Identifiability in Sparse Factor Analysis

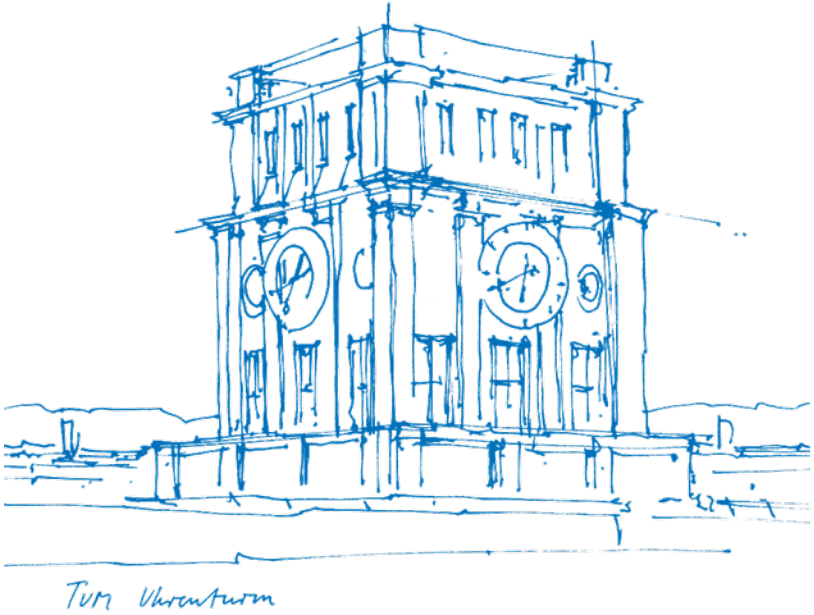
Nils Sturma

Research group Mathematical Statistics

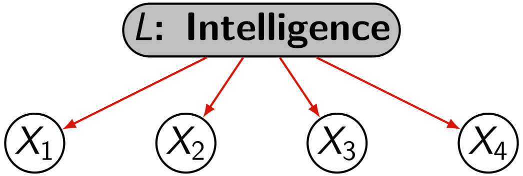
TUM School of Computation, Information and Technology

Technical University of Munich

(joint work with Mathias Drton, Miriam Kranzlmüller, Irem Portakal)



Spearman's One-Factor Model (1904)



Linear structural equations:

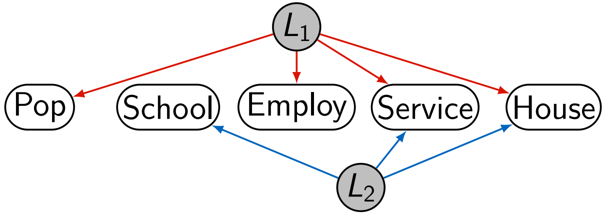
$$\begin{aligned} X_1 &= \lambda_{10} + \lambda_{1L} L + \varepsilon_1, \\ X_2 &= \lambda_{20} + \lambda_{2L} L + \varepsilon_2, \\ X_3 &= \lambda_{30} + \lambda_{3L} L + \varepsilon_3, \\ X_4 &= \lambda_{40} + \lambda_{4L} L + \varepsilon_4. \end{aligned}$$

Jointly independent errors: $\varepsilon_1, \dots, \varepsilon_4$.

$\mathrm{Var}[\varepsilon_v] = \omega_{vv} < \infty$, $\mathrm{Var}[L] = 1$.

Topic: Can we recover the "factor loadings" λ_{vL} and the "error variances" ω_{vv} from $\Sigma = \mathrm{Var}[X]$?

Socio-economic Example from Harmann (1976), Modern Factor Analysis



$$X = \Lambda L + \varepsilon, \quad \text{where } \Lambda = \begin{pmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{22} \\ \lambda_{31} & 0 \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \end{pmatrix}.$$

Observed covariance matrix (when $\mathrm{Var}[L] = I$):

$$\Sigma = \Lambda\Lambda^\top + \Omega = \begin{pmatrix} \omega_{11} + \lambda_{11}^2 & 0 & \boxed{\lambda_{11}\lambda_{31}} & \boxed{\lambda_{11}\lambda_{41}} & \lambda_{11}\lambda_{51} \\ 0 & \omega_{22} + \lambda_{22}^2 & 0 & \lambda_{22}\lambda_{42} & \lambda_{22}\lambda_{52} \\ \lambda_{11}\lambda_{31} & 0 & \omega_{33} + \lambda_{31}^2 & \boxed{\lambda_{31}\lambda_{41}} & \lambda_{31}\lambda_{51} \\ \lambda_{11}\lambda_{41} & \lambda_{22}\lambda_{42} & \lambda_{31}\lambda_{41} & \omega_{44} + \lambda_{41}^2 + \lambda_{52}^2 & \lambda_{41}\lambda_{51} + \lambda_{42}\lambda_{52} \\ \lambda_{11}\lambda_{51} & \lambda_{22}\lambda_{52} & \lambda_{31}\lambda_{51} & \lambda_{41}\lambda_{51} + \lambda_{42}\lambda_{52} & \omega_{55} + \lambda_{51}^2 + \lambda_{52}^2 \end{pmatrix}.$$

We see that

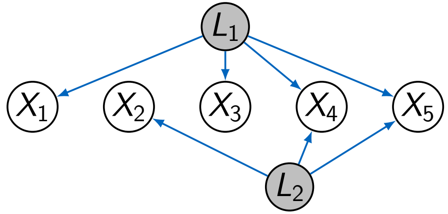
1) $\sqrt{\dfrac{\sigma_{13}\sigma_{14}}{\sigma_{34}}} = \sqrt{\dfrac{\lambda_{11}\lambda_{31}\,\lambda_{11}\lambda_{41}}{\lambda_{31}\lambda_{41}}} = \sqrt{\lambda_{11}^2} = a_1\lambda_{11}$ with $a_1 \in \{\pm 1\}$ and $\sigma_{34} = \lambda_{31}\lambda_{41} \neq 0$ 'almost surely',

2) $\dfrac{\sigma_{13}}{\sqrt{\sigma_{13}\sigma_{14}/\sigma_{34}}} = \dfrac{\lambda_{11}\lambda_{31}}{a_1\lambda_{11}} = a_1\lambda_{31}$ with $\lambda_{11} \neq 0$ 'almost surely'.

$\Longrightarrow$ Can identify $\Lambda_{\mathrm{ch}(L_1),L_1}$ up to column-sign, similarly $\Lambda_{\mathrm{ch}(L_2),L_2}$.

Factor Analysis vs. Causal Representation Learning

Sparse Factor Analysis

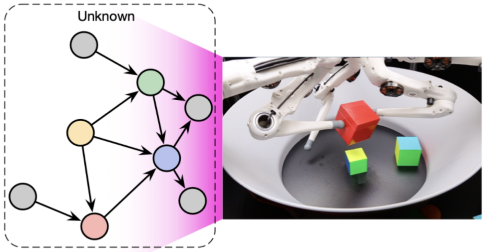


Model:

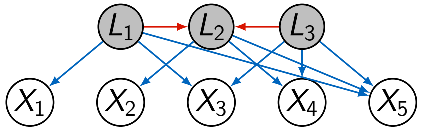
$$X = \Lambda L + \varepsilon,$$

where Λ sparse.

Causal Representation Learning



[Schölkopf et al. (2021)]



Model:

$$X = GL + \varepsilon_X,$$
$$L = ML + \varepsilon_L,$$

where G and M are sparse.

$$\Longrightarrow \quad X = \underbrace{G(I - M)^{-1}}_{=:\Lambda} \varepsilon_L + \varepsilon_X$$

Understanding sparse factor analysis is key for causal representation learning!

Setup

Variables:

Observed: $X = (X_v)_{v \in V}$ Latent: $L = (L_h)_{h \in \mathcal{H}}$

Graph:

Bipartite directed graph $G = (V \dot{\cup} \mathcal{H}, D)$, that is, $D \subseteq \mathcal{H} \times V$.

Sparse factor analysis model:

$$X = \Lambda L + \varepsilon$$

– all latent factors and error terms in (L, ε) are mutually independent, so $\Omega = \text{Var}[\varepsilon] = \text{diag}(\omega_{vv} : v \in V)$ diagonal, and $\text{Var}[L] = I$.

– parameter matrix Λ is sparse and supported over edge set D (write $\Lambda \in \mathbb{R}^D$).

Content of the Talk

Definition

Every factor analysis graph G yields a parametrization of the observed covariance matrix:

$$\tau_G : (\Lambda, \Omega) \longmapsto \Sigma \equiv \Lambda\Lambda^\top + \Omega.$$

Fiber: $\mathcal{F}_G(\Omega, \Lambda) = \{(\widetilde{\Omega}, \tilde{\Lambda}) : \tau_G(\widetilde{\Omega}, \tilde{\Lambda}) = \tau_G(\Omega, \Lambda)\}.$

The model given by G is generically sign-identifiable if

$$\mathcal{F}_G(\Omega, \Lambda) = \{(\widetilde{\Omega}, \tilde{\Lambda}) : \widetilde{\Omega} = \Omega \text{ and } \tilde{\Lambda} = \Lambda\Psi \text{ for } \Psi \in \{\pm 1\}^{|\mathcal{H}|\times|\mathcal{H}|} \text{ diagonal}\} \qquad \text{for 'almost all' } (\Lambda, \Omega).$$

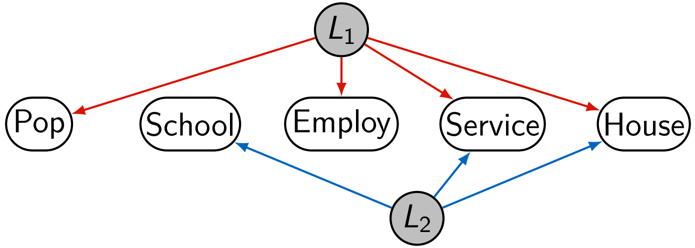
Main Contribution:

- Sufficient graphical condition for generic sign-identifiability.
- Recursive polynomial time algorithm.

  (caveat: polynomial time when bounding a cardinality in a search step)

Gröbner basis computations solve the problem ... on a small scale.

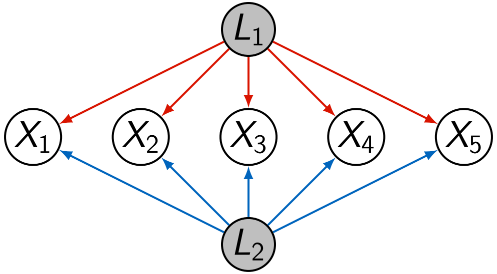
Our Software




```
# Define graph
> L = matrix(c(1, 0,
+              0, 1,
+              1, 0,
+              1, 1,
+              1, 1), 5, 2, byrow=TRUE)
> g = FactorGraph(L)
>
> # Check identifiability
> res = mID(g)
Generic Sign-Identifiability Summary
# nr. of latent nodes that are gen. sign-identifiable: 2
# gen. sign-identifiable nodes: 1, 2
```

Available at `https://github.com/MiriamKranzlmueller/id-factor-analysis`.

Rotational Indeterminacy



- No restriction on Λ:

$$\Lambda\Lambda^\top + \Omega = (\Lambda Q)(Q^\top \Lambda^\top) + \Omega \quad \text{for all } Q \text{ orthogonal.}$$

- Literature: Focus on identifiability of Ω and $\Lambda\Lambda^\top$ in full factor analysis models.

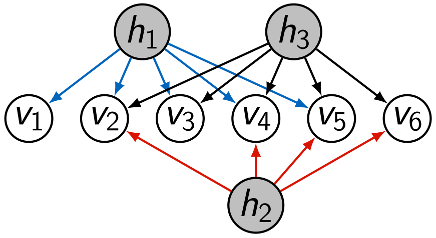
- However, if Λ is sparse, we might get generic sign-identifiability.

Zero Upper Triangular Assumption

Definition

The graph G satisfies the Zero Upper Triangular Assumption (ZUTA) if there exists an ordering $\prec$ on the latent nodes $\mathcal{H}$ such that $\mathrm{ch}(h)$ is not contained in $\cup_{\ell \succ h} \mathrm{ch}(\ell)$ for all $h \in \mathcal{H}$.

Example



$$
\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{array}
\begin{array}{c} \begin{matrix} h_1 & h_2 & h_3 \end{matrix} \\ \begin{pmatrix} * & 0 & 0 \\ * & * & * \\ * & 0 & * \\ * & * & * \\ * & * & * \\ 0 & * & * \end{pmatrix} \end{array}
\overset{\text{ZUTA}}{\rightsquigarrow}
\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{array}
\begin{array}{c} \begin{matrix} h_1 & \color{orange}{h_3} & \color{orange}{h_2} \end{matrix} \\ \begin{pmatrix} * & 0 & 0 \\ * & * & * \\ * & * & 0 \\ * & * & * \\ * & * & * \\ 0 & * & * \end{pmatrix} \end{array}
\underset{\text{upper-diag}\neq 0}{\overset{\text{upper-tri}=0}{\rightsquigarrow}}
\begin{array}{c} \\ v_1 \\ \color{orange}{v_3} \\ \color{orange}{v_2} \\ v_4 \\ v_5 \\ v_6 \end{array}
\begin{array}{c} \begin{matrix} h_1 & h_3 & h_2 \end{matrix} \\ \begin{pmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \\ * & * & * \\ * & * & * \\ 0 & * & * \end{pmatrix} \end{array}
$$

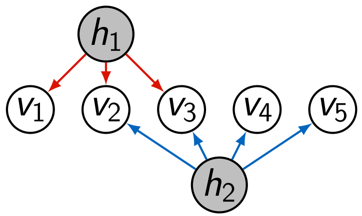
ZUTA $=$ "can permute cols and rows such that the upper right triangle of Λ is zero"
and w.l.o.g. "diagonal entries are nonzero".

Anderson-Rubin Criterion

Theorem [Anderson, Rubin (1956)]

A factor analysis graph $G = (V \cup \mathcal{H}, D)$ that satisfies ZUTA is generically sign-identifiable if for any deleted row of the symbolic matrix $\Lambda = (\lambda_{vh}) \in \mathbb{R}^D$ there exist two disjoint submatrices that are generically of rank $|\mathcal{H}|$.

Example



$$\Lambda = \begin{pmatrix} \lambda_{v_1 h_1} & 0 \\ \lambda_{v_2 h_1} & \lambda_{v_2 h_2} \\ \lambda_{v_3 h_1} & \lambda_{v_3 h_2} \\ 0 & \lambda_{v_4 h_2} \\ 0 & \lambda_{v_5 h_2} \end{pmatrix}$$

<u>Observation:</u>

Need $|V| \geq 2|\mathcal{H}| + 1$.

Examples for Inconclusiveness of Anderson-Rubin

1)

$$\begin{pmatrix} \lambda_{v_1h_1} & 0 & 0 \\ \lambda_{v_2h_1} & \lambda_{v_2h_2} & 0 \\ 0 & \lambda_{v_3h_2} & \lambda_{v_3h_3} \\ \lambda_{v_4h_1} & 0 & \lambda_{v_4h_3} \\ 0 & \lambda_{v_5h_2} & 0 \\ 0 & 0 & \lambda_{v_6h_3} \end{pmatrix}$$

[Hosszejni, Frühwirth-Schnatter (2022)]

2)

$$\begin{pmatrix} \lambda_{v_1h_1} & 0 & 0 \\ \lambda_{v_2h_1} & \lambda_{v_2h_2} & 0 \\ \lambda_{v_3h_1} & \lambda_{v_3h_2} & \lambda_{v_3h_3} \\ \lambda_{v_4h_1} & \lambda_{v_4h_2} & \lambda_{v_4h_3} \\ \lambda_{v_5h_1} & \lambda_{v_5h_2} & \lambda_{v_5h_3} \\ \lambda_{v_6h_1} & \lambda_{v_6h_2} & 0 \end{pmatrix}$$

3)

$$\begin{pmatrix} \lambda_{v_1h_1} & 0 & 0 & 0 \\ \lambda_{v_2h_1} & \lambda_{v_2h_2} & 0 & 0 \\ 0 & \lambda_{v_3h_2} & \lambda_{v_3h_3} & 0 \\ \lambda_{v_4h_1} & 0 & \lambda_{v_4h_3} & \lambda_{v_4h_4} \\ 0 & \lambda_{v_5h_2} & 0 & 0 \\ 0 & 0 & \lambda_{v_6h_3} & 0 \\ 0 & 0 & 0 & \lambda_{v_7h_4} \\ 0 & 0 & 0 & \lambda_{v_8h_4} \\ 0 & 0 & 0 & \lambda_{v_9h_4} \end{pmatrix}$$

Investigation of Anderson-Rubin

1) Anderson-Rubin: Fix $v \in V$. Find disjoint $U, W \subseteq V \setminus \{v\}$ with $|U| = |W| = |\mathcal{H}|$ such that

$$\det(\Lambda_{U,\mathcal{H}}) \neq 0 \quad \text{and} \quad \det(\Lambda_{W,\mathcal{H}}) \neq 0 \quad \text{(not the zero polynomial)}$$
$$\iff \det([\Lambda\Lambda^\top]_{U,W}) \neq 0.$$

2) Consider the matrix with exactly one diagonal entry:

$$[\Lambda\Lambda^\top]_{\{v\}\cup U,\{v\}\cup W} = \left(\begin{array}{c|c} [\Lambda\Lambda^\top]_{vv} & [\Lambda\Lambda^\top]_{v,W} \\ \hline [\Lambda\Lambda^\top]_{U,v} & [\Lambda\Lambda^\top]_{W,U} \end{array}\right) = \left(\begin{array}{c|c} [\Lambda\Lambda^\top]_{vv} & \Sigma_{v,W} \\ \hline \Sigma_{U,v} & \Sigma_{U,W} \end{array}\right).$$

3) Solve for diagonal entry $[\Lambda\Lambda^\top]_{vv}$ by Laplace expansion:

$$0 = \det([\Lambda\Lambda^\top]_{\{v\}\cup U,\{v\}\cup W}) = [\Lambda\Lambda^\top]_{vv} \underbrace{\det(\Sigma_{U,W})}_{\neq 0} - \sum_{w\in W} \text{sign}(w)\sigma_{vw} \det(\Sigma_{U,\{v\}\cup W\setminus\{w\}}).$$

4) Conclude: Solving for diagonal entries of $\Lambda\Lambda^\top$ is equivalent to solving for Ω.

$\implies$ Equivalent to solving for $\Sigma - \Omega = \Lambda\Lambda^\top$.

$\overset{\text{ZUTA}}{\implies}$ generic sign-identifiability (uniqueness of Cholesky decomposition).

Main Idea for Sparse Setup

Local approach: Can also choose U, W such that $|W| = |U| < |\mathcal{H}|$.

$\rightsquigarrow$ Ensure that $\det([\Lambda\Lambda^\top]_{U,W}) \neq 0$ and that $\det([\Lambda\Lambda^\top]_{\{v\}\cup U,\{v\}\cup W}) = 0$.

Characterization: When is $\det([\Lambda\Lambda^\top]_{A,B}) = 0$ if Λ is sparse?

$\rightsquigarrow$ Intersection-free matchings.

Definition

System of paths $\Pi = \{\pi_1, \dots, \pi_k\}$ is matching of $A = \{a_1, \dots, a_k\} \subseteq V$ and $B = \{b_1, \dots, b_k\} \subseteq V$ if

$$\pi_i = a_i \leftarrow h_i \rightarrow b_i.$$

A matching is intersection-free if all latent nodes h_i are distinct.

Lemma

For two subsets $A, B \subseteq V$ with $|A| = |B|$ it holds that $\det([\Lambda\Lambda^\top]_{A,B}) \neq 0$ if and only if there is an intersection-free matching of A and B.

Matching Criterion

Definition

Fix a latent node $h \in \mathcal{H}$. Tuple $(v, U, W, S) \in V \times 2^V \times 2^V \times 2^{\mathcal{H}\setminus\{h\}}$ satisfies the matching criterion for h if

(i) $\mathrm{pa}(v) \setminus S = \{h\}$ and $v \notin U \cup W$,

(ii) U and W are disjoint, nonempty sets of equal cardinality,

(iii) there exists an intersection-free matching of U and W that avoids S,

(iv) there does not exist an intersection-free matching of $\{v\} \cup W$ and $\{v\} \cup U$ that avoids S.

S = "solved nodes".

By (iii), $\det([\Lambda\Lambda^\top]_{U,W}) \neq 0$.

By (iv), $\det([\Lambda\Lambda^\top]_{\{v\}\cup U,\{v\}\cup W}) = 0$.

Algorithm: Recursive Solving

Theorem (M-identifiability)

If the tuple (v, U, W, S) satisfies the matching criterion with respect to h <u>*and*</u> *all nodes $\ell \in S$ are "solved before", then we can "solve" for h.*

That is, $(\widetilde{\Omega}, \widetilde{\Lambda}) \in \mathcal{F}_G(\Omega, \Lambda) \implies \widetilde{\Lambda}_{\mathrm{ch}(h),h} = \pm\Lambda_{\mathrm{ch}(h),h}$.

Algorithm (`M-ID`)

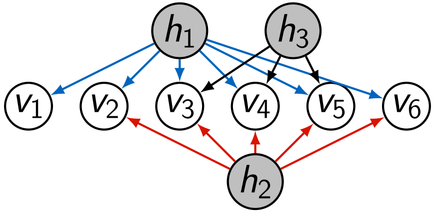
- Cycle through latent nodes h and search for tuples (v, U, W, S).
- Network-flow setup finds suitable tuples in polynomial time under a bound on $|U| = |W|$.

  Conjecture: *If we do not bound the cardinality $|U| = |W|$, then `M-ID` is NP-complete.*

Remarks

- Subsumes Anderson-Rubin.
- Together with an extension, subsumes anything we know (e.g. Bekker and ten Berge, 1997).

Example



$$\Lambda = \begin{pmatrix} \lambda_{v_1h_1} & 0 & 0 \\ \lambda_{v_2h_1} & \lambda_{v_2h_2} & 0 \\ \lambda_{v_3h_1} & \lambda_{v_3h_2} & \lambda_{v_3h_3} \\ \lambda_{v_4h_1} & \lambda_{v_4h_2} & \lambda_{v_4h_3} \\ \lambda_{v_5h_1} & \lambda_{v_5h_2} & \lambda_{v_5h_3} \\ \lambda_{v_6h_1} & \lambda_{v_6h_2} & 0 \end{pmatrix}$$

h_1: Take $v = v_1$, $S = \emptyset$ and $U = \{v_2, v_6\}$, $W = \{v_3, v_4\}$.

(iii) $v_2 \leftarrow h_1 \rightarrow v_3$, $v_6 \leftarrow h_2 \rightarrow v_4$; (iv) $\mathrm{pa}(\{v\} \cup U) \cap \mathrm{pa}(\{v\} \cup W) = \{h_1, h_2\}$.

h_2: Take $v = v_2$, $S = \{h_1\}$ and $U = \{v_3\}$, $W = \{v_6\}$.

(iii) $v_3 \leftarrow h_2 \rightarrow v_6$; (iv) $(\mathrm{pa}(\{v\} \cup U) \cap \mathrm{pa}(\{v\} \cup W)) \setminus S = \{h_2\}$.

h_3: Take $v = v_3$, $S = \{h_1, h_2\}$ and $U = \{v_4\}$, $W = \{v_5\}$.

(iii) $v_4 \leftarrow h_3 \rightarrow v_5$; (iv) $(\mathrm{pa}(\{v\} \cup U) \cap \mathrm{pa}(\{v\} \cup W)) \setminus S = \{h_3\}$.

Conclusion

- Sparse factor analysis models are the "building block" for many latent variable models.
- Latent variable models generally feature complicated parametrizations.
- Even in "simple" factor analysis models, there is still lots to explore ...
- Papers:

Sturma, Kranzlmueller, Portakal, Drton (2025).
Matching Criterion for Identifiability in Sparse Factor Analysis.
arXiv preprint arXiv:2502.02986.

Drton, Grosdos, Portakal, Sturma (2023).
Algebraic Sparse Factor Analysis.
arXiv preprint arXiv:2312.14762.

European Research Council
Established by the European Commission