Title: Half-Trek Criterion for Identifiability of Latent Variable Models

Speakers: Mathias Drton

Collection: Causal Inference & Quantum Foundations Workshop

Date: April 21, 2023 - 9:30 AM

URL: https://pirsa.org/23040124

Abstract: "Linear structural equation models relate random variables of interest via a linear equation system that features stochastic noise. The models are naturally represented by directed graphs whose edges indicate non-zero coefficients in the linear equations. In this talk I will report on progress on combinatorial conditions for parameter identifiability in models with latent (i.e., unobserved) variables. Identifiability holds if the coefficients associated with the edges of the graph can be uniquely recovered from the covariance matrix they define.

Paper:
https://doi.org/10.1214/22-AOS2221 or
https://arxiv.org/abs/2201.04457"

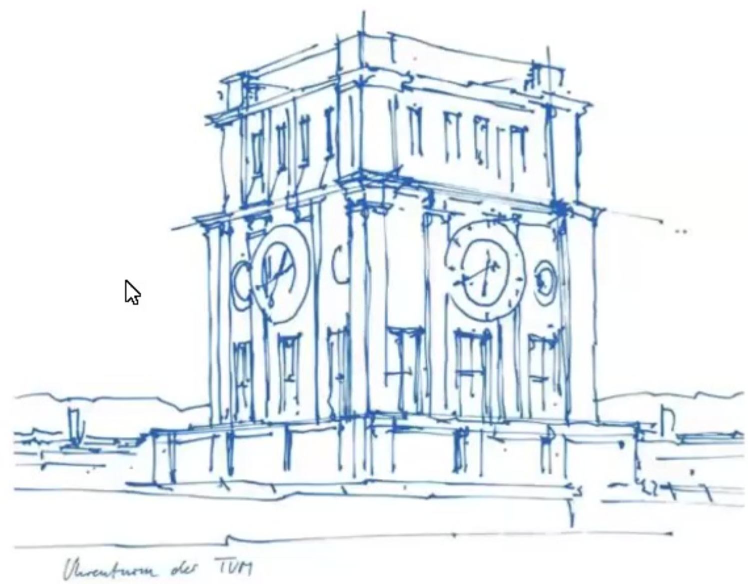
Half-Trek Criterion for Identifiability of Latent Variable Models

(actually something else seems to fit better)

Mathias Drton

Department of Mathematics

Technical University of Munich



The Annals of Statistics
2022, Vol. 50, No. 6, 3174–3196
https://doi.org/10.1214/22-AOS2221

HALF-TREK CRITERION FOR IDENTIFIABILITY OF LATENT VARIABLE MODELS

BY RINA FOYGEL BARBER[1,a], MATHIAS DRTON[2,b], NILS STURMA[2,c] AND LUCA WEIHS[3,d]

[1]*Department of Statistics, University of Chicago,* [a]*rina@uchicago.edu*
[2]*School of Computation, Information and Technology, Technical University of Munich,* [b]*mathias.drton@tum.de,* [c]*nils.sturma@tum.de*
[3]*Allen Institute for AI,* [d]*lucaw@allenai.org*

We consider linear structural equation models with latent variables and develop a criterion to certify whether the direct causal effects between the observable variables are identifiable based on the observed covariance matrix. Linear structural equation models assume that both observed and latent variables solve a linear equation system featuring stochastic noise terms. Each model corresponds to a directed graph whose edges represent the direct effects that appear as coefficients in the equation system. Prior research has developed a variety of methods to decide identifiability of direct effects in a latent projection framework, in which the confounding effects of the latent variables are represented by correlation among noise terms. This approach is effective when the confounding is sparse and effects only small subsets of the observed variables. In contrast, the new latent-factor half-trek criterion (LF-HTC) we develop in this paper operates on the original unprojected latent variable model and is able to certify identifiability in settings, where some latent variables may also have dense effects on many or even all of the observables. Our LF-HTC is an effective sufficient criterion for rational identifiability, under which the direct effects can be uniquely recovered as rational functions of the joint covariance matrix of the observed random variables. When restricting the search steps in LF-HTC to consider subsets of latent variables of bounded size, the criterion can be verified in time that is polynomial in the size of the graph.

Linear Structural Equation/Causal Models

Each model is induced by a directed graph:

Linear structural equations:

$$
\begin{aligned}
X_1 &= \lambda_{01} & & & + \varepsilon_1, \\
X_2 &= \lambda_{02} + \lambda_{12} X_1 & + \gamma_2 L_1 & & + \varepsilon_2, \\
X_3 &= \lambda_{03} + \lambda_{23} X_2 & + \gamma_3 L_1 & & + \varepsilon_3, \\
L_1 &= \lambda_{0u} & & & + \varepsilon_\ell.
\end{aligned}
$$

Independent errors:

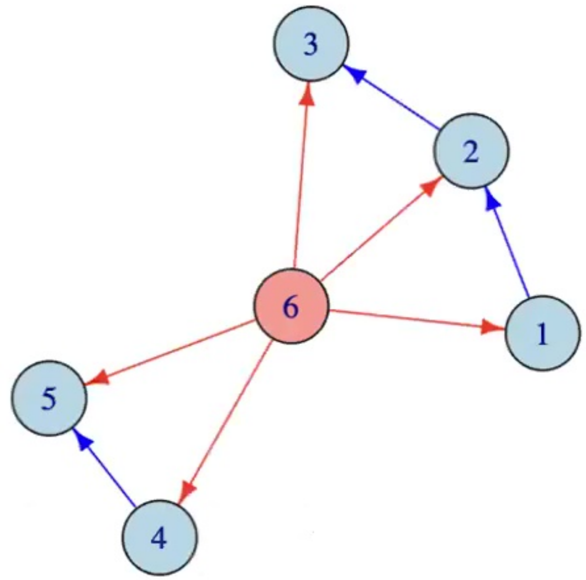
$$\varepsilon_1 \perp\!\!\!\perp \varepsilon_2 \perp\!\!\!\perp \varepsilon_3 \perp\!\!\!\perp \varepsilon_\ell$$

$$\mathrm{Var}[\varepsilon_v] = \omega_v < \infty$$

Topic of the talk: If L_1 is latent, can we recover the direct effects $(\lambda_{12}, \lambda_{23})$ from $\Sigma = \mathrm{Var}[X]$?

Our Software: SEMID (R Package)

```
# Define graph
> Lambda = matrix(c(0, 1, 0, 0, 0, 0,
+                   0, 0, 1, 0, 0, 0,
+                   0, 0, 0, 0, 0, 0,
+                   0, 0, 0, 0, 1, 0,
+                   0, 0, 0, 0, 0, 0,
+                   1, 1, 1, 1, 1, 0),
+                 6, 6, byrow=TRUE)
> observedNodes = seq(1,5)
> latentNodes = c(6)
> g = LatentDigraph(Lambda, observedNodes, latentNodes)
> plot(g)
```

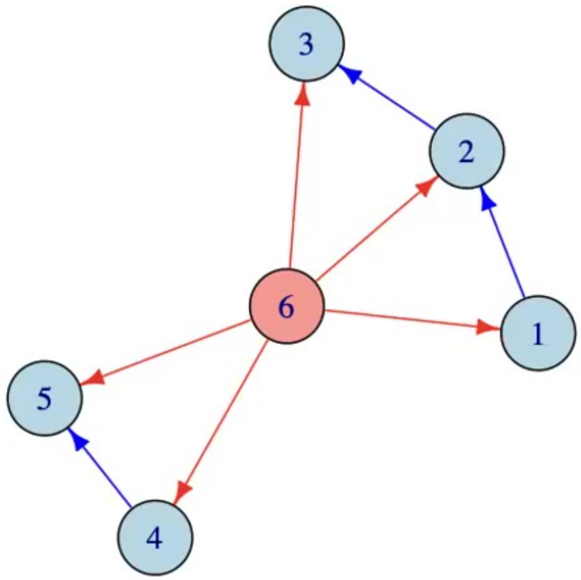


Our Software: SEMID (R Package)

```
> # Check identifiability
> res = lfhtcID(g)
> res
Call: lfhtcID(graph = g)

Latent Digraph Info
# observed nodes: 5
# latent nodes: 1
# total nr. of edges between observed nodes: 3

Generic Identifiability Summary
# nr. of edges between observed nodes shown gen. identifiable: 3
# gen. identifiable edges: 1->2, 2->3, 4->5
```



Mathias Drton (TUM)

But now something else…

SIAM J. MATRIX ANAL. APPL.
Vol. 31, No. 5, pp. 2665–2680

ON A PARAMETRIZATION OF POSITIVE SEMIDEFINITE MATRICES WITH ZEROS*

MATHIAS DRTON† AND JOSEPHINE YU‡

Abstract. We study a class of parametrizations of convex cones of positive semidefinite matrices with prescribed zeros. Each such cone corresponds to a graph whose nonedges determine the prescribed zeros. Each parametrization in this class is a polynomial map associated with a simplicial complex supported on cliques of the graph. The images of the maps are convex cones, and the maps can only be surjective onto the cone of zero-constrained positive semidefinite matrices when the associated graph is chordal and the simplicial complex is the clique complex of the graph. Our main result gives a semialgebraic description of the images of the parametrizations for chordless cycles. The work is motivated by the fact that the considered maps correspond to Gaussian statistical models with hidden variables.

ELSEVIER
Contents lists available at ScienceDirect
Linear Algebra and its Applications
www.elsevier.com/locate/laa
LINEAR ALGEBRA and its Applications

On the causal interpretation of acyclic mixed graphs under multivariate normality

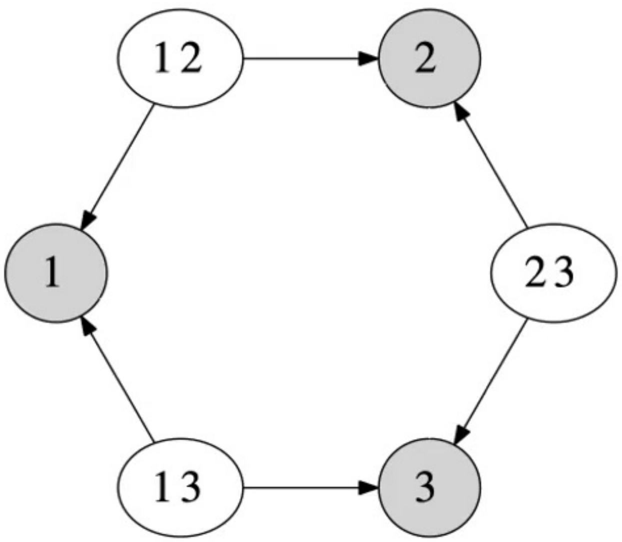
CrossMark

Christopher J. Fox[a], Andreas Käufl[b], Mathias Drton[c,*]

[a] Department of Statistics, The University of Chicago, Chicago, IL, USA
[b] Institute for Mathematics, University of Augsburg, Augsburg, Germany
[c] Department of Statistics, University of Washington, Seattle, WA, USA

Starting point for D. & Yu (2010): "Bidirected three-cycle"

Can any positive definite 3x3 correlation matrix arise in this model?

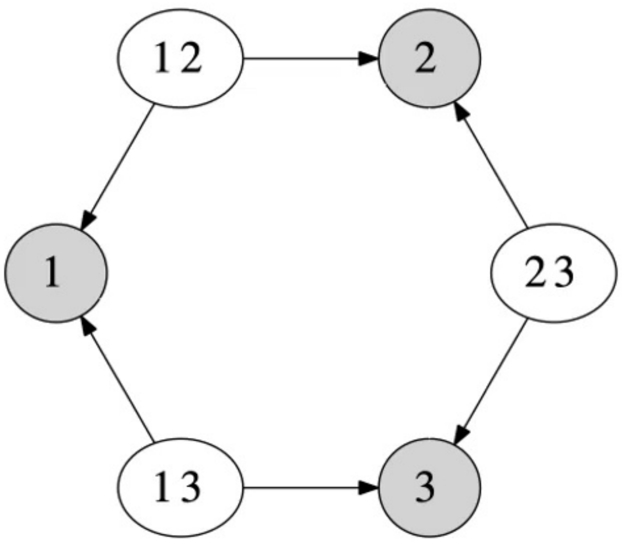


$$X_1 = \gamma_{12}L_{12} + \gamma_{13}L_{13} + \varepsilon_1$$
$$X_2 = \gamma_{21}L_{12} + \gamma_{23}L_{23} + \varepsilon_2$$
$$X_3 = \gamma_{31}L_{13} + \gamma_{32}L_{23} + \varepsilon_3$$

ε_j independent with $\mathrm{Var}[\varepsilon_j] = \gamma_j^2$

"Bidirected three-cycle"

Can any positive definite 3x3 correlation matrix arise in this model?



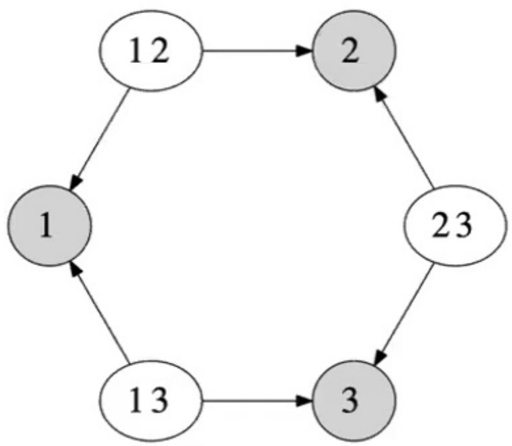
$$X_1 = \gamma_{12} L_{12} + \gamma_{13} L_{13} + \varepsilon_1$$
$$X_2 = \gamma_{21} L_{12} + \gamma_{23} L_{23} + \varepsilon_2$$
$$X_3 = \gamma_{31} L_{13} + \gamma_{32} L_{23} + \varepsilon_3$$

Spirtes et al. (1998), Richardson & Spirtes (2002):

NO… not all correlations can be "large"
(not all larger than $1/\sqrt{2} \approx 0.707$ ...)

"Bidirected three-cycle"

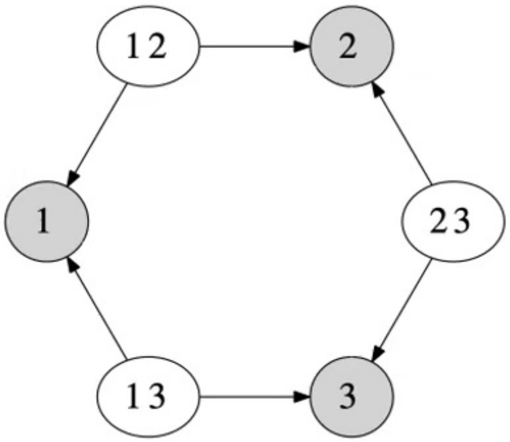
The parametrization…



$$\begin{pmatrix} \gamma_1^2 + \gamma_{12}^2 + \gamma_{13}^2 & \gamma_{12}\gamma_{21} & \gamma_{13}\gamma_{31} \\ \gamma_{12}\gamma_{21} & \gamma_2^2 + \gamma_{21}^2 + \gamma_{23}^2 & \gamma_{23}\gamma_{32} \\ \gamma_{13}\gamma_{31} & \gamma_{23}\gamma_{32} & \gamma_3^2 + \gamma_{31}^2 + \gamma_{32}^2 \end{pmatrix}$$

In fact, at least one correlation $\rho_{ij} \leq 1/2$.

“Bidirected three-cycle”



Our solution:

Correlation matrix $\mathrm{R} = (\rho_{ij})$ is in model *if and only if*

$$1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23} \geq 0.$$

"Bidirected three-cycle"

The main trick to show this (+some algebra concerning polynomial equation systems…)

$$\begin{pmatrix} \gamma_1^2+\gamma_{12}^2+\gamma_{13}^2 & \gamma_{12}\gamma_{21} & \gamma_{13}\gamma_{31} \\ \gamma_{12}\gamma_{21} & \gamma_2^2+\gamma_{21}^2+\gamma_{23}^2 & \gamma_{23}\gamma_{32} \\ \gamma_{13}\gamma_{31} & \gamma_{23}\gamma_{32} & \gamma_3^2+\gamma_{31}^2+\gamma_{32}^2 \end{pmatrix}$$

$$\begin{pmatrix} \gamma_1^2+\gamma_{12}^2+\gamma_{13}^2 & -\gamma_{12}\gamma_{21} & \gamma_{13}\gamma_{31} \\ -\gamma_{12}\gamma_{21} & \gamma_2^2+\gamma_{21}^2+\gamma_{23}^2 & \gamma_{23}\gamma_{32} \\ \gamma_{13}\gamma_{31} & \gamma_{23}\gamma_{32} & \gamma_3^2+\gamma_{31}^2+\gamma_{32}^2 \end{pmatrix}$$

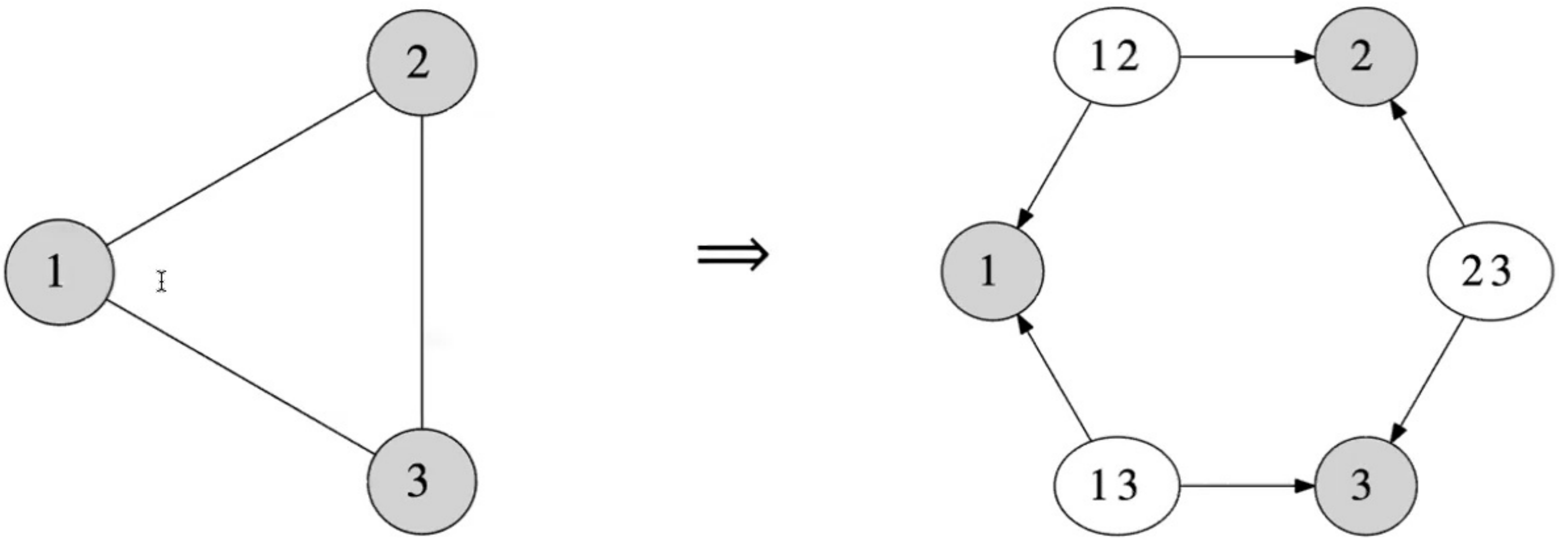
If a PD matrix is in the model, then so is its version with (1,2)-entry negated. (negate γ_{12})

$\Rightarrow$ Constraint:

Matrix with negated (1,2)-entry must remain PSD.

Perspective: Parametrizing positive definite matrices

Drton & Yu (2010): DAGs with latent source nodes, no edges among observed nodes.



Parametrizations given by simplicial complexes (mDAGs)

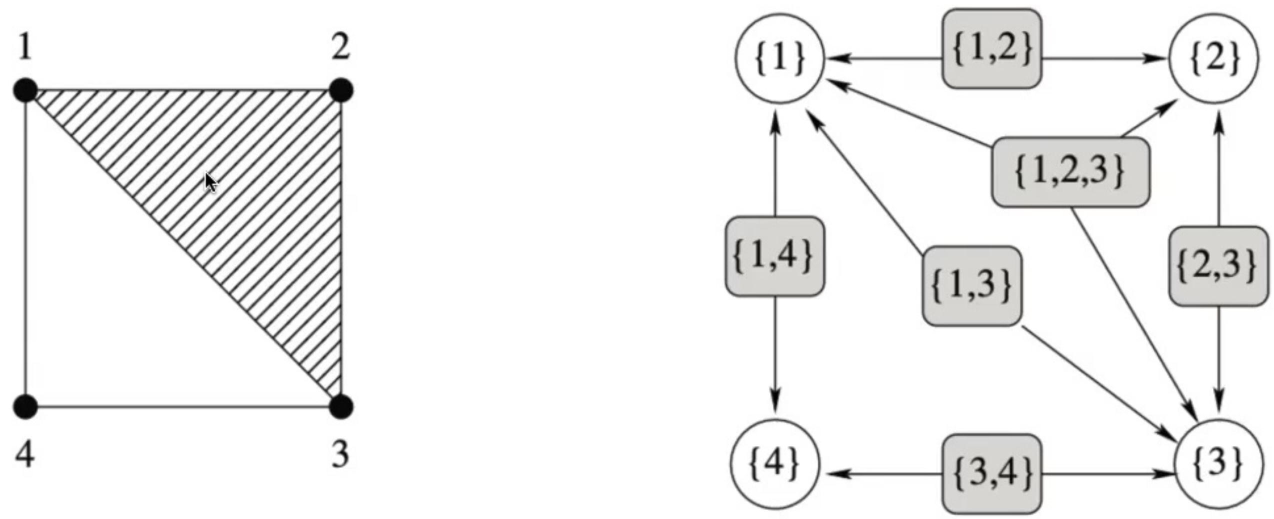


FIG. 1. *A simplicial complex (left) and the acyclic bipartite digraph corresponding to it (right).*

Motivation

- Let $\mathbb{S}^m_{\succeq}$ be set of symmetric and positive semi-definite $m \times m$ matrices
  - closed convex cone
  - extreme rays are rank one matrices: $\Sigma = vv^T$
- Let $G = (V, E)$ be a graph on $V = \{1, 2, \ldots, m\}$.
- Graphical cones:

$$\mathbb{S}_{\succeq}(G) := \{\Sigma = (\sigma_{ij}) \in \mathbb{S}^m_{\succeq} \; : \; \sigma_{ij} = 0 \;\text{ if } i \neq j \text{ and } \{i, j\} \notin E(G)\}$$

  appear in statistical models for multivariate normal distribution.

- Simple parametrizations useful for parameter estimation, prior specification, ...

Parametrization

- $\Delta =$ simplicial complex on V such that faces are cliques in G
  (e.g. $\Delta =$ edge complex of G, $\Delta =$ clique complex of G).

- $\Gamma = (\gamma_{i,F}) \in \mathbb{R}^{V \times \Delta}$ such that $\gamma_{i,F} = 0$ if $i \notin F$.

- Then $\Gamma\Gamma^T \in \mathbb{S}_{\succeq}(G)$.

- Let $\phi_\Delta : \prod_{F \in \Delta} \mathbb{R}^{|F|} \to \mathbb{S}_{\succeq}(G)$ given by $\gamma \to \Gamma(\gamma)\Gamma(\gamma)^T$

  Is ϕ_Δ surjective onto $\mathbb{S}_{\succeq}(G)$? If not, what is the image?

Parametrizations given by simplicial complexes (mDAGs)

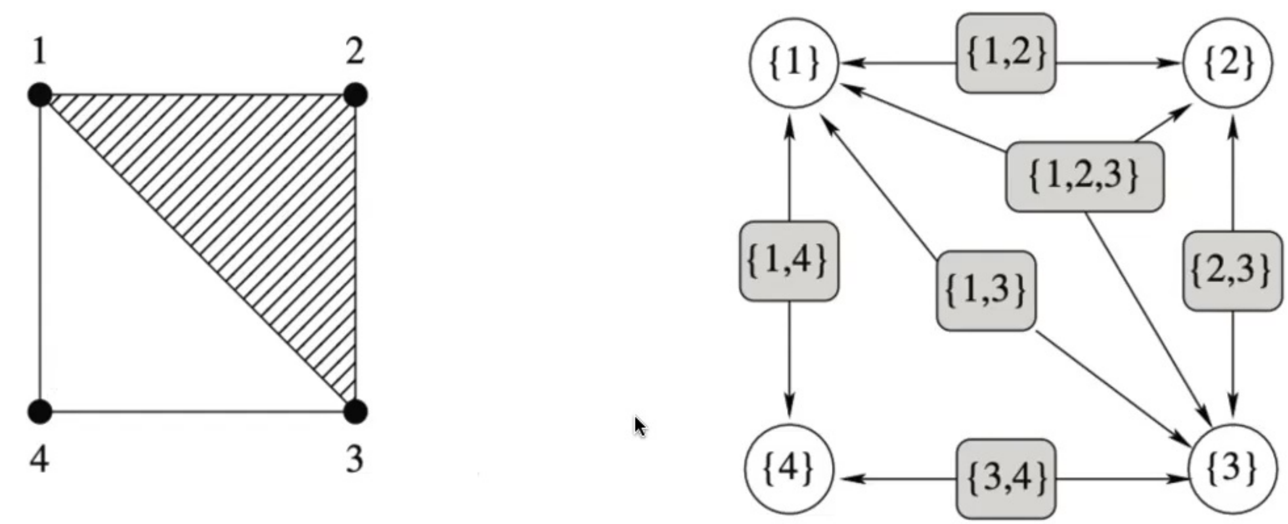


FIG. 1. *A simplicial complex (left) and the acyclic bipartite digraph corresponding to it (right).*

Example: Three-chain

- $\Delta =$ Simplicial complex whose facets are the edges $\{1,2\}$ and $\{2,3\}$ of a three-chain $1-2-3$.
- Then

$$\Gamma(\gamma) = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccccc} \{1\} & \{2\} & \{3\} & \{1,2\} & \{2,3\} \end{array} \\ \begin{pmatrix} \gamma_1 & 0 & 0 & \gamma_{12} & 0 \\ 0 & \gamma_2 & 0 & \gamma_{21} & \gamma_{23} \\ 0 & 0 & \gamma_3 & 0 & \gamma_{32} \end{pmatrix} \end{array}$$

and

$$\phi_\Delta(\gamma) = \begin{pmatrix} \gamma_1^2 + \gamma_{12}^2 & \gamma_{12}\gamma_{21} & 0 \\ \cdot & \gamma_2^2 + \gamma_{21}^2 + \gamma_{23}^2 & \gamma_{23}\gamma_{32} \\ \cdot & \cdot & \gamma_3^2 + \gamma_{32}^2 \end{pmatrix}.$$

- The map ϕ_Δ is surjective onto $\mathbb{S}_{\succeq}(G)$ in this case.

Convexity

Lemma

For any simplicial complex Δ *with underlying graph* $G(\Delta) = G$*, image of* ϕ_Δ *is a full-dimensional semialgebraic subset in* $\mathbb{S}_{\succeq}(G)$.

Theorem

For any simplicial complex Δ *on* $[m]$*, image of* ϕ_Δ *is a closed convex cone whose extreme rays are the rank one matrices of form* vv^T *with* $v \in \mathbb{R}^m$ *supported on a clique in* G.

Useful tool: Cholesky decomposition $\Sigma = LL^T$, L lower triangular.

Corollary

The convex hull of all rank one matrices in $\mathbb{S}_{\succeq}(G)$ *is* $im(\phi_\Delta)$*, where* Δ *is the clique complex of* G.

Sparsity order

Sparsity order of graph $G =$ max rank of an extreme ray of $\mathbb{S}_{\succeq}(G)$.

Theorem (Laurent, 2001, and references therein)

Let G be a graph on $[m]$, $ord(G)$ its sparsity order. Then

(i) $1 \leq ord(G) \leq m-2$,

(ii) $ord(G) = 1$ *if and only if G is chordal,*

(iii) $ord(G) = m-2$ *if and only if G is a chordless cycle, and*

(iv) *if H is an induced subgraph of G, then* $ord(H) \leq ord(G)$.

Corollary

(i) *The map ϕ_Δ is surjective onto $\mathbb{S}^m_{\succeq}(G)$ if and only if G is chordal and Δ is the clique complex of G.*

(ii) *For $\Delta = E(G)$, ϕ_Δ is surjective if and only if G is a forest.*

Chordless cycle C_m: $\Delta = E(C_m)$

Main Theorem

Let $\{k, l\}$ be any edge of the chordless cycle C_m. A matrix $\Sigma = (\sigma_{ij}) \in \mathbb{S}^m_{\succeq}(C_m)$ is in the image of ϕ_{C_m} if and only if

$$\min\{\det(\Sigma), \det(\Sigma^{(kl)})\} = \sum_{\text{matching } \Theta} (-1)^{|\Theta|} \prod_{\{i,j\}\in\Theta} \sigma_{ij}^2 \prod_{i\in[m]\setminus\Theta} \sigma_{ii} - 2\prod_{i=1}^{m} |\sigma_{i,i+1}| \quad \geq \quad 0.$$

Corollary

For a matrix $\Sigma = (\sigma_{ij}) \in \mathbb{S}^m_{\succeq}(C_m)$, the following are equivalent:

(i) *Σ is in the image of ϕ_{C_m}.*

(ii) *$\Sigma^{(ij)}$ is PSD for some edge $ij \in C_m$ with $\sigma_{ij} \neq 0$.*

(iii) *$\Sigma^{(ij)}$ is PSD for all edges $ij \in C_m$.*

Proof idea (Sufficiency)

- It suffices to show that for each $\Sigma \in \mathbb{S}^m_{\succeq}(C_m)$ satisfying the claimed condition, the equation system
$$\Gamma(\gamma)\Gamma(\gamma)^T = \Sigma$$
has a real solution.
- Σ has $2m$ free entries, γ has $3m$ (one per node, two per edge)
- Set node incidence variables γ_{ii} to zero.
- It suffices to show that
$$\gamma^2_{i,i-1} + \gamma^2_{i,i+1} = \sigma_{ii}, \qquad i = 1, \ldots, m,$$
$$\gamma_{i-1,i}\gamma_{i,i-1} = \sigma_{i-1,i}, \qquad i = 1, \ldots, m.$$
has a real solutions for any choice of $\Sigma = (\sigma_{ij}) \in \mathbb{S}^m_{\succeq}(C_m)$ satisfying the claimed condition.

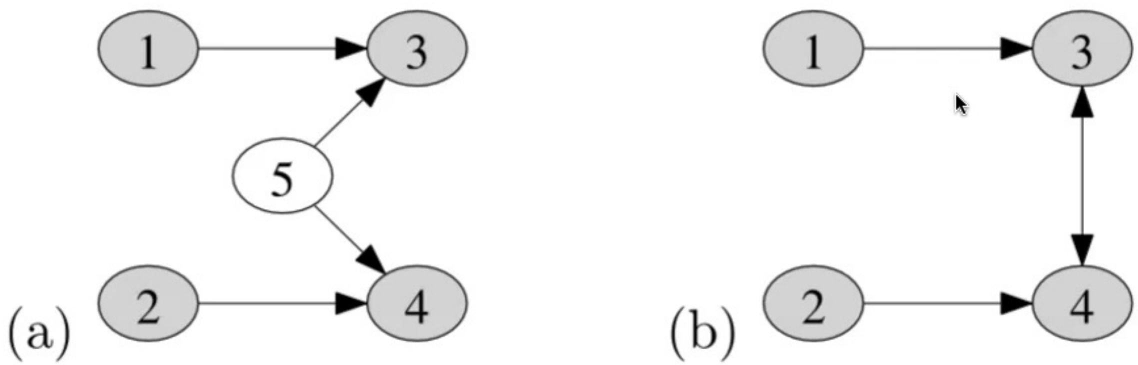
Spherical volumes

Table: Spherical volume of the image of ϕ_{C_m} as a fraction of the spherical volume of the cone $\mathbb{S}_{\succeq 0}(C_m)$.

m	3	4	5	6	7
Vol	0.78	0.90	0.95	0.98	0.99

Fox, Käufl & D. (2015)

- Consider linear Gaussian models based on mixed graphs (latent projections).
- Models postulate that confounding yields error correlations of `arbitrary' size.
- Which ones are strictly causal = they coincide with a linear Gaussian DAG model with latent variables?



Mathias Drton (TUM)

Main results

Theorem 1.2. *If an acyclic mixed graph $\mathcal{G} = (V, D, B)$ has a chordal bidirected part (V, B), then $\mathbf{N}(\mathcal{G})$ is strictly causal.*

Theorem 1.3. *Suppose the mixed graph $\mathcal{G} = (V, D, B)$ is a chain graph. Then $\mathbf{N}(\mathcal{G})$ is strictly causal if and only if the bidirected part (V, B) is chordal.*

Conclusion

- Gaussian (or rather linear covariance) version of some of the problems discussed by others. Is this right:

  Statistics $\nrightarrow$ Physics but it's not impossible that Physics $\rightarrow$ Statistics?

- Sign trick to derive inequalities between correlations in latent variable models that then differ from the model associated to a mixed graph/latent projection.
- Other approaches to get inequalities?
- If a linear Gaussian model is strictly causal, what's the ``causality index'' = smallest number of latent variables needed?
- …

European Research Council
Established by the European Commission