Matching Criterion for Identifiability in Sparse Factor Analysis

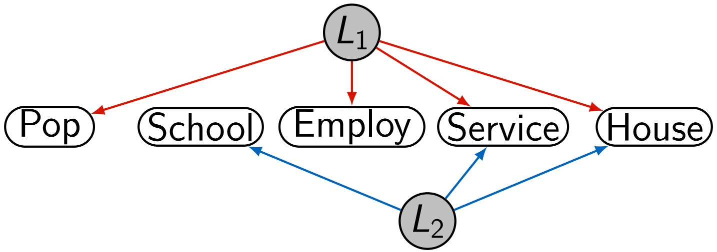
Nils Sturma

Insitute of Mathematics

EPFL

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

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



Linear structural equations:

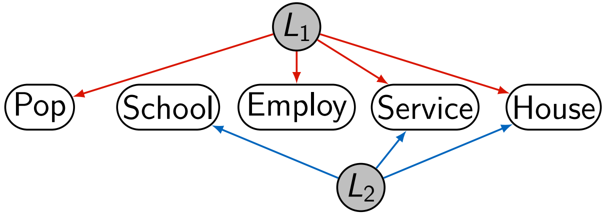
$$\begin{aligned}
X_1 &= \lambda_{10} + \textcolor{red}{\lambda_{11}} L_1 & & + \varepsilon_1, \\
X_2 &= \lambda_{20} & + \textcolor{blue}{\lambda_{22}} L_2 & + \varepsilon_2, \\
X_3 &= \lambda_{30} + \textcolor{red}{\lambda_{31}} L_1 & & + \varepsilon_3, \\
X_4 &= \lambda_{40} + \textcolor{red}{\lambda_{41}} L_1 & + \textcolor{blue}{\lambda_{42}} L_2 & + \varepsilon_4. \\
X_5 &= \lambda_{50} + \textcolor{red}{\lambda_{51}} L_1 & + \textcolor{blue}{\lambda_{52}} L_2 & + \varepsilon_5.
\end{aligned}$$

Jointly independent errors: $\varepsilon_1, \dots, \varepsilon_5$.

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

Topic: Can we recover the "factor loadings" $\textcolor{red}{\lambda_{v1}}, \textcolor{blue}{\lambda_{v2}}$ and the "error variances" ω_{vv} from $\Sigma = \mathrm{Var}[X]$?

Example: Identifiability of Factor Loadings



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

Observed covariance matrix:

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

We see that

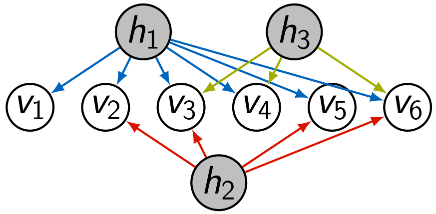
1) $\sqrt{\dfrac{\sigma_{13}\sigma_{14}}{\sigma_{34}}} = \sqrt{\dfrac{\lambda_{11}\lambda_{31}\,\lambda_{11}\lambda_{41}}{\lambda_{31}\lambda_{41}}} = \sqrt{\lambda_{11}^2} = a_1\lambda_{11}$ with $a_1 \in \{\pm 1\}$ and $\sigma_{34} = \lambda_{31}\lambda_{41} \neq 0$ 'almost surely',

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 $\Lambda_{\mathrm{ch}(L_2),L_2}$.

Identifiability of Factor Loadings

Setup



- $X = (X_v)_{v \in V}$ observed and $L = (L_h)_{h \in \mathcal{H}}$ latent variables.
- **Bipartite** Graph $G = (V \dot\cup \mathcal{H}, D)$ with $D \subseteq \mathcal{H} \times V$.
- Model: $X = \Lambda L + \varepsilon$ with Λ **sparse** according to G.

Identifiability Problem

Given a graph G and the covariance matrix $\Sigma = \Lambda\Lambda^\top + \Omega$, can we solve for Λ? Up to column-sign?

Main Contributions

- Necessary conditions: Algebraic Sparse Factor Analysis [SIAGA, 2025].
- Sufficient condition: Matching Criterion for Identifiability in Sparse Factor Analysis [arXiv:2502.02986].

  Combinatorial condition based on graph structure, efficient algorithm.

Key Idea

Determinantal approach: Fix $v \in V$. Find disjoint $U, W \subseteq V \setminus \{v\}$ with $|W| = |U|$ and ensure that

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

Solve for $[\Lambda\Lambda^\top]_{vv}$: Observe that

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

It follows by Laplace expansion:

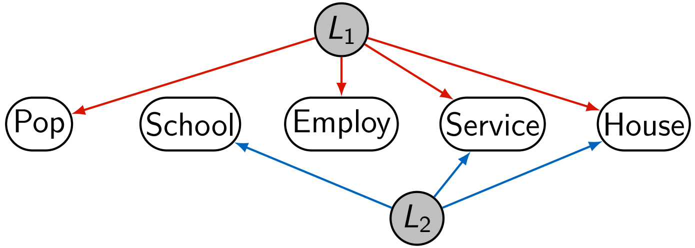
$$0 = \det([\Lambda\Lambda^\top]_{\{v\}\cup U,\{v\}\cup W}) = [\Lambda\Lambda^\top]_{vv} \underbrace{\det(\Sigma_{U,W})}_{\neq 0} - \sum_{w \in W} \operatorname{sign}(w)\sigma_{vw} \det(\Sigma_{U,\{v\}\cup W\setminus\{w\}}).$$

Solve column-wise: If $[\Lambda\Lambda^\top]_{vv} = (\Lambda_{v,h})^2$, then we recovered $\Lambda_{v,h}$ up to sign ... other entries "easy".

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

$\rightsquigarrow$ "Intersection-free matchings."

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.