

# Graphical Continuous Lyapunov Models

Philipp Maximilian Dettling

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung eines

**Doktors der Naturwissenschaften (Dr. rer. nat.)**

genehmigten Dissertation.

**Vorsitz:**

Prof. Donna Ankerst, Ph.D.

**Prüfende der Dissertation:**

1. Prof. Mathias Drton, Ph.D.
2. Prof. Dr. Niki Kilbertus

Die Dissertation wurde am 17.06.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 13.09.2024 angenommen.



# Abstract

Graphical continuous Lyapunov models offer a new perspective on modeling causally interpretable dependence structure in multivariate data by treating each independent observation as a one-time cross-sectional snapshot of a temporal process. This dissertation provides basic research for the new approach in the field of graphical modeling and is divided into three parts.

The covariance matrix for the data is obtained by solving a continuous Lyapunov equation that is parametrized by the drift matrix of the dynamic process. In this context, different statistical models postulate different sparsity patterns in the drift matrix, and it becomes a crucial problem to clarify whether a given sparsity assumption allows one to uniquely recover the drift matrix parameters from the covariance matrix of the data. We study this identifiability problem by representing sparsity patterns by directed graphs. Our main result proves that the drift matrix is globally identifiable if and only if the graph for the sparsity pattern is simple (i.e., does not contain directed two-cycles). Moreover, we present a necessary condition for generic identifiability and provide a computational classification of small graphs with up to 5 nodes.

Each graphical continuous Lyapunov model assumes the drift matrix to be sparse, with support determined by a directed graph. A natural approach to model selection in this setting is to use an  $\ell_1$ -regularization technique that, based on a given sample covariance matrix, seeks to find a sparse approximate solution to the Lyapunov equation. We study the model selection properties of the resulting lasso technique to arrive at a consistency result. Our detailed analysis reveals that the involved irrepresentability condition is surprisingly difficult to satisfy. While this may prevent asymptotic consistency in model selection, our numerical experiments indicate that even if the theoretical requirements for consistency are not met, the lasso approach is able to recover relevant structure of the drift matrix and is robust to aspects of model misspecification. The analysis concludes by applying the lasso approach in combination with the (extended) Bayesian Information Criterion to real-world data. Despite the simplicity of the approach, the method is able to recover many important connections of an among scientists accepted protein-signalling network.

Advances in Mixed Integer Quadratic Programming (MIQP) allowed the best subset selection ( $\ell_0$ -penalized) to compete with  $\ell_1$ -penalized approaches. We rigorously study the strengths and weaknesses of the best subset selection for Lyapunov models (BSSLM). First, we provide examples that show how  $\ell_1$ -penalized methods tend to produce a lot of undesired symmetry in the estimates, which can be resolved by using the BSSLM. Making the connection to the best subset selection for regression problems, we show how the problems are set up to make them feasible for MIQP solvers. Analyzing the time consumption, we suggest to settle for smaller problem sizes up to  $25 \times 25$ . In settings where the nonzero entries are clearly distinguishable

## *Abstract*

from zero, the BSSLM performs much better than the  $\ell_1$ -competitors. Furthermore, we show that the BSSLM is able to jointly estimate the drift and a diagonal volatility matrix. In particular, regarding the quality of the estimate of the drift matrix, the method shows the best performance when compared to the  $\ell_1$ -competitors in our simulation setting. To conclude, we present a potential application by estimating a protein-signaling network purely from observational data. There, we include the information regarding the diagonal of  $C$  by coloring the nodes.

# Zusammenfassung

Grafische stetige Lyapunov-Modelle bieten eine neue Perspektive für die Modellierung kausal interpretierbarer Abhängigkeitsstrukturen in multivariaten Daten, indem sie jede unabhängige Beobachtung als einmaligen Schnappschuss einer Zeitreihe im Gleichgewichtszustand behandeln. Diese Dissertation liefert Grundlagenforschung für den neuen Ansatz im Bereich der grafischen Modelle und gliedert sich in drei Teile.

Die Kovarianzmatrix für die Daten ist durch Lösen der stetigen Lyapunov-Gleichung gegeben, die durch die Driftmatrix des dynamischen Prozesses parametrisiert wird. In diesem Zusammenhang postulieren verschiedene statistische Modelle unterschiedliche Sparsity-Muster in der Driftmatrix, und eine entscheidende Frage ist, ob eine gegebene Sparsity-Annahme es erlaubt, die Einträge in der Driftmatrix eindeutig aus der Kovarianzmatrix der Daten wiederherzustellen. Wir untersuchen dieses Identifizierbarkeitsproblem, indem wir Sparsity-Muster durch gerichtete Graphen darstellen. Unser Hauptergebnis beweist, dass die Driftmatrix genau dann global identifizierbar ist, wenn der Graph für das Sparsity-Muster keine 2-Zyklen enthält. Darüber hinaus formulieren wir eine notwendige Voraussetzung für die generische Identifizierbarkeit und präsentieren eine rechnerische Klassifizierung kleiner Graphen mit bis zu 5 Knoten.

Jedes grafische stetige Lyapunov-Modell nimmt an, dass die Driftmatrix dünnbesetzt ist und das Sparsity-Muster durch einen gerichteten Graphen beschrieben werden kann. Ein natürlicher Ansatz zur Modellauswahl ist die Verwendung einer  $\ell_1$ -Regularisierungstechnik, die auf der Grundlage einer gegebenen Kovarianzmatrix der Stichprobe versucht, eine dünn besetzte Näherungslösung für die Lyapunov-Gleichung zu finden. Wir untersuchen die Modellauswahleigenschaften der resultierenden Lasso-Technik, um ein Konsistenzresultat herzuleiten. Unsere detaillierte Analyse zeigt, dass die damit verbundene Irrepresentabilitätsbedingung überraschend schwer zu erfüllen ist. Während dies möglicherweise die asymptotische Konsistenz bei der Modellauswahl verhindert, zeigen unsere numerischen Experimente, dass der Lasso-Ansatz selbst dann in der Lage ist, die relevante Struktur der Driftmatrix wiederherzustellen, wenn die theoretischen Anforderungen an die Konsistenz nicht erfüllt sind. Weiterhin ist er robust gegenüber milder Misspezifikation des Modells. Die Analyse endet mit der Anwendung des Lasso-Ansatzes in Kombination mit dem (erweiterten) Bayes'schen Informationskriterium auf reale Daten. Trotz der Einfachheit des Ansatzes ist die Methode in der Lage, viele wichtige Verbindungen eines unter Wissenschaftlern akzeptierten Protein-Signalnetzwerk korrekt zu schätzen.

Fortschritte in der gemischten ganzzahligen quadratischen Optimierung (MIQP) ermöglichten, dass Methoden mit einer  $\ell_0$ -Regularisierung, wie die beste Teilmengenauswahl, mit Methoden mit  $\ell_1$ -Regularisierung konkurrieren können. Wir untersuchen gründlich die Stärken und Schwächen der besten Teilmengenauswahl für Lyapunov-Modelle (BSSLM). Zunächst stellen wir Beispiele bereit, die zeigen, wie

## Zusammenfassung

$\ell_1$ -regularisierte Methoden dazu neigen, unerwünschte Symmetrie in den Schätzern zu erzeugen, die durch die Verwendung des BSSLM behoben werden kann. Wir stellen den Zusammenhang zur besten Teilmengenauswahl für Regressionsprobleme her und zeigen, wie die Probleme so formuliert werden können, dass sie für MIQP-Solver lösbar sind. Bei der Analyse des Zeitaufwands empfehlen wir, sich mit kleineren Problemgrößen bis zu  $25 \times 25$  zufrieden zu geben. In Szenarien, in denen die Nicht-Null-Einträge klar von Null unterscheidbar sind, schneidet die BSSLM viel besser ab als die  $\ell_1$ -Konkurrenten. Darüber hinaus zeigen wir, dass die BSSLM in der Lage ist, die Drift und eine diagonale Volatilitätsmatrix gemeinsam zu schätzen. Insbesondere hinsichtlich der Qualität der Schätzung der Driftmatrix zeigt die Methode im Vergleich zu den  $\ell_1$ -Konkurrenten in unserem Simulationssetting die beste Leistung. Abschließend stellen wir eine mögliche Anwendung vor, indem wir ein Protein-Signalnetzwerk ausschließlich anhand von Beobachtungsdaten abschätzen. Dort beziehen wir die Informationen zur Schätzung der Diagonale von  $C$  ein, indem wir die Knoten einfärben.

# Acknowledgement

I would like to thank everyone who supported me in the process of completing my doctorate.

I would especially like to thank my supervisor, Mathias Drton. Motivated by his research in the field of graphical modeling, the idea for the doctoral project emerged. With his experience, he was an excellent advisor, and we had many good conversations that helped me a lot. I also want to thank the other colleagues of the chair of mathematical statistics. Both organizational matters regarding the doctorate and the conversations with researchers who work on similar topics were very helpful for me. In particular, I would like to thank David Strieder, with whom I shared the office. He helped me with feedback and advice and was always available for a quick chat if necessary.

A special thanks also goes to the mentor of my doctoral project, Dr. Hans-Peter Reck, who always supported me with advice.

I would also like to thank the Hanns-Seidel Foundation for supporting my doctoral project. The thanks, of course, refer to the financial support and the non-material support of the foundation. I had the opportunity to attend many seminars, which broadened my horizons in addition to the very specialized doctoral project. They also led to exciting encounters with doctoral students from other disciplines.

Finally, I want to thank my family and friends for their moral support during this project.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 883818).





# Contents

<b>Abstract</b>	<b>iii</b>
<b>Zusammenfassung</b>	<b>v</b>
<b>Acknowledgement</b>	<b>vii</b>
<b>Notation and Acronyms</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Structure of the Thesis and Contributions . . . . .	3
1.3 Graphical Continuous Lyapunov Models . . . . .	4
1.4 Vectorization of the Lyapunov Equation . . . . .	6
<b>2 Parameter Identifiability</b>	<b>9</b>
2.1 Organization of the Chapter . . . . .	9
2.2 Motivation . . . . .	10
2.3 Notions of Identifiability . . . . .	12
2.4 Rank Conditions . . . . .	14
2.5 Directed Acyclic Graphs . . . . .	16
2.6 Sums of Squares Decompositions and Finer Rank Conditions . . . . .	18
2.7 Simple Cyclic Graphs . . . . .	22
2.8 Non-Simple Graphs . . . . .	24
2.9 Volatility Matrix: Diagonal vs. Non-Diagonal PD Matrix . . . . .	28
2.10 Spectral Description, Kernel and Factorization . . . . .	29
2.11 Outlook: Identifiability for Partially Unknown Volatility Matrices . . . . .	30
2.12 Summary of the Chapter . . . . .	32
<b>3 Direct Lyapunov Lasso</b>	<b>33</b>
3.1 Introduction . . . . .	33
3.1.1 The Role of Parameter Identifiability for Estimation . . . . .	35
3.1.2 Support Recovery for the Direct Lyapunov Lasso - Motivation . . . . .	35
3.1.3 Organization of the Chapter . . . . .	37
3.1.4 Notation - Chapter 3 and Chapter 4 . . . . .	37
3.2 Gram Matrix of the Direct Lyapunov Lasso . . . . .	38
3.3 Consistent Support Recovery with the Direct Lyapunov Lasso . . . . .	39
3.4 Probabilistic Analysis . . . . .	44
3.5 Irrepresentability Condition . . . . .	48
3.5.1 Theoretical Analysis of the (Sufficient) Irrepresentability Condition . . . . .	48
3.5.2 Necessity of the Weak Irrepresentability Condition . . . . .	52

Contents

3.5.3	Simulation Studies: Fulfillment of the Irrepresentability Condition and the Weak Irrepresentability Condition . . . . .	54
3.5.4	Simulation Studies: Impact of the Weak Irrepresentability Condition . . . . .	60
3.6	Simulation Studies . . . . .	63
3.7	Real World Data . . . . .	65
3.7.1	Sachs Dataset . . . . .	65
3.7.2	Direct Lyapunov Lasso with the (Extended) Bayesian Information Criterion . . . . .	67
3.7.3	Standardization of the Sachs Dataset . . . . .	68
3.7.4	Estimation Results for the Sachs Dataset with the Direct Lyapunov Lasso and (Extended) BIC . . . . .	69
3.8	Summary of the Chapter . . . . .	72
<b>4</b>	<b>Beyond the Lasso - Best Subset Selection with Mixed Integer Programming</b>	<b>73</b>
4.1	Difficulties of Existing Methods . . . . .	73
4.1.1	Revisiting Existing Methods . . . . .	73
4.1.2	Main Examples . . . . .	75
4.1.3	KKT-Conditions for the Loglik-Method . . . . .	79
4.2	Best Subset Selection with MIQP for Lyapunov Models . . . . .	81
4.2.1	Solving the MIQP with Gurobi . . . . .	83
4.2.2	Warm Starts of the BSSLM . . . . .	85
4.2.3	BSSLM - Time Consumption and Comparison of Initializations . . . . .	86
4.2.4	Comparing the BSSLM with the direct Lyapunov lasso and the Loglik-0.01 . . . . .	91
4.3	BSSLM and the Extended BIC . . . . .	96
4.3.1	Application - Sachs Dataset with the BSSLM and the Extended BIC . . . . .	99
4.4	Diagonally Unknown Volatility Matrix . . . . .	100
4.4.1	Simulations - Diagonally Unknown Volatility Matrix . . . . .	102
4.4.2	Application - Sachs Dataset with the BSSLM for Diagonally Unknown Volatility Matrix and EBIC . . . . .	107
4.5	Summary of the Chapter . . . . .	108
<b>5</b>	<b>Conclusion of the Thesis</b>	<b>109</b>
<b>A</b>	<b>Failure of Entry-Wise Concentration Inequalities</b>	<b>111</b>
<b>B</b>	<b>Additional Simulations</b>	<b>115</b>
B.1	Direct Lyapunov Lasso with BIC and EBIC . . . . .	115
B.2	Additional Simulations - Comparing the BSSLM w.r.t. Initializations . . . . .	118
B.3	BSSLM - Time Consumption and Comparison of Initializations - Additional Information . . . . .	118
<b>C</b>	<b>Auxiliary Results</b>	<b>118</b>
	<b>Bibliography</b>	<b>122</b>

## Notation and Acronyms

$A(\Sigma)$	Design matrix of Lyapunov Models when $C$ is assumed to be known.
$B(\Sigma)$	Design matrix of Lyapunov models when $C$ is assumed to be unknown.
$C$	Volatility Matrix of the Ornstein-Uhlenbeck-Process.
$K_p$	The $p \times p$ commutation matrix.
$M$	Drift matrix of the Ornstein-Uhlenbeck-Process.
$\Gamma$	Hessian matrix of Lyapunov Models.
$\hat{x}$ or $\hat{X}$	An estimate of the vector $x$ or of the matrix $X$ .
$\mathcal{M}_{G,C}$	GCLM defined by the directed graph $G$ and by a known volatility matrix $C$ .
$\mathcal{M}_G$	GCLM defined by the directed graph $G$ with unknown volatility matrix $C$ .
$\text{PD}_p$	Positive definite $p \times p$ matrices.
$\text{Stab}(E)$	Stable matrices supported over a graph with edgeset $E$ .
$\det(A)$	Determinant of the matrix $A$ .
$\ker(A)$	Kernel of the matrix $A$ .
$\text{tr}(A)$	Trace of the matrix $A$ .
$\text{vech}(A)$	Half-vectorization of the matrix $A$ .
$\text{vec}(A)$	Vectorization of the matrix $A$ .
$\ A\ _b$	$\ A\ _b = (\sum_{i=1}^p \sum_{j=1}^n  a_{ij} ^b)^{1/b}$ .
$\ A\ _b$	$\ A\ _b = \max\{\ Ax\ _b : \ x\ _b = 1\}$ .
$\otimes$	Kronecker Product.
$x^*$ or $X^*$	The data generating parameter vector or matrix when assessing an estimation method.
acc	Accuracy.
auc/auROC	Area under the roc curve.
aupr	Area under the precision curve.
BSSLM	Best subset selection for Lyapunov models.
DAG	Directed acyclic graph.
fdr	False Discovery Rate.
fpr	False Positive Rate.
GCLM	Graphical continuous Lyapunov model.
pr	Precision.
tpr	True Positive Rate.



# Chapter 1

## Introduction

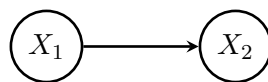
The relevance and also the availability of data is greater than ever before. In its early days, statistics was mainly descriptive and a tool for research in other fields such as medicine, psychology, or economics. Standard methods are still often used today to evaluate previously established hypotheses. At the same time, however, technological advances and developments in the field of computer science also increased the ambitions of statisticians. One of these ambitions includes understanding or even reconstructing complex networks with many interacting units in a data-driven way. For this, simultaneous observations of the units must be available, i.e., a multivariate data set.

### 1.1 Motivation

Graphical models are powerful tools for analyzing complex dependencies in multivariate observations. In particular, directed graphical models allow one to capture and explore dependencies induced by cause-effect relations [Pearl, 2009, Spirtes et al., 2000, Peters et al., 2017]. The connection to causality is made by hypothesizing that each variable is a function of parent variables and independent noise. This approach is also known as structural causal modeling or structural equation modeling. The simplest example is the equation

$$X_2 = f(X_1) + \epsilon,$$

where  $X_1, X_2$  are random variables,  $f$  is a (possibly linear) function and  $\epsilon$  is stochastic noise. This is visualized by the graph in Figure 1.1.



**Figure 1.1:**  $X_2$  is a cause of  $X_1$ .

For directed acyclic graphs (DAGs), the resulting models have simple interpretation and statistically favorable density factorization properties that facilitate large-scale analyses [Maathuis et al., 2019]. Difficulties arise when introducing cycles into the framework. Bongers et al. [2021] list several problems when allowing for cycles. For instance, they do not always have a unique solution or induce a unique observational distribution. Nevertheless, the mentioned work extends the acyclic framework to the simple cyclic graphs where the desirable properties hold under solvability assumptions.

Other problems with cycles are that they prevent density factorizations, making it more challenging to solve tasks such as computation of maximum likelihood estimates [Drton et al., 2019] or model selection [e.g., Richardson, 1996, Amendola et al., 2020]. Importantly, the interpretation of the models also becomes more involved and typically appeals to dynamic processes in a post-hoc way. For example, Fisher [1970] provided an interpretation based on data that are time averages. Alternative interpretations in terms of differential equations were suggested by Mooij et al. [2013] and Bongers and Mooij [2018].

Therefore, one might search for alternative modeling approaches that allow for cyclic models in a more organic way. Recently, Fitch [2019] proposed graphical models arising from a dynamical systems perspective. In independent work by Varando and Hansen [2020], the modeling setup of the previous work is refined, and they provide an efficient algorithm for structure learning and establish theoretical results on marginalization.

The novel idea is to start with a temporal process in equilibrium. That is, an i.i.d. sample  $X_1, \dots, X_n \in \mathbb{R}^p$  is assumed to arise from multivariate Ornstein-Uhlenbeck processes (i.e., multivariate continuous-time autoregressive processes) with  $X_i$  representing a single cross-sectional observation of the  $i$ -th process in equilibrium. Under this assumption,  $X_i$  is a multivariate normal random vector with a covariance matrix given by a (continuous) Lyapunov equation. Note that this framework considers data to be drawn from multiple time series in equilibrium and not from a single process Ornstein-Uhlenbeck process, as in related work of Gaïffas and Matulewicz [2019] or Ciolek et al. [2020].

The  $p$ -dimensional Ornstein-Uhlenbeck process is the solution to the stochastic differential equation

$$dX(t) = M(X(t) - a) dt + D dW(t), \quad (1.1)$$

where  $W(t)$  is a Wiener process and  $a \in \mathbb{R}^p$  and  $M, D \in \mathbb{R}^{p \times p}$  are non-singular parameter matrices. The *drift matrix*  $M$  is the key object of interest in the work of Fitch [2019] and Varando and Hansen [2020] as it determines the relations between the coordinates of the Ornstein-Uhlenbeck process  $X(t)$ ; see also Mogensen et al. [2018]. Provided  $M$  is stable (i.e., all eigenvalues have a strictly negative real part),  $X(t)$  admits an equilibrium distribution that is multivariate normal with a positive definite covariance matrix  $\Sigma$  determined by the continuous Lyapunov equation

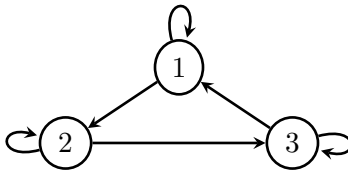
$$M\Sigma + \Sigma M^\top + C = 0, \quad (1.2)$$

where  $C = DD^\top$ . As  $M$  captures relationships among variables, it is natural to represent the connections by a (directed) graph.

For instance, the drift matrix

$$M = \begin{pmatrix} m_{11} & 0 & m_{13} \\ m_{21} & m_{22} & 0 \\ 0 & m_{32} & m_{33} \end{pmatrix} \quad (1.3)$$

translates to the graph presented in Figure 1.2.



**Figure 1.2:** The directed 3-cycle.

Assuming  $C$  to be known, every graph (i.e. sparsity pattern of  $M$ ) induces a statistical model by determining a set of covariance matrices (see Definition 1.3.2). This is also possible without assuming  $C$  to be known (see Definition 1.3.5). However, most parts of the work focus on the first setting.

## 1.2 Structure of the Thesis and Contributions

While the existing theory for structural equation models is huge, the basic questions of the “graphical model program” are yet to be asked for Graphical Continuous Lyapunov Models (GCLM). Two relevant topics are:

1. Parameter Identifiability
2. Model Selection

Topic 1) is discussed in Chapter 2. This is a fundamental theoretical question and directly proves its importance in Chapter 3 where a consistency result of a model selection method is derived. It poses the question if it is possible to uniquely recover the numerical values of the entries in the drift matrix from the true covariance matrix when fixing the nonzero pattern of  $M$  and assuming  $C$  to be known. Fixing the nonzero pattern of  $M$  coincides with considering a specific directed graph.

Topic 2) is discussed in Chapter 3 and Chapter 4. The objective is to obtain an estimate for the drift matrix or a joint estimate where the volatility matrix is also estimated.

In Chapter 3 we consider a lasso-type ( $\ell_1$ -penalized) approach that was initially proposed by Fitch [2019]. We use the convexity of the optimization problem to derive a probabilistic guarantee for support recovery. The crucial condition is an irrepresentability condition that is discussed in great detail. A potential application of Lyapunov models to real-world data is presented using a biological dataset.

In Chapter 4, we consider a variant of the best subset selection and apply it in the context of GCLMs. The idea originates from the work by Bertsimas et al. [2016] in regression contexts. The motivation is to directly control the number of active and inactive variables when seeking sparse estimates. The  $\ell_1$ -penalized method discussed in Chapter 3 is only a convenient surrogate problem and does not directly control the number of active variables. The best subset selection serves the desired purpose at the cost of a non-convex and computationally expensive problem. We rigorously analyze the strengths and weaknesses of this approach and demonstrate the superiority over the  $\ell_1$ -penalized methods in certain settings. Furthermore, we extend both the method

## Chapter 1 Introduction

discussed in Chapter 3 and the one of this chapter to jointly estimate the drift matrix and the volatility matrix.

The thesis is based on, and parts of it have been quoted verbatim from the following research articles:

**Chapter 2** is based on the publication by Dettling et al. [2023]:

P. Dettling, R. Homs, C. Améndola, M. Drton, and N. R. Hansen. Identifiability in continuous Lyapunov models. *SIAM J. Matrix Anal. Appl.*, 44(4):1799–1821, 2023.

Permission for use in the doctoral thesis granted by Managing Director Kelly Thomas (SIAM), 30.05.2024.

**Chapter 3** is based on the publication by Dettling et al. [2024]:

P. Dettling, M. Drton, and M. Kolar. On the lasso for graphical continuous Lyapunov models. In *Proceedings of the Third Conference on Causal Learning and Reasoning*, pages 514–550. PMLR, 2024.

**Chapter 4** is based on unpublished work by Dettling and Drton [2024]:

P. Dettling and M. Drton. On the best subset selection for graphical continuous Lyapunov models, 2024.

The content in **Chapter 1** is mainly an introduction and does intersect with all three papers this work is based on. Furthermore, a work on the related structural equation models by Améndola et al. [2020] was co-authored by the author of this thesis:

C. Améndola, P. Dettling, M. Drton, F. Onori, and J. Wu. Structure learning for cyclic linear causal models. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 999–1008. PMLR, 2020.

However, the paper is not part of this thesis.

## 1.3 Graphical Continuous Lyapunov Models

Before discussing the three topics above, we introduce graphical continuous Lyapunov models and make some preliminary observations. To formalize graphical continuous Lyapunov models, we need to introduce a notion of directed graphs.

**Definition 1.3.1.** A directed graph  $G$  on  $p$  nodes is defined by a pair  $(V, E)$  with  $V = [p] = \{1, \dots, p\}$  being the set of nodes and  $E = \{i \rightarrow j : i, j \in V\}$  the set of edges.

The graph displayed in Figure 1.2 is then given by  $G = (V, E)$  with  $V = \{1, 2, 3\}$  and  $E = \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, 1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1\}$ .

We consider an i.i.d. sample  $X_1, \dots, X_n \in \mathbb{R}^p$  taken from multivariate Ornstein-Uhlenbeck processes (1.1) in equilibrium. As the drift matrix  $M$  captures relationships among the coordinates of the Ornstein-Uhlenbeck process, we represent its non-zero pattern by a directed graph following the convention that

$$m_{ij} = 0 \Rightarrow j \rightarrow i \notin E.$$

For illustration, compare (1.3) with Figure (1.2).



### 1.3 Graphical Continuous Lyapunov Models

We assume  $C$  to be a known positive definite matrix. This assumption needs further explanation that we provide at a later point in this work. If all the eigenvalues of  $M$  have a strictly negative real-part, the equilibrium distribution of the Ornstein-Uhlenbeck process is Gaussian. Then, the unique covariance matrix  $\Sigma$  of the Gaussian distribution  $N(a, \Sigma)$  is determined by the continuous Lyapunov equation (1.2), see Lyapunov's theorem [Horn and Johnson, 1991, Theorem 2.2.1].

Without loss of generality we assume that the observations are centered, i.e. that  $a = 0$ . A graph, reflected by a fixed non-zero pattern of the drift matrix, together with a positive definite matrix  $C$ , defines a model.

**Definition 1.3.2.** *Let  $G = (V, E)$  be a directed graph with vertex set  $V = [p]$  and an edge set  $E$  that includes all self-loops  $i \rightarrow i$ ,  $i \in [p]$ . Given a choice of  $C \in \text{PD}_p$ , the graphical continuous Lyapunov model of  $G$  is the set of covariance matrices*

$$\mathcal{M}_{G,C} = \{\Sigma \in \text{PD}_p : M\Sigma + \Sigma M^\top = -C \text{ with } M \in \mathbb{R}^E\},$$

where we write  $\mathbb{R}^E$  for the space of matrices  $M = (m_{ij}) \in \mathbb{R}^{p \times p}$  with  $m_{ji} = 0$  whenever  $i \rightarrow j \notin E$ .

One might ask why we do not force  $M$  to be stable in Definition 1.3.2.

**Remark 1.3.3.** *Let  $\text{Stab}(E) \subseteq \mathbb{R}^E$  be the subset of stable matrices, which is always non-empty and open. When  $C$  is positive definite, the Lyapunov equation from (1.2) has a positive definite solution  $\Sigma$  if and only if  $M$  is stable [Bhaya et al., 2003, Theorem 1.1]. Hence, the definition of the model  $\mathcal{M}_{G,C}$  remains unchanged if we replace the requirement  $M \in \mathbb{R}^E$  by  $M \in \text{Stab}(E)$ .*

There is one subtlety of the Lyapunov equation that we want to mention here. It is central to correctly interpret the results throughout this work.

**Remark 1.3.4.** *If a matrix  $\Sigma$  solves the Lyapunov equation (1.2) for a pair  $(M, C)$  then  $\Sigma$  also solves the equation given by  $(\gamma M, \gamma C)$  for any  $\gamma \in \mathbb{R} \setminus \{0\}$*

$$\gamma M\Sigma + \Sigma\gamma M^\top + \gamma C = 0 \iff M\Sigma + \Sigma M^\top + C = 0. \quad (1.4)$$

Even though this might seem a bit vague at this point, a lot of the times obtaining results for a model  $\mathcal{M}_{G,C}$  implies that the results also hold for the model  $\mathcal{M}_{G,\gamma C}$  with  $\gamma \in \mathbb{R}^+$ .

In Chapter 4.4, we consider estimating  $M$  and the diagonal of  $C$  jointly. In this work, we advocate for assuming that the matrix  $C$  is diagonally unknown with  $c_{11} = 1$ . Setting  $c_{11} = 1$  takes into account the scaling invariance of the Lyapunov equation and admits for unique solutions.

**Definition 1.3.5.** *Let  $G = (V, E)$  be a directed graph with vertex set  $V = [p]$  and an edge set  $E$  that includes all self-loops  $i \rightarrow i$ ,  $i \in [p]$ . The graphical continuous Lyapunov model of  $G$  with  $C$  diagonally unknown is the set of covariance matrices*

$$\mathcal{M}_G = \{\Sigma \in \text{PD}_p : M\Sigma + \Sigma M^\top = -C \text{ for some } M \in \mathbb{R}^E, \quad (1.5)$$

$$C \in \text{PD}_p \text{ diag and } c_{11} = 1\}.$$

Estimating off-diagonal entries for  $C$  is also possible and they are represented by bidirected edges by Varando and Hansen [2020]. However, the authors mention themselves that a lot of desirable properties are only available when assuming  $C$  diagonal. For instance, assuming  $C$  diagonal, the local independence graph has the global Markov property [Mogensen et al., 2018]. With the limited theory available, assuming  $C$  to be diagonal is also a natural starting point for analyzing model geometry and extending parameter identifiability.


We end this section with an illustration of the two variants of graphical continuous Lyapunov models (GCLM).

**Example 1.3.6.** *First, we consider the models in Definition 1.3.2 where  $C$  is assumed to be known. The directed 3-cycle  $G$  with vertex set  $V = \{1, 2, 3\}$  and edge set  $E = \{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1, 2 \rightarrow 3, 3 \rightarrow 1\}$ , which is displayed on the left of Figure 1.3, encodes drift matrices of the form*

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

*Considering the models in Definition 1.3.5 where  $C$  is assumed to be diagonally unknown, the directed 3-cycle on the left in Figure 1.3 is expanded by the coloring of nodes on the right in Figure 1.3 and encodes drift matrices of the form of  $M$  and volatility matrices*

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{22} & 0 \\ 0 & 0 & c_{33} \end{pmatrix}$$

*with  $1 > c_{22} > c_{33}$ . The coloring allows for a relative comparison of the diagonal entries in  $C$ . The smallest entry is turquoise and the largest one is pink. The color progression is from small to large is: .*



**Figure 1.3:** **Left:** Directed 3-cycle. **Right:** Directed 3-cycle with information on  $C$ .

## 1.4 Vectorization of the Lyapunov Equation

The Lyapunov equation (1.2) is a matrix equation that can be vectorized in which case it takes the form of a classical linear equation system  $Ax = b$ . This comes in handy both for investigating identifiability of Lyapunov models and for model selection. In particular for model selection, the similarity to regression problems allows to modify known variable selection techniques and apply them in the context of GCLMs.

## 1.4 Vectorization of the Lyapunov Equation

We present all the notation that is needed to derive the vectorized Lyapunov equation having the form

$$A(\Sigma)\text{vec}(M) = -\text{vec}(C). \quad (1.6)$$

**Definition 1.4.1.** Let  $A$  be a  $p \times p$  matrix. The *vec-operator*  $\text{vec}(\cdot)$  transforms

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{pmatrix}$$

into a vector of length  $p^2$  by stacking the columns of  $A$  below one another, i.e.

$$\text{vec}(A) = (a_{11}, a_{21}, \dots, a_{p1}, a_{12}, a_{22}, \dots, a_{p2}, \dots, a_{1p}, a_{2p}, \dots, a_{pp})^\top.$$

When vectorizing the Lyapunov equation (1.2) we apply the *vec-operator* to a product of two matrices which can be rewritten using the Kronecker product, see the work by Horn and Johnson [1991] for instance.

**Definition 1.4.2.** Let  $A$  and  $B$  be  $p \times p$  matrices. The *Kronecker product* of  $A$  and  $B$  is a  $p^2 \times p^2$  matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1p}B \\ a_{21}B & a_{22}B & \dots & a_{2p}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}B & a_{p2}B & \dots & a_{pp}B \end{pmatrix}$$

where

$$a_{ij}B = \begin{pmatrix} a_{ij}b_{11} & a_{ij}b_{12} & \dots & a_{ij}b_{1p} \\ a_{ij}b_{21} & a_{ij}b_{22} & \dots & a_{ij}b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{ij}b_{p1} & a_{ij}b_{p2} & \dots & a_{ij}b_{pp} \end{pmatrix}.$$

Lastly, we need to transform  $\text{vec}(M^\top)$  to  $\text{vec}(M)$  to bring the summands of the Lyapunov equation together which is done using the commutation matrix, see [Magnus and Neudecker, 1999, p. 54].

**Definition 1.4.3.** The  $p \times p$  commutation matrix is given by

$$K_p = \sum_{i=1}^p \sum_{j=1}^p (e_{p,i}e_{p,j}^\top) \otimes (e_{p,j}e_{p,i}^\top),$$

It transforms  $\text{vec}(A)$  for  $A \in \mathbb{R}^{p \times p}$  to  $\text{vec}(A^\top)$ , i.e.

$$K_p \text{vec}(A) = \text{vec}(A^\top).$$

The vector  $e_{p,i}$  denotes the  $i$ -th canonical vector of dimension  $p$ .

Chapter 1 Introduction

Having introduced the above notion, we can finally vectorize the Lyapunov equation.

**Lemma 1.4.4.** *Vectorizing the Lyapunov equation (1.2), we obtain the system*

$$((\Sigma \otimes I_p) + (I_p \otimes \Sigma)K_p)\text{vec}(M) = -\text{vec}(C), \quad (1.7)$$

where  $K_p$  is the  $p \times p$  commutation matrix.

**Proof.** It holds that

$$\begin{aligned} \text{vec}(M\Sigma + \Sigma M^\top) &= \text{vec}(M\Sigma) + \text{vec}(\Sigma M^\top) \\ &= (\Sigma^\top \otimes I_p)\text{vec}(M) + (I_p \otimes \Sigma)\text{vec}(M^\top) = ((\Sigma \otimes I_p) + (I_p \otimes \Sigma)K_p)\text{vec}(M). \end{aligned}$$

□

Defining

$$A(\Sigma) := (\Sigma \otimes I_p) + (I_p \otimes \Sigma)K_p$$

we obtain the vectorized Lyapunov equation (1.6). In general, we use this definition throughout the thesis. The formulation as sum of Kronecker products involving the covariance matrix is especially useful for the probabilistic analysis in Chapter 3 and the computational study in Chapter 4. However, the analysis in Chapter 2 focuses on the rank of submatrices of  $A(\Sigma)$ . The matrix  $A(\Sigma)$ , as defined above, has redundant rows that are hindering when analyzing the rank. Therefore, we only select the rows of

$$(\Sigma \otimes I_p) + (I_p \otimes \Sigma)K_p$$

indexed by pairs  $(k, l)$  with  $k \leq l$  to obtain the matrix  $A(\Sigma)$  in Chapter 2. To avoid confusion, we clarify which version of  $A(\Sigma)$  is used in the individual chapters.

In this work we often use submatrices. For an index set  $S$ , we write  $A_{,S}$  for the submatrix of  $A$  that is obtained by selecting the columns indexed by  $S$ . The matrices  $A_S$  and  $A_{SS}$  are defined analogously by selection of rows or both rows and columns, respectively.

# Chapter 2

## Parameter Identifiability

This chapter is largely based on the publication Dettling et al. [2023]. However, Section 2.11 contains new material.

Parameter identifiability is an essential theoretical question for the validity of statistical models. Different statistical models postulate different sparsity patterns in the drift matrix, and it becomes a crucial problem to clarify whether a given sparsity assumption allows one to uniquely recover the drift matrix parameters from the covariance matrix of the data. We study this identifiability problem by representing sparsity patterns by directed graphs. Primarily, the chapter focuses on assuming the volatility matrix  $C$  to be known, which is the case for the models in Definition 1.3.2.

### 2.1 Organization of the Chapter

In Section 2.3 we introduce the notions of generic and global identifiability and make some preliminary observations. In Section 2.4, we explain the structure of the matrix  $A(\Sigma)$  that arises from (half-)vectorization of the Lyapunov equation. We also highlight how the rank of a submatrix of  $A(\Sigma)$  determines generic and global identifiability of a model. Exploiting block structure in the relevant submatrix of  $A(\Sigma)$ , we prove global identifiability for all directed acyclic graphs (DAGs) in Section 2.5. Our proof also yields that the models given by DAGs are closed algebraic subsets of  $\text{PD}_p$ , and that the models associated to complete DAGs are equal to  $\text{PD}_p$  (Corollary 2.5.4). In Section 2.6, we turn to cyclic graphs for which the relevant matrices no longer exhibit block structure. We demonstrate that for small graphs the approach studying factorizations of determinants can still be implemented using sum of squares methods to certify that the relevant polynomials are positive on  $\text{PD}_p$ . When feasible, such computations prove again that identifiable models are closed subsets of  $\text{PD}_p$ . In Section 2.7 we present our main result (Theorem 2.7.1), which proves that global model identifiability holds if the underlying graph is simple (i.e., does not contain any 2-cycle). If  $C$  is diagonal—the case of primary practical interest, then the requirement that the graph be simple is also necessary for global identifiability. Moreover, we are able to show that for all  $C \in \text{PD}_p$ , all simple graphs yield models  $\mathcal{M}_{G,C}$  that are closed algebraic subsets of  $\text{PD}_p$ . We discuss further the diagonal hypothesis on  $C$  in Section 2.9. In Section 2.8, we turn to the weaker notion of generic identifiability, for which we develop a necessary criterion and computationally classify all non-simple graphs with up to 5 nodes. The results in Section 2.10 are additional information that

comes in handy throughout this chapter. In Section 2.11 we provide a small outlook into how the results in this chapter could be used to investigate the case  $C$  unknown.

## 2.2 Motivation

The Lyapunov equation from (1.2) is a symmetric matrix equation providing  $p(p+1)/2$  constraints. In contrast, the drift matrix  $M$  is a  $p \times p$  matrix that does not need to be symmetric. Hence, without any assumptions on its structure,  $M$  is never uniquely determined by the covariance matrix  $\Sigma$  of the observations. For graphical Lyapunov models, this leads to a key identifiability question: For which sparsity patterns can the drift matrix  $M$  be recovered from the positive definite covariance matrix  $\Sigma$ ? Our treatment of this question will assume that the volatility matrix  $C$  is a known positive definite matrix. While some of our results hold for all positive definite  $C$ , others require the assumption that  $C$  is diagonal. This is a sensible assumption as it corresponds to the setting of uncorrelated noise. A special case is the assumption  $C = 2I_p$  that covers the natural setting of homoscedastic noise.

The identifiability question we pose asks if a covariance matrix  $\Sigma$  in the model  $\mathcal{M}_{G,C}$  may simultaneously solve the Lyapunov equation for more than one choice of a matrix  $M \in \mathbb{R}^E$ . In other words, we study the injectivity of the (rational) parametrization map

$$\begin{aligned} \phi_{G,C} : \text{Stab}(E) &\rightarrow \text{PD}_p \\ M &\mapsto \Sigma(M, C), \end{aligned} \tag{2.1}$$

where  $\Sigma(M, C)$  is the unique matrix  $\Sigma$  that solves the Lyapunov equation given by the stable matrix  $M$  and positive definite  $C$ . See (2.6) for details on this uniqueness.

**Example 2.2.1.** *By vectorization, the Lyapunov equation (1.2) is transformed into the linear equation system*

$$A(\Sigma)\text{vec}(M) = -\text{vech}(C), \tag{2.2}$$

where  $\text{vech}(C)$  is the half-vectorization of a fixed symmetric matrix  $C \in \text{PD}_p$ , and  $A(\Sigma)$  is a  $p(p+1)/2 \times p^2$  matrix depending on  $\Sigma$  whose form will be discussed in Section 2.4. In the case of  $p = 3$  variables the matrix  $A(\Sigma)$  equals

$$\begin{matrix} & \begin{matrix} 1 \rightarrow 1 & 1 \rightarrow 2 & 1 \rightarrow 3 & 2 \rightarrow 1 & 2 \rightarrow 2 & 2 \rightarrow 3 & 3 \rightarrow 1 & 3 \rightarrow 2 & 3 \rightarrow 3 \end{matrix} \\ \begin{matrix} (1, 1) \\ (1, 2) \\ (1, 3) \\ (2, 2) \\ (2, 3) \\ (3, 3) \end{matrix} & \left( \begin{array}{ccccccccc} 2\Sigma_{11} & 0 & 0 & 2\Sigma_{12} & 0 & 0 & 2\Sigma_{13} & 0 & 0 \\ \Sigma_{12} & \Sigma_{11} & 0 & \Sigma_{22} & \Sigma_{12} & 0 & \Sigma_{23} & \Sigma_{13} & 0 \\ \Sigma_{13} & 0 & \Sigma_{11} & \Sigma_{23} & 0 & \Sigma_{12} & \Sigma_{33} & 0 & \Sigma_{13} \\ 0 & 2\Sigma_{12} & 0 & 0 & 2\Sigma_{22} & 0 & 0 & 2\Sigma_{23} & 0 \\ 0 & \Sigma_{13} & \Sigma_{12} & 0 & \Sigma_{23} & \Sigma_{22} & 0 & \Sigma_{33} & \Sigma_{23} \\ 0 & 0 & 2\Sigma_{13} & 0 & 0 & 2\Sigma_{23} & 0 & 0 & 2\Sigma_{33} \end{array} \right) \end{matrix}$$

where the column index  $i \rightarrow j$  corresponds to entry  $m_{ji}$  of the drift matrix  $M = (m_{ij})$ .

Given a graph  $G$  with  $p(p+1)/2$  edges, unique solvability of (2.2) for  $M \in \mathbb{R}^E$  is equivalent to a certain maximal square submatrix of  $A(\Sigma)$  being invertible. This

submatrix is formed by all columns of  $A(\Sigma)$  corresponding to edges of the graph. Observe that two columns indexed by  $i \rightarrow j$  and  $k \rightarrow l$  have the same zero pattern whenever  $j = l$ . This motivates ordering the columns of  $A(\Sigma)_{\cdot, E}$  increasingly with

$$i \rightarrow j < k \rightarrow l \quad \text{if } j < l \text{ or } j = l, i < k. \quad (2.3)$$

Moreover, note that for simple graphs there is a natural pairing between pairs  $(i, j)$  with  $i \leq j$  and edges between  $i$  and  $j$ . In this case, we will order rows accordingly with their corresponding pair  $(i, j)$ .

Consider the 3-cycle  $G$  from Figure 1.2. Then unique solvability of (2.2) for  $M \in \mathbb{R}^E$  is equivalent to a submatrix of  $A(\Sigma)$  being invertible, namely, the submatrix

$$A(\Sigma)_{\cdot, E} = \begin{matrix} & \begin{matrix} 1 \rightarrow 1 & 3 \rightarrow 1 & 1 \rightarrow 2 & 2 \rightarrow 2 & 2 \rightarrow 3 & 3 \rightarrow 3 \end{matrix} \\ \begin{matrix} (1, 1) \\ (1, 3) \\ (1, 2) \\ (2, 2) \\ (2, 3) \\ (3, 3) \end{matrix} & \begin{pmatrix} 2\Sigma_{11} & 2\Sigma_{13} & 0 & 0 & 0 & 0 \\ \Sigma_{13} & \Sigma_{33} & 0 & 0 & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12} & \Sigma_{23} & \Sigma_{11} & \Sigma_{12} & 0 & 0 \\ 0 & 0 & 2\Sigma_{12} & 2\Sigma_{22} & 0 & 0 \\ 0 & 0 & \Sigma_{13} & \Sigma_{23} & \Sigma_{22} & \Sigma_{23} \\ 0 & 0 & 0 & 0 & 2\Sigma_{23} & 2\Sigma_{33} \end{pmatrix} \end{matrix}$$

To show invertibility of  $A(\Sigma)_{\cdot, E}$ , we may inspect its determinant, which factorizes as

$$|\det(A(\Sigma)_{\cdot, E})| = 2^3 \cdot \det(\Sigma) \cdot (\Sigma_{11}\Sigma_{22}\Sigma_{33} - \Sigma_{12}\Sigma_{13}\Sigma_{23}). \quad (2.4)$$

All displayed factors are positive when  $\Sigma$  is positive definite. Indeed,  $\det(\Sigma) > 0$  and the fact that  $\det(\Sigma_{ij, ij}) = \Sigma_{ii}\Sigma_{jj} - \Sigma_{ij}^2 > 0$  for all  $i \neq j$  implies that  $\Sigma_{11}^2\Sigma_{22}^2\Sigma_{33}^2 > \Sigma_{12}^2\Sigma_{13}^2\Sigma_{23}^2$ , which clarifies that the last factor is also positive. Alternatively, we can show this using the identity

$$\begin{aligned} & (\Sigma_{11}\Sigma_{22}\Sigma_{33})^2 - (\Sigma_{12}\Sigma_{13}\Sigma_{23})^2 = \\ & (\Sigma_{13}\Sigma_{23})^2 \det(\Sigma_{12, 12}) + \Sigma_{11}\Sigma_{22}\Sigma_{23}^2 \det(\Sigma_{13, 13}) + \Sigma_{11}^2\Sigma_{22}\Sigma_{33} \det(\Sigma_{23, 23}) > 0. \end{aligned}$$

We conclude that when  $G$  is the 3-cycle, then for all covariance matrices  $\Sigma \in \mathcal{M}_{G, C} \subseteq \text{PD}_3$  there is a unique matrix  $M \in \mathbb{R}^E$  such that  $\Sigma = \phi_{G, C}(M)$ . We will refer to this property as the 3-cycle defining a globally identifiable model. Note that our argument also shows that  $\mathcal{M}_{G, C} = \text{PD}_3$ .

This small example already reveals some of the subtleties arising when analyzing identifiability of continuous Lyapunov models. The problem can be reduced to determining whether a particular submatrix that is sparsely populated with covariances has full rank (see Lemma 2.4.3 and Lemma 2.6.4) but the resulting matrices have involved graph-dependent structures.

The choice of ordering in (2.3) is especially insightful for directed acyclic graphs. After sorting the nodes such that if  $i \rightarrow j$  then  $i \leq j$ , any DAG yields a block upper-triangular matrix, as in Example 2.5.2, from which identifiability for all associated models follows (Theorem 2.5.3). For cyclic graphs, however, the polynomials that appear while factoring determinants, as in (2.4), quickly increase in complexity, and it is not easy to determine whether they are non-zero. In our main result (Theorem 2.7.1) we thus consider alternative spectral arguments that use the stability of the drift matrix  $M$  in order to derive identifiability.

## 2.3 Notions of Identifiability

We begin by recalling the concept of fibers that is useful to define the different notions of identifiability we study in subsequent sections. Let  $C \in \text{PD}_p$ , and let  $\mathcal{M}_{G,C}$  be the graphical continuous Lyapunov model associated to a directed graph  $G = (V, E)$  with vertex set  $V = [p]$  and edge set  $E$ . Let  $\phi_{G,C}$  be the parametrization from (2.1). The *fiber* of a matrix  $M_0 \in \text{Stab}(E)$  is the set

$$\mathcal{F}_{G,C}(M_0) = \{M \in \text{Stab}(E) : \phi_{G,C}(M) = \phi_{G,C}(M_0)\}. \quad (2.5)$$

In other words, a fiber comprises all drift matrices  $M \in \mathbb{R}^E$  whose Lyapunov equation (for the fixed matrix  $C \in \text{PD}_p$ ) is solved by a given covariance matrix  $\Sigma$ .

We will consider three natural notions of identifiability.

**Definition 2.3.1.** *Let  $\mathcal{M}_{G,C}$  be the graphical continuous Lyapunov model given by a directed graph  $G = (V, E)$  with  $V = [p]$  and  $C \in \text{PD}_p$ . The model  $\mathcal{M}_{G,C}$  is*

- (i) globally identifiable if  $\mathcal{F}_{G,C}(M_0) = \{M_0\}$  for all  $M_0 \in \text{Stab}(E)$ ;
- (ii) generically identifiable if  $\mathcal{F}_{G,C}(M_0) = \{M_0\}$  for almost all  $M_0 \in \text{Stab}(E)$ , i.e., the matrices with  $\mathcal{F}_{G,C}(M_0) \neq \{M_0\}$  form a Lebesgue null set in  $\mathbb{R}^E$ ;
- (iii) non-identifiable if  $|\mathcal{F}_{G,C}(M_0)| = \infty$  for all  $M_0 \in \text{Stab}(E)$ .

**Remark 2.3.2.** *The generic properties we prove in this work are derived by showing that they hold outside a strict subset of  $\text{Stab}(E)$  that is described by polynomials in the entries of the drift matrix; see e.g. Lemma 2.4.3. Hence, in a generically identifiable model the exception set is not merely a set of Lebesgue measure zero, but also a lower-dimensional algebraic subset of  $\text{Stab}(E)$ .*

**Remark 2.3.3.** *Characterizing identifiability is also a key problem for standard directed graphical models; see Drton [2018] and Sullivant [2018, Chap. 16] for a discussion of the different notions of identifiability in this context. For standard graphical models, necessary and sufficient conditions for global identifiability have been obtained [Drton et al., 2011]. However, many models of interest are not globally identifiable, and much work has also gone into criteria for generic identifiability [Brito and Pearl, 2006, Kumor et al., 2019, Foygel et al., 2012, Drton and Weihs, 2016].*

The 3-cycle from Example 2.2.1 is an example of global identifiability. Under global identifiability, no two distinct stable matrices may define the same covariance matrix in the model given by the graph. Unfortunately, this is not always the case.

**Example 2.3.4.** *Consider the 2-cycle  $G = (V, E)$  with  $V = \{1, 2\}$  and  $E = \{1 \rightarrow 2, 2 \rightarrow 1\}$ . Then  $\phi_{G,C}$  maps the 4-dimensional parameter space  $\text{Stab}(E)$  to the 3-dimensional  $\text{PD}_2$ -cone. Hence, when computing any fiber we have to solve a linear system that is underdetermined, with 3 equations in 4 unknowns. Therefore,  $\mathcal{M}_{G,C}$  is non-identifiable, no matter the choice of  $C \in \text{PD}_2$ .*

The example just given generalizes as follows:

**Lemma 2.3.5.** *Let  $G = (V, E)$  be a directed graph with vertex set  $V = [p]$ , and let  $C \in \text{PD}_p$ . If  $|E| > \dim(\mathcal{M}_{G,C})$ , i.e., the number of free parameters in  $\text{Stab}(E)$  is*



## 2.3 Notions of Identifiability

greater than the dimension of the model, then  $\mathcal{M}_{G,C}$  is non-identifiable. In particular, all graphs with  $|E| > p(p+1)/2$  give non-identifiable models.

**Proof.** By the Hurwitz criterion, the set of sparse stable matrices  $\text{Stab}(E)$  is semialgebraic, see Horn and Johnson [1991, Theorem 2.3.3]. As its dimension is  $\dim(\text{Stab}(E)) = |E| > \dim(\mathcal{M}_{G,C})$ , it follows that the rational map  $\phi_{G,C}$  defined on  $\text{Stab}(E)$  is generically infinite-to-one; see, e.g., Barber et al. [2022, Lemma 2.5]. Apply Lemma 2.4.3 below to conclude that all fibers are infinite.  $\square$

A straightforward but very useful fact when studying global identifiability is that if a graph  $G = (V, E)$  yields a globally identifiable model then so does every subgraph  $H = (V, E')$ ,  $E' \subseteq E$ , that is obtained by removing edges of the form  $i \rightarrow j$  with  $i \neq j$ . We record this fact as:

**Proposition 2.3.6.** *Let  $\mathcal{M}_{G,C}$  be a globally identifiable model given by a directed graph  $G = (V, E)$  with  $V = [p]$  and  $C \in \text{PD}_p$ . Let  $E' \subset E$  be a subset of the edges. Then the model  $\mathcal{M}_{H,C}$  defined by the subgraph  $H = (V, E')$  is globally identifiable.*

**Proof.** It holds that  $\text{Stab}(E') \subseteq \text{Stab}(E)$ . Therefore, for every matrix  $M_0 \in \text{Stab}(E')$ , we have  $\mathcal{F}_{H,C}(M_0) \subseteq \mathcal{F}_{G,C}(M_0) = \{M_0\}$ , where the last equality is due to the assumed global identifiability of  $\mathcal{M}_{G,C}$ .  $\square$

In the case where  $C$  is diagonal, further conclusions can be made.

**Proposition 2.3.7.** *Let  $G = (V, E)$  be a directed graph with  $V = [p]$ . Let  $C \in \text{PD}_p$  be diagonal, and let  $I_p$  be the  $p \times p$  identity matrix. Then the models for  $C$  versus  $I_p$  are isomorphic, and so are their fibers:*

- (i)  $\mathcal{M}_{G,C} = C^{1/2}\mathcal{M}_{G,I_p}C^{1/2}$ , and
- (ii)  $\mathcal{F}_{G,C}(M) = \mathcal{F}_{G,I_p}(C^{1/2}MC^{-1/2})$  for all  $M \in \text{Stab}(E)$ .

In particular,  $\mathcal{M}_{G,C}$  is globally/generically identifiable if and only if  $\mathcal{M}_{G,I_p}$  is globally/generically identifiable.

**Proof.** Since  $C$  is diagonal, the similarity transformation  $\tau_1 : M \mapsto C^{-1/2}MC^{1/2}$  is an automorphism of  $\mathbb{R}^E$ , with  $\tau_1(\text{Stab}(E)) = \text{Stab}(E)$ . Define a second linear map  $\tau_2 : \Sigma \mapsto C^{-1/2}\Sigma C^{-1/2}$ , an automorphism of the space of symmetric matrices with  $\tau_2(\text{PD}_p) = \text{PD}_p$ . Now

$$\begin{aligned} M\Sigma + \Sigma M^\top + C = 0 &\iff \\ (C^{-1/2}MC^{1/2})(C^{-1/2}\Sigma C^{-1/2}) + (C^{-1/2}\Sigma C^{-1/2})(C^{-1/2}MC^{1/2})^\top + I_p &= 0. \end{aligned}$$

Thus,  $\mathcal{M}_{G,I_p} = \tau_2(\mathcal{M}_{G,C})$  and  $\mathcal{F}_{G,I_p}(M) = \mathcal{F}_{G,C}(\tau_1^{-1}(M))$ .  $\square$

In Proposition 2.3.6 only edges are removed when forming a subgraph. When  $C$  is diagonal we may strengthen the result to subgraphs in which we also remove vertices; compare Drton et al. [2011, Lemma 1] in the context of standard graphical models.

**Proposition 2.3.8.** *Let  $G = (V, E)$  be a directed graph with  $V = [p]$ , and let  $H = (V', E')$  be a subgraph with  $V' \subseteq V$  and  $E' \subseteq E$ . If the model  $\mathcal{M}_{G,C}$  is globally*

identifiable for a diagonal matrix  $C \in \text{PD}_p$ , then  $\mathcal{M}_{H,C'}$  is globally identifiable for all diagonal matrices  $C' \in \text{PD}_{p'}$ , where  $p' = |V'|$ .

**Proof.** By Proposition 2.3.6, it suffices to prove that removing an isolated vertex from  $G$  preserves global identifiability of the model for  $C$  diagonal. By Proposition 2.3.7, we may assume that  $C = I_p$  and  $C' = I_{p-1}$ , where  $p$  is an isolated node of  $G$ . Let  $M \in \text{Stab}(E)$ , and let  $M_{[p-1],[p-1]}$  be the submatrix comprising the first  $p-1$  rows and columns. Since  $p$  is isolated, the  $p$ th row and column of  $M$  is zero with the exception of the diagonal entry  $m_{pp}$ . It is not difficult to see that  $\Sigma = \phi_{G,I_p}(M)$  also has its  $p$ th row and column equal to zero except for the diagonal entry which equals  $\Sigma_{pp} = -1/(2m_{pp})$ . Hence, the entry  $m_{pp}$  is always uniquely determined by  $\Sigma$ , and we conclude that the cardinality of the fiber  $\mathcal{F}_{G,I_p}(M)$  is equal to the cardinality of  $\mathcal{F}_{H,I_{p-1}}(M_{[p-1],[p-1]})$ . Since every matrix in  $\text{Stab}(E')$  is a submatrix  $M_{[p-1],[p-1]}$  of a matrix  $M \in \text{Stab}(E)$ , the model  $\mathcal{M}_{H,I_{p-1}}$  is globally identifiable.  $\square$

Combining Proposition 2.3.8 with Example 2.3.4, we obtain that the graph of a globally identifiable model cannot contain any 2-cycles.

**Definition 2.3.9.** A directed graph  $G = (V, E)$  is simple if it is free of 2-cycles, i.e., there do not exist two distinct nodes  $i, j \in V$  such that  $i \rightarrow j \in E$  and  $j \rightarrow i \in E$ . Otherwise, we call  $G$  non-simple.

**Proposition 2.3.10.** If a directed graph  $G = (V, E)$ ,  $V = [p]$ , defines a globally identifiable model  $\mathcal{M}_{G,C}$  when  $C \in \text{PD}_p$  is diagonal, then  $G$  must be simple.

**Remark 2.3.11.** Proposition 2.3.8 and Proposition 2.3.10 may fail for non-diagonal  $C \in \text{PD}_p$ . See Section 2.9 for an example.

Unfortunately, similar subgraph arguments cannot be made for generic instead of global identifiability. Indeed, generic identifiability may be lost but also restored when removing an edge. Example 2.8.4 illustrates this phenomenon.

## 2.4 Rank Conditions

In this section, we discuss solving the Lyapunov equation (1.2) for the generally non-symmetric drift matrix  $M$  given the symmetric matrices  $\Sigma$  and  $C$ . We will proceed by vectorizing the Lyapunov equation, and we will state necessary and sufficient conditions for identifiability based on the ranks of submatrices of the coefficient matrix  $A(\Sigma)$  of the vectorized Lyapunov equation.

First, recall that when the matrices  $M$  and  $C$  are given, the continuous Lyapunov equation from (1.2) is uniquely solvable for the symmetric matrix  $\Sigma$  if and only if no two eigenvalues of  $M$  add up to zero. This well known fact can be shown by vectorizing the equation to

$$(I_p \otimes M + M \otimes I_p)\text{vec}(\Sigma) = -\text{vec}(C), \quad (2.6)$$

where  $\otimes$  is the Kronecker product and  $\text{vec}(\cdot)$  is the columnwise vectorization of a matrix; see, e.g., Bernstein [2018]. The coefficient matrix  $I_p \otimes M + M \otimes I_p$  is a Kronecker sum, and it follows that its eigenvalues are the pairwise sums of the eigenvalues of  $M$ .

If we now additionally assume that  $C$  is positive definite, then Lyapunov's theorem [Horn and Johnson, 1991, Theorem 2.2.1] yields that the Lyapunov equation from (1.2) has a unique positive definite solution  $\Sigma$  if and only if  $M$  is a stable matrix.

However, solving for  $M$  given two symmetric (and in our context positive definite) matrices  $\Sigma$  and  $C$  is a more difficult question. In general, it is not possible to have a unique solution for  $M$  due to the dimensionality problems mentioned in Lemma 2.3.5. The graphical perspective of the Lyapunov models motivates considering sparse matrices  $M$  and asking the solvability question in a new light, as we illustrated in Example 2.2.1.

**Lemma 2.4.1.** *Vectorizing the Lyapunov equation (1.2), we obtain the system*

$$((\Sigma \otimes I_p) + (I_p \otimes \Sigma)K_p)\text{vec}(M) = -\text{vec}(C), \quad (2.7)$$

where  $K_p$  is the  $p \times p$  commutation matrix.

The commutation matrix  $K_p$  is the symmetric permutation matrix that transforms the vectorization of a  $p \times p$  matrix to the vectorization of its transpose [Magnus and Neudecker, 1999, p. 54].

**Proof of Lemma 2.4.1.** It holds that

$$\begin{aligned} \text{vec}(M\Sigma + \Sigma M^\top) &= \text{vec}(M\Sigma) + \text{vec}(\Sigma M^\top) \\ &= (\Sigma^\top \otimes I_p)\text{vec}(M) + (I_p \otimes \Sigma)\text{vec}(M^\top) = ((\Sigma \otimes I_p) + (I_p \otimes \Sigma)K_p)\text{vec}(M). \end{aligned}$$

□

The Lyapunov equation (1.2) is symmetric and therefore  $p(p-1)/2$  equations of the equation system (2.7) are redundant.

**Definition 2.4.2.** *Given a  $p \times p$  symmetric matrix  $\Sigma$ , we define the  $p(p+1)/2 \times p^2$  matrix  $A(\Sigma)$  by selecting the rows of*

$$(\Sigma \otimes I_p) + (I_p \otimes \Sigma)K_p$$

*indexed by pairs  $(k, l)$  with  $k \leq l$ .*

Let  $\text{vech}(C) = (C_{kl} : k \leq l)$  be the half-vectorization of the symmetric matrix  $C$ . Then we can write the Lyapunov equation as

$$A(\Sigma)\text{vec}(M) = -\text{vech}(C).$$

As noted, we index the rows of  $A(\Sigma)$  by pairs  $(k, l)$  with  $k \leq l$ . To index the columns of  $A(\Sigma)$  we will use the potential edges  $i \rightarrow j$ , where we recall that the edge  $i \rightarrow j$  corresponds to the entry  $m_{ji}$  of the matrix  $M$ .

Example 2.2.1 displayed  $A(\Sigma)$  for the case of  $p = 3$ . In general, we have

$$A(\Sigma)_{(k,l),i \rightarrow j} = \begin{cases} 0, & \text{if } j \neq k, l; \\ \Sigma_{li}, & \text{if } j = k, k \neq l; \\ \Sigma_{ki}, & \text{if } j = l, l \neq k; \\ 2\Sigma_{ji}, & \text{if } j = k = l. \end{cases} \quad (2.8)$$

Any specific graphical continuous Lyapunov model assumes that  $M$  has non-zero entries only for pairs  $(j, i)$  for which the underlying graph contains the edge  $i \rightarrow j$ . We are thus led to select a subset of columns of the coefficient matrix  $A(\Sigma)$  when studying solvability of the Lyapunov equation. By the next lemma, generic and global identifiability of a graphical continuous Lyapunov model are equivalent to rank conditions on the relevant submatrix of  $A(\Sigma)$ .

**Lemma 2.4.3.** *Let  $G = (V, E)$  be a directed graph with  $V = [p]$ , and let  $C \in \text{PD}_p$ . Let  $A(\Sigma)_{\cdot, E}$  be the submatrix of  $A(\Sigma)$  obtained by selecting the columns indexed by the edges in  $E$ . Then the model  $\mathcal{M}_{G, C}$  is*

- (i) *globally identifiable if and only if  $A(\Sigma)_{\cdot, E}$  has full column rank  $|E|$  for all  $\Sigma \in \mathcal{M}_{G, C}$ ;*
- (ii) *generically identifiable if and only if there exists a matrix  $\Sigma \in \mathcal{M}_{G, C}$  such that  $A(\Sigma)_{\cdot, E}$  has full column rank  $|E|$ .*

*If  $\mathcal{M}_{G, C}$  is not generically identifiable, then it is non-identifiable.*

**Proof.** Let  $M_0 \in \text{Stab}(E)$ , and let  $\Sigma_0 = \phi_{G, C}(M_0)$  be the associated covariance matrix. The fiber  $\mathcal{F}_{G, C}(M_0)$  is the set of all matrices  $M \in \mathbb{R}^E$  with

$$A(\Sigma_0)_{\cdot, E} \text{vec}(M)_E = -\text{vech}(C), \quad (2.9)$$

where  $\text{vec}(M)_E$  is the subvector of  $\text{vec}(M)$  that comprises the entries indexed by  $(j, i)$  with  $i \rightarrow j \in E$ . Hence,  $\mathcal{F}_{G, C}(M_0) = \{M_0\}$  precisely when  $A(\Sigma_0)_{\cdot, E}$  has full column rank such that (2.9) has a unique solution. Claim (i) is now evident.

To prove (ii), note that  $A(\Sigma)_{\cdot, E}$  has full column rank if and only if the vector of all maximal minors of  $A(\Sigma)_{\cdot, E}$  is non-zero. By (2.6), the map  $\phi_{G, C}$  is a rational map. Consequently, the map taking  $M \in \text{Stab}(E)$  to the maximal minors of  $A(\phi_{G, C}(M))_{\cdot, E}$  is rational as well. Now a rational map is non-zero outside a measure zero set if and only if there exists a single point where it is non-zero. Consequently, the existence of  $\Sigma \in \mathcal{M}_{G, C}$  with  $A(\Sigma)_{\cdot, E}$  of full column rank implies generic identifiability of  $\mathcal{M}_{G, C}$ .

Finally, if  $\mathcal{M}_{G, C}$  is not generically identifiable then the column rank of  $A(\Sigma_0)_{\cdot, E}$  is strictly smaller than  $|E|$  for all  $\Sigma_0 = \phi_{G, C}(M_0) \in \mathcal{M}_{G, C}$ . The fiber  $\mathcal{F}_{G, C}(M_0) \subseteq \text{Stab}(E)$  is then the affine subspace of solutions to (2.9) of dimension  $\geq 1$ . Hence,  $|\mathcal{F}_{G, C}(M_0)| = \infty$  for all  $M_0 \in \text{Stab}(E)$ , and  $\mathcal{M}_{G, C}$  is non-identifiable.  $\square$

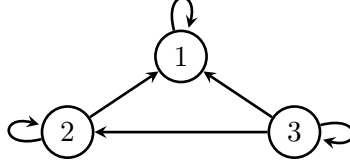
## 2.5 Directed Acyclic Graphs

In this section, we prove that all models that are given by *directed acyclic graphs* (DAGs) are globally identifiable. In our setting, a DAG is a directed graph that does not contain any directed cycles other than the always present self-loops  $i \rightarrow i$ ,  $i \in [p]$ . This case is special in that we are able to make a simple argument based on block structure in the coefficient matrix  $A(\Sigma)$ .

By Proposition 2.3.6, in order to prove global identifiability for all DAGs it suffices to treat DAGs that are complete in the sense of the following definition.

**Definition 2.5.1.** A directed simple graph  $G = (V, E)$  with  $V = [p]$  is complete if there is an edge between every pair of distinct nodes.

A simple graph that also contains all self-loops  $i \rightarrow i$ ,  $i \in [p]$ , is complete if and only if  $|E| = p(p+1)/2$ . Because vertex relabelling has no impact on identifiability, we can furthermore restrict attention to a single topological ordering. In other words, it suffices to consider the single complete DAG  $G^*$  whose edge set comprises all edges  $i \rightarrow j$  with  $i \geq j$ .



**Figure 2.1:** The complete DAG  $G^*$  on 3 nodes.

**Example 2.5.2.** Consider the case of  $p = 3$  nodes, for which the complete DAG  $G^* = (V, E^*)$  is shown in Figure 2.1. The graph encodes the drift matrix

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ 0 & m_{22} & m_{23} \\ 0 & 0 & m_{33} \end{pmatrix},$$

and the submatrix  $A(\Sigma)_{\cdot, E^*}$  is equal to

$$\begin{array}{c} \begin{matrix} 1 \rightarrow 1 & 2 \rightarrow 1 & 3 \rightarrow 1 & 2 \rightarrow 2 & 3 \rightarrow 2 & 3 \rightarrow 3 \end{matrix} \\ \begin{matrix} (1, 1) \\ (1, 2) \\ (1, 3) \\ (2, 2) \\ (2, 3) \\ (3, 3) \end{matrix} \end{array} \begin{pmatrix} 2\Sigma_{11} & 2\Sigma_{12} & 2\Sigma_{13} & 0 & 0 & 0 \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{12} & \Sigma_{13} & 0 \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & 0 & 0 & \Sigma_{13} \\ 0 & 0 & 0 & 2\Sigma_{22} & 2\Sigma_{23} & 0 \\ 0 & 0 & 0 & \Sigma_{32} & \Sigma_{33} & \Sigma_{23} \\ 0 & 0 & 0 & 0 & 0 & 2\Sigma_{33} \end{pmatrix}.$$

Up to some rows being scaled by 2, the three diagonal blocks are principal minors of the positive definite matrix  $\Sigma$ . Therefore, it holds for all  $\Sigma \in \text{PD}_3$  that

$$\begin{aligned} |\det A(\Sigma)_{\cdot, E^*}| &= \begin{vmatrix} 2\Sigma_{11} & 2\Sigma_{12} & 2\Sigma_{13} \\ \Sigma_{12} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{13} & \Sigma_{23} & \Sigma_{33} \end{vmatrix} \cdot \begin{vmatrix} 2\Sigma_{22} & 2\Sigma_{23} \\ \Sigma_{23} & \Sigma_{33} \end{vmatrix} \cdot |2\Sigma_{33}| \\ &= 2^3 \cdot \det(\Sigma) \cdot \det(\Sigma_{\{2,3\},\{2,3\}}) \cdot \Sigma_{33} > 0. \end{aligned}$$

The block structure found in Example 2.5.2 generalizes and gives the main result of this section.

**Theorem 2.5.3.** Let  $G = (V, E)$  be a DAG with  $V = [p]$ . Then the model  $\mathcal{M}_{G,C}$  is globally identifiable for every matrix  $C \in \text{PD}_p$ .

**Proof.** As noted above, it suffices to consider the complete DAG  $G^* = (V, E^*)$  whose edges are  $i \rightarrow j$  for  $i \geq j$ . Our proof then applies Lemma 2.4.3, which states that model  $\mathcal{M}_{G^*, C}$  is globally identifiable if and only if  $\det(A(\Sigma)_{\cdot, E^*}) \neq 0$  for all  $\Sigma \in \mathcal{M}_{G^*, C}$ .

In what follows, let  $\Sigma \in \text{PD}_p$ . Partition the edge set as  $E^* = E_1^* \cup E_2^* \cup \dots \cup E_p^*$ , where  $E_i^* = \{j \rightarrow i : j \geq i\}$ . Similarly, partition the row index set of  $A(\Sigma)$  into the disjoint union of the sets  $R_k = \{(k, l) : l \geq k\}$ ,  $k = 1, \dots, p$ . Inspecting (2.8), we see that the submatrix

$$A(\Sigma)_{R_k, E_i^*} = 0 \quad \text{if } k > i.$$

Hence, the matrix  $A(\Sigma)$  can be arranged in block upper-triangular form, and

$$\det(A(\Sigma)_{\cdot, E^*}) = \prod_{i=1}^p \det(A(\Sigma)_{R_i, E_i^*}).$$

Inspecting again (2.8), we find that  $A(\Sigma)_{R_i, E_i^*}$  is equal to the principal submatrix  $P(\Sigma)_{\geq i} := \Sigma_{\{i, \dots, p\}, \{i, \dots, p\}}$  but with the first row of  $P(\Sigma)_{\geq i}$  (the one indexed by  $i$ ) being multiplied by 2 in  $A(\Sigma)_{R_i, E_i^*}$ . Since all principal minors of a positive definite matrix  $\Sigma$  are positive, we obtain that

$$|\det(A(\Sigma)_{\cdot, E^*})| = 2^p \prod_{i=1}^p \det(P(\Sigma)_{\geq i}) > 0 \quad \text{for all } \Sigma \in \text{PD}_p.$$

In particular,  $A(\Sigma)_{\cdot, E^*}$  has non-vanishing determinant for all  $\Sigma \in \mathcal{M}_{G^*, C}$ .  $\square$

The proof of Theorem 2.5.3 shows that for any complete DAG  $G = (V, E)$  the matrix  $A(\Sigma)_{\cdot, E}$  is invertible for all  $\Sigma \in \text{PD}_p$ . Using this fact, the proof of the theorem reveals more information about Lyapunov models arising from DAGs.

**Corollary 2.5.4.** *Let  $G = (V, E)$  be a DAG with  $V = [p]$ . Then  $\mathcal{M}_{G, C}$  is an algebraic and thus closed subset of  $\text{PD}_p$ . If  $G$  is complete then  $\mathcal{M}_{G, C} = \text{PD}_p$ .*

**Proof.** Let  $G$  be a complete DAG. By Theorem 2.5.3, the square matrix  $A(\Sigma)_{\cdot, E}$  has full rank for all  $\Sigma \in \text{PD}_p$ . Therefore, the solution  $\text{vec}(M)$  to the vectorized Lyapunov equation (2.9) exists uniquely for all  $\Sigma \in \text{PD}_p$ . The resulting drift matrix  $M$  has the right support by construction, hence  $\mathcal{M}_{G, C} = \text{PD}_p$ .

If  $G$  is a non-complete DAG, then we may add edges to obtain a complete DAG  $\bar{G} = (V, \bar{E})$ . As  $A(\Sigma)_{\cdot, \bar{E}}$  has full column rank for all  $\Sigma \in \text{PD}_p$  the same is true for  $A(\Sigma)_{\cdot, E}$ ; recall Proposition 2.3.6. Hence, a matrix  $\Sigma \in \text{PD}_p$  is in  $\mathcal{M}_{G, C}$  if and only if  $\text{vech}(C)$  is in the column span of  $A(\Sigma)_{\cdot, E}$  if and only if the  $(|E| + 1)$ -minors of the augmented matrix  $(A(\Sigma)_{\cdot, E} \mid \text{vech}(C))$  vanish. The model  $\mathcal{M}_{G, C}$  is thus an algebraic subset: it is the set of positive definite matrices at which these minors vanish.  $\square$

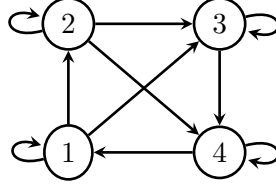
## 2.6 Sums of Squares Decompositions and Finer Rank Conditions

Directed cycles break the block-diagonal structure found for DAGs (Theorem 2.5.3) making it difficult to check rank conditions on  $A(\Sigma)$ . In this section we show that small

## 2.6 Sums of Squares Decompositions and Finer Rank Conditions

cyclic graphs can nevertheless be handled by applying sums of squares decompositions to certify positivity of subdeterminants. Moreover, we show that our rank conditions may be placed on a smaller matrix containing a basis for the kernel of  $A(\Sigma)$ .

In Example 2.2.1, we proved global identifiability for the 3-cycle by showing that the key factor  $\Sigma_{11}\Sigma_{22}\Sigma_{33} - \Sigma_{12}\Sigma_{13}\Sigma_{23}$  in the determinant of  $A(\Sigma)_{\cdot,E}$  is positive on  $\text{PD}_3$ . We were able to argue this via the positivity of  $2 \times 2$  principal minors of  $\Sigma$ . However, a direct extension of this approach to cyclic graphs with a larger number of nodes is difficult. Nevertheless, some headway can be made by exploiting the positive-definiteness of  $\Sigma$  via its Cholesky decomposition.



**Figure 2.2:** A completion of the 4-cycle.

**Example 2.6.1.** Let  $G = (V, E)$  be the completion of the 4-cycle with  $V = [4]$  and  $E = \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, 4 \rightarrow 4, 1 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 3, 2 \rightarrow 4, 3 \rightarrow 4, 4 \rightarrow 1\}$ . It is displayed in Figure 2.2. Let  $\Sigma = LL^\top$  be the Cholesky decomposition of  $\Sigma \in \text{PD}_4$  in terms of the lower-triangular matrix

$$L = \begin{pmatrix} l_{11} & 0 & 0 & 0 \\ l_{12} & l_{22} & 0 & 0 \\ l_{13} & l_{23} & l_{33} & 0 \\ l_{14} & l_{24} & l_{34} & l_{44} \end{pmatrix}$$

with  $l_{11}, l_{22}, l_{33}, l_{44} > 0$ . Then

$$|\det(A(LL^\top)_{\cdot,E})| = 16 l_{44}^2 l_{33}^2 l_{22}^4 l_{11}^6 \cdot |f(L)|,$$

where the key factor is

$$\begin{aligned} f(L) = & l_{14}^2 l_{22}^2 l_{33}^2 - l_{12} l_{14} l_{22} l_{24} l_{33}^2 + l_{12}^2 l_{24}^2 l_{33}^2 + l_{22}^2 l_{24}^2 l_{33}^2 - l_{13} l_{14} l_{22}^2 l_{33} l_{34} \\ & + l_{12} l_{14} l_{22} l_{23} l_{33} l_{34} + l_{12} l_{13} l_{22} l_{24} l_{33} l_{34} - l_{12}^2 l_{23} l_{24} l_{33} l_{34} + l_{13}^2 l_{22}^2 l_{34}^2 \\ & - 2l_{12} l_{13} l_{22} l_{23} l_{34}^2 + l_{12}^2 l_{23}^2 l_{34}^2 + l_{12}^2 l_{33}^2 l_{34}^2 + l_{22}^2 l_{33}^2 l_{34}^2 + l_{13}^2 l_{22}^2 l_{44}^2 \\ & - 2l_{12} l_{13} l_{22} l_{23} l_{44}^2 + l_{12}^2 l_{23}^2 l_{44}^2 + l_{12}^2 l_{33}^2 l_{44}^2 + l_{22}^2 l_{33}^2 l_{44}^2. \end{aligned}$$

A computer algebra system such as *Macaulay2* with the package from Cifuentes et al. [2020] quickly finds a sum of squares (SOS) decomposition for  $f$  as

$$\begin{aligned} f(L) = & \left( \frac{1}{2} l_{14} l_{22} l_{33} - \frac{1}{2} l_{12} l_{24} l_{33} - l_{13} l_{22} l_{34} + l_{12} l_{23} l_{34} \right)^2 \\ & + (-l_{13} l_{22} l_{44} + l_{12} l_{23} l_{44})^2 + (l_{12} l_{33} l_{34})^2 + (l_{12} l_{33} l_{44})^2 + (l_{22} l_{24} l_{33})^2 \\ & + (l_{22} l_{33} l_{34})^2 + (l_{22} l_{33} l_{44})^2 + \frac{3}{4} \left( l_{14} l_{22} l_{33} - \frac{1}{3} l_{12} l_{24} l_{33} \right)^2 + \frac{2}{3} (l_{12} l_{24} l_{33})^2. \end{aligned}$$

Since  $l_{22} l_{33} l_{44} > 0$ , it follows that  $f$  is strictly positive for any Cholesky factor  $L$ . Therefore,  $|\det(A(\Sigma)_{\cdot,E})| > 0$  and we conclude that  $\mathcal{M}_{G,C}$  is globally identifiable, no matter the choice of  $C \in \text{PD}_4$ .

**Remark 2.6.2.** *A polynomial being a sum of squares is a stronger requirement than the polynomial being non-zero. Therefore, we could have a non-vanishing determinant even if the considered polynomial factor failed the SOS test. However, we do not know of an example where this might be the case.*

Observe that  $\det(\Sigma) = (\det L)^2 = l_{11}^2 l_{22}^2 l_{33}^2 l_{44}^2$  appears as a factor of  $\det(A(\Sigma)_{\cdot,E})$  in all our examples so far (recall Example 2.2.1, Example 2.5.2, and Example 2.6.1). This phenomenon actually occurs for any complete simple graph (see Corollary 2.10.2) and suggests that identifiability should be encoded in a smaller matrix. Indeed, this information is carried by a specific row restriction of a matrix whose columns form a basis of the kernel of  $A(\Sigma)$ .

The kernel of  $A(\Sigma)$  is described by the following fact, straightforward to verify; see also Barnett and Storey [1967]. It parametrizes the stable matrices  $M$  that are solutions to the Lyapunov equation in terms of skew-symmetric matrices (matrices  $K$  with  $K^\top = -K$ ).

**Lemma 2.6.3.** *Consider the continuous Lyapunov equation from (1.2) for given  $\Sigma, C \in \text{PD}_p$ . Then a matrix  $M \in \mathbb{R}^{p \times p}$  solves the Lyapunov equation if and only if there exists a skew-symmetric matrix  $K$  such that*

$$M = \left( K - \frac{1}{2}C \right) \Sigma^{-1}.$$

The space of skew-symmetric matrices has dimension  $p(p-1)/2$ . Hence, for  $\Sigma \in \text{PD}_p$ , the kernel of  $A(\Sigma)$  also has dimension  $p(p-1)/2$ . We give further details about the spectral properties of  $A(\Sigma)$  in Theorem 2.10.1. The following result now gives simplified rank conditions for identifiability.

**Lemma 2.6.4.** *Let  $G = (V, E)$  be a directed graph with  $V = [p]$ , and let  $C \in \text{PD}_p$ . For every  $\Sigma \in \text{PD}_p$ , let  $H(\Sigma)$  be a  $p^2 \times p(p-1)/2$  matrix whose columns form a basis of the kernel of  $A(\Sigma)$ , and let  $H(\Sigma)_{E^c}$  be the submatrix obtained by restriction to rows corresponding to non-edges  $E^c$  of  $G$ . Then the associated model  $\mathcal{M}_{G,C}$  is*

- (i) *globally identifiable if and only if  $H(\Sigma)_{E^c}$  has full column rank  $p(p-1)/2$  for all  $\Sigma \in \mathcal{M}_{G,C}$ ;*
- (ii) *generically identifiable if and only if there exists a matrix  $\Sigma \in \mathcal{M}_{G,C}$  such that  $H(\Sigma)_{E^c}$  has full column rank  $p(p-1)/2$ .*

**Proof.** Recall from Lemma 2.4.3 that the elements of the fiber are solutions of the equation system (2.9), which has a unique solution for a given (positive definite) matrix  $\Sigma \in \mathcal{M}_{G,C}$  if and only if  $A(\Sigma)_{\cdot,E}$  has linearly independent columns. The latter condition can be rephrased as follows: the kernel of  $A(\Sigma)$  does not contain any element  $\text{vec}(M) \neq 0$  such that  $M \in \mathbb{R}^E$ . Put differently, (2.9) admits a unique solution if and only if the column span of  $H(\Sigma)$  does not contain any element  $\text{vec}(M) \neq 0$  for  $M \in \mathbb{R}^E$ . As  $H(\Sigma)$  has linearly independent columns, this latter condition is equivalent to the linear independence of the columns of the extended matrix  $(H(\Sigma) \mid \text{vec}(M))$  for any non-trivial  $M \in \mathbb{R}^E$ . It remains to be proven that this, in turn, is equivalent to the  $|E^c| \times p(p-1)/2$  submatrix  $H(\Sigma)_{E^c}$  having rank  $p(p-1)/2$ .



## 2.6 Sums of Squares Decompositions and Finer Rank Conditions

Assume that  $H(\Sigma)_{E^c}$  has rank  $p(p-1)/2$ , and consider one of its non-vanishing maximal minors. This minor can always be extended to a non-vanishing maximal minor of  $(H(\Sigma) \mid \text{vec}(M))$  by adding one of the rows corresponding to  $m_{ji} \neq 0$ . Therefore, the extended matrix has full rank.

For the converse implication, note that if  $H(\Sigma)_{E^c}$  has rank strictly less than  $p(p-1)/2$ , then there exists a (not unique) non-trivial  $M \in \mathbb{R}^E$  such that  $\text{vec}(M)$  belongs to the kernel of  $A(\Sigma)$ .  $\square$

For a convenient choice of a basis of the kernel of  $A(\Sigma)$  we may appeal to the following fact.

**Lemma 2.6.5.** *For a matrix  $\Sigma \in \text{PD}_p$ , the kernel of  $A(\Sigma)$  equals*

$$\begin{aligned} \ker A(\Sigma) &= \{\text{vec}(K\Sigma^{-1}) : K \text{ skew-symmetric}\} \\ &= \{\text{vec}(\Sigma K) : K \text{ skew-symmetric}\}. \end{aligned}$$

**Proof.** The first equality holds by Lemma 2.6.3. The second equality follows from the fact that  $K$  is skew-symmetric if and only if  $\Sigma K \Sigma$  is skew-symmetric.  $\square$

For  $1 \leq k, l \leq p$ , let  $K^{(k,l)} = e_k \otimes e_l - e_l \otimes e_k$  be the skew-symmetric matrix whose only non-zero entries are 1 in place  $(k, l)$  and  $-1$  in place  $(l, k)$ . Then the set  $\{K^{(k,l)} : k < l\}$  is a basis of the space of  $p \times p$  skew-symmetric matrices and, thus, the set  $\{\text{vec}(\Sigma K^{(k,l)}) : k < l\}$  is a basis of  $\ker A(\Sigma)$ . We may thus choose the matrix  $H(\Sigma)$  in Lemma 2.6.5 as the matrix with entries

$$H(\Sigma)_{i \rightarrow j, (k,l)} = \text{vec}(\Sigma K^{(k,l)})_{ji} = \begin{cases} -\Sigma_{lj} & \text{if } i = k, \\ \Sigma_{kj} & \text{if } i = l, \\ 0 & \text{otherwise.} \end{cases} \quad (2.10)$$

Note that we index the rows of  $H(\Sigma)$  by all possible edges of a directed graph (including self-loops), in accordance with the indexing of the columns of  $A(\Sigma)$ .

**Example 2.6.6.** *Consider the  $6 \times 9$  matrix  $A(\Sigma)$  in Example 2.2.1 corresponding to  $p = 3$ . Then the matrix from (2.10) is*

$$H(\Sigma) = \begin{pmatrix} -\Sigma_{12} & 0 & -\Sigma_{13} \\ -\Sigma_{22} & 0 & -\Sigma_{23} \\ -\Sigma_{23} & 0 & -\Sigma_{33} \\ \Sigma_{11} & -\Sigma_{13} & 0 \\ \Sigma_{12} & -\Sigma_{23} & 0 \\ \Sigma_{13} & -\Sigma_{33} & 0 \\ 0 & \Sigma_{12} & \Sigma_{11} \\ 0 & \Sigma_{22} & \Sigma_{12} \\ 0 & \Sigma_{23} & \Sigma_{13} \end{pmatrix} \begin{matrix} 1 \rightarrow 1 \\ 1 \rightarrow 2 \\ 1 \rightarrow 3 \\ 2 \rightarrow 1 \\ 2 \rightarrow 2 \\ 2 \rightarrow 3 \\ 3 \rightarrow 1 \\ 3 \rightarrow 2 \\ 3 \rightarrow 3 \end{matrix}.$$

Consider the DAG on 3 nodes given in Figure 2.1, for which the set of non-edges is  $E^c = \{1 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 3\}$ . Then

$$|\det H(\Sigma)_{E^c}| = \left| \det \begin{pmatrix} -\Sigma_{22} & 0 & -\Sigma_{23} \\ -\Sigma_{23} & 0 & -\Sigma_{33} \\ \Sigma_{13} & -\Sigma_{33} & 0 \end{pmatrix} \right| = \Sigma_{33}(\Sigma_{22}\Sigma_{33} - \Sigma_{23}^2)$$

is a product of two principal minors of  $\Sigma$ , as expected from Theorem 2.5.3.

Next, let  $E^c = \{1 \rightarrow 3, 2 \rightarrow 1, 3 \rightarrow 2\}$  be the set of non-edges of the 3-cycle from Figure 1.2. Then

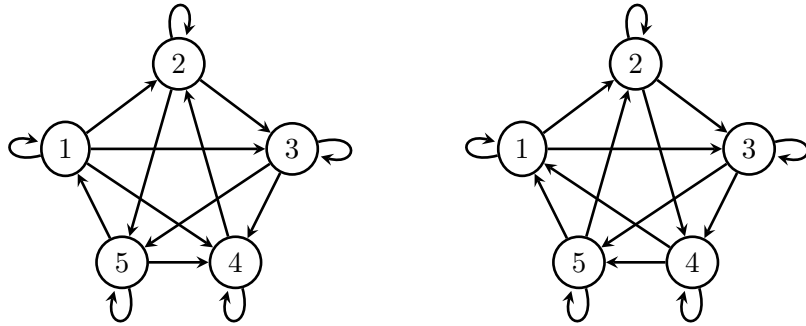
$$|\det H(\Sigma)_{E^c, \cdot}| = \left| \det \begin{pmatrix} -\Sigma_{23} & 0 & -\Sigma_{33} \\ \Sigma_{11} & -\Sigma_{13} & 0 \\ 0 & \Sigma_{22} & \Sigma_{12} \end{pmatrix} \right| = \Sigma_{11}\Sigma_{22}\Sigma_{33} - \Sigma_{12}\Sigma_{13}\Sigma_{23},$$

which is what we obtained in (2.4).

Following Example 2.6.1, we can establish global identifiability by computing an SOS decomposition of the determinant of the restricted kernel  $H(\Sigma)_{E^c, \cdot}$  using the Cholesky decomposition of  $\Sigma$ . Such computations allowed us to establish:

**Proposition 2.6.7.** *Let  $G = (V, E)$  be a simple graph with  $V = [p]$ , and let  $C \in \text{PD}_p$ . Let  $L \in \mathbb{R}^{p \times p}$  be lower-triangular. If  $p \leq 4$ , then there exists a permutation matrix  $P$  such that  $\det H(PLL^\top P^\top)_{E^c, \cdot}$  is an everywhere positive sum of squares in the entries of  $L$ , implying that  $\mathcal{M}_{G,C}$  is globally identifiable. The same is true for  $p = 5$  with the exception of two computationally intractable types of graphs, which are depicted in Figure 2.3.*

For our computer proof of the claims in the proposition, we applied the computer algebra system `Macaulay2`. For the graphs in Figure 2.3, we additionally employed `Matlab` toolboxes, but our computations did not terminate. It is natural to conjecture that Proposition 2.6.7 holds for all graphs with  $p = 5$ , and even all simple graphs.



**Figure 2.3:** The two simple cyclic graphs on 5 nodes, for which a sum of squares decomposition of the determinant of interest is computationally difficult.

## 2.7 Simple Cyclic Graphs

In this section we establish our main result: global identifiability of all Lyapunov models given by simple cyclic graphs. Moreover, we can show that simple cyclic graphs give models that are algebraic subsets of the positive definite cone. Our proofs exploit the parametrization of stable matrices  $M$  that are solutions to the Lyapunov equation in terms of skew-symmetric matrices (matrices  $K$  with  $K^\top = -K$ ); recall Lemma 2.6.3.

**Theorem 2.7.1.** *Let  $G = (V, E)$  be a directed graph with  $V = [p]$ .*

- (i) If  $G$  is simple, then the model  $\mathcal{M}_{G,C}$  is globally identifiable for all  $C \in \text{PD}_p$ .
- (ii) If  $C \in \text{PD}_p$  is diagonal, then the model  $\mathcal{M}_{G,C}$  is globally identifiable if and only if  $G$  is simple.

**Proof.** It suffices to prove (i), as (ii) then follows from Proposition 2.3.10.

To prove (i), suppose  $G$  is indeed simple. Let  $M_1, M_2 \in \text{Stab}(E)$  be any two matrices that solve the Lyapunov equation (1.2) for the same  $\Sigma \in \mathcal{M}_{G,C}$ . According to Lemma 2.6.3 there exist two skew-symmetric matrices  $K_1$  and  $K_2$  such that  $M_1 = (K_1 - \frac{1}{2}C)\Sigma^{-1}$  and  $M_2 = (K_2 - \frac{1}{2}C)\Sigma^{-1}$ . For the difference we obtain

$$M := M_1 - M_2 = (K_1 - \frac{1}{2}C)\Sigma^{-1} - (K_2 - \frac{1}{2}C)\Sigma^{-1} = (K_1 - K_2)\Sigma^{-1}.$$

The difference  $K = K_1 - K_2$  is again skew-symmetric, so that  $M$  is the product of a skew-symmetric matrix  $K$  and the positive definite matrix  $\Sigma^{-1}$ .

Consider now the square  $M^2$ . We have

$$M^2 = K\Sigma^{-1}K\Sigma^{-1}.$$

As  $\Sigma$  is positive definite, the square root  $\Sigma^{\frac{1}{2}}$  exists, and  $M^2$  is similar to

$$\Sigma^{-\frac{1}{2}}M^2\Sigma^{\frac{1}{2}} = \Sigma^{-\frac{1}{2}}K\Sigma^{-1}K\Sigma^{-\frac{1}{2}}.$$

As  $K$  is skew-symmetric,

$$\Sigma^{-\frac{1}{2}}K\Sigma^{-1}K\Sigma^{-\frac{1}{2}} = -(\Sigma^{-\frac{1}{2}}K)\Sigma^{-1}(\Sigma^{-\frac{1}{2}}K)^\top.$$

We observe that  $M^2$  is similar to a symmetric and negative semi-definite matrix. Therefore, the eigenvalues of  $M^2$  are non-positive and  $\text{tr}(M^2) \leq 0$ .

As  $M$  is supported over a simple graph, it holds for all pairs of indices  $i \neq j$  that  $m_{ij} \neq 0$  implies that  $m_{ji} = 0$ . Hence, the diagonal of  $M^2$  is given by the squared diagonal elements of  $M$ , i.e.,  $(M^2)_{ii} = m_{ii}^2$ . It follows that

$$0 \leq \sum_{i=1}^p m_{ii}^2 = \text{tr}(M^2) \leq 0,$$

which implies that  $\text{tr}(M^2) = 0$ .

Let  $\lambda_1, \dots, \lambda_p \in \mathbb{C}$  be the eigenvalues of  $M$ . The eigenvalues of  $M^2$  are then  $\lambda_1^2, \dots, \lambda_p^2$ . Since  $M^2$  is similar to a negative semi-definite matrix, all its eigenvalues satisfy  $\lambda_1^2, \dots, \lambda_p^2 \leq 0$ . Then,

$$0 = \text{tr}(M^2) = \sum_{i=1}^p \lambda_i^2 \leq 0,$$

which implies that  $\lambda_i^2 = 0$  for all  $i \in 1, \dots, p$ . But this is only true if  $\lambda_i = 0$  for all  $i = 1, \dots, p$ . Therefore, all eigenvalues of  $M$  are zero.

Observe that  $M = K\Sigma^{-1}$  is similar to  $\tilde{M} = \Sigma^{-\frac{1}{2}}K\Sigma^{-1}\Sigma^{\frac{1}{2}}$ , which is skew-symmetric since

$$\tilde{M}^\top = (\Sigma^{-\frac{1}{2}}K\Sigma^{-\frac{1}{2}})^\top = \Sigma^{-\frac{1}{2}}K^\top\Sigma^{-\frac{1}{2}} = -\Sigma^{-\frac{1}{2}}K\Sigma^{-\frac{1}{2}} = -\tilde{M}.$$

Skew-symmetric matrices are diagonalizable, and we deduce that  $M$  is similar to the zero matrix. But then  $M = 0$  and consequently  $M_1 = M_2$ , which shows that the Lyapunov equation admits a unique sparse solution.  $\square$

In addition to global identifiability, we have a generalization of Corollary 2.5.4 to general simple graphs.

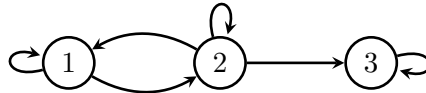
**Corollary 2.7.2.** *Let  $G = (V, E)$  be a simple graph with  $V = [p]$ . Then  $\mathcal{M}_{G,C}$  is an algebraic and thus closed subset of  $\text{PD}_p$ . If  $G$  is complete then  $\mathcal{M}_{G,C} = \text{PD}_p$ .*

**Proof.** Consider first the case where  $G$  is complete (with an edge between every pair of nodes). Let  $\Sigma_0 \in \text{PD}_p$  be an arbitrary positive definite matrix. Choosing  $M = -I_p$ , the negated identity matrix, shows that  $\Sigma_0$  belongs to the model  $\mathcal{M}_{G,C_0}$  for  $C_0 = 2\Sigma_0$ . By Theorem 2.7.1 and Lemma 2.4.3, we obtain that the determinant of  $A(\Sigma)_{\cdot,E}$  is non-zero at every matrix in  $\mathcal{M}_{G,C_0}$  and, in particular, at  $\Sigma_0$ . We conclude that  $\det(A(\Sigma)_{\cdot,E}) \neq 0$  on all of  $\text{PD}_p$ . As in the proof of Corollary 2.5.4, we deduce that  $\mathcal{M}_{G,C} = \text{PD}_p$  for all  $C \in \text{PD}_p$ .

If  $G$  is not complete, then it can be augmented to a complete graph  $\bar{G} = (V, \bar{E})$ , and we may complete the proof in analogy to the proof of Corollary 2.5.4.  $\square$

## 2.8 Non-Simple Graphs

In this section, we consider directed graphs  $G = (V, E)$  that are allowed to be non-simple, i.e., may contain a two-cycle. In our study, we restrict attention to the case where  $C \in \text{PD}_p$  is diagonal. Proposition 2.3.10 tells us that, for  $C$  diagonal, a model  $\mathcal{M}_{G,C}$  given by a non-simple graph  $G$  can never be globally identifiable. However, non-simple graphs with at most  $p(p+1)/2$  edges may still give generically identifiable models (Definition 2.3.1, Lemma 2.3.5). We are able to provide a combinatorial condition that is necessary for generic identifiability, and we computationally classify all graphs with  $p \leq 5$  nodes. Our study reveals examples for which generic identifiability depends in subtle ways on the pattern of edges.



**Figure 2.4:** Non-simple graph on 3 nodes.

We begin with a small example.

**Example 2.8.1.** *Let  $G = (V, E)$  be the graph from Figure 2.4, a 2-cycle with an additional edge pointing to a third node, and let  $C \in \text{PD}_3$  be a diagonal matrix. To inspect identifiability of  $\mathcal{M}_{G,C}$ , we may use the kernel basis of Example 2.6.6 with the*

set of non-edges  $E^c = \{1 \rightarrow 3, 3 \rightarrow 1, 3 \rightarrow 2\}$ . We find

$$\det H(\Sigma)_{E^c, \cdot} = \det \begin{pmatrix} -\Sigma_{23} & 0 & -\Sigma_{33} \\ 0 & \Sigma_{12} & \Sigma_{11} \\ 0 & \Sigma_{22} & \Sigma_{12} \end{pmatrix} = \Sigma_{23} (\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2).$$

Since  $\mathcal{M}_{G,C}$  contains positive definite matrices with both vanishing and non-vanishing  $\Sigma_{23}$ , we conclude that  $\mathcal{M}_{G,C}$  is generically (but not globally) identifiable.

Note that the matrices  $\Sigma \in \mathcal{M}_{G,C}$  with  $\Sigma_{23} = 0$  are obtained precisely from the drift matrices in the lower-dimensional set  $\{M \in \text{Stab}(E) : m_{32} = 0\}$ . Indeed, if  $m_{32} = 0$ , then the situation is as if the  $2 \rightarrow 3$  edge were removed, and we will see in Proposition 2.8.3 that this implies  $\Sigma_{23} = 0$  when  $C$  is diagonal. Conversely, when solving for  $\Sigma$  given a drift matrix  $M \in \mathbb{R}^E$  we find that  $\Sigma_{23}$  is a rational function of  $(M, C)$  whose numerator is

$$m_{32} (c_{11}m_{21}^2 \text{tr}(M) + c_{22}m_{11}^2 \text{tr}(M) + c_{22} \det(M)).$$

As  $C$  is positive definite and  $M$  stable, the second factor is negative. Thus, if  $\Sigma = \Sigma(M, C)$  is a positive definite matrix in  $\mathcal{M}_{G,C}$ , then  $\Sigma_{23} = 0$  implies  $m_{32} = 0$ .

By Lemma 2.3.5,  $|E| \leq p(p+1)/2$  is a necessary condition for generic identifiability of the model of a graph  $G = (V, E)$ . We now show how this bound may be improved by accounting for knowledge about vanishing covariances.

**Definition 2.8.2.** A trek is a sequence of edges of the form

$$l_m \leftarrow l_{m-1} \leftarrow \cdots \leftarrow l_1 \leftarrow t \rightarrow r_1 \rightarrow \cdots \rightarrow r_{n-1} \rightarrow r_n.$$

The node  $t$  is the top node of the trek. The directed paths  $l_m \leftarrow l_{m-1} \leftarrow \cdots \leftarrow l_1$  and  $r_1 \rightarrow \cdots \rightarrow r_{n-1} \rightarrow r_n$  are the left and the right side of the trek, respectively. The definition allows for one or both sides to be trivial, so directed paths and also single nodes are treks.

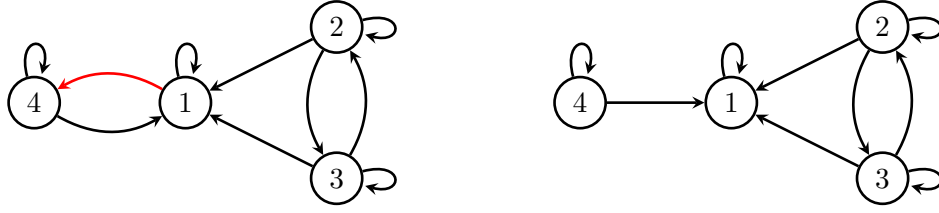
From Varando and Hansen [2020, Corollary 2.3], we deduce the following fact.

**Proposition 2.8.3.** Let  $G = (V, E)$  be a directed graph with  $V = [p]$ , and let  $C \in \text{PD}_p$  be diagonal. If there is no trek from  $i$  to  $j$  in  $G$ , then  $\Sigma_{ij} = 0$  in all matrices  $\Sigma \in \mathcal{M}_{G,C}$ .

**Example 2.8.4.** Let  $C \in \text{PD}_4$  be diagonal. Then the left graph  $G_1 = (V, E_1)$  in Figure 2.5 defines a generically identifiable model but its subgraph  $G_2 = (V, E_2)$  does not. This example stresses that global identifiability is needed in Proposition 2.3.6. But why is  $\mathcal{M}_{G_2,C}$  non-identifiable despite  $G_2$  having fewer edges? We observe that  $G_2$  contains no trek between 2 and 4 and no trek between 3 and 4. Proposition 2.8.3 yields  $\Sigma_{24} = \Sigma_{34} = 0$ . Although the  $\text{PD}_4$ -cone has dimension  $\binom{4+1}{2} = 10$ , the existence of the constraints  $\Sigma_{24} = \Sigma_{34} = 0$  implies that  $\dim(\mathcal{M}_{G_2,C}) \leq 10 - 2 = 8$ . Since  $|E_2| = 9 > 8$ , non-identifiability follows from by Lemma 2.3.5.

As a last subtlety, we emphasize that if we remove one of the edges  $2 \rightarrow 1$ ,  $3 \rightarrow 1$ , or  $4 \rightarrow 1$  of  $G_2$ , we are left again with a generically identifiable model.

The ideas in Example 2.8.4 can be generalized into a sharper necessary condition for identifiability that is a consequence of Lemma 2.3.5 and Proposition 2.8.3.



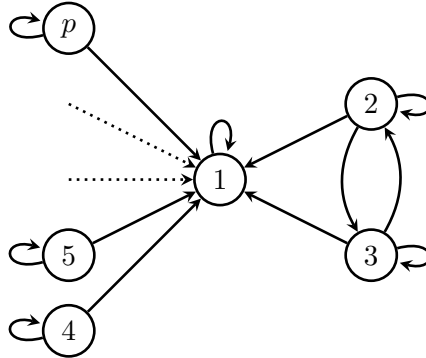
**Figure 2.5:** Left: graph  $G_1$  on 4 nodes with  $\mathcal{M}_{G_1, C}$  generically identifiable. Right: subgraph  $G_2$  of  $G_1$  such that  $\mathcal{M}_{G_2, C}$  is non-identifiable.  $C \in \text{PD}_4$  is diagonal.

**Corollary 2.8.5.** *Let  $G = (V, E)$  be a directed graph with  $V = [p]$ . If  $\mathcal{M}_{G, C}$  is generically identifiable for a diagonal matrix  $C \in \text{PD}_p$ , then it has to hold that*

$$|E| \leq \frac{p(p+1)}{2} - \#\{ \{i, j\} : i, j \in V \text{ with no trek between them} \}. \quad (2.11)$$

With this criterion in hand, we can construct graphs of arbitrary size  $p$  and fewer than  $p(p+1)/2$  edges that yield non-identifiable models.

**Corollary 2.8.6.** *Consider the graph  $G = (V, E)$  with  $p \geq 4$  nodes displayed in Figure 2.6. The model  $\mathcal{M}_{G, C}$  is non-identifiable for any diagonal  $C \in \text{PD}_p$ .*



**Figure 2.6:** Graph  $G$  with  $V = [p]$  such that  $\mathcal{M}_{G, C}$  is non-identifiable for diagonal  $C \in \text{PD}_p$ .

**Proof.** The number of parameters  $|E|$  is

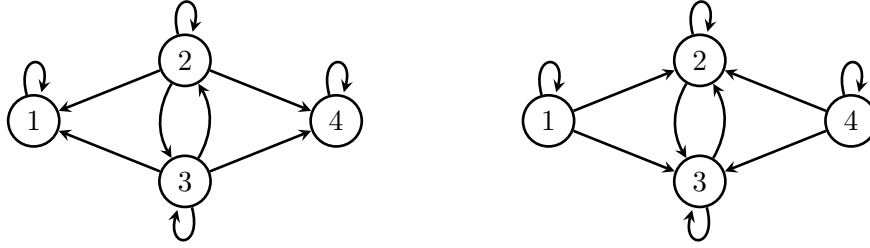
$$\begin{aligned} & 2 \text{ (edges from 2-cycle) } + p - 1 \text{ (edges pointing to node 1)} \\ & \quad + p \text{ (parameters due to the selfloops) } = 2p + 1. \end{aligned}$$

There are no treks between any pair of nodes  $\{2, \dots, p\}$  except for the pair  $(2, 3)$ . This results in  $\binom{p-1}{2} - 1$  (unordered) pairs of nodes with no trek. Corollary 2.8.5 implies that

$$\dim(\mathcal{M}_{G, C}) \leq \frac{p(p+1)}{2} - \binom{p-1}{2} + 1 = 2p.$$

□

Unfortunately, the criterion in Corollary 2.8.5 is not sufficient.



**Figure 2.7:** Left: graph fulfilling the criterion in Corollary 2.8.5, yet yields a non-identifiable model. Right: Reversing edges retains non-identifiability, due to Corollary 2.8.5, as  $\Sigma_{14} = 0$ .

**Example 2.8.7.** Let  $G_1 = (V, E)$  be the left graph in Figure 2.7. Graph  $G_1$  fulfills the necessary condition of Corollary 2.8.5 as the number of parameters is  $6 + 4 = 10$  and all pairs of nodes are connected with a trek, which is why the right side of equation (2.11) is also  $\binom{4+1}{2} = 10$ . However,  $A(\Sigma)_{\cdot, E} \in \mathbb{R}^{10 \times 10}$  does not have full rank because the columns of  $A(\Sigma)$  may be linearly combined to

$$\begin{aligned} & \Sigma_{13}A(\Sigma)_{\cdot, 2 \rightarrow 1} + \Sigma_{23}A(\Sigma)_{\cdot, 2 \rightarrow 2} + \Sigma_{33}A(\Sigma)_{\cdot, 2 \rightarrow 3} + \Sigma_{34}A(\Sigma)_{\cdot, 2 \rightarrow 4} \\ & - \Sigma_{12}A(\Sigma)_{\cdot, 3 \rightarrow 1} - \Sigma_{22}A(\Sigma)_{\cdot, 3 \rightarrow 2} - \Sigma_{23}A(\Sigma)_{\cdot, 3 \rightarrow 3} - \Sigma_{24}A(\Sigma)_{\cdot, 3 \rightarrow 4} = 0. \end{aligned}$$

Therefore, the model  $\mathcal{M}_{G_1, C}$  is non-identifiable for  $C$  diagonal despite fulfilling the necessary criterion. The right graph in Figure 2.7 yields a non-identifiable model for the simple reason that the necessary condition of Corollary 2.8.5 is violated due to the absence of a trek between nodes 1 and 4.

For smaller examples, we may check generic identifiability by choosing random drift matrices and determining whether the resulting matrix  $\Sigma$  satisfies the rank condition from Lemma 2.4.3. When this does not succeed we can check symbolically whether the corresponding restriction of the coefficient matrix  $A(\Sigma)$  or the restricted kernel basis  $H(\Sigma)$  from Lemma 2.6.4 is rank-deficient, thus implying non-identifiability. We implemented this strategy for all non-simple graphs with  $p \leq 5$  nodes and less than  $p(p+1)/2$  parameters. As justified by Proposition 2.3.7, we took  $C = I_p$  in our computations. This led to the results displayed in Table 2.1, which shows that the majority of graphs are generically identifiable. The details of the computations can be found at <https://mathrepo.mis.mpg.de/LyapunovIdentifiability>.

**Table 2.1:** Classification of models with  $p = 3, 4, 5$  nodes and  $C = I_p$ . The last column displays the number of non-identifiable models whose underlying graphs satisfy the necessary criterion for generic identifiability in Corollary 2.8.5.

nodes	total non-simple	non-identifiable	non-identifiable satisfying (2.11)
3	2	0	0
4	80	3	2
5	4862	68	37

## 2.9 Volatility Matrix: Diagonal vs. Non-Diagonal PD Matrix

This section aims at providing insight into the need of the diagonality constraint on the volatility matrix  $C \in \text{PD}_p$  of the Lyapunov equation to ensure that some of the stronger results of this chapter hold.

**Example 2.9.1.** Let  $G$  be the 2-cycle with an additional third node, so  $V = [3]$  and  $E = \{1 \rightarrow 1, 1 \rightarrow 2, 2 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3\}$ , which encodes drift matrices

$$M = \begin{pmatrix} m_{11} & m_{12} & 0 \\ m_{21} & m_{22} & 0 \\ 0 & 0 & m_{33} \end{pmatrix}.$$

Let  $C = (c_{ij}) \in \text{PD}_3$ . Clearly, the graph does not contain any treks between nodes 1 and 3, nor between nodes 2 and 3. However, a matrix  $\Sigma = \phi_{G,C}(M)$  has

$$\Sigma_{13} = \frac{c_{23}m_{12} - c_{13}(m_{22} + m_{33})}{(m_{11}m_{22} - m_{12}m_{21}) + m_{33}(m_{11} + m_{22} + m_{33})},$$

with a denominator that is positive on  $\text{Stab}(E)$  and a numerator that is constant zero only if  $c_{13} = c_{23} = 0$ . The same holds for  $\Sigma_{23}$  by symmetry. This example serves to highlight that Proposition 2.8.3 may be false when  $C$  is not diagonal. Indeed, the treks would need to be allowed to move along new edges that reflect presence of non-zero diagonal entries in  $C$ ; compare Varando and Hansen [2020].

**Example 2.9.2.** Consider again the 2-cycle with an additional third node from the previous example. Again, consider an arbitrary matrix  $C = (c_{ij}) \in \text{PD}_3$ . The kernel basis of Example 2.6.6 restricted to the set of non-edges  $E^c = \{1 \rightarrow 3, 2 \rightarrow 3, 3 \rightarrow 1, 3 \rightarrow 2\}$ , namely

$$H(\Sigma)_{E^c} = \begin{pmatrix} -\Sigma_{23} & 0 & -\Sigma_{33} \\ \Sigma_{13} & -\Sigma_{33} & 0 \\ 0 & \Sigma_{12} & \Sigma_{11} \\ 0 & \Sigma_{22} & \Sigma_{12} \end{pmatrix},$$

is rank deficient for any  $\Sigma \in \text{PD}_3$  if and only if  $\Sigma_{13} = \Sigma_{23} = 0$ . Adding this constraint to the Lyapunov equation yields  $c_{13} = c_{23} = 0$ . Therefore, it follows from Lemma 2.6.4 that  $\mathcal{M}_{G,C}$  is globally identifiable for any  $C = (c_{ij}) \in \text{PD}_3$  in which  $c_{13}$  and  $c_{23}$  do not vanish simultaneously.

Observe that this provides a counterexample to Proposition 2.3.10 and Proposition 2.3.8 when dropping the diagonality assumption. To begin with, such  $\mathcal{M}_{G,C}$  is an instance of a globally identifiable model associated to a non-simple graph. Moreover, the subgraph  $H$  obtained by removing node 3 from  $G$  defines a non-identifiable model for all positive definite volatility matrices by Lemma 2.3.5.

For the sake of completeness, note that, by Example 2.9.1,  $c_{13} = c_{23} = 0$  completely describes when the rank of  $H(\Sigma)_{E^c}$  drops for all  $\Sigma \in \mathcal{M}_{G,C}$ . In other words, the model is non-identifiable if and only if  $c_{13} = c_{23} = 0$  and globally identifiable otherwise.



## 2.10 Spectral Description, Kernel and Factorization

Here, we collect spectral properties of  $A(\Sigma)$ , derived more conveniently for its square  $p \times p$  version

$$\tilde{A}(\Sigma) = \Sigma \otimes I_p + (I_p \otimes \Sigma)K_p,$$

which features in Lemma 2.4.1. We will then use this information to clarify that  $\det(\Sigma)$  is a factor of  $\det(A(\Sigma)_{\cdot, E})$  for complete graphs which have edge sets of size  $|E| = p(p+1)/2$ ; see Corollary 2.10.2.

**Theorem 2.10.1.** *Let  $\Sigma \in \text{PD}_p$ , and let  $(\lambda_i)_{i \in [p]}$  be its eigenvalues with corresponding orthogonal eigenvectors  $(z_i)_{i \in [p]}$ .*

- (i) *The matrix  $\tilde{A}(\Sigma)$  has rank  $p(p+1)/2$ , and (2.10) gives a basis for its kernel.*
- (ii) *The transposed matrix  $\tilde{A}(\Sigma)^\top$  has rank  $p(p+1)/2$ , and a basis for its kernel is given by  $\text{vec}(e_i \otimes e_j - e_j \otimes e_i)$  for  $1 \leq i < j \leq p$ .*
- (iii) *Counting with multiplicities, the  $p(p+1)/2$  non-zero eigenvalues of  $\tilde{A}(\Sigma)$  and of  $\tilde{A}(\Sigma)^\top$  are given by the sums  $\lambda_i + \lambda_j$  for  $1 \leq i \leq j \leq p$  and for either matrix the associated set of orthogonal eigenvectors is  $\text{vec}(z_i \otimes z_j + z_j \otimes z_i)$  for  $1 \leq i \leq j \leq p$ .*

**Proof.** (i) follows from (2.10), and (ii) follows from the symmetry of the Lyapunov (matrix) equation.

For (iii), the claim about  $\tilde{A}(\Sigma)$  follows from the calculation

$$\begin{aligned} & (z_i \otimes z_j + z_j \otimes z_i)\Sigma + \Sigma(z_i \otimes z_j + z_j \otimes z_i)^\top \\ &= [\lambda_j(z_i \otimes z_j) + \lambda_i(z_j \otimes z_i)] + [\lambda_i(z_i \otimes z_j) + \lambda_j(z_j \otimes z_i)] \\ &= (\lambda_i + \lambda_j)(z_i \otimes z_j + z_j \otimes z_i). \end{aligned}$$

The transpose  $\tilde{A}(\Sigma)^\top = \Sigma \otimes I_p + K_p(I_p \otimes \Sigma)$  encodes the Lyapunov equation with  $M$  replaced by  $M^\top$  and the claim about  $\tilde{A}(\Sigma)^\top$  follows from the symmetry of the matrices  $z_i \otimes z_j + z_j \otimes z_i$ . The orthogonality of the eigenvectors holds because

$$\text{tr}((z_i \otimes z_j + z_j \otimes z_i)(z_k \otimes z_l + z_l \otimes z_k)) = 0$$

unless  $\{i, j\} = \{k, l\}$ . □

As a consequence of Theorem 2.10.1, we can conclude information regarding the factorization of the determinant of  $A(\Sigma)_{\cdot, E}$  when  $|E| = p(p+1)/2$  such that  $A(\Sigma)_{\cdot, E}$  is a square matrix.

**Corollary 2.10.2.** *Let  $G = (V, E)$  be a directed graph with  $V = [p]$  and  $|E| = p(p+1)/2$ . The polynomials  $\det(\Sigma)$  and  $\det(H(\Sigma)_{E^c, \cdot})$  are factors of  $\det(A(\Sigma)_{\cdot, E})$ .*

**Proof.** The zero set of the determinant  $\det(\Sigma)$  is the set of singular symmetric matrices. Since  $\det(\Sigma)$  is an irreducible polynomial, every polynomial that vanishes at all singular matrices must be a polynomial multiple of  $\det(\Sigma)$ . Hence, it suffices to show that  $\det(A(\Sigma)_{\cdot, E}) = 0$  for all singular matrices  $\Sigma$ . So let  $\Sigma$  be a singular matrix. Then there exists an eigenvalue  $\lambda_i = 0$  with  $i \in [p]$ . Using Theorem 2.10.1 this implies that the eigenvalue  $\lambda_i + \lambda_i$  of  $\tilde{A}(\Sigma)$  is zero (the theorem is written for  $\Sigma$  positive definite

but the fact we used also holds for  $\Sigma$  singular). Hence,  $\text{rank}(\tilde{A}(\Sigma)) \leq p(p+1)/2 - 1$  which implies that  $\text{rank}(A(\Sigma)_{\cdot,E}) \leq p(p+1)/2 - 1$  and thus  $\det(A(\Sigma)_{\cdot,E}) = 0$ .

The fact that  $\det(H(\Sigma)_{E^c,\cdot})$  is a factor of  $\det(A(\Sigma)_{\cdot,E})$  follows from the proof of Lemma 2.6.4.  $\square$

## 2.11 Outlook: Identifiability for Partially Unknown Volatility Matrices

Compared to standard directed graphical models, our approach starting with  $C$  known is in the spirit of the work of Peters and Bühlmann [2014] on homoscedasticity. From an application perspective, assuming  $C$  to be known is a limitation. Ideally, one would like to estimate  $M$  and  $C$  simultaneously. This requires a solid theoretical basis including identifiability theory for that case. There are no “cheap” results in the sense that the results Theorem 2.5.3 and Theorem 2.7.1 can immediately be extended to the case  $C$  unknown. Nevertheless, some considerations regarding the structure of  $A(\Sigma)$  used in Theorem 2.5.3 might be helpful when investigating this case. The objective of this section is not to provide detailed identifiability theory for the case  $C$  unknown but to showcase that the general patterns of this chapter also reoccur if the volatility matrix is assumed to be (partially) unknown. We present a simple and straightforward example in this section.

We consider a setup where the diagonal entries in  $C$  are assumed to be known, but unknown off-diagonal entries in  $C$  might exist. This results in a mixed graph  $G = (V, E, B)$  where  $V$  and  $E$  are as in Definition 1.3.1 and  $B = \{i \leftrightarrow j : i, j \in V\}$  is the set of blunt edges with  $C_{ij} = 0$  implying  $i \leftrightarrow j \notin B$ . More details on blunt edges are given in the work of Varando and Hansen [2020]. The models associated to such graphs  $G = (V, E, B)$  are then given by the set of covariance matrices

$$\tilde{\mathcal{M}}_G = \{\Sigma \in \text{PD}_p : M\Sigma + \Sigma M^\top = -C \text{ with } (M, C) \in \mathbb{R}^E \times \text{PD}_p^{\text{diag}=1}(B)\}, \quad (2.12)$$

where  $\text{PD}_p^{\text{diag}=1}(B)$  are the positive-definite matrices supported over  $B$  where the diagonal elements  $c_{11}, \dots, c_{pp}$  are set to 1.

**Remark 2.11.1.** *The diagonal elements of the volatility matrix  $C$  are fixed to be one without loss of generality. Other choices for the diagonal elements are possible too.*

Again, we can ask if it is possible to uniquely recover the entries in  $M$  and  $C$  from the equilibrium covariance matrix  $\Sigma$  if we fix the support of  $M$  and  $C$ . This can be proven in certain cases simply by using the observations made in Theorem 2.5.3.

**Corollary 2.11.2.** *Let  $G = (V, E, B)$  be a mixed graph and let  $\tilde{E} = \{(i, j) : j \rightarrow i \in E \text{ and } i \neq j\}$  and let  $\tilde{B} = \{(i, j) : i \leftrightarrow j \in B \text{ and } i \neq j\}$ . If  $\tilde{E} \sqcup \tilde{B} = \{(i, j) : i, j \in 1, \dots, p \text{ and } i < j\}$ , we have that  $A(\Sigma)_{-\tilde{B}, \tilde{E}}$  has a block-structure with principal minors of  $\Sigma \in \text{PD}_p$  as blocks. The subscript  $-\tilde{B}$  indicates that the rows indexed by  $\tilde{B}$  are dropped. Moreover provided  $\Sigma \in \text{PD}_p$ , a unique solution for the entries in  $M \in \text{Stab}(E)$  and  $C \in \text{PD}_p^{\text{diag}=1}(B)$  exists.*

**Proof.** Consider the complete directed acyclic graph  $H = (V, D)$  where  $D$  is used to denote the set of directed edges of  $G$ . By Theorem 2.5.3, the matrix  $A(\Sigma)_{\cdot,D}$  has

## 2.11 Outlook: Identifiability for Partially Unknown Volatility Matrices

an upper triangular blockstructure with principal minors of  $\Sigma$  of decreasing size as diagonal blocks. In the equation system

$$A(\Sigma)_{\cdot,D} \text{vec}(M)_D = -\text{vech}(C)$$

each row of  $A(\Sigma)_{\cdot,D}$  corresponds to an element of  $\text{vech}(C)$ . To solve for the unknown entries in  $C$ , the rows indexed by  $\tilde{B}$  are required. As the matrix  $A(\Sigma)_{\cdot,D}$  has an upper triangular blockstructure, removing the rows indexed by  $\tilde{B}$  and the columns indexed by  $\{i \rightarrow j : (j, i) \in \tilde{B}\}$  preserves the blockstructure. The resulting equation system

$$A(\Sigma)_{-\tilde{B},E} \text{vec}(M)_E = -\text{vech}(C)_{-\tilde{B}},$$

is uniquely solvable for  $\text{vec}(M)_E$ .  $\square$

**Example 2.11.3.** Consider the mixed graph  $G = (V, E, B)$  with  $V = \{1, 2, 3\}$  and  $E = \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, 2 \rightarrow 1, 3 \rightarrow 1\}$  and  $B = \{1 \leftrightarrow 2, 2 \leftrightarrow 3, 3 \leftrightarrow 2\}$  presented in Figure 2.8. We omit drawing the self-loops.

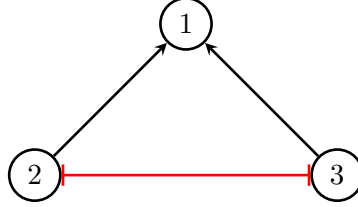


Figure 2.8:  $G$ , a DAG with a blunt edge.

The corresponding volatility matrix  $C \in \text{PD}_p^{\text{diag}=1}(B)$  is

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & c_{23} \\ 0 & c_{23} & 1 \end{pmatrix}.$$

Then, we have

$$\Delta C = \begin{pmatrix} 1 & \star & \star \\ 0 & 1 & \star \\ 0 & c_{23} & 1 \end{pmatrix} \quad \text{and} \quad \text{vech}(C) = (1 \ 0 \ 0 \ 1 \ c_{23} \ 1)^\top.$$

Recall that the matrix  $A(\Sigma)_{\cdot,D}$  for the complete DAG  $\tilde{G} = (V, D)$  presented in Example 2.5.2 where we use  $D$  to denote the set of  $\tilde{G}$

$$A(\Sigma)_{\cdot,D} = \begin{matrix} & \begin{matrix} 1 \rightarrow 1 & 2 \rightarrow 1 & 3 \rightarrow 1 & 2 \rightarrow 2 & 3 \rightarrow 2 & 3 \rightarrow 3 \end{matrix} \\ \begin{matrix} (1,1) \\ (1,2) \\ (1,3) \\ (2,2) \\ (2,3) \\ (3,3) \end{matrix} & \begin{pmatrix} 2\Sigma_{11} & 2\Sigma_{12} & 2\Sigma_{13} & 0 & 0 & 0 \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{12} & \Sigma_{13} & 0 \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & 0 & 0 & \Sigma_{13} \\ 0 & 0 & 0 & 2\Sigma_{22} & 2\Sigma_{23} & 0 \\ 0 & 0 & 0 & \Sigma_{32} & \Sigma_{33} & \Sigma_{23} \\ 0 & 0 & 0 & 0 & 0 & 2\Sigma_{33} \end{pmatrix} \end{matrix}.$$

The equation resulting from row (2,3) is needed to determine the entry  $c_{23}$  and the directed edge  $3 \rightarrow 2$  is missing in Figure 2.8. We obtain

$$A(\Sigma)_{-\tilde{B},E} = \begin{matrix} & \begin{matrix} 1 \rightarrow 1 & 2 \rightarrow 1 & 3 \rightarrow 1 & 2 \rightarrow 2 & 3 \rightarrow 3 \end{matrix} \\ \begin{matrix} (1,1) \\ (1,2) \\ (1,3) \\ (2,2) \\ (3,3) \end{matrix} & \begin{pmatrix} 2\Sigma_{11} & 2\Sigma_{12} & 2\Sigma_{13} & 0 & 0 \\ \Sigma_{12} & \Sigma_{22} & \Sigma_{23} & \Sigma_{12} & 0 \\ \Sigma_{13} & \Sigma_{23} & \Sigma_{33} & 0 & \Sigma_{13} \\ 0 & 0 & 0 & 2\Sigma_{22} & 0 \\ 0 & 0 & 0 & 0 & 2\Sigma_{33} \end{pmatrix} \end{matrix}$$

which has a blocktriangular structure with the diagonal elements being principle minors.

## 2.12 Summary of the Chapter

This chapter addresses the fundamental problem of whether, up to joint scaling, the parameters of the dynamic process can be identified from the covariance matrix of the cross-sectional equilibrium observations. Our main contribution shows that simple graphs yield globally identifiable models, and that the graph being simple is necessary and sufficient for global identifiability in the case where the volatility matrix  $C$  is diagonal. Moreover, we are able to show that the models of simple graphs are closed algebraic subsets of the positive definite cone. In particular, the models of complete simple graphs equal the entire positive definite cone.

Our analysis of directed acyclic graphs (DAGs) highlights block structure in the coefficient matrix for the Lyapunov equation. This leads to a straightforward proof of global identifiability and also reveals that the determinant studied in our rank conditions is a positive sum of squares in the entries of a Cholesky factor. This sum of squares property was also observed in small cyclic graphs.

While we were able to characterize global identifiability, we know less about generic identifiability of graphical Lyapunov models. Our results include an effective necessary but not sufficient graphical criterion for non-simple graphs to be generically identifiable. We also obtain a computational classification of graphs with up to 5 nodes, and we hope that future research will lead to an improved understanding of generic identifiability of the models we considered.

# Chapter 3

## Direct Lyapunov Lasso

This chapter is largely based on the publication by Dettling et al. [2024]. However, Section 3.1 and Section 3.7 contain new material. We omit drawing self-loops in this Chapter and in Chapter 4.

### 3.1 Introduction

In this chapter, we thoroughly analyze the direct Lyapunov lasso which is an intuitive, easy-to-implement and convex model selection method for Lyapunov models. The idea was raised by Fitch [2019] and by Varando and Hansen [2020] in independent work.

In Chapter 2, we vectorize the Lyapunov equation for the purpose of analyzing the identifiability question. At the same time, the vectorization reveals a similarity of Lyapunov models and classical regression problems.

For a better comparison, we briefly introduce (sparse) regression. Of course, there exist many works on that topic. However, the work by Hastie et al. [2015] provides a complete overview over sparse regression.

They consider the classical regression setup with  $n$  observations of multiple variables that are contained within the rows of the design matrix  $X$  of size  $n \times p$ . One row of the design matrix is given by  $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  where the individual entries are the  $p$  predictor variables. Based on the observation, a linear regression model postulates that the outcome  $y_i$  depends linearly on the predictors assuming

$$y_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \epsilon_i. \quad (3.1)$$

The unknown parameters are the intercept  $\beta_0$  and the vector  $\beta = (\beta_1, \dots, \beta_p)^\top$  and  $\epsilon_i$  is an additive noise term. Naturally, the question arises how to estimate  $\beta$  based on a given design matrix  $X$  and a response vector  $y$ . Neglecting the intercept and using matrix notation, the least squares estimate is given by solving

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2n} \|X\beta - y\|_2^2.$$

However, this method does not work if the number of variables is larger than the sample size ( $p > n$ ). The reason is visible when deriving the closed form solution for  $\hat{\beta}$ . Simple calculations yield

$$\hat{\beta} = (X^\top X)^{-1} X^\top y.$$

The matrix  $X^\top X$  is not invertible in the setting where  $p > n$ . There is a solution to this problem by assuming that only some of the variables are active. The so-called Lasso method was introduced in statistics by Tibshirani [1996] although the idea existed previously in natural sciences. A detailed summary of the Lasso method is given by Hastie et al. [2015]. Additionally to minimizing the squared  $\ell_2$ -norm, an  $\ell_1$ -penalty is added. In Lagrangian form the optimization problem is

$$\hat{\beta}_{Lasso} = \arg \min_{\beta \in \mathbb{R}^{p^2}} \frac{1}{2n} \|X\beta - y\|_2^2 + \lambda \|\beta\|_1,$$

where  $\lambda > 0$  is a tuning parameter. The Lasso usually shrinks a lot of entries in  $\hat{\beta}_{Lasso}$  and allows for a solution even in high-dimensional setting ( $p > n$ ).

When thinking about model selection in graphical models, one aims for graphs that only have a limited number of edges to make them meaningful connections. Obviously, this comes with the assumption that the true underlying structure is sparse. Moreover, high-dimensional regimes might occur. This motivates applying a Lasso-type method for graphical continuous Lyapunov models. Consider the vectorized Lyapunov equation

$$A(\Sigma)\text{vec}(M) = -\text{vec}(C).$$

Based on data, we can calculate the sample covariance matrix  $\hat{\Sigma}$  and obtain an estimated version of  $A(\Sigma)$  that takes on the role of the design matrix  $X$ . We assume the matrix  $C$  to be known and therefore  $-\text{vec}(C)$  coincides with the response vector  $y$  in the regression setup. We want to obtain an estimate for the drift matrix  $M$  which is why  $\text{vec}(M)$  coincides with the vector  $\beta$  in the classical regression setup. This leads to the optimization problem

$$\arg \min_{M \in \mathbb{R}^{p \times p}} \frac{1}{2} \|A(\Sigma)\text{vec}(M) + \text{vec}(C)\|_2^2 + \lambda \|M\|_1, \quad (3.2)$$

which can be computed analogously to the classical Lasso regression. At first glance, one might wonder why it is interesting to analyze this problem which seems to be just another Lasso problem. From a probabilistic perspective, there is quite a difference. In the classical regression setup (3.1), additive noise is considered whereas the uncertainty is contained in the estimation of the matrix  $A(\Sigma)$  in Lyapunov models. This makes the probabilistic analysis very different. Moreover, the matrix  $A(\Sigma)$  is of fixed size ( $p^2 \times p^2$ ) and does not contain the observations as rows, but has a predetermined structure with covariances as entries. As we show in this chapter, this has further implications on the theoretical results. At the same time, the computational results are also influenced by the different problem structure. Overall, the problem is quite different with the pleasant aspect that the methods to compute estimates are analogous to those for Lasso regression.

### 3.1.1 The Role of Parameter Identifiability for Estimation

In Chapter 2, we investigate parameter identifiability for graphical continuous Lyapunov models. This is a fundamental theoretical question when thinking about estimation. To obtain unique estimates, we have to be able to uniquely recover the entries in  $M$  (and possibly  $C$ ) when provided the covariance matrix  $\Sigma$  and when fixing the support of  $M$  (and possibly  $C$ ). Otherwise, even if we select the correct tuning parameter  $\lambda$  in (3.2), there would be no possibility of obtaining a unique estimate. We briefly recall the main aspects of the previous chapter.

It is evident that the equilibrium distribution of the observations does not uniquely determine the pair of drift and volatility matrices  $(M, C)$  as we have discussed in Remark 1.3.4. In particular, if  $(M, C)$  solves the Lyapunov equation, so does any scalar multiple of  $(M, C)$ . As scaling does not change the support of the drift matrix  $M$ , we solve this problem by assuming that  $C$  is fully known. Even after reducing to the case of a known volatility matrix  $C$ , the drift matrix  $M$  is not identifiable without exploiting further structure, such as sparsity. Indeed, the Lyapunov equation is a symmetric matrix equation with  $(p + 1)p/2$  individual equations, whereas  $M$  contains  $p^2$  unknown parameters. However,  $M$  becomes identifiable when it is known to be suitably sparse. In particular, we show that all simple cyclic graphs are globally identifiable. Many graphs with two-cycles permit almost sure unique recovery when the sparse entries of the drift matrix are selected randomly according to a continuous distribution, although here we cannot yet offer a concise sufficient condition.

### 3.1.2 Support Recovery for the Direct Lyapunov Lasso - Motivation

We study an  $\ell_1$ -regularization method for estimating the support of the drift matrix  $M$  from an i.i.d. sample consisting of centered observations  $X_1, \dots, X_n \in \mathbb{R}^p$ . Let

$$\hat{\Sigma} = \hat{\Sigma}^{(n)} = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top \quad (3.3)$$

be the sample covariance matrix. The *direct Lyapunov lasso* finds a sparse estimate of  $M$  as a solution of the convex optimization problem

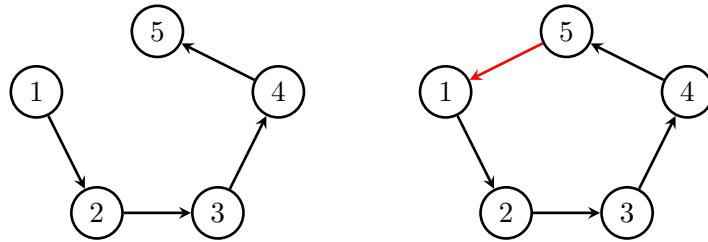
$$\min_{M \in \mathbb{R}^{p \times p}} \frac{1}{2} \|M \hat{\Sigma} + \hat{\Sigma} M^\top + C\|_F^2 + \lambda \|M\|_1 \quad (3.4)$$

with tuning parameter  $\lambda > 0$ . This method is considered in numerical experiments by Fitch [2019] as well as by Varando and Hansen [2020] who additionally explore non-convex methods based on regularizing Gaussian likelihood or a Frobenius loss. The direct Lyapunov lasso yields matrices in  $\mathbb{R}^{p \times p}$  that can be non-stable. If a stable estimate is required in such a case, one can appeal to projection onto the set of stable matrices; e.g., using techniques by Noferini and Poloni [2021].

Before developing a detailed analysis of the direct Lyapunov lasso, we present an example that illustrates the behavior of estimates for growing sample size and highlights the impact of the irrepresentability condition.

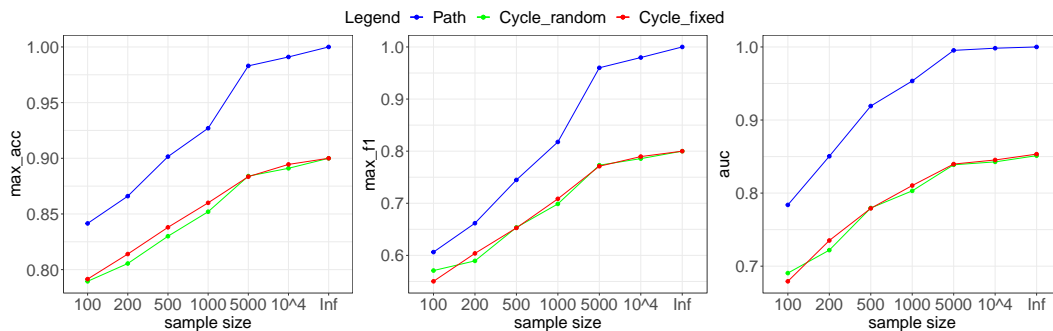
**Example 3.1.1.** *Let  $G_1$  be the path from 1 to 5, and let  $G_2$  be the 5-cycle obtained by adding the edge  $5 \rightarrow 1$ ; see Figure 3.1. For  $G_1$  we define a (well-conditioned)*

stable matrix  $M_1^*$  by setting the diagonal to  $(-2, -3, -4, -5, -6)$  and the four nonzero subdiagonal entries to 0.65. For  $G_2$ , we consider two cases. In the first case, we add the entry  $m_{15} = 0.65$  to  $M_1^*$  to obtain the matrix  $M_2^*$ . We then draw 100 samples of size  $n = 100, 200, 500, 1000, 5000, 10^4, 10^5, \infty$  from  $N(0, \Sigma_j^*)$  for  $j = 1, 2$ , where  $\Sigma_j^*$  is the covariance matrix obtained from  $M_j^*$ . When  $n = \infty$ , the population covariance matrices are taken as input to the method. In the second case, we generate 100 stable matrices  $M_{2,1}^*, \dots, M_{2,100}^*$  from  $M_1^*$  by selecting 100 entries  $m_{15}$  according to a uniform distribution on  $[0.5, 1]$ . Let  $\Sigma_{2,1}^*, \dots, \Sigma_{2,100}^*$  be the corresponding equilibrium covariance matrices. In the second case, we generate one sample from  $N(0, \Sigma_j^*)$ ,  $j = 1, (2, 1), \dots, (2, 100)$  for each of the sample sizes given above. direct Lyapunov lasso is used for support recovery, with the penalty parameter  $\lambda$  chosen on a grid  $\lambda_1 = \lambda_{\max}/10^4, \dots, \lambda_{100} = \lambda_{\max}$  that is equidistant on the log-scale. The value  $\lambda_{\max}$  is the minimal  $\lambda$ -value such that the estimate is diagonal. To implement the direct Lyapunov lasso, we use the R package `glmnet`, which runs a coordinate descent algorithm for fitting the Lasso, see Friedman et al. [2010]. For each data set, we calculate the maximum accuracy, the maximum  $F_1$ -score and the area under the ROC curve. We give the details for the metrics in Definition 3.5.11. Figure 3.2 plots the performance measures, averaged over the 100 datasets in each pairing of setup and sample size. There the blue curves refer to  $G_1$ , the red curves to  $G_2$  with  $m_{15} = 0.65$  fixed, and the green curves to  $G_2$  with  $m_{15}$  chosen randomly. We observe that for every sample size and performance measure, the direct Lyapunov lasso performs better for the path  $G_1$  than for the cycle  $G_2$ . When the sample size is  $n = 10^4$ , we observe an almost perfect recovery of  $G_1$ . However, increasing the sample size when recovering  $G_2$  does not result in perfect recovery. The choice of  $m_{15} = 0.65$  is not particularly unfortunate—averaging over various completions does not improve the metrics. We conclude that while learning useful structure in either case, the direct Lyapunov lasso is consistent only for the considered path. Our subsequent analysis explains this behavior, which is a consequence of the failure of the irrepresentability condition in (3.15).



**Figure 3.1:** Left: The graph  $G_1$ , a path 1 to 5. Right: The graph  $G_2$ , the 5-cycle.





**Figure 3.2:** Performance measures for sample sizes  $n = 10^1, \dots, 10^4, \infty$  and models given by the graphs  $G_1$  and  $G_2$  from Figure 3.1, one choice of edgeweights for the path, one choice of edgeweights for the cycle (Cycle fixed) and 100 random completions to the 5-cycle (Cycle random). Left: maximal accuracy, Middle: maximal  $F_1$ -score, Right: area under the ROC curve.

### 3.1.3 Organization of the Chapter

We first connect the direct Lyapunov lasso to more standard lasso problems by vectorizing the Lyapunov equation and describing the structure of the Hessian matrix for the smooth part of the direct Lyapunov lasso objective (Section 3.2). In Section 3.3, we derive a deterministic guarantee for support recovery based on the primal-dual witness approach (Theorem 3.3.1). We then extend this guarantee to a statistical consistency result (Corollary 3.3.3), where the solution  $\hat{M}$  of (3.4) is shown to converge in the max norm at a rate  $\|\hat{M} - M^*\|_\infty = O(\sqrt{dp/n})$  with  $d$  being the number of nonzero entries in the true drift matrix  $M^*$ . The necessary probabilistic analysis is based on the concentration results for the spectral norm of the sample covariance matrix, from which we are able to deduce the concentration results for the direct Lyapunov lasso Hessian (Section 3.4). The consistency result depends on an irrepresentability condition, which turns out to be more subtle than in the classical lasso regression. As we explore in Section 3.5, the condition is highly dependent on the structure of the graph associated with the true signal and appears to be particularly restrictive for graphs with directed cycles. At least for DAGs (directed acyclic graphs) we are able to always construct matrices at which the condition holds. In Section 3.6 we show that the direct Lyapunov lasso is somewhat robust to both misspecification of the volatility matrix  $C$  and the irrepresentability condition being not fulfilled. Towards the end of the chapter, we apply the direct Lyapunov lasso to real world data. In Section 3.7, we first analyze two Bayesian information criteria on synthetic data and then proceed to present estimates of a protein-signalling network purely from observational data. Despite the simplicity of the approach, some estimates recover most of the important connections in the network. We conclude the chapter with a discussion in Section 3.8.

### 3.1.4 Notation - Chapter 3 and Chapter 4

Let  $b \in [1, \infty]$ . The  $\ell_b$ -norm of  $v \in \mathbb{R}^p$  is  $\|v\|_b = (\sum_{i=1}^p |v_i|^b)^{1/b}$ , with  $\|v\|_\infty = \max_{1 \leq i \leq n} |v_i|$ . We may apply this vector norm to a matrix  $A = (a_{ij}) \in \mathbb{R}^{p \times p}$  and obtain the norm  $\|A\|_b = (\sum_{i=1}^p \sum_{j=1}^n |a_{ij}|^b)^{1/b}$ . In particular,  $\|A\|_F := \|A\|_2$  is the Frobenius norm. We denote the associated operator norm by  $\|A\|_b = \max\{\|Ax\|_b : \|x\|_b = 1\}$ .

Specifically, we use  $\|A\|_2$  to denote the spectral norm, given by the maximal singular value of  $A$ , and  $\|A\|_\infty = \max_{1 \leq i \leq p} \sum_{j=1}^p |a_{ij}|$  to denote the maximum absolute row sum.

### 3.2 Gram Matrix of the Direct Lyapunov Lasso

In this section, we rewrite the smooth part of the objective of the direct Lyapunov lasso from (3.4) in terms of the vectorized drift matrix and present the resulting Hessian matrix.

The Lyapunov equation from (1.2) is a linear matrix equation and may be rewritten as

$$A(\Sigma) \text{vec}(M) + \text{vec}(C) = 0, \quad (3.5)$$

where the  $p^2 \times p^2$  matrix  $A(\Sigma)$  has its rows and columns indexed by pairs  $(i, j) \in \{1, \dots, p\}^2$  and takes the form

$$A(\Sigma) = (\Sigma \otimes I_p) + (I_p \otimes \Sigma)K^{(p,p)}. \quad (3.6)$$

We have  $\text{vec}(M\Sigma) = (\Sigma \otimes I_p)\text{vec}(M)$  and  $\text{vec}(\Sigma M^\top) = (I_p \otimes \Sigma)K^{(p,p)}\text{vec}(M)$ . By the symmetry of the Lyapunov equation,  $A(\Sigma)$  has two copies of each row corresponding to an off-diagonal entry in the Lyapunov equation. Retaining this redundancy will be helpful for later arguments, as it preserves the Kronecker product structure in (3.6).

**Example 3.2.1.** When  $p = 3$ , the matrix  $A(\Sigma)$  is a  $9 \times 9$  matrix and has the form

$$\begin{array}{c} \begin{matrix} (1, 1) & (1, 2) & (1, 3) & (2, 1) & (2, 2) & (2, 3) & (3, 1) & (3, 2) & (3, 3) \end{matrix} \\ \begin{pmatrix} (1, 1) & 2\Sigma_{11} & 0 & 0 & 2\Sigma_{12} & 0 & 0 & 2\Sigma_{13} & 0 & 0 \\ (1, 2) & \Sigma_{21} & \Sigma_{11} & 0 & \Sigma_{22} & \Sigma_{12} & 0 & \Sigma_{23} & \Sigma_{13} & 0 \\ (1, 3) & \Sigma_{31} & 0 & \Sigma_{11} & \Sigma_{23} & 0 & \Sigma_{12} & \Sigma_{33} & 0 & \Sigma_{13} \\ (2, 1) & \Sigma_{21} & \Sigma_{11} & 0 & \Sigma_{22} & \Sigma_{12} & 0 & \Sigma_{23} & \Sigma_{13} & 0 \\ (2, 2) & 0 & 2\Sigma_{21} & 0 & 0 & 2\Sigma_{22} & 0 & 0 & 2\Sigma_{23} & 0 \\ (2, 3) & 0 & \Sigma_{31} & \Sigma_{21} & 0 & \Sigma_{23} & \Sigma_{22} & 0 & \Sigma_{33} & \Sigma_{23} \\ (3, 1) & \Sigma_{31} & 0 & \Sigma_{11} & \Sigma_{23} & 0 & \Sigma_{12} & \Sigma_{33} & 0 & \Sigma_{13} \\ (3, 2) & 0 & \Sigma_{31} & \Sigma_{21} & 0 & \Sigma_{23} & \Sigma_{22} & 0 & \Sigma_{33} & \Sigma_{23} \\ (3, 3) & 0 & 0 & 2\Sigma_{31} & 0 & 0 & 2\Sigma_{23} & 0 & 0 & 2\Sigma_{33} \end{pmatrix} \end{array}.$$

Rows with an italicized index correspond to strictly upper triangular entries in the Lyapunov equation from (1.2).

Define the Gram matrix

$$\Gamma(\Sigma) := A(\Sigma)^\top A(\Sigma) \in \mathbb{R}^{p^2 \times p^2} \quad (3.7)$$

and the vector

$$g(\Sigma) := -A(\Sigma)\text{vec}(C) \in \mathbb{R}^{p^2}. \quad (3.8)$$

Omitting a constant from the objective function, the direct Lyapunov lasso problem from (3.4) may be reformulated as

$$\min_{M \in \mathbb{R}^{p \times p}} \frac{1}{2} \text{vec}(M)^\top \Gamma(\hat{\Sigma}) \text{vec}(M) - g(\hat{\Sigma})^\top \text{vec}(M) + \lambda \|\text{vec}(M)\|_1. \quad (3.9)$$

### 3.3 Consistent Support Recovery with the Direct Lyapunov Lasso

As noted in the introduction, one difficulty that arises in the analysis of the solution of (3.9) is the fact that the Gram matrix has entries that are quadratic polynomials in  $\Sigma$  with  $p$  terms (i.e., the number of terms scales with the size of the problem). This fact can be seen in the appearance of  $\Sigma^2$  in the following formula for the Gram matrix.

**Lemma 3.2.2.** *The Gram matrix for a given covariance matrix  $\Sigma$  is equal to*

$$\Gamma(\Sigma) = A(\Sigma)^\top A(\Sigma) = 2(\Sigma^2 \otimes I_p) + (\Sigma \otimes \Sigma)K^{(p,p)} + K^{(p,p)}(\Sigma \otimes \Sigma).$$

**Proof.** Apply the rules  $(A \otimes B)^\top = (A^\top \otimes B^\top)$ ,  $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$  and  $K^{(p,p)}(A \otimes B)K^{(p,p)} = B \otimes A$  to deduce that

$$\begin{aligned} A(\Sigma)^\top A(\Sigma) &= [(\Sigma \otimes I_p) + K^{(p,p)}(I_p \otimes \Sigma)][(\Sigma \otimes I_p) + (I_p \otimes \Sigma)K^{(p,p)}] \\ &= 2(\Sigma^2 \otimes I_p) + (\Sigma \otimes \Sigma)K^{(p,p)} + K^{(p,p)}(\Sigma \otimes \Sigma). \end{aligned}$$

□

### 3.3 Consistent Support Recovery with the Direct Lyapunov Lasso

In this section, we now provide a probabilistic guarantee that the direct Lyapunov lasso is able to recover the support of the true population drift matrix that defines the data-generating distribution. This result is based on a slightly adapted version of [Lin et al., 2016, Theorem 1] that we also present in this section.

We start by introducing some more notation. The matrix  $M^*$  denotes the true drift matrix in (1.1) and  $\hat{M}$  denotes the solution of the direct Lyapunov lasso problem in (3.4). The support of  $M^*$  is the set of all indices of nonzero elements and is denoted by

$$S \equiv S(M^*) = \{(j, k) : M_{jk}^* \neq 0\}.$$

We write  $d = |S|$  for the size of the support of  $M^*$ . The support set of the estimate  $\hat{M}$  is written

$$\hat{S} \equiv S(\hat{M}) = \{(j, k) : \hat{M}_{jk} \neq 0\}.$$

Let  $\hat{\Gamma} = \Gamma(\hat{\Sigma})$ ,  $\Gamma^* = \Gamma(\Sigma^*)$ ,  $\hat{g} = g(\hat{\Sigma})$ ,  $g^* = g(\Sigma^*)$ . Furthermore, let  $\Delta_\Gamma = \hat{\Gamma} - \Gamma^*$  and  $\Delta_g = \hat{g} - g^*$ , and define the quantities

$$c_{\Gamma^*} = \|(\Gamma_{SS}^*)^{-1}\|_\infty \text{ and } c_{M^*} = \|\text{vec}(M^*)\|_\infty.$$

The definition of  $c_{\Gamma^*}$  requires  $\Gamma_{SS}^*$  to be invertible, which is an implicit assumption on the identifiability of the parameters; recall Section 3.1.1.

We provide the deterministic result that Corollary 3.3.3 is based on. We adapt Theorem 1 by Lin et al. [2016] to arrive at our deterministic result. This requires resolving only a few differences, as we describe in Remark 3.3.2. The underlying construction for the proof is the Primal-Dual-Witness (PDW) method [Wainwright, 2009].

**Theorem 3.3.1.** *Let  $M^* \in \text{Stab}_p$  be the true drift matrix, and let  $S$  be its support. Assume that  $\Gamma_{SS}^*$  is invertible and that the irrepresentability condition*

$$\|\Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}\|_\infty < 1 - \alpha \quad (3.10)$$

*holds with parameter  $\alpha \in (0, 1]$ . Furthermore, assume that  $\hat{\Gamma}$  is a matrix such that*

$$\|(\Delta_\Gamma)_{\cdot S}\|_\infty < \epsilon_1, \quad \|\Delta_g\|_\infty < \epsilon_2,$$

*with  $\epsilon_1 \leq \alpha/(6c_{\Gamma^*})$ . If*

$$\lambda > \frac{3(2 - \alpha)}{\alpha} \max\{c_{M^*}, \epsilon_1, \epsilon_2\},$$

*then the following statements hold:*

- a) *The LSGE  $\hat{M}$  is unique, has its support included in the true support ( $\hat{S} \subseteq S$ ), and satisfies*

$$\|\hat{M} - M^*\|_\infty < \frac{2c_{\Gamma^*}}{2 - \alpha} \lambda.$$

- b) *If*

$$\min_{\substack{1 \leq j < k \leq m \\ (j,k) \in S}} |M_{jk}^*| > \frac{2c_{\Gamma^*}}{2 - \alpha} \lambda,$$

*then  $\hat{S} = S$  and  $\text{sign}(\hat{M}_{jk}) = \text{sign}(M_{jk}^*)$  for all  $(j, k) \in S$ .*

**Proof.** The proof is very similar to the proof of Theorem 1 by Lin et al. [2016]. However, there are a few subtle differences and missing explanations that we add in this proof. For all the calculations that are already carried out by Lin et al. [2016], we refer to the original manuscript for these passages.

We use the PDW technique to prove the result. The estimate  $\hat{M}$  satisfies the KKT conditions

$$\hat{\Gamma} \text{vec}(\hat{M}) - \hat{g} + \lambda \hat{z} = 0, \quad (3.11)$$

where  $\hat{z} \in \partial \|\text{vec}(\hat{M})\|_1$  is an element of the subdifferential of the  $\ell_1$ -norm, that is, the elements of the vector  $\hat{z} \in \mathbb{R}^{p^2}$  satisfy that elements

$$\hat{z}_{(i,j)} = \begin{cases} \text{sign}(\text{vec}(\hat{M})_{(i,j)}) & \text{if } \text{vec}(\hat{M})_{(i,j)} \neq 0, \\ \in [-1, 1] & \text{if } \text{vec}(\hat{M})_{(i,j)} = 0. \end{cases}$$

Here, we index  $\hat{z}$  by pairs  $(i, j)$  with  $1 \leq i, j \leq p$ . The optimization problem in (3.9) is convex as  $\Gamma$  is positive semidefinite by construction, and the KKT conditions are necessary and sufficient for a solution to be optimal for the problem. The PDW technique constructs, in three steps, a primal-dual pair  $(\hat{M}, \hat{z})$  that satisfies (3.11) and has the support of  $\hat{M}$  contained in  $S$ .

Since the true signal  $M^* \in \text{Stab}_p$  and  $C \in PD_p$ , there exists a unique positive definite  $\Sigma^*$  determined by the continuous Lyapunov equation in (1.2). As a result

$$\Gamma^* \text{vec}(M^*) - g^* = 0,$$

### 3.3 Consistent Support Recovery with the Direct Lyapunov Lasso

and we can rewrite the KKT conditions in (3.11) in the following block form

$$\begin{bmatrix} \Gamma_{SS}^* & \Gamma_{SS^c}^* \\ \Gamma_{S^cS}^* & \Gamma_{S^cS^c}^* \end{bmatrix} \begin{bmatrix} (\Delta_M)_S \\ (\Delta_M)_{S^c} \end{bmatrix} + \begin{bmatrix} (\Delta_\Gamma)_{SS} & (\Delta_\Gamma)_{SS^c} \\ (\Delta_\Gamma)_{S^cS} & (\Delta_\Gamma)_{S^cS^c} \end{bmatrix} \begin{bmatrix} \text{vec}(\hat{M})_S \\ \text{vec}(\hat{M})_{S^c} \end{bmatrix} + \begin{bmatrix} (\Delta_g)_S \\ (\Delta_g)_{S^c} \end{bmatrix} + \lambda \begin{bmatrix} \hat{z}_S \\ \hat{z}_{S^c} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where  $\Delta_M = \text{vec}(\hat{M}) - \text{vec}(M^*)$ . We now construct a pair  $(\hat{M}, \hat{z})$  that satisfies the equation.

**Step 1.** We solve the restricted optimization problem

$$\text{vec}(\tilde{M}) = \arg \min_{\text{vec}(M)_{S^c}=0} \frac{1}{2} \text{vec}(M)^\top \hat{\Gamma} \text{vec}(M) - \hat{g}^\top \text{vec}(M) + \lambda \|\text{vec}(M)\|_1. \quad (3.12)$$

Since  $\Gamma_{SS}^*$  is invertible, under our assumptions,  $\hat{\Gamma}_{SS}$  is also invertible. The matrix  $\hat{\Gamma}_{SS}$  can be expressed as

$$\hat{\Gamma}_{SS} = \Gamma_{SS}^* + (\hat{\Gamma}_{SS} - \Gamma_{SS}^*) = \Gamma_{SS}^* + (\Delta_\Gamma)_{SS}$$

Factoring out  $\Gamma_{SS}^*$ , we obtain

$$\hat{\Gamma}_{SS} = \Gamma_{SS}^* (I_{|S|} + (\Gamma_{SS}^*)^{-1} (\Delta_\Gamma)_{SS})$$

where  $I_{|S|}$  denotes the identity matrix of size  $|S| \times |S|$ . Then, the matrix  $\hat{\Gamma}_{SS}$  is invertible if

$$\rho((\Gamma_{SS}^*)^{-1} (\Delta_\Gamma)_{SS}) < 1.$$

This is true as the spectral norm is bounded by the maximum absolute row sum norm and

$$\|(\Gamma_{SS}^*)^{-1} (\Delta_\Gamma)_{SS}\|_\infty \leq \|(\Gamma_{SS}^*)^{-1}\|_\infty \|(\Delta_\Gamma)_{SS}\|_\infty < 1$$

with the second inequality being true because of  $\|(\Delta_\Gamma)_{SS}\|_\infty < \epsilon_1 < \alpha/6c_{\Gamma^*} < 1/c_{\Gamma^*}$  and  $c_{\Gamma^*} = \|(\Gamma_{SS}^*)^{-1}\|_\infty$ .

Therefore, the solution  $\text{vec}(\tilde{M})$  is unique. Furthermore, we have

$$(\text{vec}(\tilde{M}))_S = (\hat{\Gamma}_{SS})^{-1} (\hat{g}_S - \lambda \text{sign}((\text{vec}(\tilde{M}))_S)).$$

Let  $\tilde{\Delta}_M = \text{vec}(\tilde{M}) - \text{vec}(M^*)$ . Following the proof of Theorem 1 by Lin et al. [2016], we have

$$\|\tilde{\Delta}_M\|_\infty \leq \frac{c_{\Gamma^*}}{1 - \alpha/6} \cdot \frac{6 - \alpha}{3(2 - \alpha)} \lambda = \frac{2c_{\Gamma^*}}{2 - \alpha} \lambda. \quad (3.13)$$

**Step 2.** Let  $\tilde{z}_S = \text{sign}(\text{vec}(\tilde{M})_S)$ . Then  $\tilde{z}_S \in \partial \|\text{vec}(\tilde{M})\|_1$ .

**Step 3.** Let

$$\begin{aligned} \tilde{z}_{S^c} = \frac{1}{\lambda} \left[ -\Gamma_{S^cS}^* (\Gamma_{SS}^*)^{-1} ((\Delta_\Gamma)_{SS} \text{vec}(\tilde{M})_S + (\Delta_g)_S) + (\Delta_\Gamma)_{S^cS} \text{vec}(\tilde{M})_S \right. \\ \left. + (\Delta_g)_{S^c} + \lambda \Gamma_{S^cS}^* (\Gamma_{SS}^*)^{-1} \text{sign}(\text{vec}(\tilde{M})_S) \right]. \end{aligned} \quad (3.14)$$

We show that  $\|\tilde{z}_{S^c}\|_1 < 1$ , which is a dual feasibility condition. Once this is shown, we have that the pair  $(\text{vec}(\tilde{M}), \tilde{z})$  satisfies (3.11) by construction, and  $(\text{vec}(\hat{M}), \hat{z}) = (\text{vec}(\tilde{M}), \tilde{z})$  is the solution to the optimization problem in (3.9). Furthermore, Lemma 1 of Wainwright [2009] implies that the strict dual feasibility implies that  $\hat{S} \subseteq S$ . Following Theorem 1 by Lin et al. [2016], we have

$$\begin{aligned} \|\tilde{z}_{S^c}\|_\infty &\leq \underbrace{\frac{2-\alpha}{\lambda} \|(\Delta_\Gamma)_{\cdot S} \text{vec}(M^*)\|_\infty}_{G_1} + \underbrace{\frac{2-\alpha}{\lambda} \|(\Delta_\Gamma)_{\cdot S}\|_\infty \|\Delta_S\|_\infty}_{G_2} \\ &\quad + \underbrace{\frac{2-\alpha}{\lambda} \|\Delta_g\|_\infty}_{G_3} + (1-\alpha). \end{aligned}$$

For  $G_1$ , we have that

$$G_1 \leq \frac{2-\alpha}{\lambda} \|(\Delta_\Gamma)_{\cdot S}\|_\infty \|\text{vec}(M^*)\|_\infty = \frac{2-\alpha}{\lambda} c_{M^*} \epsilon_1 \leq \frac{\alpha}{3}.$$

For  $G_3$ , we have that

$$G_3 = \frac{2-\alpha}{\lambda} \|\Delta_g\|_\infty < \frac{2-\alpha}{\lambda} \epsilon_2 \leq \frac{\alpha}{3}.$$

Finally, for  $G_2$ , we have that

$$G_2 < \frac{2-\alpha}{\lambda} \cdot \frac{\alpha}{6c_{\Gamma^*}} \cdot \epsilon_1 \cdot \frac{c_{\Gamma^*}}{1-\alpha/6} \cdot \frac{6-\alpha}{3(2-\alpha)} < \frac{\alpha}{3}.$$

Combining these bounds, we have that  $\|\tilde{z}_{S^c}\|_\infty < 1$ , which establishes the strict dual feasibility.

Finally, for any  $(j, k) \in S$ , we have that

$$|\hat{M}_{jk}| \geq |M_{jk}^*| - |\hat{M}_{jk} - M_{jk}^*| > \min_{\substack{1 \leq j < k \leq p \\ (j,k) \in S}} |M_{jk}^*| - \|\text{vec}(\hat{M}) - \text{vec}(M^*)\|_\infty > 0,$$

which shows that  $\hat{S} = S$ . □

**Remark 3.3.2.** *The distinction between Theorem 3.3.1 and Theorem 1 of Lin et al. [2016] lies in the steps of our analysis that involve the maximal absolute row sum norm in the bound for the difference between the estimated Hessian  $\hat{\Gamma}$  and the true Hessian  $\Gamma^*$ . In Lin et al. [2016], the bound was based on the maximal entry. This difference requires adjustments in certain steps of the proof. Consequently, our above proof also provides a more detailed explanation of certain arguments that were omitted by Lin et al. [2016], but are of greater significance in our work. For example, we address the issue of invertibility of  $\hat{\Gamma}$ . We also indicate when parts of the proof by Lin et al. [2016] are unaffected to ensure clarity and consistency.*

Theorem 3.3.1 provides a deterministic result for the estimation error and support recovery under a general bound on  $\Delta_\Gamma$  and  $\Delta_g$ . It leads to the following probabilistic result.

### 3.3 Consistent Support Recovery with the Direct Lyapunov Lasso

**Corollary 3.3.3.** *Suppose that the data are generated as  $n$  i.i.d. draws from the Gaussian equilibrium distribution of a  $p$ -dimensional Ornstein-Uhlenbeck process defined by a drift matrix  $M^* \in \text{Stab}_p$  and a matrix  $C \in \text{PD}_p$ . Let  $S$  be the support of  $M^*$ . Assume that  $\Gamma_{SS}^*$  is invertible and that the irrepresentability condition*

$$\|\Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}\|_\infty < 1 - \alpha \quad (3.15)$$

holds for  $\alpha \in (0, 1]$ . Let  $c_{\Sigma^*} = \|\Sigma^*\|_2$ ,  $c_C = \|\text{vec}(C)\|_2$ ,

$$\begin{aligned} \tilde{c} = \max & \left\{ \frac{4 \max\{1, c_{\Sigma^*}^2\} (4 + 8c_{\Sigma^*})^2}{c_3}, 16c_1^2 c_{\Sigma^*}^2 (4 + 8c_{\Sigma^*})^2, \right. \\ & \left. \frac{16 \max\{1, c_{\Sigma^*}^2\} c_C^2}{c_3}, 64c_1^2 c_{\Sigma^*}^2 c_C^2 \right\}, \\ c_* = & \frac{6}{\alpha} c_{\Gamma^*}, \end{aligned}$$

where  $\{c_i\}_{i=1}^3$  are universal constants (from Theorem 3.4.4 below) with  $c_1 > \max\{1, \|\Sigma^*\|_2\}$ . Suppose the sample size satisfies  $n > \tau_1 \tilde{c} dp \max\{c_*^2, 1/4\}$  for  $\tau_1 > 1$ , and the regularization parameter is chosen as

$$\lambda > \frac{3c_{M^*}(2 - \alpha)}{\alpha} \sqrt{\frac{\tau_1 \tilde{c} dp}{n}}.$$

Then the following statements hold with probability at least  $1 - c_2 \exp(-\tau_1 p)$ :

- a) The minimizer  $\hat{M}$  is unique, has its support included in the true support ( $\hat{S} \subset S$ ), and satisfies

$$\|\hat{M} - M^*\|_\infty < \frac{2c_{\Gamma^*}}{2 - \alpha} \lambda.$$

- b) Furthermore, if

$$\min_{\substack{1 \leq j < k \leq m \\ (j,k) \in S}} |M_{jk}^*| > \frac{2c_{\Gamma^*}}{2 - \alpha} \lambda,$$

then  $\hat{S} = S$  and  $\text{sign}(\hat{M}_{jk}) = \text{sign}(M_{jk}^*)$  for all  $(j, k) \in S$ .

The Corollary follows from Theorem 3.3.1 together with the concentration results we obtain in Section 3.4. We defer the proof of Corollary 3.3.3 to the end of Section 3.4.

The reader may be surprised by the sample size requirement of  $n = \Omega(dp)$ ; recall that  $d = |S|$  is the size of the support of  $M^*$ . Since  $S$  includes the diagonal of  $M^*$ , we have  $d \geq p$ . Under sparsity, however,  $dp$  is not much larger than the number of unknown parameters  $p^2$ .

This said,  $\Omega(dp)$  is far larger than the sample size requirement a reader may be familiar with from the glasso for learning undirected conditional independence, which is on the order of  $d^2 \log p$  but with  $d$  being the maximum number of nonzero entries in any row of a true precision matrix [Ravikumar et al., 2011]. This allows for far higher-dimensional settings, but crucially relies on the glasso having a Hessian that concentrates well entry-wise and a simple connection between the covariance matrix and the sparse precision matrix. In contrast, the Lyapunov Lasso has a denser Hessian/Gram matrix that includes entries that become heavier-tailed as the dimension  $p$  grows.

### 3.4 Probabilistic Analysis

The direct Lyapunov lasso depends on the loss being sufficiently close to its population version in the sense of  $\Delta_\Gamma = \hat{\Gamma} - \Gamma^*$  and  $\Delta_g = \hat{g} - g^*$  being sufficiently small. In this section, we bound  $\Delta_\Gamma$  and  $\Delta_g$  in terms of  $\Delta_\Sigma = \hat{\Sigma} - \Sigma^*$  and, subsequently, use a concentration inequality for  $\|\Delta_\Sigma\|_2$  to probabilistically bound  $\Delta_\Gamma$  and  $\Delta_g$ .

Deriving an inequality for  $\hat{\Gamma}$  is most critical as the matrix contains sums of products of covariances and a careful analysis is required to obtain a non-trivial requirement on the sample size. Let  $\Gamma(\Sigma) = \Gamma_1(\Sigma) + \Gamma_2(\Sigma)$ , where

$$\Gamma_1(\Sigma) = 2(\Sigma^2 \otimes I_p) \quad \text{and} \quad \Gamma_2(\Sigma) = (\Sigma \otimes \Sigma)K^{(p,p)} + K^{(p,p)}(\Sigma \otimes \Sigma).$$

**Lemma 3.4.1.** *Let  $c_{\Sigma^*} = \|\Sigma^*\|_2$ . Then*

$$\|\Gamma_1(\hat{\Sigma}) - \Gamma_1(\Sigma^*)\|_2 \leq 2\|\Delta_\Sigma\|_2^2 + 4c_{\Sigma^*}\|\Delta_\Sigma\|_2.$$

**Proof.** Using that  $\|A \otimes B\|_2 = \|A\|_2\|B\|_2$ , we obtain that

$$\begin{aligned} \|\Gamma_1(\hat{\Sigma}) - \Gamma_1(\Sigma^*)\|_2 &= 2\|(\hat{\Sigma}^2 - (\Sigma^*)^2) \otimes I_p\|_2 \\ &= 2\|\hat{\Sigma}^2 - (\Sigma^*)^2\|_2 \\ &\leq 2\|\Delta_\Sigma\|_2^2 + 2\|\Delta_\Sigma\Sigma^*\|_2 + 2\|\Sigma^*\Delta_\Sigma\|_2. \end{aligned}$$

Since the spectral norm of a symmetric matrix is the absolute maximal eigenvalue, and the eigenvalues of a squared matrix are the squared eigenvalues of the original matrix, we find as claimed that

$$\|\Gamma_1(\hat{\Sigma}) - \Gamma_1(\Sigma^*)\|_2 \leq 2\|\Delta_\Sigma\|_2^2 + 4\|\Sigma^*\|_2\|\Delta_\Sigma\|_2. \quad \square$$

**Lemma 3.4.2.** *Let  $c_{\Sigma^*} = \|\Sigma^*\|_2$ . Then*

$$\|\Gamma_2(\hat{\Sigma}) - \Gamma_2(\Sigma^*)\|_2 \leq 2\|\Delta_\Sigma\|_2^2 + 4c_{\Sigma^*}\|\Delta_\Sigma\|_2.$$

**Proof.** The commutation matrix  $K^{(p,p)}$  is an orthonormal matrix. Therefore,  $\|K^{(p,p)}\|_2 = 1$  and

$$\|K^{(p,p)}(\hat{\Sigma} \otimes \hat{\Sigma} - \Sigma^* \otimes \Sigma^*)\|_2 = \|(\hat{\Sigma} \otimes \hat{\Sigma} - \Sigma^* \otimes \Sigma^*)K^{(p,p)}\|_2 = \|\hat{\Sigma} \otimes \hat{\Sigma} - \Sigma^* \otimes \Sigma^*\|_2.$$

We obtain that

$$\begin{aligned} \|\Gamma_2(\hat{\Sigma}) - \Gamma_2(\Sigma^*)\|_2 &\leq 2\|\hat{\Sigma} \otimes \hat{\Sigma} - \Sigma^* \otimes \Sigma^*\|_2 \\ &= 2\|\Delta_\Sigma \otimes \Delta_\Sigma + \Delta_\Sigma \otimes \Sigma^* + \Sigma^* \otimes \Delta_\Sigma + \Sigma^* \otimes \Sigma^* - \Sigma^* \otimes \Sigma^*\|_2 \\ &\leq 2\|\Delta_\Sigma \otimes \Delta_\Sigma\|_2 + 2\|\Delta_\Sigma \otimes \Sigma^*\|_2 + 2\|\Sigma^* \otimes \Delta_\Sigma\|_2 \\ &\leq 2\|\Delta_\Sigma\|_2^2 + 4\|\Sigma^*\|_2\|\Delta_\Sigma\|_2, \end{aligned}$$

which was the claim. □



For a matrix  $A \in \mathbb{R}^{p \times d}$ , it holds that  $\|A\|_\infty \leq \sqrt{d}\|A\|_2$ . Then it follows from Lemma 3.4.1 and Lemma 3.4.2 that

$$\|(\Delta_\Gamma)_{\cdot S}\|_\infty \leq \sqrt{d} (4\|\Delta_\Sigma\|_2^2 + 8c_{\Sigma^*}\|\Delta_\Sigma\|_2). \quad (3.16)$$

We note that bounding  $\|(\Delta_\Gamma)_{\cdot S}\|_\infty$  using  $\|(\Delta_\Gamma)_{\cdot S}\|_\infty$ , as was done in Lin et al. [2016], leads to a worse bound. While such an approach might seem simpler, it does not exploit the structure of the Hessian  $\Gamma$  in Lemma 3.2.2.

We now provide a bound on  $\|\Delta_g\|_\infty$ .

**Lemma 3.4.3.** *We have  $\|\Delta_g\|_\infty \leq 2c_C\|\Delta_\Sigma\|_2$ , where  $c_C = \|\text{vec}(C)\|_2$ .*

**Proof.** Similar to the proof of Lemma 3.4.1 and Lemma 3.4.2, we have

$$\begin{aligned} \|\Delta_g\|_\infty &\leq \|\Delta_g\|_2 \\ &\leq c_C \|\Sigma^* \otimes I_p - (I_p \otimes \Sigma^*)K^{(p,p)} - \hat{\Sigma} \otimes I_p + (I_p \otimes \hat{\Sigma})K^{(p,p)}\|_2 \\ &\leq c_C (\|I_p \otimes (\hat{\Sigma} - \Sigma^*)\|_2 + \|(\hat{\Sigma} - \Sigma^*) \otimes I_p\|_2) \quad (\text{since } \|K^{(p,p)}\|_2 = 1) \\ &= 2c_C \|\Delta_\Sigma\|_2. \end{aligned}$$

□

The bounds in (3.16) and Lemma 3.4.3 depend on the spectral norm of  $\Delta_\Sigma$ . We adapt Theorem 6.5 in Wainwright [2019] to our setting to upper bound  $\|\Delta_\Sigma\|_2$  under the assumption that  $(x_i)_{i=1}^n$  are sub-Gaussian.

**Theorem 3.4.4** (Theorem 6.5. in Wainwright [2019]). *Suppose that  $(X_i)_{i=1}^n$  are  $\sigma$  sub-Gaussian random variables. Then the sample covariance matrix  $\hat{\Sigma}$  in (3.3) satisfies*

$$\mathbb{P} \left( \frac{\|\hat{\Sigma} - \Sigma^*\|_2}{\sigma^2} \geq c_1 \left\{ \sqrt{\frac{p}{n}} + \frac{p}{n} \right\} + \delta \right) \leq c_2 \exp(-c_3 n \min\{\delta, \delta^2\}) \quad \forall \delta \geq 0,$$

where  $\{c_j\}_{j=0}^3$  are universal constants.

**Corollary 3.4.5.** *Let  $\{c_j\}_{j=1}^3$  be the universal constants from Theorem 3.4.4, but ensuring that  $c_1 > \max\{1, 1/\|\Sigma^*\|_2\}$ . Let  $(X_i)_{i=1}^n$  be Gaussian random variables. For any  $\epsilon \in (4c_1\|\Sigma^*\|_2\sqrt{p/n}, 2)$ , we have*

$$\mathbb{P} \left( \|\hat{\Sigma} - \Sigma^*\|_2 \geq \epsilon \right) \leq c_2 \exp \left( -\frac{c_3}{4 \max(1, \|\Sigma^*\|_2^2)} n \epsilon^2 \right).$$

**Proof.** A Gaussian random vector is sub-Gaussian with parameter  $\sigma = \|\Sigma^*\|_2$ .

Set  $\delta = \min \left( \frac{\epsilon}{2\|\Sigma^*\|_2}, \frac{\epsilon}{2} \right)$ . Since  $\frac{p}{n} < \frac{\epsilon^2}{16c_1^2\|\Sigma^*\|_2^2}$ , we have

$$\begin{aligned} \|\Sigma^*\|_2 \left( c_1 \left\{ \sqrt{\frac{p}{n}} + \frac{p}{n} \right\} + \delta \right) &< c_1 \|\Sigma^*\|_2 \left\{ \frac{\epsilon}{4c_1\|\Sigma^*\|_2} + \frac{\epsilon^2}{16c_1^2\|\Sigma^*\|_2^2} \right\} + \frac{\epsilon}{2} \\ &= \frac{\epsilon}{4} + \frac{\epsilon^2}{16c_1\|\Sigma^*\|_2} + \frac{\epsilon}{2} < \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Since  $\delta < 1$ , it holds that  $\delta^2 < \delta$ . Then

$$\begin{aligned} \mathbb{P} \left( \|\hat{\Sigma} - \Sigma^*\|_2 \geq \epsilon \right) &\leq \mathbb{P} \left( \|\hat{\Sigma} - \Sigma^*\|_2 \geq \|\Sigma^*\|_2 \left( c_1 \left\{ \sqrt{\frac{p}{n}} + \frac{p}{n} \right\} + \delta \right) \right) \\ &\leq c_2 \exp(-c_3 n \delta^2) = c_2 \exp \left( -\frac{c_3}{4 \max(1, \|\Sigma^*\|_2^2)} n \epsilon^2 \right). \end{aligned}$$

□

We finally have the following result.

**Lemma 3.4.6.** *In the event that*

$$\|\Delta_\Sigma\|_2 = \|\hat{\Sigma} - \Sigma^*\|_2 < \min \left\{ \frac{\epsilon_1}{\sqrt{d}(4 + 8c_{\Sigma^*})}, \frac{\epsilon_2}{2c_C} \right\}$$

*it holds that*

$$\|(\Delta_\Gamma).s\|_\infty < \epsilon_1 \quad \text{and} \quad \|\Delta_g\|_\infty < \epsilon_2.$$

**Proof.** The result follows directly from (3.16), where  $\|\Delta_\Sigma\|_2^2 \leq \|\Delta_\Sigma\|_2$ , and Lemma 3.4.3. □

With this preparation we can complete the proof of our main result. Using the preparation in Section 3.4, we prove the main result.

**Proof of Theorem 3.3.3.** We prove the result in three steps.

1) It has to hold that

$$\frac{\epsilon}{\sqrt{d}(4 + 8c_{\Sigma^*})}, \frac{\epsilon}{2c_C} \in \left( 4c_1 \|\Sigma^*\|_2 \sqrt{p/n}, 2 \right).$$

2) Then Corollary 3.4.5 gives us that

$$\|\Delta_\Sigma\|_2 < \min \left\{ \frac{\epsilon}{\sqrt{d}(4 + 8c_{\Sigma^*})}, \frac{\epsilon}{2c_C} \right\}$$

with probability at least  $1 - c_2 \exp(-\tau_1 p)$ . Then  $\|(\Delta_\Gamma).s\|_\infty < \epsilon$  and  $\|\Delta_g\|_\infty < \epsilon$ , using Lemma 3.4.6.

3) We verify that  $\epsilon \leq \frac{\alpha}{6c_{\Gamma^*}}$  under the assumption on the sample size. Then, the result follows from Theorem 3.3.1.

In the following, we go through the steps in detail.

1) Using the lower bound on the sample size, it holds that

$$\begin{aligned}
 \frac{\epsilon}{\sqrt{d}(4 + 8c_{\Sigma^*})} &= \frac{\sqrt{\tau_1 \tilde{c} dp/n}}{\sqrt{d}(4 + 8c_{\Sigma^*})} \\
 &< \frac{\sqrt{\tau_1 \tilde{c} dp / \tau_1 \tilde{c} dp \max\{c_*^2, 1/4\}}}{\sqrt{d}(4 + 8c_{\Sigma^*})} \\
 &= \frac{\sqrt{1/\max\{c_*^2, 1/4\}}}{\sqrt{d}(4 + 8c_{\Sigma^*})} \\
 &\leq \sqrt{1/\max\{c_*^2, 1/4\}} \\
 &\leq \sqrt{4} = 2.
 \end{aligned}$$

Using  $\tau_1 \geq 1$ , we obtain

$$\begin{aligned}
 \frac{\epsilon}{\sqrt{d}(4 + 8c_{\Sigma^*})} &= \frac{\sqrt{\tau_1 \tilde{c} dp/n}}{\sqrt{d}(4 + 8c_{\Sigma^*})} \\
 &> \frac{\sqrt{\tilde{c}}\sqrt{p/n}}{(4 + 8c_{\Sigma^*})} \\
 &\geq \frac{\sqrt{(4 + 8c_{\Sigma^*})^2 16c_1^2 c_{\Sigma^*}^2 \sqrt{p/n}}}{(4 + 8c_{\Sigma^*})} \\
 &= 4c_1 c_{\Sigma^*} \sqrt{p/n}.
 \end{aligned}$$

2) Using Corollary 3.4.5 we obtain

$$\begin{aligned}
 &\mathbb{P}\left(\|\Delta_{\Sigma}\|_2 \geq \frac{\epsilon}{\sqrt{d}(4 + 8c_{\Sigma^*})}\right) \\
 &\leq c_2 \exp\left(-\frac{c_3}{4 \max(1, c_{\Sigma^*}^2)} n \frac{\tau_1 \tilde{c} dp/n}{d(4 + 8c_{\Sigma^*})^2}\right) \\
 &\leq c_2 \exp\left(-\frac{c_3 \tilde{c}}{4 \max(1, c_{\Sigma^*}^2)(4 + 8c_{\Sigma^*})^2} \tau_1 p\right) \\
 &\leq c_2 \exp(-\tau_1 p)
 \end{aligned}$$

3) We verify that  $\epsilon \leq \frac{\alpha}{6c_{\Gamma^*}}$  under the assumption on the sample size.

$$\begin{aligned}
 \epsilon &= \sqrt{\tau_1 \tilde{c} dp/n} \\
 &\leq \sqrt{\tau_1 \tilde{c} dp / \tau_1 \tilde{c} dp \max\{c_*^2, 1/4\}} \\
 &= \sqrt{1/\max\{c_*^2, 1/4\}} \\
 &\leq \frac{\alpha}{6c_{\Gamma^*}}
 \end{aligned}$$

For the same choice of  $\epsilon$  and  $\epsilon/2c_C$  steps 1) - 3) can be carried out analogously and we obtain

$$\mathbb{P}\left(\|\Delta_{\Sigma}\|_2 \geq \frac{\epsilon}{2c_C}\right) \leq c_2 \exp(-\tau_1 dp).$$

The result follows by applying Theorem 3.3.1.  $\square$

### 3.5 Irrepresentability Condition

The irrepresentability condition is vital for Theorem 3.3.1. The condition is well-known from the standard lasso regression, but appears to be much more subtle in the Lyapunov model. In regression and in the Lyapunov model, the irrepresentability condition makes an assumption about the Gram matrix in light of the signal. However, the Gram matrix in regression depends solely on the predictors, whereas the Gram matrix for the Lyapunov model is obtained from the matrix  $A(\Sigma)$  which depends on the signal itself; recall Example 3.2.1. This section is structured in 3 parts. First, we prove that for every support corresponding to a directed acyclic graph (DAG), there exists a drift matrix that fulfills the irrepresentability condition from Theorem 3.3.1. Second, we present a theoretical result that a weaker notion of the irrepresentability condition is necessary for consistent support recovery. Third, we investigate both irrepresentability conditions by the means of simulations. Naturally, we observe that the irrepresentability condition presented in Theorem 3.3.1 is fulfilled less often than its weaker notion introduced in Section 3.5.2. Nevertheless, the frequency with which the irrepresentability conditions are met for randomly drawn signals is surprising. In a last step, we present simulation results suggesting that despite the fact that only the necessity of the weak irrepresentability condition is proven in this thesis, the weak irrepresentability condition being fulfilled results in an extremely positive influence on support recovery.

In our study of the irrepresentability condition, we will consider the case where the volatility matrix  $C$  is a multiple of the identity, specifically, we assume  $C = 2I_p$  throughout this section. (Other diagonal matrices  $C$  would also be tractable for analysis and would yield analogous conclusions.) Before proceeding, we recall that a matrix  $M^* \in \text{Stab}_p$  with support  $S$  satisfies the irrepresentability condition if

$$\rho(M^*) := \|\Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}\|_\infty \quad (3.17)$$

is strictly smaller than 1; the condition in (3.15) stated an explicit gap  $\alpha > 0$ . In the sequel, we will refer to the number  $\rho(M^*)$  as the *irrepresentability constant* of  $M^*$ .

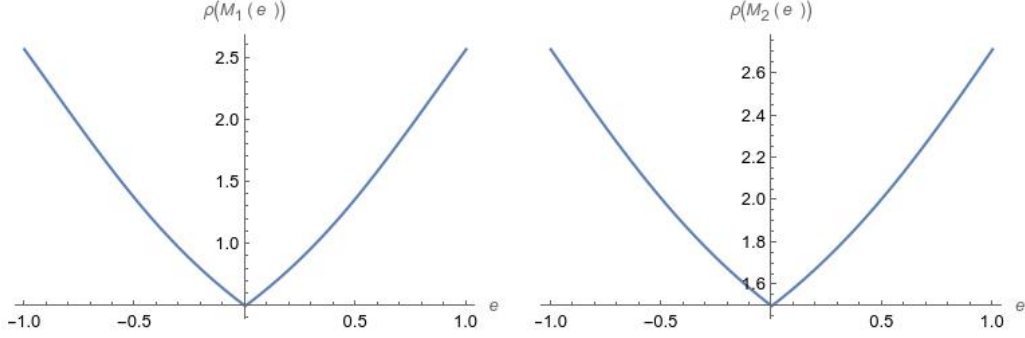
#### 3.5.1 Theoretical Analysis of the (Sufficient) Irrepresentability Condition

In standard lasso regression, the irrepresentability condition is fulfilled when each irrelevant predictor exhibits little correlation with the active predictors. In particular, the condition would hold in a neighborhood of a diagonal Gram matrix. Under the Lyapunov model, it is not obvious how to suggest points for which the irrepresentability condition may be fulfilled. By the analogy with regression, natural candidates are obtained from drift matrices that are close to diagonal, for which the resulting covariance matrices are close to diagonal as well. Such candidates are the topic of the analysis presented in this subsection.

**Example 3.5.1.** Consider the graph  $G = (V, E)$  in Figure 3.4, a path on 3 nodes. For small  $e \in \mathbb{R}$ , we define two stable matrices  $M_1(e)$  and  $M_2(e)$  with support given by  $G$ . We set their diagonals to  $\text{diag}(M_1(e)) = (-1/2, -1, -3/2)$  and  $\text{diag}(M_2(e)) = (-3/2, -1, -1/2)$ , respectively, and we set all non-zero off-diagonal entries equal to  $e$ . Note that the diagonal of  $M_2(e)$  is the reverse of the diagonal of  $M_1(e)$ . In Figure

### 3.5 Irrepresentability Condition

3.3, we plot the two irrepresentability constants  $\rho(M_1(e))$  and  $\rho(M_2(e))$  as functions of the off-diagonal value  $e$ . We observe that irrepresentability holds in a neighborhood of  $M_1(0)$ , but not around  $M_2(0)$ .



**Figure 3.3:** Values of the irrepresentability constants  $\rho(M_1(e))$  and  $\rho(M_2(e))$  for the two matrices from Figure 3.4 plotted against the size of the off-diagonal entries  $e$ . Left:  $\rho(M_1(e))$  where  $\text{diag}(M_1(e)) = (-1/2, -1, -3/2)$ . Right:  $\rho(M_2(e))$  where  $\text{diag}(M_2(e)) = (-3/2, -1, -1/2)$ .

In the example just presented, the order of diagonal entries is seen to impact whether irrepresentability holds near a diagonal matrix. As we prove in the theorem below this fact is not a coincidence but rather a general phenomenon.

Let  $S \subseteq \{(i, j) : 1 \leq i, j \leq p\}$  be a given support set. We say that the irrepresentability condition for support  $S$  holds uniformly over a set  $U \subset \text{Stab}_p$  if there exists  $\alpha > 0$  such that  $\rho(M^*) \leq 1 - \alpha$  for all  $M^* \in U$  with support  $S(M^*) = S$ . By our convention, the edge set of a directed graph  $G = (V, E)$  determines the support set  $S_G = \{(j, i) : i \rightarrow j \in E\}$ .



**Figure 3.4:** Directed graph on 3 nodes.

**Theorem 3.5.2.** *Let  $G = (V, E)$  be a graph with  $p$  nodes. Let  $M^0 = \text{diag}(-d_1, \dots, -d_p)$  be a stable diagonal matrix. Then, the irrepresentability condition for support  $S_G$  holds uniformly over a neighborhood of  $M^0$  if and only if*

$$d_i < d_j \text{ for every edge } i \rightarrow j \in E.$$

*In particular, it is necessary that the graph  $G$  is a DAG.*

**Proof.** Let  $\Sigma^0 = \Sigma(M^0, C)$  be the covariance matrix associated to the drift matrix  $M^0$ . As we assume that  $C = 2I_p$ , we have

$$\Sigma^0 = -(M^0)^{-1} = \text{diag}(1/d_1, \dots, 1/d_p).$$

Writing  $\Gamma^0 = \Gamma(\Sigma^0)$  for the resulting Gram matrix, we define the *local* irrepresentability constant

$$\tilde{\rho}_G(M^0) = \|\Gamma_{S_G^c S_G}^0 (\Gamma_{S_G S_G}^0)^{-1}\|_\infty.$$

If a small open ball around  $M^0$  contains a matrix  $M$ , then the ball also contains all matrices that are obtained from  $M$  by negating one or more of the off-diagonal entries. Hence, by continuity, the irrepresentability condition for support  $S_G$  holds uniformly over a neighborhood of  $M^0$  if and only if (i) the submatrix  $\Gamma_{S_G S_G}^0 = (\Gamma^0)_{S_G S_G}$  is invertible and (ii)  $\tilde{\rho}_G(M^0) < 1$ .

Since  $\Sigma^0$  is diagonal, plugging it into the coefficient matrix from (3.6) gives a symmetric matrix with entries

$$A(\Sigma^0)_{(i,j),(k,l)} = \begin{cases} 2/d_l & \text{if } i = j = k = l, \\ 1/d_l & \text{if } i = k, j = l \text{ and } k \neq l, \\ 1/d_l & \text{if } i = l, j = k \text{ and } k \neq l, \\ 0 & \text{otherwise.} \end{cases}$$

The entries of the Gram matrix  $\Gamma^0 = \Gamma(\Sigma^0)$  are the inner products of the columns of  $A(\Sigma^0)$ . That is,

$$\Gamma_{(i,j),(k,l)}^0 = \begin{cases} 4/d_l^2 & \text{if } i = j = k = l, \\ 2/d_l^2 & \text{if } i = k, j = l \text{ and } k \neq l, \\ 2/(d_k d_l) & \text{if } i = l, j = k \text{ and } k \neq l, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the only off-diagonal entries in  $\Gamma^0$  occur when the row index is  $(i, j)$  and the column index is  $(j, i)$  with  $i \neq j$ . We display the matrices  $A(\Sigma^0)$  and  $\Gamma^0$  for a graph with  $p = 3$  nodes in Example 3.5.3.

*Case I: Graph contains a two-cycle.* Suppose  $G$  contains a two-cycle, say  $k \rightarrow l \rightarrow k$  with  $k \neq l$ . The two edges on the cycle index two columns of  $A(\Sigma^0)$  that are linearly dependent. Indeed, the column indexed by  $(k, l)$  has only two nonzero entries in rows  $(k, l)$  and  $(l, k)$ , both of which are equal to  $d_l$ , and the same holds for the column indexed  $(l, k)$  except that the common value of its two nonzero entries is  $d_k$ . The columns  $(k, l)$  and  $(l, k)$  of  $\Gamma^0$  are similarly linearly dependent. Therefore, the submatrix  $\Gamma_{S_G S_G}^0$  fails to be invertible, if the graph  $G$  contains a two-cycle. Consequently, the irrepresentability condition holds uniformly over a neighborhood of  $M^0$  only if  $G$  is free of two-cycles, in which case we call  $G$  *simple*.

*Case II. Graph is simple.* In the rest of the proof suppose that  $G$  is simple. In this case, the submatrix  $\Gamma_{S_G S_G}^0$  is diagonal with entries

$$\Gamma_{(k,l),(k,l)}^0 = \begin{cases} 4/d_l^2 & \text{if } k = l, \\ 2/d_l^2 & \text{if } k \neq l, \end{cases}$$

where  $l \rightarrow k$  is an edge of  $G$ . The second submatrix of interest,  $\Gamma_{S_G^c S_G}^0$ , also has only one nonzero entry in each column. If  $l \rightarrow k$  is an edge, indexing column  $(k, l)$ , then the entry is

$$(\Gamma_{S_G^c S_G}^0)_{(l,k),(k,l)} = 2/(d_k d_l).$$

### 3.5 Irrepresentability Condition

Note that  $G$  being simple implies that  $k \rightarrow l$  is not an edge of  $G$ . Multiplying the second submatrix to the inverse of the first, we obtain that

$$\begin{aligned} & (\Gamma_{S_G^c, S_G}^0 (\Gamma_{S_G, S_G}^0)^{-1})_{(i,j),(l,k)} \\ &= \begin{cases} d_k/d_l & \text{if } (i,j) = (k,l) \text{ and } (l,k) \in S_G, (k,l) \in S_G^c, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Since  $\tilde{\rho}_G(M^0)$  is obtained via the maximum absolute row sum, we have  $\tilde{\rho}_G(M^0) < 1$  if and only if  $d_i/d_j < 1$  for all pairs  $(j,i) \in S_G$ , or equivalently, all edges  $i \rightarrow j \in E$ , as the theorem claims. If  $G$  contains a cycle of at least length 3, there exists a sequence of edges in  $E$  such that  $i_1 \rightarrow i_2 \rightarrow i_3 \rightarrow \dots \rightarrow i_m \rightarrow i_1$  with  $i_1, \dots, i_m \in V$ . Then, we have  $\tilde{\rho}_G(M^0) < 1$  if and only if

$$d_{i_1}/d_{i_2} < 1, \quad d_{i_2}/d_{i_3} < 1, \quad \dots \quad d_{i_{m-1}}/d_{i_m} < 1, \quad d_{i_m}/d_{i_1} < 1.$$

Multiplying yields

$$d_{i_1}/d_{i_2} \cdot d_{i_2}/d_{i_3} \cdot \dots \cdot d_{i_{m-1}}/d_{i_m} \cdot d_{i_m}/d_{i_1} = 1$$

which contradicts that all individual quotients are smaller than one.  $\square$

We illustrate the matrix calculations in the proof of Theorem 3.5.2 for a graph on  $p = 3$  nodes.

**Example 3.5.3.** We consider the 3-chain  $G = (V, E)$  displayed in Figure 3.4, and the matrices

$$M^0 = \text{diag}(-d_1, -d_2, -d_3) \quad \text{and} \quad \Sigma^0 = \text{diag}(1/d_1, 1/d_2, 1/d_3).$$

Ordering rows as

$(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)$  and columns as

$(1, 1), (2, 1), (3, 1), (1, 2), (2, 2), (3, 2), (1, 3), (2, 3), (3, 3)$ , we find

$$A(\Sigma^0) = \begin{pmatrix} 2/d_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/d_1 & 0 & 1/d_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/d_1 & 0 & 0 & 0 & 1/d_3 & 0 & 0 \\ 0 & 1/d_1 & 0 & 1/d_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2/d_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/d_2 & 0 & 1/d_3 & 0 \\ 0 & 0 & 1/d_1 & 0 & 0 & 0 & 1/d_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/d_2 & 0 & 1/d_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2/d_3 \end{pmatrix}$$

and for  $\Gamma^0$  using the labelling  $(1, 1), (2, 1), (3, 1), (1, 2), (2, 2), (3, 2), (1, 3), (2, 3), (3, 3)$  both for rows and columns we obtain

$$\begin{pmatrix} 4/d_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2/d_1^2 & 0 & 2/d_1d_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2/d_1^2 & 0 & 0 & 0 & 2/d_1d_3 & 0 & 0 \\ 0 & 2/d_1d_2 & 0 & 2/d_2^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4/d_2^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2/d_2^2 & 0 & 2/d_2d_3 & 0 \\ 0 & 0 & 2/d_1d_3 & 0 & 0 & 0 & 2/d_3^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2/d_2d_3 & 0 & 2/d_3^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4/d_3^2 \end{pmatrix}.$$

Since

$$S_G = \{(1, 1), (2, 1), (2, 2), (3, 2), (3, 3)\} \quad \text{and} \\ S_G^c = \{(3, 1), (1, 2), (1, 3), (2, 3)\}$$

we obtain

$$(\Gamma_{S_G S_G}^0)^{-1} = \text{diag}(d_1^2/4, d_1^2/2, d_2^2/4, d_2^2/2, d_3^2/4), \\ \Gamma_{S_G^c S_G}^0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2/d_1d_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2/d_2d_3 & 0 \end{pmatrix},$$

and

$$\Gamma_{S_G^c S_G}^0 (\Gamma_{S_G S_G}^0)^{-1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & d_1/d_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_2/d_3 & 0 \end{pmatrix}.$$

To have  $\|\Gamma_{S_G^c S_G}^0 (\Gamma_{S_G S_G}^0)^{-1}\|_\infty < 1$ , we need  $d_1/d_2 < 1$  and  $d_2/d_3 < 1$ . With the edges  $1 \rightarrow 2$  and  $2 \rightarrow 3$  present in  $G$ , this requirement coincides with the statement of Theorem 3.5.2.

### 3.5.2 Necessity of the Weak Irrepresentability Condition

In Theorem 3.3.1 we show that the irrepresentability condition

$$\|\Gamma_{S^c S}^* (\Gamma_{SS}^*)^{-1}\|_\infty \leq (1 - \alpha), \quad \alpha \in (0, 1)$$

is sufficient for model selection consistency. As we show in the subsequent Proposition, a weaker version of the condition is indeed necessary for model selection consistency.

**Definition 3.5.4.** Let  $M^* \in \text{Stab}_p$  and  $S = S(M)$  its corresponding support set. Then, the weak irrepresentability condition is fulfilled if

$$\|\Gamma_{S^c S}^* (\Gamma_{SS}^*)^{-1} \text{sign}(\text{vec}(M^*))_S\|_\infty \leq 1. \quad (3.18)$$



### 3.5 Irrepresentability Condition

We would like to address a small subtlety regarding the relation of the irrepresentability condition to the weak irrepresentability condition.

**Remark 3.5.5.** Consider a matrix  $M^* \in \text{Stab}_p$  fulfilling the irrepresentability condition (3.15), then it also fulfills the weak irrepresentability condition (3.18). The reasoning is that by multiplying  $\Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}$  with  $\text{sign}(\text{vec}(M^*))_S$  the absolute values of the entries in a row of  $\Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}$  are added up in the “worst case”. By applying  $\|\cdot\|_\infty$  the maximum value is chosen. That is exactly what  $\|\Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}\|_\infty$  is.

If the slightly weaker condition (3.18) is violated and the entries in the drift matrix fulfill a minimal signal strength condition, we cannot recover the correct support asymptotically.

**Proposition 3.5.6.** Consider the setting of Corollary 3.3.3. Let  $M^* \in \text{Stab}_p$  with  $S = S(M^*)$  such that

$$\min_{\substack{1 \leq j < k \leq p \\ (j,k) \in S}} |M_{jk}^*| > \frac{2c_{\Gamma^*}}{2 - \alpha} \lambda$$

holds and that the weak irrepresentability condition (3.18) is violated. For a fixed positive definite matrix  $C$ , the equilibrium distribution for  $M^*$  is given by  $\mathcal{N}(0, \Sigma^*)$ . Let  $X_1, \dots, X_p \in \mathbb{R}^p$  be an i.i.d sample of centered observations and let

$$\hat{\Sigma}^n = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top$$

be the sample covariance. We denote the estimate obtained by the direct Lyapunov lasso (3.4) using  $\hat{\Sigma}^n$  by  $\hat{M}^n$ . Then it holds that

$$\mathbb{P}(S(\hat{M}^n) = S(M^*)) \longrightarrow 0 \quad \text{for } n \rightarrow \infty.$$

**Proof.** The proof is based on the proof of Theorem 3.3.1. Since the optimization problem (3.4) is convex, the KKT - conditions

$$\hat{\Gamma}^n \text{vec}(\hat{M}^n) - \hat{g}^n + \lambda \hat{z}^n = 0, \quad (3.19)$$

with

$$\hat{z}_{(i,j)}^n = \begin{cases} \text{sign}(\text{vec}(\hat{M}^n)_{(i,j)}) & \text{if } \text{vec}(\hat{M}^n)_{(i,j)} \neq 0, \\ \in [-1, 1] & \text{if } \text{vec}(\hat{M}^n)_{(i,j)} = 0, \end{cases}$$

are necessary and sufficient for optimality of  $\hat{M}^n$ . Assume that  $S(\hat{M}^n) = S(M^*)$ . Then,  $\hat{M}^n$  is the unique solution of the support restricted problem (3.12) and following the calculations in the proof of Theorem 3.3.1, the subgradient  $\hat{z}_{S^c}^n$  is given by

$$\hat{z}_{S^c}^n = \frac{1}{\lambda} \left[ -\Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}((\Delta_{\Gamma}^n)_{SS} \text{vec}(\hat{M}^n)_S + (\Delta_g^n)_S) + (\Delta_{\Gamma}^n)_{S^c S} \text{vec}(\hat{M}^n)_S + (\Delta_g^n)_{S^c} + \lambda \Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1} \text{sign}(\text{vec}(\hat{M}^n)_S) \right]. \quad (3.20)$$

We need  $\|\hat{z}_{S^c}^n\|_\infty \leq 1$  for  $\hat{M}^n$  to satisfy the KKT-conditions (3.19). Using Lemma 3.4.6 together with Corollary 3.4.5, we obtain that  $\Delta_g^n \xrightarrow{P} 0$  and that  $\Delta_\Gamma^n \xrightarrow{P} 0$ . Moreover, the inequality (3.13) holds for  $n$  large enough for  $\hat{M}^n$  resulting in

$$\|\text{vec}(\hat{M}^n)_S - \text{vec}(M^*)_S\|_\infty \leq \frac{2c_{\Gamma^*}}{2-\alpha}\lambda.$$

Then, we obtain for the weak irrepresentability condition that

$$\Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}\text{sign}(\text{vec}(\hat{M}^n)_S) = \Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}\text{sign}(\text{vec}(M^*)_S).$$

Therefore, we obtain

$$\|\hat{z}_{S^c}^n\|_\infty \xrightarrow{P} \|\Gamma_{S^c S}^*(\Gamma_{SS}^*)^{-1}\text{sign}(\text{vec}(M^*)_S)\|_\infty > 1.$$

Asymptotically, the subgradient condition is violated for  $\hat{M}^n$  with  $S(\hat{M}^n) = S(M^*)$  and probability 1. Hence,

$$\mathbb{P}(S(\hat{M}^n) = S(M^*)) \longrightarrow 0$$

as the sample size  $n \rightarrow \infty$ . □

It is also easily possible to construct drift matrices fulfilling the weak irrepresentability condition (3.18).

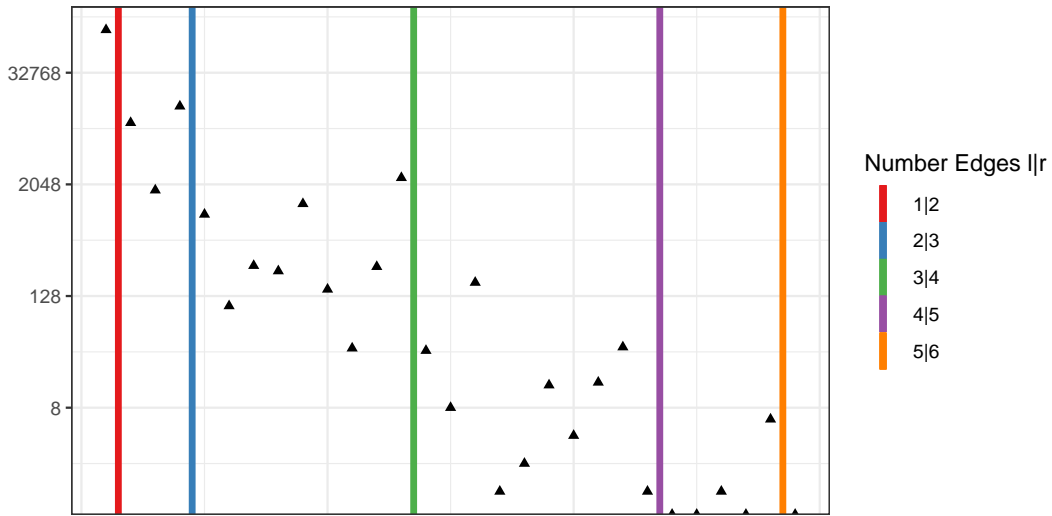
**Remark 3.5.7.** *The same construction as in Theorem 3.5.2 is applicable to the weak irrepresentability condition (3.18).*

### 3.5.3 Simulation Studies: Fulfillment of the Irrepresentability Condition and the Weak Irrepresentability Condition

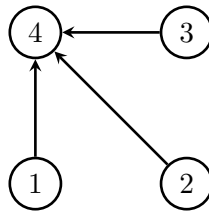
In this section, we want to answer two urgent questions. We have shown that for every DAG there exist non-trivial stable drift matrices such that the irrepresentability condition (3.15) holds. The same is possible for the weak irrepresentability condition (3.18). These signals were constructed to be in a neighborhood of diagonal matrices whose diagonal entries are ordered in accordance with the topological ordering of the DAG. As the size of the graphs increases, this diagonal ordering becomes more restrictive. Moreover, there might be signals that have a different diagonal ordering, but still fulfill the irrepresentability condition. Therefore, the first question is how often the conditions are fulfilled when selecting random drift matrices according to a predetermined distribution.

Given a graph  $G = (V, E)$ , we generate signals  $M^* \in \text{Stab}_p(E)$  by drawing from the uniform distribution on the subset of matrices in  $\text{Stab}_p(E)$  that have all entries in  $[-1, 1]$ . The sampling is carried out by rejection sampling, with rejection of matrices that are not stable.

We consider connected graphs with  $p = 2, 3, 4$  nodes and at most  $p(p+1)/2$  edges. This includes all DAGs but also many cyclic graphs. Furthermore, we only consider one labeling of vertices for every graph. For every graph, we check for one million simulated signals  $M^*$  if  $\rho(M^*) < 1$  and store the signals that meet the irrepresentability



**Figure 3.5:** Frequency of the irrepresentability condition (3.15) being met for one million simulated stable matrices  $M^*$  for DAGs up to 4 nodes. The number of edges is given by the coloring.



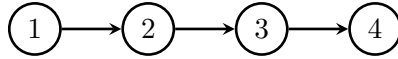
**Figure 3.6:** The graph on three nodes with highest frequency of simulated signals satisfying irrepresentability.

condition (3.15). The frequency of signals that meet the irrepresentability condition is shown in Figure 3.8.

The frequency with which the irrepresentability condition is fulfilled decreases with increasing number of edges. The decrease is not monotonic in the number of edges, since the restrictiveness is tied to whether an edge adds a new condition on the quotient of the diagonal elements as presented in Theorem 3.5.2. An investigation of the drift matrices in Figure 3.5 shows that those who fulfill the irrepresentability condition (3.15) all have a diagonal ordering according to our theoretical result.

**Example 3.5.8.** Consider the graph shown in Figure 3.6. The drift matrices supported on this graph have the highest frequency of irrepresentability among the graphs with three edges in Figure 3.5. Since there is no edge between the nodes  $\{1, 2, 3\}$ , the only conditions on the diagonal are  $d_1/d_4 < 1$ ,  $d_2/d_4 < 1$  and  $d_3/d_4 < 1$ . Translated, this means that  $d_4$  has to be bigger than  $d_1, d_2, d_3$ .

Following Theorem 3.3.1, the conditions on the diagonal elements for the drift matrices supported in Figure 3.7 are  $d_1/d_2 < 1$ ,  $d_2/d_3 < 1$  and  $d_3/d_4 < 1$ . In particular, these conditions also contain the requirement that  $d_4$  has to be bigger than  $d_1, d_2, d_3$ . In

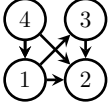
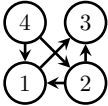
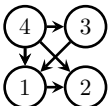
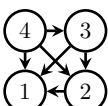


**Figure 3.7:** The path from 1 to 4.

*addition, they contain the requirement that  $d_3$  has to be bigger than  $d_1, d_2$  and that  $d_2$  has to be bigger than  $d_1$ .*

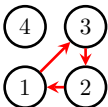
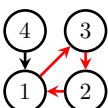
Another important observation is that the condition is extremely restrictive when selecting stable drift matrices according to a uniform distribution. In Figure 3.5 we observe that already if a graph on 4 nodes has 3 or more edges, the irrepresentability condition is only fulfilled in less than 1 % of the cases. There even exist some graphs for which the irrepresentability condition is never met. These graphs are displayed in Table 3.1. We tried to find stable drift matrices by applying the above mentioned selection procedure ten million times to these critical graphs. For only two of the graphs we were able to select drift matrices fulfilling the irrepresentability condition. Theorem 3.5.2 guarantees that there must exist stable drift matrices supported over the two remaining graphs. Using Theorem 3.5.2, we put one choice for each of the two graphs in red in Table 3.1.

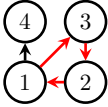
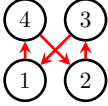
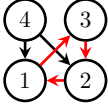
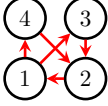
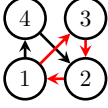
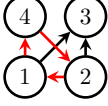
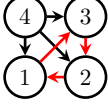
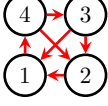
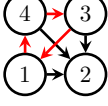
**Table 3.1:** **Left:** The four graphs where none of the one million randomly selected drift matrices  $M$  fulfilled the irrepresentability condition in Figure 3.5. **Right:** Drawing another ten million drift matrices, we obtain for the second and third graph drift matrices that fulfill the irrepresentability condition (black). For the first and fourth graph, we use Theorem 3.5.2 to construct drift matrices that fulfill the irrepresentability condition (red).

	$\begin{pmatrix} -0.5 & 0 & 0 & 0.05 \\ 0.05 & -1 & 0.05 & 0.05 \\ 0.05 & 0 & -0.75 & 0 \\ 0 & 0 & 0 & -0.25 \end{pmatrix}$
	$\begin{pmatrix} -0.584860503 & 0.03949857 & 0.0000000 & -0.05605342 \\ 0.000000000 & -0.35729470 & 0.0000000 & -0.00303305 \\ 0.005031837 & -0.08209815 & -0.7782385 & 0.000000000 \\ 0.000000000 & 0.000000000 & 0.0000000 & -0.22854795 \end{pmatrix}$
	$\begin{pmatrix} -0.7388917 & 0.0000000 & -0.1277403 & 0.01491351 \\ -0.1184546 & -0.9615896 & 0.0000000 & -0.09631827 \\ 0.0000000 & 0.0000000 & -0.4652617 & 0.04858871 \\ 0.0000000 & 0.0000000 & 0.0000000 & -0.23807701 \end{pmatrix}$
	$\begin{pmatrix} -1 & 0.05 & 0.05 & 0.05 \\ 0 & -0.75 & 0.05 & 0.05 \\ 0 & 0 & -0.5 & 0.05 \\ 0 & 0 & 0 & -0.25 \end{pmatrix}$

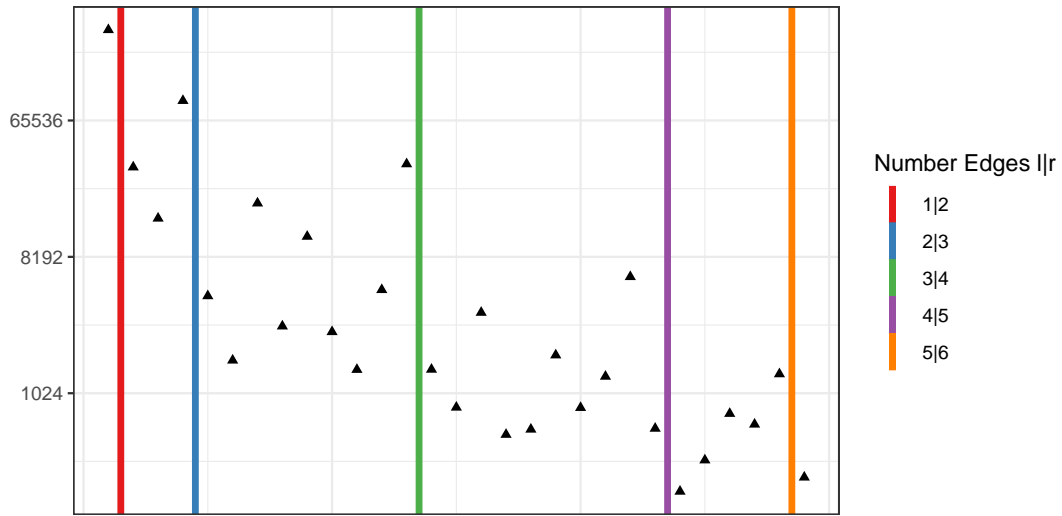
We also carried out the simulation study for simple cyclic graphs. None of the cyclic graphs on 4 nodes fulfilled the irrepresentability condition for ten million randomly selected drift matrices for each graph structure. This is not a proof that the irrepresentability condition (3.15) is never met for a cyclic graph, but at least a strong computational evidence. In a next step we carry out the same sampling procedure for graphs on 4 nodes for the weak irrepresentability condition (3.18) than we did previously for the irrepresentability condition (3.15). The results are displayed in Figure 3.8.

**Table 3.2:** **Left:** All simple cyclic graphs with 4 nodes, up to relabelling of the nodes. Edges on cycles are highlighted in red. **Right:** Specific choice of matrices  $M$  matching the graph on the left and fulfilling the weak irrepresentability condition (3.18), all entries are rounded to 10 digits.

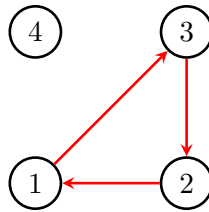
	$\begin{pmatrix} -0.0444620792 & -0.5733500496 & 0.0000000000 & 0.0000000000 \\ 0.0000000000 & -0.0153532191 & 0.0054622865 & 0.0000000000 \\ 0.8317033453 & 0.0000000000 & -0.8824298000 & 0.0000000000 \\ 0.0000000000 & 0.0000000000 & 0.0000000000 & -0.3405775614 \end{pmatrix}$
	$\begin{pmatrix} -0.9780979650 & 0.1042322782 & 0.0000000000 & 0.3752107187 \\ 0.0000000000 & -0.7998522464 & -0.4260628200 & 0.0000000000 \\ 0.2079165080 & 0.0000000000 & -0.6517819995 & 0.0000000000 \\ 0.0000000000 & 0.0000000000 & 0.0000000000 & -0.8112314143 \end{pmatrix}$

	$\begin{pmatrix} -0.6792729949 & -0.6022921619 & 0.0000000000 & 0.0000000000 \\ 0.0000000000 & -0.1733464822 & 0.5762203289 & 0.0000000000 \\ 0.0383909321 & 0.0000000000 & -0.1785332798 & 0.0000000000 \\ 0.2089620568 & 0.0000000000 & 0.0000000000 & -0.6556593408 \end{pmatrix}$
	$\begin{pmatrix} -0.5008390141 & 0.0000000000 & -0.3301411900 & 0.0000000000 \\ 0.0000000000 & -0.0754047022 & 0.0000000000 & -0.2224099669 \\ 0.0000000000 & 0.9894780936 & -0.8953534714 & 0.0000000000 \\ -0.4568265276 & 0.0000000000 & 0.0000000000 & -0.6545859827 \end{pmatrix}$
	$\begin{pmatrix} -0.9852473154 & 0.0237436080 & 0.0000000000 & -0.1801203806 \\ 0.0000000000 & -0.9146776730 & -0.6301784553 & -0.3625553502 \\ 0.0314035588 & 0.0000000000 & -0.7371845325 & 0.0000000000 \\ 0.0000000000 & 0.0000000000 & 0.0000000000 & -0.2936787312 \end{pmatrix}$
	$\begin{pmatrix} -0.6168078599 & -0.4643970933 & 0.0000000000 & 0.0000000000 \\ 0.0000000000 & -0.8265482867 & 0.0118716909 & 0.4726413568 \\ 0.3998511671 & 0.0000000000 & -0.8792877044 & 0.0000000000 \\ -0.5496377517 & 0.0000000000 & 0.0000000000 & -0.7865214688 \end{pmatrix}$
	$\begin{pmatrix} -0.2066421132 & -0.0034684981 & 0.1383411973 & 0.0000000000 \\ 0.0000000000 & -0.9617960961 & 0.0000000000 & -0.7641737331 \\ 0.0000000000 & -0.3169060163 & -0.7561623598 & 0.0000000000 \\ -0.7012514030 & 0.0000000000 & 0.0000000000 & -0.2419070452 \end{pmatrix}$
	$\begin{pmatrix} -0.8234110032 & -0.6069790549 & 0.0000000000 & 0.0000000000 \\ 0.0000000000 & -0.4768311884 & 0.0000000000 & -0.5430481988 \\ -0.1151224086 & 0.5541216009 & -0.8947804412 & 0.0000000000 \\ -0.1818817416 & 0.0000000000 & 0.0000000000 & -0.6244826200 \end{pmatrix}$
	$\begin{pmatrix} -0.7566250684 & 0.1517044385 & 0.0000000000 & 0.0068894741 \\ 0.0000000000 & -0.9917302341 & 0.5077337530 & 0.3153799707 \\ 0.0895817326 & 0.0000000000 & -0.7472212519 & -0.1730670566 \\ 0.0000000000 & 0.0000000000 & 0.0000000000 & -0.3600410065 \end{pmatrix}$
	$\begin{pmatrix} -0.8680259003 & 0.4557597358 & -0.0925138230 & 0.0000000000 \\ 0.0000000000 & -0.9139470784 & -0.1607573517 & 0.3138186112 \\ 0.0000000000 & 0.0000000000 & -0.9212171654 & -0.9521876550 \\ -0.5101859323 & 0.0000000000 & 0.0000000000 & -0.2475099666 \end{pmatrix}$
	$\begin{pmatrix} -0.6688544271 & 0.0000000000 & -0.7215559445 & 0.0000000000 \\ -0.4272868899 & -0.9967063963 & 0.0374428187 & -0.8531300114 \\ 0.0000000000 & 0.0000000000 & -0.6779836947 & -0.5781906121 \\ -0.6749138949 & 0.0000000000 & 0.0000000000 & -0.5980373188 \end{pmatrix}$

Comparing the results in Figure 3.8 with those in Figure 3.5, we observe that the weak irrepresentability condition is fulfilled much more often than the irrepresentability condition. The reason is that the sign vector in (3.18) enables fortunate cancellation. Moreover, this allows us to find a suitable drift matrix for every simple cyclic graph on 4 nodes. In Table 3.2, we list all cyclic graphs on 4 nodes together with examples of drift matrices that satisfy the irrepresentability condition. The selection of graphs



**Figure 3.8:** Frequency of the weak irrepresentability condition (3.18) being met for one million simulated stable matrices  $M^*$  for DAGs up to 4 nodes. The number of edges is given by the coloring.



**Figure 3.9:** 3-cycle in a four node setting.

includes all graphs that contain at least one directed cycle and are simple (i.e., do not contain a two-cycle).

Calculations are carried out with the statistical software R. A natural suspicion is that these very few matrices were only selected due to numerical imprecision. In addition, one might wonder if the 10 digits are really necessary. Example 3.5.9 provides more insight using a representative from Table 3.2.

**Example 3.5.9.** For the graph in Figure 3.9 (or first row of Table 3.2) the matrix

$$M = \begin{pmatrix} -0.0444620792 & -0.5733500496 & 0.0000000000 & 0.0000000000 \\ 0.0000000000 & -0.0153532191 & 0.0054622865 & 0.0000000000 \\ 0.8317033453 & 0.0000000000 & -0.8824298000 & 0.0000000000 \\ 0.0000000000 & 0.0000000000 & 0.0000000000 & -0.3405775614 \end{pmatrix}$$

fulfills the weak irrepresentability condition. The margins to satisfy the weak irrepresentability condition are thin. Rounding the entries of  $M$  potentially yields matrices  $M$  that do not satisfy the weak irrepresentability condition. The matrix  $M$  displayed in this example results in a value for the left side of (3.18) of 0.9960339 while the 2 - digit version yields a value of 1.011801, i.e. the longer version fulfills the weak

irrepresentability condition while the shorter version does not. This is the reason for the long displays in Table 3.2. However, for the matrix  $M$  in this Example, we are able to rationalize the entries with a tolerance of 0.0001 to obtain

$$M_R = \begin{pmatrix} -2/45 & -43/75 & 0 & 0 \\ 0 & -1/65 & 1/183 & 0 \\ 84/101 & 0 & -15/17 & 0 \\ 0 & 0 & 0 & -31/91 \end{pmatrix}$$

fulfilling the weak irrepresentability condition with all calculations being carried out rationally in *Mathematica* [Wolfram Research, Inc., 2022]. This allays the concern that these matrices only exist due to numerical imprecision in the calculations.

Summing up the situation for simple cyclic graphs, extensive computation was necessary to present an example for every simple cyclic graph up to four nodes. We were unable to discern the structure that would suggest how to construct such examples in general.

The last class of graphs that misses are the non-simple graphs. We omit discussing them in detail, but the work by Dettling et al. [2023] suggests that there exist drift matrices  $M^*$  supported over non-simple graphs such that  $\Gamma_{SS}^*$  is invertible. This leaves the possibility for drift matrices fulfilling the irrepresentability condition. We observed that the drift matrices that satisfy the weak irrepresentability condition fulfill the diagonal ordering of Theorem 3.3.1 for the ‘‘DAG part’’ of the graph over which the drift matrix is supported.

### 3.5.4 Simulation Studies: Impact of the Weak Irrepresentability Condition

Corollary 3.3.3 ensures that if the irrepresentability condition (3.15) is fulfilled and some assumptions about minimal signal strength and sample size hold, we are able to recover the support of a drift matrix correctly when applying the direct Lyapunov lasso (3.4). We were not able to prove this for the weak irrepresentability condition (3.18), only its necessity in Proposition 3.5.6 in case a minimal signal requirement is fulfilled. Nevertheless, the condition is quite close to the sufficient condition and is fulfilled much more often, as we show in Section 3.5.3. Therefore, we want to investigate the impact of the fulfillment of the weak irrepresentability condition on support recovery. The positive computational results in this section also translate to the irrepresentability condition as every drift matrix fulfilling the irrepresentability condition also fulfills the weak irrepresentability condition.

For every DAG on 4 nodes, we select 10 drift matrices fulfilling the weak irrepresentability condition. The selection procedure is the same that we use to obtain Figure 3.8 (uniform distribution of stable matrices with entries between -1 and 1). Furthermore, we select 100 stable drift matrices supported over the DAGs that do not necessarily fulfill the irrepresentability condition. Based on the drift matrices  $M^*$  and the Lyapunov equation (1.2) with  $C = 2I_p$ , we calculate the equilibrium covariance matrices  $\Sigma^*$ . We then sampled the data with  $n = 100$  from the normal distributions  $\mathcal{N}(0, \Sigma^*)$ . Then, we apply the direct Lyapunov lasso (3.4) along a regularization path

$$\lambda_1 = \lambda_{\max}, \dots, \lambda_{100} = \frac{\lambda_{\max}}{10^4}$$



### 3.5 Irrepresentability Condition

where  $\lambda_{\max}$  is chosen on an initial grid such that  $\hat{M}$  is diagonal. For the estimates  $\hat{M}_1, \dots, \hat{M}_{100}$  obtained along the regularization path, we calculate some basic metrics regarding support recovery of the data generating  $M^*$ .

**Definition 3.5.10.** Let  $\hat{M} \in \mathbb{R}^{p \times p}$  be an estimate and let  $M^*$  be the estimation target. Then, we define

$$\begin{aligned} tp &= |\{\hat{M}_{ij} : \hat{M}_{ij} \neq 0 \text{ and } M_{ij}^* \neq 0\}|, \\ fp &= |\{\hat{M}_{ij} : \hat{M}_{ij} \neq 0 \text{ and } M_{ij}^* = 0\}|, \\ tn &= |\{\hat{M}_{ij} : \hat{M}_{ij} = 0 \text{ and } M_{ij}^* = 0\}|, \\ fn &= |\{\hat{M}_{ij} : \hat{M}_{ij} = 0 \text{ and } M_{ij}^* \neq 0\}|. \end{aligned}$$

While these metrics already provide some insights, there exist more refined metrics to evaluate the performance of a structure learning algorithm.

**Definition 3.5.11.** Let  $\hat{M} \in \mathbb{R}^{p \times p}$  be an estimate and let  $M^*$  be the estimation target and let

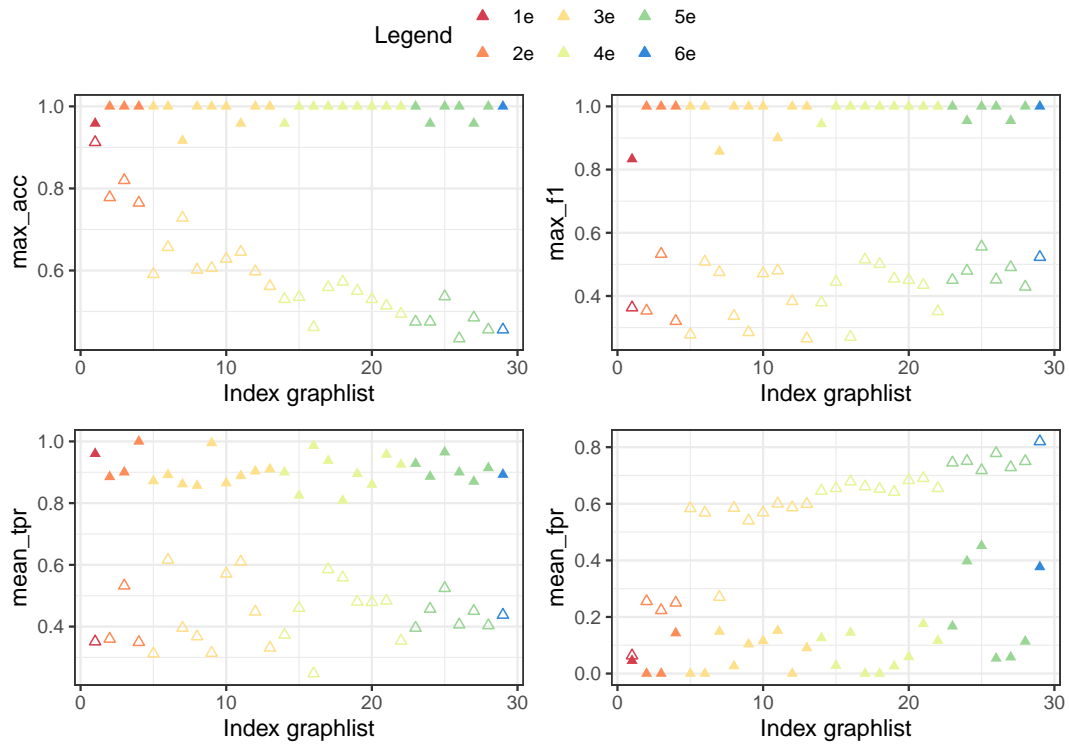
$tp, fp, tn, fn$  be defined as in Definition 3.5.10. Then, we define

$$\begin{aligned} \mathbf{tpr} \text{ (true positive rate)} &= \frac{tp}{tp + fn}, \\ \mathbf{fpr} \text{ (false positive rate)} &= \frac{fp}{fp + tn}, \\ \mathbf{acc} \text{ (accuracy)} &= \frac{tp + tn}{tp + tn + fp + fn}, \\ \mathbf{f_1\text{-score}} &= \frac{2tp}{2tp + fp + fn}, \\ \mathbf{pr} \text{ (precision)} &= \frac{tp}{tp + fp}. \end{aligned}$$

Calculating  $tpr$  and  $fpr$  for all regularization parameters, we define the roc curve as plotting  $tpr$  vs.  $fpr$  with  $fpr$  ranging from 0 to 1 using interpolation and extrapolation if necessary. The **auc roc** or just **auc** is then defined as the area under the roc curve. Calculating  $pr$  and  $tpr$  for all regularization parameters, we define the  $pr$  curve as plotting  $pr$  vs.  $tpr$  with  $tpr$  ranging from 0 to 1 using interpolation and extrapolation if necessary. The **aupr** curve is then defined as the area under the precision curve.

For the estimates  $\hat{M}_1, \dots, \hat{M}_{100}$  obtained for each DAG and for each initial drift matrix  $M^*$ , we calculate the metrics mean  $tpr$ , mean  $fpr$  and max  $acc$ , max  $f_1$ -score. All metrics are then averaged over the 10 drift matrices that satisfy the weak irrepresentability condition per DAG or over the 100 randomly selected drift matrices, respectively. The results are displayed in Figure 3.10. The empty triangles correspond to the average over the randomly selected drift matrices while the full triangles correspond to the average over the drift matrices fulfilling the weak irrepresentability condition.

Generally, there are many subtleties to be discovered in the plots. For conciseness, we limit our discussion to the key observation that across all metrics, the results for the signals that fulfill the weak irrepresentability condition are almost perfect and much

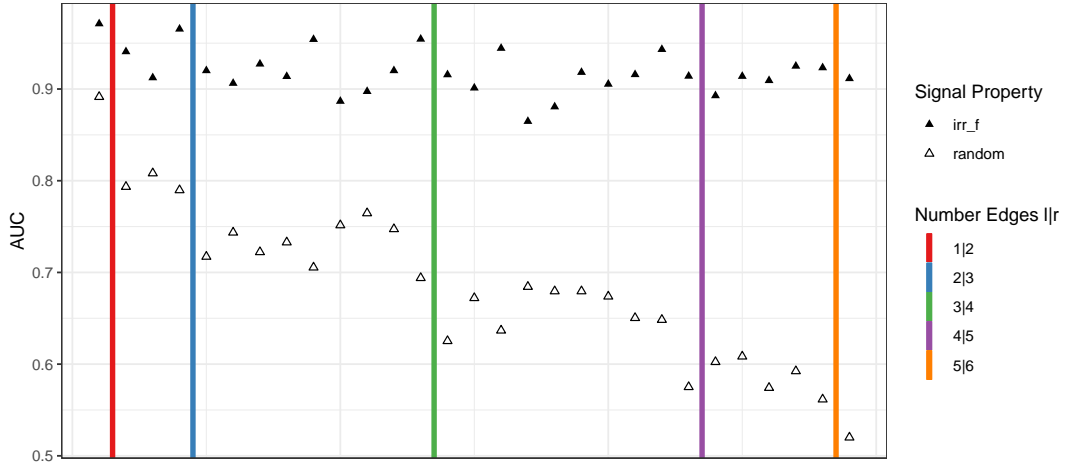


**Figure 3.10:** Four metrics measuring the quality of the estimate for DAGs with up to 4 nodes. The number of edges is given by the coloring. **Empty:** irrepresentability condition in general not fulfilled, **Full:** weak irrepresentability condition fulfilled.

better than for randomly selected ones. Of course, for graphs with fewer edges, more randomly selected drift matrices already fulfill the weak irrepresentability condition, which explains why the difference is not severe.

Lastly, we present the results for the area under the roc curve (auc) using the exact same simulation setup as for Figure 3.10. The auc is particularly insightful as the roc curve is obtained by plotting the trade-off of tpr vs. fpr. An auc value of 0.5 means that the method applied performs badly (random guessing), while a value of 1 is optimal. For drift matrices fulfilling the weak irrepresentability condition, we observe that the auc is above 0.9 for almost all graphs that fulfill the weak irrepresentability condition while the performance is very poor for randomly selected ones.

We do not include further simulations for cyclic graphs in the above setting, which is mainly because we already struggle to find 10 drift matrices supported over cyclic graphs fulfilling the weak irrepresentability condition. In particular, we struggle to find 10 “really different” drift matrices that do not only differ by a small margin in the individual entries.



**Figure 3.11:** The auc values for DAGs with up to 4 nodes. The number of edges is given by the coloring. **Empty:** irrepresentability condition in general not fulfilled, **Full:** weak irrepresentability condition fulfilled.

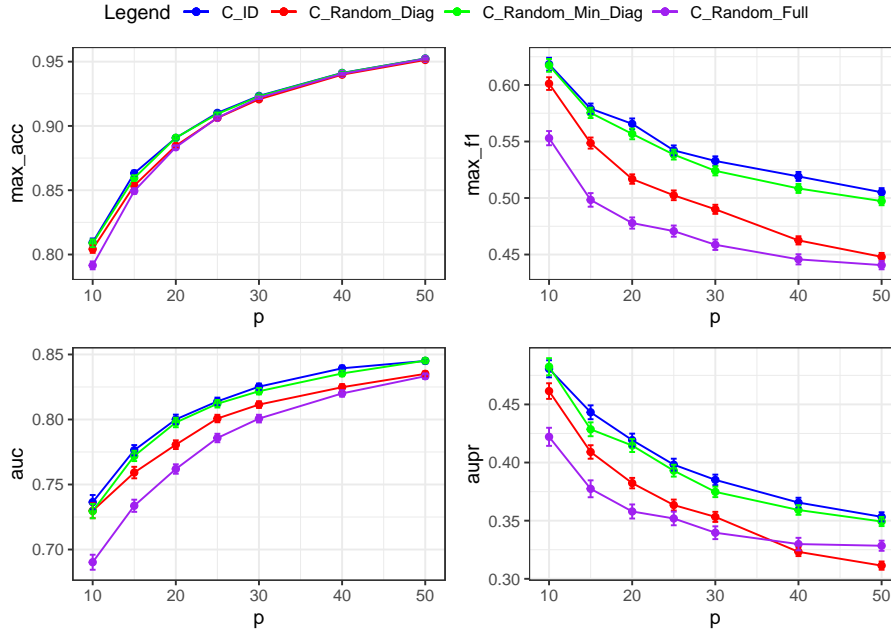
### 3.6 Simulation Studies

In this section, we present simulation studies that provide insight into the performance of the direct Lyapunov lasso in seemingly unfavorable settings. First, most drift matrices do not satisfy the irrepresentability condition; compare Section 3.5.3. Second, while our assumption that  $C$  is fixed up to a scalar multiple is made similarly in the related case where actual time series data is considered [Gaïffas and Matulewicz, 2019], it is an assumption that may be overly simple for many applications. Nevertheless, our simulations suggest robustness of the direct Lyapunov lasso to the irrepresentability condition not being fulfilled and to mild misspecification of the volatility matrix  $C$ , where by robustness we mean that a part of signal is being learned correctly.

For the simulations in this section, we use a similar setting as in Varando and Hansen [2020]. Each stable matrix  $M$  was generated with  $M_{ij} = \omega_{ij}\epsilon_{ij}$  for  $i \neq j$  and  $M_{ii} = -\sum_{j \neq i} |M_{ij}| - |\epsilon_{ii}|$  where  $\omega_{ij} \sim \text{Bernoulli}(d)$  and  $\epsilon_{ij} \sim N(0, 1)$ . Unlike in Varando and Hansen [2020], we consider four different choices for  $C$ . The label in brackets corresponds to the one used in Figure 3.12.

- 1) We choose  $C = 2I_p$  (C\_ID).
- 2) We choose  $C$  diagonal with  $C_{ii} \sim \text{Unif}[0.5, 4]$  (C\_Random\_Diag).
- 3) We choose  $C$  diagonal with  $C_{ii} \sim \text{Unif}[2, 4]$  (C\_Random\_Min\_Diag).
- 4) We choose  $C$  symmetric but non-diagonal. Let  $\tilde{\omega}_{ij} \sim \text{Bernoulli}(2/p)$  and  $\tilde{\epsilon}_{ij} \sim N(0, 1)$  be independent random variables,  $i, j = 1, \dots, p$ . Then the off-diagonal entries of  $C$  are set to  $C_{ij} = \tilde{\omega}_{ij}\tilde{\epsilon}_{ij} + \tilde{\omega}_{ji}\tilde{\epsilon}_{ji}$  and the diagonal entries to  $C_{ii} = \sum_{j \neq i} |C_{ij}| + |\tilde{\epsilon}_{ii}| + 0.5$  (C\_Random\_Full).

For each  $k \in \{1, 2, 3, 4\}$  and  $p = \{10, 15, 20, 25, 30, 40, 50\}$ , the edge probability is set as  $d = k/p$ . For each choice of  $C$ , we generate 100 pairs of signals  $(M, C)$ . We generate  $n = 1000$  observations from a multivariate Gaussian distribution with covariance matrix solving the Lyapunov equation for  $(M, C)$ . Note that  $p^2 > n$  for  $p = \{40, 50\}$



**Figure 3.12:** The maximum accuracy (top left), maximum  $F_1$ -score (top right), area under the ROC curve (bottom left) and area under the precision curve (bottom right) in support recovery with the direct Lyapunov lasso using parameter  $C = 2I_p$ . The data has been generated using the choices 1)  $C_{ID}$ , 2)  $C_{Random\_Diag}$ , 3)  $C_{Random\_Min\_Diag}$ , and 4)  $C_{Random\_Full}$ . The error bars are the estimated standard errors of the average of a metric for a specific problem size over the 400 randomly generated drift matrices.

which corresponds to the high-dimensional setting. Then we apply the direct Lyapunov lasso with  $C = 2I_p$  for model selection. The results are calculated along the  $\lambda$ -grid:

$$0 < \frac{\lambda_{\max}}{10^4} = \lambda_1 < \dots < \lambda_{100} = \lambda_{\max}, \quad (3.21)$$

with  $\lambda_{\max}$  being the minimal  $\lambda$ -value such that  $M$  is diagonal. We compute the maximum accuracy, the maximum  $F_1$ -score, the area under the ROC curve and the area under the precision curve; more details on the metrics are given in Definition 3.5.11. The metrics are averaged over the 4 different sparsity levels  $k$  and the 100 randomly selected drift matrices  $M$ . The results are shown in Figure 3.12.

Choice 1) for  $C$  is used when applying direct Lyapunov lasso for model selection. Thus, it is natural to expect the best results for this choice. Choices 2) and 3) allow for variability on the diagonal. The second choice allows for larger differences ( $\text{Unif}[0.5, 4]$ ) in the size of the diagonal entries, while the third choice is more conservative ( $\text{Unif}[2, 4]$ ). Choices 1) and 3) perform best in our simulations. We observe that there are few differences in all metrics among the choices between choice 1) and choice 3), indicating that the direct Lyapunov lasso with  $C = 2I_p$  possesses a certain robustness to the exact diagonal matrix  $C$  of the data generating model. This is true for all  $p \in \{10, 15, 20, 25, 30, 40, 50\}$ . For choice 2), we observe that the results in all metrics except maximum accuracy fall with increasing  $p$ . For  $p = 40$  and especially for  $p = 50$ , the results for all metrics are similar to choice 4). Choice 4) allows for data generating

models for which  $C$  is no longer diagonal. For this choice, the worst results are to be expected as the matrix  $C$  used for data generation is much different from the one used for estimation. Another interesting point revealed by the simulations is that although the irrepresentability condition is not satisfied in almost any of the signals, it is still possible to get estimates that recover much of the support of the drift matrix.

## 3.7 Real World Data

Previous simulations in Section 3.6 assess the performance of the direct Lyapunov lasso on synthetic data and along a path of regularization parameters  $\lambda$ . However, when applying the method to a real-world dataset, the objective is to provide a single estimate of a network. This can be engineered by applying the Bayesian information criterion (BIC) or its extended version (EBIC). In this section, we revisit both of them in the context of Lyapunov models and showcase the simplicity and strength of the approach by applying the method to the famous Sachs dataset by Sachs et al. [2005]. This dataset has been a testing field for graphical models for many years and was, in particular, analyzed by Fitch [2019] and Varando and Hansen [2020] in the context of Lyapunov models.

### 3.7.1 Sachs Dataset

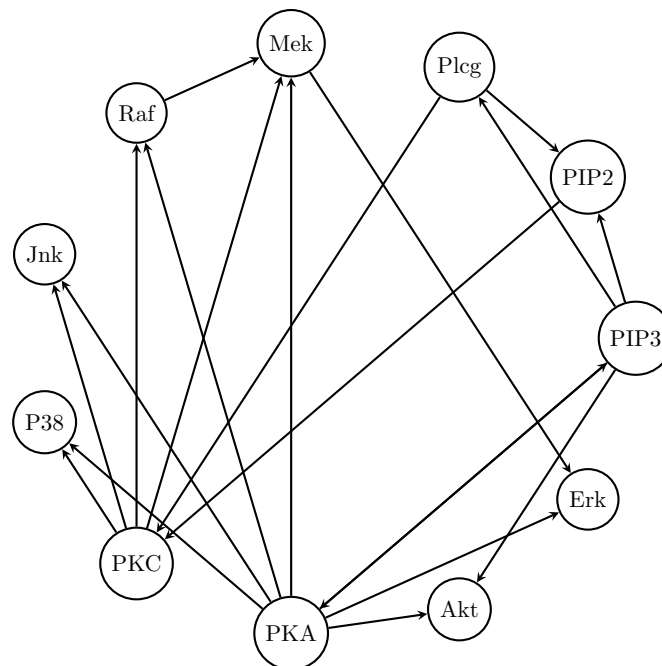
The famous dataset collected and first analyzed by Sachs et al. [2005] has become a testing field for graphical models and model selection algorithms. Some recent examples are the review paper of causal discovery methods based on Graphical Models by Glymour et al. [2019], the application of classical structure equation models allowing for cycles by Amendola et al. [2020] or the application of Lyapunov models and a specific model selection technique by Varando and Hansen [2020]. The dataset consists of flow cytometry measurements of 11 phosphorylated proteins and phospholipids in human T-cells captured under different experimental conditions, resulting in 14 independent datasets of varying sizes ( $n = 707$  to  $n = 927$ ). When flow cytometry is applied, the cells are destroyed during the measurement process, and hence, the measurements are collected at one point in time. Every sample consists of the quantitative and simultaneous measurements of the 11 phosphorylated proteins and phospholipids of single cells. The way in which the dataset was collected, and previous works in the context of structural equation models make it an interesting choice for Lyapunov models (Definition 1.3.2) despite the strong and ambitious parametric assumption.

As the main focus in this work lies on the mathematical properties of Lyapunov models and not so much on the application, we refer to the datasets by numbering them. Table 3.3 provides the assignment of actual names of the datasets (different experimental conditions) in <https://zenodo.org/records/7681811> to the numbers used in this work.

**Table 3.3:** Listing the datasets obtained under different experimental conditions (perturbations) presented by Sachs et al. [2005]. The table displays the ordering used in this work.

Perturbation/Filename	Number of the Dataset
b2camp.csv	Dataset 1
cd3cd28_aktinhib.csv	Dataset 2
cd3cd28_g0076.csv	Dataset 3
cd3cd28_icam2.csv	Dataset 4
cd3cd28_ly.csv	Dataset 5
cd3cd28_psitect.csv	Dataset 6
cd3cd28_u0126.csv	Dataset 7
cd3cd28.csv	Dataset 8
cd3cd28icam2_aktinhib.csv	Dataset 9
cd3cd28icam2_g0076.csv	Dataset 10
cd3cd28icam2_ly.csv	Dataset 11
cd3cd28icam2_psitect.csv	Dataset 12
cd3cd28icam2_u0126.csv	Dataset 13
pma.csv	Dataset 14

The protein-signaling network displayed in Figure 3.13 shows an among scientist accepted network for molecule interactions for the datasets by Sachs et al. [2005]. The authors also mention that there is some ambiguity regarding some connections. Nevertheless, we refer to this network as “ground truth”.



**Figure 3.13:** Among scientists accepted signaling molecule interactions for the dataset by Sachs et al. [2005].

### 3.7.2 Direct Lyapunov Lasso with the (Extended) Bayesian Information Criterion

The first few steps are identical to Section 3.6. Based on a data matrix  $X \in \mathbb{R}^{n \times p}$ , we calculate the sample covariance matrix  $\hat{\Sigma}$  on the standardized data. Setting  $C = 2I_p$ , we apply the direct Lyapunov lasso along a regularization path. We set the regularization path to be the logarithmic sequence

$$\lambda_1 = \lambda_{\max}, \dots, \lambda_{100} = \frac{\lambda_{\max}}{10^4}.$$

where  $\lambda_{\max}$  is chosen such that the estimated drift matrix is diagonal. Using the direct Lyapunov lasso, we compute the estimates  $\hat{M}_1, \dots, \hat{M}_{100}$  along this grid. Extracting the non-zero structure, each estimate  $\hat{M}_i$  defines a directed graph  $G_i$  and thus a model  $\mathcal{M}_{G_i, 2I_p}$ . To decide which model to pick, we use the BIC/extended BIC (Bayesian Information Criterion). First, we recall the idea of the classical BIC. Initially, the BIC criterion for model selection was proposed by Schwarz [1978]. The purpose of this criterion is to select a unique model out of many models of different dimensions.

The idea behind the BIC criterion is to maximize the posterior probability of a model given the data  $x_1, \dots, x_n$  which we denote by  $\mathbb{P}(\mathcal{M}_{G,C} | x_1, \dots, x_n)$ . Direct application of Bayes Theorem yields that

$$\mathbb{P}(\mathcal{M}_{G,C} | x_1, \dots, x_n) \propto \mathbb{P}(x_1, \dots, x_n | \mathcal{M}_{G,C}) \mathbb{P}(\mathcal{M}_{G,C}).$$

If we assume that all models are equally likely ( $\mathbb{P}(\mathcal{M}_{G,C}) = \text{const.}$ ), maximizing  $\mathbb{P}(\mathcal{M}_{G,C} | x_1, \dots, x_n)$  is the same as maximizing the likelihood  $L$  integrated over the parameter space

$$\mathbb{P}(x_1, \dots, x_n | \mathcal{M}_{G,C}) = \int_{\text{Stab}_p^C(E)} L(M | x_1, \dots, x_n) g_{G,C}(M) dM$$

where we denote by  $\text{Stab}_p^C(E)$  the set of parameters and by  $g_{G,C}(M)$  the density function of the parameters associated with model  $\mathcal{M}_{G,C}$ . Taking the logarithm and applying a second-order Taylor series expansion, we arrive at

$$\log(\mathbb{P}(x_1, \dots, x_n | \mathcal{M}_{G,C})) = \log L(\hat{M} | x_1, \dots, x_n) - \frac{|E| + p}{2} \log n,$$

where  $L(\hat{M} | x_1, \dots, x_n)$  is the maximized likelihood function. In the literature, the BIC criterion is ultimately defined as the minimizing  $-2 \log(\mathbb{P}(x_1, \dots, x_n | \mathcal{M}_{G,C}))$ , that is

$$BIC(\mathcal{M}_{G,C}) = (|E| + p) \log n - 2 \log \hat{L}, \quad (3.22)$$

where  $\hat{L}$  is the abbreviation for  $L(\hat{M} | x_1, \dots, x_n)$ . For more details on the derivation, we refer to [Ghosh et al., 2007, Section 6.1.1].

To apply the BIC criterion to the path of Lasso solutions  $\hat{M}_1, \dots, \hat{M}_{100}$  defining the models  $\mathcal{M}_{G_1, 2I_p}, \dots, \mathcal{M}_{G_{100}, 2I_p}$ , we first minimize two times the negative Gaussian log-likelihood

$$L(M) = n(-\log \det((\Sigma(M, 2I_p))^{-1}) + \text{tr}(\hat{\Sigma}(\Sigma(M, 2I_p))^{-1}))$$

for these models. Then, we plug  $\hat{L}$  into (3.22) and select the model that has the minimal BIC score.

The BIC criterion has a tendency to produce estimates with too many edges, i.e., it produces false positives. This can be seen in our experiments with the Sachs data (Figure 3.14). Therefore, Chen and Chen [2008] propose an extended BIC criterion for model selection that takes into account both the number of unknown parameters and the complexity of the model space and is particularly useful for large model spaces with varying dimensions. We explain the issue with the BIC for Lyapunov models and provide a reasoning for the formulation of the extended BIC for Lyapunov models; a more general view on the issue is given in [Chen and Chen, 2008, Section 2].

The main problem of the BIC in its original formulation is that it is not equally likely to select models of different dimensions. There exist exactly  $p^2 - p$  models  $\mathcal{M}_{G,C}$  with  $|E| = 1$  while there already exist  $\binom{p^2-p}{2} = \frac{(p^2-p)!}{2!(p^2-p-2)!} = \frac{(p^2-p)(p^2-p-1)}{2}$  models  $\mathcal{M}_{G,C}$  with  $|E| = 2$ . This strong increase in the number of models of the same dimensionality continues till  $|E| = \lfloor (p^2 - p)/2 \rfloor$ . To only consider identifiable models, we restrict ourselves to models with  $|E| \leq \lfloor (p^2 - p)/2 \rfloor$ . In particular, when the number of variables is high compared to the sample size, we have to aim for sparse models. In these cases, assigning much higher probabilities to more complex models is not desirable. We try to mitigate this problem by putting less weight on more complex models and more on sparse ones.

Note the way we count our models is such the diagonal elements in the drift matrix are always assumed to be present. This results for a model  $\mathcal{M}_{G,C}$  in the probability

$$\mathbb{P}(\mathcal{M}_{G,C}) = \frac{1}{\binom{p}{2}} \frac{1}{\binom{p^2-p}{|E|}} \propto \frac{1}{\binom{p^2-p}{|E|}}.$$

If the model dimension is not too big, we can approximate the binomial coefficient by

$$\binom{p^2-p}{|E|} \approx p^{2|E|}.$$

Using this approximation, we obtain

$$-2 \log(\mathbb{P}(\mathcal{M}_{G,C})) \approx 2 \log p^{2|E|} = 4|E| \log p.$$

Introducing an additional tuning parameter  $\gamma \in (0, 1)$  to regulate between the classical BIC ( $\gamma = 0$ ) and the full penalization of the extended BIC ( $\gamma = 1$ ), we obtain

$$EBIC_\gamma(\mathcal{M}_{G,C}) = (|E| + p) \log n + 4\gamma|E| \log p - 2 \log \hat{L}. \quad (3.23)$$

Both for the BIC and EBIC, we provide additional simulations with synthetic data in Appendix B.1 that indicate consistency for the undirected structure for these model selection methods.

### 3.7.3 Standardization of the Sachs Dataset

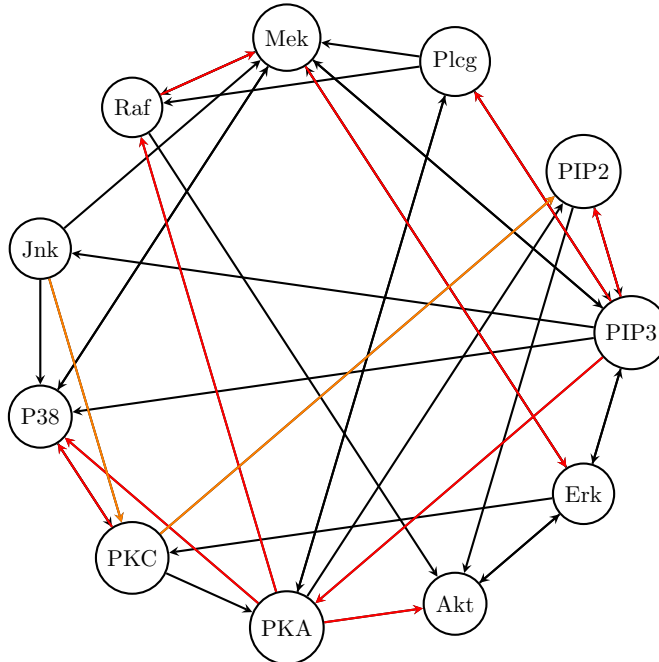
For one simulation, we consider one of the 14 datasets given by the matrix  $X \in \mathbb{R}^{n \times 11}$  where one row corresponds to one observation of the flow cytometry measurements and



the columns are the 11 phosphorylated proteins and phospholipids. We standardize every column of  $X$  by calculating

$$X_{:,i}^{std} = \frac{X_{:,i} - \mu}{\sigma}$$

where  $\mu$  is the mean and  $\sigma$  the standard deviation of  $X_{:,i}$ . This standardization has a practical motivation. The expression levels of the individual proteins differ in size. The observations for Mek are around 1, while those for PKA are in the thousands. Without standardization, we observe that some relevant connections present in the among scientist accepted network (“ground truth” by Sachs et al. [2005]) cannot be estimated. The reason is that these small entries are shrunk to zero by the  $\ell_1$ -penalty. The standardization allows to recover more connections of the ground truth network. A justification could lie in interpreting the standardization as a weighted penalty with the weights being dependent on the estimated covariances.



**Figure 3.14:** Estimated Sachs Network using the direct Lyapunov lasso and the BIC criterion (Dataset 7)

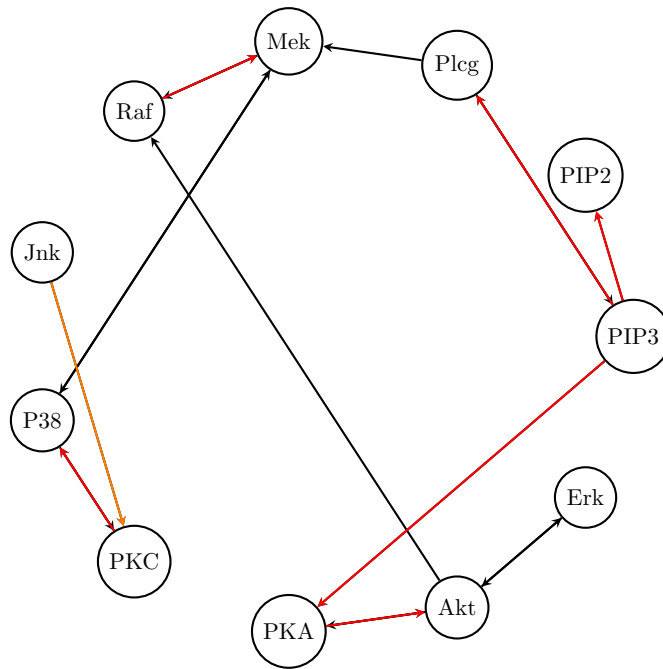
### 3.7.4 Estimation Results for the Sachs Dataset with the Direct Lyapunov Lasso and (Extended) BIC

In Figure 3.14 we present the estimate of the protein signalling network using dataset 7 in Table 3.3 and the BIC criterion for scoring. The datasets in Sachs et al. [2005] also contain a graph that shows the conventionally accepted signaling molecule interactions to which we refer as ground truth. A visualization of the ground truth is given in Figure 3.13. There exists doubt regarding some connections in this network; see, for instance, [Ramsey and Andrews, 2018, Section 2]. Therefore, there is some ambiguity in all comparisons between our estimates and the ground truth. Note that we do not draw self-loops in this section. The graph in Figure 3.14 shows all the edges estimated

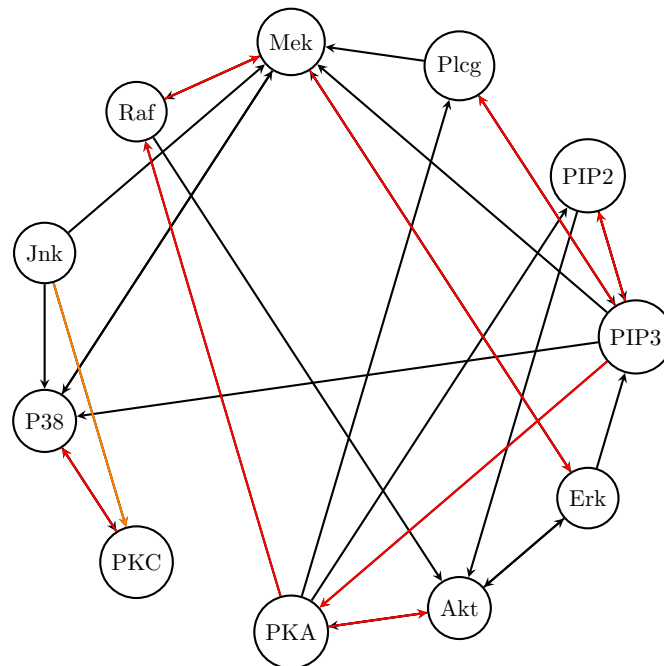
by the combination of the direct Lyapunov lasso and BIC. The red edges are the edges that are correctly recovered by the estimate. The orange edges are the edges where the reversed edges are present in the ground truth. The black edges are additional edges that are not present in the ground truth. The ground truth has 20 edges, while our estimate has 37 edges. This drastic difference is not immediately visible when comparing the networks. The reason is that our estimate also contains a few 2-cycles that are only present between PIP3 and PKA in the ground truth. Only counting the edges of the skeleton, i.e. not counting the 2-cycles twice, there exist 27 connections in the estimated network. The authors Sachs et al. [2005] categorize the connections in the network by how certain they are. Subsequently, they provide detailed explanations of well-established mechanisms that we use to evaluate our estimate. Among the correctly estimated connections are the direct enzyme-substrate relationships  $\text{PKA} \rightarrow \text{Raf}$ ,  $\text{Raf} \rightarrow \text{Mek}$ ,  $\text{Mek} \rightarrow \text{Erk}$  and  $\text{PIP3} \rightarrow \text{Pcl-}\gamma$ . Only the connection  $\text{Pcl-}\gamma \rightarrow \text{PIP2}$  is missing in Figure 3.14, but the pathway  $\text{Pcl-}\gamma \rightarrow \text{PIP3} \rightarrow \text{PIP2}$  suggests the presence of this interaction. Another interesting relationship is the influence of PKC and PKA on P38 and Jnk. We observe that the influence of PKC and PKA on P38 is correctly estimated, while the influence of PKC on Jnk is reversed in the estimated network. Of course, there exist also parts of the network that are not correctly captured by our estimate. In particular, in our estimate, PKA has outgoing edges to PIP2, Akt, P38 and Raf while only the edges  $\text{PKA} \rightarrow \text{Akt}$ ,  $\text{PKA} \rightarrow \text{P38}$  and  $\text{PKA} \rightarrow \text{Raf}$  are also present in the ground truth and the edges  $\text{PKA} \rightarrow \text{Erk}$ ,  $\text{PKA} \rightarrow \text{Mek}$  and  $\text{PKA} \rightarrow \text{PIP3}$  are not present in our estimate, but in the ground truth. Overall, Lasso together with BIC for scoring provides a reasonable estimate that recovers many important connections in the network, but tends to produce estimates with too many edges.

We introduce the EBIC criterion in Section 3.7.2 as a possibility to deal with the tendency of the BIC criterion to include too many variables into the estimate. With this new criterion, we revisit dataset 7 in Table 3.3. The steps are exactly the same as for the classical BIC, but we use the extended BIC (3.23) in the final step for scoring. The model that minimizes  $EBIC_\gamma$  with  $\gamma = 1$  is selected. The estimated graph is displayed in Figure 3.15. The estimate has 17 edges and 11 connections when counting the 2-cycles only once. Among the correctly estimated connections are the direct enzyme-substrate relationships  $\text{Raf} \rightarrow \text{Mek}$  and  $\text{PIP3} \rightarrow \text{Pcl-}\gamma$ . The connections  $\text{Pcl-}\gamma \rightarrow \text{PIP2}$  and  $\text{PKA} \rightarrow \text{Raf}$  are missing in Figure 3.15, but the  $\text{Pcl-}\gamma \rightarrow \text{PIP3} \rightarrow \text{PIP2}$  and  $\text{Pcl-}\gamma \rightarrow \text{PIP3} \rightarrow \text{PIP2}$  pathways suggest the presence of these interactions. Only the connection  $\text{Mek} \rightarrow \text{Erk}$  is not present at all. In general, direct Lyapunov lasso with extended BIC is an intuitive and easy-to-implement method that produces a sparse estimate with most edges (or their reverse) present in the ground truth, and even some additional edges such as  $\text{Akt} \rightarrow \text{Raf}$  can be interpreted as connecting pieces of meaningful pathways.

Based on this experiment, we observe that both BIC and the extended BIC have their advantages and disadvantages. While the BIC leads to an estimate that contains more true connections among its edges, it also produces many false positives. The estimate using the extended BIC has fewer edges and also misses a couple of well-established connections, but produces fewer false positives. Using the tuning parameter  $\gamma \in (0, 1)$  one can try to find the “sweet spot”. We present an example for  $\gamma = 1/2$  in Figure 3.16.



**Figure 3.15:** Estimated Sachs Network using the direct Lyapunov lasso and the extended BIC criterion with  $\gamma = 1$  for scoring (Dataset 7)



**Figure 3.16:** Estimated Sachs Network using the direct Lyapunov lasso and the extended BIC criterion with  $\gamma = 1/2$  for scoring (Dataset 7)

This estimate contains most of the important edges of Figure 3.14, but does not produce as many false positives.

### 3.8 Summary of the Chapter

We investigated model selection properties of the direct Lyapunov lasso when applied to data distributed according to the graphical continuous Lyapunov model. Although the optimization problem that direct Lyapunov lasso solves is similar to the lasso penalized linear regression objective, there are several surprising differences. The role of the matrix  $A(\Sigma)$ , which is analogous to the design matrix in the regression setting, is subtle under the Lyapunov model. We established a reasonable bound on sample complexity by careful investigation of the Hessian matrix whose elements are sums of  $p$  products of covariances. Furthermore, while the irrepresentability condition can be assumed in the linear regression setting, this is not the case for the model considered here. Indeed, our detailed analysis of the irrepresentability condition illustrates the reasons for its restrictiveness. We formulated conditions under which the irrepresentability condition is guaranteed to hold for DAGs based on the topological ordering of the nodes and provided insight into why a similar result is difficult to obtain in the presence of cyclic structures. Simulations further provided evidence to the extent to which one can hope that the irrepresentability condition is satisfied for a randomly drawn signals. In fact, we are not able to present a drift matrix supported over a cyclic graph that fulfills the irrepresentability condition. We show that a slightly weaker notion of the irrepresentability is necessary for asymptotic support recovery. This condition is fulfilled much more often than the irrepresentability condition and we are even able to present drift matrices fulfilling the weak irrepresentability condition for all simple cyclic graph up to 4 nodes. Interestingly, the weak notion already has a strong effect on the quality of the estimation results. Despite the irrepresentability conditions being rarely fulfilled for randomly selected drift matrices and the problem of misspecification of the volatility matrix when applying the direct Lyapunov lasso, we showed that the method is quite robust and performs decently in seemingly unfavorable settings. We investigated the direct Lyapunov lasso alongside with the (Extended) Bayesian Information Criterion to select specific drift matrices along a regularization grid. We observe that the undirected structure is recovered very well, but the tendency to produce estimates with a lot of symmetry prevents fully satisfying results regarding the directed structure. We applied this combination of direct Lyapunov lasso and (Extended) Bayesian Information Criterion onto the Sachs dataset. Despite the mentioned issues with symmetry, the method manages to recover important structures of a protein-signaling network purely based on observational data.

## Chapter 4

# Beyond the Lasso - Best Subset Selection with Mixed Integer Programming

Recently, Bertsimas et al. [2016] demonstrated with the help of more efficient solvers that the best subset selection represents an alternative to the classic  $\ell_1$ -penalized lasso regression. They connect the best subset selection to Mixed Integer Quadratic Programs (MIQP) for which efficient solvers exist [Gurobi Optimization, LLC, 2023]. The authors even claim clear superiority of the best subset method over the lasso in all circumstances. This ambitious view is put into perspective by Hastie et al. [2020b], and a more nuanced picture emerges. Nevertheless, the advantages of the best subset method raised interest from statisticians working in the field of graphical modeling. For instance, Gao et al. [2023] consider a variant of the best subset selection in the context of structural equation models. Additionally, there is an ongoing effort to improve the algorithms solving best subset problems [Zhu et al., 2020].

This motivates applying the best subset selection to graphical continuous Lyapunov models. In this chapter, we show that the method can be used when aiming to estimate the drift matrix and also when estimating the drift and the volatility matrix. When setting up the best subset selection for estimating both matrices, we obtain a variant of the direct Lyapunov lasso (3.2) that is also able to estimate the volatility matrix as a by-product. However, the main focus lies on the best subset selection.

### 4.1 Difficulties of Existing Methods

As introduced, this chapter aims to present a new method for model selection for graphical continuous Lyapunov models based on the best subset selection. Using concrete examples, we illustrate that this method can offer advantages over the  $\ell_1$ -penalized methods direct Lyapunov lasso (3.2) and the likelihood-based method by Varando and Hansen [2020]. In particular, the estimates of the  $\ell_1$ -penalized methods contain a lot of symmetry.

#### 4.1.1 Revisiting Existing Methods

First, we briefly revisit the existing methods for structure learning for GCLMs. All methods rely on an estimated version of the covariance matrix which is given by the

sample covariance matrix

$$\hat{\Sigma} = \hat{\Sigma}^{(n)} = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top \quad (4.1)$$

in all instances. In order to connect the estimation of sparse graphs with sparse regression, it is helpful to consider the vectorized Lyapunov equation

$$A(\Sigma)\text{vec}(M) + \text{vec}(C) = 0, \quad (4.2)$$

where  $A(\Sigma) := (\Sigma \otimes I_p) + (I_p \otimes \Sigma)K^{(p,p)}$  is a  $p^2 \times p^2$  matrix with covariances as entries. We denote by  $K^{(p,p)}$  the  $p^2 \times p^2$  commutation matrix and by  $\text{vec}(\cdot)$  the vec-operator. For more details, we refer to Section 1.4. We show that the direct Lyapunov lasso

$$\arg \min_{M \in \mathbb{R}^{p \times p}} \frac{1}{2} \|A(\Sigma)\text{vec}(M) + \text{vec}(C)\|_2^2 + \lambda \|M\|_1 \quad (4.3)$$

is an intuitive and easy to implement method to perform model selection for GCLMs. The convexity of the optimization problem allows for consistency results as Corollary 3.3.3. However, we also mention the downsides of the approach. The weak irrepresentability condition is necessary for consistent support recovery, see Proposition 3.5.6. In particular, consistent support recovery is only proven if the irrepresentability condition holds, see Corollary 3.3.3.

Recall the irrepresentability condition (3.15)

$$\|\Gamma_{S^c S}^* (\Gamma_{SS}^*)^{-1}\|_\infty < 1 - \alpha \quad (4.4)$$

and the weak irrepresentability condition (3.18)

$$\|\Gamma_{S^c S}^* (\Gamma_{SS}^*)^{-1} \text{sign}(\text{vec}(M^*))_S\|_\infty \leq 1. \quad (4.5)$$

Both conditions turn out to be extremely restrictive. For further details and numerical experiments we defer to Section 3.5.3. Subsequently, we show that a variant of the best subset selection presented by Bertsimas et al. [2016] is able to recover the correct support even in the unfavorable settings where it is theoretically not possible for the direct Lyapunov lasso.

We also revisit one of the optimization problems by Varando and Hansen [2020]. The loss function considered is the negative Gaussian log-likelihood

$$L(M, C) = \log \det(\Sigma(M, C)) + \text{tr}(\hat{\Sigma} \Sigma(M, C)), \quad (4.6)$$

where  $\Sigma(M, C)$  is the solution to the Lyapunov equation (1.2) for given matrices  $M$  and  $C$ . For the purpose of variable selection, an  $\ell_1$ -penalty is added. In addition, a penalty that regulates how close the estimated  $C$  is to the identity is included. The optimization problem is given by

$$\begin{aligned} \text{argmin } & L(\Sigma(M, C)) + \lambda \|\text{vec}(M)\|_1 + \kappa \|\text{vec}(C) - \text{vec}(I_p)\|_F^2 \\ \text{s.t. } & M \text{ stable and } C \text{ diagonal.} \end{aligned} \quad (4.7)$$

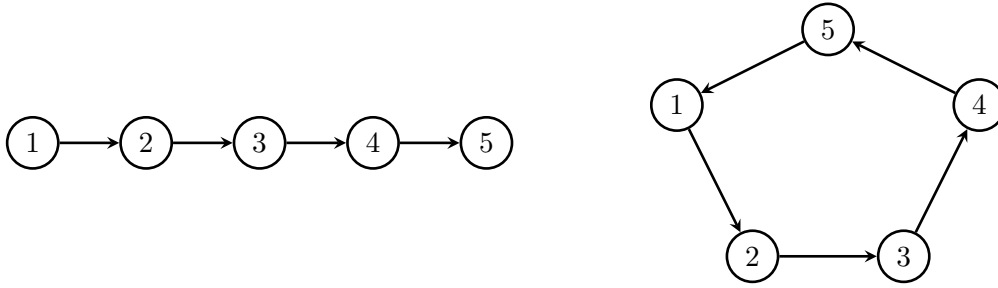
### 4.1.2 Main Examples

Without discussing the method in detail at this point, we present concrete examples to show that the best subset selection for Lyapunov Models (BSSLM) can offer advantages over the direct Lyapunov lasso and over the method of Varando and Hansen [2020].

First, we consider the path from 1 to 5 displayed in Figure 4.1. We choose the entries of the drift matrix supported over the path to be

$$M_{\text{path}}^* = \begin{pmatrix} -2 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 & -2 \end{pmatrix}. \quad (4.8)$$

Then, we use the Lyapunov equation (1.2) with  $C = 2I_p$  to calculate the equilibrium covariance matrix  $\Sigma_{\text{path}}^*$ . We do not sample data from the equilibrium distribution  $\mathcal{N}(0, \Sigma_{\text{path}}^*)$ , but directly use  $\Sigma_{\text{path}}^*$  for estimation mimicking the setting of “ $n = \infty$ ”. We apply three estimation methods for model selection.



**Figure 4.1:** **Left:** The path 1 to 5. **Right:** The 5-cycle.

For the direct Lyapunov lasso we calculate the estimates along the  $\lambda$ -grid

$$0 < \frac{\lambda_{\max}}{10^2} = \lambda_1 < \dots < \lambda_{100} = \lambda_{\max} \quad (4.9)$$

with  $\lambda_{\max}$  being the minimal  $\lambda$ -value such that  $M$  is diagonal. For the Gaussian likelihood based method “loglik-0.01” of Varando and Hansen [2020], we use  $\hat{M}_0 = -0.5 \cdot (\Sigma_{\text{path}}^*)^{-1}$  as initialization and a lambda grid

$$0 < \frac{\lambda_{\max}}{10^4} = \lambda_1 < \dots < \lambda_{200} = \lambda_{\max}$$

with  $\lambda_{\max}$  being again the minimal  $\lambda$ -value such that  $M$  is diagonal. Further details on the method are given in [Varando and Hansen, 2020]. Finally, we apply the best subset method where the number of active variables is directly controlled by a sparsity tuning parameter  $k$ . We calculate the estimates for values of  $k = 1, \dots, 15$ . The diagonal is always included and a value of  $k = 1$  results in one off-diagonal element being selected. For all methods and estimates along the regularization paths, we check if at least one estimate on the path fulfills  $fp = fn = 0$  (Definition 3.5.10), i.e. that the support of the data generating  $M_{\text{path}}^*$  is correctly recovered.

In general, all calculations are deterministic, however there are some aspects making the calculations for the best subset approach non-deterministic. This is due to the time-limit and current machine workload, for more details we refer to the documentation of Gurobi [Gurobi Optimization, LLC, 2023]. Therefore, we carry out the calculations for the best subset method 100 times and count how often we obtain a perfect result.

**Example 4.1.1.** *Initially, we calculate the irrepresentability condition (3.15) and the weak irrepresentability condition (3.18). For the left side of the irrepresentability condition we obtain a value of 1.726225 and for the weak irrepresentability condition a value of 1.208417. Both conditions are violated. In fact, the direct Lyapunov lasso (3.2) is not able to recover the support of  $M_{path}^*$  in our simulations. The main issue of the estimates  $\hat{M}_{\lambda_1}^{DL}, \dots, \hat{M}_{\lambda_{100}}^{DL}$  is that a nonzero off-diagonal entry in position  $(i, j)$  often results in a nonzero off-diagonal entry in position  $(j, i)$ . Below we present the first three estimates of the direct Lyapunov lasso starting with the regularization parameter  $\lambda_{100}$  that yields the diagonal matrix. Already for  $\lambda_{98}$  and small absolute value of the off-diagonal elements, we observe that both the entry  $(2, 3)$  and the entry  $(3, 2)$  are included in the support. The entries are marked in red. We have*

$$\hat{M}_{\lambda_{100}}^{DL} = \begin{pmatrix} -1.61 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & -1.51 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & -1.52 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & -1.58 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -1.88 \end{pmatrix},$$

$$\hat{M}_{\lambda_{99}}^{DL} = \begin{pmatrix} -1.63 & 0.04 & 0.00 & 0.00 & 0.00 \\ 0.00 & -1.51 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & -1.52 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & -1.58 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -1.88 \end{pmatrix},$$

$$\hat{M}_{\lambda_{98}}^{DL} = \begin{pmatrix} -1.64 & 0.08 & 0.00 & 0.00 & 0.00 \\ 0.00 & -1.52 & \mathbf{0.005} & 0.00 & 0.00 \\ 0.00 & \mathbf{0.03} & -1.54 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.03 & -1.59 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -1.88 \end{pmatrix}.$$

The pattern is observed along the whole path. The estimate for  $\lambda_{50}$  is

$$\hat{M}_{\lambda_{50}}^{DL} = \begin{pmatrix} -1.81 & \mathbf{0.32} & 0.00 & 0.00 & 0.00 \\ \mathbf{0.57} & -1.91 & \mathbf{0.20} & 0.00 & 0.00 \\ 0.00 & \mathbf{0.68} & -1.92 & \mathbf{0.08} & 0.00 \\ 0.00 & 0.00 & \mathbf{0.78} & -1.96 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.78 & -2.18 \end{pmatrix}.$$

Except for the entry  $(4, 5)$ , all other entries on the subdiagonal have their counterpart on the superdiagonal. The estimates obtained by the direct Lyapunov lasso are not



#### 4.1 Difficulties of Existing Methods

entirely bad as the skeleton (i.e. the undirected structure) is estimated correctly. One might even argue that the estimate  $\hat{M}_{\lambda_{50}}^{DL}$  contains the correct directionality of the edges by comparing the magnitude of the subdiagonal and the superdiagonal. All entries in the subdiagonal are smaller than those in the superdiagonal.

Subsequently, we investigate the other  $\ell_1$ -penalized method by Varando and Hansen [2020].

**Example 4.1.2.** Similarly, the “loglik-0.01” does not produce an optimal result along the regularization path. Below we present the most interesting segment. The first nonzero off-diagonal elements appear for  $\lambda_{189}$ . As for the direct Lyapunov lasso, the estimate has nonzero entries on the sub- and superdiagonal. In this case only (1, 2), (2, 1). The best estimate in terms of support is then the next estimate on the path for  $\lambda_{188}$  as it contains the full nonzero subdiagonal and the only wrong entry is (1, 2). We have

$$\hat{M}_{\lambda_{190}}^{ML} = \begin{pmatrix} -1.80 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & -1.88 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & -1.88 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & -1.91 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -2.07 \end{pmatrix},$$

$$\hat{M}_{\lambda_{189}}^{ML} = \begin{pmatrix} -1.38 & \mathbf{0.21} & 0.00 & 0.00 & 0.00 \\ \mathbf{0.25} & -1.05 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.44 & -1.21 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.46 & -1.64 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -2.06 \end{pmatrix},$$

$$\hat{M}_{\lambda_{188}}^{ML} = \begin{pmatrix} -1.34 & \mathbf{0.20} & 0.00 & 0.00 & 0.00 \\ \mathbf{0.25} & -0.96 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.42 & -1.14 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.47 & -1.59 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.28 & -2.07 \end{pmatrix}.$$

Progressing the path, we observe that more entries in the superdiagonal of the estimate are included in its support. For instance, consider the estimate for  $\lambda_{130}$  below

$$\hat{M}_{\lambda_{130}}^{ML} = \begin{pmatrix} -1.26 & \mathbf{0.22} & 0.00 & 0.00 & 0.00 \\ \mathbf{0.32} & -0.79 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.43 & -0.96 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.57 & -1.40 & \mathbf{0.02} \\ 0.00 & 0.00 & 0.00 & \mathbf{0.67} & -2.00 \end{pmatrix}.$$

If we further continue the path, we observe nonzero entries not even matching the skeleton of the path.

In the spirit of the direct Lyapunov lasso, we tried to find a concrete counterexample where perfect support recovery is not possible. However, the non-convexity of the optimization problem makes theoretical analysis extremely difficult. We provide insights in Section 4.1.3.

Third, we present the results of the best subset method.

**Example 4.1.3.** For 94 out of 100 attempts, the best subset method produces an estimate that recovers the support of  $M_{path}^*$  correctly. Of course, we obtain these estimates for  $k = 4$  and it is even true that

$$\hat{M}_{k=4}^{BS} = M_{path}^* = \begin{pmatrix} -2 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 & -2 \end{pmatrix}.$$

Finally, we carry out the same calculations, but consider a drift matrix that is supported over the 5-cycle in Figure 4.1. We choose the same order of magnitude for the entries in the drift matrix as for the path which results in

$$M_{cycle}^* = \begin{pmatrix} -2 & 0 & 0 & 0 & 1 \\ 1 & -2 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 & -2 \end{pmatrix}.$$

Running the best subset method we obtain 52 estimates that recover the support and the magnitude of the entries while 48 do not estimate the support correctly. These estimates are mostly the cycle in the reverse direction. The reason is that

$$\Sigma(M_{cycle}^*, C) = \Sigma((M_{cycle}^*)^\top, C).$$

Obviously, if two drift matrices with different underlying graph structure and the same number of edges yield the same covariance matrix, no method can guarantee to recover the support correctly. However, the best subset method does what it is expected to do. The other two methods fail by estimating to recover the support of the cycle correctly. They do not fail by estimating the reversed cycle, but a similar problem as for the path occurs.

The results in this section are in line with the observations in Appendix B.1 where we observe that the direct Lyapunov lasso is much better in recovering the undirected structure than the directed structure. In summary, we find that the best subset method leads to predominantly very good results. It is able to recover the support of the data generating drift matrices correctly while the direct Lyapunov lasso and the loglik-0.01 fail to do so. Natural limitations due to model geometry exist.

Obviously, these are very specific examples that do not allow any general conclusions to be drawn. However, they definitely provide motivation to study the method in more detail. Before we study this further, we explain the difficulties of the theoretical analysis of the method by Varando and Hansen [2020].

### 4.1.3 KKT-Conditions for the Loglik-Method

For the direct Lyapunov lasso (3.2), the KKT-conditions are a necessary and sufficient criterion for support recovery, see Theorem 3.3.1. The optimization problem (4.7) is non-convex. Therefore, the KKT-conditions are not sufficient, but they are still necessary for optimality. The optimization problem considered by Varando and Hansen [2020] in matrix notation is

$$\underset{(M,C)}{\operatorname{argmin}} \log \det \Sigma + \operatorname{tr}(\hat{\Sigma}\Sigma^{-1}) + \lambda\|M\|_1 + \kappa\|C - I\|_F^2. \quad (4.10)$$

The results for this method in Section 4.1.2 suggest that perfect support recovery is not possible for the drift matrix presented in (4.8) and for  $C = 2I_5$ . Unlike for the direct Lyapunov lasso where calculating the irrepresentability conditions suffices, more work is required for the method by Varando and Hansen [2020]. In this section, we present two intriguing examples. First, we show in Example 4.1.5 that deciding whether the KKT-conditions are fulfilled is already extremely difficult for  $p = 3$ . In Example 4.1.6 we present a  $2 \times 2$  example where perfect support recovery is not possible for the direct Lyapunov lasso, but a solution with the correct support for the loglik- $\kappa$  method exists (4.10). First, we have to calculate all derivatives for the summands of (4.10). They are given in Lemma C.0.3. Using the derivative, the following fact is an immediate consequence.

**Remark 4.1.4.** *Let  $G$  be a non-empty directed graph and let  $M$  be a stable matrix that is supported over  $G$ . Furthermore, we consider  $C \in \operatorname{PD}_p$ . Using the true covariance matrix  $\Sigma = \Sigma(M, C)$ , the KKT-conditions of (4.10) are not fulfilled for the pair  $(M, C)$ .*

The natural follow-up question is to ask if there exists another drift matrix supported over the same graph that fulfills the KKT-conditions. We showcase how difficult it is to solve this question for a  $3 \times 3$  matrix.

**Example 4.1.5.** *Consider the data generating drift and volatility matrix*

$$M = \begin{pmatrix} -2 & 0 & 0 \\ 1 & -2 & 0 \\ 0 & 1 & -2 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad (4.11)$$

where we choose  $\hat{\Sigma} = \Sigma^* = \Sigma(M, C)$ . Is it possible that a pair

$$M_{\text{var}} = \begin{pmatrix} -d_1 & 0 & 0 \\ m_1 & -d_2 & 0 \\ 0 & m_2 & -d_3 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad (4.12)$$

fulfills the KKT condition with  $d_1, \dots, d_3 > 0$ ? The subsequent computations are symbolic and can be carried out with various computer algebra programs. First, we calculate

$$(\nabla_{\Sigma} L)_{var} = (\Sigma(M_{var}, C))^{-1} - (\Sigma(M_{var}, C))^{-1} \hat{\Sigma} (\Sigma(M_{var}, C))^{-1}.$$

Then, we obtain the gradient of the log-likelihood-loss (C.3) w.r.t.  $M$  by

$$(\nabla_M L)_{var} = (2\Sigma(M_{var}, C)\Sigma(M_{var}^T, (\nabla L)_{var})).$$

This is a (very) long rational expression in the variables  $d_1, d_2, d_3, m_1, m_2$ . The KKT-conditions (sum of the three parts: gradient loss + subgradient drift matrix + gradient of squared  $\ell_2$ -norm for  $C$ , see Lemma C.0.3) are zero if all entries in  $(\nabla_M L)_{var}$  are the same (up to sign) such that they can be shrunk to zero by the subgradient (C.5). The subgradient is  $\lambda \text{sign}(M_{ij})$  if  $M_{ij} \neq 0$ . Therefore, the entries  $(1, 2), (2, 3), (1, 1), (2, 2), (3, 3)$  of  $(\nabla_M L)_{var}$  need to have the same absolute value. If one does not penalize the diagonal, it reduces to  $(1, 2), (2, 3)$  being the same.

In addition, the subgradient of the remaining entries is given by  $\lambda[-1, 1]$  if  $M_{ij} = 0$ . That means the absolute values of the entries  $(1, 3), (2, 1), (2, 3), (3, 1)$  of  $(\nabla_M L)_{var}$  need to be smaller than the absolute value of the entries  $(1, 2), (2, 3), (1, 1), (2, 2), (3, 3)$ . If we find choices  $d_1, d_2, d_3, m_1, m_2$  such that all of this holds, a matrix  $M_{var}$  fulfills the KKT-conditions. If one does not penalize the diagonal, the entries  $(1, 1), (2, 2), (3, 3)$  of  $(\nabla_M L)_{var}$  need to be zero. Still the entries  $(1, 3), (2, 1), (2, 3), (3, 1)$  of  $(\nabla_M L)_{var}$  need to be smaller than the absolute value of the entries  $(1, 2), (2, 3)$ .

We try to find parameters  $d_1, d_2, d_3, m_1, m_2$  such that the pair  $(M, C)$  displayed in (4.12) fulfills the KKT-conditions. Here, we write out the conditions when not penalizing the diagonal. We denote the entry  $(i, j)$  in  $(\nabla_M L)_{var}$  by  $G_{i,j}$ . The KKT-conditions imply that

$$\begin{aligned} G_{1,1} &= 0, \quad G_{2,2} = 0, \quad G_{3,3} = 0 \quad \text{and} \quad G_{2,1} = G_{3,2} \\ G_{1,2} &< |G_{2,1}|, \quad G_{1,3} < |G_{2,1}|, \quad G_{2,3} < |G_{2,1}| \quad \text{and} \quad G_{3,1} < |G_{2,1}| \\ G_{1,2} &> -|G_{2,1}|, \quad G_{1,3} > -|G_{2,1}|, \quad G_{2,3} > -|G_{2,1}| \quad \text{and} \quad G_{3,1} > -|G_{2,1}|. \end{aligned}$$

Additionally, it has to hold that  $d_1, d_2, d_3 > 0$ . At first glance it might seem surprising, but all attempts solving this system failed due to its computational complexity. A reason why these problems are so hard to solve is given in [Basu et al., 2006, Chapter 11]. Quantifier Elimination which is used for solving this problem can at worst be doubly exponential. We want to mention that it is computationally feasible to analyze the problem when setting the diagonal entries to a fixed value. We consider

$$M_{var} = \begin{pmatrix} -2 & 0 & 0 \\ m_1 & -2 & 0 \\ 0 & m_2 & -2 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad (4.13)$$

where in fact no choice of parameters  $m_1, m_2 \in \mathbb{R}$  exists such that the KKT-conditions are met.

## 4.2 Best Subset Selection with MIQP for Lyapunov Models

The previous example shows how hard it is to analyze problems as small as  $p = 3$ . Although the setup is very idealistic, we present a  $2 \times 2$  example where we make an interesting observation.

**Example 4.1.6.** *Equivalently to the  $3 \times 3$  case, we consider*

$$M = \begin{pmatrix} -2 & 0 \\ 1 & -2 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

*The left side of the irrepresentability condition (3.15) has a numeric value of 1.625 and the left side of the weak irrepresentability condition (3.18) has a value of 1.125. Therefore, the KKT-conditions for the direct Lyapunov lasso cannot be fulfilled. Can we find parameters  $d_1, d_2, m_1$  such that*

$$M_{var} = \begin{pmatrix} -d_1 & 0 \\ m_1 & -d_2 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

*fulfill the KKT-conditions for the loglik- $\kappa$  method with  $M_{var}$  being stable? First, we calculate  $(\nabla_M L)_{var}$  with the entries being rational functions in the variables  $d_1, d_2, m_1$ . The KKT-conditions imply that*

$$\begin{aligned} G_{1,1} &= 0 \text{ and } G_{2,2} = 0, \\ G_{1,2} &< |G_{2,1}|, \\ G_{1,2} &> -|G_{2,1}|. \end{aligned}$$

*More details on the notation are given in Example 4.1.5. Additionally it has to hold that  $d_1, d_2 > 0$ . This system can be solved and the solution is given by*

$$\tilde{M} \approx \begin{pmatrix} -1.98438 & 0 \\ 0.149253 & -1.79699 \end{pmatrix}.$$

*To illustrate that the computations are already surprisingly complex for the  $2 \times 2$  case, the unrounded expression for  $m_1$  is given by*

$$m_1 = \frac{48678659497325 - 9047381575\sqrt{13860921}}{100467168307456}.$$

*Here a solution with the same support exists.*

Overall, the method by Varando and Hansen [2020] is extremely hard to analyze theoretically as already the seemingly trivial structure of the path from 1 to 3 results in very difficult and lengthy rational functions in the gradient of the loss function.

## 4.2 Best Subset Selection with MIQP for Lyapunov Models

In this section, we introduce the best subset selection for Lyapunov models. The general setup is similar to the work by Bertsimas et al. [2016], who considered the best subset selection in regression settings. However, there exist a lot of subtle differences that we solve. We explain how to phrase the problem such that it is suitable for large-scale numerical computations using the commercial optimizer Gurobi [Gurobi Optimization, LLC, 2023]. We consider different warm starts and compare them

both with respect to the time consumption and result-wise. Ultimately, we compare the method with the  $\ell_1$ -penalized methods in various simulation settings. There, we observe quite a few differences when compared to the regression setting.

We already know from the direct Lyapunov lasso (3.2) that optimization problems with an objective function being the penalized squared Frobenius norm of the continuous Lyapunov equation (1.2) are computationally identical to classical sparse regression problems. Similar to the best subset problem (1.1) of Bertsimas et al. [2016], we can formulate the optimization problem

$$\begin{aligned} \arg \min_{M \in \mathbb{R}^{p \times p}} \|A(\Sigma)\text{vec}(M) + \text{vec}(C)\|_2^2 \\ \text{s.t. } \|\text{vec}(M)\|_{0,\text{off}} \leq k \end{aligned} \quad (4.14)$$

with

$$\|\text{vec}(M)\|_{0,\text{off}} = \sum_{\substack{i,j=1 \\ i \neq j}}^p \mathbb{1}_{\{M_{i,j} \neq 0\}}.$$

for GCLMs. From a dynamical systems perspective, many stable signals possess negative diagonal entries that are (possibly) significantly larger than the off-diagonal entries. This motivates only penalizing the off-diagonal elements and ensuring they are selected more carefully. The problem (4.14) is non-convex. Moreover, the best subset problems are shown to be NP-hard in [Natarajan, 1995]. According to the current knowledge, there is no solution to these problems in polynomial time (Presumption:  $\text{NP} \neq \text{P}$ ). Despite this limitation, researchers are searching for algorithms that solve these problems with high probability in a reasonable amount of time. The authors Bertsimas et al. [2016] show that the classical best subset selection can be phrased as a MIQP. More details on Mixed Integer Optimization are given by Bertsimas and Weismantel [2005]. A general form of a MIQP with binary variables is

$$\begin{aligned} \min_x x^\top Qx + c^\top x \\ \text{s.t. } \begin{cases} Ax \leq b, \\ x_i \in \{0, 1\} & i \in I, \\ l \leq x_i \leq r & i \notin I, \end{cases} \end{aligned} \quad (4.15)$$

where  $A \in \mathbb{R}^{n \times p}$ ,  $c \in \mathbb{R}^p$ ,  $b \in \mathbb{R}_\infty^n$ ,  $l \in \mathbb{R}_\infty^p$ ,  $r \in \mathbb{R}^p$  and  $Q \in \mathbb{R}^{p \times p}$  positive semidefinite are the parameters. The variables  $x_i$  indexed by  $I$  are binary, while those not in  $I$  are continuous. There are several solvers that can deal with problems of this type. Explicitly, we want to mention the optimizers SCIP [Bestuzheva et al., 2021], CPLEX [Cplex, 2009] and Gurobi [Gurobi Optimization, LLC, 2023]. We use the latter in this work. To our knowledge, Gurobi is currently still one of the fastest software packages to solve MIQPs of the form (4.15). In [Bertsimas et al., 2016, Section 2.1], the development of MIO solvers till 2013 is explained. A detailed discussion about the ins and outs of the solvers for these problems would exceed the scope of this work.

Naturally, the question arises what the connection of MIQPs and problem (4.14) is as the variables  $\text{vec}(M)$  are all continuous. We can formulate (4.14) as MIQP.

**Proposition 4.2.1.** *We consider the binary variables  $z_{(i,j)} \in \{0, 1\}$  with  $i, j \in \{1, \dots, p\}$  and  $i \neq j$ . Then, a solution to*

$$\begin{aligned} \arg \min_{M, z} & \|A(\Sigma)\text{vec}(M) + \text{vec}(C)\|_2^2 & (4.16) \\ \text{s.t.} & \begin{cases} -\mathcal{B}_U z_{(i,j)} \leq \text{vec}(M)_{(i,j)} \leq \mathcal{B}_U z_{(i,j)} & \text{for } i, j \in 1, \dots, p \text{ with } i \neq j, \\ z_{(i,j)} \in \{0, 1\} & \text{for } i, j \in 1, \dots, p \text{ with } i \neq j, \\ \sum_{\substack{i,j=1 \\ i \neq j}}^p z_{(i,j)} \leq k \end{cases} \end{aligned}$$

is also a solution to (4.14) for a sufficiently large constant  $\mathcal{B}_U$ .

**Proof.** The idea behind this formulation, presented in Bertsimas et al. [2016], is that if  $\text{vec}(\hat{M})$  is a solution to (4.16), we have that  $\|\text{vec}(\hat{M})\|_{\infty, \text{off}} \leq \mathcal{B}_U$ . The binary variables possess the function of “selecting” variables. If  $z_{(i,j)} = 1$ , then  $|\text{vec}(\hat{M})_{(i,j)}| \leq \mathcal{B}_U$  and if  $z_{(i,j)} = 0$ , then  $\text{vec}(\hat{M})_{(i,j)} = 0$ . Choosing  $\mathcal{B}_U$  sufficiently large, the solution of (4.16) is also a solution of (4.14).  $\square$

Naively, one might think that choosing a large  $\mathcal{B}_U$  and solving the problem is a good idea. However, this is not the case. A sensible choice of  $\mathcal{B}_U$  is necessary to obtain good lower bounds.

The authors Bertsimas et al. [2016] present two formulations of the problem (4.16) that are then used for the actual computations. The reason behind this is that the MIO solver of Gurobi Optimization, LLC [2023] deals best with an objective function that has a small-dimensional quadratic objective. Writing out the squared  $\ell_2$ -norm of problem (8) of Bertsimas et al. [2016], we obtain a quadratic objective involving the matrix  $X^\top X$  which is of size  $p \times p$  and might be large when considering high-dimensional settings ( $p > n$ ). Therefore, they use one formulation (2.5) that has a quadratic objective involving the matrix  $X^\top X$  for the cases when ( $p \leq n$ ) and a second formulation (2.6) that has a quadratic objective of size  $n \times n$  making use of the design matrix  $X$  having  $n$  rows for the cases ( $p > n$ ). For Lyapunov models, we consider the  $p^2 \times p^2$  matrix  $A(\Sigma)$  that takes on the role of the design matrix. No matter what the sample size is, we are “in between” these formulations. Therefore, we only consider the first formulation. Preliminary computations that are not included in this work show that there is no advantage. The second formulation seems to be slightly worse for Lyapunov models.

### 4.2.1 Solving the MIQP with Gurobi

In this section, we adapt formulation (2.5) of Bertsimas et al. [2016] in such a way that it is suitable to solve (4.14) and explain how to input it into Gurobi [Gurobi Optimization, LLC, 2023]. They use the notion of specifically ordered sets which ensures that at most  $k$  variables  $\text{vec}(M)$  can be nonzero. This can directly be passed to the Gurobi optimizer and leads to a specific problem structure. As this type of problem formulation is not conducive to this work, we forego it and stick to the classic formulation in the style of (4.16).

**Problem 4.2.2.** Adding a constraint on the  $\ell_1$ -norm of the off-diagonal entries in  $\text{vec}(M)$  and different constants for the diagonal and off-diagonal entries compared to (4.16), we formulate

$$\begin{aligned} \arg \min_{M,z} & \frac{1}{2} \text{vec}(M)^\top (A(\Sigma)^\top A(\Sigma)) \text{vec}(M) + \langle A(\Sigma)^\top \text{vec}(C), \text{vec}(M) \rangle \\ & + \frac{1}{2} \|\text{vec}(C)\|_2^2 \\ \text{s.t.} & \begin{cases} z_{(i,j)} \in \{0, 1\} \quad \text{for } i, j \in 1, \dots, p \quad \text{with } i \neq j, \\ \sum_{\substack{i,j=1 \\ i \neq j}}^p z_{(i,j)} \leq k, \\ -\mathcal{B}_U^{\text{off}} z_{(i,j)} \leq \text{vec}(M)_{(i,j)} \leq \mathcal{B}_U^{\text{off}} z_{(i,j)} \quad \text{for } i, j \in 1, \dots, p \quad \text{with } i \neq j, \\ -\mathcal{B}_U^{\text{diag}} \leq \text{vec}(M)_{(i,i)} \leq \mathcal{B}_U^{\text{diag}} \quad \text{for } i \in 1, \dots, p, \\ \|\text{vec}(M)\|_{1,\text{off}} \leq \mathcal{B}_l. \end{cases} \end{aligned} \quad (4.17)$$

We further comment on the exact choices of the constants  $B_U^{\text{off}}$ ,  $B_U^{\text{diag}}$  and  $B_l$  in Section 4.2.2. Provided suitable constants, a solution of (4.17) is also a solution of (4.14). To implement (4.17) in  $\mathbb{R}$  [R Core Team, 2021], we make use of code that has been used to implement the numerical experiments in Hastie et al. [2020b]. They implemented the exact problem formulations of Bertsimas et al. [2016] while there are a few differences in (4.17) that we want to mention here. Gurobi allows for a MIQP of the form

$$\begin{aligned} \min_x & x^\top Qx + q^\top x \\ \text{s.t.} & \begin{cases} Ax \text{ comp } b \quad \text{comp consists of } \leq, =, \\ l \leq x \leq u, \\ \text{some or all } x \text{ must take integer values.} \end{cases} \end{aligned} \quad (4.18)$$

**Lemma 4.2.3.** Problem 4.2.2 can be written in form of (4.18).

*Proof.* To apply the Gurobi optimizer, we restructure the problem. We define

$$\begin{aligned} \tilde{A}(\Sigma) &= (A(\Sigma)_{\cdot,(1,2),(1,3),\dots,(2,1),(2,3),\dots,(p,p-1)} | A(\Sigma)_{\cdot,(1,1),(2,2),\dots,(p,p)}), \\ \tilde{\text{vec}}(M) &= (m_{21}, m_{31}, \dots, m_{12}, m_{32}, \dots, m_{p-1,p}, m_{11}, m_{22}, \dots, m_{pp})^\top. \end{aligned} \quad (4.19)$$

The vector of variables is

$$x = (\tilde{\text{vec}}(M), z_{12}, z_{13}, \dots, z_{21}, z_{23}, z_{p,p-1})^\top.$$

Then we set  $Q = \tilde{A}(\Sigma)^\top \tilde{A}(\Sigma)$  and  $q^\top = -2\tilde{A}(\Sigma)^\top \text{vec}(C)$ . The linear constraint is implemented by choosing

$$A = \begin{pmatrix} I & Z & -\mathcal{B}_U^{\text{off}} I \\ -I & Z & -\mathcal{B}_U^{\text{off}} I \\ & z_v & o_v \end{pmatrix},$$



## 4.2 Best Subset Selection with MIQP for Lyapunov Models

where  $I$  is the identity matrix of size  $p^2 - p \times p^2 - p$ ,  $Z$  is the zero matrix of size  $p \times p$ ,  $z_v$  is the vector of zeros of length  $p^2$  and  $o_v$  is the vector of ones of length  $p^2 - p$ . To complete the linear constraint, we set

$$b = \underbrace{(0, \dots, 0, k)}_{2(p^2-p)}^\top.$$

The constant bounds on the variables are set to

$$l = \underbrace{(-\mathcal{B}_U^{\text{off}}, \dots, -\mathcal{B}_U^{\text{off}})}_{p^2-p}, \underbrace{(-\mathcal{B}_U^{\text{diag}}, \dots, -\mathcal{B}_U^{\text{diag}})}_p, \underbrace{(0, \dots, 0)}_{p^2-p}^\top,$$

$$u = \underbrace{(\mathcal{B}_U^{\text{off}}, \dots, \mathcal{B}_U^{\text{off}})}_{p^2-p}, \underbrace{(\mathcal{B}_U^{\text{diag}}, \dots, \mathcal{B}_U^{\text{diag}})}_p, \underbrace{(1, \dots, 1)}_{p^2-p}^\top.$$

□

From now on, we refer to solving Problem 4.2.2 as best subset selection for Lyapunov Models (BSSLM).

### 4.2.2 Warm Starts of the BSSLM

Prior to solving the MIQP, an initial estimate can be passed to the MIO (Mixed Integer Optimization) solver. There are several options to provide a prior estimate for a warm start. A comparison between cold starts and warm starts for the classical best subset problem in regression is made in [Bertsimas et al., 2016, Section 5.2.2]. We discuss the following options for Lyapunov models:

- a) Warm start using a projected gradient descent as in [Bertsimas et al., 2016].
  - b) Warm start using the direct Lyapunov lasso (3.2).
  - c) Cold start initializing with a vector where all entries are set to a constant value.
- a) The warm start option presented in [Bertsimas et al., 2016] is a projected gradient descent that is based on Appendix C.0.1 and uses ideas from projected gradient descent methods in first-order convex optimization problems by Nesterov [2013], see Appendix C.0.2. This results for Lyapunov models in the Algorithm:

Input: A parameter  $L$ , a convergence tolerance  $\epsilon$ , the matrices  $A(\Sigma)$  and  $C$  with  $\Sigma$  and  $C$  being positive-definite.

- 1) Initialize with the vector  $\text{vec}(M)_1$  that contains the  $k$  largest entries of

$$\text{vec}(M)_{\text{init}} = \frac{-A(\Sigma)^\top \text{vec}(C)}{\text{colsums}(A(\Sigma)^2)},$$

where  $A(\Sigma)^2$  and the division are elementwise operations.

- 2) For  $m \geq 1$ , we obtain

$$\text{vec}(M)_{m+1} \in H_k \left( \text{vec}(M)_m - \frac{2}{L} A(\Sigma)^\top (A(\Sigma) \text{vec}(M)_m + \text{vec}(C)) \right),$$

where  $H_k$  is defined in C.0.1.

3) Repeat step 2), until

$$||A(\Sigma)\text{vec}(M)_m + \text{vec}(C)||_2^2 - ||A(\Sigma)\text{vec}(M)_{m+1} + \text{vec}(C)||_2^2 \leq \epsilon.$$

Detailed analysis of the convergence of the above algorithm is given in [Bertsimas et al., 2016, Section 3.1.].

b) Another possibility is to calculate a warm start using the direct Lyapunov lasso (3.2). The direct Lyapunov lasso as model selection technique is discussed in great detail in Chapter 3. It is computationally much faster than solving a MIQP which motivates using the method as a warm start. For this purpose we run the direct Lyapunov lasso along a  $\lambda$ -grid

$$0 < \frac{\lambda_{\max}}{10^4} = \lambda_1 < \dots < \lambda_{100} = \lambda_{\max}$$

with  $\lambda_{\max}$  being again the minimal  $\lambda$ -value such that  $M$  is diagonal. Then, we select the estimate along the path for which the number of nonzero entries is closest to  $k$ , the parameter restricting the number of active variables in (4.14).

c) One might be curious how important is the warm start for the MIO solver in the context of Lyapunov models. For classical regression problems, [Bertsimas et al., 2016, Table 1, Figure 3] suggest that warm starts are beneficial both time wise and result wise. However, despite some computational similarities, the matrix  $A(\Sigma)$ , contrary to a classical design matrix, possesses a very specific structure (4.2) that might affect the impact of warm starts.

Problem 4.2.2 contains constant bounds on the parameters  $\text{vec}(M)$ . Based on the warm starts a)-c), we select these constants. For choices a) and b) we select

$$B_U^{\text{off}} = 2\max_{\substack{(i,j) \\ i \neq j}}(|(\text{vec}(M)_{\text{best}})_{(i,j)})|$$

with  $\text{vec}(M)_{\text{best}}$  being the estimate obtained by a) and b). Then, we define

$$\begin{aligned} \mathcal{B}_l &= (p^2 - p)B_U^{\text{off}}, \\ \mathcal{B}_U^{\text{diag}} &= 2\max_{(i,j)}(|(\text{vec}(M)_{\text{best}})_{(i,j)})|. \end{aligned}$$

For the cold start, i.e. initialization c) we choose  $\mathcal{B}_U^{\text{off}} = \mathcal{B}_U^{\text{diag}} = 10000$ . This choice is arbitrary but much larger than any entry in the drift matrices  $M$  in our simulations. It reflects that by initializing with a constant vector, no data dependant estimate is available that could provide a sensible choice for the constant bounds. Based on this choice for  $\mathcal{B}_U^{\text{off}}, \mathcal{B}_U^{\text{diag}}$ , we choose the parameter  $\mathcal{B}_l$  as for a) and b).

### 4.2.3 BSSLM - Time Consumption and Comparison of Initializations

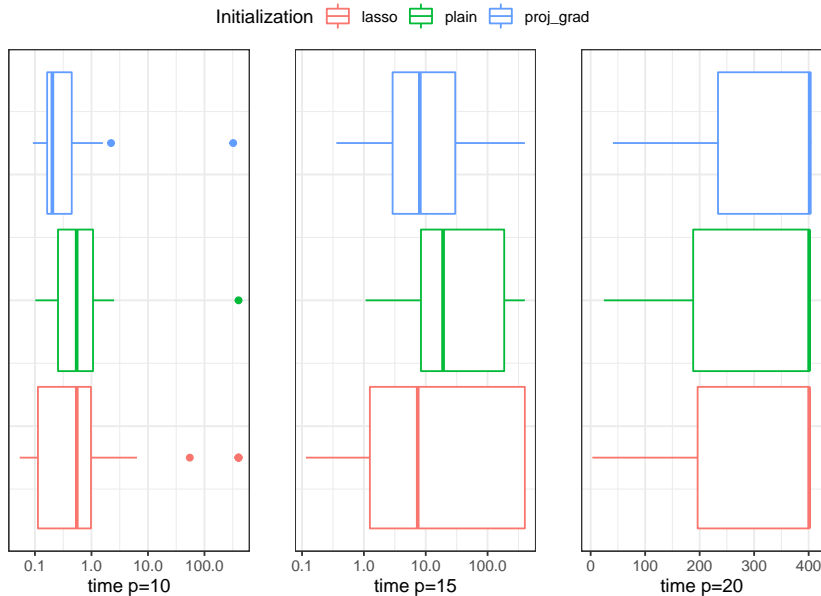
In this section, we study the time consumption of the BSSLM for the three initialization methods in Section 4.2.2 and compare them with the direct Lyapunov lasso. In particular, we observe that the optimistic view on time consumption in regression

## 4.2 Best Subset Selection with MIQP for Lyapunov Models

settings by Bertsimas et al. [2016] does not transfer to GCLMs. However, warm starts have also proven to be beneficial for GCLMs. Moreover, we illustrate how much slower the method is compared to the direct Lyapunov lasso (3.2).

The simulation setting we consider is similar to Section 3.6. Each stable matrix  $M$  is generated with  $M_{ij} = \omega_{ij}\epsilon_{ij}$  for  $i \neq j$  and  $M_{ii} = -\sum_{j \neq i} |M_{ij}| - |\epsilon_{ii}|$  where  $\omega_{ij} \sim \text{Bernoulli}(d)$  and  $\epsilon_{ij} \sim N(0, 1)$ . We fix the matrix  $C = 2I_p$ . For each  $m \in \{1, 2, 3, 4\}$  and  $p = \{10, 15, 20\}$ , the edge probability is set as  $d = m/p$ . For each value of  $m$ , we generate 10 pairs of signals  $(M, C)$ . Then, we generate  $n = 1000$  observations from a multivariate Gaussian distribution with covariance matrix solving the Lyapunov equation for  $(M, C)$ . We apply the BSSLM (4.17) with initialization a)-c) from Section 4.2.2 for  $p = \{10, 15, 20\}$  to all 40 datasets. We choose the parameter  $k$  to be the number of nonzero entries in the true drift matrix  $M$  that is used to generate the dataset. To solve the problem, we use the Gurobi optimizer [Gurobi Optimization, LLC, 2023] along with the R package `gclm` by Hastie et al. [2020b] where the functions are adjusted such that they are feasible for our problem. The maximal computation time for the Gurobi optimizer is set to 400 seconds per value of  $k$ . For the direct Lyapunov lasso (3.2), we calculate estimates along a regularization path of length 100. The path is chosen as in (4.9) with the difference that the start is set to  $\lambda_{\max}/10^4$ .

First, we compare the time consumption of the three initializations a)-c). We measure the total time required to calculate the BSSLM estimate for  $M$  with the respective initialization method. Therefore, the maximal computation time can be slightly above the 400 seconds we set as the time limit for the solver. We compare the initialization methods for  $p = \{10, 15, 20\}$  using boxplots in Figure 4.2. The x-axes for  $p = \{10, 15\}$  are logarithmized with base 10. More detailed summary statistics for Figure 4.2 are given in Appendix B.3.



**Figure 4.2:** Boxplots summarizing the time used to calculate the solution of (4.17) for one value of  $k$  across 40 randomly selected drift matrices and for the three initializations in Section 4.2.2. We have a) `proj_grad`, b) `lasso` and c) `plain`.

Neglecting the subtle differences, we observe a straightforward pattern. For almost all drift matrices, the BSSLM is solved within less than a second for  $p = 10$  for all initialization methods. The time consumption is already significantly higher for  $p = 15$ , but almost all problems are still solved within 400 seconds. However, for  $p = 20$ , we observe that for the majority of drift matrices, the BSSLM is not solved within 400 seconds. The projected gradient descent initialization that is the favorite of Bertsimas et al. [2016] also performs best for Lyapunov models. This can be seen in particular for  $p = \{10, 15\}$ .

**Remark 4.2.4.** *We want to emphasize that even  $p = 20$  is relatively problematic regarding time consumption. This is not so apparent in the work by Bertsimas et al. [2016], where they consider vectors rather than matrices. This means that graphs with  $p = 20$  for Lyapunov models have to be compared with regression problems with  $p = 400$ . Moreover, the matrices we consider are sparse but not to such an extreme degree as in most of their examples and makes the task more challenging.*

The above discussion raises two questions. First, what does it mean if we call the MIQP solved? Second, is the result unusable if the MIQP solver cannot certify optimality?

The MIP (MIQP) solver terminates when the gap between the lower and upper objective bound is less than  $\text{MIPGap}$  times the absolute value of the incumbent objective value [Gurobi Optimization, LLC, 2023, MIPGap]. If this is the case, we call the MIQP solved. In a small simulation study in Appendix B.2 with MIPgaps ranging from 0.001 to 0.1, we conclude that setting the MIPgap to 0.01 is the best choice from a practical perspective. Answering the second question requires more detailed simulations regarding the quality of the estimates for the different sample sizes.

Before we present more detailed simulations for the different initialization methods, we want to compare them to the computation time of the direct Lyapunov lasso. The computation is carried out using R along with the `glmnet` package [Friedman et al., 2010]. The computation times for the full path of length 100 are displayed in Table 4.1.

**Table 4.1:** Summary statistics of the time used to calculate the full path of length 100 of the direct Lyapunov lasso (3.2) across 40 randomly selected drift matrices  $M^*$ .

prob. size	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
p=10	0.004046	0.005453	0.007455	0.025579	0.016832	0.450402
p=15	0.01184	0.01850	0.03562	0.43298	0.18698	4.00151
p=20	0.03920	0.05945	0.11445	0.64046	0.30020	9.67137

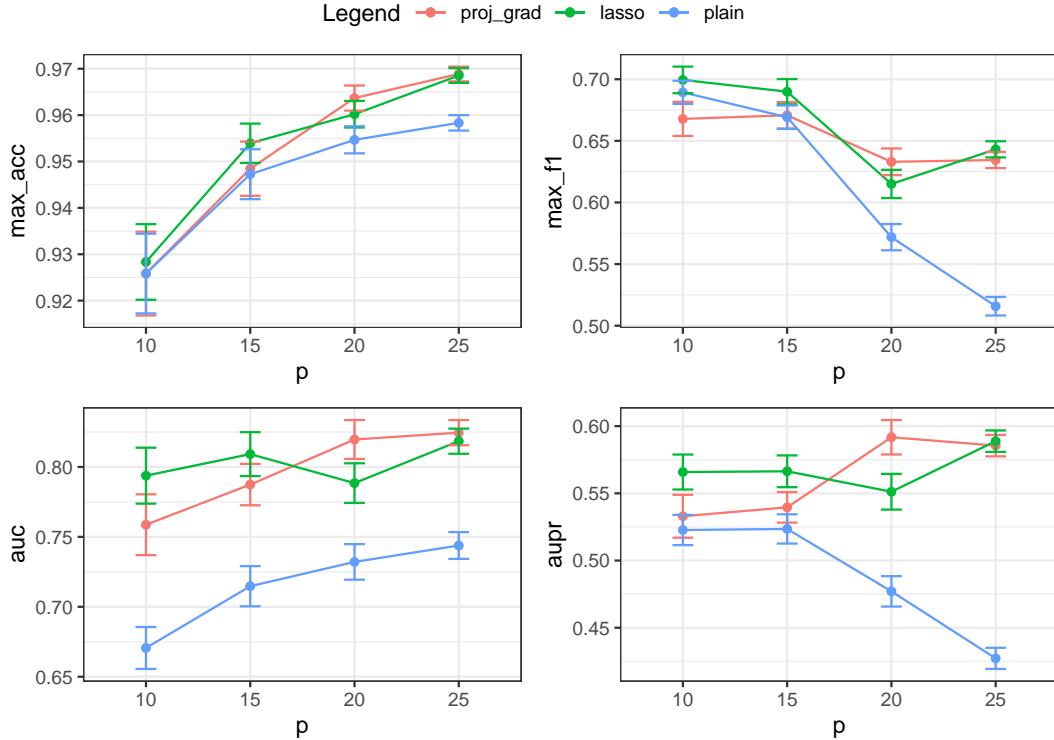
Despite calculating estimates along a path of length 100, the direct Lyapunov lasso is much faster than the BSLMM. The authors Bertsimas et al. [2016] consider more favorable settings where this downside is not that apparent. The consequence for graphical continuous Lyapunov models is that unless there are major technological advances, larger problems need to be solved using the direct Lyapunov lasso (3.2) or the `loglik-0.01` method (4.7).

Subsequently, we analyze the quality of the results in the above simulation setting using the four well-known metrics that are used by Varando and Hansen [2020] and in

## 4.2 Best Subset Selection with MIQP for Lyapunov Models

Section 3.6. The metrics are the maximum accuracy (max acc) and the maximum  $f_1$ -score (max\_f1). They are calculated as the maximum along the grid of regularization parameters. The other two metrics are the area under the roc curve (auc) and the area under the precision curve (aupr), which provide an average of the regularization path. Details on the metrics are provided in Definition 3.5.11.

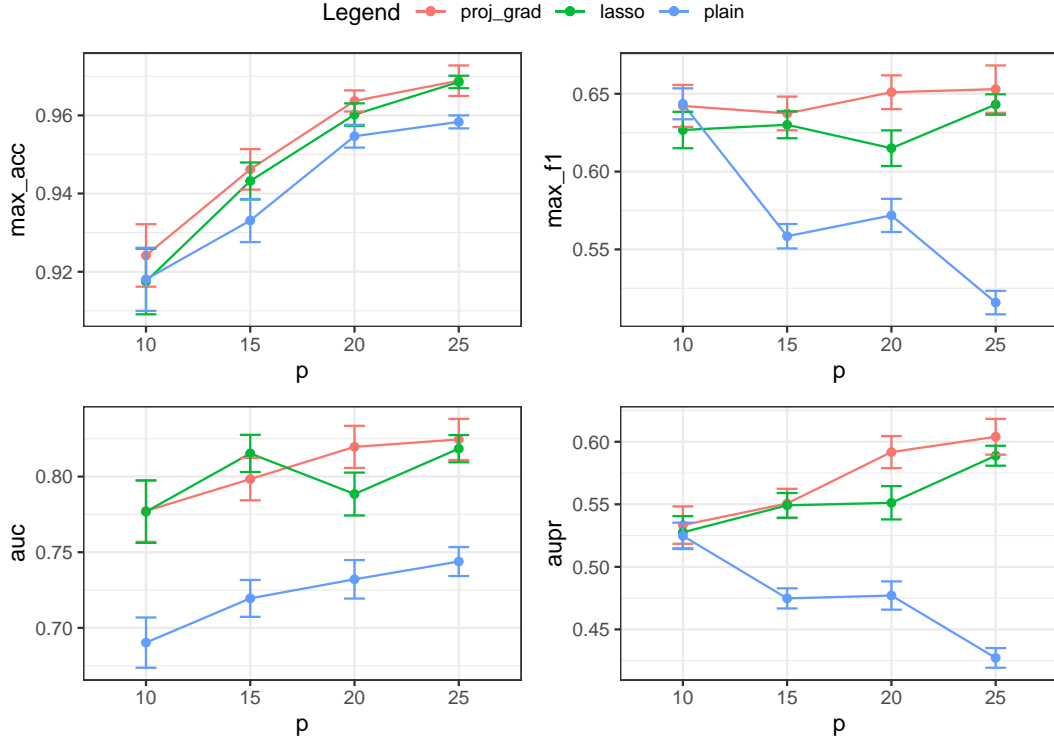
Unlike for the time comparison, we also include  $p = 25$ . We investigate two settings. First, we select 10 equidistant sparsity levels  $k$  starting with  $k = 1$  and ranging to  $k = \binom{p+1}{2}/3$ . The reason behind this maximal  $k$ -value is that, in most cases, we do not know the correct number of nonzero entries beforehand. Ideally, we would like to include every possible sparsity level in the grid. However, the extremely high time consumption does not allow for such a scenario. The upper limit for the sparsity level is set to  $k = \binom{p+1}{2}/3$  as more than  $\binom{p}{2}$  entries result in a non-identifiable graph, see Lemma 2.3.5. We choose the maximal number of sparsity parameters to be smaller as we assume the underlying structure to be sparse. We set the computation time to 100 seconds per value of  $k$ . The results are displayed in Figure 4.3.



**Figure 4.3:** Four evaluation metrics comparing support recovery for (4.17) with initializations a)-c) described in section 4.2.2 with  $C = 2I_p$  where data has been generated using  $C = 2I_p$ . The path is of length ten and ranges from  $k = 1$  to  $k = \binom{p+1}{2}/3$ . The maximal computation time per value of  $k$  is 100 seconds. Initialization proj\_grad is a), lasso is b), and plain is c).

We observe that the cold “plain” start is much worse than the warm starts with the projected gradient descent or the direct Lyapunov lasso. Regarding the warm starts, both methods exhibit similar performance. A comparison with Figure 3.12 reveals that although most estimates for  $p = \{20, 25\}$  are labeled suboptimal by the

solver, the simulation results can compete with those of the direct Lyapunov lasso. We provide a more detailed comparison in Section 4.2.4. Second, we select 15 equidistant sparsity levels  $k$  starting with  $k = 1$  and ranging to  $k = \binom{p+1}{2}/3$  and set the maximal computation time to 250 seconds per sparsity level  $k$ . These results are displayed in Figure 4.4.



**Figure 4.4:** Four evaluation metrics comparing support recovery for (4.17) with initializations a)-c) described in section 4.2.2 with  $C = 2I_p$  where data has been generated using  $C = 2I_p$ . The path is of length 15 and ranges from  $k = 1$  to  $k = \binom{p+1}{2}/3$ . The maximal computation time per value of  $k$  is 250 seconds. Initialization proj\_grad is a), lasso is b), and plain is c).

The general ordering of the methods remains the same. A warm start is clearly advantageous. Overall, the metrics for this expensive simulation run are similar when compared to Figure 4.3. This is unsurprising for  $p = \{10, 15\}$  as many computations finish within 100 seconds. However, it is pretty interesting that the improvement for  $p = \{20, 25\}$  is also not significant. There is a mild improvement for the projected gradient descent initialization, which seems to dominate the lasso initialization. This, in conjunction with the slightly better time consumption for  $p = 15$ , leads us to the suggestion to use the projected gradient descent initialization for the warm start.

Note that it is possible to push the computational boundaries of the BSSLM further. We omit it when presenting Figure 4.3 and Figure 4.4. The purpose is to show that when the optimality of the estimate for a certain MIPGap is not achieved, this does not result in the estimate being useless. The results can keep up with those of the direct Lyapunov lasso presented in Section 3.6. The clear conclusion of the section is that warm starts are definitely required, with the projected gradient descent initialization

having a slight edge. The time consumption of the BSSLM seems to be its biggest issue.

#### 4.2.4 Comparing the BSSLM with the direct Lyapunov lasso and the Loglik-0.01

In this section, we compare the BSSLM with the direct Lyapunov lasso (3.2) and the loglik-0.01 method (4.7). The loglik-0.01 method jointly estimates  $M$  and  $C$ , while the other methods only estimate the drift matrix  $M$ . In this section, we only compare the quality of the estimated  $M$ . We consider different simulation settings to investigate in which scenario the BSLMM performs best and where the  $\ell_1$ -penalized methods have their advantages. Our results confirm that in settings where the active variables are clearly recognizable, the BSSLM is superior to the  $\ell_1$ -penalized methods. The setting where the  $\ell_1$ -penalized methods perform best in the work by Hastie et al. [2020b] can not easily be translated to GCLMs and proves to lead to similarly bad performances across all methods. However, we show that generally, the  $\ell_1$ -penalized methods can offer advantages when the number of active variables is relatively big and, at the same time, the size of the entries is small.

First, we outline the general simulation setup and then give the specifics for the individual simulation runs. We consider problem sizes  $p = \{10, 15, 20, 25\}$ . We select 100 drift matrices  $M$  according to the respective setting. The volatility matrix  $C = 2I_p$  in all settings. Then, we generate  $n = 1000$  observations from a multivariate Gaussian distribution with a covariance matrix solving the Lyapunov equation for  $(M, C)$ . Subsequently, the following three estimation methods are applied.

- a) `Best_Subset` stands for the Best Subset Selection for Lyapunov Models (BSSLM): Solving Problem 4.2.2 using an equidistant grid of sparsity levels  $k$  ranging from  $k = 1$  to  $k = \binom{p+1}{2}/3$  of length 20. We select  $C = 2I_p$ . The time limit is set to 400 seconds per value of  $k$ . The initialization method is the projected gradient descent with the number of runs being set to 50 and the maximum number of iterations to 1000. The initialization method is method a) in Section 4.2.2. The computations are carried out using an adapted version of the `best-subset` package by Hastie et al. [2020a]. The MIQP solver is the one by Gurobi Optimization, LLC [2023].
- b) `loglik-0.01`: Solving (4.7) using the negative Gaussian log-likelihood (4.6) where the parameter  $\kappa$  is set to 0.01. The maximum number of iterations is set to 1000. The path of regularization parameters is set to be

$$0 < \frac{\lambda_{\max}}{10^4} = \lambda_1 < \dots < \lambda_{100} = \lambda_{\max} \quad (4.20)$$

with  $\lambda_{\max}$  being chosen such that the matrix  $M$  is diagonal. The computations are carried out using the `gclm` package by Varando [2020].

- c) `Direct_Lasso` stands for the direct Lyapunov lasso: Solving (3.2) while setting  $C = 2I_p$  and using the same regularization path as for the loglik-0.01 method (4.9). The maximum number of iterations is set to 10000. This is the default value in the `glmnet` package Friedman et al. [2010] that is used for computations.

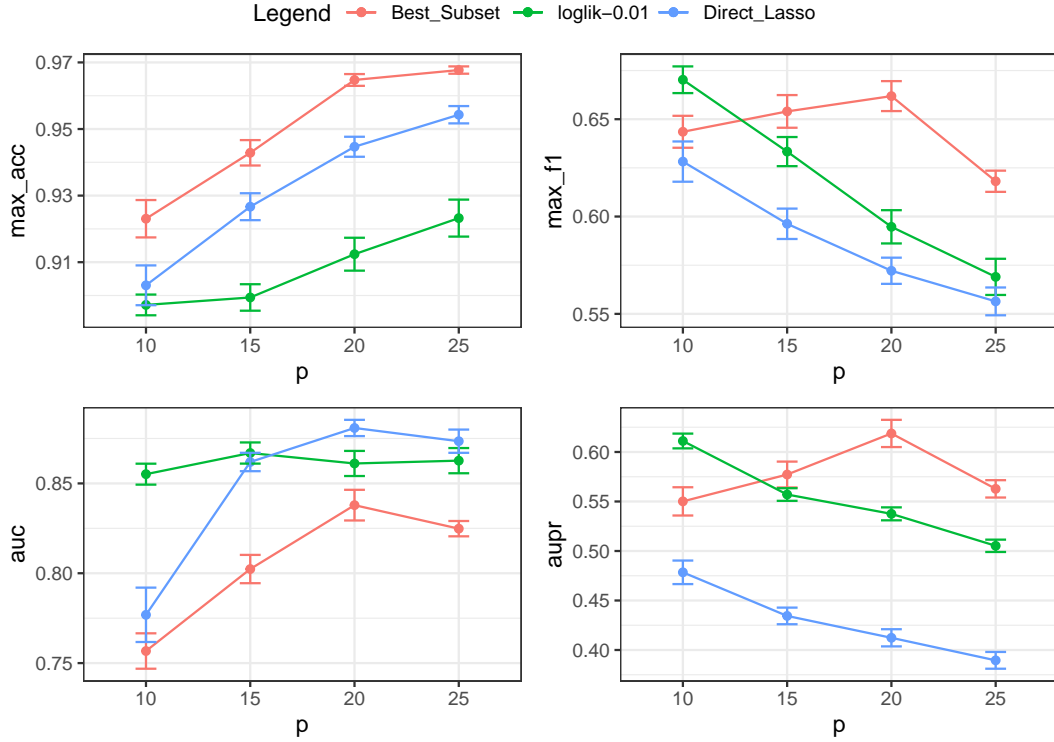
For these problem sizes, the methods b) and c) can be calculated relatively quickly on a standard local computer. However, the BSSLM has a total computation time of multiple days. Therefore, the simulations are carried out on the CoolMUC2 Cluster with 28-way Haswell-based nodes and 64 GB RAM per node at Leibniz-Rechenzentrum (LRZ) supercomputing facility ([www.lrz.de](http://www.lrz.de)). Unless there are major technical advances, we find that the BSSLM is only applicable to problems up to sizes  $30 \times 30$  at maximum. For an exhaustive simulation study that we present here, the limit seems to be  $25 \times 25$ . Now, we present the different simulation settings in which we apply the methods a)-c).

- 1) We consider drift matrices  $M$  selected as in Section 4.2.3. The difference is that we consider 25 randomly selected drift matrices for the 4 sparsity levels  $m/p$  with  $m \in \{1, 2, 3, 4\}$ . This setting leads to drift matrices with entries that having varying sizes. However, no extreme setting is considered.
- 2) We consider drift matrices  $M$  with an edge probability of 0.15 and off-diagonal entries selected according to  $\text{Unif}[1,2]$ . The diagonal of the drift matrices is chosen as the negative absolute row sum of the off-diagonal entries minus a small safety margin to make them stable. This setting is the ideal scenario for variable selection. The drift matrices are reasonably sparse, and the nonzero entries are all relatively equal in size. Furthermore, they match the size of the diagonal entries in  $C = 2I_p$ .
- 3) We consider drift matrices  $M$  with an edge probability of 0.3 and off-diagonal entries selected according to  $N(0, \sigma^2)$  with  $\sigma = 0.1$  and using  $C = 2I_p$ . The diagonal of the drift matrices is chosen as the negative absolute row sum of the off-diagonal entries minus a small safety margin to make them stable. This setting is supposed to be the opposite of 2). The edge probability is double the one in 2), and almost all entries in the drift matrix are much smaller than those in  $C = 2I_p$ . This setting is less desirable for variable selection.
- 4) We consider drift matrices  $M$  with an edge probability of 0.3. One-third of these entries are set to 1, and two third are selected as a decreasing sequence  $(2^{-\frac{n}{4}})_{n \in \mathbb{N}}$ . The diagonal of the drift matrices is chosen as the negative absolute row sum of the off-diagonal entries minus a small safety margin to make them stable. In this setting, the drift matrices are reasonably dense, and the entries vary a lot in size. This is the setting where the Lasso method in the work by Hastie et al. [2020b] proved to be superior to the best subset method in the context of sparse regression. However, there is a subtle difference. For graphical continuous Lyapunov models the number of active variables is usually much higher than in their examples which results in some entries being very small due to the exponential decrease.

The metrics for evaluation are those that are used in Section 4.2.3. Namely, we use the maximum accuracy (max acc) and the maximum  $f_1$ -score (max\_f1). The other two metrics are the area under the roc curve (auc) and the area under the precision curve (aupr). Details on the metrics are provided in Definition 3.5.11. The results for setting 1) are displayed in Figure 4.5.



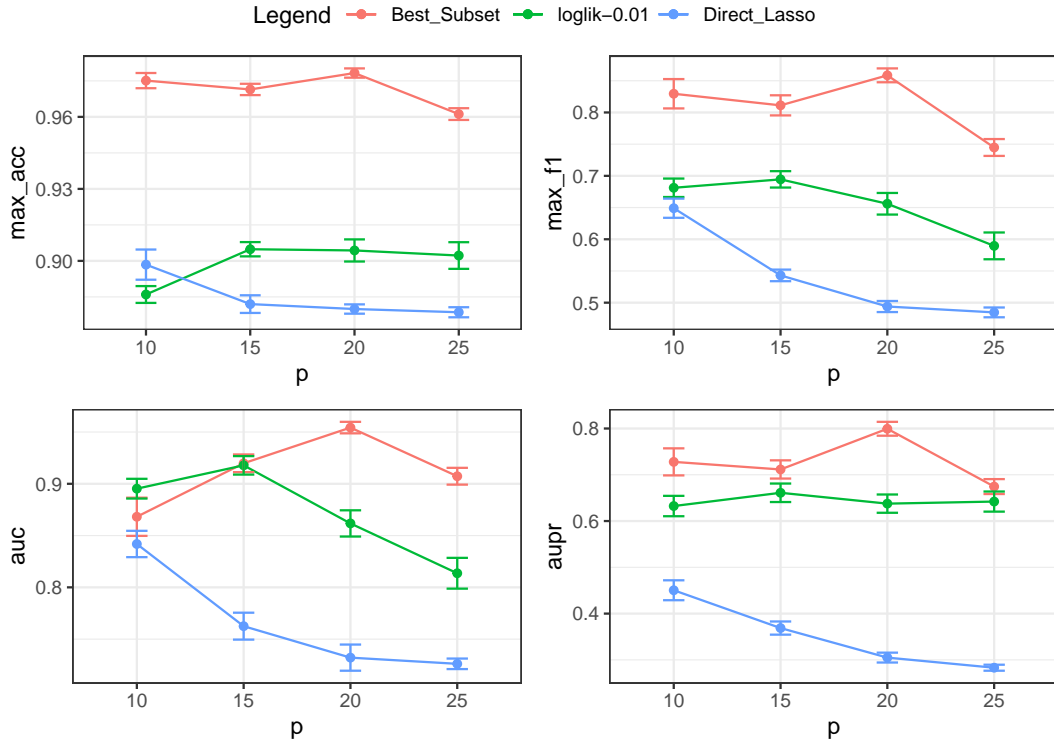
## 4.2 Best Subset Selection with MIQP for Lyapunov Models



**Figure 4.5:** Comparison of 3 methods for model selection for Lyapunov models. We consider the simulation setting 1) in Section 4.2.4. The graphic shows four evaluation metrics comparing support recovery for the best subset method (4.17), the loglik-0.01 (4.7), and the direct Lyapunov lasso (3.2) using  $C = 2I_p$  for the best subset method and the direct Lyapunov lasso.

There are a couple of interesting observations to be made. First, we observe that the best subset approach performs better than the  $\ell_1$ -penalized methods for  $p = \{15, 20, 25\}$  except for the auc. The reason why the auc score is worse is because the metric focuses both on the correctly classified true positives and true negatives and calculates an “average” along the regularization path. The  $\ell_1$ -penalized methods tend to have a lot of symmetry in their nonzero pattern. For instance, consider the Examples in Section 4.1. As the number of zero entries is larger than the number of nonzero entries, this results in decent ratios and an overall good auc score. The proportion of zero entries to nonzero entries is increasing with increasing problem size. This seems to favor the best subset method except for  $p = 25$  where the metrics auc, max  $f_1$ -score and aupr decrease a little. This might be because the computation time does not suffice to produce optimal estimates. The loglik-0.01 method is superior to the direct Lyapunov lasso overall. The results for setting 2) are displayed in Figure 4.6.

This is the dream setting for the best subset method. The drift matrices possess entries that are more or less equal in size and, more importantly, are similar to the entries of  $C = 2I_p$  in size. No small entries exist for which it might be hard for a  $\ell_0$ -penalized method to distinguish if an entry is zero. Moreover, the drift matrices are relatively sparse. The simulation results reflect this. The BSSLM dominates the



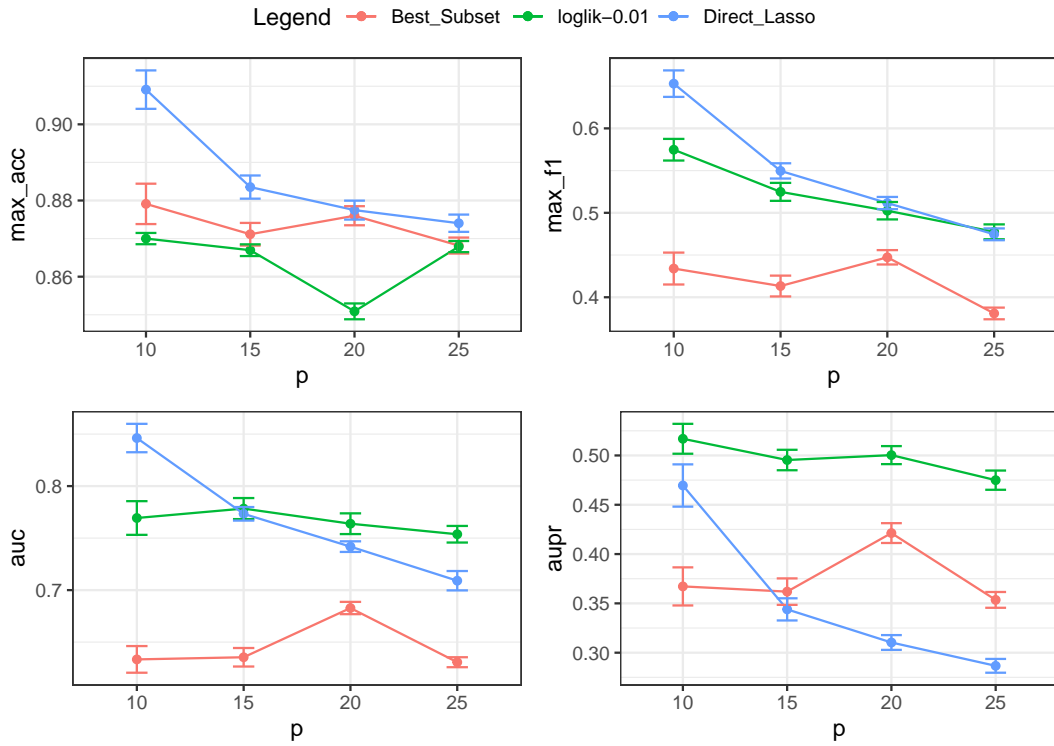
**Figure 4.6:** Comparison of 3 methods for model selection for Lyapunov models. We consider the simulation setting 2) in Section 4.2.4. The graphic shows four evaluation metrics comparing support recovery for the best subset method (4.17), the loglik-0.01 (4.7), and the direct Lyapunov lasso (3.2) using  $C = 2I_p$  for the best subset method and the direct Lyapunov lasso.

$\ell_1$ -penalized methods. The loglik-0.01 method is superior to the direct Lyapunov lasso in this setting. The results for setting 3) are displayed in Figure 4.7.

Setting 3) is challenging. The drift matrices are twice as dense as in setting 2). The normal distribution with mean zero and small standard deviation leads to the majority of the entries being close to zero while those in  $C = 2I_p$  are much larger. Moreover, there is quite some variability in the size of the entries when compared relatively. Overall, the loglik-0.01 method produces the best results. The direct Lyapunov lasso performs well for the smaller problem sizes  $p = \{10, 15\}$ , but its performance falls off for  $p = \{20, 25\}$ . The best subset method is for no metric and for no problem sizes better than the  $\ell_1$ -penalized methods. It is worth pointing out that all methods are performing worse as in setting 2). However, the decrease in performance is most drastic for the BSSLM. The results for setting 4) are displayed in Figure 4.8.

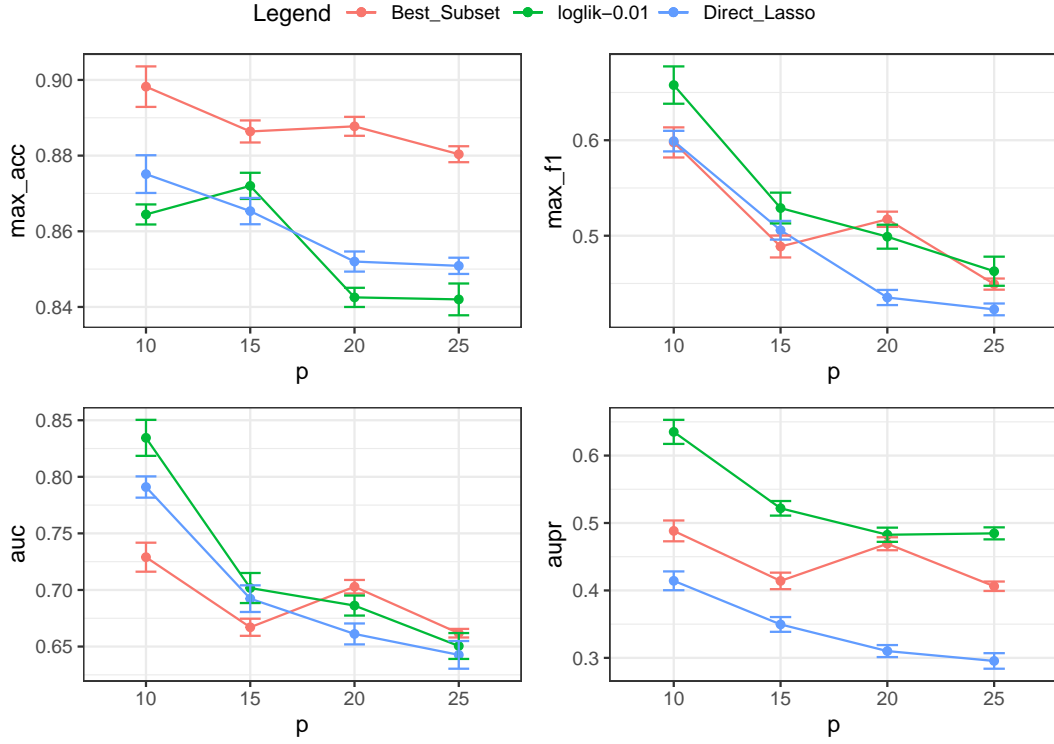
In classical sparse regression, this is the setting where Hastie et al. [2020b] show that the lasso regression is able to outperform the best subset method. Of course, the simulation setting for GCLMs is slightly different and the problem is only computationally a classical lasso problem. Indeed, the results do not allow the same conclusion. No method does particularly well. The best subset method is not inferior to the other methods. The decrease in all metrics for increasing problem size can be traced back

## 4.2 Best Subset Selection with MIQP for Lyapunov Models



**Figure 4.7:** Comparison of 3 methods for model selection for Lyapunov models. We consider the simulation setting 3) in Section 4.2.4. The graphic shows four evaluation metrics comparing support recovery for the best subset method (4.17), the loglik-0.01 method by (4.7) and the direct Lyapunov lasso (3.2) using  $C = 2I_p$  for the best subset method and the direct Lyapunov lasso.

to the simulation setting. With increasing problem size, the additional entries in the drift matrices become smaller due to the exponential decrease.



**Figure 4.8:** Comparison of 3 methods for model selection for Lyapunov models. We consider the simulation setting 4) in Section 4.2.4. The graphic shows four evaluation metrics comparing support recovery for the best subset method (4.17), the loglik-0.01 method by (4.7) and the direct lyapunov lasso (3.2) using  $C = 2I_p$  for the best subset method and the direct lyapunov lasso.

### 4.3 BSSLM and the Extended BIC

In Section 4.2.4, we show that the BSSLM is superior to the  $\ell_1$ -penalized model selection methods, in particular, when there is a clear distinction between zero and nonzero entries in the true drift matrix. This is done by calculating metrics along a grid of sparsity levels  $k$ . The ultimate goal of a model selection method is to produce one estimate. For this purpose, we use the extended Bayesian information criterion (EBIC) in Section 3.7. In this section, we investigate the EBIC together with the BSSLM on two specific structures. This reveals subtle issues with structure recovery.

For a more detailed explanation of the application of the EBIC for tuning parameter selection we defer to [Chen and Chen, 2008, Foygel and Drton, 2010] and for GCLMs to Section 3.7.2. We recall the central criterion. The general concept is to minimize the two times negative Gaussian log-likelihood

$$L(M) = n[\log \det(\Sigma(M, 2I_p)) + \text{tr}(\hat{\Sigma}(\Sigma(M, 2I_p))^{-1})] \quad (4.21)$$

for all models that are obtained by restricting the support of  $M$  for the considered sparsity levels  $k$ . The models are labelled as  $G_j$  for  $j \in I$  with  $I$  being the indices of the considered sparsity levels. The minima of (4.21) are denoted by  $\hat{L}_j$ . Using the



**Figure 4.9:** Left: The path 1 to 5. Right: The Double Path from 1 to 5.

convention that  $E_j$  is the edge set of  $G_j$  and substituting the values into

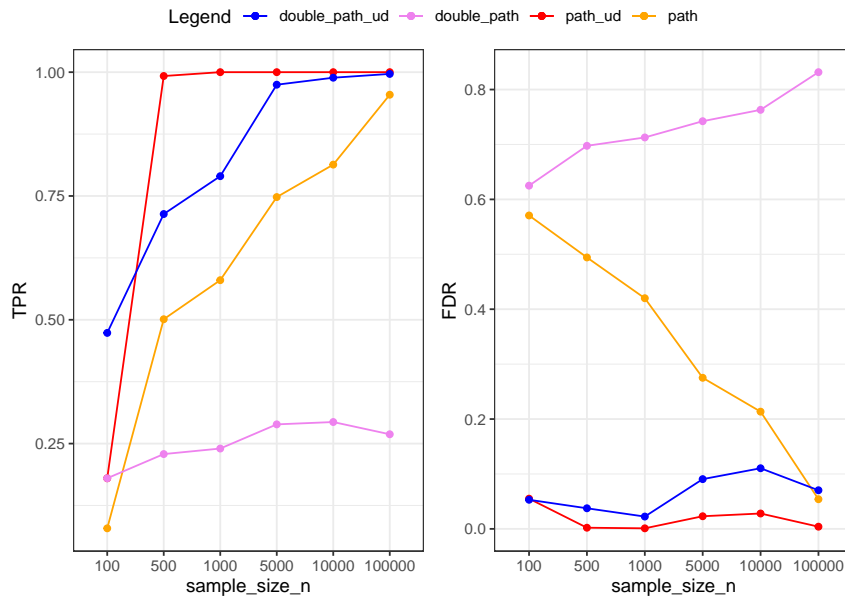
$$EBIC_\gamma(G_j) = (|E_j| + p) \log n + 4\gamma|E_j| \log p + \hat{L}_j, \quad (4.22)$$

one selects the graph with the lowest score. We consider two graph structures. First, the path of length 11. Second, the “double path” of length 11 where the path is extended by edges connecting nodes  $i$  and  $i+2$  with a directed edge where  $i \in 1, \dots, 9$ . A visualization for  $p = 5$  is given in Figure 4.9.

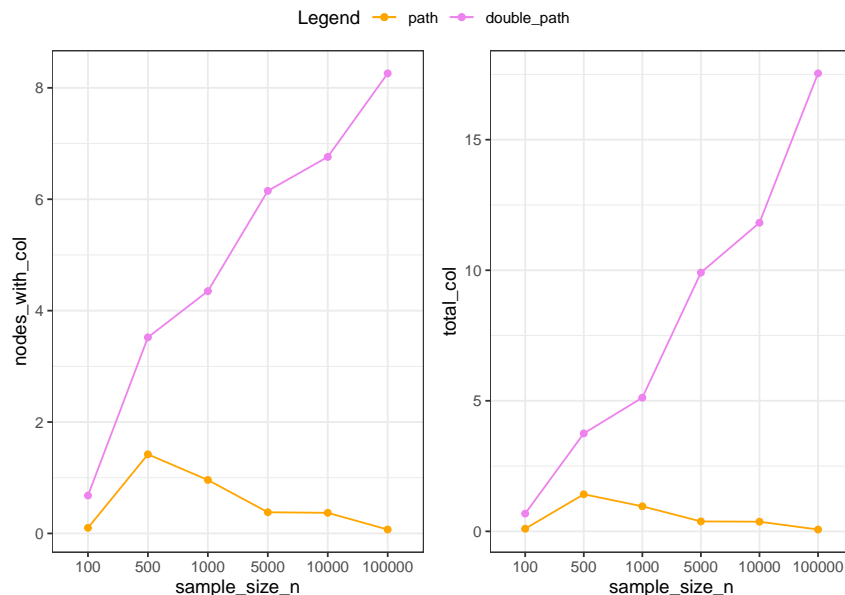
We choose drift matrices with the diagonal entries in  $M_{\text{path}}^*$  and  $M_{\text{doublepath}}^*$  set to -2 and the off-diagonal entries set to 1. Choosing  $C = 2I_p$  and solving the continuous Lyapunov equation (1.2), we obtain the covariance matrices  $\Sigma_{\text{path}}^*$  and  $\Sigma_{\text{doublepath}}^*$ . Using them, we generate 100 datasets of sizes  $n = \{100, 500, 1000, 5000, 10000, 100000\}$  for both graphs. We select an equidistant grid of 20  $k$ -values ranging from  $k = 1$  to  $k = \binom{p+1}{2}/3 + p$ . We make sure that the true number of nonzero entries of the drift matrices is included in the grid of  $k$ -values, sometimes resulting in a grid of length 21. The computation time per value of  $k$  is set to 400 seconds. Then, we use the EBIC criterion with  $\gamma = 1$  to select one structure. We calculate the true positive rate (TPR) and false discovery rate (FDR) and average out over the 100 datasets for each sample size and for both graph structures. The TPR is defined in Definition 3.5.11, and the FDR is given by  $fp/(fp + tp)$ . Additionally, we carry out the same calculation where we only take into account the skeleton (undirected structure) of the graphs. For instance, an edge from  $2 \rightarrow 1$  is classified as  $tp$  despite only the edge  $1 \rightarrow 2$  being present in the graph. The results are displayed in Figure 4.10.

Consistency has been proven for BIC-type criteria for Gaussian graphical models [Foygel and Drton, 2010]. The natural question is how the EBIC behaves for Lyapunov models. The blue curve displays the TPR and FDR if we take out the directionality of the edges and simply calculate the metrics for the undirected structure for the double path. The red curve displays the same for the path. The pink curve is the result of the double path for the directed structure. The orange curve displays the same for the path. The results for the undirected structure show the consistency that is desired. With increasing sample size, the TPR tends to be one. This happens faster for the path than for the double path, which is quite natural as the double path has a more complicated structure. The FDR is relatively close to zero, but some false positives always seem to be present. Regarding the directed structure, we observe similar behavior for the path. The TRP is increasing and almost one for  $n = 100000$ . The FDR decreases with the increasing sample size and is almost zero for  $n = 100000$ . The convergence is much slower than for the undirected structure. The results for the directed structure of the double path are most surprising. The TPR is increasing only very slightly for increasing sample size and is overall very low. The FDR is even increasing. At first glance, one might think that the method does not work at all. However, how does this align with the great results when considering the undirected structure? A more detailed investigation of the estimates shows that the direction of some edges is reversed. This is also reflected by the number of colliders displayed

in Figure 4.11. A collider triple in  $G = (V, E)$  is a triple of vertices  $(i, j, k)$  such that there are directed edges from  $i$  to  $j$  and from  $k$  to  $j$ , i.e. a structure of the form  $i \rightarrow j \leftarrow k$ .



**Figure 4.10:** TPR (true positive rate) and FDR (false discovery rate) of the estimated drift matrix obtained with the best subset method (4.17) and EBIC with  $\gamma = 1$  for increasing sample size. The results are averaged over 100 datasets generated by the same drift matrix.



**Figure 4.11:** Number of nodes with collider and the total number of colliders of the estimated drift matrix obtained with the best subset method (4.17) and EBIC with  $\gamma = 1$  for increasing sample size. The results are averaged over 100 datasets generated by the same drift matrix.

We observe that the number of colliders goes to zero for the path while the number of colliders is increasing with increasing sample size for the double path. This is not because the results get worse per se with increasing sample size. More edges are included with increasing sample size, but with the wrong direction.

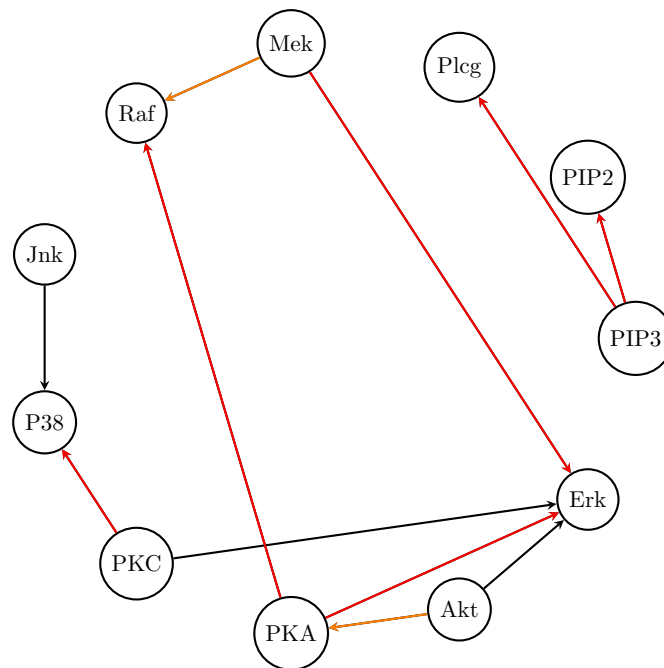
Sometimes, there seem to exist issues with the direction of the edges. Despite this exact condition not being necessary, we want to mention that the left side of the irrepresentability condition (3.15) is much higher ( $\approx 4.7$ ) than 1 for the double path. In the subsequent application example, we also indicate the edges where the reverse direction is correct.

### 4.3.1 Application - Sachs Dataset with the BSSLM and the Extended BIC

In this section, we showcase that the BSSLM with the EBIC is a method that can recover important connections of an among scientist-accepted protein-signaling network purely from observational data. The network has initially been analyzed by Sachs et al. [2005]. There exist better methods if the only goal is to provide the best network estimate. Of course, this is also due to the limited theory on the relatively new Lyapunov models. For more details, we defer to Section 3.7. However, the way the dataset has been collected makes it a suitable and interesting application case for these models. This is for two reasons. First, there exist feedback loops in the among scientists accepted network that are always problematic to deal with in classical structural equation modeling setups. Second, the cells have been destroyed in the measurement process, which is in line with snapshots of a temporal process in equilibrium. This motivated the authors Fitch [2019], Varando and Hansen [2020] and ourselves in Section 3.7 to study the dataset in the context of GCLMs. We refer for the details regarding the dataset and standardization to the mentioned section.

We use a grid of 21  $k$ -values from  $k^* - 10$  to  $k^* + 10$  with  $k^*$  being the true number of connections in the protein-signaling network, i.e.,  $k^* = 20$ . This reflects that some prior guesses on the number of connections might be available. The computation time per value of  $k$  is set to 400 seconds. The extended BIC criterion is used to select the graph structure. We present the estimate for dataset 2 in Figure 4.12.

In total, the estimate has 11 edges which are way fewer than the 20 edges of the among scientist accepted protein-signalling network in Figure 3.13. Although there is some debate about some connections in the network, we refer to the network in Figure 3.13 as “ground truth”. The edges of the estimate that are marked in red are also present in the ground truth and for the orange edges the reverse edge is present. Out of the edges present in the estimate, we have six edges that are estimated completely correctly and two more where the reverse direction is present in Figure 3.13. For the additional edges, we have that Jnk and P38 are connected via a trek  $JNK \leftarrow PKC \rightarrow P38$  and that Akt and Erk are connected via the trek  $Akt \leftarrow PKA \rightarrow Erk$ . The implication of the absence of such a trek is discussed in [Varando and Hansen, 2020, Section 2.3.], but as both the ground truth and the estimate connect these pairs of variables with a trek, there is at least no evidence for a difference in the covariance matrices zero pattern. The edge  $PKC \rightarrow ERK$  is missing in the ground truth, too. However, the pathway  $PKC \rightarrow Raf \rightarrow Mek \rightarrow ERK$  is present.



**Figure 4.12:** Estimated Sachs Network using the BSSLM (4.17) assuming  $C = 2I_p$  and selecting the structure according to the extended BIC (3.23) with  $\gamma = 1$  (Dataset 2)

Among the correctly estimated connections are the direct enzyme-substrate relationships  $PKA \rightarrow Raf$ ,  $Mek \rightarrow Erk$ , and  $PIP3 \rightarrow PIP2$  and the connection  $Raf \rightarrow Mek$  is estimated with the wrong directionality. Only the connection  $PIP3 \rightarrow Plcg$  is missing, but a trek via  $PIP2$  is present. Of course, quite a few connections are missing in the network presented in Figure 4.12, but the majority of the edges are present in the ground truth; some are reversed, and even the ones that are not present can be explained by pathways or treks. Overall, the estimate reveals a lot of important connections with the correct directionality and produces a low number of false positives. One might even argue to which extent these edges are false positives.

### 4.4 Diagonally Unknown Volatility Matrix

Both methods, the direct Lyapunov lasso (3.2) and the best subset method (4.14) are adaptable such that  $M$  and  $C$  can be estimated jointly. A detailed theoretical discussion requires more theory for Lyapunov models, such as the extension of the results on parameter identifiability in Chapter 3. However, computations can be carried out relatively easily, and we present a simulation study that indicates the potential of this generalization. We limit ourselves to the analysis of  $C$  diagonally unknown.

The idea to adapt the optimization problems is quite simple and based on the idea by Varando and Hansen [2020]. We add an additional penalty term to both objective functions that regulate how close the matrix  $C$  is to the identity matrix  $I_p$ . The larger the parameter  $\kappa$ , the closer the matrix  $C$  to  $I_p$ . The direct Lyapunov lasso for  $C$



#### 4.4 Diagonally Unknown Volatility Matrix

unknown is given by

$$\arg \min_{\substack{M, C \in \mathbb{R}^{p \times p} \\ C \text{ diag}}} \frac{1}{2} \|A(\Sigma)\text{vec}(M) + \text{vec}(C)\|_2^2 + \lambda \|M\|_1 + \frac{\kappa}{2} \|\text{vec}(C) - \text{vec}(I_p)\|_2^2 \quad (4.23)$$

and the best subset selection (BSSLM) for  $C$  unknown is given by

$$\arg \min_{\substack{M, C \in \mathbb{R}^{p \times p} \\ C \text{ diag}}} \|A(\Sigma)\text{vec}(M) + \text{vec}(C)\|_2^2 + \kappa \|\text{vec}(C) - \text{vec}(I_p)\|_2^2 \quad (4.24)$$

*s.t.*  $\|\text{vec}(M)\|_{0,\text{off}} \leq k.$

We transform the optimization problems such that they follow a more standard form of a quadratic program that eases analysis and computations. The transformation appears in the master thesis by Szekeres [2023] that the author of this thesis co-supervised. When optimizing over  $M$  and  $C$  diagonal jointly, we consider a vector of variables that is given by

$$\text{vec}(M, C) = (m_{11}, m_{21}, m_{31}, \dots, m_{p1}, \dots, m_{1p}, \dots, m_{pp}, c_{11}, c_{22}, \dots, c_{pp})^\top \in \mathbb{R}^{p^2+p}.$$

For  $C$  known, we only vectorize

$$\text{vec}(M\Sigma + \Sigma M^\top) = A(\Sigma)\text{vec}(M),$$

while here, we need to find a matrix  $B(\Sigma)$  such that

$$\text{vec}(M\Sigma + \Sigma M^\top + C) = B(\Sigma)\text{vec}(M, C).$$

The matrix  $B(\Sigma)$  is the matrix  $A(\Sigma)$  with additional columns that emerge from including the diagonal elements of  $C$  into the vector of variables. The  $p^2 \times (p^2 + p)$  matrix  $B(\Sigma)$  is given by

$$B(\Sigma)_{ij} = \begin{cases} A(\Sigma)_{ij} & \text{if } i, j \in 1, \dots, p^2, \\ 1 & \text{if } i = (k-1)p + k, j = p^2 + k, k = 1, \dots, p, \\ 0 & \text{otherwise.} \end{cases}$$

This notation allows to formulate

$$\|A(\Sigma)\text{vec}(M) + \text{vec}(C)\|_2^2 = \text{vec}(M, C)^\top B(\Sigma)^\top B(\Sigma)\text{vec}(M, C).$$

Additionally, we have to rephrase

$$\|\text{vec}(C) - \text{vec}(I_p)\|_2^2 = \text{vec}(C)^\top \text{vec}(C) - 2\text{vec}(C)^\top \text{vec}(I_p) + \text{vec}(I_p)^\top \text{vec}(I_p).$$

For this purpose, we introduce

$$\Omega_{ij} = \begin{cases} \kappa & \text{if } i = j \text{ and } i, j \geq p^2, \\ 0 & \text{otherwise,} \end{cases}$$

and the notation that  $\mathbf{1}_p$  is the vector of ones of length  $p$  and that  $\mathbf{0}_{\mathbf{p} \times \mathbf{p}}$  is the matrix of zeros of size  $p \times p$ . This yields

$$\frac{\kappa}{2} \|\text{vec}(C) - \text{vec}(I_p)\|_2^2 = \frac{1}{2} \text{vec}(M, C)^\top \Omega \text{vec}(M, C) - \kappa(\mathbf{0}_{\mathbf{p} \times \mathbf{p}}, \mathbf{1}_p)^\top \text{vec}(M, C) + \mathbf{1}_p^\top \mathbf{1}_p.$$

Using the notation that  $D(\Sigma) = B(\Sigma)^\top B(\Sigma) + \Omega$ , we can rephrase the optimization problems for the direct Lyapunov lasso and the best subset method.

**Direct lyapunov lasso for  $C$  diagonally unknown:**

$$\arg \min_{\substack{M, C \in \mathbb{R}^{p \times p} \\ C \text{ diag}}} \frac{1}{2} \text{vec}(M, C)^\top D(\Sigma) \text{vec}(M, C) - \kappa(\mathbf{0}_{\mathbf{p} \times \mathbf{p}}, \mathbf{1}_p)^\top \text{vec}(M, C) + \lambda \|M\|_1 \quad (4.25)$$

**Best subset selection (BSSLM) for  $C$  diagonally unknown:**

$$\arg \min_{\substack{M, C \in \mathbb{R}^{p \times p} \\ C \text{ diag}}} \text{vec}(M, C)^\top D(\Sigma) \text{vec}(M, C) - 2\kappa(\mathbf{0}_{\mathbf{p} \times \mathbf{p}}, \mathbf{1}_p)^\top \text{vec}(M, C) \quad (4.26)$$

*s.t.*  $\|\text{vec}(M)\|_{0, \text{off}} \leq k$

These optimization problems can easily be solved using the statistic software R [R Core Team, 2021]. For the best subset selection, we additionally require a solver for MIQPs. As for  $C$  known, we use the Gurobi optimizer Gurobi Optimization, LLC [2023]. We do not want to go into too many technical details regarding the computations as this is similar to the case  $C$  known. However, there are some subtle changes. The optimization problem for the direct Lyapunov lasso cannot be passed to the `glmnet` function [Friedman et al., 2010]. Therefore, the optimization package `smde` that is able to deal with more general  $\ell_1$ -penalized quadratic forms is used [Hansen, 2014]. Using a mildly modified version of (4.17) works for the best subset selection. Moreover, we fix  $c_{11} = 1$  to take into account the scaling invariance, see Remark 1.3.4.

**4.4.1 Simulations - Diagonally Unknown Volatility Matrix**

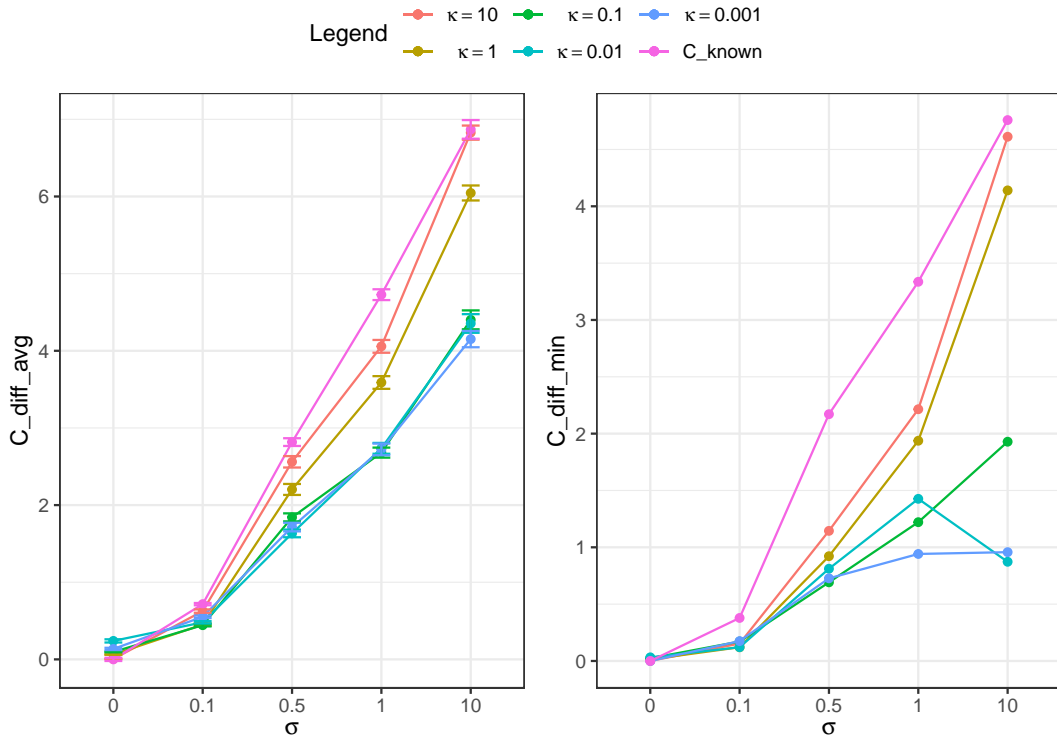
In this section, we show how the BSSLM for  $C$  diagonally unknown provides more information on the unknown diagonal of the matrix  $C$ . At the same time, the way the problem is set up with the tuning parameter  $\kappa$ , the estimation results for  $C$  are not perfect. We investigate which choice of the tuning parameter leads to the most favorable results for  $M$  and  $C$ . To conclude, we present a small simulation study that indicates that the BSSLM can offer advantages over the direct Lyapunov lasso and the `loglik-0.01` method for  $C$  diagonally unknown.

The problems get computationally even harder when solving for  $M$  and  $C$ . Therefore, we only consider problems of size  $p = 10$ . For the drift matrices, we consider the setting outlined at the beginning of Section 4.2.3 with the difference that the edge probability is set to  $d = 2/10$ . We consider 100 choices of drift matrices  $M^*$ . Simultaneously, we select 100 diagonal matrices  $C^*$  where the entries are selected according to  $\mathcal{N}(1, \sigma)$  with  $\sigma \in \{0, 0.1, 0.5, 1, 10\}$ . We hereby truncate smaller values at 0.1. Then, we generate

#### 4.4 Diagonally Unknown Volatility Matrix

$n = 1000$  observations from a multivariate Gaussian distribution with covariance matrix solving the Lyapunov equation for the pairs  $(M^*, C^*)$ . We apply (4.26) using an equidistant grid of sparsity levels  $k$  ranging from  $k = 1$  to  $k = \binom{p+1}{2}/3$  of length 20. The time limit is set to 400 seconds per value of  $k$ . The initialization method is the projected gradient descent with the number of runs being set to 50 and the maximum number of iterations to 1000. We carry out the estimation procedure for  $\kappa = \{0.001, 0.01, 0.1, 1, 10, \infty\}$ . With “ $\infty$ ” we refer to the case  $C$  known.

First, we standardize the estimated volatility matrix  $\hat{C}$  by multiplying it with  $c_{11}^*$  to put both on the same scale. We calculate  $C_{\text{diff}} = \|\hat{C} \cdot c_{11}^* - C^*\|_1$  to measure the difference between the estimated and the true covariance matrix. We average out  $C_{\text{diff}}$  over the 100 choices of pairs  $(M^*, C^*)$ , but also calculate the minimum. We plot both metrics for the different values of  $\kappa$  where the x-axis is given by the increasing standard deviation  $\sigma$ . We display the results in Figure 4.13.



**Figure 4.13:** The method applied is the BSSLM for  $C$  diagonally unknown (4.26). **Left:**  $C_{\text{diff}}$  averaged out over 100 randomly selected pairs  $(M^*, C^*)$ . **Right:** The minimum of  $C_{\text{diff}}$  across 100 randomly selected pairs  $(M^*, C^*)$ .

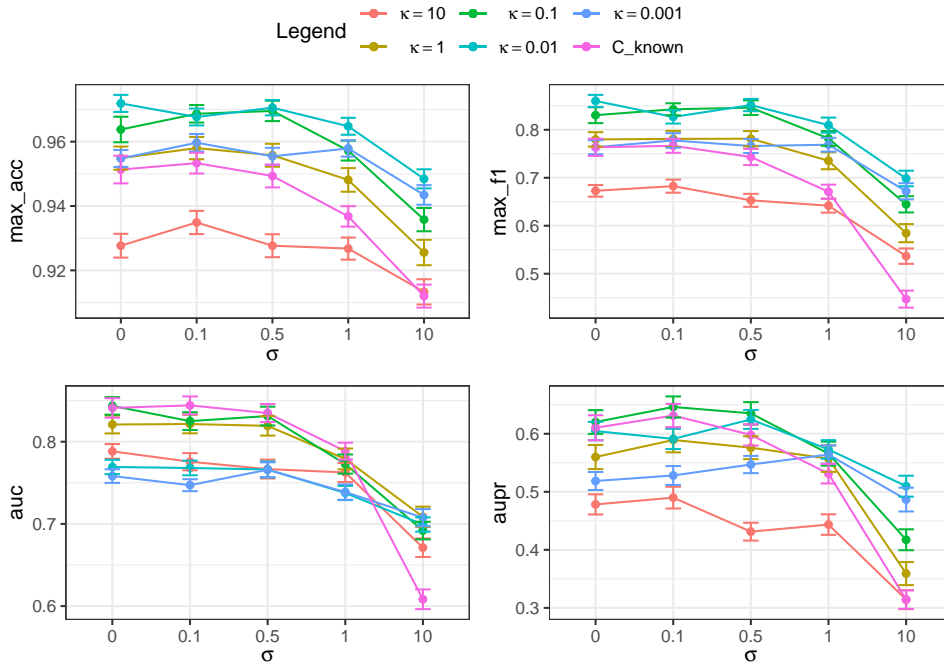
We observe that the tendencies for the average and minimum  $C_{\text{diff}}$  across 100 randomly selected pairs  $(M^*, C^*)$  are pretty similar. The larger the parameter  $\kappa$  is, the closer the matrix  $C$  is to  $I_p$ . As we expect, the case  $C_{\text{known}}$  leads to the worst results as the matrix  $C$  is not estimated. The only exception is  $\sigma = 0$ , which is clear as there is nothing estimated, and the matrix used for data generation equals the one for estimation ( $C = I_p$ ). The larger choices for  $\kappa$ , namely  $\kappa = \{1, 10\}$  produce significantly worse results than the smaller choices  $\kappa = \{0.001, 0.01, 0.1\}$ . Even though the difference between the smaller values of  $\kappa$  and the case  $C$  known is visible, the

estimation results are far from perfect. Regarding the minimal difference, we observe for  $\kappa = \{0.01, 0.001\}$  and  $\sigma = \{0.5, 1, 10\}$  that the results are four times better than for  $C$  known. A deeper look into the estimates reveals that the way the penalty is formed, the entries are mostly correctly ordered but are closer to the identity matrix than those in  $C^*$ . This trade-off is made to make it a feasible optimization problem.

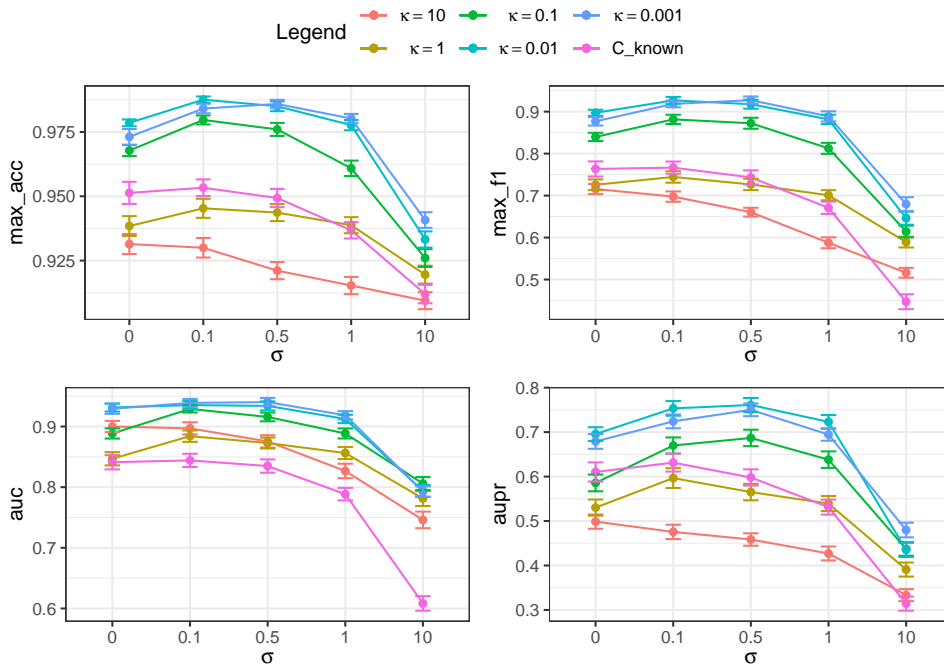
Estimating  $(M, C)$  not only provides information on the volatility matrix but might also be beneficial for the estimation results for the drift matrix  $M$ . The simulations used for Figure 4.13 also produce estimation results for  $M$ . We use the metrics maximum accuracy (`max_acc`), maximum  $f_1$ -score (`max_f1`), the area under the ROC curve (`auc`), and the area under the precision curve (`aupr`). More details are provided in Definition 3.5.11. We display the simulation results in Figure 4.14.

Interestingly, the results for `C_known` are not the worst ones per se. The results for  $\kappa = 10$  are the worst, which is natural as this choice of  $\kappa$  essentially mimics the case `C_known`, but with a more involved problem structure. Not across all metrics, but in particular for the `max_acc` and the `max_f1`-score, the lower  $\kappa$ -values seem to produce better results than `C_known` for  $\sigma = 0$ . The flexible setup for smaller  $\kappa$ -values might be favorable in some instances. Generally, we observe that all approaches are much worse for  $\sigma = 10$  than for the milder changes in the diagonal. However, the decrease for `C_known` is much worse than for any other method. The results for the smaller  $\kappa$ -values are stable for  $\sigma = \{0, 0.1, 0.5\}$ . The decrease for  $\sigma = \{1, 10\}$  for  $\kappa = 0.01$  is the mildest. For Figure 4.14 we use datasets of size  $n = 1000$ . To provide more insights, we carry out the same calculations for  $n = \infty$  which means that no data is sampled, but the true covariance matrix  $\Sigma^*$  is directly inputted into the BSSLM. The results are displayed in Figure 4.15.

#### 4.4 Diagonally Unknown Volatility Matrix

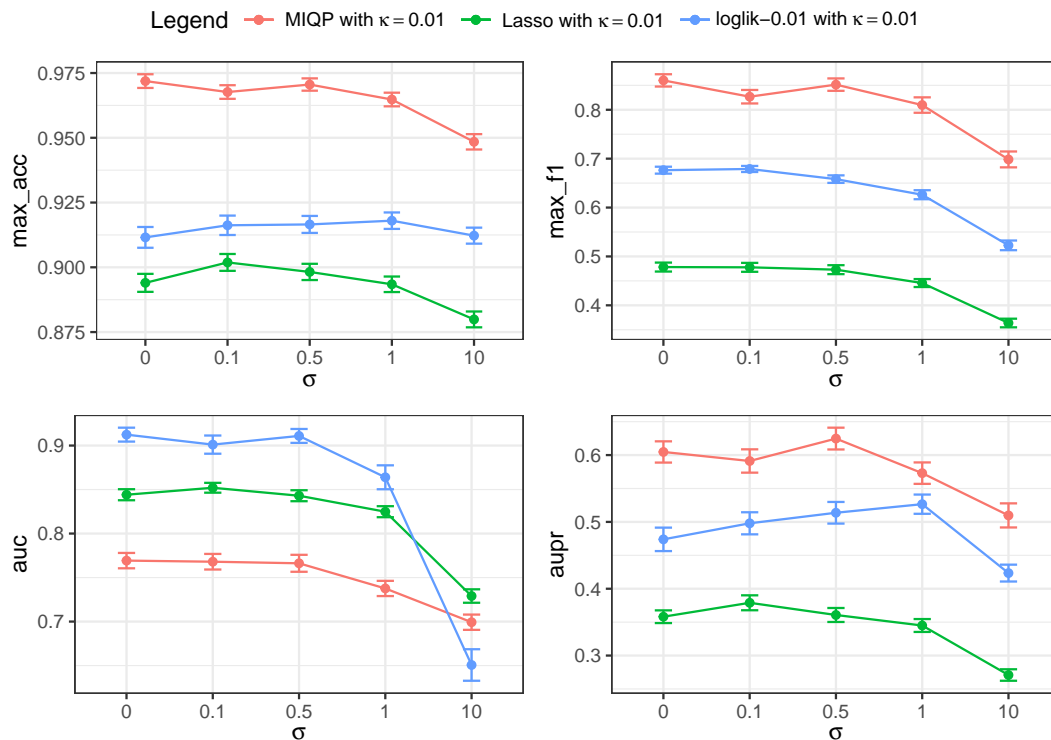


**Figure 4.14:** The method applied is the BSSLM for  $C$  diagonally unknown (4.26) and using datasets of size  $n = 1000$ . The graphic shows four evaluation metrics comparing support recovery averaged out over 100 randomly selected pairs  $(M^*, C^*)$ .



**Figure 4.15:** The method applied is the BSSLM for  $C$  diagonally unknown (4.26) and using the true covariance matrices  $\hat{\Sigma}$  as input. The graphic shows four evaluation metrics comparing support recovery averaged out over 100 randomly selected pairs  $(M^*, C^*)$ .

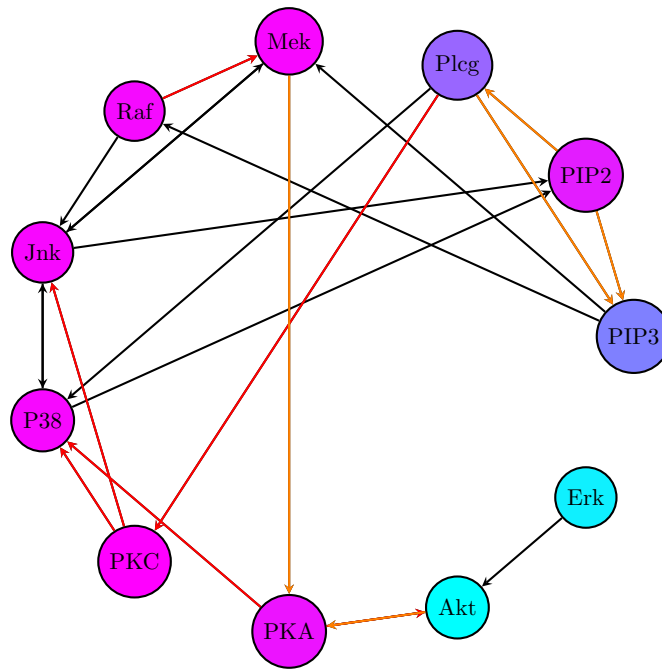
The results confirm the observations that were previously made but are a little more stable, which is not surprising as “perfect” data is used. The dominance of the method of for  $\kappa = \{0.001, 0.01\}$  over  $C$ -known is even more obvious. To conclude this section, we use the exact same simulation setting, but we compare the BSSLM for  $C$  diagonally unknown (4.26) with the direct Lyapunov lasso for  $C$  diagonally unknown (4.25) and with the loglik-0.01 method (4.7). The  $\kappa$ -value is set to 0.01 for all methods. The results are displayed in Figure 4.16.



**Figure 4.16:** Comparing the model selection methods loglik-0.01 (4.7), direct Lyapunov lasso (4.25) and the BSSLM (4.26) using four evaluation metrics comparing support recovery averaged out over 100 randomly selected pairs  $(M^*, C^*)$ .

We observe that the BSSLM performs best for all metrics except for the auc. This holds true no matter what the variance of the diagonal entries of  $C^*$  is. The inferior results for the auc are also observed in Figure 4.5. The  $\ell_1$ -penalized methods tend to produce estimates with many symmetric entries, for instance consider the examples in Section 4.1. Oftentimes, they are present for large parts of the grid. This results for relatively sparse graphs in a high tpr while the fpr is moderate. Therefore, the auc values are quite good. The BSSLM includes variables more carefully which results in inferior auc values, but better aupr values.

Overall, the results show that there exist cases where the BSSLM for  $C$  diagonally unknown can offer advantages over the  $\ell_1$ -penalized methods and also provides information about the volatility matrix  $C$ .



**Figure 4.17:** Estimated Sachs Network using the BSSLM for  $C$  diagonally unknown (4.26) and selecting the structure according to the extended BIC (3.23) with  $\gamma = 1$  (Dataset 1)

#### 4.4.2 Application - Sachs Dataset with the BSSLM for Diagonally Unknown Volatility Matrix and EBIC

This short section is an extension of Section 4.3.1 using the BSSLM for  $C$  diagonally unknown (4.26). The simulation setup is exactly the same as in Section 4.3.1. The only difference is that we jointly estimate the drift matrix  $M$  and the diagonal entries of the volatility matrix  $C$ . The relative comparison of the entries in  $C$  is displayed using the coloring presented in Figure 1.3. The results are displayed in Figure 4.17.

The estimate has more edges than the one for  $C$  known for dataset 2 presented in Figure 4.12. Again, the estimate contains many important connections as the direct enzyme-substrate relationship  $\text{Raf} \rightarrow \text{Mek}$ . For the other enzyme-substrate connection  $\text{PKA} \rightarrow \text{Raf}$ ,  $\text{Mek} \rightarrow \text{Erk}$  and  $\text{Pcl-}\gamma \rightarrow \text{PIP}_2$  are either the reverse directions or treks via other nodes present. The estimate contains an edge from  $\text{Pcl-}\gamma \rightarrow \text{PIP}_3$  which the authors Sachs et al. [2005] mention could be another potential true connection of the network. It is also worth mentioning that the two “triangles”  $\text{Jnk}, \text{P38}, \text{PKC}$  and  $\text{Pcl-}\gamma, \text{PIP}_2$  and  $\text{PIP}_3$  are well-estimated and connected contrary to the estimate for  $C$  known. On the other side, the estimate contains more false positives. We observe that most of the diagonal entries of the volatility matrix are estimated to be relatively equal in size (pink), two are medium-sized (violet), and two are relatively small (turquoise). Despite there not being a direct interpretation of these values by Sachs et al. [2005] or by Varando and Hansen [2020] for this modeling setup, they provide additional information and lead to a more stable estimation procedure.

## 4.5 Summary of the Chapter

In this chapter, we connected the best subset selection, that recently gained popularity due to modern solvers, with graphical continuous Lyapunov models. We showed that the problem sizes that are considered in sparse regression settings do not translate to the estimation of the slightly denser drift matrices in the context of GCLMs. Therefore, the time consumption seems to be the bottleneck for a broader range of applications. We conclude that problems up to sizes  $25 \times 25$  can extensively be studied in a reasonable amount of time when only estimating the drift matrix. Furthermore, the strength of the BSSLM becomes particularly apparent when the active variables are clearly distinguishable from the zero entries. There the method clearly dominates the  $\ell_1$ -penalized methods. We also extended the BSSLM and the direct Lyapunov lasso to the case  $C$  diagonally unknown where we showed that the BSSLM is also able to dominate the  $\ell_1$ -penalized methods. Moreover, information about the diagonal of the volatility matrix is provided. As a by-product we discovered that the  $\ell_1$ -penalized methods tend to produce estimates with a lot of symmetry where the undirected structure is much better estimated than the directed structure.



# Chapter 5

## Conclusion of the Thesis

### Content and Takeaway of the Thesis

The idea for graphical continuous Lyapunov models and the direct Lyapunov lasso is relatively new and was raised by Fitch [2019]. In independent work by Varando and Hansen [2020], an additional likelihood-based estimation approach was introduced. However, the theory for the GCLMs was very limited prior to this thesis. The goal of this thesis was to kick-start the theoretical discussion of Lyapunov models and, at the same time, explore alternative approaches to model selection.

In Chapter 2, we address a fundamental question for statistical models, namely, parameter identifiability. We consider the setting where the volatility matrix is assumed to be known. The main result is the proof that a GCLM based on a simple graph is globally identifiable. This comes with many subtle observations concerning the structure of  $A(\Sigma)$  and with an equivalence result regarding global identifiability for simple graphs and diagonal volatility matrices. Furthermore, we provide a necessary criterion for the generic identifiability of non-simple graphs. The results of this chapter increase the validity of GCLMs as there exist unique solutions for the parameters in  $\mathcal{M}_{G,C}$ , which becomes immediately necessary in Chapter 3. Furthermore, the overall gain in information about the structure of  $A(\Sigma)$  might come in handy for further theoretical analysis.

In Chapter 3, we show that the Direct Lyapunov Lasso can be, under certain assumptions, a consistent estimation method for the drift matrix  $M$  when assuming the volatility matrix  $C$  to be known. Despite the restrictiveness of the irrepresentability condition, it is important to note that a consistent model selection method for GCLMs exists. In addition, the probabilistic result has a reasonable sample size requirement given the complexity of the structure of the “design matrix”  $A(\Sigma)$ . The structure of  $A(\Sigma)$  is also the cause why the irrepresentability condition is much more subtle than in classical regression settings. We provide a detailed study in the context of GCLMs.

We start Chapter 4 by mentioning one of the main problems of the Direct Lyapunov Lasso. The method tends to produce estimates with a lot of symmetry where the undirected structure is much better estimated than the directed structure. This motivated studying a variant of the best subset selection as a model selection method for GCLMs. The problem allows direct control of the number of active variables but is computationally expensive. Despite its computational bottleneck, we show that the method is able to outperform the  $\ell_1$ -penalized methods in certain settings. This is particularly obvious if the active variables (nonzero entries in the true drift matrix) are clearly distinguishable from zero. We adapted the Direct Lyapunov Lasso from

Chapter 3 and we also adapted the BSSLM such that they are able to jointly estimate the drift matrix  $M$  and the diagonal of the volatility matrix  $C$ . For the BSSLM, we gain information about the diagonal elements in  $C$ , but at the same time, we obtain excellent simulation results for the drift matrix  $M$  in the considered setting.

We applied the Direct Lyapunov Lasso and the BSSLM in Chapter 3 and Chapter 4 onto a dataset containing the simultaneous measurements of 11 proteins. The way the dataset is collected makes it an interesting application for GCLMs despite the ambitious parametric assumptions. Using the extended Bayesian information criterion in combination with the Direct Lyapunov Lasso or the BSSLM, we were able to present estimates that recover large parts of an among scientists accepted protein-signalling network.

### Potential Subjects of further Research

The theory for structural equation models has been developed in a period of multiple decades [Maathuis et al., 2019]. As the graphical continuous Lyapunov models are supposed to compete with them, quite a few questions need to be studied. This thesis is a good starting point for further analysis.

Starting from Chapter 2, the natural follow-up is to analyze parameter identifiability for  $C$  (partially) unknown. A starting point could be analyzing  $C$  diagonally unknown. Our observations regarding the structure of  $A(\Sigma)$  might be particularly helpful.

A related topic is the distributional equivalence of graphical continuous Lyapunov models. For structural equation models, this is done by Nowzohour et al. [2017], for instance. Solving this question immediately impacts other aspects of the graphical models. For model selection, it is beneficial if sets of models can be formed that best explain the observational data that is considered.

We extensively studied the model selection methods Direct Lyapunov Lasso in Chapter 3 and the BSSLM in Chapter 4. A detailed study on support recovery for the case  $C$  unknown could be an exciting subject of further research. Moreover, the symmetry of the Direct Lyapunov Lasso estimates and the fact that the undirected structure is very well estimated motivates a theoretical investigation of support recovery for the undirected structure.

# Appendix A

## Failure of Entry-Wise Concentration Inequalities

We mention in Chapter 3 that the careful analysis of the Hessian  $\Gamma = A(\Sigma)^\top A(\Sigma)$  in Section 3.2 and the subsequent derivation of concentration results using spectral properties of the Hessian in Section 3.4 is required to arrive at a reasonable sample size requirement in Corollary 3.3.3. Previously, we attempted a derivation of an entry-wise concentration result. This is of course possible, however, yields a much higher sample size requirement. In this section, we present the entry-wise analysis that reveals some difficulties when dealing with the direct Lyapunov lasso compared to Lasso regression or usual Gaussian graphical models; for instance, consider the work by Lin et al. [2016]. We show that the complexity obtained by the entry-wise concentration inequalities ( $n = \Omega(\log p \tilde{d}^2 p^2)$ ) is much higher than the one that we obtain in Corollary 3.3.3 ( $n = \Omega(dp)$ ). By  $\tilde{d}$  we denote the total number of non-zero entries in the true drift matrix  $M^*$

As in Section 3.4, the derivation of the concentration result is based on a concentration result for Gaussian covariance matrices, [Ravikumar et al., 2011, Lemma 1].

**Lemma A.0.1.** *Consider a zero-mean random vector  $(X_1, \dots, X_p)$  with covariance  $\Sigma^*$  such that each  $X_i / \sqrt{\Sigma_{ii}^*}$  is sub-Gaussian with parameter  $\sigma$ . Given  $n$  i.i.d. samples, the associated sample covariance  $\hat{\Sigma}^n$  satisfies the tail bound*

$$\mathbb{P}[|\hat{\Sigma}_{ij}^n - \Sigma_{ij}^*| > \epsilon] \leq 4 \exp \left\{ -\frac{n\epsilon^2}{128(1+4\sigma^2)^2 \max_i (\Sigma_{ii}^*)^2} \right\}$$

for all  $\epsilon \in (0, \max_i (\Sigma_{ii}^*) 8(1+4\sigma^2))$ .

Throughout the remainder of this section, we use the notation

$$\Delta_\Sigma := \hat{\Sigma} - \Sigma^* \quad \text{and} \quad c^* = \|\Sigma^*\|_\infty \tag{A.1}$$

with  $\Delta_\Gamma$  and  $\Delta_g$  being defined equivalently. Applying a union bound, the above result can be transformed to a concentration result for  $\|\Delta_\Sigma\|_\infty$ .

**Lemma A.0.2.** *Under the assumptions of A.0.1 it holds that*

$$\mathbb{P}(\|\Delta_\Sigma\|_\infty > \epsilon) \leq 4 \binom{p+1}{2} \exp \left\{ -\frac{n\epsilon^2}{128(1+4\sigma^2)^2 \max_i (\Sigma_{ii}^*)^2} \right\}.$$

*Proof.* Union bound and application of Lemma A.0.1. □

## Appendix A Failure of Entry-Wise Concentration Inequalities

At this point an improvement may be possible by deriving or applying direct concentration inequalities for  $\|\Delta_\Sigma\|_\infty$ . To our knowledge, there is no such result yet. The subsequent calculations to arrive at concentration inequalities for  $g, \Gamma$  are deterministic. Similar to Lemma 3.4.2 and Lemma 3.4.3, we have to express  $\|\Delta_\Gamma\|_\infty$  and  $\|\Delta_g\|_\infty$  by  $\|\Delta_\Sigma\|_\infty$  to apply Lemma A.0.2.

**Lemma A.0.3.** *Let  $\Delta_\Sigma$  and  $c^*$  be defined as in (A.1). For the estimation error of the product of two covariances it holds that*

$$|\hat{\Sigma}_{ij}\hat{\Sigma}_{kl} - \Sigma_{ij}^*\Sigma_{kl}^*| \leq 2c^*\|\Delta_\Sigma\|_\infty + \|\Delta_\Sigma\|_\infty^2. \quad (\text{A.2})$$

**Proof.** We define  $w_{ij} = \hat{\Sigma}_{ij} - \Sigma_{ij}^*$ , then

$$\begin{aligned} \hat{\Sigma}_{ij}\hat{\Sigma}_{kl} - \Sigma_{ij}^*\Sigma_{kl}^* &= (\Sigma_{ij}^* + w_{ij})(\Sigma_{kl}^* + w_{kl}^*) - \Sigma_{ij}^*\Sigma_{kl}^* \\ &= \Sigma_{ij}^*w_{kl} + w_{ij}\Sigma_{kl}^* + w_{ij}w_{kl} \end{aligned}$$

Using the definition of  $c^*$  and  $\delta$  we obtain

$$|\hat{\Sigma}_{ij}\hat{\Sigma}_{kl} - \Sigma_{ij}^*\Sigma_{kl}^*| \leq c^*\|\Delta_\Sigma\|_\infty + \|\Delta_\Sigma\|_\infty c^* + \|\Delta_\Sigma\|_\infty^2 = 2c^*\|\Delta_\Sigma\|_\infty + \|\Delta_\Sigma\|_\infty^2$$

□

Applying a union bound yields a straightforward result for  $\Gamma$  as well.

**Lemma A.0.4.** *Let  $\Delta_\Sigma, \Delta_\Gamma$  and  $c^*$  be defined as in (A.1). For the maximal error in  $\Gamma$  it holds that*

$$\|\Delta_\Gamma\|_\infty \leq (2p+2)(2c^*\|\Delta_\Sigma\|_\infty + \|\Delta_\Sigma\|_\infty^2).$$

**Proof.** Note that the most critical entries in  $\Gamma = A(\Sigma)^\top A(\Sigma)$  consist of a sum of length  $p$  of products of two covariances where one coefficient is 4 and the remaining are 2 (A visualization of the structure of  $A(\Sigma)$  for  $p=3$  is given in Example 3.2.1). Thus, using Lemma A.0.3 we can bound

$$\begin{aligned} \|\Delta_\Gamma\|_\infty &\leq 4(2c^*\|\Delta_\Sigma\|_\infty + \|\Delta_\Sigma\|_\infty^2) + 2(p-1)(2c^*\|\Delta_\Sigma\|_\infty + \|\Delta_\Sigma\|_\infty^2) \\ &= (2p+2)(2c^*\|\Delta_\Sigma\|_\infty + \|\Delta_\Sigma\|_\infty^2). \end{aligned}$$

□

In fact, most of the entries in  $\Gamma$  are not of the type of the worst case entries in Lemma A.0.4, but can be bounded by  $2(2c^*\|\Delta_\Sigma\|_\infty + \|\Delta_\Sigma\|_\infty^2)$ . This come in handy when bounding  $\|\Delta_\Gamma\|_\infty$  or  $\|(\Delta_\Gamma)_{\cdot,S}\|_\infty$  for  $S = \{(i,j) : i,j \in [p]\}$ . Depending on the length of the entries included in the rows of  $\Gamma_{\cdot,S}$  we either add  $(2c^*\delta + \delta)$  or  $(2p+2)(2c^*\delta + \delta)$  out of the modular system to the bound.

Finally, we formulate the concentration inequality for  $\Gamma$ . We will require that  $\frac{\epsilon}{2p+6} < 1$ . This is only for technical reasons and is always fulfilled for a small  $\epsilon$ .

**Lemma A.0.5.** *Let the same assumption apply as in Lemma A.0.1. Furthermore, let  $\frac{\epsilon}{2p+6} < 1$  and let  $\Delta_\Sigma, \Delta_\Gamma, c^*$  be as in (A.1). Let  $c^{**} = \sqrt{(c^*)^2 + 1} + c^*$ . Then it holds that*

$$\mathbb{P}(\|\Delta_\Gamma\|_\infty > \epsilon) < 4 \binom{p+1}{2} \exp \left\{ -\frac{n(\epsilon/(2p+6)c^{**})^2}{128(1+4\sigma^2)^2 \max_i (\Sigma_{ii}^*)^2} \right\}.$$

**Proof.** Making use of the inequality presented in Lemma A.0.4 we obtain

$$\mathbb{P}(\|\Delta_\Gamma\|_\infty > \epsilon) \leq \mathbb{P}((2p+6)(2c^*\|\Delta_\Gamma\|_\infty + \|\Delta_\Gamma\|_\infty^2) > \epsilon).$$

To bound  $\|\Delta_\Gamma\|_\infty$  we solve the quadratic equation

$$\|\Delta_\Gamma\|_\infty^2 + 2c^*\|\Delta_\Gamma\|_\infty - \frac{\epsilon}{(2p+2)} = 0.$$

After simplification we obtain

$$\|\Delta_\Gamma\|_\infty = -c^* \pm \sqrt{(c^*)^2 + \frac{\epsilon}{2p+2}}.$$

Since  $\|\Delta_\Gamma\|_\infty$  is supposed to be positive, we only take

$$\|\Delta_\Gamma\|_\infty = -c^* + \sqrt{(c^*)^2 + \frac{\epsilon}{2p+2}}$$

into account. Therefore

$$\mathbb{P}((2p+6)(2c^*\|\Delta_\Gamma\|_\infty + \|\Delta_\Gamma\|_\infty^2) > \epsilon) = \mathbb{P}\left(\|\Delta_\Sigma\|_\infty > -c^* + \sqrt{(c^*)^2 + \frac{\epsilon}{2p+2}}\right).$$

Under the assumption that  $\frac{\epsilon}{2p+2} < 1$  we obtain

$$-c^* + \sqrt{(c^*)^2 + \frac{\epsilon}{2p+2}} > \frac{\epsilon/(2p+2)}{\sqrt{(c^*)^2 + 1} + c^*} = \frac{\epsilon/(2p+2)}{c^{**}}.$$

In total we can bound

$$\mathbb{P}(\|\Delta_\Gamma\|_\infty > \epsilon)$$

through

$$\mathbb{P}\left(\|\Delta_\Sigma\|_\infty > \frac{\epsilon}{(2p+2)c^{**}}\right).$$

Employing Lemma A.0.2 leads to

$$\mathbb{P}\left(\|\Delta_\Sigma\|_\infty > \frac{\epsilon}{(2p+2)c^{**}}\right) \leq 4 \binom{p+1}{2} \exp \left\{ -\frac{n(\epsilon/(2p+2)c^{**})^2}{128(1+4\sigma^2)^2 \max_i (\Sigma_{ii}^*)^2} \right\}.$$

□

We define  $C_{\max} := \max_i |C_{ii}|$ . Remember that we assume  $C$  to be diagonal as we only consider directed graphs. We formulate the concentration inequality for  $g$  which places less stringent conditions on the sample size than the one for  $\Gamma$ .

## Appendix A Failure of Entry-Wise Concentration Inequalities

**Lemma A.0.6.** *Let the same assumption apply as in Lemma A.0.1 and let  $\delta$ ,  $c^*$  be as in (A.1), then it holds that*

$$\mathbb{P}(\|\Delta_g\|_\infty > \epsilon) \leq 4 \binom{p+1}{2} \exp \left\{ -\frac{n(\epsilon/2C_{\max})^2}{128(1+4\sigma^2)^2 \max_i (\Sigma_{ii}^*)^2} \right\}.$$

**Proof.** Inserting the definition of  $g$  we obtain

$$\begin{aligned} \|\Delta_g\|_\infty &= 2\|\text{vec}(C)^T A(\hat{\Sigma}) - \text{vec}(C)^T A(\Sigma^*)\|_\infty \\ &= 2\|\text{vec}(C)^T \cdot (A(\hat{\Sigma}) - A(\Sigma^*))\|_\infty \\ &\leq 2C_{\max} \cdot \|\Delta_\Sigma\|_\infty. \end{aligned}$$

Therefore

$$\begin{aligned} \mathbb{P}(\|\Delta_g\|_\infty > \epsilon) &\leq \mathbb{P}\left(\|\Delta_\Sigma\|_\infty > \frac{\epsilon}{2C_{\max}}\right) \\ &\leq 4 \binom{p+1}{2} \exp \left\{ -\frac{n(\epsilon/2C_{\max})^2}{128(1+4\sigma^2)^2 \max_i (\Sigma_{ii}^*)^2} \right\}. \end{aligned}$$

□

The Lemma above indicates that the concentration inequality for  $g$  is not as important as the concentration inequality for  $\Gamma$  when aiming for precision with a minimal amount of samples required. The exact amount is ultimately determined when using the concentration inequalities on an adapted version of Theorem 1 in Lin et al. [2016] or, phrased differently, on an adapted version of Theorem 3.3.1 to arrive at a probabilistic guarantee. We do not present the whole Theorem, but only the requirements that influence the sample size when deriving a probabilistic result. Instead of requiring that  $\|\Delta_\Gamma\|_\infty < \epsilon_1$  for  $\epsilon_1 > 0$  and  $\epsilon_1 \leq \alpha/(6c_{\Gamma^*})$  as in Theorem 3.3.1, it has to hold that  $\|\Delta_\Gamma\|_\infty < \epsilon_1$  and  $2\tilde{d}\epsilon_1 \leq \alpha/(6c_{\Gamma^*})$  with  $\tilde{d}$  being the total number of non-zero entries in the true drift matrix  $M^*$ .

**Lemma A.0.7.** *Let*

$$\epsilon_1 = \sqrt{\frac{\tilde{c}(\log p^{\tau_1} + \log 4)}{n}},$$

then  $P(\|\Delta_\Gamma\|_\infty > \epsilon_1) < 1$  if

$$\tau_1 \geq \left(\frac{\log 4}{\log p} + 2\right) \max \left\{ (c^{**})^2 (2p+2)^2, 4C_{\max}^2 \right\}.$$

**Proof.** Without changing the order of magnitude of the overall result we substitute the factor  $\binom{p+1}{2}$  with  $p^2$  in Lemma A.0.5 as  $\binom{p+1}{2} \leq p^2$ . Inserting  $\epsilon_1$  into the bound

$$\mathbb{P}(\|\Delta_\Gamma\|_\infty > \epsilon_1) < 4^{1-1/(2p+2)^2(c^{**})^2} p^{2-\tau_1/(2p+2)^2(c^{**})^2} < 4p^{2-\tau_1/(2p+2)^2(c^{**})^2} \quad (\text{A.3})$$

We choose  $\tau_1$  such that the right side in (A.3) is less than 1 making it a valid probabilistic statement. We require that  $p^{2-\frac{\tau_1}{(2p+2)^2(c^{**})^2}}$  is less than  $\frac{1}{4}$ . Therefore, we have to solve

$$p^x \leq \frac{1}{4} \iff x \leq -\frac{\log 4}{\log p}.$$

It has to hold that

$$2 - \frac{\tau_1}{(2p+2)^2(c^{**})^2} \leq -\frac{\log 4}{\log p}.$$

Therefore

$$\tau_1 \geq \left( \frac{\log 4}{\log p} + 2 \right) (c^{**})^2 (2p+2)^2.$$

This implies that  $\mathbb{P}(\|\Delta_\Gamma\|_\infty > \epsilon_1) < 1$  and more precisely

$$\mathbb{P}(\|\Delta_\Gamma\|_\infty > \epsilon_1) < 4^{1-1/(2p+2)^2(c^{**})^2} p^{2-\tau_1/(2p+2)^2(c^{**})^2}$$

for our choice of  $\tau_1$ . □

We mentioned that it also has to hold that  $2\tilde{d}\epsilon_1 \leq \alpha/(6c_{\Gamma^*})$ . Using

$$n > 4\tilde{c}c_1^2\tilde{d}^2(\log p^{\tau_1} + \log 4) \tag{A.4}$$

this holds true. In total this results in requiring that  $n = \Omega(\log p \tilde{d}^2 p^2)$  which is much worse than  $n = \Omega(dp)$  in Corollary 3.3.3. This is quite intriguing as entry-wise concentration inequalities are oftentimes used in the context of Lasso problems [Hastie et al., 2015] and in the context of Gaussian graphical models [Ravikumar et al., 2011].

## Appendix B

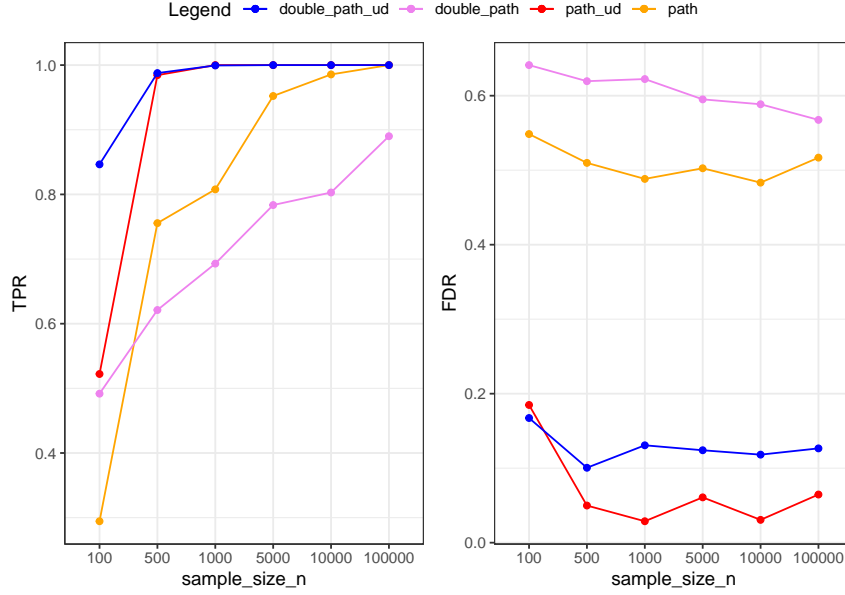
### Additional Simulations

#### B.1 Direct Lyapunov Lasso with BIC and EBIC

In this section, we apply the direct Lyapunov lasso with BIC or EBIC on synthetic data. Consider the graph structures in Figure B.1.1. Instead of 5 nodes, we consider these graphs with 11 nodes matching the number of variables in the dataset by Sachs et al. [2005]. We fix  $C = 2I_p$  and consider  $M^*$  supported over the path or the double path. Furthermore, we set the diagonal entries in  $M^*$  to be -2 and choose 1 for the off-diagonal entries. We generate 100 times  $n = \{100, 500, 1000, 5000, 10000, 100000\}$  observations from a multivariate Gaussian distribution  $N(0, \Sigma^*)$  with covariance matrix  $\Sigma^* = \Sigma(M^*, C)$  solving the Lyapunov equation for  $(M^*, C)$ . As introduced, we apply the direct Lyapunov lasso with the BIC criterion or the EBIC criterion to obtain an estimated drift matrix  $\hat{M}$ . For each value of  $n$  and for both graph structures, we calculate the true positive rate (TPR) and the false discovery rate (FDR) comparing the estimate  $\hat{M}$  with the true drift matrix  $M^*$  based on the directed graph and the skeleton. For the TPR we refer to Definition 3.5.11 and the FDR is given by  $fp/(fp + tp)$ .



**Figure B.1.1:** Left: The path 1 to 5. Right: The Double Path from 1 to 5.

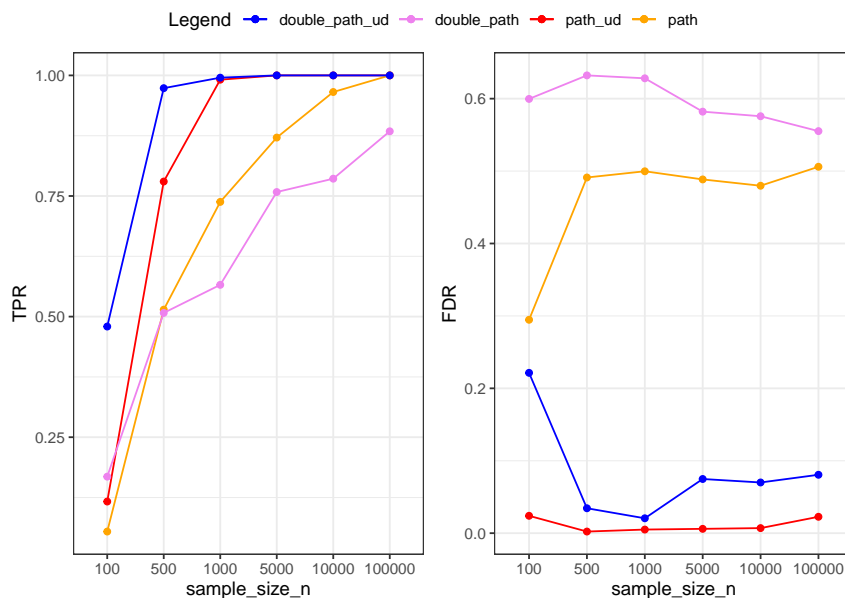


**Figure B.1.2:** TPR (true positive rate) and FDR (false discovery rate) of the estimated drift matrix obtained with the direct Lyapunov lasso (3.4) and BIC for increasing sample size. The results are averaged over 100 datasets generated by the same drift matrix.

In Figure B.1.2 we present the estimation results for the BIC criterion. We label the results only using the skeleton by “ud” for “undirected structure”. With increasing sample size we observe an increase in the TPR. For the path and the skeleton of the double path, we observe a TPR of 1 for  $n = 100000$ . This translates to all positive entries in the true drift matrix  $M^*$  being recovered by the estimate  $\hat{M}$ . Only for the directed double path not all nonzero entries are recovered, but we observe an increasing trend. The plot of the FDR is intriguing. When only considering the directed path and the directed double path, we observe a surprisingly high FDR of around 0.5 that does not decrease substantially for increasing sample size. This means that only half of the nonzero entries present in  $\hat{M}$  are also present in  $M^*$ . At first glance, this might seem quite bad. However, when looking at the results for the undirected structure, we observe that the FDR is quite low for both the path and the double path for  $n \geq 500$ . This can only be explained in one way. Many of the false positive entries in  $\hat{M}$  coincide with the reversed edges of the path and the double path. The tendency of the direct Lyapunov lasso to select symmetric estimates reoccurs in Section 4.1.



## B.1 Direct Lyapunov Lasso with BIC and EBIC



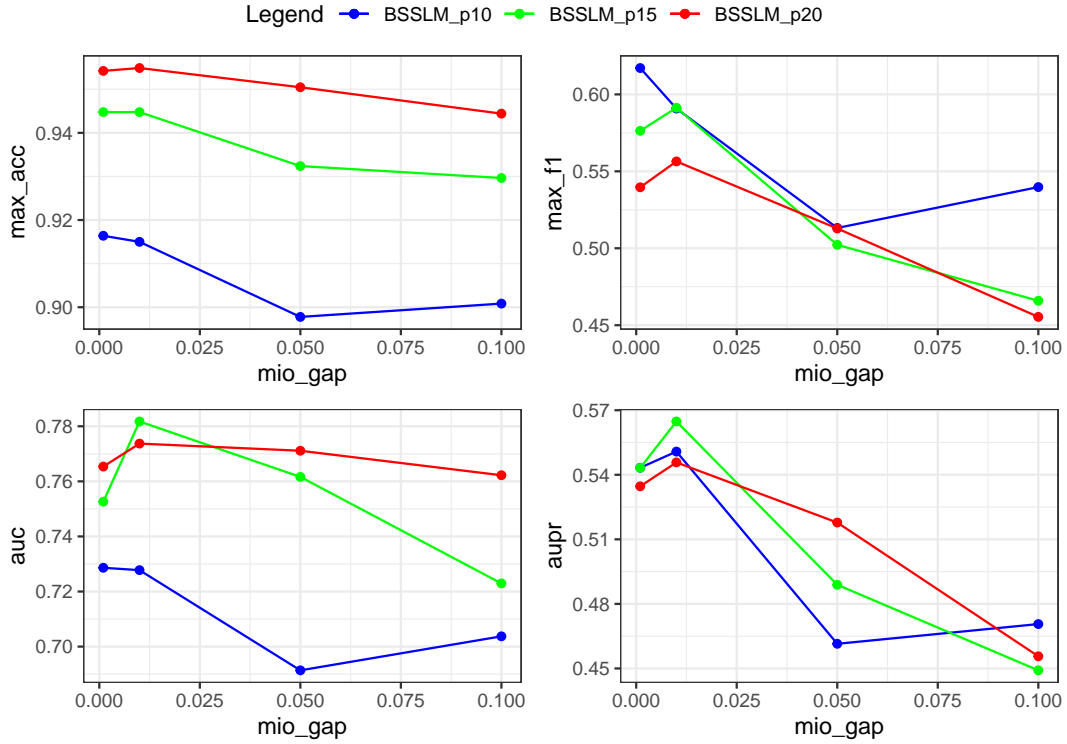
**Figure B.1.3:** TPR (true positive rate) and FDR (false discovery rate) of the estimated drift matrix obtained with the direct Lyapunov lasso (3.4) and EBIC with  $\gamma = 1$  for increasing sample size. The results are averaged over 100 datasets generated by the same drift matrix.

In Figure B.1.3 we present the estimation results for the EBIC criterion with  $\gamma = 1$ . The general behaviour that we observed for the direct Lyapunov lasso with the BIC criterion in Figure B.1.2 also applies to the simulations with EBIC. The differences are very subtle. The TPR increases slightly faster with the BIC criterion than with the EBIC criterion. While the TPR is already around 1 for the skeleton of the path and double path for  $n = 500$ , this only happens with the EBIC for  $n = 1000$ . At the same time, we observe a slightly lower FDR for the EBIC criterion. This holds true for  $n \geq 500$ , but is particularly interesting for  $n = 500, n = 1000$  as the Sachs data set if of sizes  $n = 707$  to  $n = 927$ .

Overall, both methods seem to be a decent choice for model selection alongside with direct Lyapunov lasso when considering synthetic data. In particular, estimating the undirected structure seems to work very well. The theoretical derivation of the BIC and EBIC in Section 3.7.2 is also somewhat reflected by the simulation results. The BIC tends to produce estimates with a higher TPR for sample sizes around  $n = 500, 1000$ . At the same time, the EBIC results in a lower FDR for these sample sizes. Asymptotically, the TPR of the directed and undirected structure is increasing for both selection methods with the TPR being 1 for the undirected structure for higher sample sizes. Neglecting the false positives that are reversed edges of those being present in the true drift matrix, we observe an overall low FDR.

## B.2 Additional Simulations - Comparing the BSSLM w.r.t. Initializations

Here, we present simulation results for varying MIPGap [Gurobi Optimization, LLC, 2023]. We consider the same simulation setting as in Section 4.2.3. The metrics considered are displayed in Definition 3.5.11. The results are displayed in Figure B.2.1.



**Figure B.2.1:** The method applied is the BSSLM (4.17) using  $C = 2I_p$  and using the projected gradient descent initialization a) in Section 4.2.2. The results are displayed for the varying tuning parameter MIPGap (mio\_gap). The number in the label refers to the problem size  $p$ .

We observe that across all problem sizes there is little to no difference between MIPGap 0.001 and MIPGap 0.01. For coarser MIPGaps, there is a more significant decrease in some metrics. Therefore, we select the MIPGap 0.01 as the solution can be certified optimal in more cases due to the milder requirement.

## B.3 BSSLM - Time Consumption and Comparison of Initializations - Additional Information

In this section, we provide additional information for Section 4.2.3.

**Table B.1:** Summary statistics of the time used to calculate the solution of (4.17) for one value of  $k$  across 40 randomly selected drift matrices  $M^*$ . The initialization methods are proj\_grad, lasso and plain which are labelled as methods a)-c) in Section 4.2.2.

prob. size	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
p=10 proj_grad	0.0920	0.1643	0.2032	8.4511	0.4515	322.5318
p=10 lasso	0.0537	0.1130	0.5519	42.0854	0.9855	400.6142
p=10 plain	0.1014	0.2562	0.5500	10.7348	1.0670	401.0197
p=15 proj_grad	0.3585	2.9118	8.0261	78.2780	30.0132	401.1775
p=15 lasso	0.116	1.249	7.452	123.788	400.380	400.771
p=15 plain	1.066	8.463	19.068	117.058	207.980	400.708
p=20 proj_grad	40.81	233.67	401.82	324.86	403.26	404.77
p=20 lasso	3.19	196.30	400.97	306.33	401.17	402.36
p=20 plain	24.33	188.15	400.92	308.06	401.22	401.67

## Appendix C

### Auxiliary Results

In this section, we present auxiliary results that are needed in this work. The first two results concern the understanding of the projected gradient descent initialization in Section 4.2.2. They are already presented in the work by Bertsimas et al. [2016] when the initialization method was introduced for the Best Subset method for regression problems. In Lemma C.0.3 we present the derivative of eq. (4.10) which is the optimization problem introduced by Varando and Hansen [2020].

**Proposition C.0.1** (Proposition 3 in [Bertsimas et al., 2016]). *If a vector  $\text{vec}(\hat{M})$  is an optimal solution to*

$$\text{vec}(\hat{M}) \in \underset{\|\text{vec}(M)\|_0 \leq k}{\text{arg min}} \|\text{vec}(M) - c\|_2^2, \quad (\text{C.1})$$

*then  $\text{vec}(M)$  retains the  $k$  largest (in absolute value) elements of  $c \in \mathbb{R}^p$  and sets the rest to zero, i.e., if  $|c_{(1)}| \geq |c_{(2)}| \geq \dots \geq |c_{(p)}|$ , denote the ordered values of the absolute values of the vector  $c$ , then*

$$\text{vec}(\hat{M})_i = \begin{cases} c_i, & \text{if } i \in \{(1), \dots, (k)\}, \\ 0, & \text{otherwise,} \end{cases}$$

*where  $\text{vec}(\hat{M})_i$  is the  $i$ -th coordinate of  $\text{vec}(\hat{M})$ . We will denote the set of solutions to problem (C.1) by the notation  $H_k(c)$ .*

**Proposition C.0.2** (Proposition 4 [Bertsimas et al., 2016] based on [Nesterov, 2013]). *For a convex function  $g(\text{vec}(M))$  satisfying*

$$\|\nabla g(\text{vec}(M)) - \nabla g(\text{vec}(\tilde{M}))\| \leq l \|\text{vec}(M) - \text{vec}(\tilde{M})\|$$

## Appendix C Auxiliary Results

and for any  $L \geq l$ , we have

$$g(\text{vec}(\tilde{M})) \leq Q_L(\text{vec}(\tilde{M}), \text{vec}(M)) := g(\text{vec}(M)) + \frac{L}{2} \|\text{vec}(\tilde{M}) - \text{vec}(M)\| \quad (\text{C.2})$$

$$+ \langle \nabla g(\text{vec}(M)), \text{vec}(\tilde{M}) - \text{vec}(M) \rangle$$

for all  $\text{vec}(\tilde{M}), \text{vec}(M)$  with equality holding at  $\text{vec}(\tilde{M}) = \text{vec}(M)$ .

Following [Bertsimas et al., 2016, p. 829] and applying Proposition C.0.1, it holds that

$$\arg \min_{\|\text{vec}(\tilde{M})\|_0 \leq k} Q_L(\text{vec}(\tilde{M}), \text{vec}(M)) = H_k \left( \text{vec}(M) - \frac{1}{L} \nabla g(\text{vec}(M)) \right).$$

**Lemma C.0.3.** *The derivative for the first summand in eq. (4.10) is*

$$\nabla_{(M,C)} L(\Sigma(M, C)) = (2\Sigma(M, C)\Sigma(M^\top, \nabla L), 2\Sigma(M^\top, \nabla L)) \quad (\text{C.3})$$

where

$$\nabla L(\Sigma) = \frac{dL(\Sigma)}{d\Sigma} = \Sigma^{-1} - \Sigma^{-1} \hat{\Sigma} \Sigma^{-1}.$$

The derivative for the second summand is

$$\nabla_{(M,C)} \lambda \|M\|_1 = \left( \hat{Z}, \mathbf{0} \right) \quad (\text{C.4})$$

where  $\mathbf{0}$  is a  $p \times p$  matrix of zeros.

$$\hat{Z}_{ij} = \begin{cases} \text{sign}(M_{ij}) & \text{if } M_{ij} \neq 0, \\ \in [-1, 1] & \text{if } M_{ij} = 0. \end{cases} \quad (\text{C.5})$$

The derivative of the third summand is

$$\nabla_{(M,C)} \kappa \|C - I_p\|_F^2 = \kappa (\mathbf{0}, C^\top - 2I_p).$$

**Proof.** For the function  $L(\Sigma(M, C)) = \log \det \Sigma(M, C) + \text{tr}(\hat{\Sigma}(\Sigma(M, C))^{-1})$ , we obtain by differentiating with respect to  $(M, C)$  and using Proposition 3.1 by Varando and Hansen [2020]:

$$\nabla_{(M,C)} L(\Sigma(M, C)) = (2\Sigma(M, C)\Sigma(M^\top, \nabla L), 2\Sigma(M^\top, \nabla L))$$

By  $\nabla L$  we denote the derivative of the function  $\Sigma \rightarrow L(\Sigma)$ . Using the identity  $(\frac{d}{dX} \text{tr}(AX^{-1}B)) = -X^{-1}BAX^{-1}$  given in [Bernstein, 2018, Proposition 10.7.3.] we obtain

$$\frac{d}{d\Sigma} \text{tr}(\hat{\Sigma}\Sigma^{-1}) = -\Sigma^{-1}\hat{\Sigma}\Sigma^{-1}.$$

Using the identity ( $\frac{d}{dX} \log \det AXB = B(AXB)^{-1}A$ ) given in [Bernstein, 2018, Proposition 10.7.2.] we obtain

$$\frac{d}{d\Sigma} \log \det \Sigma = \Sigma^{-1}.$$

This yields

$$\nabla L(\Sigma) = \frac{dL(\Sigma)}{d\Sigma} = \Sigma^{-1} - \Sigma^{-1} \hat{\Sigma} \Sigma^{-1}.$$

The second summand is directly obtained by calculating the subgradient of the  $\ell_1$ -norm. For the third summand, we calculate

$$\begin{aligned} \frac{d}{dC} \|C - I_p\|_F^2 &= \frac{d}{dC} \text{tr}((C - I_p)^\top (C - I_p)) \\ &= \frac{d}{dC} \text{tr}(C^\top C - C^\top I_p - I_p C + I_p) \\ &= \frac{d}{dC} \text{tr}(C^\top C) - \frac{d}{dC} \text{tr}(C^\top I_p) - \frac{d}{dC} \text{tr}(I_p C) \\ &= C^\top - I_p - I_p \\ &= C^\top - 2I_p. \end{aligned}$$

This results in

$$\nabla_{(M,C)} \kappa \|C - I_p\|_F^2 = \kappa(\mathbf{0}, C^\top - 2I_p).$$

□

## Bibliography

- C. Amendola, P. Dettling, M. Drton, F. Onori, and J. Wu. Structure learning for cyclic linear causal models. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 999–1008. PMLR, 2020.
- R. F. Barber, M. Drton, N. Sturma, and L. Weihs. Half-trek criterion for identifiability of latent variable models. *Ann. Statist.*, 50(6):3174–3196, 2022.
- S. Barnett and C. Storey. Analysis and synthesis of stability matrices. *J. Differential Equations*, 3:414–422, 1967.
- S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in real algebraic geometry*, volume 10 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, second edition, 2006.
- D. S. Bernstein. *Scalar, vector, and matrix mathematics*. Princeton University Press, Princeton, NJ, 2018.
- D. Bertsimas and R. Weismantel. *Optimization over integers*. Athena Scientific, 2005.
- D. Bertsimas, A. King, and R. Mazumder. Best subset selection via a modern optimization lens. *Ann. Statist.*, 44(2):813–852, 2016.
- K. Bestuzheva, M. Besançon, W.-K. Chen, A. Chmiela, T. Donkiewicz, J. van Doornmalen, L. Eifler, O. Gaul, G. Gamrath, A. Gleixner, L. Gottwald, C. Graczyk, K. Halbig, A. Hoen, C. Hojny, R. van der Hulst, T. Koch, M. Lübbecke, S. J. Maher, F. Matter, E. Mühmer, B. Müller, M. E. Pfetsch, D. Rehfeldt, S. Schlein, F. Schlösser, F. Serrano, Y. Shinano, B. Sofranac, M. Turner, S. Vigerske, F. Wegscheider, P. Wellner, D. Weninger, and J. Witzig. The scip optimization suite 8.0. Technical Report 21-41, ZIB, Takustr. 7, 14195 Berlin, 2021.
- A. Bhaya, E. Kaszkurewicz, and R. Santos. Characterizations of classes of stable matrices. *Linear Algebra and Its Applications*, 374:159–174, 11 2003.
- S. Bongers and J. M. Mooij. From random differential equations to structural causal models: the stochastic case. *arXiv.org preprint*, 1803.08784, 2018.
- S. Bongers, P. Forré, J. Peters, and J. M. Mooij. Foundations of structural causal models with cycles and latent variables. *Ann. Statist.*, 49(5):2885–2915, 2021.
- C. Brito and J. Pearl. Graphical condition for identification in recursive sem. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 47–54, 2006.
- J. Chen and Z. Chen. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, 95(3):759–771, 09 2008.

- D. Cifuentes, T. Kahle, and P. Parrilo. Sums of squares in macaulay2. *Journal of Software for Algebra and Geometry*, 10:17–24, 03 2020.
- G. Ciolek, D. Marushkevych, and M. Podolskij. On Dantzig and Lasso estimators of the drift in a high dimensional Ornstein-Uhlenbeck model. *Electron. J. Stat.*, 14(2):4395–4420, 2020.
- I. I. Cplex. V12. 1: User’s manual for cplex. *International Business Machines Corporation*, 46(53):157, 2009.
- P. Dettling and M. Drton. On the best subset selection for graphical continuous lyapunov models. unpublished, 2024.
- P. Dettling, R. Homs, C. Améndola, M. Drton, and N. R. Hansen. Identifiability in continuous Lyapunov models. *SIAM J. Matrix Anal. Appl.*, 44(4):1799–1821, 2023.
- P. Dettling, M. Drton, and M. Kolar. On the lasso for graphical continuous lyapunov models. In *Proceedings of the Third Conference on Causal Learning and Reasoning*, pages 514–550. PMLR, 2024.
- M. Drton. Algebraic problems in structural equation modeling. In *The 50th anniversary of Gröbner bases*, volume 77 of *Adv. Stud. Pure Math.*, pages 35–86. Math. Soc. Japan, Tokyo, 2018.
- M. Drton and L. Weihs. Generic identifiability of linear structural equation models by ancestor decomposition. *Scandinavian Journal of Statistics*, 43(4):1035–1045, 2016.
- M. Drton, R. Foygel, and S. Sullivant. Global identifiability of linear structural equation models. *The Annals of Statistics*, 39(2):865–886, 2011.
- M. Drton, C. Fox, and Y. S. Wang. Computation of maximum likelihood estimates in cyclic structural equation models. *Ann. Statist.*, 47(2):663–690, 2019.
- F. M. Fisher. A correspondence principle for simultaneous equation models. *Econometrica*, 38(1):73–92, 1970.
- K. E. Fitch. Learning directed graphical models from Gaussian data. *arXiv*, abs/1906.08050, 2019.
- R. Foygel and M. Drton. Extended bayesian information criteria for gaussian graphical models. In *Proceedings of the 24th Annual Conference on Neural Information Processing Systems 2010*, pages 604–612. Curran Associates, Inc., 2010.
- 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, 2012.
- J. Friedman, R. Tibshirani, and T. Hastie. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.
- M. Gao, W. M. Tai, and B. Aragam. Optimal neighbourhood selection in structural equation models. *arXiv*, abs/2306.02244, 2023.
- S. Gaïffas and G. Matulewicz. Sparse inference of the drift of a high-dimensional ornstein–uhlenbeck process. *Journal of Multivariate Analysis*, 169:1–20, 2019.

## BIBLIOGRAPHY

- J. Ghosh, M. Delampady, and T. Samanta. *An Introduction to Bayesian Analysis: Theory and Methods*. Springer Texts in Statistics. Springer New York, 2007.
- C. Glymour, K. Zhang, and P. Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10, 06 2019.
- Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2023. URL <https://www.gurobi.com>.
- N. Hansen. smde: Sparse multivariate differential equations. <https://rdrr.io/rforge/smde/>, 2014.
- T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC, 2015.
- T. Hastie, R. Tibshirani, and R. Tibshirani. Best subset selection and related tools. <https://github.com/ryantibs/best-subset>, 2020a.
- T. Hastie, R. Tibshirani, and R. J. Tibshirani. Best subset, forward stepwise or Lasso? Analysis and recommendations based on extensive comparisons. *Statist. Sci.*, 35(4): 625–626, 2020b.
- R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1991.
- D. Kumor, B. Chen, and E. Bareinboim. Efficient identification in linear structural causal models with instrumental cutsets. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- L. Lin, M. Drton, and A. Shojaie. Estimation of high-dimensional graphical models using regularized score matching. *Electron. J. Stat.*, 10(1):806–854, 2016.
- M. Maathuis, M. Drton, S. Lauritzen, and M. Wainwright, editors. *Handbook of graphical models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL, 2019.
- J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1999.
- S. W. Mogensen, D. Malinsky, and N. R. Hansen. Causal learning for partially observed stochastic dynamical systems. In *Proceedings of the 34th conference on Uncertainty in Artificial Intelligence (UAI)*, pages 350–360. PMLR, 2018.
- J. M. Mooij, D. Janzing, and B. Schölkopf. From ordinary differential equations to structural causal models: The deterministic case. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 440–448. PMLR, 2013.
- B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234, 1995.
- Y. Nesterov. Gradient methods for minimizing composite functions. *Math. Program.*, 140(1):125–161, 2013.



- V. Noferini and F. Poloni. Nearest  $\Omega$ -stable matrix via Riemannian optimization. *Numer. Math.*, 148(4):817–851, 2021.
- C. Nowzohour, M. H. Maathuis, R. J. Evans, and P. Bühlmann. Distributional equivalence and structure learning for bow-free acyclic path diagrams. *Electron. J. Stat.*, 11(2):5342–5374, 2017.
- J. Pearl. *Causality*. Cambridge University Press, Cambridge, second edition, 2009.
- J. Peters and P. Bühlmann. Identifiability of Gaussian structural equation models with equal error variances. *Biometrika*, 101(1):219–228, 2014.
- J. Peters, D. Janzing, and B. Schölkopf. *Elements of causal inference*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2017.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://www.R-project.org/>.
- J. Ramsey and B. Andrews. Fask with interventional knowledge recovers edges from the sachs model. *arXiv*, abs/1805.03108, 2018.
- P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing  $\ell_1$ -penalized log-determinant divergence. *Electron. J. Stat.*, 5:935–980, 2011.
- T. Richardson. A discovery algorithm for directed cyclic graphs. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 454–461. Morgan Kaufmann, San Francisco, CA, 1996.
- K. Sachs, O. Perez, D. Pe’er, D. A. Lauffenburger, and G. P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.
- G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.
- P. Spirtes, C. Glymour, and R. Scheines. *Causation, prediction, and search*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, second edition, 2000.
- S. Sullivant. *Algebraic statistics*, volume 194 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.
- N. M. Szekeres. Graphical continuous lyapunov models with unknown volatility. Master’s thesis, Technical University of Munich, Sep 2023. co-supervised by P.Dettling.
- R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.
- G. Varando. gclm. <https://github.com/cran/gclm>, 2020.

## BIBLIOGRAPHY

- G. Varando and N. R. Hansen. Graphical continuous Lyapunov models. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, pages 989–998, 2020.
- M. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using  $\ell_1$ -constrained quadratic programming (lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202, 2009.
- M. J. Wainwright. *High-dimensional statistics*, volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019.
- Wolfram Research, Inc. Mathematica version 13.2, 2022. Champaign, IL, 2022.
- J. Zhu, C. Wen, J. Zhu, H. Zhang, and X. Wang. A polynomial algorithm for best-subset selection problem. *Proceedings of the National Academy of Sciences*, 117(52):33117–33123, 12 2020.