Treks, tetrads & third-order tensors of moments

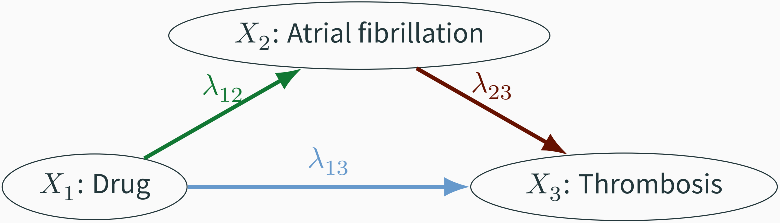
Polynomial relations for linear graphical models beyond covariance matrices

Alex Grosdos

joint with Carlos Améndola, Mathias Drton, Roser Homs, Elina Robeva

ETH–UCPH–TUM Workshop on Graphical Models

Directed arrows capture causal relations between random variables



translating to equations

$$
\begin{aligned}
X_1 &= && && \varepsilon_1 \\
X_2 &= \lambda_{12} X_1 && && + \varepsilon_2 \\
X_3 &= \lambda_{13} X_1 && + \lambda_{23} X_2 && + \varepsilon_3
\end{aligned}
$$

Structural equation models

A graph $G = (V, E)$ gives rise to structural equations

$$X_i = \sum_{j \in \text{pa}(i)} \lambda_{ji} X_j + \varepsilon_i, \quad i \in V,$$

- ε_i represent stochastic errors with $\mathbb{E}[\varepsilon_i] = 0$,
- λ_{ji} are unknown parameters forming a matrix $\Lambda = (\lambda_{ij})$.

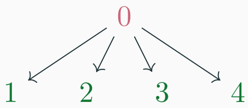
The corresponding moment tensor model is

$$\begin{aligned}\mathcal{M}^{(2,3)}(G) = \{(S = (I-\Lambda)^{-T}\Omega^{(2)}(I-\Lambda)^{-1},\\ T = \Omega^{(3)} \bullet (I-\Lambda)^{-1} \bullet (I-\Lambda)^{-1} \bullet (I-\Lambda)^{-1}) :\\ \Omega^{(2)} \text{ is } n \times n \text{ positive definite diagonal matrix},\\ \Omega^{(3)} \text{ is } n \times n \times n \text{ diagonal 3-way tensor, and } \Lambda \in \mathbb{R}^E\}.\end{aligned}$$

This makes (statistical) sense for non-Gaussian random variables.

Goal: characterize polynomial relations among elements in the model.

Let 0 be a hidden variable and $1, 2, 3, 4$ be observed ones.



Then the following tetrad relations [Drton, Sturmfels, Sullivant 05; Peters, Janzing, Schölkopf 17] hold among the elements of the covariance matrix

$$s_{12}s_{34} - s_{13}s_{24} \quad s_{12}s_{34} - s_{14}s_{23} \quad s_{13}s_{24} - s_{14}s_{23}.$$

Proof Evaluate the expressions via

$$s_{ij} \mapsto \omega_0 \lambda_{0i} \lambda_{0j}.$$

They all evaluate to zero.

Why higher moments?

The covariance matrix structure reveals statistical information, e.g.

has covariance ideal $I = \langle s_{13}s_{22} - s_{12}s_{23} \rangle$ corresponding to the statement $X_1 \perp\!\!\!\perp X_3 | X_2$...

Proposition One can deduce the skeleton and the colliders of the graph from the covariance matrix.

...but so do the graphs

Proposition [Wang, Drton 18] The whole graph can be reconstructed from covariance matrices and 3^{rd}-order tensors.

A *trek* with top v between i and j is formed by joining two paths sharing a source v

$$i = i_l \leftarrow \cdots \leftarrow i_1 \leftarrow v \rightarrow j_1 \rightarrow \ldots \rightarrow j_r = j$$

and gives rise to a monomial

$$a_v(\lambda_{vi_1}\lambda_{i_1i_2}\ldots\lambda_{i_{l-1}i_l})(\lambda_{vj_1}\lambda_{j_1j_2}\ldots\lambda_{j_{r-1}j_r}).$$

An n-*trek* between vertices $k_1, \ldots, k_n$ is a collection of directed paths $T = (P_1, \ldots, P_n)$, where P_r has sink k_r and they all share the same top vertex as source $v = \mathrm{top}(T)$.

$$\begin{array}{lllllll} & \nearrow & k_{11} \longrightarrow k_{12} \longrightarrow \cdots \rightarrow k_{1l_1} = k_1 \\ v & \longrightarrow & k_{21} \longrightarrow k_{22} \longrightarrow \cdots \rightarrow k_{2l_2} = k_2 \\ & \searrow & k_{31} \longrightarrow k_{32} \longrightarrow \cdots \rightarrow k_{3l_2} = k_3 \end{array}$$

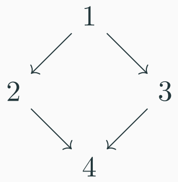
A trek is minimal if the top node is the only one that appears in all paths of the trek. For a graph G, let $T(i_1, \ldots, i_n)$ be the set of minimal n-treks between $i_1, \ldots, i_n$.

Consider the polynomial map ϕ_G:

$$\begin{aligned}
\mathbb{C}[s_{ij}, t_{ijk} \mid 1 \le i \le j \le k \le n] &\to \mathbb{C}[a_i, b_i, \lambda_{ij} \mid i \to j \in E] \\
s_{ij} &\mapsto \sum_{\tau \in T(i,j)} a_{\mathrm{top}(\tau)} \prod_{k \to l \in \tau} \lambda_{kl}, \\
t_{ijk} &\mapsto \sum_{\tau \in T(i,j,k)} b_{\mathrm{top}(\tau)} \prod_{m \to l \in \tau} \lambda_{ml}.
\end{aligned}$$

Example

$$\begin{aligned}
s_{ii} &\mapsto a_i \quad t_{iii} \mapsto b_i \\
s_{13} &\mapsto a_1 \lambda_{13} \\
s_{14} &\mapsto a_1 \lambda_{12} \lambda_{24} + a_1 \lambda_{13} \lambda_{34} \\
t_{123} &\mapsto b_1 \lambda_{12} \lambda_{13}
\end{aligned}$$



$$\begin{aligned} s_{ij} &\mapsto \sum_{\tau \in T(i,j)} a_{\mathrm{top}(\tau)} \prod_{k \to l \in \tau} \lambda_{kl} \\ t_{ijk} &\mapsto \sum_{\tau \in T(i,j,k)} b_{\mathrm{top}(\tau)} \prod_{m \to l \in \tau} \lambda_{ml} \end{aligned}$$

Proposition [Sullivant 08; Améndola, Drton, G, Homs & Robeva 22] For a directed graph G, let ϕ_G be the map given by the simple trek rule. Then the vanishing ideal $I^{(2,3)}(G) := \mathcal{J}(\mathcal{M}^{(2,3)}(G))$ of the model is

$$I^{(2,3)}(G) = \ker \phi_G.$$

Corollary [Améndola, Drton, G, Homs & Robeva 22] If G is a tree, $I^{(2,3)}(G)$ is a toric ideal, i.e. it is generated by binomials.

Vanishing minors

Let $i, j \in V$ be two vertices such that a 2-trek between i and j exists. Define

$$A_{ij} := \begin{bmatrix} s_{ik_1} & \cdots & s_{ik_r} & t_{i\ell_1 m_1} & \cdots & t_{i\ell_q m_q} \\ s_{jk_1} & \cdots & s_{jk_r} & t_{j\ell_1 m_1} & \cdots & t_{j\ell_q m_q} \end{bmatrix},$$

where

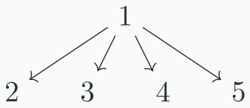
- $k_1, \dots, k_r$ are all vertices such that $\text{top}(i, k_a) = \text{top}(j, k_a)$ and
- (l_1, m_1),... ,(l_q, m_q) are all pairs of vertices such that $\text{top}(i, l_b, m_b) = \text{top}(j, l_b, m_b)$.

Proposition [Améndola, Drton, G, Homs & Robeva 22] For a tree G, the following polynomials are in $I^{(2,3)}(G)$:

- s_{ij} such that there is no 2-trek between i and j,
- t_{ijk} such that there is no 3-trek between i, j and k,
- the 2-minors of A_{ij}, for all (i, j) with a 2-trek between them.

Proposition [Améndola, Drton, G, Homs & Robeva 22] All quadratic binomials in $I^{(2,3)}(G)$ are linear combinations of 2-minors of matrices A_{ij}.

Example The binomial $f = s_{23}t_{145} - s_{45}t_{123}$ lies in $I^{(2,3)}(G)$. It is a sum of minors from A_{13}, A_{14} and A_{15}.

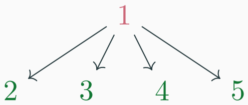


Theorem [Améndola, Drton, G, Homs & Robeva 22] All binomials in $I^{(2,3)}(G)$ are generated by quadratic binomials, i.e. $I^{(2,3)}(G)$ is generated by the matrices A_{ij} (plus vanishing indeterminates).

Proof A distance reduction argument for binomials in the ideal, showing that matrix minors are a Markov basis.

Application: trees with hidden variables

Let $H \cup O$ be a partition of the nodes of the DAG G. The hidden nodes H are said to be *upstream* from the observed nodes O in G if there are no edges $o \to h$ in G with $o \in O$ and $h \in H$.



Lemma The ideal $I^{(2,3)}(G)$ is homogeneous w.r.t. the grading:

$$\begin{aligned} \deg s_{ij} &= (1, 1 + \text{number of elements in the multiset } \{i, j\} \text{ in } O) \\ \deg t_{ijk} &= (1, \text{number of elements in the multiset } \{i, j, k\} \text{ in } O). \end{aligned}$$

Proposition For a tree G, the observed variable ideal $I_O^{(2,3)}(G)$ is generated by the minors of the submatrices of A_{ij} with i, j both in O, with columns indexed by k and (l, m) where k, l, m are all in O.

Theorem [Améndola, Drton, G, Homs & Robeva 22] Let J be the ideal generated by the linear generators of $I^{(2,3)}(G)$ and matrices A_{ij} such that there is a directed path between i and j. Then

$$\mathcal{M}^{(2,3)}(G) = V(J) \cap PD(n).$$

In particular, pick $(S,T) \in \mathcal{M}^{(2,3)}(G)$. For $i \to j \in E$, let $\lambda_{ij} = \frac{s_{ij}}{s_{ii}}$, coming from A_{ij}. Then one can show

$$S' = (I-\Lambda)^T S (I-\Lambda) \quad \text{and} \quad T' = T \bullet (I-\Lambda) \bullet (I-\Lambda) \bullet (I-\Lambda)$$

are diagonal.

Example Let G be $1 \to 2$, $1 \to 3$, $1 \to 4$, $1 \to 5$. Computations show

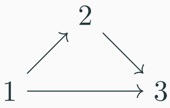
$$I^{(2,3)}(G) = (J : s_{11}^{\infty})$$

and

$$\mathcal{M}^{(2,3)}(G) = V(I^{(2,3)}(G)) \cap PD(5) = V(J) \cap PD(5).$$

A forest of non-trees

The ideal of the 3×3 minors of



$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 11 & 12 & 13 & 22 & 23 \end{array} \\ \begin{pmatrix} s_{11} & s_{12} & t_{111} & t_{112} & t_{113} & t_{122} & t_{123} \\ s_{12} & s_{22} & t_{112} & t_{122} & t_{123} & t_{222} & t_{223} \\ s_{13} & s_{23} & t_{113} & t_{123} & t_{133} & t_{223} & t_{223} \end{pmatrix} \end{array}$$

is given by

the 2×2 minors of

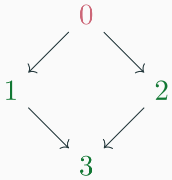
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 11 & 12 & 13 \end{array} \\ \begin{pmatrix} s_{11} & t_{111} & t_{112} & t_{113} \\ s_{12} & t_{112} & t_{122} & t_{123} \\ s_{13} & t_{113} & t_{123} & t_{133} \end{pmatrix} \end{array}$$

and the determinant of

$$\begin{array}{c} \\ \emptyset \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 13 & 22 & 23 \end{array} \\ \begin{pmatrix} s_{13} & s_{22} & s_{23} \\ t_{123} & t_{222} & t_{223} \\ t_{133} & t_{223} & t_{223} \end{pmatrix} \end{array}$$

A forest on non-trees

The ideal of the graph



is given by the maximal minors of

$$
\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 1 & 2 & 11 & 12 & 13 & 22 & 23 \end{array} \\
\begin{pmatrix}
s_{11} & s_{12} & t_{111} & t_{112} & t_{113} & t_{122} & t_{123} \\
s_{12} & s_{22} & t_{112} & t_{122} & t_{123} & t_{222} & t_{223} \\
s_{13} & s_{23} & t_{113} & t_{123} & t_{133} & t_{223} & t_{233}
\end{pmatrix}
\end{array}.
$$

Given a polytree G, the third-order moment polytope is

$P_G^{(3)} = \mathrm{conv}\left(e_{ijk} : i, j, k \text{ such that a 3-trek between } i, j \text{ and } k \text{ exists}\right)$

where $e_{ijk} \in \mathbb{R}^{|V|+|E|}$ is the vector of exponents of the monomial $\phi_G(t_{ijk}) = b_{\mathrm{top}(i,j,k)} \prod_{l \to m \in \mathcal{T}(i,j,k)} \lambda_{lm} \in \mathbb{R}[b_l, \lambda_{lm}]$.

Theorem The third-order moment polytope $P_G^{(3)}$ is the solution to

$$z_l \geq 0 \text{ for all } l \in V,$$

$$y_{lm} \geq 0 \text{ for all } l \to m \in E,$$

$$\sum_{l \in V} z_l = 1,$$

$$2z_l + \sum_{h \in pa(l)} y_{hl} - y_{lm} \geq 0 \text{ for all m such that } l \to m \in E,$$

$$3z_l + \sum_{h \in pa(l)} y_{hl} - \sum_{m \in ch(l)} y_{lm} \geq 0 \text{ for all } l \in V.$$

Open problems

- What about non-trees?
- Sparse latent factor analysis (current project with Drton, Portakal, Sturma)
- Latent factor analysis in higher dimensions (see also Ardiyansyah, Sodomaco 22)
- Euclidean distance degrees to the varieties
- **[Your favorite graphical model here]**

Summary

- Graphical models are richer in the non-Gaussian setting, it is meaningful to study higher-order moment tensors
- The trek rules can be extended for h.o.m. and one can obtain binomial (matrix minors) descriptions of the ideals
- The hidden variable ideals are given by some of the binomials
- Lots of open questions …

Reference:
Améndola, Drton, G, Homs & Robeva
Third-Order Moment Varieties of Linear Non-Gaussian Graphical Models

Thank you!