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

Dimension of Sparse Factor Analysis Models

Carolina Kornitzer

Thesis for the attainment of the academic degree

Bachelor of Science

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

Supervisor:
Prof. Mathias Drton, Ph.D.

Advisor:
Dr. Nils Sturma

Submitted:
Munich, 15.01.2025

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

Munich, 15.01.2025

Kornitzer C.

Abstract

Factor analysis is a statistical tool for modeling observed variables and their correlations in terms of underlying independent, unobserved factors. This thesis focuses on sparse factor analysis models, which are characterized by factors that do not necessarily influence all observed variables. Recent results have shown that for such models, the dimension can always be upper-bounded. When a certain level of sparsity is present, a lower bound can also be determined. Clearly, in cases where both bounds coincide, a dimension formula is obtained. In this thesis, we seek to develop software in R to determine these bounds computationally, as this may already suffice to get the model's dimension. As a practical application, we perform simulations to identify the smallest graph in terms of the number of nodes and edges for which the bounds vary, as well as to identify the graphs that have expected dimension despite differing bounds. For the latter case of graphs, we will make use of the computer algebra system MACAULAY2 to determine the dimension, as differing bounds do not provide definitive information. Moreover, we extend our dimension analysis to a particular case of sparse factor analysis models with dependent, unobserved factors.

Zusammenfassung

Faktorenanalyse ist ein statistisches Verfahren zur Modellierung beobachteter Variablen und ihrer Korrelationen in Bezug auf zugrunde liegende unabhängige, unbeobachtete Faktoren. Diese Arbeit befasst sich mit dünnbesetzten Faktorenanalyse-Modellen, die dadurch charakterisiert sind, dass die Faktoren nicht alle beobachteten Variablen beeinflussen müssen. Jüngste Ergebnisse haben gezeigt, dass für solche Modelle die Dimension immer nach oben beschränkt werden kann. Wenn ein bestimmter Grad an Dünnbesetzung vorhanden ist, kann auch eine untere Schranke bestimmt werden. In Fällen, in denen beide Schranken übereinstimmen, erhält man eine Dimensionsformel. In dieser Arbeit werden wir eine Software in R entwickeln, um die Schranken computerbasiert zu berechnen. Oftmals genügt dies, um die Dimension des Modells zu erhalten. Als praktische Anwendung führen wir Simulationen durch, um den kleinsten Graphen in Bezug auf die Anzahl an Knoten und Kanten zu identifizieren, für den die Schranken nicht übereinstimmen, und um diejenigen Graphen zu identifizieren, die trotz abweichender Schranken die erwartete Dimension haben. Für diesen letzten Fall nutzen wir das Computeralgebrasystem MACAULAY2, um die Dimension zu bestimmen, da abweichende Schranken keine endgültige Information liefern. Darüber hinaus erweitern wir unsere Analyse auf einen konkreten Fall von dünnbesetzten Modellen mit abhängigen Faktoren.

Contents

1 Introduction

Factor analysis is a powerful statistical technique where unobserved random variables, termed *factors*, affect linearly the behavior of observed random variables along with some independent noise. It has its origins in the early 1900s when Spearman conceptualized the idea now known as factor analysis to study intelligence structures (Spearman, 1904). Since then, factor analysis has contributed valuable insights for analyzing multivariate data in a variety of fields, including psychology and medicine (Iwanaga et al., 2020; Petrinovich and Hardyck, 1964), economics (Jurczak and Jurczak, 2021) and sociology (Petersen et al., 1964). In practice, the number of factors is significantly smaller than the number of observed variables, which explains the common characterization of factor analysis as a dimension-reduction tool (Drton et al., 2007). In particular, let $X \in \mathbb{R}^p$ be the vector of observed variables X_i, and let $Y \in \mathbb{R}^m$ represent the vector of factors Y_i. Then factor analysis relies on the assumption that X can be defined as the solution of equations

$$X = \Lambda Y + \varepsilon,$$

where $\Lambda \in \mathbb{R}^{p \times m}$ is the unknown *factor loading matrix* and $\varepsilon \in \mathbb{R}^p$ denotes the vector of noise terms ε_i. The factors Y_i are usually not only assumed to be independent of the noise ε but also mutually independent. It is furthermore assumed that the factors Y_i are scaled to have mean zero and variance one and that $\mathbb{E}[\varepsilon_i] = 0$ and $\mathsf{Var}[\varepsilon_i] =: \omega_{ii} \in (0, \infty)$.
Our principal focus in this work will be the covariance matrix of X, which, given our previously made assumptions, is given by

$$\Sigma := \mathsf{Cov}[X] = \Lambda\Lambda^\top + \Omega,$$

where Ω is a diagonal matrix with entries $\omega_{ii} = \mathsf{Cov}[\varepsilon_i]$. More precisely, we will identify each factor analysis model with the set of its covariance matrices and aim to determine this set's dimension. Significant differences in terms of dimension properties emerge when distinguishing between *full* factor analysis models, where all entries of Λ are nonzero, and *sparse* factor analysis models, where the coefficients λ_{vh} of Λ are allowed to be zero. While the dimension of traditional full factor models can be easily obtained by counting parameters, the dimension of sparse factor models can potentially drop, making the process of dimension determination less immediate. However, a recent advance has shown that for any model an upper bound on the dimension can be evaluated. For models with a certain level of sparsity, we can additionally derive a lower bound. In cases where both bounds coincide, a dimension formula is obtained (Drton et al., 2024). This approach significantly accelerates the process of manually determining the dimension in many cases and motivates the central task of this thesis: Implementing efficient algorithms in R for computational bound evaluation, with the final goal of computationally deriving the model's dimension when possible. For cases where the bounds are insufficient to obtain the dimension, we demonstrate how to compute the dimension directly using the computer algebra system MACAULAY2.
The thesis is structured into four major parts. Chapter 2 covers basic algebraic background and notions of factor analysis, necessary for a proper understanding of the topic. We then summarize in Chapter 3 the bounds on the dimension of sparse factor analysis models that were obtained in Drton et al. (2024), illustrating the results with an example. In Chapter 4 we examine thoroughly the computational aspects, which includes presenting algorithms in R to evaluate the bounds on the dimension of models, comparing their complexity, running simulations and analyzing their results. Finally, in Chapter 5 we drop the previously made assumption of independent factors, leading to a brief study of the dimension of a particular sparse factor analysis graph with dependent factors.

2 Preliminaries

2.1 Algebraic background

We will begin by introducing some common algebraic results which will provide the necessary foundation for a proper understanding of the subject. The first proposition establishes a connection between the dimension of a map's image and the rank of the corresponding Jacobian matrix. This is Proposition 16.1.7 from Sullivant (2018), specifically tailored to the case where the function maps into the space $\mathbb{R}^n$.

Proposition 2.1. *Let $\Theta \subseteq \mathbb{R}^d$ with $\dim \Theta = d$ and suppose that τ is a rational map $\tau : \Theta \longrightarrow \mathbb{R}^n$. Then $\dim \mathrm{Im}(\tau)$ is equal to the rank of the Jacobian matrix evaluated at a generic point:*

$$J(\tau) \;=\; \begin{pmatrix} \frac{\partial \tau_1}{\partial \theta_1} & \cdots & \frac{\partial \tau_1}{\partial \theta_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial \tau_n}{\partial \theta_1} & \cdots & \frac{\partial \tau_n}{\partial \theta_d} \end{pmatrix}. \tag{1}$$

Remark 2.2. Let $s = \dim \mathrm{Im}(\tau)$ be the generic rank of $J(\tau)$ and denote the set of all $s \times s$ minors of $J(\tau)$ by $\mathcal{G} = \{g_1, ..., g_k\} \subset \mathbb{R}[\theta_1, ..., \theta_d]$ with $k = \binom{d}{s}\binom{n}{s}$. Then every $l \times l$ minor of $J(\tau)$ with $l > s$ vanishes, while at least one $s \times s$ minor must be nonzero (Bapat, 2012, Section 2.3). Furthermore, let us denote by $\mathcal{N} = \{\theta \in \Theta : \operatorname{rank} J(\tau(\theta)) < s\}$ the set of all critical points of τ. By definition, for all $\theta \in \mathcal{N}$, every $s \times s$ minor must vanish. Consequently, we can express the set of critical points as the intersection of the vanishing sets of $g_i \in \mathcal{G}$:

$$\mathcal{N} = \bigcap_{i \in [k]} V(g_i) = \bigcap_{i \in [k]} \{\theta \in \Theta : g_i(\theta) = 0\}.$$

As stated in the Lemma of Okamoto (1973, Section 2), each vanishing set $V(g_i)$ has Lebesgue measure zero, which implies that $\mathcal{N}$ is also of measure zero. We conclude that $\mathrm{rank}(J(\tau(\theta))) = s$ for almost all $\theta \in \Theta$.

For a given polynomial map $\tau : \mathbb{R}^d \longrightarrow \mathbb{R}^n$, we can define the following ring map

$$\begin{aligned} \psi_\tau : \mathbb{R}[x_1, ..., x_n] &\longrightarrow \mathbb{R}[y_1, ..., y_d] \\ x_i &\longmapsto \tau_i(y_1, ..., y_d), \end{aligned} \tag{2}$$

referred to as the *dual map* of τ throughout the thesis. It can indeed be interpreted as a sort of dual to τ, in the sense that it maps the coordinates of the codimension of τ to the corresponding polynomial expression. The next proposition states an important property about the kernel of ψ_τ. We denote the vanishing ideal by $\mathcal{I}(\cdot)$.

Proposition 2.3. *Let $\tau : \mathbb{R}^d \longrightarrow \mathbb{R}^n$ be a polynomial map and let ψ_τ be the corresponding dual map as defined in (2). It holds that*

$$\mathrm{kernel}(\psi_\tau) = \mathcal{I}(\mathrm{Im}(\tau)). \tag{3}$$

Proof. We show inclusion in both directions.
"$\subseteq$" Let $g \in \mathrm{kernel}(\psi_\tau)$, then

$$0 = \psi_\tau(g(x_1, ..., x_n)) = g(\tau_1(y_1, ..., y_d), ..., \tau_n(y_1, ..., y_d)).$$

In particular, this holds for all $y \in \mathbb{R}^d$. Since $\{\tau(y) | y \in \mathbb{R}^d\} = \mathrm{Im}(\tau)$, we get that g vanishes for all elements of $\mathrm{Im}(\tau)$, which implies $g \in \mathcal{I}(\mathrm{Im}(\tau))$.

"$\supseteq$" Let $g \in \mathcal{I}(\mathrm{Im}(\tau))$. In order to prove that $g \in \mathrm{kernel}(\psi_\tau)$, we must show $0 = \psi_\tau(g(x_1, ..., x_n))$ for all $x \in \mathbb{R}^n$. Fixing any $x \in \mathbb{R}^n$ yields

$$\psi_\tau(g(x_1, ..., x_n)) = g(\tau_1(y_1, ..., y_d), ..., \tau_n(y_1, ..., y_d)) = 0,$$

with the last equality following from $g \in \mathcal{I}(\mathrm{Im}(\tau))$. This completes the proof. □

We will now review the study of semi-algebraic sets, which are sets defined by a combination of polynomial equations and inequalities. The results are based on Chapter 2 of Bochnak et al. (1998), unless where otherwise stated. They will become particularly relevant in Section 4.3, where we give an approach to algebraically compute the dimension of a factor analysis model.

Definition 2.4. *A semi-algebraic subset of $\mathbb{R}^n$ is a finite union of sets of the form*

$$\{x \in \mathbb{R}^n | g_1(x) = \ldots = g_l(x) = 0, h_1(x) > 0, \ldots, h_m(x) > 0\},$$

where $g_1(x), \ldots, g_l(x), h_1(x), \ldots, h_m(x)$ are in $\mathbb{R}[X_1, \ldots, X_n]$.

Having established a formal definition of semi-algebraic sets, we can now proceed to state one of their main properties: stability under projections. This valuable characteristic, known as *Tarski-Seidenberg Theorem*, guarantees that mapping a semi-algebraic set to a lower dimensional space preserves its semi-algebraic structure.

Theorem 2.5. *(Tarski-Seidenberg Theorem) Let A denote a semi-algebraic subset of the space $\mathbb{R}^{n+1}$ and let $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ be the projection on the space of the first n coordinates. Then $\pi(A)$ is a semi-algebraic subset of $\mathbb{R}^n$.*

Proof. A proof of this theorem can be found in Bochnak et al. (1998, Theorem 2.2.1). □

We can now extend beyond the general projection property of semi-algebraic sets, leading to a corollary that particularly highlights the behavior of a semi-algebraic set's image under polynomial mappings.

Corollary 2.6. *(Coste, 2002, Corollary 2.4)*

(1) If A is a semi-algebraic subset of $\mathbb{R}^{n+k}$, its image by the projection on the space of the first n coordinates is a semi-algebraic subset of $\mathbb{R}^n$.

(2) If A is a semi-algebraic subset of $\mathbb{R}^m$ and $F : \mathbb{R}^m \to \mathbb{R}^n$ a polynomial mapping, then the direct image $F(A)$ is a semi-algebraic subset of $\mathbb{R}^n$.

Proof. The first statement is trivially obtained by induction on k. For the second statement, write F as the composition of the graph function and a projection

$$F : \mathbb{R}^m \xrightarrow{\Gamma_F} \mathbb{R}^m \times \mathbb{R}^n \xrightarrow{\pi} \mathbb{R}^n.$$

We note that $\Gamma_F(A) = \{(x, y) \in \mathbb{R}^m \times \mathbb{R}^n \mid x \in A \text{ and } y = F(x)\}$ is a semi-algebraic subset of $\mathbb{R}^m \times \mathbb{R}^n$ and that $F(A)$ is its projection onto $\mathbb{R}^n$. The result follows by applying the first statement. □

To further study semi-algebraic sets, we will now focus on their dimension, which is defined below.

Definition 2.7. *Let $A \subseteq \mathbb{R}^n$ be a semi-algebraic set. Denote by $\mathbb{R}[X_1, \ldots, X_n]/\mathcal{I}(A)$ the ring of polynomial functions on A. The dimension of A is defined by*

$$\dim(A) = \dim(\mathbb{R}[X_1, \ldots, X_n]/\mathcal{I}(A)). \tag{4}$$

We close this section by considering the *Zariski closure* of a semi-algebraic set $A \subseteq \mathbb{R}^n$, which is the smallest algebraic subset of $\mathbb{R}^n$ containing A (Coste, 2002, Section 3.3.2). The vanishing ideal of a Zariski closure always coincides with the vanishing ideal of the original set. Note that this result is widely considered self-evident, and no explicit proof was found in the literature. For completeness, we provide a proof here.

Proposition 2.8. *Let $A \subseteq \mathbb{R}^n$ be a semi-algebraic set. Then*

$$\mathcal{I}(A) = \mathcal{I}(\mathrm{clos}_{\mathrm{zar}}(A)), \tag{5}$$

where $\mathrm{clos}_{\mathrm{zar}}(A)$ is the Zariski closure of A.

Proof. We show inclusion in both directions.
"$\supseteq$" Since $A \subseteq \mathrm{clos}_{\mathrm{zar}}(A)$, we get that $\mathcal{I}(\mathrm{clos}_{\mathrm{zar}}(A)) \subseteq \mathcal{I}(A)$.
"$\subseteq$" Suppose $g \in \mathcal{I}(A)$. Then g vanishes on a set $Q \subseteq \mathbb{R}^n$ which contains A. By definition, Q must be Zariski-closed and $\mathrm{clos}_{\mathrm{zar}}(A) \subseteq Q$. This implies that g vanishes on the set $\mathrm{clos}_{\mathrm{zar}}(A)$, and therefore $g \in \mathcal{I}(\mathrm{clos}_{\mathrm{zar}}(A))$ as needed. □

The next proposition follows immediately.

Proposition 2.9. *Let $A \subseteq \mathbb{R}^n$ be a semi-algebraic set. Then*

$$\dim(A) = \dim(\mathrm{clos}_{\mathrm{zar}}(A)). \tag{6}$$

Proof. Since $\mathcal{I}(A) = \mathcal{I}(\mathrm{clos}_{\mathrm{zar}}(A))$, we have

$$\dim(A) \overset{(4)}{=} \dim(\mathbb{R}[X_1, \ldots, X_n]/\mathcal{I}(A)) \overset{(5)}{=} \dim(\mathbb{R}[X_1, \ldots, X_n]/\mathcal{I}(\mathrm{clos}_{\mathrm{zar}}(A))) \overset{(4)}{=} \dim(\mathrm{clos}_{\mathrm{zar}}(A)).$$

□

2.2 Factor analysis models

The results in this section rely on the work of Drton et al. (2024). We begin by introducing some necessary terminology and notation related to directed graphs, which will be used in this thesis.
A *directed Graph G* can be described as a tuple $(\mathcal{H} \cup V, D)$, where V and $\mathcal{H}$ represent finite and disjoint sets of observed and latent nodes, respectively. For the entirety of this thesis, we will fix $|V| = p$ and $|\mathcal{H}| = m$. The set $D \subseteq \mathcal{H} \times V$ refers to the edges. We write $h \to v \in D$, whenever $(h, v) \in D$ and say that h is a parent of its child v. Note that in all chapters of this thesis except for Chapter 5, we will exclusively consider bipartite graphs where edges point from latent to observed nodes. Furthermore, we will denote by $\mathrm{pa}(v) = \{h \in \mathcal{H} : h \to v \in D\}$ the set of all *parents* of a node $v \in V$ and by $\mathrm{ch}(h) = \{v \in V : h \to v \in D\}$ the set of all *children* of a node $h \in \mathcal{H}$. For the set that contains all pairs consisting of 2 distinct nodes of V, we write $C(V, 2) := \{\{v, w\} : v, w \in V, v \neq w\}$. Then the set of *joint parents* of a pair $\{u, v\} \in C(V, 2)$ can be declared as $\mathrm{jpa}(\{u, v\}) = \{h \in \mathcal{H} : h \in \mathrm{pa}(u) \cap \mathrm{pa}(v)\}$. Finally, let $h \in \mathcal{H}$ be a latent node, then $C(V, 2)_h = \{\{v, w\} \in C(V, 2) : h \in \mathrm{jpa}(\{v, w\})\}$ refers to the set of all 2-pairs of nodes that share h as a joint parent.
We now formalize that each factor analysis graph encodes a factor analysis model that can be identified with the set of its covariance matrices. We write $\mathbb{R}^D$ for the set of real $|V| \times |\mathcal{H}|$ matrices $\Lambda = (\lambda_{vh})$ with support D, where $\lambda_{vh} = 0$ if $h \to v \notin D$. Additionally, we denote the subset of diagonal positive definite matrices by $\mathbb{R}^p_{>0} \subset \mathrm{PD}(p)$, where $\mathrm{PD}(p)$ is the set of all positive definite matrices in $\mathbb{R}^{p \times p}$.

Definition 2.10. *Let $G = (V \cup \mathcal{H}, D)$ be a factor analysis graph with $|V| = p$ and $|\mathcal{H}| = m$. As a model of the covariance matrix, the factor analysis model determined by G is the image $F(G) = \mathrm{Im}(\tau_G)$ of the parametrization map*

$$\begin{aligned} \tau_G : \mathbb{R}^p_{>0} \times \mathbb{R}^D &\longrightarrow \mathrm{PD}(p) \\ (\Omega, \Lambda) &\longmapsto \Omega + \Lambda\Lambda^\top. \end{aligned} \tag{7}$$

Our main subject of interest is the dimension of the covariance model $F(G)$ of a sparse factor analysis graph G. Determining the dimension of a full factor analysis model involves simply counting parameters, since they are always of *expected dimension*, $\min\{|V| + |D|, \binom{|V|+1}{2}\}$ (Drton et al., 2007, Theorem 2). However, this intuitive approach is usually not applicable for sparse factor analysis graphs, as their dimension may drop. The following example serves well to introduce the sparse factor analysis models that we will be working with.

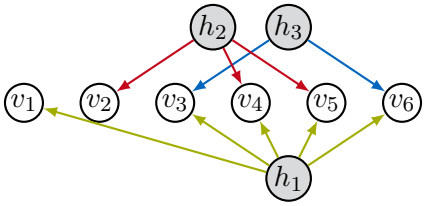

Figure 1 Sparse factor analysis graph with lower model dimension than number of parameters.

Example 2.11. Consider the graph in Figure 1. Clearly, its expected dimension equals $|V| + |D| = 16$. The covariance model $F(G)$ is a subset of the space of symmetric 6×6 matrices which has dimension $\binom{7}{2} = 21$. Performing matrix multiplication, we can easily convince ourselves that in every covariance matrix there are three zero entries, $\sigma_{v_1v_2} = \sigma_{v_2v_3} = \sigma_{v_2v_6} = 0$. This yields a trivial upper bound of 18. We will see in Example 3.4 that the model's actual dimension is 15. Hence, we observe that in the case of this particular sparse factor analysis model, dimension drops by one.

To determine the actual dimension of a sparse factor analysis model, we will, rather than just counting parameters, make use of the connection between the dimension of $F(G)$ and the Jacobian matrix of τ_G. This was established in Proposition 2.1. Thus, let us consider the dual map of τ_G which expresses the entries of the covariance matrix $\Sigma = \Omega + \Lambda\Lambda^\top$ with $\Omega = \mathrm{diag}(\omega_{vv}) \in \mathbb{R}^p_{>0}$ and $\Lambda = (\lambda_{vh}) \in \mathbb{R}^D$, and which is given by

$$\begin{aligned} \psi_{\tau_G} : \mathbb{R}[\sigma_{uv} | u, v \in V] &\longrightarrow \mathbb{R}[\omega_{vv}, \lambda_{vh} | v \in V, h \in \mathcal{H}] \\ (\sigma_{uv}) &\longmapsto \begin{cases} \sum_{h \in \mathrm{jpa}(\{u,v\})} \lambda_{uh}\lambda_{vh} & \text{if } u \neq v, \\ \omega_{uu} + \sum_{h \in \mathrm{pa}(u)} \lambda_{uh}^2 & \text{if } u = v. \end{cases} \end{aligned} \tag{8}$$

We can now easily study the Jacobian matrix of τ_G that has the following structure

$$J = \begin{array}{c} \\ u \\ \{u,v\} \end{array} \begin{array}{c} \begin{array}{cc} \omega & \lambda \end{array} \\ \begin{pmatrix} I_p & C \\ 0 & B \end{pmatrix} \end{array},$$

where the first rows indexed by u indicate the derivatives of σ_{uu} and the lower rows indexed by $\{u, v\}$ with $u \neq v$ refer to the derivatives of σ_{uv}. The derivatives of the entries σ_{uu} with respect to ω are given by

$$\frac{\partial \sigma_{uu}}{\partial \omega_{vv}} = \begin{cases} 1 & \text{if } u = v, \\ 0 & \text{else,} \end{cases}$$

which explains the unit matrix I_p in the upper left block of J. The entries of the matrix B in the lower right block are given by

$$\frac{\partial \sigma_{uv}}{\partial \lambda_{zh}} = \begin{cases} \lambda_{vh} & \text{if } z = u \text{ and } h \in jpa(\{u,v\}), \\ \lambda_{uh} & \text{if } z = v \text{ and } h \in jpa(\{u,v\}), \\ 0 & \text{else.} \end{cases} \tag{9}$$

Recalling Proposition 2.1, the dimension of the covariance model $F(G)$ coincides with the rank of the Jacobian matrix and thus

$$\dim F(G) = p + \mathrm{rank}(B). \tag{10}$$

For our purpose of formalizing bounds on the dimension of sparse factor analysis models, we will introduce the idea of *valid* collections which are characterized by pairwise disjoint components whose cardinalities do not exceed the number of children of the corresponding factor.

Definition 2.12. *Let $G = (V \cup \mathcal{H})$ be a factor analysis graph and let $\mathcal{A} = (A_h)_{h\in\mathcal{H}}$ be a collection of observed 2-pairs, that is, $A_h \subseteq C(V, 2)$. We say that the collection $\mathcal{A}$ is* valid *if*

(i) $A_h \subseteq C(V, 2)_h$ with cardinality $|A_h| \leq |\operatorname{ch}(h)|$ for all $h \in \mathcal{H}$, and

(ii) the collection is pairwise disjoint, i.e., $A_h \cap A_\ell = \emptyset$ for $h \neq \ell$.

Moreover, we say that $\sum_{h\in\mathcal{H}} |A_h|$ is the sum of cardinalities *of a valid collection.*

The next definition allows us to identify a particular class of factor analysis graphs that possess a minimal level of sparsity.

Definition 2.13. *A factor analysis graph and its associated model satisfy the* Zero Upper Triangular Assumption (ZUTA) *if there exists a relabeling of the latent nodes $\mathcal{H} = \{h_1, \ldots, h_m\}$ such that $\operatorname{ch}(h_i)$ is not contained in $\bigcup_{j>i} \operatorname{ch}(h_j)$ for all $i = 1, \ldots, m$. In this case, there is then a relabeling of the observed nodes $V = \{v_1, \ldots, v_p\}$ such that $v_i \in \operatorname{ch}(h_i)$ and $v_i \notin \bigcup_{j>i} \operatorname{ch}(h_j)$ for all $i = 1, \ldots, m$.*

Example 2.14. Consider the graph in Figure 1. Both the latent and observed nodes are already ordered in a way that satisfies ZUTA. Note that several other relabelings of latent and observed nodes are also possible for ZUTA to be satisfied, such as interchanging the roles of h_2 and h_3 as well as v_2 and v_3.

The ZUTA condition guarantees that the factor loading matrix Λ can be rearranged, by permuting its columns and rows, such that all entries above the main diagonal are zero, while all diagonal entries are nonzero. Furthermore, ZUTA ensures the existence of a more refined version of valid collections, referred to as *ZUTA-compliant* collections.

Definition 2.15. *Suppose that ZUTA is satisfied. A valid collection $\mathcal{A} = (A_h)_{h\in\mathcal{H}}$ of 2-pairs is* ZUTA-compliant *if $\{v_i, w\} \in A_{h_i}$ for all $w \in \operatorname{ch}(h_i) \setminus \{v_i\}$ and for all $i \in [m]$.*

Remark 2.16. Consider the sets $C_i := \{\{v_i, w\} : w \in \operatorname{ch}(h_i) \setminus \{v_i\}\}$ for $i \in [m]$. If the ZUTA assumption holds, existence of a ZUTA-compliant collection follows directly, given that the collection $\mathcal{A} = (A_h)_{h\in\mathcal{H}}$, where $A_{h_i} = C_i$, satisfies all conditions of Definition 2.15. In fact, this collection contains the minimum amount of 2-pairs that are necessary to form a ZUTA-compliant collection, with each component A_{h_i} having cardinality $|\operatorname{ch}(h_i)| - 1$. For every further ZUTA-compliant collection $\mathcal{A} = (A_h)_{h\in\mathcal{H}}$, it must hold that $C_i \subseteq A_{h_i}$ for all $i \in [m]$. Recalling Definition 2.12, we get that the cardinality of A_{h_i} can be at most equal to $|\operatorname{ch}(h_i)|$. This leads to the conclusion that there are 2 possibilities for the components of a ZUTA-compliant collection, namely

$$A_{h_i} = \begin{cases} C_i \cup \{S_i\} \text{ or} \\ C_i, \end{cases}$$

where $S_i := \{u_i, w_i\} \in C(V, 2)_h \setminus C_i$ such that an empty intersection between the components is maintained. We define the index sets $I^{(=)} = \{i \in [m] : |A_{h_i}| = |\operatorname{ch}(h_i)|\}$ and $I^{(<)} = [m] \setminus I^{(=)}$. Clearly, if $A_{h_i} = C_i \cup \{S_i\}$ then $i \in I^{(=)}$, but if $A_{h_i} = C_i$ it holds that $i \in I^{(<)}$.

3 Bounds on the dimension

Having established the previous results, we are now able to properly study the dimension of sparse factor analysis models. Therefore, we summarize the approaches for evaluating the upper and lower bound on a model's dimension as described in Drton et al. (2024), and give a brief sketch of the corresponding proofs. We will then demonstrate both proof strategies with an illustrative example. It is important to note that if the upper bound coincides with the lower bound, we immediately get a result for the model's dimension. Otherwise, the bounds do not suffice to provide a precise value of the dimension. This issue is tackled in Section 4.3.

3.1 Upper bound on the dimension

We begin by giving an upper bound on the dimension of sparse factor analysis models in the next theorem. It relies on finding a valid collection $\mathcal{A} = (A_h)_{h\in\mathcal{H}}$ with maximal sum of cardinalities $\sum_{h\in\mathcal{H}} |A_h|$ and can be applied to any sparse factor model, regardless of whether ZUTA is satisfied.

Theorem 3.1. *Let $G = (V \cup \mathcal{H}, D)$ be a factor analysis graph. Let $\mathcal{A} = (A_h)_{h\in\mathcal{H}}$ be a valid collection of 2-pairs such that the sum of cardinalities $\sum_{h\in\mathcal{H}} |A_h|$ is maximal among all valid collections. Then*

$$\dim(F(G)) \leq |V| + \sum_{h\in\mathcal{H}} |A_h|.$$

Outline of the Proof: We refer to Drton et al. (2024, Theorem 2.9) for a detailed version of the proof, and provide a summary of the key steps in the following. Recalling Equation (10), we aim to show that $\operatorname{rank}(B) \leq \sum_{h\in\mathcal{H}} |A_h|$, as this directly implies the statement. Let $\lambda_i := (\lambda_{\mathrm{ch}(h_i),h_i}) \in \mathbb{R}^{|\mathrm{ch}(h_i)|}$ and $\mathcal{A}^{\complement} := C(V,2) \setminus (\bigcup_{h\in\mathcal{H}} A_h)$. Furthermore, we recall the index sets $I^{(=)}$ and $I^{(<)}$ that were defined in Remark 2.16. For the purpose of this proof, we arrange columns and rows of the matrix B as

$$B = \begin{array}{c} \\ A_{h_1} \\ \vdots \\ A_{h_m} \\ \mathcal{A}^{\complement} \end{array} \begin{array}{c} \begin{array}{ccc} \lambda_1 & \cdots & \lambda_m \end{array} \\ \left(\begin{array}{ccc} B_{1,1} & \cdots & B_{1,m} \\ \vdots & & \vdots \\ B_{m,1} & \cdots & B_{m,m} \\ \hline B_{\mathcal{A}^{\complement},1} & \cdots & B_{\mathcal{A}^{\complement},m} \end{array}\right) \end{array}. \tag{11}$$

The proof strategy is as follows. First, we verify that the submatrix of B, consisting of the rows indexed by the pairs $R \in \mathcal{A}^{\complement}$ and the columns indexed by λ_i, where $i \in I^{(<)}$, is zero. Define the set

$$J^0 = \{i \in I^{(=)} : B_{\mathcal{A}^{\complement},i} \neq 0\} = \{i \in I^{(=)} : h_i \in \mathrm{jpa}(R) \text{ for some } R \in \mathcal{A}^{\complement}\},$$

and, for all $k \geq 1$, define

$$J^k = \{j \in I^{(=)} : \text{there is } i \in J^{k-1} \text{ such that } h_j \in \mathrm{jpa}(R) \text{ for some } R \in A_{h_i}\}.$$

Since $h_j \in \mathrm{jpa}(R)$ for all $R \in A_{h_j}$, it must hold that $J^k \subseteq J^{k+1}$ for all $k \geq 0$. Additionally, there must be a $k^* \geq 0$ such that the sequence $J^0 \subseteq J^1 \subseteq \ldots$ stabilizes with $J^{k^*} = J^{k^*+1} = \ldots$. Define $\overline{J} := J^{k^*}$. Assuming that the previous statements hold true, we proceed to show that the matrix blocks $B_{j,i}$ with $j \in \overline{J}$ and $i \in [m] \setminus \overline{J}$ are also equal to zero. Finally, we conclude that after restructuring the matrix B, its rank cannot exceed $\sum_{h\in\mathcal{H}} |A_h|$.

Example 3.2. Consider the graph in Figure 1. A maximal, valid collection is given by

$$
\begin{aligned}
A_{h_1} &= \{\{v_1, v_3\}, \{v_1, v_4\}, \{v_1, v_5\}, \{v_1, v_6\}, \{v_3, v_4\}\},\\
A_{h_2} &= \{\{v_2, v_4\}, \{v_2, v_5\}, \{v_4, v_5\}\},\\
A_{h_3} &= \{\{v_3, v_6\}\}.
\end{aligned}
$$

Clearly, $I^{(=)} = \{1, 2\}$ and $I^{(<)} = \{3\}$. The only relevant 2-pairs in $\mathcal{A}^{\complement}$ that index a nonzero row in B are $\{\{v_3, v_5\}, \{v_4, v_6\}, \{v_5, v_6\}\} \subset \mathcal{A}^{\complement}$. In the following discussion of this example, $\mathcal{A}^{\complement}$ will refer exclusively to this set of three 2-pairs.

Claim 1: $B_{\mathcal{A}^{\complement},i} = 0$ *for all* $i \in I^{(<)}$.

Recalling Equation (9) and Equation (11), B has the following structure

$$
B = \begin{array}{c} \\ \{v_1,v_3\} \\ \{v_1,v_4\} \\ \{v_1,v_5\} \\ \{v_1,v_6\} \\ \{v_3,v_4\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_4,v_5\} \\ \{v_3,v_6\} \\ \{v_3,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \end{array}
\begin{array}{c}
\begin{array}{ccccccccccc} \lambda_{v_1,h_1} & \lambda_{v_3,h_1} & \lambda_{v_4,h_1} & \lambda_{v_5,h_1} & \lambda_{v_6,h_1} & \lambda_{v_2,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_3,h_3} & \lambda_{v_6,h_3} \end{array} \\
\left(\begin{array}{ccccc|ccccc}
\lambda_{v_3,h_1} & \lambda_{v_1,h_1} & & & & & & & & \\
\lambda_{v_4,h_1} & & \lambda_{v_1,h_1} & & & & & & & \\
\lambda_{v_5,h_1} & & & \lambda_{v_1,h_1} & & & & & & \\
\lambda_{v_1,h_1} & & & & \lambda_{v_6,h_1} & & & & & \\
\hdashline
 & \lambda_{v_4,h_1} & \lambda_{v_3,h_1} & & & & & & & \\
\hdashline
 & & & & & \lambda_{v_4,h_2} & \lambda_{v_2,h_2} & & & \\
 & & & & & \lambda_{v_5,h_2} & & \lambda_{v_2,h_2} & & \\
\hdashline
 & & \lambda_{v_5,h_1} & \lambda_{v_4,h_1} & & & \lambda_{v_5,h_2} & \lambda_{v_4,h_2} & & \\
\hdashline
 & \lambda_{v_6,h_1} & & & \lambda_{v_3,h_1} & & & & \lambda_{v_6,h_3} & \lambda_{v_3,h_3} \\
\hline
 & \lambda_{v_5,h_1} & & \lambda_{v_3,h_1} & & & & & & \\
 & & \lambda_{v_6,h_1} & & \lambda_{v_4,h_1} & & & & & \\
 & & & \lambda_{v_6,h_1} & \lambda_{v_5,h_1} & & & & &
\end{array}\right)
\end{array},
$$

where void entries are zero. We see that the lower right block $B_{\mathcal{A}^{\complement},3}$ is zero as required. Hence, Claim 1 is already satisfied.

Claim 2: $B_{j,i} = 0$ *for all* $j \in \overline{J}, i \in [m] \setminus \overline{J}$.

The sequence $(J_k)_{k\in\mathbb{N}}$ is given by $J^0 = J^k = \{1\} \quad \forall k \in \mathbb{N}$. Consequently, the sequence stabilizes already for $k^* = 0$. That said, according to Claim 2, it must hold that $B_{1,2} = B_{1,3} = 0$ which coincides with the results seen in the prior matrix B.

Claim 3: *The rank of the matrix B cannot exceed* $\sum_{h\in\mathcal{H}} |A_h|$.

As a last step, we rearrange our matrix appropriately by placing the rows corresponding to $\mathcal{A}^{\complement}$ in between the rows corresponding to A_{h_1} and A_{h_2}. Then the matrix B is given by

$$
B = \begin{array}{c} \\ \{v_1,v_3\} \\ \{v_1,v_4\} \\ \{v_1,v_5\} \\ \{v_1,v_6\} \\ \{v_3,v_4\} \\ \{v_3,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_4,v_5\} \\ \{v_3,v_6\} \end{array}
\begin{array}{c}
\begin{array}{ccccccccccc} \lambda_{v_1,h_1} & \lambda_{v_3,h_1} & \lambda_{v_4,h_1} & \lambda_{v_5,h_1} & \lambda_{v_6,h_1} & \lambda_{v_2,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_3,h_3} & \lambda_{v_6,h_3} \end{array} \\
\left(\begin{array}{ccccc|ccccc}
\lambda_{v_3,h_1} & \lambda_{v_1,h_1} & & & & & & & & \\
\lambda_{v_4,h_1} & & \lambda_{v_1,h_1} & & & & & & & \\
\lambda_{v_5,h_1} & & & \lambda_{v_1,h_1} & & & & & & \\
\lambda_{v_1,h_1} & & & & \lambda_{v_6,h_1} & & & & & \\
 & \lambda_{v_4,h_1} & \lambda_{v_3,h_1} & & & & & & & \\
 & \lambda_{v_5,h_1} & & \lambda_{v_3,h_1} & & & & & & \\
 & & \lambda_{v_6,h_1} & & \lambda_{v_4,h_1} & & & & & \\
 & & & \lambda_{v_6,h_1} & \lambda_{v_5,h_1} & & & & & \\
\hline
 & & & & & \lambda_{v_4,h_2} & \lambda_{v_2,h_2} & & & \\
 & & & & & \lambda_{v_5,h_2} & & \lambda_{v_2,h_2} & & \\
 & & \lambda_{v_5,h_1} & \lambda_{v_4,h_1} & & & \lambda_{v_5,h_2} & \lambda_{v_4,h_2} & & \\
 & \lambda_{v_6,h_1} & & & \lambda_{v_3,h_1} & & & & \lambda_{v_6,h_3} & \lambda_{v_3,h_3}
\end{array}\right)
\end{array}.
$$

The rank of this block matrix is limited by the sum of the minimum of the number of nonzero rows and columns of the upper left block plus the minimum of the number of nonzero rows and columns of the lower right block, i.e.

$$\operatorname{rank}(B) \leq 5 + 4 = \sum_{h \in \mathcal{H}} |A_h|.$$

Therefore, in this example the upper bound from Theorem 3.1 is 15.

3.2 Lower bound on the dimension

Unlike Theorem 3.1, which applies to any sparse factor analysis graph, the theorem for determining a lower bound on the dimension is valid only for graphs that meet the ZUTA Assumption of Definition 2.13.

Theorem 3.3. *Let $G = (V \cup \mathcal{H}, D)$ be a factor analysis graph. Suppose that ZUTA is satisfied and let $\mathcal{A} = (A_h)_{h \in \mathcal{H}}$ be a valid collection that is ZUTA-compliant. Then*

$$\dim(F(G)) \geq |V| + \sum_{h \in \mathcal{H}} |A_h|.$$

Outline of the Proof: For a detailed version of the proof we refer to Drton et al. (2024, Theorem 2.12), and provide a summary of the key steps in the following.
Let us first explain why it is enough to find a generic choice of Λ, denoted by $\Lambda^0 = (\lambda^0_{v_i,h_j}) \in \mathbb{R}^D$, for which the rank of $B^0 = B(\Lambda^0)$ is at least $r = \sum_{h \in \mathcal{H}} |A_h|$. Applying the conclusion of Remark 2.2 to the submatrix B of the Jacobian matrix, we find that the rank of B equals its generic rank $s = \dim(F(G)) - |V|$ almost everywhere . Assume we find a Λ^0 for which $\operatorname{rank}(B^0) \geq r$, then we must distinguish between two cases. Either Λ^0 is a critical point, and thus $s > \operatorname{rank}(B^0)$, which, recalling Equation (10), implies

$$\dim(F(G)) = |V| + s > |V| + \operatorname{rank}(B^0) \geq |V| + r,$$

or Λ^0 is not a critical point and hence $s = \operatorname{rank}(B^0)$, which yields

$$\dim(F(G)) = |V| + s = |V| + \operatorname{rank}(B^0) \geq |V| + r.$$

Thus, finding a generic choice of parameters Λ^0 such that $\operatorname{rank}(B(\Lambda^0)) \geq r$ indeed concludes the proof. Recall all sets defined in Remark 2.16 and let $\mathcal{S} = \{S_i : i \in I^{(=)}\}$ as well as $\mathcal{A}^{\complement} = C(V,2) \setminus (\bigcup_{h \in \mathcal{H}} A_h)$. Moreover, define $\operatorname{ch}(h_i)^- = \operatorname{ch}(h_i) \setminus \{v_i\}$ and $\lambda_i^- = \lambda_{\operatorname{ch}(h_i)^-, h_i} \in \mathbb{R}^{|\operatorname{ch}(h_i)|-1}$. For the purpose of this proof, we rearrange the matrix B as

$$B = \begin{array}{c} \\ C_1 \\ \vdots \\ C_m \\ \mathcal{S} \\ \mathcal{A}^{\complement} \end{array} \begin{array}{c} \begin{array}{cccccc} \lambda_{v_1,h_1} & \cdots & \lambda_{v_m,h_m} & \lambda_1^- & \cdots & \lambda_m^- \end{array} \\ \left(\begin{array}{ccc|ccc} \lambda_1^- & & & B_{1,1} & & \\ & \ddots & & \vdots & \ddots & \\ & & \lambda_m^- & B_{m,1} & \cdots & B_{m,m} \\ \hline & & & B_{\mathcal{S},1} & \cdots & B_{\mathcal{S},m} \\ & & & B_{\mathcal{A}^{\complement},1} & \cdots & B_{\mathcal{A}^{\complement},m} \end{array}\right) \end{array}, \tag{12}$$

where void entries are zero. We now define all $\lambda^0_{v_i,h_i}$ with $i \in [m]$ and all $\lambda^0_{u_i,h_i}, \lambda^0_{w_i,h_i}$ with $\{u_i, w_i\} \in \mathcal{S}$ to be set to one. The remaining entries of the matrix Λ^0 will be zero. We proceed by showing that through row reduction, we can eliminate all nonzero non-diagonal entries in the upper right block of B^0, while ensuring that the upper left block remains unchanged. Additionally, eliminating all nonzero entries in the rows indexed by $\mathcal{S}$ will introduce fill-ins in the corresponding left block of B^0, resulting in it having full row rank equal to $|\mathcal{S}|$. This shows that $\operatorname{rank}(B(\Lambda^0)) \geq \sum_{i \in [m]} |C_i| + |\mathcal{S}| = r$.

Example 3.4. Consider the graph in Figure 1 and the ZUTA-compliant, valid collection

$$A_{h_1} = \{\{v_1, v_3\}, \{v_1, v_4\}, \{v_1, v_5\}, \{v_1, v_6\}, \{v_3, v_4\}\},$$
$$A_{h_2} = \{\{v_2, v_4\}, \{v_2, v_5\}, \{v_4, v_5\}\},$$
$$A_{h_3} = \{\{v_3, v_6\}\}.$$

We have $C_1 = \{\{v_1, v_3\}, \{v_1, v_4\}, \{v_1, v_5\}, \{v_1, v_6\}\}$, $C_2 = \{\{v_2, v_4\}, \{v_2, v_5\}\}$, $C_3 = \{\{v_3, v_6\}\}$, as well as $I^{(=)} = \{1, 2\}$, $I^{(<)} = \{3\}$ and $S = \{S_1, S_2\}$ with $S_1 = \{v_3, v_4\}$ and $S_2 = \{v_4, v_5\}$. Furthermore, $\lambda_1^- = (\lambda_{v_3,h_1}, \lambda_{v_4,h_1}, \lambda_{v_5,h_1}, \lambda_{v_6,h_1})$, $\lambda_2^- = (\lambda_{v_4,h_2}, \lambda_{v_5,h_2})$, $\lambda_3^- = (\lambda_{v_6,h_3})$. Recalling Equation (9) and Equation (12) we conclude that the matrix B can be written as

$$B = \begin{array}{c} \\ \{v_1,v_3\} \\ \{v_1,v_4\} \\ \{v_1,v_5\} \\ \{v_1,v_6\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_3,v_6\} \\ \{v_3,v_4\} \\ \{v_4,v_5\} \\ \{v_3,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \end{array} \begin{array}{c} \begin{array}{cccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_3,h_1} & \lambda_{v_4,h_1} & \lambda_{v_5,h_1} & \lambda_{v_6,h_1} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_6,h_3} \end{array} \\ \left(\begin{array}{cccccccccc} \lambda_{v_3,h_1} & & & \lambda_{v_1,h_1} & & & & & & \\ \lambda_{v_4,h_1} & & & & \lambda_{v_1,h_1} & & & & & \\ \lambda_{v_5,h_1} & & & & & \lambda_{v_1,h_1} & & & & \\ \lambda_{v_6,h_1} & & & & & & \lambda_{v_1,h_1} & & & \\ \hdashline & \lambda_{v_4,h_2} & & & & & & \lambda_{v_2,h_2} & & \\ & \lambda_{v_5,h_2} & & & & & & & \lambda_{v_2,h_2} & \\ \hdashline & & \lambda_{v_6,h_3} & \lambda_{v_6,h_1} & & & \lambda_{v_3,h_1} & & & \lambda_{v_3,h_3} \\ \hline & & & \lambda_{v_4,h_1} & \lambda_{v_3,h_1} & & & & & \\ & & & & \lambda_{v_5,h_1} & \lambda_{v_4,h_1} & & \lambda_{v_5,h_2} & \lambda_{v_4,h_2} & \\ \hline & & & \lambda_{v_5,h_1} & & \lambda_{v_3,h_1} & & & & \\ & & & & \lambda_{v_6,h_1} & & \lambda_{v_4,h_1} & & & \\ & & & & & \lambda_{v_6,h_1} & \lambda_{v_5,h_1} & & & \end{array}\right) \end{array},$$

where void entries are zero. To obtain the matrix B^0 we set the entries λ_{v_1,h_1}, λ_{v_2,h_2}, λ_{v_3,h_3}, λ_{v_3,h_1}, λ_{v_4,h_1}, λ_{v_4,h_2}, λ_{v_5,h_2} equal to one and get

$$B^0 = \begin{array}{c} \\ \{v_1,v_3\} \\ \{v_1,v_4\} \\ \{v_1,v_5\} \\ \{v_1,v_6\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_3,v_6\} \\ \{v_3,v_4\} \\ \{v_4,v_5\} \\ \{v_3,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \end{array} \begin{array}{c} \begin{array}{cccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_3,h_1} & \lambda_{v_4,h_1} & \lambda_{v_5,h_1} & \lambda_{v_6,h_1} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_6,h_3} \end{array} \\ \left(\begin{array}{cccccccccc} 1 & & & 1 & & & & & & \\ 1 & & & & 1 & & & & & \\ & & & & & 1 & & & & \\ & & & & & & 1 & & & \\ \hdashline & 1 & & & & & & 1 & & \\ & 1 & & & & & & & 1 & \\ \hdashline & & & & & & 1 & & & 1 \\ \hline & & & 1 & 1 & & & & & \\ & & & & & 1 & & 1 & 1 & \\ \hline & & & & & 1 & & & & \\ & & & & & & 1 & & & \\ & & & & & & & & & \end{array}\right) \end{array}.$$

Claim 1: *By row reduction, the upper right block of B_0 can be transformed into a diagonal matrix of size $\sum_{i\in[m]} |C_i|$, while no fill-in occurs in the upper left block of B_0.*

For Claim 1 to be satisfied, we need to eliminate the nonzero entry with row index $\{v_3, v_6\}$ and column index λ_{v_6,h_1} by subtracting the fourth row indexed by $\{v_1, v_6\}$ from it. Clearly, this does not create any fill-ins in the upper left block of B^0. The upper right block is then diagonal and of size $\sum_{i\in[3]} |C_i| = 7$

Claim 2: *By row reduction, the submatrix of B_0 that consists of the rows indexed by S can be transformed such that the left block with columns indexed by $\lambda_{v_i,h_i}, ..., \lambda_{v_m,h_m}$ is of full row rank, and the right block with columns indexed by $\lambda_1^-, ..., \lambda_m^-$ is zero.*

As for Claim 2, we eliminate the nonzero entries in the rows indexed by $\{v_3, v_4\}$ and $\{v_4, v_5\}$. Both transformations mentioned in Claim 1 and Claim 2, result in B^0 having the form

$$
B^0 = \begin{array}{c} \\ \{v_1,v_3\} \\ \{v_1,v_4\} \\ \{v_1,v_5\} \\ \{v_1,v_6\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_3,v_6\} \\ \{v_3,v_4\} \\ \{v_4,v_5\} \\ \{v_3,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_3,h_1} & \lambda_{v_4,h_1} & \lambda_{v_5,h_1} & \lambda_{v_6,h_1} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_6,h_3} \end{array} \\
\left(\begin{array}{cccccccccc}
1 & & & 1 & & & & & & \\
1 & & & & 1 & & & & & \\
& & & & & 1 & & & & \\
& & & & & & 1 & & & \\
\hdashline
& 1 & & & & & & 1 & & \\
& 1 & & & & & & & 1 & \\
\hdashline
& & & & & & & & & 1 \\
\hline
-2 & & & & & & & & & \\
& -2 & & & & & & & & \\
\hline
& & & & & 1 & & & & \\
& & & & & & 1 & & & \\
& & & & & & & & &
\end{array}\right)
\end{array}.
$$

Clearly, Claim 2 holds for this matrix, and we can conclude that the matrix consisting of all except the last three rows is of full row rank, which is why

$$rank(B) \geq rank(B^0) \geq \sum_{i\in[3]} |C_i| + |S| = \sum_{i\in[3]} |A_{h_i}| = 9.$$

Hence, the lower bound from Theorem 3.3 is given by 15. Recalling Example 3.2, we observe that the upper and lower bound on the dimension coincide. This confirms that the dimension is indeed 15, as claimed in Example 2.11.

Remark 3.5. Note that there can potentially be several relabelings of observed and latent nodes such that the ZUTA condition is satisfied. Each relabeling can give a different lower bound in Theorem 3.3. To get as close as possible to the actual value of the dimension, we are interested in finding the maximal lower bound that can be found among all ZUTA-compliant, valid collections of all possible ZUTA relabelings.

4 Computation

In this section, we will focus on the computational aspects of the thesis. We begin by presenting the algorithms developed in R (R Core Team, 2024) to compute the upper and lower bound on the dimension of a sparse factor analysis model, as proposed in Theorem 3.1 and Theorem 3.3. This also includes a comparison of different approaches, illustrative examples and a complexity analysis. For a detailed review on the complexity of algorithms, we refer to Chapter 9 of Cormen et al. (2009). We cover the case of dimension determination for models with differing bounds by proposing a solution using the open-source computer algebra system MACAULAY2 (Grayson and Stillman). Furthermore, we apply all algorithms to run simulations and summarize the results obtained. Note, that the computations were performed on a single thread of an Apple M1 processor (3.2 GHz) with 8GB unified memory.

4.1 Algorithm to compute the upper bound

An intuitive approach to evaluate the upper bound on the dimension of a given sparse factor analysis graph is described in R pseudocode by Algorithm 1. In line 2 we begin by computing all $\binom{|V|}{2}$ 2-pairs of observed nodes. The function `generateAllCollections` is the core of Algorithm 1 and generates, as the name suggests, all possible collections of the just obtained 2-pairs. The function generates a total of $(2^{\binom{|V|}{2}} - 1)^{|\mathcal{H}|}$ collections. However, this number is extremely large even for graphs of moderate size, and the calculation crashes due to insufficient memory and long running times. For instance, the graph in Figure 2 (a), with only five observed nodes, already has $1.070.599.167$ distinct collections.

Algorithm 1 Evaluate the upper bound on the dimension of a sparse factor analysis graph

Input: sparse factor loading matrix Λ of $G = (\mathcal{H} \cup V, D)$, a factor analysis graph.
Output: upper bound on the dimension of the model corresponding to G.

1: $MaxCollection \leftarrow \{\}$.
2: $allpairs \leftarrow$ `generateAllpairs(`Λ`)`.
3: $allCollections \leftarrow$ `generateAllCollections(`Λ`,`$allpairs$`)`.
4: **for** $A \in allCollections$ **do**
5: **if** A *is valid* **and** $|A| > |MaxCollection|$ **then**
6: $MaxCollection \leftarrow A$.
7: **end if**
8: **end for**
9: $upperbound \leftarrow |MaxCollection| + |V|$.

Let us derive the theoretical worst-case complexity of Algorithm 1. Generating all 2-pairs of observed nodes in line 2 has complexity $\mathcal{O}(|V|^2)$. To compute all collections, we first use the function *powerSet* from the package *rje* to generate the power set of the set of all 2-pairs once, which has cardinality $2^{\binom{|V|}{2}}$ (Evans, 2022). The operation's complexity is given by $\mathcal{O}(\binom{|V|}{2}2^{\binom{|V|}{2}}) = \mathcal{O}(|V|^2 2^{|V|^2})$. We then take the $|\mathcal{H}|$- fold Cartesian product of the power set to obtain all collections, adding a complexity of $\mathcal{O}(2^{|V|^2|\mathcal{H}|})$. Let $s \in \mathbb{N}$ such that $|\operatorname{ch}(h_i)| \leq s$ for all $i \in \{1, ..., m\}$. For each collection, we verify three conditions in line 5, two of which involve simple cardinality checks. The computationally most expensive operation in line 5 is examining whether the components are pairwise disjoint, which has a complexity of $\mathcal{O}(s^2|\mathcal{H}|^2)$. Thus, iterating through every element of $allCollections$ and verifying all conditions of line five determines the running time of the Algorithm with $\mathcal{O}(2^{|V|^2|H|}s^2|\mathcal{H}|^2)$. This very high theoretical complexity is consistent with the issue previously described, and with the observed results shown in the second column of Table 1.

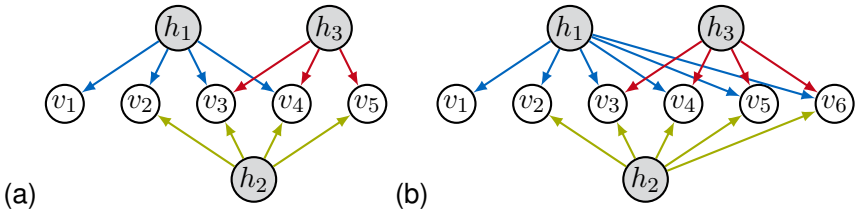

Figure 2 (a) ZUTA graph with 5 observed variables and 3 latent factors. (b) ZUTA graph with 6 observed variables and 3 latent factors.

To obtain results for at least all graphs with up to seven observed variables and up to three latent factors, it is necessary to use a more efficient method than the one presented in Algorithm 1. One possibility is demonstrated in Algorithm 2, which only computes collections of 2-pairs that satisfy Definition 2.12 and chooses one with maximal sum of cardinalities. The function `getchildrenpairs` generates a list of size m where each component corresponds to one latent node and contains all 2-pairs of children of the respective latent node. Then `getvalidcomponents` creates a list, also of size m, where each entry contains all combinations of the just generated children pairs of the corresponding latent node that satisfy condition (i) of Definition 2.12. For this purpose, the efficient function *comboGeneral* from the package *RcppAlgos* is used (Wood, 2024). By default, it sorts the combinations from smallest to largest. The Algorithm proceeds with a total of m loops, one for each latent factor. The loops verify that the second condition (ii) of Definition 2.12 holds, making sure that the resulting collection is valid. When the innermost loop is reached, the variable $MaxCollection$ is updated only if the resulting collection has higher sum of cardinalities than the one currently stored in $MaxCollection$. A crucial improvement in terms of efficiency is achieved by going through each loop in reverse order, since the largest combinations come last, as previously mentioned. This enables the algorithm to find a maximal valid collection of 2-pairs much faster.

Algorithm 2 Evaluate the upper bound on the dimension of a sparse factor analysis graph

Input: sparse factor loading matrix Λ of $G = (\mathcal{H} \cup V, D)$, a factor analysis graph.
Output: upper bound on the dimension of the model corresponding to G.

$MaxCollection \leftarrow \{\}$.
$childrenpairs \leftarrow$ `getchildrenpairs(`Λ`)`.
$allcomponents \leftarrow$ `getvalidcomponents(`Λ`,` $childrenpairs$`)`.
for $A_1 \in allcomponents[1]$ **do**
  for $A_2 \in allcomponents[2]$ **do**
    if $A_1 \cap A_2 = \emptyset$ **then**
      for $A_3 \in allcomponents[3]$ **do**
        if $A_1 \cap A_3 = \emptyset$ **and** $A_2 \cap A_3 = \emptyset$ **then**
          $\vdots$
          for $A_m \in allcomponents[m]$ **do**
            if $A_i \cap A_m = \emptyset$ **for all** $i < m$ **and** $|MaxCollection| < \sum_{h \in [m]} |A_h|$ **then**
              $MaxCollection \leftarrow (A_1, A_2, ..., A_m)$.
            end if
          end for
          $\vdots$
        end if
      end for
    end if
  end for
end for
$upperbound \leftarrow |MaxCollection| + |V|$.

Let us consider again the graph in Figure 2 (a). Then

$$|allcomponents[1]| = \binom{6}{1} + \binom{6}{2} + \binom{6}{3} + \binom{6}{4} = 56$$
$$|allcomponents[2]| = \binom{6}{1} + \binom{6}{2} + \binom{6}{3} + \binom{6}{4} = 56$$
$$|allcomponents[3]| = \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 7.$$

Consequently, using the approach in Algorithm 2, we iterate over at most $56 * 56 * 7 = 21.952$ collections. In the R code, additional break statements were added so that if $|MaxCollection| = \sum_{h_i \in \mathcal{H}} |\operatorname{ch}(h_i)|$, all m loops are immediately terminated, since this is already as high as cardinality is allowed to get; recall Definition 2.12. A comparison of the running times of Algorithm 1, Algorithm 2 without break statements and Algorithm 2 with break statements, all applied to the graphs in Figure 1 and Figure 2, is shown in Table 1. While the influence of the break conditions is subtle for the graphs in Figure 1 and Figure 2 (a), it becomes more significant as the number of edges increases. Note that the break condition only improves performance in the case where a maximal valid collection $\mathcal{A} = (A_{h_i})_{i \in \mathcal{H}}$ with $\sum_{h_i \in \mathcal{H}} |A_{h_i}| = \sum_{h_i \in \mathcal{H}} |\operatorname{ch}(h_i)|$ exists. The average run time of Algorithm 2 with break statements for graphs with the same number of edges is shown in Table 3. Note that only ZUTA graphs with up to seven observed nodes and up to three latent nodes were considered.

Graph	Alg. 1	Alg. 2 without break statements	Alg. 2 with break statements
Figure 1	Failed: Out of Memory	0.162201 secs	0.010082 secs
Figure 2 (a)	Failed: Out of Memory	0.437795 secs	0.397758 secs
Figure 2 (b)	Failed: Out of Memory	57.91415 mins	0.060979 secs

Table 1 Comparison of running times.

Theorem 4.1. *Let $G = (\mathcal{H} \cup V, D)$ be a factor analysis graph with $|\mathcal{H}| = m$ and $|V| = p$. The upper bound on the dimension of $F(G)$ given in Theorem 3.1 equals C if and only if Algorithm 2 returns C. The algorithm has complexity at most $\mathcal{O}(s^{2sm+2}m)$ where $s \in \mathbb{N}$ such that $s \geq |\operatorname{ch}(h_i)|$ for all $i \in \{1, ..., m\}$.*

Proof. We start by analyzing the worst-case complexity of the algorithm. First, we note that for each latent node h_i, there are $\binom{|\operatorname{ch}(h_i)|}{2}$ pairs of children. Let $s \in \mathbb{N}$ such that $s \geq |\operatorname{ch}(h_i)|$ for all $i \in \{1, ..., m\}$. Then

$$|childrenpairs[i]| = \binom{|\operatorname{ch}(h_i)|}{2} = \mathcal{O}(s^2).$$

Let $r_i := \min(|\operatorname{ch}(h_i)|, |childrenpairs[i]|)$. Note that if h_i has exactly two children, we get $r_i = 1$. If h_i has strictly more than two children, it holds that $r_i = |\operatorname{ch}(h_i)|$. In any case, this implies $r_i \leq s$ for all $i \in \{1, ..., m\}$. Now, the number of elements in $allcomponents[i]$ equals the number of subsets of $childrenpairs[i]$ with cardinality at most r_i. This is

$$|allcomponents[i]| = \sum_{k=1}^{r_i} \binom{|childrenpairs[i]|}{k} = \mathcal{O}(s^{2r_i}) \leq \mathcal{O}(s^{2s}).$$

The complexity of verifying that two components A_i, A_j are disjoint is given by

$$|A_i||A_j| \leq |\operatorname{ch}(h_i)||\operatorname{ch}(h_j)| = \mathcal{O}(s^2).$$

Thus, iterating through all m loops in lines 4-20, has complexity

$$\mathcal{O}(s^{2s}s^{2s}(s^2 + s^{2s}(2s^2 + s^{2s}(3s^2 + s^{2s}(\dots((m-2)s^2 + s^{2s}((m-1)s^2 + ms + 1)))))))$$
$$= \mathcal{O}(s^{2sm+2}m).$$

We get that the lines 4-20 of Algorithm 2 run in exponential time $\mathcal{O}(s^{2sm+2}m)$. Note that the function `getvalidcomponents` has complexity $\mathcal{O}(ms^{2s})$ using the same thoughts as before. Clearly, generating $childrenpairs$ is not of higher complexity and can thus be neglected. We can conclude that the algorithm's total complexity is given by $\mathcal{O}(s^{2sm+2}m)$.

Next, we show that the algorithm indeed returns an upper bound on the model's dimension. Suppose that the upper bound on the dimension of $F(G)$ in Theorem 3.1 is given by C. Then there must be at least one valid collection $\mathcal{A} = (A_1, ..., A_m)$ with sum of cardinalities equal to $C - |V| =: \tilde{C}$ and no valid collection with a sum of cardinalities exceeding $\tilde{C}$ can be found. By construction of the variable $allcomponents$, each component A_i of $\mathcal{A}$ has to be contained in the respective list $allcomponents[i]$. The algorithm iterates over all possible combinations of components, including the combination possibility given by $\mathcal{A}$. Thus, at some stage during the execution, the collection $\mathcal{A}$ will be examined. By definition, all components of $\mathcal{A}$ are pairwise disjoint and we reach the innermost loop. Now one of two cases can occur. Either the sum of cardinalities of $\mathcal{A}$ is larger than the sum of cardinalities of the collection currently stored in $MaxCollection$, forcing the algorithm to overwrite $MaxCollection$ by $\mathcal{A}$. Or a different valid collection $\tilde{\mathcal{A}}$ with $|\tilde{\mathcal{A}}| = |\mathcal{A}|$ has already been identified preventing $MaxCollection$ from being overwritten. However, in both cases a valid collection with sum of cardinalities $\tilde{C}$ is ultimately assigned to $MaxCollection$. Note that the case that $|\mathcal{A}|$ is strictly smaller than the sum of cardinalities of the collection stored in $MaxCollection$ cannot occur, since this would imply the existence of a valid collection $\mathcal{B}$ with $|\mathcal{B}| > \tilde{C}$. This contradicts what was previously stated. We conclude that the Algorithm returns $upperbound = |V| + \tilde{C}$ as needed.

Conversely, suppose the algorithm returns C. This translates to $|MaxCollection| = C - |V| =: \tilde{C}$. Consequently, the variable $MaxCollection$ must have been updated for the last time by a collection $\mathcal{A} = (A_1, ..., A_m)$ with cardinality $|\mathcal{A}| = \tilde{C}$. Additionally, it is ensured that the components of $\mathcal{A}$ are pairwise disjoint, since this is being verified in each level of the nested for-loop. Given that the function `getvalidcomponents` exclusively allows components that satisfy condition (i) of Definition 2.12, we can conclude that $\mathcal{A}$ is valid. It remains to show that $\mathcal{A}$ has maximal sum of cardinalities among all other valid collections. However, this follows easily given that in the case of $\mathcal{B}$ being a valid collection with $|\mathcal{B}| > |\mathcal{A}|$, we would have overwritten the variable $MaxCollection$ with $\mathcal{B}$ and the output of the algorithm would be strictly larger than C, since

$$|\mathcal{B}| + |V| > |\mathcal{A}| + |V| = C.$$

This leads to a contradiction. We conclude that $\mathcal{A}$ satisfies all conditions of Theorem 3.1, and therefore $C = |V| + \tilde{C}$ is indeed the upper bound on the model's dimension. □

Remark 4.2. We emphasize that although Algorithm 1 and Algorithm 2 both have a very high theoretical worst-case complexity, the running time of Algorithm 2 is significantly smaller in practice compared to Algorithm 1, see Table 1. Optimizations, such as the break statements, the reverse iteration through the loops, and the generation of only valid collections played a significant role in accelerating the process of finding the largest valid collection.

4.2 Algorithm to compute the largest lower bound

We now propose an algorithm that computes the largest lower bound on the dimension of a sparse ZUTA factor analysis graph, as determined by Theorem 3.3. To find the largest among all lower bounds, every possible relabeling of the factor loading matrix Λ has to be taken into consideration; recall Remark 3.5. Therefore, we first apply the function `getRelabelings` which generates all ZUTA relabelings of Λ. We proceed to iterate over all ZUTA relabelings and use `generateValidZUTACollections` to construct for each relabeling its valid, ZUTA-compliant collections. Note that the additional requirement of ZUTA compliance significantly reduces the number of collections to evaluate. After generating all these collections, we loop through them and update the variable $MaxCollection$ whenever a collection with larger sum of cardinalities is identified. Returning to our graph in Figure 2 (a), after running the function `getRelabelings`, we find that only three ZUTA relabelings of Λ must be examined. Executing `generateValidZUTACollections` for each relabeling reveals that two of them have exactly three

Graph	# ZUTA relabelings	# ZUTA compliant valid collections	running time Alg. 3
Figure 2 (a)	3	10	0.08212 secs
Figure 2 (b)	1	34	0.04112 secs
Figure 1	27	255	0.31838 secs

Table 2 Comparison of the total number of ZUTA relabelings, the total number of ZUTA compliant collections, and running times.

valid, ZUTA-compliant collections, while the third relabeling has four. Thus, in total we iterate over ten collections which results in the algorithm terminating quickly as can be seen in Table 2. The average run time of Algorithm 3 for graphs with the same number of edges is shown in Table 3. Note that only ZUTA graphs with up to seven observed nodes and up to three latent nodes were considered.

Algorithm 3 Evaluate the largest lower bound on the dimension of a sparse ZUTA factor analysis graph

Input: sparse factor loading matrix Λ of $G = (\mathcal{H} \cup V, D)$, a factor analysis graph that satisfies ZUTA.
Output: largest lower bound on the dimension of the model corresponding to G.

1: $MaxCollection \leftarrow \{\}$.
2: $allRelabelings \leftarrow$ `getRelabelings(`Λ`)`.
3: **for** $A \in allRelabelings$ **do**
4: $ZUTACollections \leftarrow$ `generateValidZUTACollections(`A`)`.
5: **for** $B \in ZUTACollections$ **do**
6: **if** $|B| > |MaxCollection|$ **then**
7: $MaxCollection \leftarrow B$.
8: **end if**
9: **end for**
10: **end for**
11: $lowerbound \leftarrow |MaxCollection| + |V|$.

Theorem 4.3. *Let $G = (\mathcal{H} \cup V, D)$ be a sparse factor analysis graph with $|\mathcal{H}| = m$ and $|V| = p$ that satisfies ZUTA. The largest lower bound on the dimension of $F(G)$, as determined by Theorem 3.3 equals C if and only if Algorithm 3 returns C. The algorithm has complexity at most $\mathcal{O}(|\mathcal{H}|!|V|!s^{2m+3}m^3)$ where $s \in \mathbb{N}$ such that $s \geq |\operatorname{ch}(h_i)|$ for all $i \in \{1, ..., m\}$.*

Proof. In terms of complexity, we note that there are $|\mathcal{H}|!$ permutations of columns, as well as $|V|!$ permutations of rows of the factor loading matrix Λ. The function, `getRelabelings` generates all possible $|\mathcal{H}|!|V|!$ matrix permutations, loops through them, and checks if the matrix is in ZUTA form. This is done by verifying that all diagonal entries are equal to one, while the upper triangle is zero. This adds a running time of $\mathcal{O}(|\mathcal{H}|^2)$. Thus, the total complexity of `getRelabelings` is given by $\mathcal{O}(|\mathcal{H}|!|V|!|\mathcal{H}|^2)$, while the cardinality of the variable $allRelabelings$ is bounded by $|\mathcal{H}|!|V|!$.
Let $s \in \mathbb{N}$ such that $|\operatorname{ch}(h_i)| \leq s$ for all $i \in [m]$. Then for each component A_{h_i} of a ZUTA compliant collection, there exist at most

$$1 + \binom{|\operatorname{ch}(h_i)| - 1}{2} \leq s^2$$

distinct possibilities if $|\operatorname{ch}(h_i)| > 2$, and one possibility, if $|\operatorname{ch}(h_i)| = 2$. This implies that the cardinality of the variable $ZUTACollections$ is bounded by s^{2m}. Let us derive the complexity of generating this variable in line 4. The function `generateValidZUTACollections` begins by generating the potentially relevant collections, of which there are at most s^{2m}, as previously stated. The complexity of this operation is $\mathcal{O}(s^{2m})$, proportional to the number of collections generated. It proceeds to loop through these collections and checks for each whether the components are pairwise disjoint. Recall that the complexity of verifying that two components A_i, A_j are disjoint is given by

$$|A_i||A_j| \leq |\operatorname{ch}(h_i)||\operatorname{ch}(h_j)| = \mathcal{O}(s^2).$$

For each collection, checking if its m components are pairwise disjoint, has a complexity of $\mathcal{O}(s^2m^2)$. We conclude that generating $ZUTACollections$ in line 4 has at most complexity $\mathcal{O}(s^{2m+2}m^2)$. The cardinality check in line 6 has complexity $\mathcal{O}(sm)$. To summarize, we have shown that the nested for loop in line 3-10 has at most complexity $\mathcal{O}(|\mathcal{H}|!|V|!s^{2m+3}m^3)$ which determines the complexity of Algorithm 3.
Next, we show that the algorithm indeed returns the largest lower bound on the dimension of a sparse factor analysis model. Suppose that C is the largest lower bound on the dimension of $F(G)$ found by applying Theorem 3.3 to all possible ZUTA relabelings. Then there must exist at least one valid, ZUTA-compliant collection $\mathcal{A}$ corresponding to a ZUTA relabeling of G, such that $|\mathcal{A}| = C - |V| =: \tilde{C}$. The sum of cardinalities of all further valid, ZUTA-compliant collections among all possible ZUTA relabelings cannot exceed $\tilde{C}$. Since we generate all ZUTA relabelings of G in line 2 as well as all corresponding valid, ZUTA-compliant collections in line 4, we must examine at some stage of the algorithm the collection $\mathcal{A}$ with the largest sum of cardinalities among all valid, ZUTA-compliant collections of all relabelings. At that point, either the collection stored in the variable $MaxCollection$ already has cardinality $\tilde{C}$ and $MaxCollection$ remains unchanged, or the collection stored in $MaxCollection$ has smaller cardinality, in which case $MaxCollection$ is overwritten by $\mathcal{A}$. Anyhow, the resulting sum of cardinalities of $MaxCollection$ is given by $\tilde{C}$ and the algorithm correctly returns the largest lower bound $C = |V| + \tilde{C}$.
Conversely, assume that the algorithm returns C. Then the variable $MaxCollection$ must have been overwritten the last time by a ZUTA-compliant, valid collection $\mathcal{A}$ with $|\mathcal{A}| = C - |V| =: \tilde{C}$ of some ZUTA relabeling of G. By construction, all elements in $ZUTACollections$ are both, valid and ZUTA-compliant, which allows us to apply Theorem 3.3. We obtain that C is indeed a lower bound on the dimension of $F(G)$. It remains to show that it is the largest. Therefore, we assume that $\mathcal{D}$ is another valid, ZUTA-compliant collection, possibly corresponding to a different ZUTA relabeling than $\mathcal{A}$, with $|\mathcal{D}| > |\mathcal{A}|$. Per construction, $\mathcal{D}$ has to be in the variable $ZUTACollections$ of the respective relabeling. Since $\mathcal{D}$ satisfies all conditions to overwrite the variable $MaxCollection$, the output of Algorithm 3 would be

$$|V| + |\mathcal{D}| > |V| + |\mathcal{A}| = C$$

This contradicts our previously made assumption, and we conclude that C is indeed the largest lower bound on the model's dimension that can be identified with Theorem 3.3. □

Remark 4.4. Note that in practice we almost never have to iterate over all $|\mathcal{H}|!|V|!$ permutation possibilities of the factor loading matrix Λ, since in the majority of cases there are only very few combinations of column and row permutations of Λ with the desired properties. Thus, finding the maximal lower bound is usually significantly faster than predicted by the theoretical worst-case running time, see Table 2 for reference.

4.3 Computing dimensions in Macaulay2

The purpose of this section is to demonstrate a method for computing the dimension of a factor analysis model, using the computer algebra system MACAULAY2. This is primarily relevant for models whose upper and lower bound, determined with Algorithm 2 and Algorithm 3, do not coincide.
We recall the algebraic background studied in Section 2.1, as well as the map τ_G in (7) and its dual ψ_{τ_G} in (8). To computationally derive $\dim(\mathrm{Im}(\tau_G))$ we must first understand that the set $\mathrm{Im}(\tau_G)$ is semi-algebraic. To get there, we define $\Omega = (\omega_{ij}) \in \mathbb{R}^{p\times p}$ and $\Lambda = (\lambda_{ij}) \in \mathbb{R}^{p\times m}$ and rewrite the domain of τ_G as

$$\begin{aligned}\mathbb{R}^p_{>0} \times \mathbb{R}^D = \{(\Omega,\Lambda) \in \mathbb{R}^{p\times p} \times \mathbb{R}^{p\times m} \mid g_{ij}(\Omega,\Lambda) = 0 \text{ for } i \in [p],\ j \in [m],\ j \to i \notin D \text{ and} \\ \tilde{g_{ij}}(\Omega,\Lambda) = 0 \text{ for } i \in [p],\ j \in [p],\ i \neq j \text{ and } h_{ii}(\Omega,\Lambda) > 0 \text{ for } i \in [p]\},\end{aligned}$$

where $g_{ij}(\Omega,\Lambda) = \lambda_{ij}$, $\tilde{g_{ij}}(\Omega,\Lambda) = \omega_{ij}$ and $h_{ii}(\Omega,\Lambda) = \omega_{ii}$. By Definition 2.4, $\mathbb{R}^p_{>0} \times \mathbb{R}^D$ is a semi-algebraic set. Since τ_G can be considered a polynomial mapping, Corollary 2.6 implies that $\mathrm{Im}(\tau_G)$ is also a semi-algebraic set. Bringing together these observations with the results obtained in Section 2.1 yields

$$\begin{aligned}&\dim(\mathbb{R}[\sigma_{ij}|1 \le i < j \le p]/\ker(\psi_{\tau_G})) \overset{(3)}{=} \dim(\mathbb{R}[\sigma_{ij}|1 \le i < j \le p]/\mathcal{I}(\mathrm{Im}(\tau_G))) \\ &\overset{(5)}{=} \dim(\mathbb{R}[\sigma_{ij}|1 \le i < j \le p]/\mathcal{I}(clos_{zar}(\mathrm{Im}(\tau_G))) \overset{(4)}{=} \dim(clos_{zar}(\mathrm{Im}(\tau_G)) \overset{(6)}{=} \dim(\mathrm{Im}(\tau_G)).\end{aligned} \tag{13}$$

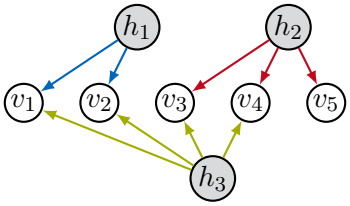

Figure 3 Smallest graph with differing lower and upper bounds, in terms of number of nodes and edges.

In MACAULAY2, applying the function `dim(·)` to an ideal `I` of a ring `R` computes the *Krull dimension* of the quotient ring `R/I`. In particular, $\ker(\psi_{\tau_G})$ is an ideal (Karpfinger, 2024, Lemma 15.1). Thus, computing its dimension in MACAULAY2 will actually return the dimension of $\mathbb{R}[\sigma_{ij}|1 \leq i < j \leq p]/\ker(\psi_{\tau_G})$. Now, with Equation (13) we get that this dimension coincides with $\dim(\mathrm{Im}(\tau_G))$. We summarize that in order to compute $\dim(\mathrm{Im}(\tau_G))$, it suffices to compute $\dim(\ker(\psi_{\tau_G}))$ in MACAULAY2.

Example 4.5. Consider the graph depicted in Figure 3. It corresponds to the smallest graph, in terms of number of nodes and edges, for which the upper bound and largest lower bound on the dimension, determined with Algorithm 2 and Algorithm 3, do not coincide. This graph is analyzed in more detail in Section 4.4.2. The code snippet below can be executed to determine the actual dimension of the corresponding model. First, the source and target ring, as well as the factor loading matrix and its transpose are defined. One can proceed by computing the corresponding covariance matrix and by defining the dual map ψ_{τ_G}. As previously stated, computing the dimension of the dual map's kernel returns the dimension of the factor analysis model. Running this code reveals that the model's dimension is 12.

```
-- define source and target ring of the dual map
S = QQ[x11, x12, x13, x14, x15, x22, x23,
       x24, x25, x33, x34, x35, x44, x45, x55];
T = QQ[omega1, omega2, omega3, omega4, omega5,
       lambda11, lambda21, lambda31, lambda41, lambda51,
       lambda12, lambda22, lambda32, lambda42, lambda52,
       lambda13, lambda23, lambda33, lambda43, lambda53];

-- define the factor loading matrix
Lambda = matrix{{lambda11, 0       , lambda13},
                {lambda21, 0       , lambda23},
                {0       , lambda32, lambda33},
                {0       , lambda42, lambda43},
                {0       , lambda52, 0       }};

LambdaT = transpose Lambda;

-- compute the covariance matrix
Cov = Lambda * LambdaT +
      diagonalMatrix {omega1, omega2, omega3, omega4, omega5};

-- compute the dual map, its kernel and the dimension
dualmap = map(T,S, matrix{{Cov_(0,0), Cov_(0,1), Cov_(0,2),
                           Cov_(0,3), Cov_(0,4), Cov_(1,1),
                           Cov_(1,2), Cov_(1,3), Cov_(1,4),
                           Cov_(2,2), Cov_(2,3), Cov_(2,4),
                           Cov_(3,3), Cov_(3,4), Cov_(4,4)}});
I = kernel(dualmap);
actualdim = dim(I);
```

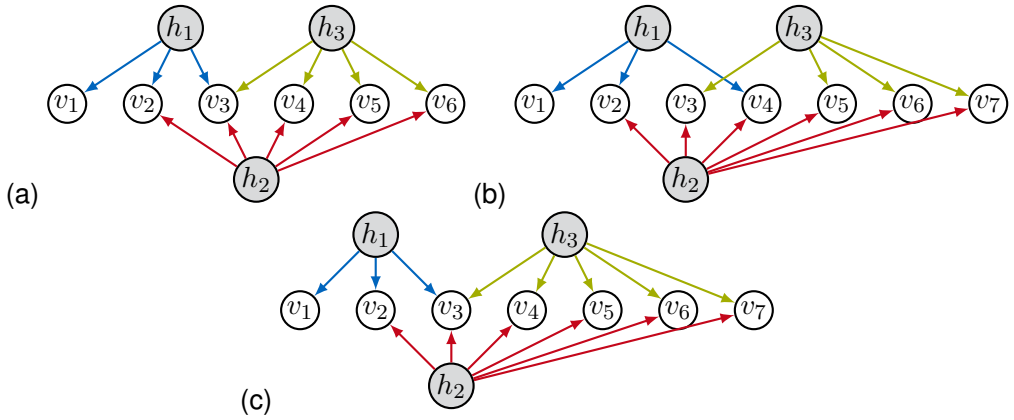

Figure 4 Graphs that have expected dimension, although upper and lower bounds do not coincide.

4.4 Simulations

4.4.1 Results on graphs up to three latent and seven observed nodes

We will now use the results from the previous sections to investigate the dimension properties of all ZUTA-graphs up to three latent nodes and up to seven observed nodes. After generating these graphs, we compute for each the upper bound and lower bound on the dimension using Algorithm 2 and Algorithm 3, as well as their expected and actual dimension. Calculation of the latter was performed using MACAULAY2, as illustrated in Example 4.5, for all graphs where the resulting upper bound and lower bound differed. In these particular cases, algebraic computation is necessary, since the theorems discussed in Section 3 are not sufficient to determine the actual dimension.

Computation in R revealed that the bounds on the dimension of six graphs differed. Among these, three graphs were identified in MACAULAY2 to have expected dimension and are depicted in Figure 4. Their upper bound and lower bound are 18, 20, 21 and 17, 19, 20 respectively, while their actual dimension equals 18, 20, 21. These graphs are of particular interest as the theorems in Section 3 fail to recognize expected dimension, given that the dimension's lower bound is not tight.

Let us inspect the graph in Figure 4 (a) as an example, to confirm that its true dimension is 18. Similarly as described in the proof of Theorem 3.3 it suffices to find a generic parameter choice, such that the rank of the block matrix B equals 12. The matrix B is given by

$$B = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_1,v_3\} \\ \{v_2,v_3\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_2,v_6\} \\ \{v_3,v_4\} \\ \{v_3,v_5\} \\ \{v_3,v_6\} \\ \{v_4,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \end{array} \begin{array}{c} \begin{array}{ccccccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_2,h_1} & \lambda_{v_3,h_1} & \lambda_{v_3,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_6,h_2} & \lambda_{v_4,h_3} & \lambda_{v_5,h_3} & \lambda_{v_6,h_3} \end{array} \\ \left(\begin{array}{ccccccccccccc} \lambda_{v_2,h_1} & & & \lambda_{v_1,h_1} & & & & & & & & \\ \lambda_{v_3,h_1} & & & & \lambda_{v_1,h_1} & & & & & & & \\ \hline & \lambda_{v_3,h_2} & & \lambda_{v_3,h_1} & \lambda_{v_2,h_1} & \lambda_{v_2,h_2} & & & & & & \\ & \lambda_{v_4,h_2} & & & & & \lambda_{v_2,h_2} & & & & & \\ & \lambda_{v_5,h_2} & & & & & & \lambda_{v_2,h_2} & & & & \\ & \lambda_{v_6,h_2} & & & & & & & \lambda_{v_2,h_2} & & & \\ \hline & & \lambda_{v_4,h_3} & & & \lambda_{v_4,h_2} & \lambda_{v_3,h_2} & & & \lambda_{v_3,h_3} & & \\ & & \lambda_{v_5,h_3} & & & \lambda_{v_5,h_2} & & \lambda_{v_3,h_2} & & & \lambda_{v_3,h_3} & \\ & & \lambda_{v_6,h_3} & & & \lambda_{v_6,h_2} & & & \lambda_{v_3,h_2} & & & \lambda_{v_3,h_3} \\ \hline & & & & & & \lambda_{v_5,h_2} & \lambda_{v_4,h_2} & & \lambda_{v_5,h_3} & \lambda_{v_4,h_3} & \\ & & & & & & \lambda_{v_6,h_2} & & \lambda_{v_4,h_2} & \lambda_{v_6,h_3} & & \lambda_{v_4,h_3} \\ & & & & & & & \lambda_{v_6,h_2} & \lambda_{v_5,h_2} & & \lambda_{v_6,h_3} & \lambda_{v_5,h_3} \end{array}\right) \end{array},$$

where void entries are zero. Assume first that we set all parameters to one and transform the upper right block into a diagonal matrix. Then the matrix B is given by

$$B^0 = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_1,v_3\} \\ \{v_2,v_3\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_2,v_6\} \\ \{v_3,v_4\} \\ \{v_3,v_5\} \\ \{v_3,v_6\} \\ \{v_4,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \end{array} \begin{array}{c} \begin{array}{cccccccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_2,h_1} & \lambda_{v_3,h_1} & \lambda_{v_3,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_6,h_2} & \lambda_{v_4,h_3} & \lambda_{v_5,h_3} & \lambda_{v_6,h_3} \end{array} \\ \left(\begin{array}{cccccccccccc} 1 & & & 1 & & & & & & & & \\ 1 & & & & 1 & & & & & & & \\ -2 & 1 & & & & 1 & & & & & & \\ & 1 & & & & & 1 & & & & & \\ & 1 & & & & & & 1 & & & & \\ & 1 & & & & & & & 1 & & & \\ 2 & -2 & 1 & & & & & & & 1 & & \\ 2 & -2 & 1 & & & & & & & & 1 & \\ 2 & -2 & 1 & & & & & & & & & 1 \\ & & & & & & 1 & 1 & & 1 & 1 & \\ & & & & & & 1 & & 1 & 1 & & 1 \\ & & & & & & & 1 & 1 & & 1 & 1 \end{array}\right) \end{array}.$$

Consider the last three rows indexed by $\{v_i, v_j\}$ with $i \in \{4,5\}, j \in \{5,6\}, i \neq j$. To eliminate all nonzero entries in these rows, we subtract from each the corresponding four rows indexed by $\{v_2, v_i\}$, $\{v_2, v_j\}$, $\{v_3, v_i\}$, $\{v_3, v_j\}$. The structure of the matrix immediately indicates that these row-reduction operations will create identical fill-ins in each row of the lower left block.

$$B^0 = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_1,v_3\} \\ \{v_2,v_3\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_2,v_6\} \\ \{v_3,v_4\} \\ \{v_3,v_5\} \\ \{v_3,v_6\} \\ \{v_4,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \end{array} \begin{array}{c} \begin{array}{cccccccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_2,h_1} & \lambda_{v_3,h_1} & \lambda_{v_3,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_6,h_2} & \lambda_{v_4,h_3} & \lambda_{v_5,h_3} & \lambda_{v_6,h_3} \end{array} \\ \left(\begin{array}{cccccccccccc} 1 & & & 1 & & & & & & & & \\ 1 & & & & 1 & & & & & & & \\ -2 & 1 & & & & 1 & & & & & & \\ & 1 & & & & & 1 & & & & & \\ & 1 & & & & & & 1 & & & & \\ & 1 & & & & & & & 1 & & & \\ 2 & -2 & 1 & & & & & & & 1 & & \\ 2 & -2 & 1 & & & & & & & & 1 & \\ 2 & -2 & 1 & & & & & & & & & 1 \\ -4 & 2 & -2 & & & & & & & & & \\ -4 & 2 & -2 & & & & & & & & & \\ -4 & 2 & -2 & & & & & & & & & \end{array}\right) \end{array}.$$

This shows that setting all parameters to one is ineffective here, since it results in B having rank 10 which causes the Jacobian to have rank 16. This is strictly smaller than the expected dimension. To obtain full row rank, we must make a different parameter choice, such that the resulting fill-ins in the lower left block do not create linearly dependent rows. Thus, we will now consider setting at least two parameters to different values. Indeed, by assigning, for example, λ_{v_5,h_3} to two and λ_{v_6,h_3} to three, while setting all remaining entries to one, we achieve the desired result. The matrix's structure with those new entries, after removing all nonzero entries below the diagonal in the upper right block is given by

$$B^0 = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_1,v_3\} \\ \{v_2,v_3\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_2,v_6\} \\ \{v_3,v_4\} \\ \{v_3,v_5\} \\ \{v_3,v_6\} \\ \{v_4,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \end{array} \begin{array}{c} \begin{array}{cccccccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_2,h_1} & \lambda_{v_3,h_1} & \lambda_{v_3,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_6,h_2} & \lambda_{v_4,h_3} & \lambda_{v_5,h_3} & \lambda_{v_6,h_3} \end{array} \\ \left(\begin{array}{cccccccccccc} 1 & & & 1 & & & & & & & & \\ 1 & & & & 1 & & & & & & & \\ -2 & 1 & & & & 1 & & & & & & \\ & 1 & & & & & 1 & & & & & \\ & 1 & & & & & & 1 & & & & \\ & 1 & & & & & & & 1 & & & \\ 2 & -2 & 1 & & & & & & & 1 & & \\ 2 & -2 & 2 & & & & & & & & 1 & \\ 2 & -2 & 3 & & & & & & & & & 1 \\ & & & & & & 1 & 1 & & 2 & 1 & \\ & & & & & & 1 & & 1 & 3 & & 1 \\ & & & & & & & 1 & 1 & & 3 & 2 \end{array}\right) \end{array}.$$

Proceeding as before, consider the rows indexed by $\{v_i, v_j\}$, with $i \in \{4,5\}, j \in \{5,6\}, i \neq j$. We subtract from each of these rows, the rows indexed by $\{v_2, v_i\}, \{v_2, v_j\}$, as well as λ_{v_j,h_3} times the row indexed by $\{v_3, v_i\}$, and λ_{v_i,h_3} times the row indexed by $\{v_3, v_j\}$, respectively. This yields

$$B^0 = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_1,v_3\} \\ \{v_2,v_3\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_2,v_6\} \\ \{v_3,v_4\} \\ \{v_3,v_5\} \\ \{v_3,v_6\} \\ \{v_4,v_5\} \\ \{v_4,v_6\} \\ \{v_5,v_6\} \end{array} \begin{array}{c} \begin{array}{cccccccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_2,h_1} & \lambda_{v_3,h_1} & \lambda_{v_3,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_6,h_2} & \lambda_{v_4,h_3} & \lambda_{v_5,h_3} & \lambda_{v_6,h_3} \end{array} \\ \left(\begin{array}{cccccccccccc} 1 & & & 1 & & & & & & & & \\ 1 & & & & 1 & & & & & & & \\ -2 & 1 & & & & 1 & & & & & & \\ & 1 & & & & & 1 & & & & & \\ & 1 & & & & & & 1 & & & & \\ & 1 & & & & & & & 1 & & & \\ 2 & -2 & 1 & & & & & & & 1 & & \\ 2 & -2 & 2 & & & & & & & & 1 & \\ 2 & -2 & 3 & & & & & & & & & 1 \\ -6 & 4 & -4 & & & & & & & & & \\ -8 & 6 & -6 & & & & & & & & & \\ -10 & 8 & -12 & & & & & & & & & \end{array}\right) \end{array}.$$

We can now easily convince ourselves that the fill-ins that occurred result in the lower left block having full rank of three, since

$$\begin{pmatrix} -6 & 4 & -4 \\ -8 & 6 & -6 \\ -10 & 8 & -12 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} -6 & 4 & -4 \\ 0 & \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 & -4 \end{pmatrix}.$$

Consequently, the generic parameter choice $\lambda_{v_5,h_3} = 2$ and $\lambda_{v_6,h_3} = 3$ leads to the matrix block B having rank 12. The Jacobian matrix has full rank of 18 which confirms that the graph in Figure 4 (a) has expected dimension. For the graphs (b) and (c) in Figure 4 we can use the same argumentation as above. With setting $\lambda_{v_6,h_3} = 2$ and $\lambda_{v_7,h_3} = 3$ and assigning all remaining entries of graph (b) to one, the resulting Jacobian has rank 20. For graph (c) the same parameter choice, $\lambda_{v_6,h_3} = 2$ and $\lambda_{v_7,h_3} = 3$, results in a Jacobian of rank 21, as needed.

Bringing together the results computed in MACAULAY2 and R we obtain Table 3 which summarizes, for how many of all generated ZUTA graphs, the expected dimension and the actual dimension coincide, and for how many the upper bound and the lower bound coincide, grouped by the number of edges. Additionally the average run times of Algorithm 2 and Algorithm 3 are displayed for each number of edges.

# edges	# ZUTA graphs	# graphs with expected dim.	# graphs with coinciding bounds	avg. run time Alg. 2 in secs	avg. run time Alg. 3 in secs
2	1	1	1	6.40e-04	6.90e-04
3	1	1	1	6.90e-04	8.20e-04
4	3	1	3	7.70e-04	9.99e-03
5	4	2	4	1.02e-03	2.25e-02
6	11	3	11	1.16e-03	3.16e-02
7	18	5	18	1.82e-03	3.37e-01
8	32	7	32	1.69e-03	1.09e+00
9	48	9	47	3.46e-03	2.17e+00
10	60	11	59	1.39e-02	3.02e+00
11	66	25	65	1.19e-01	3.06e+00
12	60	33	59	8.52e-01	2.53e+00
13	50	35	49	5.71e+00	1.61e+00
14	30	25	29	2.60e+01	1.06e+00
15	18	17	18	8.80e-03	5.93e-01
16	8	8	8	7.12e-03	3.47e-01
17	4	4	4	1.94e-02	1.52e-01
18	1	1	1	2.36e-01	8.54e-02
Total	415	188	409	-	-

Table 3 Results on all ZUTA graphs with at least two children per latent node and with $|V| \leq 7$ and $|\mathcal{H}| \leq 3$. The average running times of Algorithm 2 and Algorithm 3 are given in seconds.

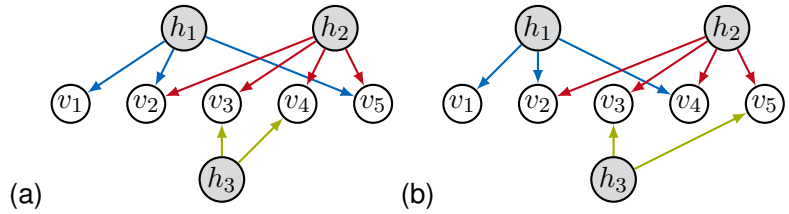

Figure 5 (a) & (b) The only two relabelings of the graph in Figure 3 that satisfy ZUTA .

4.4.2 Smallest graph for which the bounds differ

The simulation performed in Section 4.4.1 revealed that the graph in Figure 3 is the smallest graph, in terms of number of nodes and edges, where the upper bound and the largest lower bound do not coincide. Note that for the purpose of this example, we will also examine the graphs in Figure 5 (a) and (b) which depict the only two relabeled versions of the graph in Figure 3 that are in ZUTA form. The smallest upper bound found is 13, whereas the largest lower bound among both ZUTA relabeling possibilities is 12. In Example 4.5 we showed that the model's dimension is 12 and thus coincides with the largest lower bound. Let us now retrace the bound evaluation. Consider the graph in Figure 3 and the maximal, valid collection

$$\begin{aligned} A_{h_1} &= \{\{v_1, v_2\}\}, \\ A_{h_2} &= \{\{v_3, v_4\}, \{v_3, v_5\}, \{v_4, v_5\}\}, \\ A_{h_3} &= \{\{v_1, v_3\}, \{v_1, v_4\}, \{v_2, v_3\}, \{v_2, v_4\}\}. \end{aligned} \tag{14}$$

It follows that $A^C = \{\{v_1, v_5\}, \{v_2, v_5\}\}$. Our index sets are defined by $I^{(=)} = \{2, 3\}$ and consequently $I^{(<)} = \{1\}$. In order to determine the maximal rank of the Jacobian of the map τ_G, we need to inspect its submatrix B. Recalling Equation (9) and Equation (11) yields

$$B = \begin{array}{c} \\ \{v_1, v_2\} \\ \{v_3, v_4\} \\ \{v_3, v_5\} \\ \{v_4, v_5\} \\ \{v_1, v_3\} \\ \{v_1, v_4\} \\ \{v_2, v_3\} \\ \{v_2, v_4\} \\ \{v_1, v_5\} \\ \{v_2, v_5\} \end{array} \begin{array}{c} \begin{array}{cccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_1} & \lambda_{v_3,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_1,h_3} & \lambda_{v_2,h_3} & \lambda_{v_3,h_3} & \lambda_{v_4,h_3} \end{array} \\ \left(\begin{array}{ccccccccc} \lambda_{v_2,h_1} & \lambda_{v_1,h_1} & & & & \lambda_{v_2,h_3} & \lambda_{v_1,h_3} & & \\ & & \lambda_{v_4,h_2} & \lambda_{v_3,h_2} & & & & \lambda_{v_4,h_3} & \lambda_{v_3,h_3} \\ & & \lambda_{v_5,h_2} & & \lambda_{v_3,h_2} & & & & \\ & & & \lambda_{v_5,h_2} & \lambda_{v_4,h_2} & & & & \\ & & & & & \lambda_{v_3,h_3} & & \lambda_{v_1,h_3} & \\ & & & & & \lambda_{v_4,h_3} & & & \lambda_{v_1,h_3} \\ & & & & & & \lambda_{v_3,h_3} & \lambda_{v_2,h_3} & \\ & & & & & & \lambda_{v_4,h_3} & & \lambda_{v_2,h_3} \\ & & & & & & & & \\ & & & & & & & & \end{array}\right) \end{array},$$

where void entries are zero. Since B has eight nonzero rows, we conclude

$$\operatorname{rank}(B) \leq 8 = \sum_{h \in \mathcal{H}} |A_h|. \tag{15}$$

Next we derive the maximal lower bound on the dimension. In order to get there, we aim to identify among all ZUTA-compliant, valid collections of all ZUTA relabelings, one with maximal cardinality. Let us consider the ZUTA graph in Figure 5 (a) first. According to Definition 2.15 the following 2-pairs are mandatory in the respective components of any valid, ZUTA-compliant collection

$$\begin{aligned} C_1 &= \{\{v_1, v_2\}, \{v_1, v_5\}\}, \\ C_2 &= \{\{v_2, v_3\}, \{v_2, v_4\}, \{v_2, v_5\}\}, \\ C_3 &= \{\{v_3, v_4\}\}. \end{aligned}$$

Hence, $C(V,2)_{h_1} \setminus C_1 = \{\{v_2, v_5\}\}$, as well as $C(V,2)_{h_2} \setminus C_2 = \{\{v_3, v_4\}, \{v_3, v_5\}, \{v_4, v_5\}\}$ and $C(V,2)_{h_3} \setminus C_3 = \emptyset$. Recalling Remark 2.16, we are interested in extending as many sets C_i as possible by exactly one 2-pair from the corresponding set $C(V,2)_{h_i} \setminus C_i$. Clearly, C_3 cannot be increased any further, so we are forced to set $A_{h_3} := C_3$. However, C_1 cannot be extended either, as the only remaining 2-pair of children of h_1 is $\{v_2, v_5\}$, which is already an element of C_2. It is crucial to understand that in this specific example there's no ZUTA-compliant, valid collection where $|A_{h_1}| = |\operatorname{ch}(h_1)|$ and we thus have to set $A_{h_1} := C_1$. Contrarily to the previous two components, one can add exactly one of the 2-pairs $\{\{v_3, v_5\}, \{v_4, v_5\}\} \subseteq C(V,2)_{h_2} \setminus C_2$ to C_2 as neither appears in A_{h_1} nor A_{h_3}. This shows that the collection $\mathcal{A} = (A_h)_{h \in \mathcal{H}}$ with

$$\begin{aligned} A_{h_1} &= C_1 = \{\{v_1, v_2\}, \{v_1, v_5\}\}, \\ A_{h_2} &= C_2 \cup \{\{v_3, v_5\}\} = \{\{v_2, v_3\}, \{v_2, v_4\}, \{v_2, v_5\}, \{v_3, v_5\}\}, \\ A_{h_3} &= C_3 = \{\{v_3, v_4\}\} \end{aligned}$$

satisfies all necessary criteria to be valid and ZUTA-compliant and has maximal sum of cardinalities, $\sum_{h \in \mathcal{H}} |A_h| = 7$, among all ZUTA-compliant collections of the relabeling given in Figure 5 (a). For completeness, we verify that no larger ZUTA-compliant collection can be found for the relabeled graph in Figure 5 (b). Following the same argumentation as before with the roles of the nodes v_4 and v_5 interchanged, we get that $\tilde{\mathcal{A}} = (\tilde{A}_h)_{h \in \mathcal{H}}$ with

$$\begin{aligned} \tilde{A}_{h_1} &= \{\{v_1, v_2\}, \{v_1, v_4\}\}, \\ \tilde{A}_{h_2} &= \{\{v_2, v_3\}, \{v_2, v_4\}, \{v_2, v_5\}, \{v_3, v_4\}\}, \\ \tilde{A}_{h_3} &= \{\{v_3, v_5\}\} \end{aligned}$$

is valid, ZUTA-compliant and of maximal cardinality. However, since $\sum_{h \in \mathcal{H}} |A_h| = \sum_{h \in \mathcal{H}} |\tilde{A}_h|$ we may proceed with either collection, as both will return the same maximal lower bound. In the following, we will continue with the collection $\mathcal{A} = (A_h)_{h \in \mathcal{H}}$. In our specific case, the matrix B can be written as

$$B = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_1,v_5\} \\ \{v_2,v_3\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_3,v_4\} \\ \{v_3,v_5\} \\ \{v_4,v_5\} \end{array} \begin{array}{c} \begin{array}{ccccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_2,h_1} & \lambda_{v_5,h_1} & \lambda_{v_3,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_4,h_3} \end{array} \\ \left(\begin{array}{ccc|ccccccc} \lambda_{v_2,h_1} & & & \lambda_{v_1,h_1} & & & & & \\ \lambda_{v_5,h_1} & & & & \lambda_{v_1,h_1} & & & & \\ \hline & \lambda_{v_3,h_2} & & & & \lambda_{v_2,h_2} & & & \\ & \lambda_{v_4,h_2} & & & & & \lambda_{v_2,h_2} & & \\ & \lambda_{v_5,h_2} & & & & & & \lambda_{v_2,h_2} & \\ \hline & & \lambda_{v_4,h_3} & & & \lambda_{v_4,h_2} & \lambda_{v_3,h_2} & & \lambda_{v_3,h_3} \\ \hline & & & & & \lambda_{v_5,h_2} & & \lambda_{v_3,h_2} & \\ & & & & & & \lambda_{v_5,h_2} & \lambda_{v_4,h_2} & \end{array}\right) \end{array};$$

recall Equation (9) and Equation (12). Setting the parameters λ_{v_1,h_1}, λ_{v_2,h_2}, λ_{v_3,h_3}, λ_{v_3,h_2}, λ_{v_5,h_2} equal to one, and all remaining entries to zero, results in B^0 having the form

$$B^0 = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_1,v_5\} \\ \{v_2,v_3\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_3,v_4\} \\ \{v_3,v_5\} \\ \{v_4,v_5\} \end{array} \begin{array}{c} \begin{array}{ccccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_2,h_1} & \lambda_{v_5,h_1} & \lambda_{v_3,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_4,h_3} \end{array} \\ \left(\begin{array}{ccc|ccccccc} & & & 1 & & & & & \\ & & & & 1 & & & & \\ \hline & 1 & & & & 1 & & & \\ & & & & & & 1 & & \\ & 1 & & & & & & 1 & \\ \hline & & & & & & 1 & & 1 \\ \hline & & & & & 1 & & 1 & \\ & & & & & & 1 & & \end{array}\right) \end{array}.$$

We proceed by row reduction and subtract the rows indexed by $\{v_2, v_3\}$ and $\{v_2, v_5\}$ from the row indexed by $\{v_3, v_5\}$ to eliminate all nonzero entries in this row's right block.

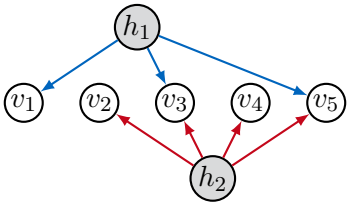

Figure 6 Same graph as in Figure 5 (a), after removing h_3 and swapping v_2 and v_3.

$$
B^0 = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_1,v_5\} \\ \{v_2,v_3\} \\ \{v_2,v_4\} \\ \{v_2,v_5\} \\ \{v_3,v_4\} \\ \{v_3,v_5\} \\ \{v_4,v_5\} \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} \lambda_{v_1,h_1} & \lambda_{v_2,h_2} & \lambda_{v_3,h_3} & \lambda_{v_2,h_1} & \lambda_{v_5,h_1} & \lambda_{v_3,h_2} & \lambda_{v_4,h_2} & \lambda_{v_5,h_2} & \lambda_{v_4,h_3} \end{array} \\
\left(\begin{array}{ccc:cccccc:cc}
 & & & 1 & & & & & & \\
 & & & & 1 & & & & & \\ \hdashline
 & 1 & & & & 1 & & & & \\
 & & & & & & 1 & & & \\
 & 1 & & & & & & 1 & & \\ \hdashline
 & & & & & & 1 & & 1 & \\ \hline
 & -2 & & & & & & & & \\ \hline
 & & & & & & 1 & & &
\end{array}\right)
\end{array}.
$$

After these two operations, the resulting fill-in in the row indexed by $\{v_3, v_5\}$ ensures that the submatrix of B^0 consisting of all but the last row has full row rank. Consequently,

$$\operatorname{rank}(B) \geq \operatorname{rank}(B^0) \geq 7. \tag{16}$$

The result $13 \overset{(15)}{\geq} \dim(F(G)) = \operatorname{rank}(B) + |V| \overset{(16)}{\geq} 12$ aligns with the outcome of the computation in R.

Remark 4.6. Note that the inequality between the upper bound and lower bound on the dimension of the graph in Figure 3 arises because we were not able to find a valid, ZUTA-compliant collection with the same sum of cardinalities as the maximal valid collection presented in Equation (14). However, if we were to consider the graph in Figure 6, which is obtained by removing the latent node h_3 from the graph in Figure 5 (a), and by permuting the latent nodes v_2 and v_3, ensuring that ZUTA is satisfied, results will differ significantly. To be more precise, its upper bound and lower bound on the dimension align. We can convince ourselves that

$$
\begin{aligned}
A_{h_1} &= \{\{v_1, v_3\}, \{v_1, v_5\}, \{v_3, v_5\}\}, \\
A_{h_2} &= \{\{v_2, v_3\}, \{v_2, v_4\}, \{v_2, v_5\}, \{v_4, v_5\}\}
\end{aligned}
$$

is a valid, ZUTA-compliant collection with maximal sum of cardinalities since $|A_{h_1}| = |\operatorname{ch}(h_1)|$ as well as $|A_{h_2}| = |\operatorname{ch}(h_2)|$. The main difference is that having only two latent nodes enables, in this case, a relabeling of the nodes where $v_2 \notin \operatorname{ch}(h_1)$. This adjustment avoids the exclusion of the third 2-pair of children of h_1, which was previously necessary. We conclude that the dimension's lower bound equals its upper bound and consequently

$$\dim(F(G)) = |V| + \sum_{h \in \mathcal{H}} |A_h| = 12.$$

5 Dependent factors

In this part of the thesis we will drop the previously made assumption of independent latent nodes. Instead, we will consider models with a parametrization that includes a latent covariance matrix Φ, which is distinct from the identity matrix. Therefore, we let $G = (V \cup \mathcal{H}, D, B)$ be a mixed graph, where V and $\mathcal{H}$ are finite and disjoint sets of observed and latent nodes, respectively, and $D \subseteq \mathcal{H} \times V$, $B \subseteq \mathcal{H} \times \mathcal{H}$ are two sets of edges. The set D contains all directed edges $h \to v$, that point from latent to observed nodes, while the pairs in B represent bidirected edges $h \leftrightarrow \tilde{h}$ between latent nodes. The pairs in B have no orientation, which is why $h \leftrightarrow \tilde{h} \in B$ if and only if $\tilde{h} \leftrightarrow h \in B$.
For a definition of the factor analysis model, which is identified with the set of its covariance matrices, we let $\mathbb{R}^D$, $\mathbb{R}^p_{>0}$ and $\mathrm{PD}(p)$ denote the same sets as in Section 2.2. Additionally, we will define the set $\mathbb{R}^B = \{\Phi \in \mathrm{PD}(m) : \Phi_{hh} = 1 \text{ and } \Phi_{h\tilde{h}} = 0 \text{ if } h \leftrightarrow \tilde{h} \notin B, h \neq \tilde{h}\}$.

Definition 5.1. *Let $G = (V \cup \mathcal{H}, D, B)$ be a mixed factor analysis graph with $|V| = p$ and $|\mathcal{H}| = m$. As a model of the covariance matrix, the factor analysis model determined by G is the image $F(G) = \mathrm{Im}(\tau_G)$ of the parametrization map*

$$\begin{aligned} \tau_G : \mathbb{R}^p_{>0} \times \mathbb{R}^B \times \mathbb{R}^D &\longrightarrow \mathrm{PD}(p) \\ (\Omega, \Phi, \Lambda) &\longmapsto \Omega + \Lambda\Phi\Lambda^\top. \end{aligned} \tag{17}$$

From now on, τ_G will always refer to the map in (17). Recall Proposition 2.1 which states that the dimension of the model $F(G)$ coincides with the maximal rank of the Jacobian matrix of τ_G. A natural upper bound on the dimension is always given by the model's expected dimension, $\min\{|V| + |D| + |B|, \binom{|V|+1}{2}\}$.
In this thesis, the influence of dependent latent nodes on a model's dimension, will be analyzed exclusively for a very particular case of graphs, which we will refer to as *2-chain graphs*. They are characterized by each latent node having two *pure* children and by a tridiagonal latent covariance matrix.

Definition 5.2. *(Huang et al., 2022, Definition 2) An observed node $v \in V$ is a pure child of a latent node $h \in \mathcal{H}$ in a graph G, if $\mathrm{pa}(v) = \{h\}$. That is, v has no other parents than h.*

Definition 5.3. *Let $G = (V \cup \mathcal{H}, D, B)$ be a mixed graph. We say that G is a* 2-chain graph *if*

(i) each latent node h_i with $i \in [m]$ has exactly two children, where both are pure children, and

(ii) $B = \{(h_i, h_{i+1}) | i \in [m-1]\}$.

There is then a relabeling of the observed nodes $V = \{v_1, \ldots, v_p\}$ such that $\mathrm{ch}(h_i) = \{v_{2i-1}, v_{2i}\}$ for all $i \in [m]$.

Let G be a 2-chain graph with m factors. The corresponding latent covariance matrix $\Phi \in \mathbb{R}^B$ is symmetric and tridiagonal. To simplify notation, we will from now on replace the entries $\phi_{h_i h_{i+1}}$ and $\phi_{h_{i+1} h_i}$ with ϕ_i for all $i \in [m-1]$. Thus, the latent covariance matrix of G can be written as

$$\Phi = \begin{pmatrix} 1 & \phi_1 & & & & \\ \phi_1 & 1 & \phi_2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \phi_{m-2} & 1 & \phi_{m-1} & \\ & & & \phi_{m-1} & 1 & \end{pmatrix}, \tag{18}$$

where void entries are zero.

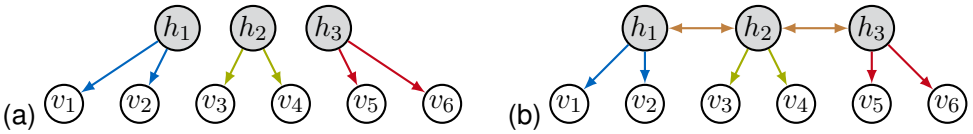

Figure 7 (a) A graph where each latent node has exactly 2 children and both are pure. (b) The 2-chain graph that corresponds to the graph in Figure 7 (a).

Example 5.4. Consider the graph in Figure 7 (a). For simplicity, we identify in this example the nodes $v_1, \ldots, v_6$ with the integers $1, \ldots, 6$, and the factors h_1, h_2, h_3 with the integers $1, 2, 3$. Clearly, each latent node has exactly two children, where both are pure. The model's expected dimension is given by 12. Computing the actual dimension by following the approach described in Section 4.3 returns 9, indicating that dimension drops by three. Let us instead consider the corresponding 2-chain graph, depicted in Figure 7 (b).The six observed nodes, $v_1, \ldots, v_6$, are already ordered as desired. Furthermore, the expected dimension is 14 and the latent covariance matrix Φ is given by

$$\Phi = \begin{pmatrix} 1 & \phi_1 & 0 \\ \phi_1 & 1 & \phi_2 \\ 0 & \phi_2 & 1 \end{pmatrix}.$$

This gives rise to the covariance matrix $\Sigma \in F(G)$ of the form

$$\Sigma = (\sigma_{uv}) = \begin{pmatrix} \omega_{11}+\lambda_{11}^2 & \lambda_{11}\lambda_{21} & \lambda_{11}\lambda_{32}\phi_1 & \lambda_{11}\lambda_{42}\phi_1 & 0 & 0 \\ \lambda_{11}\lambda_{21} & \omega_{22}+\lambda_{21}^2 & \lambda_{21}\lambda_{32}\phi_1 & \lambda_{21}\lambda_{42}\phi_1 & 0 & 0 \\ \lambda_{11}\lambda_{32}\phi_1 & \lambda_{21}\lambda_{32}\phi_1 & \omega_{33}+\lambda_{32}^2 & \lambda_{32}\lambda_{42} & \lambda_{32}\lambda_{53}\phi_2 & \lambda_{32}\lambda_{63}\phi_2 \\ \lambda_{11}\lambda_{42}\phi_1 & \lambda_{21}\lambda_{42}\phi_1 & \lambda_{32}\lambda_{42} & \omega_{44}+\lambda_{42}^2 & \lambda_{42}\lambda_{53}\phi_2 & \lambda_{42}\lambda_{63}\phi_2 \\ 0 & 0 & \lambda_{32}\lambda_{53}\phi_2 & \lambda_{42}\lambda_{53}\phi_2 & \omega_{55}+\lambda_{53}^2 & \lambda_{53}\lambda_{63} \\ 0 & 0 & \lambda_{32}\lambda_{63}\phi_2 & \lambda_{42}\lambda_{63}\phi_2 & \lambda_{53}\lambda_{63} & \omega_{66}+\lambda_{63}^2 \end{pmatrix}.$$

We will verify in Example 5.6 that the graph in Figure 7 (b) has expected dimension.

Let G be a 2-chain graph whose observed nodes are already ordered as stated in Definition 5.3. In what follows, to avoid confusion, $h_{v_i} \in \mathcal{H}$ will refer to the only parent of the observed node $v_i \in V$. We aim to generalize the parametrization of the entries $\sigma_{v_iv_j}$ of a matrix $\Sigma \in F(G)$. Since Σ is symmetric, it suffices to consider the entries $\sigma_{v_iv_j}$ for $i \leq j$. In particular, for $\Omega = \operatorname{diag}(\omega_{v_iv_i}) \in \mathbb{R}^p_{>0}$, $\Lambda = (\lambda_{v_i,h_{v_i}}) \in \mathbb{R}^D$ and $\Phi = (\phi_{h_{v_i}h_{v_j}}) \in \mathbb{R}^B$, we have

$$\sigma_{v_iv_j} = \begin{cases} \lambda_{v_i,h_{v_i}}\lambda_{v_j,h_{v_j}} & \text{if } i \neq j \text{ and } \operatorname{jpa}(\{v_i,v_j\}) = \{h_{v_i}\} = \{h_{v_j}\} \neq \emptyset, \\ \lambda_{v_i,h_{v_i}}\lambda_{v_j,h_{v_j}}\phi_k & \text{if } i \in \{2k-1,2k\} \text{ and } j \in \{2k+1,2k+2\} \text{ and } k \in [m-1], \\ \omega_{v_iv_i}+\lambda_{v_i,h_{v_i}}^2 & \text{if } i = j, \\ 0 & \text{else.} \end{cases} \tag{19}$$

Similarly as in the previous chapters, we need to inspect the Jacobian matrix for dimension evaluation. Its structure is given by

$$J = \begin{array}{r} \\ v_i \\ \{v_i,v_j\} \end{array} \begin{array}{c} \begin{array}{ccc} \omega & \phi & \lambda \end{array} \\ \begin{pmatrix} I_p & 0 & * \\ 0 & C_1 & C_2 \end{pmatrix} \end{array}.$$

The upper rows correspond to the derivatives of $\sigma_{v_iv_i}$, while the lower rows indicate the derivatives of the entries $\sigma_{v_iv_j}$ for $i < j$. The unit matrix of rank p in the upper left matrix block results from

$$\frac{\sigma_{v_iv_i}}{\omega_{v_jv_j}} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else.} \end{cases}$$

Recalling Definition 2.1, we obtain that $\dim(F(G)) = p + \operatorname{rank}(C)$, where $C = (C_1|C_2)$ is the submatrix consisting of the two lower right matrix blocks, C_1, C_2, of the Jacobian. The entries of C_1 are given by

$$\frac{\sigma_{v_i v_j}}{\phi_k} = \begin{cases} \lambda_{v_i,h_{v_i}} \lambda_{v_j,h_{v_j}} & \text{if } i \in \{2k-1, 2k\} \text{ and } j \in \{2k+1, 2k+2\}, \\ 0 & \text{else.} \end{cases} \tag{20}$$

The entries of C_2 are given by

$$\frac{\sigma_{v_i v_j}}{\lambda_{v_z,h_{v_z}}} = \begin{cases} \lambda_{v_i,h_{v_i}} & \text{if } j = z \text{ and } \operatorname{jpa}(\{v_i, v_j\}) = \{h_{v_i}\} = \{h_{v_j}\} \neq \emptyset, \\ \lambda_{v_j,h_{v_j}} & \text{if } i = z \text{ and } \operatorname{jpa}(\{v_i, v_j\}) = \{h_{v_i}\} = \{h_{v_j}\} \neq \emptyset, \\ \lambda_{v_j,h_{v_j}} \phi_k & \text{if } i = z \text{ and } i \in \{2k-1, 2k\}, j \in \{2k+1, 2k+2\} \text{ and } k \in [m-1], \\ \lambda_{v_i,h_{v_i}} \phi_k & \text{if } j = z \text{ and } i \in \{2k-1, 2k\}, j \in \{2k+1, 2k+2\} \text{ and } k \in [m-1], \\ 0 & \text{else.} \end{cases} \tag{21}$$

The next theorem states that the model of any 2-chain graph has expected dimension.

Theorem 5.5. *Let $G = (V \cup \mathcal{H}, D, B)$ be a 2-chain graph. Then the corresponding factor analysis model has expected dimension.*

Proof. We begin by proving Theorem 5.5 for a 2-chain graph with $|\mathcal{H}| = 1$ separately. Therefore, let $G = (V \cup \mathcal{H}, D, B)$ be a 2-chain graph such that $|\mathcal{H}| = 1$. Its expected dimension equals three. Clearly, $B = \emptyset$ and $\Phi = (1) \in \mathbb{R}^{1\times 1}$. The matrix block C is given by

$$C \;=\; \{v_1, v_2\} \begin{pmatrix} \lambda_{v_1,h_1} & \lambda_{v_2,h_1} \\ \lambda_{v_2,h_1} & \lambda_{v_1,h_1} \end{pmatrix}$$

and has generic rank equal to one. Thus, $\dim(F(G)) = p + \operatorname{rank}(C) = 3$ and G has indeed expected dimension.

Now let G be a 2-chain graph with at least two latent nodes. In that case, expected dimension always equals $|V| + |D| + |B|$. Given that the upper left matrix block I_p of the Jacobian matrix has rank $|V| = p$, it remains to show, that the block C has generically rank $r = |D| + |B| = 3m - 1$. Applying a similar argument as in the proof of Theorem 3.3, it suffices to find a generic choice of parameters, $\Lambda^0 \in \mathbb{R}^D$ and $\Phi^0 \in \mathbb{R}^B$, for which the rank of $C^0 = C(\Lambda^0, \Phi^0)$ equals r. Recalling Definition 5.3, we can relabel the observed nodes, such that $P_k := \operatorname{ch}(h_k) = \{v_{2k-1}, v_{2k}\}$ for all $k \in [m]$. Note, that h_k refers to the same factor as $h_{v_{2k-1}}$ and $h_{v_{2k}}$ previously. Let us specify the tuples $\phi = (\phi_k)_{k\in[m-1]}$, $\lambda^- = (\lambda_{v_{2k-1},h_k})_{k\in[m]}$ and $\lambda^+ = (\lambda_{v_{2k},h_k})_{k\in[m]}$. Moreover, we define the sets $\mathcal{R}^- = \{\{v_{2k-1}, v_{2k+1}\} : k \in [m-1]\}$ and $\mathcal{R}^+ = \{\{v_{2k}, v_{2k+2}\} : k \in [m-1]\}$. We write $\mathcal{R}^{\complement} = C(V, 2) \setminus \Big(\big(\bigcup_{k\in[m]} P_k\big) \cup \mathcal{R}^+ \cup \mathcal{R}^- \cup \{v_1, v_4\}\Big)$. For the purpose of this proof, the submatrix C will be rearranged as

$$C = \begin{array}{c} \\ P_1 \\ P_2 \\ \vdots \\ P_m \\ \mathcal{R}^+ \\ \{v_1,v_4\} \\ \mathcal{R}^- \\ \mathcal{R}^{\complement} \end{array} \begin{pmatrix} \lambda^+ & & \phi & & \lambda^+ \\ \lambda_1^+ & & & & \lambda_1^- \\ & \lambda_2^+ & & & & \lambda_2^- \\ & & \ddots & & & & \ddots \\ & & & \lambda_m^+ & & & & \lambda_m^- \\ \hline & & & \lambda_1^+\lambda_2^+ & & & \lambda_2^+\phi_1 \;\; \lambda_1^+\phi_1 & & \\ & & & & \ddots & & & \ddots \;\; \ddots & \\ & & & & & \lambda_{m-1}^+\lambda_m^+ & & & \lambda_m^+\phi_{m-1} \;\; \lambda_{m-1}^+\phi_{m-1} \\ \hline \lambda_2^+\phi_1 & & & \lambda_1^-\lambda_2^+ & & & \lambda_1^-\phi_1 & & \\ \lambda_2^-\phi_1 & \lambda_1^-\phi_1 & & \lambda_1^-\lambda_2^- & & & & & \\ & \ddots & \ddots & & \ddots & & & & \\ & & \lambda_m^-\phi_{m-1} \;\; \lambda_{m-1}^-\phi_{m-1} & & & \lambda_{m-1}^-\lambda_m^- & & & \\ \hline & & * & & * & & & * & \end{pmatrix}, \tag{22}$$

where void entries are zero; recall Equation (20) and Equation (21). We choose to set all parameters in Λ^0 and Φ^0 equal to one. This immediately implies, that the submatrix of C^0 consisting of only the rows $T \in \left(\left(\bigcup_{k\in[m]} P_k\right) \cup \mathcal{R}^+\right)$ and the columns λ^-, ϕ is diagonal and of size $2m-1$. Therefore it must have full row rank of $2m-1$.

Claim 1: By row reduction, the submatrix of C^0 that consists of the rows indexed by $\mathcal{R}^-$ can be transformed such that the right block with columns indexed by λ^+ is of full row rank, and the left and middle blocks with columns indexed by λ^- and ϕ are zero.

Fix any row indexed by $\{v_{2k-1}, v_{2k+1}\} \in \mathcal{R}^-$. This row contains three potential nonzero entries given by $\lambda_{v_{2k+1},h_{k+1}}\phi_k$, $\lambda_{v_{2k-1},h_k}\phi_k$ and $\lambda_{v_{2k-1},h_k}\lambda_{v_{2k+1},h_{k+1}}$. They occur at the entries with column indices λ_{v_{2k-1},h_k}, $\lambda_{v_{2k+1},h_{k+1}}$ and ϕ_k. All three nonzero entries are equal to one, given our parameter choice. We eliminate those entries by subtracting the rows indexed by P_k, P_{k+1} and $\{v_{2k}, v_{2k+2}\} \in \mathcal{R}^+$. The relevant submatrix of C^0 is given by

$$\begin{array}{r} \\ P_k \\ P_{k+1} \\ \{v_{2k}, v_{2k+2}\} \\ \{v_{2k-1}, v_{2k+1}\} \end{array} \begin{array}{c} \begin{array}{ccccc} \lambda_{v_{2k-1},h_k} & \lambda_{v_{2k+1},h_{k+1}} & \phi_k & \lambda_{v_{2k},h_k} & \lambda_{v_{2k+2},h_{k+1}} \end{array} \\ \left(\begin{array}{ccc|cc} \lambda_{v_{2k},h_k} & 0 & 0 & \lambda_{v_{2k-1},h_k} & 0 \\ 0 & \lambda_{v_{2k+2},h_{k+1}} & 0 & 0 & \lambda_{v_{2k+1},h_{k+1}} \\ 0 & 0 & \lambda_{v_{2k},h_k}\lambda_{v_{2k+2},h_{k+1}} & \lambda_{v_{2k+2},h_{k+1}}\phi_k & \lambda_{v_{2k},h_k}\phi_k \\ \hline \lambda_{v_{2k+1},h_{k+1}}\phi_k & \lambda_{v_{2k-1},h_k}\phi_k & \lambda_{v_{2k-1},h_k}\lambda_{v_{2k+1},h_{k+1}} & 0 & 0 \end{array}\right) \end{array}.$$

Setting all parameters equal to one yields

$$\begin{array}{r} \\ P_k \\ P_{k+1} \\ \{v_{2k}, v_{2k+2}\} \\ \{v_{2k-1}, v_{2k+1}\} \end{array} \begin{array}{c} \begin{array}{ccccc} \lambda_{v_{2k-1},h_k} & \lambda_{v_{2k+1},h_{k+1}} & \phi_k & \lambda_{v_{2k},h_k} & \lambda_{v_{2k+2},h_{k+1}} \end{array} \\ \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 0 & 0 \end{array}\right) \end{array}.$$

Performing the row-reduction as previously described creates two fill-ins. One occurs in the entry with row index $\{v_{2k-1}, v_{2k+1}\}$ and column index λ_{v_{2k},h_k}. The other one is created in the entry with row index $\{v_{2k-1}, v_{2k+1}\}$ and column index $\lambda_{v_{2k+2},h_{k+1}}$. Both fill-ins are equal to -2. Thus, after elimination the relevant submatrix of C^0 is given by

$$\begin{array}{r} \\ P_i \\ P_{k+1} \\ \{v_{2k}, v_{2k+2}\} \\ \{v_{2k-1}, v_{2k+1}\} \end{array} \begin{array}{c} \begin{array}{ccccc} \lambda_{v_{2k-1},h_k} & \lambda_{v_{2k+1},h_{k+1}} & \phi_i & \lambda_{v_{2k},h_k} & \lambda_{v_{2k+2},h_{k+1}} \end{array} \\ \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ \hline 0 & 0 & 0 & -2 & -2 \end{array}\right) \end{array}.$$

We have shown that, after elimination, for an arbitrary row of $\mathcal{R}^-$, indexed by $\{v_{2k-1}, v_{2k+1}\}$, the entries indexed by the columns $\lambda_{v_{2k},h_k} \in \lambda^+$ and $\lambda_{v_{2k+2},h_{k+1}} \in \lambda^+$ are both -2. All remaining entries of the row equal zero. This implies that the submatrix of C^0 consisting of the rows in $\mathcal{R}^-$ is given by

$$\begin{array}{r} \\ \mathcal{R}^- \end{array} \begin{array}{c} \begin{array}{cc|c} \lambda^- & \phi & \overbrace{\lambda_{v_2,h_1} \quad \lambda_{v_4,h_2} \quad \dots \quad \lambda_{v_{2m},h_m}}^{\lambda^+} \end{array} \\ \left\{\begin{array}{r} \{v_1, v_3\} \\ \vdots \\ \{v_{2m-3}, v_{2m-1}\} \end{array}\right. \left(\begin{array}{cc|ccc} & & -2 & -2 & \\ & & & \ddots & \ddots \\ & & & & -2 \quad -2 \end{array}\right) \end{array}.$$

Clearly, the right block with column index λ^+ has full row rank $|\mathcal{R}^-| = m-1$. The left and middle block are zero. This proves Claim 1.

Claim 2: By row reduction, the row indexed by $\{v_1, v_4\}$ can be transformed, such that the submatrix of C^0 consisting of the rows indexed by $\mathcal{R}^-$ and $\{v_1, v_4\}$ has full row rank.

We want to subtract the rows indexed by P_1 and $\{v_2, v_4\} \in \mathcal{R}^+$ from the row indexed by $\{v_1, v_4\}$. Hence, the relevant part of C^0 is given by the submatrix

$$\begin{array}{r} \\ P_1 \\ \{v_2,v_4\} \\ \{v_1,v_4\} \end{array} \begin{array}{c} \begin{array}{cccc} \lambda_{v_1,h_1} & \phi_1 & \lambda_{v_2,h_1} & \lambda_{v_4,h_2} \end{array} \\ \left(\begin{array}{c|c|cc} \lambda_{v_2,h_1} & 0 & \lambda_{v_1,h_1} & 0 \\ 0 & \lambda_{v_2,h_1}\lambda_{v_4,h_2} & \lambda_{v_4,h_2}\phi_1 & \lambda_{v_2,h_1}\phi_1 \\ \hline \lambda_{v_4,h_2}\phi_1 & \lambda_{v_4,h_2}\lambda_{v_2,h_1} & 0 & \lambda_{v_1,h_1}\phi_1 \end{array}\right) \end{array}.$$

Setting all parameters to one and performing the row reductions yields

$$\begin{array}{r} \\ P_1 \\ \{v_2,v_4\} \\ \{v_1,v_4\} \end{array} \begin{array}{c} \begin{array}{cccc} \lambda_{v_1,h_1} & \phi_1 & \lambda_{v_2,h_1} & \lambda_{v_4,h_2} \end{array} \\ \left(\begin{array}{c|c|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ \hline 0 & 0 & -2 & 0 \end{array}\right) \end{array}.$$

We now use Claim 1 of this proof and conclude, that the submatrix of C^0 consisting of the rows indexed by $\mathcal{R}^-$ and $\{v_1, v_4\}$ has the following structure

$$\begin{array}{r} \\ \{v_1,v_4\} \\ \\ \mathcal{R}^- \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} \lambda^- & \phi & & \lambda^+ \end{array} \\ \left(\begin{array}{c|c|ccc} & -2 & & & \\ \hline & -2 & -2 & & \\ & & \ddots & \ddots & \\ & & & -2 & -2 \end{array}\right) \end{array}.$$

We can easily convince ourselves, that by subtracting from every row indexed by a 2-pair in $\mathcal{R}^-$ the row above, we obtain a diagonal matrix block

$$\begin{array}{r} \\ \{v_1,v_4\} \\ \\ \mathcal{R}^- \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} \lambda^- & \phi & & \lambda^+ \end{array} \\ \left(\begin{array}{c|c|ccc} & -2 & & & \\ \hline & & -2 & & \\ & & & \ddots & \\ & & & & -2 \end{array}\right) \end{array},$$

which is of full row rank m. Putting together Claim 1 and Claim 2, we see that after all eliminations, the submatrix of C^0 consisting of all rows indexed by a 2-pair $T \in \left(\left(\bigcup_{k\in[m]} P_k\right) \cup \mathcal{R}^+ \cup \{v_1, v_4\} \cup \mathcal{R}^-\right)$ has the form

$$\begin{array}{r} \\ P_1 \\ \vdots \\ P_m \\ \\ \mathcal{R}^+ \\ \\ \{v_1,v_4\} \\ \\ \mathcal{R}^- \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} \lambda^- & & \phi & & & \lambda^+ \end{array} \\ \left(\begin{array}{ccc|ccc|ccc} 1 & & & & & & 1 & & \\ & \ddots & & & & & & \ddots & \\ & & 1 & & & & & & 1 \\ \hline & & & 1 & & & 1 & 1 & \\ & & & & \ddots & & & \ddots & \ddots \\ & & & & & 1 & & & 1 \\ \hline & & & & & & -2 & & \\ \hline & & & & & & & -2 & \\ & & & & & & & & \ddots \\ & & & & & & & & -2 \end{array}\right) \end{array},$$

and is of full rank $r = m + (m-1) + 1 + (m-1) = 3m - 1$. Therefore, we found a particular choice of parameters for which the Jacobian matrix has rank $|V| + |D| + |B|$. This concludes the proof. □

Example 5.6. Let us consider the graph depicted in Figure 7 (b). We aim to illustrate the row reductions that were performed in the proof of Theorem 5.5. The 2-pairs of pure children are given by $P_1 = \{v_1, v_2\}$, $P_2 = \{v_3, v_4\}$ and $P_3 = \{v_5, v_6\}$. Furthermore, we have $\lambda^- = (\lambda_{v_1,h_1}, \lambda_{v_3,h_2}, \lambda_{v_5,h_3})$ and $\lambda^+ = (\lambda_{v_2,h_1}, \lambda_{v_4,h_2}, \lambda_{v_6,h_3})$, as well as $\mathcal{R}^- = \{\{v_1, v_3\}, \{v_3, v_5\}\}$ and $\mathcal{R}^+ = \{\{v_2, v_4\}, \{v_4, v_6\}\}$. The only 2-pairs in $\mathcal{R}^{\complement}$ that index a nonzero row are $\{\{v_2, v_3\}, \{v_3, v_6\}, \{v_4, v_5\}\} \subset \mathcal{R}^{\complement}$. Now, the matrix in (22) is given by

$$
C = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_3,v_4\} \\ \{v_5,v_6\} \\ \{v_2,v_4\} \\ \{v_4,v_6\} \\ \{v_1,v_4\} \\ \{v_1,v_3\} \\ \{v_3,v_5\} \\ \{v_2,v_3\} \\ \{v_3,v_6\} \\ \{v_4,v_5\} \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} \lambda_{v_1,h_1} & \lambda_{v_3,h_2} & \lambda_{v_5,h_3} & \phi_1 & \phi_2 & \lambda_{v_2,h_1} & \lambda_{v_4,h_2} & \lambda_{v_6,h_3} \end{array} \\
\left(\begin{array}{ccc|cc|ccc}
\lambda_{v_2,h_1} & & & & & \lambda_{v_1,h_1} & & \\
& \lambda_{v_4,h_2} & & & & & \lambda_{v_3,h_2} & \\
& & \lambda_{v_6,h_3} & & & & & \lambda_{v_5,h_3} \\
\hdashline
& & & \lambda_{v_2,h_1}\lambda_{v_4,h_2} & & \lambda_{v_4,h_2}\phi_1 & \lambda_{v_2,h_1}\phi_1 & \\
& & & & \lambda_{v_4,h_2}\lambda_{v_6,h_3} & & \lambda_{v_6,h_3}\phi_2 & \lambda_{v_4,h_2}\phi_2 \\
\hline
\lambda_{v_4,h_2}\phi_1 & & & \lambda_{v_1,h_1}\lambda_{v_4,h_2} & & & \lambda_{v_1,h_1}\phi_1 & \\
\hdashline
\lambda_{v_3,h_2}\phi_1 & \lambda_{v_1,h_1}\phi_1 & & \lambda_{v_1,h_1}\lambda_{v_3,h_2} & & & & \\
& \lambda_{v_5,h_3}\phi_2 & \lambda_{v_3,h_2}\phi_2 & & \lambda_{v_3,h_2}\lambda_{v_5,h_3} & & & \\
\hline
& \lambda_{v_2,h_1}\phi_1 & & \lambda_{v_2,h_1}\lambda_{v_3,h_2} & & \lambda_{v_3,h_2}\phi_1 & & \\
& \lambda_{v_6,h_3}\phi_2 & & & \lambda_{v_3,h_2}\lambda_{v_6,h_3} & & & \lambda_{v_3,h_2}\phi_2 \\
& & \lambda_{v_4,h_2}\phi_2 & & \lambda_{v_4,h_2}\lambda_{v_5,h_3} & & \lambda_{v_5,h_3}\phi_2 &
\end{array}\right)
\end{array},
$$

where void entries are zero. By setting all parameters to one, we obtain the matrix C^0 that has the form

$$
C^0 = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_3,v_4\} \\ \{v_5,v_6\} \\ \{v_2,v_4\} \\ \{v_4,v_6\} \\ \{v_1,v_4\} \\ \{v_1,v_3\} \\ \{v_3,v_5\} \\ \{v_2,v_3\} \\ \{v_3,v_6\} \\ \{v_4,v_5\} \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} \lambda_{v_1,h_1} & \lambda_{v_3,h_2} & \lambda_{v_5,h_3} & \phi_1 & \phi_2 & \lambda_{v_2,h_1} & \lambda_{v_4,h_2} & \lambda_{v_6,h_3} \end{array} \\
\left(\begin{array}{ccc|cc|ccc}
1 & & & & & 1 & & \\
& 1 & & & & & 1 & \\
& & 1 & & & & & 1 \\
\hdashline
& & & 1 & & 1 & 1 & \\
& & & & 1 & & 1 & 1 \\
\hline
1 & & & 1 & & & 1 & \\
\hdashline
1 & 1 & & 1 & & & & \\
& 1 & 1 & & 1 & & & \\
\hline
& 1 & & 1 & & 1 & & \\
& 1 & & & 1 & & & 1 \\
& & 1 & & 1 & & 1 &
\end{array}\right)
\end{array}.
$$

The submatrix of C^0 consisting of the columns λ^-, ϕ, and the rows $T \in \left(\left(\bigcup_{k\in[3]} P_k\right) \cup \mathcal{R}^+\right)$ is diagonal and has full row rank of five. This coincides with what was stated in the proof. We proceed to perform the row reduction described in Claim 1, which corresponds to eliminating all ones in the rows indexed by $\{v_1, v_3\} \in \mathcal{R}^-$ and $\{v_3, v_5\} \in \mathcal{R}^-$. We get

$$
C^0 = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_3,v_4\} \\ \{v_5,v_6\} \\ \{v_2,v_4\} \\ \{v_4,v_6\} \\ \{v_1,v_4\} \\ \{v_1,v_3\} \\ \{v_3,v_5\} \\ \{v_2,v_3\} \\ \{v_3,v_6\} \\ \{v_4,v_5\} \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} \lambda_{v_1,h_1} & \lambda_{v_3,h_2} & \lambda_{v_5,h_3} & \phi_1 & \phi_2 & \lambda_{v_2,h_1} & \lambda_{v_4,h_2} & \lambda_{v_6,h_3} \end{array} \\
\left(\begin{array}{ccc|cc|ccc}
1 & & & & & 1 & & \\
& 1 & & & & & 1 & \\
& & 1 & & & & & 1 \\
\hdashline
& & & 1 & & 1 & 1 & \\
& & & & 1 & & 1 & 1 \\
\hline
1 & & & 1 & & & 1 & \\
\hdashline
& & & & & -2 & -2 & \\
& & & & & & -2 & -2 \\
\hline
& 1 & & 1 & & 1 & & \\
& 1 & & & 1 & & & 1 \\
& & 1 & & 1 & & 1 &
\end{array}\right)
\end{array}.
$$

The entries of the row $\{v_1, v_3\} \in \mathcal{R}^-$ equal -2 if the column index is λ_{v_2,h_1} or λ_{v_4,h_2}, and zero otherwise. Similarly, the entries of the row with index $\{v_3, v_5\} \in \mathcal{R}^-$ are -2 if the column index is λ_{v_4,h_2} or λ_{v_6,h_3}, and zero otherwise. Therefore, the right block of the submatrix with rows in $\mathcal{R}^-$ has full row rank of two. The left and middle block of the submatrix are both zero and Claim 1 is satisfied. We proceed to eliminate the ones in the row indexed by $\{v_1, v_4\}$, which yields

$$
C^0 = \begin{array}{c} \\ \{v_1,v_2\} \\ \{v_3,v_4\} \\ \{v_5,v_6\} \\ \{v_2,v_4\} \\ \{v_4,v_6\} \\ \{v_1,v_4\} \\ \{v_1,v_3\} \\ \{v_3,v_5\} \\ \{v_2,v_3\} \\ \{v_3,v_6\} \\ \{v_4,v_5\} \end{array}
\begin{array}{c}
\begin{array}{cccccccc} \lambda_{v_1,h_1} & \lambda_{v_3,h_2} & \lambda_{v_5,h_3} & \phi_1 & \phi_2 & \lambda_{v_2,h_1} & \lambda_{v_4,h_2} & \lambda_{v_6,h_3} \end{array} \\
\left(\begin{array}{cccccccc}
1 & & & & & 1 & & \\
& 1 & & & & & 1 & \\
& & 1 & & & & & 1 \\
\hdashline
& & & 1 & & 1 & 1 & \\
& & & & 1 & & 1 & 1 \\
\hline
& & & & & -2 & & \\
\hdashline
& & & & & -2 & -2 & \\
& & & & & & -2 & -2 \\
\hline
& 1 & & 1 & & 1 & & \\
& 1 & & & 1 & & & 1 \\
& & 1 & & 1 & & 1 &
\end{array}\right)
\end{array}.
$$

Claim 2 is satisfied, given that the submatrix of C^0 consisting of the rows indexed by $\mathcal{R}^-$ and $\{v_1, v_4\}$ has full row rank of three. We see that the matrix consisting of all but the last three rows has full row rank of eight. To summarize, we found a generic choice of parameters $\Lambda^0 \in \mathbb{R}^D$ and $\Phi^0 \in \mathbb{R}^B$ that satisfies $\operatorname{rank}(C(\Lambda^0, \Phi^0)) \geq 8$. As the expected dimension of the model is 14, we get

$$14 \geq \dim(F(G)) = p + \operatorname{rank}(C) \geq p + 8 = 14.$$

We can confirm that the model has expected dimension.

6 Conclusion

Motivated by the challenges arising from sparsity, the aim of this thesis was to provide a detailed study of the dimension of sparse factor analysis models from both an algebraic and computational perspective.
We began by reviewing the necessary algebraic background and learned early on about the importance of the Jacobian matrix for the analysis. Moreover, we identified each model with its set of covariance matrices, and justified ways to bound the dimension of this set, building on recent results from Drton et al. (2024). The approach of Drton et al. (2024) for evaluating the upper bound can be applied to any arbitrary sparse factor analysis model and relies on finding a valid collection with maximal sum of cardinalities. On the other hand, deriving a lower bound, following the idea suggested by Drton et al. (2024), is only possible for models that satisfy the ZUTA condition and involves identifying a valid, ZUTA-compliant collection. In cases where the upper bound and the largest lower bound coincide, we get a dimension formula.
We implemented the concepts in R to enable computational evaluation of the bounds. A large part of this thesis consisted of a detailed analysis of the algorithms that were developed, including their complexity and efficiency. It was necessary to incorporate strategic code improvements to obtain results for relatively large graphs. These improvements comprised iterating backwards through the loops, adding break conditions, and generating only relevant collections. All these enhancements contributed to successfully examine the dimension properties of all sparse factor analysis graphs up to seven observed variables and up to three latent factors. We proceeded to run simulations and created an overview that summarized the number of graphs for which the bounds coincide, as well as the number of graphs with expected dimension, both grouped by the number of edges. In cases where the largest lower bound differed from the upper bound, we were initially uncertain whether the model had expected dimension or not. This gave rise to the necessity of outsourcing the task of dimension evaluation for certain models to MACAULAY2. We therefore briefly explained a core functionality of the computer algebra system and how it can be beneficial to our analysis. We then demonstrated how the knowledge gap regarding the dimension of models with varying bounds was closed using MACAULAY2 code. After finalizing the previously incomplete overview, we recognized that in certain cases, factor analysis models have expected dimension despite differing bounds. These cases were of particular interest as they accentuate weaknesses in the theorems and point out the need of alternative approaches that cover all graphs appropriately.
Another interesting result was the identification of the smallest graph, for which the corresponding bounds differed. Already for a graph with five observed and three latent nodes, the bounds failed to coincide. We thoroughly analyzed the reason for this deviation and learned that in the case of this particular graph, no valid, ZUTA-compliant collection could be determined with the same sum of cardinalities as the valid collection found for the upper bound.
In the last chapter, we extended our analysis of dimension to models with dependent factors. In that case, the relationship among the latent nodes is described by a latent covariance matrix that is not the identity matrix. As an example, we focused exclusively on 2-chain graphs. For these very particular graphs, we were able to prove that the corresponding models always have expected dimension. However, a general study of the dimension of sparse factor analysis models with dependent factors remains an open problem.

Bibliography

Bapat, R. B. (2012). *Linear algebra and linear models*, third edition. Springer-Verlag, London.

Bochnak, J., Coste, M., and Roy, M.-F. (1998). *Real algebraic geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*. Springer-Verlag, Berlin.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to algorithms*, third edition. MIT Press, Cambridge, MA.

Coste, M. (2002). *An Introduction to Semi-Algebraic Geometry*. Available at: `https://perso.univ-rennes1.fr/michel.coste/polyens/SAG.pdf`.

Drton, M., Grosdos, A., Portakal, I., and Sturma, N. (2024). Algebraic Sparse Factor Analysis. *arXiv preprint arXiv:2312.14762*.

Drton, M., Sturmfels, B., and Sullivant, S. (2007). Algebraic factor analysis: tetrads, pentads and beyond. *Probability Theory and Related Fields*, 138 (3-4): 463–493.

Evans, R. (2022). *rje: Miscellaneous Useful Functions for Statistics*. R package version 1.12.1.

Grayson, D. R. and Stillman, M. E. *Macaulay2, a software system for research in algebraic geometry*.

Huang, B., Low, C. J. H., Xie, F., Glymour, C., and Zhang, K. (2022). Latent Hierarchical Causal Structure Discovery with Rank Constraints. *arXiv preprint arXiv:2210.01798*.

Iwanaga, K., Umucu, E., Wu, J.-R., Yaghmaian, R., Lee, H.-L., Fitzgerald, S., and Chan, F. (2020). Assessing vocational outcome expectancy in individuals with serious mental illness: a factor-analytic approach. *Journal of Mental Health*, 29 (1): 116–123.

Jurczak, J. and Jurczak, G. (2021). Application of factor analysis in company management – selected examples related to competitiveness and market success. *Engineering Management in Production and Services*, 13 (3): 25–36.

Karpfinger, C. (2024). *Algebra: Gruppen – Ringe – Körper*, sixth edition. Springer Spektrum, Berlin.

Okamoto, M. (1973). Distinctness of the Eigenvalues of a Quadratic form in a Multivariate Sample. *The Annals of Statistics*, 1 (4): 763–765.

Petersen, R. J., Komorita, S. S., and Quay, H. C. (1964). Determinants of Sociometric Choices. *The Journal of Social Psychology*, 62 (1): 65–75.

Petrinovich, L. and Hardyck, C. (1964). Behavioral changes in parkinson patients following surgery: A factor analytic study. *Journal of Chronic Diseases*, 17 (3): 225–233.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, 15 (2): 201–293.

Sullivant, S. (2018). *Algebraic statistics*, volume 194 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.

Wood, J. (2024). *RcppAlgos: High Performance Tools for Combinatorics and Computational Mathematics*. R package version 2.8.5.