasso for graphical continuous Lyapunov models. *3rd Conference on Causal Learning and Reasoning (CLeaR), Proceedings of Machine Learning Research*, 236:514–550, 2024.

P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, 26(1):363–397, 1998. ISSN 0090-5364. doi: 10.1214/aos/1030563990.

K. Domino, P. Gawron, and L. Pawela. Efficient computation of higher-order cumulant tensors. *SIAM Journal on Scientific Computing*, 40(3):A1590–A1610, 2018. doi: 10.1137/17M1149365.

M. Drton. *Algebraic problems in structural equation modeling*, volume 77 of *Advanced studies in pure mathematics*. Mathematical Society of Japan, 2018.

M. Drton and T. S. Richardson. Binary models for marginal independence. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(2):287–309, 2008.

M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on algebraic statistics*. Springer, 2009. ISBN 9783764389055.

G. Ewald. *Combinatorial convexity and algebraic geometry*, volume 168 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996. ISBN 0-387-94755-8. doi: 10.1007/978-1-4612-4044-0.

K. Fitch. Learning directed graphical models from Gaussian data. *arXiv:1906.08050*, 2020.

R. Foygel, J. Draisma, and M. Drton. Supplement to "half-trek criterion for generic identifiability of linear structural equation models", 2012a. ISSN 00905364, 21688966.

R. Foygel, J. Draisma, and M. Drton. Half-trek criterion for generic identifiability of linear structural equation models. *The Annals of Statistics*, 40(3):1682 – 1713, 2012b. doi: 10.1214/12-AOS1012.

W. Fulton. *Introduction to toric varieties*, volume 131 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1993. ISBN 0-691-00049-2. doi: 10.1515/9781400882526. The William H. Roever Lectures in Geometry.

Z. Gajic and M. T. J. Qureshi. *Lyapunov Matrix Equation in System Stability and Control*, volume 195. Elsevier, 1995.

L. D. García-Puente, S. Spielvogel, and S. Sullivant. Identifying causal effects with computer algebra. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, UAI'10, page 193–200. AUAI Press, 2010.

D. Geiger, C. Meek, and B. Sturmfels. On the toric algebra of graphical models. *The Annals of Statistics*, 34(3):1463 – 1492, 2006. doi: 10.1214/009053606000000263.

A. K. Goharshady, S. Hitarth, F. Mohammadi, and H. J. Motwani. Template-based program synthesis using stellensätze. *arXiv:2209.03602*, 2022.

C. Görgen, A. Maraj, and L. Nicklasson. Staged tree models with toric structure. *Journal of Symbolic Computation*, 113:242–268, 2022. ISSN 0747-7171. doi: https://doi.org/10.1016/j.jsc.2022.04.006.

G. Gorin, J. J. Vastola, M. Fang, and L. Pachter. Interpretable and tractable models of transcriptional noise for the rational design of single-molecule quantification experiments. *Nature Communications*, 13(1):7620, 2022.

D. R. Grayson and M. E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at http://www2.macaulay2.com.

T. N. E. Greville. Some applications of the pseudoinverse of a matrix. *SIAM Review*, 2 (1):15–22, 1960. ISSN 00361445.

V. Guan, J. Janssen, H. Rahmani, A. Warren, S. Zhang, E. Robeva, and G. Schiebinger. Identifying drift, diffusion, and causal structure from temporal snapshots. *arXiv:2410.22729*, 2024.

V. Guan, J. Janssen, N. Lanzetti, A. Terpin, G. Schiebinger, and E. Robeva. Gradient-flow SDEs have unique transient population dynamics. *arXiv:2505.21770*, Oct. 2025. doi: 10.48550/arXiv.2505.21770.

N. R. Hansen. A trek rule for the Lyapunov equation. *Algebraic Statistics*, 16:95–112, 09 2025. doi: 10.2140/astat.2025.16.95.

S. Hoşten and S. Sullivant. A finiteness theorem for Markov bases of hierarchical models. *Journal of Combinatorial Theory. Series A*, 114(2):311–321, 2007. ISSN 0097-3165. doi: 10.1016/j.jcta.2006.06.001.

S. Hoşten and S. Sullivant. Gröbner bases and polyhedral geometry of reducible and cyclic models. *Journal of Combinatorial Theory, Series A*, 100(2):277–301, 2002. ISSN 0097-3165. doi: https://doi.org/10.1006/jcta.2002.3301.

A. Hyttinen, F. Eberhardt, and P. O. Hoyer. Learning linear cyclic causal models with latent variables. *Journal of Machine Learning Research*, 13(109):3387–3439, 2012.

M. Jacobsen. A brief account of the theory of homogeneous Gaussian diffusions in finite dimensions. In *Vol. 1 Proceedings of the Third Finnish-Soviet Symposium on Probability Theory and Mathematical Statistics, Turku, Finland*, pages 86–94, Berlin, Boston, 1993. De Gruyter. ISBN 9783112314203. doi: doi:10.1515/9783112314203-007.

Z. J. Jurek. An integral representation of operator-selfdecomposable random variables. *Bulletin de l'Académie Polonaise des Sciences. Série des Sciences Mathématiques*, 30, 01 1982.

T. Kahle, K. Kubjas, M. Kummer, and Z. Rosen. The geometry of rank-one tensor completion. *SIAM Journal on Applied Algebra and Geometry*, 1(1):200–221, 2017.

G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nature reviews Molecular cell biology*, 9(10):770–780, 2008.

F. J. Király, L. Theran, R. Tomioka, and T. Uno. The algebraic combinatorial approach for low-rank matrix completion. *arXiv:1211.4116*, 2012.

F. J. Király, L. Theran, and R. Tomioka. The algebraic combinatorial approach for low-rank matrix completion. *Journal of Machine Learning Research*, 16(1):1391–1436, 2015.

T. C. Koopmans. Identification problems in economic model construction. *Econometrica*, 17(2):125–144, 1949. ISSN 00129682, 14680262.

T. C. Koopmans and O. Reiersøl. The Identification of Structural Characteristics. *The Annals of Mathematical Statistics*, 21(2):165 – 181, 1950. doi: 10.1214/aoms/1177729837.

R. Kozlov and O. Kozlova. Investigation of stability of nonlinear continuous-discrete models of economic dynamics using vector Lyapunov function. ii. *Journal of Computer and Systems Sciences International*, 48:370–378, 2009.

K. Kubjas and Z. Rosen. Matrix completion for the independence model. *Journal of Algebraic Statistics*, 8(1):1–21, 2017. ISSN 1309-3452. doi: 10.18409/jas.v8i1.50.

S. L. Lauritzen and T. S. Richardson. Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64 (3):321–348, 2002.

B. Liu. Identifiability of VAR(1) model in a stationary setting. *arXiv:2504.03466*, 2025.

H. Lombardi, D. Perrucci, and M.-F. Roy. An elementary recursive bound for effective Positivstellensatz and Hilbert's 17th problem. *Mem. Amer. Math. Soc.*, 263(1277): v+125, 2020. ISSN 0065-9266. doi: 10.1090/memo/1277.

L. Lorch, A. Krause, and B. Schölkopf. Causal modeling with stationary diffusions. In *Proceedings of the 27th international conference on artificial intelligence and statistics*, volume 238 of *Proceedings of machine learning research*, pages 1927–1935. PMLR, May 2024.

L. Lorch, J. Zhang, C. Bunne, A. Krause, B. Schölkopf, and C. Uhler. Latent causal diffusions for single-cell perturbation modeling. *arXiv:2601.15341*, 2026.

A. M. Lyapunov. The general problem of the stability of motion. *International journal of control*, 55(3):531–534, 1992.

J. R. Magnus. On differentiating eigenvalues and eigenvectors. *Econometric Theory*, 1 (2):179–191, 1985. doi: 10.1017/S0266466600011129.

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons, 3rd edition, 2019. ISBN 978-1-119-54120-2.

A. Markham, C. O. Recke, and N. R. Hansen. Causal inference with latent operator-selfdecomposable distributions. *Working paper*, 2026+.

H. Masuda. On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli*, 10(1):97 – 120, 2004.

P. McCullagh. *Tensor Methods in Statistics: Second Edition*. Dover Publications, 2018. ISBN 9780486832692.

G. Mesters and P. Zwiernik. Non-independent component analysis. *The Annals of Statistics*, 52(6):2506 – 2528, 2024. doi: 10.1214/24-AOS2373.

G. Michailidis and F. d'Alché Buc. Autoregressive models for gene regulatory network inference: Sparsity, stability and causality issues. *Mathematical Biosciences*, 246(2): 326–334, 2013.

W. Niemiro and Ł. Rajkowski. Local dependence graphs for discrete time processes. In *Conference on Causal Learning and Reasoning*, pages 772–790. PMLR, 2023.

K. M. Passino, A. N. Michel, and P. J. Antsaklis. Lyapunov stability of a class of discrete event systems. *IEEE transactions on automatic control*, 39(2):269–279, 2002.

J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, Cambridge, MA, USA, 2017. ISBN 9780262037310.

K. B. Petersen, M. S. Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.

J. C. Rajapakse and P. A. Mundra. Stability of building gene regulatory networks with sparse autoregressive models. *BMC Bioinformatics*, 12(13):S17, 2011.

C. O. Recke and N. R. Hansen. Identifiability and estimation in continuous Lyapunov models. *arXiv: 2603.17142*, 2026.

C. O. Recke, S. Lumpp, N. Kushnerchuk, J. Oldekop, J. Li, J. I. Coons, and E. Robeva. Identifiability in graphical discrete Lyapunov models. *arXiv: 2601.21818*, 2026.

O. Reiersøl. Identifiability of a linear relation between variables which are subject to error. *Econometrica*, 18(4):375–389, 1950. ISSN 00129682, 14680262.

E. Robeva and J.-B. Seby. Multi-trek separation in linear structural equation models. *SIAM Journal on Applied Algebra and Geometry*, 5(2):278–303, 2021. doi: 10.1137/20M1316470.

M. Rohbeck, B. Clarke, K. Mikulik, A. Pettet, O. Stegle, and K. Ueltzhöffer. Bicycle: Intervention-based causal discovery with cycles. In *Proceedings of the third conference on causal learning and reasoning*, volume 236 of *Proceedings of machine learning research*, pages 209–242. PMLR, Apr. 2024.

K.-i. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2nd edition, 2013.

K.-i. Sato and M. Yamazato. Stationary processes of Ornstein-Uhlenbeck type. In *Probability Theory and Mathematical Statistics*, pages 541–551, Berlin, Heidelberg, 1983. Springer. ISBN 978-3-540-38701-5. doi: 10.1007/BFb0072949.

K.-i. Sato and M. Yamazato. Operator-selfdecomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type. *Stochastic Processes and their Applications*, 17(1):73–100, 1984.

D. Schkoda, E. Robeva, and M. Drton. Causal discovery of linear non-Gaussian causal models with unobserved confounding. *arXiv: 2408.04907*, 2024.

M. G. Sethuraman, R. Lopez, R. Mohan, F. Fekri, T. Biancalani, and J.-C. Hütter. NODAGS-Flow: Nonlinear cyclic causal structure learning. In *International Conference on Artificial Intelligence and Statistics*, pages 6371–6387. PMLR, 2023.

S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen. A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(72):2003–2030, 2006.

C. A. Sims. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48, 1980.

A. Singer and M. Cucuringu. Uniqueness of low-rank matrix completion by rigidity theory. *SIAM Journal on Matrix Analysis and Applications*, 31(4):1621–1641, 2010.

A. Sokol and N. R. Hansen. Causal interpretation of stochastic differential equations. *Electron. J. Probab.*, 19(100):1–24, 2014.

F. Sottile. Toric ideals, real toric varieties, and the moment map. In *Topics in algebraic geometry and geometric modeling*, volume 334 of *Contemporary Mathematics*, pages 225–240. American Mathematical Society, Providence, RI, 2003. doi: 10.1090/conm/334/05984.

N. Sturma, M. Kranzlmueller, I. Portakal, and M. Drton. Matching criterion for identifiability in sparse factor analysis. *Psychometrika*, 2025.

S. Sullivant. *Algebraic Statistics*. Graduate Studies in Mathematics. American Mathematical Society, 2018. ISBN 9781470435172.

S. Sullivant, K. Talaska, and J. Draisma. Trek separation for gaussian graphical models. *The Annals of Statistics*, 38(3):1665 – 1685, 2010. doi: https://doi.org/10.1214/09-AOS760.

G. H. Thomson. The proof or disproof of the existence of general ability. *British Journal of Psychology*, 9(3):321, 1919.

D. Tramontano, Y. Kivva, S. Salehkaleybar, M. Drton, and N. Kiyavash. Causal effect identification in lvLiNGAM from higher-order cumulants. *arXiv: 2506.05202*, 2025.

K. Urbanik. Lévy's probability measures on euclidean spaces. *Studia Mathematica*, 44 (2):119–148, 1972.

A. W. v. d. Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. ISBN 9780511802256.

L. van der Krieke, F. J. Blaauw, A. C. Emerencia, H. M. Schenk, J. P. J. Slaets, E. H. Bos, P. de Jonge, and B. F. Jeronimus. Temporal dynamics of health and well-being: A crowdsourcing approach to momentary assessments and automated generation of personalized feedback. *Psychosomatic medicine*, 79(2):213–223, 2017.

G. van Seeventer and S. Salehkaleybar. Sign identifiability of causal effects in stationary stochastic dynamical systems. *arXiv: 2603.08311*, 2026.

G. Varando and N. R. Hansen. Graphical continuous Lyapunov models. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pages 989–998. PMLR, 03–06 Aug 2020.

L. Wang, N. Trasanidis, T. Wu, G. Dong, M. Hu, D. E. Bauer, and L. Pinello. Dictys: dynamic gene regulatory network dissects developmental continuum with single-cell multiomics. *Nature Methods*, 20(9):1368–1378, 2023.

Y. S. Wang and M. Drton. Causal discovery with unobserved confounding and non-Gaussian data. *Journal of Machine Learning Research*, 24(271):1–61, 2023.

S. J. Wolfe. A characterization of certain stochastic integrals. *Stochastic Processes and their Applications (Tenth conference on stochastic processes and their applications: Montréal, Canada, 23–28 August 1981)*, 12(2):117–170, 1982. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(82)90039-4.

S. Wright. The method of path coefficients. *The annals of mathematical statistics*, 5(3): 161–215, 1934.

X. Xu and Q.-W. Wang. On the solutions of a class of tensor equations. *Linear and Multilinear Algebra*, 70(21):6141–6154, 2022.

W. C. Young, K. Y. Yeung, and A. E. Raftery. Identifying dynamical time series model parameters from equilibrium samples, with application to gene regulatory networks. *Statistical modelling*, 19(4):444–465, 2019.

J. Zbigniew and M. David. *Operator-Limit Distributions in Probability Theory*. Wiley Series in Probability and Mathematical Statistics. Probability and Mathematical Statistics. John Wiley, New York, 1993. ISBN 978-0-471-58595-4.