Mathematisches Forschungsinstitut Oberwolfach

Report No. 55/2022

DOI: 10.4171/OWR/2022/55

Algebraic Structures in Statistical Methodology

Organized by
Mathias Drton, München
Thomas Kahle, Magdeburg
Seth Sullivant, Raleigh
Caroline Uhler, Zürich/Cambridge MA

4 December – 10 December 2022

Abstract. Algebraic structures arise naturally in a broad variety of statistical problems, and numerous fruitful connections have been made between algebra and discrete mathematics and research on statistical methodology. The workshop took up this theme with a particular focus on algebraic approaches to graphical models, causality, axiomatic systems for independence and non-parametric models.

Mathematics Subject Classification (2020): 62R01, 62H22, 62F10, 62G05, 62P10, 62P20, 13P15, 14M25, 14Q20.

Introduction by the Organizers

The workshop *Algebraic Structures in Statistical Methodology*, organised by Mathias Drton (München), Thomas Kahle (Magdeburg), Seth Sullivant (Raleigh) and Caroline Uhler (Zürich) was attended by an international group of researchers from algebraic statistics. With 24 participants, there was a strong thematic focus.

The week started off with a stimulating lecture by Steffen Lauritzen who presented two open problems in the area of graphical models. Discussions about these problems, in particular about distributions whose support is a distributive lattice, could be heard over dinner, on the hike, and next to several blackboards. Over the course of the week, each participant gave exactly one talk. The themes of the meeting and these talks can be roughly grouped as follows.

Conditional independence, graphical models and causality. The concept of conditional independence is fundamental to statistical modeling and inference in general. The logical structures that underlie conditional independence have

been the subject of intense investigation. Seminal work of Pearl and Verma proposed the semi-graphoid axioms for an axiomatization of conditional independence, which later was proven impossible. Natural axiom systems like semi-graphoids and gaussoids have since become objects of independent interest both from a statistical and logical inference perspective. Conditional independence also plays a key role in the area of graphical models, notably in the development of algorithms for inferring a causal network from data. Generally, the combinatorics and algebra related to modeling with graphs arose in many different talks.

Geometry of models and likelihood. One of the pillars on which algebraic statistics rests is the observation that statistical models are often algebraic varieties and that this perspective provides insights into the behavior of estimation/testing methods. Several talks in this meeting focused on the complexity of the likelihood method, measures of distance to models and non-parametric estimation through log-concavity assumptions.

Applications. Algebraic statistics continuously draws inspiration from novel applications in biology, genomics, economics, game theory, theoretical physics, etc. The talks of Solus, Kubjas, Portakal and Zwiernik drew inspiration/showcased some of these applications and introduced them to the community.

The atmosphere at the workshop was very lively and engaging. Many participants expressed relief to be on such a productive and inspiring meeting after another year that was still influenced by the remnants of the COVID19-pandemic. The hike was well attended despite chilly weather. In the evenings there were foosball matches and football watching (a world cup took place this winter!). The participants also enjoyed browsing through (online) archives of activities at Oberwolfach. Participants Steffen Lauritzen and Milan Studený shared memories from a similarly themed 1995 workshop organized at MFO by Friedrich Pukelsheim and Michael Perlman: Algebraic Methods in Multivariate Statistical Analysis. It is nice to see that key ideas from this time are carried forward by a diverse, energetic and open community.

Upon conclusion of the workshop, Thomas Kahle and Carlos Améndola agreed to write a snapshot about an exciting and accessible topic from algebraic statistics. Elizabeth Gross (Hawaii) was a Simons visiting professor at TU Munich and Serkan Hoşten (San Francisco) at OvGU Magdeburg.

The workshop was a key event for the community of algebraic statistics. We are looking forward to much new research and many new collaborations emerging from the workshop.

Acknowledgement: The MFO and the workshop organizers would like to thank the Simons Foundation for supporting Elizabeth Gross and Serkan Hoşten in the "Simons Visiting Professors" program at the MFO.

Workshop: Algebraic Structures in Statistical Methodology

Table of Contents

Abstracts

Two open problems in graphical models of algebraic nature

Steffen Lauritzen

Factorization of totally positive Markov distributions. Let $\mathcal{X} = \{0,1\}^V$ where V is a finite set and $G = (V, E)$ be a simple, undirected graph with V as vertex set. We further let $x_A = (x_v)_{v\in A}$ denote the coordinate projections from $\mathcal{X}$ to $\mathcal{X}_A = \{0,1\}^A$ for subsets $A \subseteq V$. Recall that a probability distribution $P \in \mathbb{P}(\mathcal{X})$ is said to be (globally) *Markov with respect to* G if for all finite and disjoint subsets A, B, C of V it holds that

$$A \perp_G B \,|\, C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C$$

where $\perp_G$ denotes separation in G.

We let $M(G)$ denote the set of Markov probability distributions on $\mathcal{X}$. Further we let $M_+(G)$ denote the strictly positive elements of $M(G)$. A distribution P is said to *factorize with respect to* G if there are functions $\psi_A : \mathcal{X}_A \mapsto \mathbb{R}$ such that the density p of P with respect to counting measure has the form

$$p(x) = \prod_{A\in\mathcal{A}} \psi_A(x_A),$$

where $\mathcal{A}$ denotes the set of subsets that are *complete* with respect to G, i.e. $u \sim v$ for all $u \neq v \in A$. The set of Markov distributions that factorize shall be denoted $M_F(G)$. Finally, we let $M_E(G)$ denote the distributions that are *extended Markov with respect to* G, i.e. the pointwise topological closure of $M_F(G)$ or, equivalently, of $M_+(G)$. It is well-known ([3], p. 42 ff.) that

$$M_+(G) \subseteq M_F(G) \subseteq M_E(G) \subseteq M(G).$$

All inclusions are strict for a generic graph. We shall also be interested in the subclass of *Ising models* $M_I(G) \subseteq M_F(G)$, which are those that factorize using only pairs, i.e. edges of G; in other words those that have the form

$$p(x) = \prod_{e\in E} \psi_e(x_e).$$

The article [1] gives general conditions for $P \in M_E(G)$ to factorize, chiefly formulated in terms of a condition on the *support of* P.

Now consider a distribution P that is *multivariate totally positive of order two*, i.e. its density p satisfies

$$p(x \vee y)p(x \wedge y) \geq p(x)p(y) \quad \text{for all } x, y \in \mathcal{X}, \tag{1}$$

where the maximum and minimum are taken coordinatewise. Let $M_2(G)$ denote the Markov distributions that also satisfy (1).

Conjecture 1. *$M_I(G) \cap M_2(G)$ is closed under pointwise limits.*

Conjecture 2. *$M_2(G) \subseteq M_F(G)$.*

The results in [1] clearly indicate that conditions for factorization of $P \in M(G)$ only relate to the support of P it is worth noting that (1) clearly implies that the support of a totally positive distribution is a *lattice*, it seems natural to strengthen the conjectures further to

Conjecture 3. *If the support of P is a lattice and $P \in M(G)$, then $P \in M_F(G)$.*

See also [4] for details on binary totally positive distributions.

Convergence of neighbourhood coordinate descent. The direct maximum likelihood problem for a Gaussian graphical model may be formulated as follows, where S denotes the empirical covariance matrix.

$$\begin{aligned} &\underset{K}{\text{maximize}} && \ell(K) = \log\det(K) - \operatorname{tr}(KS) \\ &\text{subject to} && K \in \mathbb{S}^{d\times d}_{\succ}(G), \end{aligned} \tag{2}$$

where $\mathbb{S}^{d\times d}_{\succ}(G)$ denotes the set of positive definite matrices K with $K_{uv} = 0$ for all $uv \notin E(G)$. The dual of this optimization problem is ([5])

$$\begin{aligned} &\underset{\Sigma}{\text{minimize}} && \ell^*(\Sigma) = -\log\det(\Sigma) - d \\ &\text{subject to} && \Sigma \in \mathbb{S}^{d\times d}_{\succ}, \quad \Sigma_{uv} = S_{uv} \text{ for all } uv \in E(G) \end{aligned} \tag{3}$$

It holds that Σ is the unique optimizer of (3) if and only if $K = \Sigma^{-1}$ is the unique optimizer of (2). Algorithms for solving (3) typically work by coordinate descent, for example by choosing $u \in V$ and writing

$$\det\Sigma = \det\Sigma_{\setminus u,\setminus u}\left\{\Sigma_{uu} - \Sigma_{u,\setminus u}(\Sigma_{\setminus u,\setminus u})^{-}\Sigma_{\setminus u,u}\right\}, \tag{4}$$

where $(\Sigma_{\setminus u,\setminus u})^{-}$ is any generalized inverse to $\Sigma_{\setminus u,\setminus u}$ and then optimizing the second factor by letting

$$\tilde{\Sigma}_{\mathrm{bd}^c(u),u} = \Sigma_{\mathrm{bd}^c(u),\mathrm{bd}(u)}(\Sigma_{\mathrm{bd}(u),\mathrm{bd}(u)})^{-}S_{\mathrm{bd}(u),u}.$$

We shall refer to this as the *NCD algorithm* (neighbourhood coordinate descent). This algorithm is convergent if S has full rank. After the update we then have

$$\det\tilde{\Sigma} = \left\{\Sigma_{uu} - S_{u,\mathrm{bd}(u)}(\Sigma_{\mathrm{bd}(u),\mathrm{bd}(u)})^{-}S_{\mathrm{bd}(u),u}\right\}\det\Sigma_{\setminus u,\setminus u}.$$

If $\det\Sigma_{\setminus u,\setminus u} = 0$, we will still have $\det\Sigma = 0$ after the update. However, since

$$\operatorname{rank}\Sigma = \operatorname{rank}\Sigma_{\setminus u,\setminus u} + \operatorname{rank}\left\{\Sigma_{uu} - \Sigma_{u,\setminus u}(\Sigma_{\setminus u,\setminus u})^{-}\Sigma_{\setminus u,u}\right\}$$

then if

$$\Sigma_{uu} - \Sigma_{u,\setminus u}(\Sigma_{\setminus u,\setminus u})^{-}\Sigma_{\setminus u,u} = 0 < \Sigma_{uu} - S_{u,\mathrm{bd}(u)}(\Sigma_{\mathrm{bd}(u),\mathrm{bd}(u)})^{-}S_{\mathrm{bd}(u),u},$$

we would have

$$\operatorname{rank}\tilde{\Sigma} = \operatorname{rank}\Sigma_{\setminus u,\setminus u} + 1,$$

so if $\operatorname{rank}(\Sigma_{\setminus u,\setminus u}) = \operatorname{rank}(\Sigma)$, the rank will increase by one.

Gross and Sullivant show in [2] that if $\operatorname{rank}(S) \geq n$ and the n-core of G is empty, then the MLE exists. The question is whether

Conjecture 4. *If rank$(S) \geq n$ and the n-core of G is empty, then the NCD algorithm converges to the MLE.*

This conjecture is probably more risky than those concerning multivariate total positivity but we have no empirical evidence to the contrary.

References

[1] Geiger, D., Meek, C. and Sturmfels, B. (2006). On the toric algebra of graphical models. *The Annals of Statistics,* **34**, 1463–1492.

[2] Gross, E. and Sullivant, S. (2018). The maximum likelihood threshold of a graph. *Bernoulli*, 24:386–407.

[3] Lauritzen, S. (1996). *Graphical Models*. Clarendon Press, Oxford.

[4] Lauritzen, S., Uhler, C., and Zwiernik, P. (2021). Total positivity in exponential families with application to binary variables. *The Annals of Statistics,* **49**, 1436–1459.

[5] Uhler, C. (2019). Gaussian graphical models. In Maathuis, M., Drton, M., Lauritzen, S., and Wainwright, M., editors, *Handbook of Graphical Models*, pages 217–238. CRC Press, Boca Raton, FL.

Combinatorial and algebraic perspectives on the marginal independence structure of Bayesian networks

Pratik Misra

Directed acyclic graphs (DAGs) are commonly used in modern data science and artificial intelligence to represent the conditional independence structure, and even the causal relations, underlying complex systems of jointly distributed random variables. In recent years, DAG models have become a cornerstone of the field of *causal inference*, in which one aims to learn the cause-effect relations in a given complex system and then estimate the causal effect of one variable within the system on another. The first step in such a causal analysis is the process of *causal discovery* in which one aims to infer a DAG to which the data-generating distribution is Markov. Such a DAG then serves as an estimate for the underlying causal structure of the system. However, based on observational data alone, one can only learn an equivalence class of DAGs that contain the causal structure, known as a *Markov equivalence class*.

In many practical applications, such as in the medical and biological sciences, once a Markov equivalence class is learned, additional data is collected via *interventional experiments* (such as randomized controlled trials), which can then be used to distinguish the true causal system from within its Markov equivalence class. Such experiments typically target a subset of variables in the system, and the choice of these targets affects the number of elements within the class that can be rejected as candidates for the true causal system. Since such experiments are costly, it is desirable to have effective methods for identifying good candidates for targeting. One such method is to identify a single set of targets for individual intervention by estimating a set of possible source nodes in the true underlying causal system. Since $X_v \perp\!\!\!\perp X_w$ in a distribution Markov to a DAG $\mathcal{D}$ for any two source nodes v, w of $\mathcal{D}$, it is reasonable to identify the collection of all *marginally independent* nodes in the system; i.e., all pairs v, w for which $X_v \perp\!\!\!\perp X_w$. In a DAG model, the set of all such nodes fulfilling any such marginal independence statements will form a disjoint union of sets $A_1, \ldots, A_k$, each containing a unique

source node. Hence, identifying the marginal independence structure of a DAG model will yield a good set of targets for interventional experiments.

We consider the problem of estimating the marginal independence structure of a Bayesian network from observational data. A typical approach as seen in popular causal discovery algorithms (like the Greedy Equivalence Search) is to divide the space of DAGs into Markov Equivalence classes and then learn the essential graph representative of the class. We divide the space of DAGs with n nodes into unconditional equivalence classes (UECs), where two DAGs are said to be *unconditionally equivalent* if they have the same marginal independence statements. Now, for any DAG $\mathcal{D}$, the *unconditional dependence graph* (UDG) of $\mathcal{D}$ is the undirected graph $\mathcal{U}$ having the same number of vertices as $\mathcal{D}$, and (u, v) is an edge in $\mathcal{U}$ if and only if there exists a trek between u and v in $\mathcal{D}$. Each UEC can be uniquely represented by a UDG.

The unconditional dependence graphs can be uniquely characterized by using some graphical properties called *intersection* and *independence* number. We prove that an undirected graph is a UEC representative if and only their intersection and independence number are equal. This result allows us to provide a monomial representation to any UEC representative $\mathcal{U}$ as

$$\mathcal{U} = x_{i_1|A_1} x_{i_2|A_2} \cdots x_{i_k|A_k},$$

where k is the number of cliques in the minimal edge clique cover of $\mathcal{U}$ and i_j are the vertices which lie in exactly one maximal clique $i_j \cup A_j$. We define a monomial homomorphism ϕ which maps each $x_{i|A}$ to a product of indeterminants corresponding to the index $i|A$. We then obtain a Grobner basis for $ker(\phi)$ by constructing a *meet-join* ideal corresponding to ϕ. The motive behind this construction is that the binomials obtained in the Grobner basis can be used to move between any two graphs within a fiber. Now, in order to explore the space of UEC representatives beyond a given fiber, we introduce some new moves (binomials) which we name as *out-of-fiber add, out-of-fiber delete, merge* and *split*. The main result is as follows:

Theorem 1. *Let $\mathcal{U}_n$ be the collection of all the undirected graphs on n nodes representing nonempty UECs. Then for any two graphs $\mathcal{U}$ and $\mathcal{U}'$ in $\mathcal{U}_n$, there exists a sequence of within fiber, out-of-fiber, and/or merge and split moves that connects $\mathcal{U}$ and $\mathcal{U}'$.*

Using these moves, we implement an MCMC method called *GrUES (Gröbner-based Un-conditional Equivalence Search)* and applied to synthetic Gaussian data. We obtain that GrUES recovers the true marginal independence structure via a BIC-optimal or a *MAP* estimate at a higher rate than simple independence tests while also yielding an estimate of the posterior, for which the 20% HPD credible sets include the true structure at a high rate for graphs with density at least 0.5.

Towards standard imsets for maximal ancestral graphs

Robin Evans
(joint work with Zhongyi Hu)

Background. Imsets, introduced by Studený (see [5] for details), are an algebraic method for representing conditional independence models. Formally an imset is an integer-valued vector, with entries indexed by the power set of a collection of random variables V. The conditional independence $X_A \perp\!\!\!\perp X_B \mid X_C$ is represented by the *semi-elementary imset* $u_{\langle A,B|C\rangle} := \delta_C - \delta_{A\cup C} - \delta_{B\cup C} + \delta_{A\cup B\cup C}$, where δ_W is an imset with an entry 1 for the set W, and 0 elsewhere. An imset is said to be *structural* if, after multiplication by some natural number, it can be written as a sum of these semi-elementary imsets. In this case we say that the imset represents a model defined by the corresponding conditional independences.

Imsets have many attractive properties, and they can represent arbitrary probabilistic conditional independence models. They are particularly nice when applied to directed acyclic graph (DAG) models. These are models represented by graphs in which the vertices represent variables, and all edges are directed ($\rightarrow$) such that there are no directed cycles. Independences arise when an edge is not present between two vertices. DAG models are useful for causal inference, but have a disadvantage that two distinct DAGs may represent the same independence model; this issue is referred to as *Markov equivalence*.

In particular, the so-called *standard imset* for a DAG represents the model implied by that DAG exactly, and is invariant to the particular DAG that was used to generate it. Hence, the standard imset is also a label for the Markov equivalence class of the graph. An alternative label for the Markov equivalence class is the *characteristic imset*, which is given by a Möbius transform of the standard imset. It separately has the nice property that all the entries are either 0 or 1 for any DAG model.

Standard Imsets for Ancestral Graphs. DAGs are commonly used in causal inference, but a drawback is that they assume *causal sufficiency*, i.e. that all the causally important variables are observed. *Maximal ancestral graphs* are an extension to DAGs with an additional bidirected ($\leftrightarrow$) edge type, and they represent the conditional independence model induced by a DAG when some variables are unobserved. They are therefore very common in causal inference, because it is generally much more plausible not to assume that the system under study is causally sufficient. An example of this is shown in Figure 1. We present a proposed extension to standard imsets for MAG models, by first defining the characteristic imset using the parametrizing set representation of [3]. This is a very natural thing to do, because the two representations coincide for DAGs. Since the parametrizing set is also a label for the Markov equivalence class of a MAG, that property is retained by the new characteristic and standard imsets.

For the MAG in Figure 1, the standard imset turns out to be

$$\begin{aligned} u_{\mathcal{G}} &= \delta_{\emptyset} - \delta_{\{1,2\}} - \delta_{\{1,3\}} + \delta_{\{1,2,3\}} - \delta_{\{3,4\}} + \delta_{\{1,3,4\}} \\ &= \left\{\delta_{\emptyset} - \delta_{\{1,2\}} - \delta_{\{3\}} + \delta_{\{1,2,3\}}\right\} + \left\{\delta_{\{3\}} - \delta_{\{1,3\}} - \delta_{\{3,4\}} + \delta_{\{1,3,4\}}\right\} \\ &= u_{\langle 12,3\rangle} + u_{\langle 4,1|3\rangle}, \end{aligned}$$

where sets are abbreviated in the obvious way. The two independences $X_1, X_2 \perp\!\!\!\perp X_3$ and $X_4 \perp\!\!\!\perp X_1 \mid X_3$ are precisely the same as the MAG model that the graph $\mathcal{G}$ implies.

We show that the same pattern holds for many such graphs, though there is a subclass for which the method fails. We show that if the resulting imset is structural, then it always defines a subset of the independences implied by the graph. We also prove that the models are the same for what we term *simple MAGs*; that is, MAGs where any triple of the form $i \leftrightarrow k \leftrightarrow j$ has a directed path between i and j. It also works for a large class of purely bidirected models.

Scoring and Search. Andrews et al. [1] show that the maximum log-likelihood of a MAG model can be approximated by taking the inner-product of the vector of entropies over all subsets of variables, and our standard imset. Consequently, we can approximately compute the BIC in a much more efficient manner than would be possible by calculating the MLEs exactly. Given that the standard imset for simple MAGs represents precisely the model that the graph itself does, we develop a greedy score-based algorithm over this subclass.

The equivalence class of MAGs is determined by three things: which pairs of vertices are adjacent; which *unshielded* triples i, k, j (i.e. with k adjacent to i, j but i, j not adjacent) are *colliders* (i.e. both the edges have an arrowhead at k); and *discriminating paths*, which can determine the orientation of certain shielded triples. Therefore our search considers two graphs to be neighbours if they differ either by a single adjacency, or by whether an unshielded triple is a collider or not. If such a change creates a new discriminating path, then the two possible statuses for the shielded triple are also considered, and this recurses if necessary. Our results show that we do considerably better then a similar algorithm introduced by [2]. That algorithm involves computing the MLEs exactly, and their search is restricted in that, initially at least, it only allows new unshielded triples to be non-colliders.

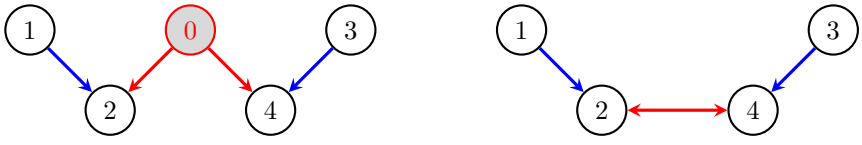

FIGURE 1. Left: a DAG with five vertices, where 0 represents a latent variable. Right: the MAG that corresponds to this DAG with 0 removed.

References

[1] Bryan J. Andrews, Gregory F. Cooper, Thomas S. Richardson, and Peter Spirtes. The m-connecting imset and factorization for ADMG models. *arXiv preprint* `https://arxiv.org/abs/2207.08963`, 2022.

[2] T. Claassen and I. G. Bucur. Greedy equivalence search in the presence of latent confounders. In *Proceedings of the 38th Conference on Uncertainty in Artificial Intelligence (UAI-22)*, volume 180, pages 443–452. PMLR, 2022.

[3] Z. Hu and R. J. Evans. Faster algorithms for Markov equivalence. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI-20)*, volume 124, pages 739–748. PMLR, 2020.

[4] Z. Hu and R. J. Evans. Towards imsets for maximal ancestral graphs. *arXiv preprint* `https://arxiv.org/abs/2208.10436`, 2022.

[5] Milan Studený. *Probabilistic Conditional Independence Structures*. Springer Verlag, 2005.

Algebraic methods for 3D genome reconstruction

Kaie Kubjas

(joint work with Diego Cifuentes, Jan Draisma, Oskar Henriksson and Annachiara Korchmaros)

The three-dimensional (3D) genome structure plays an important role in gene regulation. One of the main approaches to infer the 3D genome structure is from contact matrices that record interactions between different regions (loci) of the genome. In the case of diploid organisms the contact data is often unphased, which means that one cannot differentiate between contacts for homologous chromosomes. This talk is about partially-phased population contact data. Partially-phased contact data means that for some loci one can assign contacts to a maternal or paternal homolog. These loci are called unambiguous loci and the rest of the loci are called ambiguous loci. Population data means that the data is from a collection of cells.

We model the chromosome as a string of beads such that in the diploid case there is a pair of beads corresponding to one locus. We call a bead unambiguous (resp. ambiguous) if the corrsponding locus is unambiguous (resp. ambiguous). We assume that there is a power law $c_{ij} = d_{ij}^{\alpha}$, where $\alpha < 0$, between the pairwise distances d_{ij} between beads and corresponding contacts c_{ij}. The first main result of our work states that given positions of at least six unambiguous beads and a rational negative α, there are finitely many possibilities for the rest of the beads up to rigid transformations that give the observed contacts. A similar result holds if the contacts are noisy. We conjecture that in the noiseless case, knowing locations of seven unambiguous beads one can uniquely recover the locations of the rest of the beads.

In the fully ambiguous case when there are no unambiguous loci, we prove a similar result in the noiseless case for $\alpha = -2$. Namely, we show that given positions of at least twelve pairs of ambiguous beads, there are finitely many possibilities for the rest of the beads up to rigid transformations that give the observed contacts. The fully ambiguous case when $\alpha = 2$ was studied in [1].

In the second part of the talk, I discuss an algorithm to obtain a 3D reconstruction when $\alpha = -2$. This algorithm consists of four steps:

(1) A semidefinite programming based approach to recover the locations of unambiguous beads.
(2) A homotopy continuation based approach to recover the locations of ambiguous beads.
(3) A local optimization step with the initialization taken from step 2.
(4) A clustering step to disambiguate between beads in a bead pair.

The first steps solves the well-studied Euclidean distance geometry problem. We use the ChromSDE method from [2] to solve it. The second step uses local information from several sets of six unambiguous beads to reconstruct the ambiguous beads, motivated by the earlier finiteness result. In this step we choose several sets of six beads and use them to find all finitely many positions for the rest of the beads. When choosing different sets of six beads, then in general we get different solution sets for the ambiguous beads, but there is one solution that is similar in all these solution sets. This is the solution that one is looing for. Since the second step uses only local information, then we also run one step of local optimization with the initialization from the second step. The benefit of the third step is that it uses global information, but in general it can be difficult to find good initializations. Finally we use the observation that homologous chromosomes occupy different cell nucleus territories to assign beads in an ambiguous bead pair to a maternal or paternal chromosome.

References

[1] Anastasiya Belyaeva, Kaie Kubjas, Lawrence J Sun, and Caroline Uhler. Identifying 3d genome organization in diploid organisms via euclidean distance geometry. *SIAM Journal on Mathematics of Data Science*, 4(1):204–228, 2022.
[2] ZhiZhuo Zhang, Guoliang Li, Kim-Chuan Toh, and Wing-Kin Sung. 3d chromosome modeling with semi-definite programming and hi-c data. *Journal of computational biology*, 20(11):831–846, 2013.

On various (combinatorial) ways to describe faces of the cone of supermodular functions

Milan Studený

The motivation for this research is a long-term (ambitious) goal to characterize extreme rays of the cone of standardized *supermodular functions* in a combinatorial way so that one is able to determine/compute them in case of six or more basic variables, which is impossible by standard polyhedral-geometry-based procedures.

What was presented in the talk was the first gradual step on the way to the intended alternative computational procedure. The message is that the extreme rays of that cone *do have* relevant combinatorial interpretation. In fact, they admit several mutually equivalent combinatorial descriptions, which can be extended to all non-empty faces of the cone of supermodular functions.

Let N be a finite non-empty *basic set* of variables, intentionally regarded as an unordered set, while $[n] := \{1, \ldots, n\}$ denotes an ordered set of integers between 1 and $n \in \mathbb{N}$, where $n = |N|$. There are deeper reasons to distinguish N and $[n]$. The symbol $\mathcal{P}(N) := \{A : A \subseteq N\}$ will denote the power set of N.

Definition 1. *An* enumeration *of N is a one-to-one mapping $\varepsilon : [n] \to N$ from $[n]$ onto N. A record of the form $|\,\varepsilon(1)\,|\ldots|\,\varepsilon(n)\,|$ can specify such an enumeration. It can be interpreted as a total order on N in two ways: $\varepsilon(1) \prec \ldots \prec \varepsilon(n)$ is the ascending way, while $\varepsilon(1) \succ \ldots \succ \varepsilon(n)$ is the descending way. The set of (all) enumerations of N will be denoted by $\Upsilon(N)$. Notice that $\Upsilon(N)$ is not a group, unlike the set of permutations of N or the set of permutations of $[n]$.*
The rank vector *for an enumeration ε is the vector $[\varepsilon_{-1}(\ell)]_{\ell \in N}$ in $\mathbb{R}^N$, denoted by ρ_ε in a geometric context (of $\mathbb{R}^N$). The* permutohedron *in $\mathbb{R}^N$, denoted by $\Pi(N)$, is the convex hull of the rank vectors for (all) enumerations of N:*

$$\Pi(N) := \mathrm{conv}\,(\,\{\rho_\varepsilon \,:\, \varepsilon \in \Upsilon(N)\}\,)\,.$$

Every non-empty face of the permutohedron $\Pi(N)$ corresponds to an ordered partition of the basic set N into non-empty blocks [1].

Say that enumerations $\varepsilon, \eta \in \Upsilon(N)$ differ by an adjacent transposition *if there exists $1 \le i < n$ such that $\varepsilon(i) = \eta(i+1)$, $\varepsilon(i+1) = \eta(i)$ and $\varepsilon(k) = \eta(k)$ for remaining $k \in [n] \setminus \{i, i+1\}$. This is equivalent to the condition that $[\rho_\varepsilon, \rho_\eta]$ is a geometric edge (= 1-dimensional face) of $\Pi(N)$. One can interpret $\Upsilon(N)$ as an undirected* permutohedral graph *(over N) in which edges are determined by adjacent transpositions. This is a connected undirected graph and the following graphical concepts make sense for it. A* geodetic *between nodes $\varepsilon, \eta \in \Upsilon(N)$ is a walk between ε and η which has the shortest possible length among such walks. We say that a node $\sigma \in \Upsilon(N)$* is between nodes *$\varepsilon \in \Upsilon(N)$ and $\eta \in \Upsilon(N)$ if σ belongs to some geodetic between ε and η. A set $S \subseteq \Upsilon(N)$ is named* geodetically convex *if, for any $\varepsilon, \eta \in S$, all nodes between them belong to S.*

Using the method of Galois connections one is able to introduce a lattice of enumeration sets interpretable as posets on N. If $\varepsilon : [n] \to N$ is an enumeration of N and $(u, v) \in N \times N$ then we say that u strictly precedes *v in ε and write $u \prec_\varepsilon v$ if $\varepsilon_{-1}(u) < \varepsilon_{-1}(v)$. A set $S \subseteq \Upsilon(N)$ belongs to the* posets-based lattice *if either $S = \emptyset$ or there exists a (strict version of a) partial order $\prec$ on N such that*

$$S = \{\,\varepsilon \in \Upsilon(N) \,:\, u \prec_\varepsilon v \quad \textit{whenever } u \prec v\,\}\,.$$

Theorem 2. [7] *A set belongs to the posets-based lattice iff it is geodetically convex.*

Definition 3. *A* set function over *N is a map $w : \mathcal{P}(N) \to \mathbb{R}$, that is, $w \in \mathbb{R}^{\mathcal{P}(N)}$. A (transferable-utility coalitional)* game *is modeled by a set function w over N satisfying $w(\emptyset) = 0$. Given $w \in \mathbb{R}^{\mathcal{P}(N)}$ and a triplet $(A, B|C)$ of pairwise disjoint subsets of N, the respective "supermodular" difference expression is*

$$\Delta w\,(A, B|C) \;:=\; w(A \cup B \cup C) + w(C) - w(A \cup C) - w(B \cup C)\,.$$

A supermodular function *w over N can equivalently be defined by the condition $\Delta w\,(A, B|C) \geq 0$ for each triplet $(A, B|C)$ of pairwise disjoint subsets of N. The cone of* supermodular games *over N will be denoted by $\Diamond(N)$.*

Given $w \in \Diamond(N)$, its core *is a (bounded) polyhedron in $\mathbb{R}^N$ defined by*

$$C(w) \;:=\; \{\, [x_\ell]_{\ell\in N} \in \mathbb{R}^N \;:\; \sum_{\ell\in N} x_\ell = w(N) \;\&\; \sum_{\ell\in S} x_\ell \geq w(S) \text{ for any } S \subseteq N \,\}\,.$$

A classic result in game theory by Shapley [4] *is the characterization of the set* $\mathrm{ext}\,(C(w))$ *of vertices of the core of $w \in \Diamond(N)$ through enumerations. Given $w \in \Diamond(N)$ and $\varepsilon \in \Upsilon(N)$, the respective* marginal vector *in $\mathbb{R}^N$ is given by*

$$\varphi^w(\varepsilon) = [x_\ell]_{\ell\in N},\ \text{where } x_{\varepsilon(i)} \;:=\; w\,(\bigcup_{j=1}^{i} \varepsilon(j)\,) \;-\; w\,(\bigcup_{j=1}^{i-1} \varepsilon(j)\,) \text{ for } i \in [n]\,.$$

Lemma 4. [4],[6] *Given $w \in \Diamond(N)$, one has* $\mathrm{ext}\,(C(w)) = \{\, \varphi^w(\varepsilon) \;:\; \varepsilon \in \Upsilon(N) \,\}$.

Definition 5. *Given $w \in \Diamond(N)$, the* rank test *induced by w is*

$$\boldsymbol{\Upsilon}(w) \;:=\; \{\, \varphi^w_{-1}(x) \;:\; x \in \mathrm{ext}\,(C(w)) \,\},\ \text{being a partition of } \Upsilon(N).$$

If $x \in \mathbb{R}^N$ then $\mathcal{T}^w_x \;:=\; \{\, S \subseteq N \;:\; \sum_{\ell\in S} x_\ell = w(S) \,\}$ denotes the respective tightness class. The core structure *induced by w is defined by*

$$\mathbb{T}(w) \;:=\; \{\, \mathcal{T}^w_x \;:\; x \in \mathrm{ext}\,(C(w)) \,\},\ \text{being a covering of } \mathcal{P}(N).$$

Finally, the independence structure *induced by w is*

$$\mathcal{I}(w) \;:=\; \{\; (A, B|C) \;:\; \Delta w\,(A, B|C) = 0 \,\},$$
$$\text{being a class of triplets of pairwise disjoint subsets of } N.$$

Here are some characterizations of inclusion of faces of $\Diamond(N)$.

Theorem 6. [7] *Given $m, r \in \Diamond(N)$, the following conditions are equivalent:*

(i) *the face of $\Diamond(N)$ generated by m is included in the face generated by r,*
(ii) *the normal fan of $C(r)$ (in $\mathbb{R}^N$) refines the normal fan of $C(m)$,*
(iii) *the rank test $\boldsymbol{\Upsilon}(r)$ refines $\boldsymbol{\Upsilon}(m)$: $\forall\, S \in \boldsymbol{\Upsilon}(r)\ \exists\, T \in \boldsymbol{\Upsilon}(m) :\ S \subseteq T$,*
(iv) *the core structure $\mathbb{T}(r)$ refines $\mathbb{T}(m)$: $\forall\, \mathcal{T} \in \mathbb{T}(r)\ \exists\, \mathcal{S} \in \mathbb{T}(m) :\ \mathcal{T} \subseteq \mathcal{S}$,*
(v) *$\mathcal{I}(r) \subseteq \mathcal{I}(m)$: $\Delta r\,(A, B|C) = 0 \;\Rightarrow\; \Delta m\,(A, B|C) = 0$ for any $(A, B|C)$.*

Thus, various combinatorial objects allow one to describe (non-empty) faces of $\Diamond(N)$. It seems that the most appropriate ones from the "computational" point of view are the *rank tests*, alternatively be viewed as subgraphs of the permutohedral graph. The future research plan consists of three steps (= research directions):

(1) Characterize in combinatorial terms those rank tests $\boldsymbol{\Upsilon}$ for which a supermodular game $w \in \Diamond(N)$ exists such that $\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}(w)$.
(2) Characterize sub-maximal such ranks tests: these precisely correspond to the extreme rays of the standardized supermodular cone.
(3) Propose an algorithm for generating *all* such sub-maximal rank tests.

Several partial results in this context have already been achieved. For example, it was shown that $S \subseteq \Upsilon(N)$ can occur as a component of $\boldsymbol{\Upsilon}(w)$ for $w \in \Diamond(N)$ iff it is geodetically convex. Additionally, a simple efficient linear criterion to recognize sub-maximal rank tests formulated in terms of core structures in available [6].

Results by Morton *et al.* [3] say that (supermodular) rank tests correspond to particular independence structures, namely semi-graphoids, known as *structural semi-graphoids*. A related question from [5] was whether every sub-maximal semi-graphoid is structural. Nevertheless, Hemmecke *et al.* [2] showed that this is not the case in case of 5 basic variables. As a part of the talk an even more specific (and relevant) open question was raised:

Question 7. *Is it true that every sub-maximal structural semi-graphoid over N, where $|N| = 5$, is also a sub-maximal semi-graphoid?*

The question was answered negatively by Tobias Boege with help of his computer programme. The supermodular game w over $N = \{a, b, c, d, e\}$ inducing the counter-example is as follows: $w(N) = 7$, $w(S) = 4$ for $S \subseteq N$ with $|S| = 4$,

- $w(S) = 1$ for S of the form $\{a, b, c\}$, $\{a, b, d\}$, $\{a, b, e\}$, $\{a, e\}$, $\{b, c\}$, $\{b, d\}$,
- $w(S) = 2$ for other $S \subseteq N$ with $|S| = 3$, and
- $w(S) = 0$ for the remaining sets $S \subseteq N$.

Acknowledgements. The work on this topic has been supported from a former GAČR project n. 19-04579S.

References

[1] L. J. Billera, A. Sarangarajan. The combinatorics of permutation polytopes. In *Formal Power Series and Algebraic Combinatorics*, DISMACS Series in Discrete Mathematics and Theoretical Computer Science 24, AMS, Providence 1996, pp. 1–23.

[2] R. Hemmecke, J. Morton, A. Shiu, B. Sturmfels, O. Wienand. Three counter-examples on semi-graphoids. *Combinatorics, Probability and Computing* 17 (2008) 239–257.

[3] J. Morton, L. Pachter, A. Shiu, B. Sturmfels, O. Wienand. Convex rank tests and semi-graphoids. *SIAM Journal on Discrete Mathematics* 23 (2009) 1117–1134.

[4] L. S. Shapley. Cores of convex games. *International Journal of Game Theory* 1 (1971/72) 11–26.

[5] M. Studený. *Probabilistic Conditional Independence Structures*. Springer, London 2005.

[6] M. Studený, T. Kroupa. Core-based criterion for extreme supermodular functions. *Discrete Applied Mathematics* 206 (2016) 122–151.

[7] M. Studený. On combinatorial descriptions of faces of the cone of supermodular functions. Manuscript in preparation, 2023.

The singular locus for linear compartment models and structural equation models

Elizabeth Gross

(joint work with Mark Curiel, Nicolette Meshkat, and Anne Shiu)

Mathematical models, whether deterministic or probabilistic, are often parametric with parameter values that need to be inferred from the data. The parameter identifiability problem concerns the question of which parameters can be determined uniquely from data. In this talk, we explore identifiability for two types of models used in epidemiology [3] that are described by directed graphs: linear compartment models and structural equation models. This talk is based on joint work with Nicolette Meshkat and Anne Shiu [2] and preliminary work with Mark Curiel and Nicolette Meshkat.

Linear compartment models are systems of parameterized linear ordinary differential equations. They are described by an underlying directed graph, where the vertices represent compartments and edges represent exchanges or flows between compartments, and two distinguished subsets of vertices, one subset representing inputs and, the other, outputs. In epidemiology, the compartments may represent groups of infected, susceptible, and recovered individuals, and the edges represent progression between the compartments.

For linear compartment models, using standard differential algebra techniques, the question of whether a given model is generically locally identifiable is equivalent to asking whether the Jacobian matrix of a certain coefficient map, arising from input-output equations, is generically full rank. A natural next step is to study the set of parameter values where the Jacobian matrix drops in rank, which we refer to as the *locus of non-identifiable parameter values*, or, for short, *the singular locus*. In this talk, we discuss how a defining equation of the singular locus can be used to determine when submodels are generically locally identifiable. We also give the singular-locus equation for two families of linear compartmental models, cycle and mammillary (star) models with input and output in a single compartment. We also state a conjecture for the corresponding equation for a third family: catenary (path) models [2].

Conjecture 1. *Assume $n \geq 2$. For the n-compartment catenary (path) model in Figure 1, the equation of the singular locus is:*

$$a_{12}^{n-1}(a_{21}a_{23})^{n-2}(a_{32}a_{34})^{n-3}\dots(a_{n-1,n-2}a_{n-1,n}) \ . \tag{1}$$

After exploring the singular locus for linear compartmental models, we then turn to structural equation models, another class of models used to understand causal processes [1]. Linear structural equation models are multivariate statistical models encoded by mixed graphs. For structural equation models, the local identifiabilty question amounts to testing if the parameters of the model can be determined from the covariance matrix, or more precisely, if the mapping from the parameters to the covariance matrix entries is generically finite-to-one. This amounts to checking the rank of the Jacobian of this mapping evaluated at a generic point. However, we

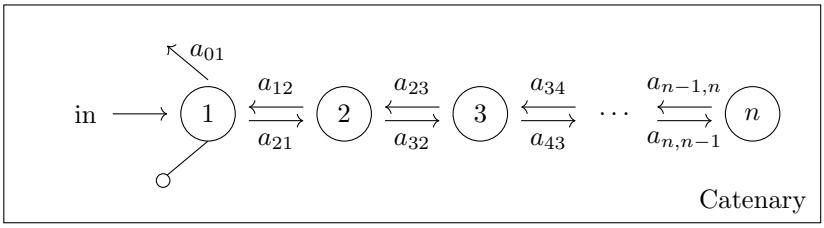

FIGURE 1. The **catenary** (path) model with n compartments, in which compartment 1 has an input, output, and leak.

can ask the same question that we asked for linear compartment models: for what parameter values does this Jacobian matrix drop in rank? Analogously to linear compartment models, we can define the singular locus, whose defining equations are given by the determinant of the Jacobian matrix when it is square, or more generally, by its minors. We conclude the talk by considering some families of graphs and finding the corresponding formulas for the singular locus equations.

REFERENCES

[1] Mathias Drton et al. Algebraic problems in structural equation modeling. *The 50th anniversary of Gröbner bases*, pages 35–86, 2016.

[2] Elizabeth Gross, Nicolette Meshkat, and Anne Shiu. Identifiability of linear compartmental models: The singular locus. *Advances in Applied Mathematics*, 133:102268, 2022.

[3] Joshua Havumaki and Marisa C Eisenberg. Using compartmental models to simulate directed acyclic graphs to explore competing causal mechanisms underlying epidemiological study data. *Journal of the Royal Society Interface*, 17(167):20190675, 2020.

Likelihood Geometry of Correlation Models

CARLOS AMÉNDOLA
(joint work with Piotr Zwiernik)

We consider centered Gaussian models $\mathcal{M} \subset \mathcal{S}_+^n$ where $\mathcal{S}_+^n$ denotes the set of $n \times n$ positive definite matrices, so that $\Sigma \in \mathcal{M}$ represents a Gaussian correlation matrix, i.e., a covariance matrix with $\Sigma_{ii} = 1$ for $i = 1, \ldots, n$. We study the geometry of the maximum likelihood estimation problem for the model $\mathcal{M}$ consisting of all $n \times n$ correlation matrices as well as linear submodels that encode additional symmetries.

The bivariate case $n = 2$ of $\mathcal{M} = \left\{ \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \mid -1 < \rho < 1 \right\}$ is classical in statistics, e.g. [2, 3]. If the sample covariance matrix $S = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$ then we determine exactly when the likelihood function has three real critical points (regions I, II, IV in Figure 1) and precisely when they are positive definite (region I in Figure 1).

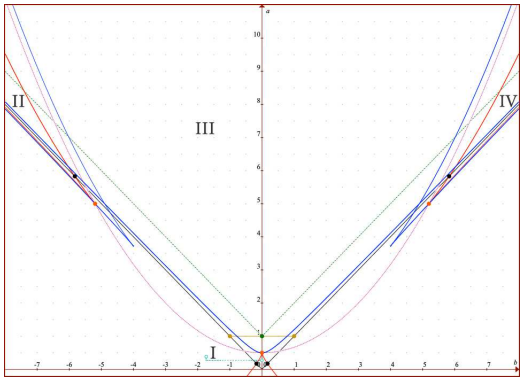

FIGURE 1. Likelihood geometry of the bivariate correlation model.

The *ML degree* for $n = 2$ is 3, but the *dual ML degree* is only 2. We also introduce the SSL degree as a new measure of algebraic complexity, which is 4 for $n = 2$. The following table shows these degrees for small n.

n	1	2	3	4	5	6	7	8	9
SSL degree	1	4	28	292	?	?	?	?	?
ML degree	1	3	15	109	1077	13695	?	?	?
dual ML degree	1	2	5	14	43	144	522	2028	8357

A natural open question is, can we find a (recursive) formula for these sequences?

We also prove that for the equicorrelation model all three degree invariants remain constant with n.

More details about this talk are found in the joint work with Piotr Zwiernik (University of Toronto) [1].

REFERENCES

[1] Carlos Améndola, Piotr Zwiernik (2021). *Likelihood Geometry of Correlation Models*. Le Matematiche 76(2), pp. 559–583.

[2] Ole E. Barndorff-Nielsen, David R. Cox (1994). *Inference and asymptotics*, Chapman and Hall.

[3] Maurice G. Kendall, Alan Stuart (1967). *The advanced theory of statistics: Inference and relationship*, Vol. 2. London.

Toric Ideals of Characteristic Imsets via Quasi-Independence Gluing

Benjamin Hollering
(joint work with Joseph Johnson, Irem Portakal, and Liam Solus)

Given a directed acyclic graph (DAG) $\mathcal{G} = ([n], E)$ with vertices $[n] := \{1, \dots, n\}$ and edges E and a collection of jointly distributed random variables $(X_1, \dots, X_n)$ with probability density function $f(x_1, \dots, x_n)$, we say that $f(x_1, \dots, x_n)$ is *Markov* to the DAG $\mathcal{G}$ if

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i | x_{\mathrm{pa}_{\mathcal{G}}(i)}),$$

where $\mathrm{pa}_{\mathcal{G}}(i) = \{j \in [n] : j \to i \in E\}$ is the set of *parents* of i in $\mathcal{G}$. A central problem in causal discovery is to recover an unknown DAG whose joint distribution best represents observational data but it is well known that two different DAGs may encode the same set of conditional independence statements, in which case the DAGs are called Markov equivalent. Recently, a new geometric perspective on causal discovery algorithms has emerged which uses *characteristic imsets* (CIM) [4] to embed DAGs in Euclidean space in the following way.

Definition 1. *Given a DAG $\mathcal{G} = ([n], E)$, the* characteristic imset *of $\mathcal{G}$ is*

$$c_{\mathcal{G}} : \{S \subseteq [n] : |S| \geq 2\} \longrightarrow \{0, 1\};$$

$$c_{\mathcal{G}} : S \longmapsto \begin{cases} 1 & \text{if there exists } i \in S \text{ such that } S \setminus \{i\} \subseteq \mathrm{pa}_{\mathcal{G}}(i), \\ 0 & \text{otherwise.} \end{cases}$$

As is discussed in [2, 3], characteristic imsets allow us to rephrase the problem of causal discovery as a linear programming problems over the polytope

$$\mathrm{CIM}_n := \mathrm{conv}(c_{\mathcal{G}} \in \mathbb{R}^{2^n - n - 1} \mid \mathcal{G} = ([n], E) \text{ a DAG}).$$

In my talk I introduced the characteristic imset ideal I_G which is the toric ideal associated to the lattice polytope $\mathrm{CIM}_G = \mathrm{conv}(c_{\mathcal{G}} : \mathcal{G} \text{ has skeleton } G)$ where $G = ([n], E)$ is an undirected graph. We introduced and studied this object because the algebraic structure of a toric ideal can be helpful for understanding the polyhedral structure of the associated polytope. Our main focus was to determine a Gröbner basis for the ideal I_G. To do this we introduced a new operation on homogeneous ideals called a *quasi-independence gluing* (QIG) which generalizes the *toric fiber product*.

Definition 2. *Let $Q \subset [r] \times [s]$ and $I \subset \mathbb{K}[x_j \mid j \in [r]]$ and $J \subset \mathbb{K}[y_k \mid k \in [s]]$ be homogeneous ideals. The* quasi-independence gluing *of I and J with respect to Q is*

$$I \times_Q J := \phi_Q^{-1}(I + J)$$

where ϕ_Q is the map

$$\phi_Q : \mathbb{K}[z_{jk} \mid (j, k) \in Q] \to \mathbb{K}[x_j, y_k \mid j \in [r], k \in [s]]$$
$$z_{jk} \mapsto x_j y_k.$$

We showed that under a technical combinatorial condition called Q-homogeneity, that a Gröbner basis for $I \times_Q J$ can be computed by *lifting* the Gröbner bases of I and J which generalizes Sullivant's similar results for toric fiber products [5].

Theorem 3. [1] *Let $F \subseteq I$ be a weakly Q-homogeneous Gröbner basis for I with respect to the weight ω_1 and $G \subseteq J$ be a weakly Q-homogeneous Gröbner basis for J with respect to the weight ω_2. Then*

$$\mathrm{Lift}(F) \cup \mathrm{Lift}(G) \cup H_Q$$

is a pseudo-Gröbner basis for $I \times_Q J$ with respect to the weight $\omega^T B_Q$ where $B_Q \in \mathbb{Z}^{(r+s)\times \#Q}$ is the matrix of exponents of the map ϕ_Q and $\omega = (\omega_1, \omega_2)$.

Lastly, we showed that this operation can be used iteratively compute a Gröbner basis for the characteristic imset ideal I_G when G is a tree and thus one can quickly compute a Gröbner basis for I_G via repeated lifting.

Theorem 4 ([1])**.** *Let $T = ([p], E)$ be a tree, $e = u - v$ be a non-leaf edge of T. Then $I_T = I_{T_u} \times_Q I_{T_v}$ where Q is the set $Q = \{\mathrm{part}(\mathcal{T}, e) \mid \mathcal{T} \in \mathrm{Pat}(T)\}$.*

Corollary 5 ([1])**.** *The Gröbner basis of I_T for any tree T can be quickly determined via iterated quasi-independence gluing of star tree ideals.*

References

[1] Benjamin Hollering, Joseph Johnson, Irem Portakal, and Liam Solus. Toric ideals of characteristic imsets via quasi-independence gluing, 2022.

[2] Svante Linusson, Petter Restadh, and Liam Solus. Greedy causal discovery is geometric. *To appear in SIAM Journal on Discrete Mathematics*, 2022.

[3] Svante Linusson, Petter Restadh, and Liam Solus. On the edges of characteristic imset polytopes. *arXiv preprint* `https://arxiv.org/abs/2209.07579`, 2022.

[4] Milan Studený, Raymond Hemmecke, and Silvia Lindner. Characteristic imset: a simple algebraic representative of a bayesian network structure. In *Proceedings of the 5th European workshop on probabilistic graphical models*, pages 257–264. HIIT Publications, 2010.

[5] Seth Sullivant. Toric fiber products. *J. Algebra*, 316(2):560–577, 2007.

Non-independent component analysis

Piotr Zwiernik

(joint work with Geert Mesters)

Consider the equation $AY = X$, where X, Y are random vectors in $\mathbb{R}^d$ and $A \in \mathbb{R}^{d\times d}$ is an invertible matrix. Here the random variable X is not-directly observed but we assume that its components are independent. Moreover, without loss of generality we can assume $\mathbb{E}X = 0$ and $\mathrm{var}(X) = I_d$. In practice the goal is to recover the matrix A from the observations of Y. Therefore, the fundamental identifiability question is whether it is possible to identify A knowing the distribution of Y. For example, knowing the covariance $\mathrm{var}(Y) = \Sigma$ of Y allows us to write

$$A^\top A \;=\; \Sigma^{-1},$$

which shows that from the second order moments of Y we can identify A up to

$$\{QA: \ Q \text{ is orthogonal}\}.$$

However, if Y has Gaussian distribution then we cannot do better.

A seminal result in the literature on the Independent Component Analysis states that, if at most one component of X is Gaussian, then A is identified up to sign and permutation from the distribution of Y (Comon, 1994). This means that we are not able to exactly identify A but we identify the finite set

$$\{QA: \ Q \text{ is a signed permutation}\}.$$

This is satisfactory because all matrices in this finite set look similar.

Our study of this problem was motivated by relaxing the independence assumption on the components of X. There is a big gap between X having uncorrelated components ($\text{var}(X) = I_d$) and these components being independent. As we noted, in several interesting situations independence is rather special. For example, in elliptical distributions if some components are independent it follows that the underlying vector must be Gaussian (Kelker 1970).

In this paper we replace the independence assumption with restrictions on the higher order moments/cumulants of X. Conceptually the idea is simple. Using the covariance of Y we know that every candidate for A is of the form QA for an orthogonal matrix Q. Since $AY = X$, for QA to remain a valid candidate we need that in the equation $(QA)Y = QX$ the random vector QX satisfies the given conditions on X. Clearly $\mathbb{E}(QX) = 0$ and $\text{var}(QX) = I_d$ so the two basic assumptions are always satisfied. The idea now is to impose restrictions on higher order moments/cumulants of X so that QX will not satisfy these restrictions unless Q is a signed permutation matrix. In this way we can recover a more algebraic version of Comon's result.

Let $T \in S^r(\mathbb{R}^d)$ be a symmetric $d \times \cdots \times d$ real tensor of order r. Here T will represent the tensor of all moments/cumulants of X of order r. It is important to note that the moments/cumulants of QX are obtained from T by the standard multilinear action, and so they are equal to $(Q, \ldots, Q) \cdot T$, where

$$[(Q, \ldots, Q) \cdot T]_{i_1 \cdots i_r} = \sum_{j_1, \ldots, j_r} Q_{i_1 j_1} \cdots Q_{i_r j_r} T_{j_1 \cdots j_r}.$$

We consider zero restrictions on T. This can be motivated in various ways. For example, if the components of X are independent then T must be a diagonal tensor. If the components of X are mean independent ($\mathbb{E}(X_i|X_j) = \mathbb{X}_i$ for all $i \neq j$) then the cumulant tensor T satisfies $T_{ij\cdots j} = 0$ for all $i \neq j$. So the basic question is as follows: Suppose that $V \subset S^r(\mathbb{R}^d)$ is a linear subspace defined by some coordinates being zero. Let T be a generic element of V. For which V the equation $(Q, \ldots, Q) \cdot T \in V$ holds only for sign permutation matrices? We show that this property holds, for example, when V is the set of diagonal tensors. In this case the genericity condition on T requires that at most one diagonal entry of T is zero. Another example is given by, what we call, reflectionally invariant tensors. These are the tensors $W \subset S^r(\mathbb{R}^d)$ such that $T \in W$ if and only if

$(D, \ldots, D) \cdot T \in W$ for every signed diagonal matrix D (note that r must be even for this set to be non-trivial).

Triangulations of Cosmological Polytopes

Liam Solus
(joint work with Martina Juhnke-Kubitze, and Lorenzo Venturello)

Arkani-Hamed, Benincasa and Postnikov [1] introduced the cosmological polytope of a connected, undirected graph $G = (V, E)$, where V is a finite set of *vertices* and E is a finite collection of *edges*; i.e., pairs ij such that $i, j \in V$. From the perspective of physics, the graph G can be interpreted as a Feynman diagram, in which case the cosmological polytope provides a geometric model for the computation of the contribution of the system G to the wavefunction of certain cosmological models. Such models permit us to assume that the graph G is connected. However, no additional constraints need to be placed on G. For instance, E can be a finite multiset of edges; that is, G need not be simple.

The cosmological polytope of the graph G is a convex polytope residing in the real-Euclidean space $\mathbb{R}^{|V|+|E|}$, for which we denote the standard basis vectors as x_i and x_e for all $i \in V$, $e \in E$. The *cosmological polytope* of G is then defined as

$$\mathcal{C}_G = \text{conv}\{x_i + x_j - x_e, x_i - x_j + x_e, -x_i + x_j + x_e \ : \ e = ij \in E\}.$$

The relevance of the polytope $\mathcal{C}_G$ to the computation of wavefunctions for cosmological models arises via a recently established connection between scattering amplitudes and a generalization of convex polytopes called *positive geometries* [2]. Positive geometries (and hence convex polytopes) admit a unique differential form having only logarithmic singularities along the boundary of the positive geometry [2]. This form is called its *canonical form*. We denote the canonical form of a positive geometry X by Ω_X. In the case of the cosmological polytope, having the canonical form of $\mathcal{C}_G$ is sufficient to compute the contribution of the Feynman diagram G to the wavefunction of interest. Hence, it is desirable to have methods for computing the canonical form of $\mathcal{C}_G$ for any graph G.

One approach to computing the canonical form of a positive geometry X that has been applied successfully in various settings is to consider a subdivision of X, say $Y_1, \ldots, Y_m$ [4] . We then have that

$$\Omega_X = \Omega_{Y_1} + \cdots + \Omega_{Y_m}.$$

In the case that X is a convex polytope, the subdivision $Y_1, \ldots, Y_m$ is a polyhedral subdivision of X. Hence, to compute the canonical form of $\mathcal{C}_G$ one could identify a subdivision of this polytope.

From the algebraic perspective, perhaps the most natural approach to finding such a subdivision for every G is to identify a Gröbner basis for the toric ideal associated to the lattices points in the polytope $\mathcal{C}_G$. The initial terms of the generators in such a basis correspond to the minimal non-faces of a regular unimodular triangulation of $\mathcal{C}_G$ [5]. A triangulation also has the advantage that all facets

$Y_1, \ldots, Y_m$ of the subdivision are simplices. This feature turns out to be useful as it implies that the canonical form of each Y_i may be expressed as

$$\Omega_{Y_i} = \frac{\omega}{f_1 \cdots f_r},$$

where $f_1, \ldots, f_r$ are the facet-defining equations of the simplex Y_i, and ω is an associated regular form [2]. Since the facet-defining equations of Y_i are easily recoverable from its set of vertices, the above results give a quick recipe for computing the canonical form of the polytope $\mathcal{C}_G$.

In this talk, we present the following results relevant to this method.

Theorem 1. *The cosmological polytope $\mathcal{C}_G$ of any undirected, connected graph G has a regular unimodular triangulation.*

The proof of Theorem 1 is constructive. We identify a family of term orders, called *good term orders*, that can be shown to exist for every G. For such term orders we recover a Gröbner basis for the toric ideal associated to $\mathcal{C}_G$ with squarefree initial ideal.

While this approach yields the existence of such a triangulation, the presentation is in the form of the minimal non-faces, as opposed to an explicit characterization of the facets of the triangulation. The latter is needed in order to compute the desired canonical forms. We show that such characterizations can be found in terms of certain decorated graphs for special instances, including the path graph, cycles and trees. In the case of paths and cycles, the characterizations admit straightforward enumeration, allowing us to recover the normalized volume of the polytope $\mathcal{C}_G$ for these graphs.

Theorem 2. *The normalized volume for the cosmological polytope $\mathcal{C}_{I_{n+1}}$ for the path on $n+1$ vertices is 4^n, and the normalized volume of $\mathcal{C}_{C_n}$ for the n-cycle C_n on n vertices is $4^n - 2^n$.*

These formulas tell us the number of summands in our computation of the canonical form of $\mathcal{C}_G$. In the case of the path, this result recovers the formula previously identified by Kühne and Monin [3]. The formula for the n-cycle extends their results.

A note on the connection to statistics. While the results in this talk pertain mainly to quantum physics, the results are presented at this workshop in hopes to stimulate interest in this topic related to possible causal interpretations of these polytopes. In [1], it is noted that, instead of interpreting G as a Feynman diagram, one could interpret the graph as a discretized representation of the intersection of light cones. The notion of causation in physics is typically studied from this perspective; where an event may have a causal effect on another event if the first event lies in the past light cone of the second. The interpretation of the cosmological polytope from this perspective is presented in [1]. Given that the statistical community has made substantial progress in the last few decades towards an ever-improving probabilistic theory of causality, one may hope that statisticians studying causal inference find use for the cosmological polytope. Such uses may

provide insights into a long-standing endeavour in quantum physics: explaining how causation is embedded in a Feynman diagram.

References

[1] Nima Arkani-Hamed, Paolo Benincasa, and Alexander Postnikov. *Cosmological polytopes and the wavefunction of the universe.* arXiv preprint arXiv:1709.02813, 2017.

[2] Nima Arkani-Hamed, Yuntao Bai, and Thomas Lam. *Positive geometries and canonical forms.* Journal of High Energy Physics, 2017(11):1–124, 2017.

[3] Lukas Kühne and Leonid Monin. *Faces of cosmological polytopes.* arXiv preprint arXiv:2209.08069, 2022.

[4] Thomas Lam. *An invitation to positive geometries.* arXiv preprint arXiv:2208.05407, 2022.

[5] Bernd Sturmfels. *Grobner bases and convex polytopes*, volume 8. American Mathematical Soc., 1996.

Moment Varieties of Linear Non-Gaussian Graphical Models

Alexandros Grosdos

(joint work with Carlos Améndola, Mathias Drton, Roser Homs, and Elina Robeva)

Featuring prominently in a variety of applications, directed graphical models (DAGs) [2] capture intuitive cause-effect relations among a set of random variables by hypothesizing that each variable is a noisy function of its causes. For a number of statistical tasks, such as model selection, it has proven useful to obtain insights about the algebraic structure of the moments of the joint distributions in the graphical model for a given graph [1]. A prominent example are results on algebraic relations among second moments, i.e., covariances, in models that postulate linear functional relationships among the variables [4]. This is a first systematic study on algebraic relations that also involve higher moments of such a model.

Let $G = (V, E)$ be DAG, and let $(X_i, i \in V)$, be a collection of random variables that represent statistical observations indexed by the vertices in V. The graph G gives rise to the *linear structural equation model* consisting of the joint distributions of all random vectors $X = (X_i, i \in V)$ such that

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

where the ε_i are mutually independent random variables representing stochastic errors. The errors are assumed to have expectation $\mathbb{E}[\varepsilon_i] = 0$, finite variance $\omega_i^{(2)} = \mathbb{E}[\varepsilon_i^2] > 0$, and finite third moment $\omega_i^{(3)} = \mathbb{E}[\varepsilon_i^3]$. The coefficients λ_{ji} are unknown real-valued parameters, and we fill them into a matrix $\Lambda = (\lambda_{ji}) \in \mathbb{R}^{V \times V}$ by adding a zero entry when $(j, i) \notin E$.

The model. The covariance matrix S and the third moment tensor T of X are given by

$$S = (s_{ij}) = (I - \Lambda)^{-T}\Omega^{(2)}(I - \Lambda)^{-1},$$
$$T = (t_{ijk}) = \Omega^{(3)} \bullet (I - \Lambda)^{-1} \bullet (I - \Lambda)^{-1} \bullet (I - \Lambda)^{-1},$$

where $\Omega^{(2)}$ is a diagonal matrix in the set $PD(V)$ of positive-definite matrices and similarly $\Omega^{(3)} \in \mathrm{Sym}_3(V)$ is a diagonal symmetric tensor. Here $\bullet$ denotes the Tucker product, see [3].

Let $G = (V, E)$ be a DAG. The *third-order moment model* of G is the set $\mathcal{M}^{\leq 3}(G)$ that comprises all pairs of covariance matrices and third moment tensors that are realizable under the linear structural equation model given by G. That is,

$$\mathcal{M}^{\leq 3}(G) = \big\{\big((I - \Lambda)^{-T}\Omega^{(2)}(I - \Lambda)^{-1},\ \Omega^{(3)} \bullet (I - \Lambda)^{-1} \bullet (I - \Lambda)^{-1} \bullet (I - \Lambda)^{-1}\big) :$$
$$\Omega^{(2)} \in PD(V) \text{ diagonal},\ \Omega^{(3)} \in \mathrm{Sym}_3(V) \text{ diagonal},\ \Lambda \in \mathbb{R}^E\big\}.$$

Furthermore, the *third-order moment ideal* of G is the ideal $\mathcal{I}^{\leq 3}(G)$ of polynomials in the entries $S = (s_{ij})$ and $T = (t_{ijk})$ that vanish when $(S, T) \in \mathcal{M}^{\leq 3}(G)$.

Simple Trek Parametrization. By introducing a new set of indeterminates a_i, b_i for each vertex in the graph, we obtain the shorter *simple trek rule parametrization*:

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

where $\mathcal{T}$ is the set of all simple treks. Then

Proposition 1 (Simple trek rule)**.** *Let $G = (V, E)$ be a DAG, and let ϕ_G the ring morphism above. Then the map ϕ_G induces a parametrization of the model $\mathcal{M}^{\leq 3}(G)$, and, therefore, $\mathcal{I}^{\leq 3}(G) = \ker \phi_G$.*

Corollary 2. *If G is a polytree, then the ideal $\mathcal{I}^{\leq 3}(G)$ is toric.*

Low-Rank Trek-Matrices. Let G be a polytree. Let $i, j \in V$ be two vertices such that a 2-trek between i and j exists. We define the *trek-matrix* between i and j as

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

where

- $k_1, \ldots, k_r$ are vertices such that $\mathrm{top}(i, k_a) = \mathrm{top}(j, k_a)$ for $a = 1, \ldots, r$, and
- $(l_1, m_1), \ldots, (l_q, m_q)$ are such that $\mathrm{top}(i, l_b, m_b) = \mathrm{top}(j, l_b, m_b)$ for $b = 1, \ldots, q$.

The next result explains how to cut out the variety.

Theorem 3. *Let G be a polytree and J be the ideal generated by s_{ij} such that no 2-trek between i and j exists, t_{ijk} such that no 3-trek between i, j and k exists, as well as all 2-minors of the matrices $A_{i,j}$ for $i \to j \in E$. Then,*

$$\mathcal{M}^{\leq 3}(G) = \mathcal{V}(J) \cap (PD(V) \times \mathrm{Sym}_3(V)).$$

To obtain the generators of the ideal one needs further polynomials arising from minors:

Theorem 4. *The third-order moment ideal $\mathcal{I}^{\leq 3}(G)$ of the model $\mathcal{M}^{\leq 3}(G)$ is generated by the linear generators of $\mathcal{I}^{\leq 3}(G)$ and the minors of matrices A_{ij} for all i, j such that there exists a trek between them.*

References

[1] Mathias Drton. Algebraic problems in structural equation modeling. In *The 50th anniversary of Gröbner bases*, volume 77 of *Adv. Stud. Pure Math.*, pages 35–86. Math. Soc. Japan, Tokyo, 2018.

[2] Marloes Maathuis, Mathias Drton, Steffen Lauritzen, and Martin Wainwright, editors. *Handbook of graphical models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL, 2019.

[3] Elina Robeva and Jean-Baptiste Seby. Multi-trek separation in linear structural equation models. *SIAM J. Appl. Algebra Geom.*, 5(2):278–303, 2021.

[4] Seth Sullivant. Algebraic geometry of Gaussian Bayesian networks. *Advances in Applied Mathematics*, 40(4):482–513, 2008.

Weak maximum likelihood threshold of colored Gaussian graphical models

Roser Homs

(joint work with Olga Kuznetsova)

Given an undirected Gaussian graphical model, a very natural question arises: how many observations do we need to ensure the existence of the maximum likelihood estimator (MLE)? In 1993, Buhl [1] proved that, when the underlying graph is chordal, the MLE exists with probability one whenever the sample size is at least the maximal clique size, and it cannot exist otherwise. However, when the graph is not chordal, existence of the MLE is not completely understood in general for a certain range of observations.

In particular, it can occur that its probability is strictly between zero and one. Given a graph, the minimal number of observations from which the MLE exists almost surely is called the *maximum likelihood threshold* (MLT). If we only require existence with strictly positive probability, we talk about the *weak maximum likelihood threshold* (WMLT).

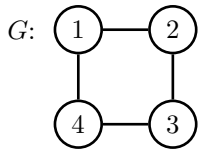

FIGURE 1. In the graphical model arising from a 4-cycle, the MLE exists with probability $p = 1$ for at least 3 observations, with $p \in (0, 1)$ for 2 observations and it does not exist with 1 observation ($2 = \mathrm{WMLT}(G) < \mathrm{MLT}(G) = 3$).

Colored graphical models. Colored graphical models are linear concentration models arising from undirected graphs with a coloring in its vertices and edges. More formally, a coloring of the graph $G = (V, E)$ with r different colors in the vertices and s different colors in the edges yields partitions $V = V_1 \sqcup V_2 \sqcup \cdots \sqcup V_r$ and $E = E_1 \sqcup E_2 \sqcup \cdots \sqcup E_s$. Let $\mathcal{K}_G$ be the set of all concentration matrices in the cone of positive definite matrices $\mathbb{S}^m_{\succ 0}$ that satisfy $K_{\alpha\beta} = 0$ whenever the edge $\alpha - \beta$ is not present in the graph and, additionally, equalities $K_{\alpha\alpha} = K_{\beta\beta}$ hold for any pair of vertices α, β in the same vertex color class V_i (resp. $K_{\alpha\beta} = K_{\gamma\delta}$ holds for any pair of edges $\alpha - \beta, \gamma - \delta$ in a common color class E_j).

From the statistical perspective, equal colors in the vertices yield equal inverse partial variances of the corresponding random variables and equal colors in the edges result in equal conditional independence restrictions. These additional symmetries allow for the existence of the MLE with fewer observations than in the case of the model arising from the corresponding uncolored graph.

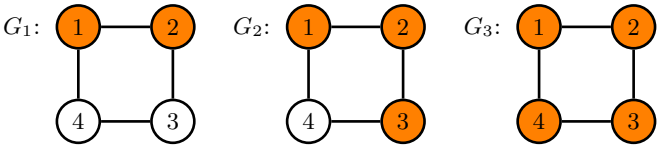

FIGURE 2. Adding symmetries to the vertices of the 4-cycle yields all remaining combinations of WMLT and MLT in 4-cycles: $\mathrm{WMLT}(G_1) = \mathrm{MLT}(G_1) = 2$, $1 = \mathrm{WMLT}(G_2) < \mathrm{MLT}(G_2) = 2$ and $\mathrm{WMLT}(G_3) = \mathrm{MLT}(G_3) = 1$.

Maximum likelihood thresholds have been investigated using algebraic geometry [2], rigidity theory [3] or the score matching estimator [4] (see [5] for a connection among different approaches), but remain largely misunderstood.

Geometry of maximum likelihood and the elimination criterion [2]**.** Given a graph $G = (V, E)$ with r different colors in the vertices and s different colors in the edges, and a sample covariance matrix $S \in \mathbb{S}^m_{\succeq 0}$, its sufficient statistics can be computed via the projection

$$\begin{array}{rccl} \pi_G: & \mathbb{S}^m_{\succeq 0} & \longrightarrow & \mathbb{R}^{r+s} \\ & S & \longmapsto & \left(\sum_{\alpha\in V_1} s_{\alpha\alpha}, \ldots, \sum_{\alpha\in V_r} s_{\alpha\alpha}, \sum_{\alpha-\beta\in E_1} s_{\alpha\beta}, \ldots, \sum_{\alpha-\beta\in E_s} s_{\alpha\beta} \right). \end{array}$$

The open convex cone $\mathcal{C}_G := \pi_G\left(\mathbb{S}^m_{\succ 0}\right)$ is known as the *cone of sufficient statistics*. A consequence of colored graphical models being exponential families is that the MLE of S exists and is unique if and only if its sufficient statistics lie in $\mathcal{C}_G$. In particular, existence occurs for any positive-definite matrix S (see e.g. [6, Theorem 2.1.14]). All sample covariance matrices are positive semi-definite, hence they will lie in the (topological) closure of the cone of sufficient statistics. Therefore, the question in terms of existence of the MLE for a rank-deficient covariance matrix is whether it is projected into the interior of the cone or right into the topological boundary $\partial\mathcal{C}_G = \pi_G(\mathbb{S}^m_{\succeq 0})\backslash\pi_G(\mathbb{S}^m_{\succ 0})$.

Algebraic geometry provides tools to study the (Zariski closure of the) projection $\pi_G\left(\mathbb{S}^m_{\preceq n}\right)$ of matrices of rank at most n.

Definition 1. *Let $S = (s_{ij})$ be an $(m \times m)$ symmetric matrix of unknowns and $J_{(G,n)}$ be the ideal in $\mathbb{R}[s_{ij}, t_k]$ generated by*

- $t_i - \sum_{\alpha\in V_i} s_{\alpha\alpha}$ *for* $i \in \{1, \ldots, r\}$, $t_{r+j} - \sum_{\alpha\beta\in E_j} s_{\alpha\beta}$ *for* $j \in \{1, \ldots, s\}$;
- $(n+1)$*-minors of* S.

The ideal of sufficient statistics of rank at most n *is the elimination ideal* $I_{(G,n)} := J_{(G,n)} \cap \mathbb{R}[t_1, \ldots, t_{r+s}]$.

Theorem 2. [2, Theorem 3.3 - Elimination criterion] *If $I_{(G,n)} = (0)$, then the MLE exists with probability 1 for n observations.*

Weak maximum likelihood threshold. We report on preliminary work with Olga Kuznetsova exploring the case when $I_{(G,n)}$ is not the zero ideal. In this situation, it is crucial to study the intersection of $V(I_{(G,n)})$ with the cone of sufficient statistics $\mathcal{C}_G$. More precisely, we need to understand when a tuple $(t_1, \ldots, t_{r+s})$ of sufficient statistics arising from a rank n positive semi-definite matrix S has a full-rank positive definite completion with the same sufficient statistics. In other words, whether exists $S_0 \in \mathbb{S}^m_{\succ 0}$ such that $\pi_G(S_0) = \pi_G(S) = (t_1, \ldots, t_{r+s})$. To this end, it will be helpful to consider the Cholesky decomposition $S = LL^t$.

Proposition 3. *Assume $I_{(G,n+1)} = (0)$, $I_{(G,n)} = \langle f_1, \ldots, f_u\rangle \neq (0)$ and let polynomials f_i be expressed in variables l_{ij}, where $L = (l_{ij})$ is a lower triangular matrix with positive values in the diagonal.*

- *If some f_i is a non-vanishing (possibly negative) sum of squares, then $WMLT(G) = MLT(G) = n + 1$.*
- *If there exists a positive definite $S_0 \in \mathbb{S}^m_{\succ 0}$ such that $\pi_G(S_0) \in V(I_{(G,n)})$, then $n = WMLT(G) \leq MLT(G) = n + 1$.*

This result allows us to complete the computation of the probability of existence of the MLE for n observations of any 4-cycle [2, Table 2], as it is displayed in Fig. 3 for Graph 9 in the previously cited table, and implement algorithms to explicitely construct rank-deficient matrices for which the MLE exists when such probability is strictly between 0 and 1.

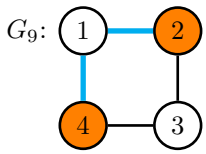

FIGURE 3. $I_{(G_9,1)} = \langle f_1, f_2 \rangle$ with $f_1 = -4t_2t_3 + t_5^2 + t_6^2$. For any $S = LL^t \in \mathbb{S}^m_{\succ 0}$, $-f_1$ evaluates to $4(-l_{13}l_{22}+l_{12}l_{23})^2+4(l_{12}l_{33})^2+4(-l_{14}l_{23}+l_{13}l_{24})^2+4(-l_{14}l_{33}+l_{13}l_{34})^2+4(l_{13}l_{44})^2+4(l_{22}l_{33})^2+4(-l_{24}l_{33}+l_{23}l_{34})^2+4(l_{23}l_{44})^2+4(l_{33}l_{44})^2$, a non-vanishing SOS because $l_{11}, l_{22}, l_{33}, l_{44} > 0$. Therefore, $\mathrm{WMLT}(G_9) = \mathrm{MLT}(G_9) = 2$.

Conjecture 4. *The two cases in Proposition 3 are the only possible scenarios.*

REFERENCES

[1] Søren L. Buhl, *On the Existence of Maximum Likelihood Estimators for Graphical Gaussian Models*, Scandinavian Journal of Statistics **20** (1993), 263–270.
[2] Caroline Uhler, *Geometry of maximum likelihood estimation in Gaussian graphical models*, Annals of Statistics **40** (2012), 238–261.
[3] Daniel Irving Bernstein, Sean Dewar, Steven J. Gortler, Anthony Nixon, Meera Sitharam, Louis Theran, *Maximum likelihood thresholds via graph rigidity*, arXiv:2108.02185 (2021).
[4] Petter G.M. Forbes, Steffen Lauritzen, *Linear estimating equations for exponential families with application to Gaussian linear concentration models*, Linear Algebra Appl. **473** (2015), 261–283.
[5] Elizabeth Gross, Seth Sullivant, *The maximum likelihood threshold of a graph*, Bernoulli **24** (2018), 386–407.
[6] Mathias Drton, Bernd Sturmfels, Seth Sullivant, *Lectures on algebraic statistics*, Oberwolfach Seminars **38** (2008).

Staged trees and rational MLEs

ELIANA DUARTE

The goal of this note is to highlight the role of staged tree models in understanding the defining equations of certain classes of discrete statistical models and also their role in the study of toric models with rational MLE.

One principle in Algebraic Statistics is that many statistical models are algebraic varieties and we study them by using tools from algebraic geometry, commutative algebra and combinatorics. In this field, toric varieties play a special role because they correspond to discrete regular exponential families in statistics

[21]. Many statistical models are defined in terms of polynomial parametrizations, here we restrict to parametrizations inside a probability simplex, it is of interest to know what is the prime ideal that defines the closed image of the parametrization and if such ideal is toric. This is a difficult problem, many times the best hope is to find an ideal whose vanishing locus intersected with the open probability simplex is equal to the model. This perspective has been useful to study a wide class of discrete and Gaussian models [21, 10].

In the study of statistical models that are also varieties, the MLdegree of the model is an algebraic measure of the complexity of maximum likelihood estimation for that model [17]. The discrete statistical models with MLdegree 1 are of particular interest because they are the ones for which the maximum likelihood estimator can be written as a rational function of the data. It is an open problem to classify (in statistical terms) all discrete statistical models with MLdegree 1, even in the case where we restrict to toric varieties. A general classification for complex algebraic varieties was given by Huh [16].

Context-specific conditional independence models. Let $(X_1, \ldots, X_p)$ be a vector of discrete random variables and f a joint probability distribution for this vector, for A, B, C, S disjoint subsets of $[p]$, we say that the subvectors X_A, X_B are *contextually independent* given the subvector X_S in the context $X_C = \mathbf{x}_C$, written $X_A \perp\!\!\!\perp X_B | X_S, X_C = \mathbf{x}_c$, if for all outcomes $(\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_S)$ of (X_A, X_B, X_S) and the fixed outcome $\mathbf{x}_C$ of X_C we have

$$f(\mathbf{x}_A | \mathbf{x}_B, \mathbf{x}_C, \mathbf{x}_S) = f(\mathbf{x}_A | \mathbf{x}_C, \mathbf{x}_S).$$

Staged tree models are a generalization of discrete Bayesian networks, they are useful to encode context-specific conditional independence statements (CSI statements) among random variables like the one just defined [20]. These models are defined via a recursive factorization property that yields a polynomial parametrization of the model, similar to discrete Bayesian networks. The defining equations for staged tree models where first considered in Görgen's PhD thesis [5]. Subsequently, Duarte and Görgen classified the staged tree models that are defined by toric ideals [9] and Ananiadi and Duarte constructed Gröbner bases in this case [1]. The treatment in [9] is a generalization of the framework in [14, Theorem 8]. A characterization of the discrete Bayesian networks that are defined by toric ideals is given by [15] and a new characterization of these models, in terms of staged trees, was given in [11].

Motivated by the fact that the class of staged tree models is very flexible and causal relations are difficult to read from such representations, Duarte and Solus introduce the notion of a CStree [12]. CStree models are a restricted version of staged tree models and they can alternatively be represented by a collection of context DAGs. Duarte and Solus prove that for these models, and their interventional extensions, model equivalence can be determined in analogous fashion as for DAG models. Namely, using Verma and Pearl's criterion that two models are equivalent if and only if they have the same skeleton and v-structures [22]. In terms of defining equations, the class of CStree models that are also toric, where

studied in [8]. Their results show that these class of models, called *decomposable CSmodels*, have the same algebraic and combinatorial properties of decomposable graphical models.

Discrete models with MLdegree 1. From the maximum likelihood perspective, staged tree models are interesting because they are a wide class of models with MLdegree 1 [6]. Models with MLdegree 1 can be completely described in terms of their *Horn matrices*. These were introduced to Algebraic Statistics by Huh in [16], but date back to the work of Kapranov [18] and Horn [19]. They were also later used by Clarke and Cox to characterize the family of polytopes with stritct linear precision [3]. This family of polytopes appears in Geometric Modeling to construct parametrized patches of curves, surfaces, and higher dimensional analogs. The connection between Geometric Modeling and Algebraic Statistics was made earlier by Garcia-Puente and Sottile [13, Proposition 5.1]. This proposition states that a polytope together with a choice of weights for the lattice points has rational linear precision if and only the associated toric model has MLdegree 1. In 2D, the polytopes with rational linear precision where characterized by [2].

The connection between Algebraic Statistics and Geometric Modeling was further exploited in [7]. The authors show that the models in 2D that correspond to the classification of the polytopes in [2] are all represented by a more general class of staged tree model. They introduce this type of model formally and identify the class of such models that are also toric varieties. These results yield a wide class of examples of polytopes with rational MLE in higher dimensions. Moroever, the authors also identify polytopes for which the Horn matrices can be described completely in terms of the lattice distance functions of the faces of the polytope and the primitive collections of the normal fan.

References

[1] L. Ananiadi and E. Duarte. *Gröbner bases for staged trees.* Algebraic Statistics **12**, 2021, 1–20.

[2] H.C. Graf von Bothmer, K. Ranestad, Kristian and F. Sottile *Linear precision for toric surface patches* Found. Compu. Math. **10**, 2010, 37–66.

[3] P. Clarke and D. Cox: Moment maps, strict linear precision, and maximum likelihood degree one, *A*dvances in Mathematics **370**, 2020, 107233.

[4] Rodrigo Collazo, Christiane Görgen and Jim Smith: *Chain Event Graphs*, CRC Press, Taylor & Francis, 2018.

[5] C. Görgen, An algebraic characterisation of staged trees: their geometry and causal implications., Ph.D. thesis, University of Warwick, 2017, available at `http://webcat.warwick.ac.uk/record=b3084201~S15.`

[6] E. Duarte, O. Marigliano, and B. Sturmfels. *Discrete statistical models with rational maximum likelihood estimators* Bernoulli. **27**(1), 2021, 135–154.

[7] Families of polytopes with rational linear precision in higher dimension I. Davies, E. Duarte, I. Portakal and M.S. Sorea. *Families of polytopes with rational linear precision in higher dimensions*, 2022, 1–52.

[8] Y. Alexandr, E. Duarte and J. Vill. Decomposable context-specific models. *arXiv preprint* `https://arxiv.org/abs/2210.11521`, 2022.

[9] E. Duarte and C. Görgen: Equations defining probability tree models, *Journal of Symbolic Computation*, **99**, 2020, 127–146

[10] M. Drton, B. Sturmfels and S. Sullivant: *Lectures on Algebraic Statistics*, Oberwolfach Seminars, Vol 40, Birkhäuser, Basel, 2009.
[11] E. Duarte and L. Solus A new characterization of discrete decomposable models *Proc. Amer. Math. Soc.* **151** (2023), 1325–1338
[12] E. Duarte and L. Solus Representation of Context-Specific Causal Models with Observational and Interventional Data *arXiv preprint* `https://arxiv.org/abs/2101.09271`, 2021.
[13] Luis Garcia-Puente and Frank Sottile: Linear precision for parametric patches, *Advances in Computational Mathematics* **33** (2010) 191–214.
[14] L. D. Garcia, M. Stillman, and B. Sturmfels. *Algebraic geometry of Bayesian networks.* Journal of Symbolic Computation 39.3-4 (2005): 331–355.
[15] D. Geiger, C. Meek, and B. Sturmfels. On the toric algebra of graphical models. *Ann. Statist.*, 34(3):1463–1492, 2006.
[16] June Huh: Varieties with maximum likelihood degree one, *J. Algebr. Statistics* **5** (2014) 1–17. Available at `https://repository.iit.edu/islandora/object/islandora%3A1007817`.
[17] June Huh and Bernd Sturmfels: Likelihood geometry, in *Combinatorial Algebraic Geometry*, pages 63–117. Springer, 2014.
[18] Mikhael Kapranov: A characterization of A-discriminantal hypersurfaces in terms of the logarithmic Gauss map, *Mathematische Annalen* **290** (1991) 277–285.
[19] Horn. *Über die Convergenz der hypergeometrischen Reihen zweier und dreier Veränderlichen* Mathematische Annalen 34 (1889): 544–600.
[20] Jim Q. Smith and Paul E. Anderson: Conditional independence and chain event graphs, *Artificial Intelligence* **172** (2008) 42–68.
[21] S. Sullivant. *Algebraic statistics*, volume 194 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.
[22] T. Verma and J. Pearl. *Causal networks: Semantics and expressiveness.* Machine intelligence and pattern recognition. Vol. 9. North-Holland, 1990. 69-76.

Maximum likelihood degrees, Euclidean distance degrees, and the topology underneath

Jose Israel Rodriguez

Algebraic Statistics: A variety is a model. A mantra of *algebraic statistics* is that a statistical model can be realized as the restriction of an algebraic variety to a semi-algebraic set. For discrete models this semi-algebraic set is the probability simplex while for Gaussian models it is the positive definite cone. With this perspective, the statistical inference method *maximum likelihood estimation* can be viewed as solving a system of polynomial equations. For discrete models, the *likelihood equations* and methods to solve them were introduced by Hoşten, Khetan, and Sturmfels [4]. These equations are with respect to observed data and a model. The number of solutions for generic data is called the *maximum likelihood (ML) degree* of the model. The ML degree bounds the algebraic complexity of solving MLE.

Moreover, the ML degree is related to the topology of the model's Zariski closure with Huh expressing the ML degree of a smooth model as an Euler characteristic of a variety [5]. This Euler characteristic result was generalized by Rodriguez and Wang to singular models to prove a conjectured formula [3] for the ML degree of mixtures of independence models [11]. In the context of graphical models, these

mixtures are known as a *naive Bayes model* and are expressed as two observed variables and one hidden variable, see [6].

Nearest Point Problems. Solving nearest point problems (NPP) is important for engineering, data science, and physics. They also arise in statistics as a maximum likelihood estimate [12, Chapter 7.1] for the general multivariate Gaussian model. The *Euclidean distance (ED) degree* [2] gives an algebraic measure of complexity for NPP. Moreover, in [2] they show that the ED degree of a projective variety, under some assumptions, is the sum of the polar classes [2, Theorem 5.4]. Using CSM classes, [1] derives a formula for the ED degree of a smooth projective variety. The question on how to express the ED degree of an affine (smooth or singular) variety as an Euler characteristic is addressed in [7].

Theorem 1 (Maxim-Rodriguez-Wang)**.** *For $u \in \mathbb{C}^n$, let $d_u(x) := \sum_{i=1}^n (x_i - u_i)^2$. If $X \subset \mathbb{C}^n$ is a smooth affine complex algebraic variety and $(u_0, u_1, \ldots, u_n) \in \mathbb{C}^{n+1}$ is generic, then*

$$\mathrm{EDdeg}(X) = (-1)^{\dim_{\mathbb{C}} X} \chi(X \setminus Q_u), \tag{1}$$

with $Q_u := \{x \in \mathbb{C}^n \mid d_u(x) = u_0\}$.

This result for smooth varieties is used in a computer vision context to find the ED degree of n-view triangulation [2, Conjecture 3.4]. Formula (1) relates the degree of an ideal of critical points to the topology of a variety. The left side counts solutions to a certain system of equations, while the right side involves the Euler characteristic χ, a basic topological invariant used in several areas of pure mathematics and now in *applied algebraic geometry*.

Thus, solving polynomial systems and determining Euler characteristics give two distinct approaches for finding ED degrees. Moreover, these methods can be used in conjunction by having the Euler characteristic tools provide stopping criteria for *monodromy solvers*.

Variety of Critical Points: From bidegree to sectional. Monodromy solvers compute ED degrees by considering the *variety of critical points*, the closure of

$$\{(x, u) \in \mathbb{C}^n_x \times \mathbb{C}^n_u : x \text{ is a critical point of } d_u|_{X_{reg}}\}, \tag{2}$$

as a branched cover of $\mathbb{C}^n_u$. The idea is to use the cover's *monodromy action* to populate the fiber over a point. If the point is general, then the fiber's size is the ED degree of X. More broadly, we want information about the closure of (2).

If X is an irreducible affine cone, then the closure of (2) in $\mathbb{P}^{n-1}_x \times \mathbb{P}^{n-1}_u$ is called the *projective ED correspondence of X*, denoted $\mathcal{E}_X$. The variety $\mathcal{E}_X$ has *bidegree*

$$b_0(X)s^{n-1} + b_1(X)s^{n-2}t + \cdots + b_{n-1}(X)t^{n-1} \in \mathbb{Z}[s,t]/\langle s^n, t^n \rangle, \tag{3}$$

where $b_i(X)$ is the number of points of intersection of $\mathcal{E}_X$ with $L \times M$ such that L, M are general linear subspaces of $\mathbb{P}^{n-1}$ of dimensions $n-1-i$ and i. The ED degree of X is $b_0(X)$, and the trailing term of (3) with nonzero coefficient gives the degree of X. A natural question to ask is, "What information do bidegrees give us about optimization?".

An answer is given in terms of *sectional ED degrees*. Define the *ith sectional ED degree of* X to be the ED degree of $X \cap L$ where L is a general codimension i linear space. Under some conditions, [2] relates the bidegree (3) to the sectional ED degrees so that one can be obtained from the other. Such an exchange of information from bidegree to sectional holds for other objective functions. Recent results prove this for maximum likelihood estimation [10] and for linear objective functions on varieties [9]. The next step is to use these results to improve monodromy solvers and to develop homotopies for specific applications, e.g. phylogenetics.

Non-generic Data: Positive dimensional sets of critical points. For MLE and NPP, working with non-generic data can lead to having an infinite number of critical points. For instance, every point on the circle is a critical point of the distance function when the data is taken to be the circle's center. In the context for Euclidean distance degrees, we can write

$$d_{u+t\epsilon}(x) = d_u(x) - t\ell(x) + c, \quad t \in \mathbb{C}, \quad \epsilon \in \mathbb{C}^n,$$

with $\ell(x) = 2\sum_{i=1}^n \epsilon_i x_i$, c is a constant with respect to x, and ϵ_i is generic. We consider limits of critical points as t is taken to zero. These endpoints may cluster, tend to infinity or go into possibly positive dimensional components of the set of critical points of d_u. Describing the limits of critical points is the content of the next theorem.

Theorem 2 (Maxim-Rodriguez-Wang)**.** *There exists a stratification of* $X \subset \mathbb{C}^n$ *into locally closed irreducible subvarieties* $X_1, \ldots, X_s$ *such that*

$$\lim_{t\to 0} \mathrm{Crit}(d_{u+t\epsilon}|_{X_{\mathrm{reg}}}) = \sum_{i=1}^{s} n_i \cdot \mathrm{Crit}(\ell|_{X_i}),$$

where Crit *denotes the set of critical points and the numbers* n_i *are positive integers, or a critical point of* $d_{u+t\epsilon}|_{X_{\mathrm{reg}}}$ *tends to infinity as* $t \to 0$.

A stronger version of this statement is proved in [8] where d_u is *any* nonconstant polynomial function and the n_i are topological invariants. The power of the statement comes from the fact that the stratification is *independent of* ℓ and the right hand side can be calculated without taking a limit. With an array of examples and motivated by numerical algebraic geometry, this talk will discuss topological methods to compute ML and ED degrees, leading to this question. Can we use Euler characteristics to describe the ML degree of multivariate Gaussian models when the data is a low rank (non-generic) sample covariance matrix.

References

[1] P. Aluffi and C. Harris. The Euclidean distance degree of smooth complex projective varieties. *Algebra Number Theory*, 12(8):2005–2032, 2018.

[2] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. R. Thomas. The Euclidean distance degree of an algebraic variety. *Found. Comput. Math.*, 16(1):99–149, 2016.

[3] J. Hauenstein, J. I. Rodriguez, and B. Sturmfels. Maximum likelihood for matrices with rank constraints. *J. Algebr. Stat.*, 5(1):18–38, 2014.

[4] S. Hoşten, A. Khetan, and B. Sturmfels. Solving the likelihood equations. *Found. Comput. Math.*, 5(4):389–407, 2005.

[5] J. Huh. The maximum likelihood degree of a very affine variety. *Compos. Math.*, 149(8):1245–1266, 2013.
[6] K. Kubjas, E. Robeva, and B. Sturmfels. Fixed points EM algorithm and nonnegative rank boundaries. *Ann. Statist.*, 43(1):422–461, 2015.
[7] L. G. Maxim, J. I. Rodriguez, and B. Wang. Euclidean distance degree of the multiview variety. *SIAM J. Appl. Algebra Geom.*, 4(1):28–48, 2020.
[8] L. G. Maxim, J. I. Rodriguez, and B. Wang. A Morse theoretic approach to non-isolated singularities and applications to optimization. *J. Pure Appl. Algebra*, 226(3):Paper No. 106865, 23, 2022.
[9] L. G. Maxim, J. I. Rodriguez, B. Wang, and L. Wu. Linear optimization on varieties and Chern-Mather classes. *arXiv preprint* `https://arxiv.org/abs/2208.09073`, 2022.
[10] L. G. Maxim, J. I. Rodriguez, B. Wang, and L. Wu. Logarithmic cotangent bundles, Chern-Mather classes, and the Huh-Sturmfels Involution conjecture. *arXiv preprint* `https://arxiv.org/abs/2202.00554`, 2022.
[11] J. I. Rodriguez and B. Wang. The maximum likelihood degree of mixtures of independence models. *SIAM J. Appl. Algebra Geom.*, 1(1):484–506, 2017.
[12] S. Sullivant. *Algebraic statistics*, volume 194 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.

Nonlinear Algebra in Game Theory

Irem Portakal
(joint work with Marie Brandenburg, Ben Hollering, Javier Sendra-Arranz, and Bernd Sturmfels)

Introduction. In 1950, Nash published a very influential two-page paper [6] proving the existence of Nash equilibria for any finite game. The proof uses an elegant application of the Kakutani fixed-point theorem (a generalization of Brouwer fixed-point theorem) from the field of topology. This opened a new horizon not only in game theory but also in areas such as economics, computer science, evolutionary biology, quantum mechanics, and social sciences. It has, however, been noted that in some cases the Nash equilibrium fails to predict the most beneficial outcome for all players. To address this, another mathematician from knot theory, Aumann, introduced the concept of correlated equilibria in 1974 [1], which is a generalization of Nash equilibrium. Still, this solution concept has proved to be inefficient for certain games, e.g. the prisoner's dilemma. In this 2-player game, two prisoners are put in solitary confinement rooms and are offered two choices: cooperate or defect. If one of them cooperates and the other defects, the one who cooperated will serve three years in prison and the other will be set free. If both of them cooperate, they will both serve one year of prison, but if they both defect, they will both serve two years. Nash equilibrium and correlated equilibrium propose that both of them should defect. Although mutual defection is a rational outcome from a self-interested perspective, it is not the most beneficial or Pareto optimal. If the prisoners had the opportunity to communicate, they would have reached the decision of mutual cooperation. Ergo, as an attempt to rationalize mutual cooperation, the dependency equilibrium was proposed by the philosopher Spohn in 2003. It posits a scenario where there are means of communication between

all players and the dependency equilibrium occurs as the players maximize their conditional expected payoffs [9]. The following quote [10] is crucial, as it was the starting point for the journey to this study:

> "The computation of dependency equilibria seems to be a messy business. Obviously, it requires one to solve quadratic equations in two-person games, and the more persons, the higher the order of the polynomials we become entangled with. All linear ease is lost. Therefore, I cannot offer a well-developed theory of dependency equilibria."

As a positive instance of history recurrence, nonlinear algebra promises an innovative way to extend the horizons in game theory and to construct many bridges between the shores of the big continents in the mathematical world, with new applications in economics, computer science, and social sciences. The strongest assurance for this promise is the universality theorems for Nash, correlated and Spohn conditional independence equilibria [4, 11, 7]. This is a report based on joint works with Marie Brandenburg, Ben Hollering, Javier Sendra-Arranz, and Bernd Sturmfels [8, 7, 3], where we observe how (computational) algebraic geometry, convex geometry, algebraic statistics, and combinatorics are indispensable for studying undiscovered territories of game theory.

Results. We work in the setting of n-player normal form games X. The ith player can select from d_i pure strategies. The game is specified by n payoff tables $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ where each is a tensor of format $d_1 \times d_2 \times \cdots \times d_n$ whose entries are arbitrary real numbers. The entry $X^{(i)}_{j_1 j_2 \cdots j_n} \in \mathbb{R}$ represents the payoff for player i if player 1 chooses pure strategy j_1, player 2 chooses pure strategy j_2, etc. A joint probability distribution P is a tensor of format $d_1 \times d_2 \times \cdots \times d_n$ whose entries are positive reals that sum to 1. The entry $p_{j_1 j_2 \cdots j_n}$ is the probability that player 1 chooses pure strategy j_1, player 2 chooses pure strategy j_2, etc.

A joint probability distribution is called a *correlated equilibrium* if no player can raise their expected payoff by breaking their part of the (agreed) joint distribution while assuming that the other players adhere to their own recommendations. Aumann shows [2] that this definition is equivalent to the following: A tensor P in the probability simplex $\Delta_{d_1 \cdots d_n - 1}$ is a *correlated equilibrium* for a game X if and only if

$$\sum_{j_1=1}^{d_1} \cdots \sum_{j_i=1}^{\widehat{d_i}} \cdots \sum_{j_n=1}^{d_n} \left(X^{(i)}_{j_1 \cdots j_{i-1} k j_{i+1} \cdots j_n} - X^{(i)}_{j_1 \cdots j_{i-1} l j_{i+1} \cdots, j_n} \right) p_{j_1 \cdots j_{i-1} k j_{i+1} \cdots j_n} \geq 0.$$

for all $k, l \in [d_i]$, and for all $i \in [n]$. The set of all such equilibria is the *correlated equilibrium polytope* P_X of the game X. We examine the combinatorial types of P_X for the next unknown case after 2×2 games, using oriented matroids and computations in `Mathematica 13.0` and `SageMath 9.6`. All computations can be found in MATHREPO [5].

Theorem 1 ([3, Theorem 5.9])**.** *Let X be a (2×3)-game and P_X be the associated correlated equilibrium polytope. Then one of the following holds:*

- P_X *is a point,*
- P_X *is of maximal dimensional* 5 *and of a unique combinatorial type,*
- *There exists a* (2×2)*-game* X' *such that* $P_{X'}$ *has maximal dimensional* 3 *is and combinatorially equivalent to* P_X*.*

Spohn introduced in 2003 [9] the concept of *dependency equilibria* where the players simultaneously maximize their *conditional expected payoffs*; appending communication between players. The conditional expected payoff of Player i, in case they choose strategy $k \in [d_i]$:

$$\sum_{j_1=1}^{d_1} \cdots \widehat{\sum_{j_i=1}^{d_i}} \cdots \sum_{j_n=1}^{d_n} X^{(i)}_{j_1 \cdots k \cdots j_n} \frac{p_{j_1 \cdots k \cdots j_n}}{p_{+\cdots+k+\cdots+}}.$$

A tensor P inside the probability simplex $\Delta^\circ_{d_1 \cdots d_n - 1}$ is a *dependency equilibrium* for X if the conditional expected payoff of each Player i is independent of their choice $k \in [d_i]$. For $i = 1, 2, \ldots, n$, we define a matrix with d_i rows and two columns:

$$M_i = \begin{bmatrix} \vdots & \vdots \\ p_{+\cdots+k+\cdots+} & \sum_{j_1=1}^{d_1} \cdots \widehat{\sum_{j_i=1}^{d_i}} \cdots \sum_{j_n=1}^{d_n} X^{(i)}_{j_1 \cdots k \cdots j_n} p_{j_1 \cdots k \cdots j_n} \\ \vdots & \vdots \end{bmatrix}$$

Equivalently, the dependency equilibria for X are tensors $P \in \Delta^\circ_{d_1 \cdots d_n - 1}$, where each M_i has rank one. Let $V = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$ be the real vector space of all tensors and let $\mathbb{P}(V)$ denote the corresponding projective space. We write $\mathcal{V}_X$ for the projective subvariety of $\mathbb{P}(V)$ that is given by requiring $M_1, \ldots, M_n$ to have rank one. We call $\mathcal{V}_X$ the *Spohn variety* of the game X.

Theorem 2 ([8, Theorem 3.2, Theorem 3.4])**.** *For a generic game* X*, the Spohn variety* $\mathcal{V}_X$ *is irreducible of codimension* $d_1 + d_2 + \cdots + d_n - n$ *and degree* $d_1 d_2 \ldots d_n$*. The intersection of* $\mathcal{V}_X$ *with the Segre variety in the open simplex* Δ *is precisely the set of totally mixed Nash equilibria for* X*. If* $n = d_1 = d_2 = 2$ *then the Spohn variety* $\mathcal{V}_X$ *is an elliptic curve. In all other cases, the Spohn variety* $\mathcal{V}_X$ *is rational, represented by a map onto* $(\mathbb{P}^1)^n$ *with linear fibers.*

To explore the fact that dependency equilibria are abundant, we focus our attention to the intersection of $\mathcal{V}_X$ with statistical models, more precisely with the model defined by the conditional independence statements arising from the dependencies of the players of X. The construction of this intersection is explained in detail in [8, Chapter 6] and we call the resulting variety the *Spohn conditional independence (CI) variety.* Consequently, we bring the concept of Nash and dependency equilibrium under the roof of graphical models, i.e. game theory and algebraic statistics are meeting for the first time. More importantly, this study is the first attempt at a kind of unification of non-cooperative and cooperative game theory mentioned in [9, Section 6.3] from the perspective of algebraic statistics as

well as with methods from computational algebraic geometry. We prove [8, Conjecture 6.3] for Spohn CI varieties for one-edge models and a similar universality theorem as of Datta [4] for Nash equilibrium:

Theorem 3 ([7, Theorem 15])**.** *Let $S \subseteq \mathbb{R}^n$ be a real affine algebraic variety defined by m polynomials with $m < n$. Then, there exists a N-person game with binary choices such that an affine open subset of the Spohn CI variety for the one-edge model is isomorphic to S.*

References

[1] R. J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96, 1974.

[2] R. J. Aumann. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica: Journal of the Econometric Society*, 1–18, 1987.

[3] M. Brandenburg, B. Hollering, and I. Portakal. *Combinatorics of Correlated Equilibria. arXiv preprint* `https://arxiv.org/abs/2209.13938`, 2022.

[4] R. S. Datta. Universality of Nash equilibria. *Math. of Operations Research* **28**, 424–432, 2003.

[5] MATHREPO Mathematical Data and Software. `https://mathrepo.mis.mpg.de/correlated-equilibrium`, 2022.

[6] J. F. Nash. Equilibrium Points in N-Person Games. *Proceedings of the National Academy of Sciences of the United States of America*, 36 1, 48–9, 1950.

[7] I. Portakal and J. Sendra-Arranz. Nash conditional independence curve. *arXiv preprint* `https://arxiv.org/abs/2206.07000`, 2022.

[8] I. Portakal and B. Sturmfels. Geometry of dependency equilibria. *Rendiconti dell'Istituto di Matematica dell'Università di Trieste*, 54(3), 2022.

[9] W. Spohn. Dependency equilibria and the causal structure of decision and game situations. *Homo Oeconomicus* 20, 195–255, 2003.

[10] W. Spohn. Dependency equilibria. *Philosophy of Science*, 74(5):775–789, 2007.

[11] Y. Viossat, E. Solan, and E. Lehrer. Equilibrium payoffs of finite games. *Journal of Mathematical Economics*, vol. 47, n°1, p. 48–53, 2011.

Dimensions of Higher Order Factor Analysis Models

Muhammad Ardiyansyah

(joint work with Luca Sodomaco)

The factor analysis model is a statistical model where a certain number of hidden random variables, called factors, affect linearly the behavior of another set of observed random variables, with additional random noise. In the factor analysis model, each observed random variable X_i is a linear combination of the hidden factors Y_j with some independent noise, namely

$$X = \Lambda Y + \varepsilon \tag{1}$$

for some unknown coefficient matrix $\Lambda = (\lambda_{ij}) \in \mathbb{R}^{p \times m}$, whose entries are sometimes referred as *factor loadings*, and for some noise random vector $\varepsilon \in \mathbb{R}^p$. In particular, several observed variables X_i might be measures of the same factor Y_j. The factor analysis model may be regarded as a special instance of a much more

general graphical model, where the components of a certain random vector Z interact with each other, and their interaction is encoded by the edges of a directed acyclic graph with vertex set equal to the components of Z. In our setting, the random vector Z is the joint vector (X, Y), and the interactions between X and Y are described by a directed bipartite graph where all edges are directed from elements of Y to elements of X.

The main assumption of the factor analysis model is that the factors and the noise are Gaussian random variables. This implies that the feasible set lies in the cone of positive semidefinite matrices. In this paper, we do not assume that the factors and the noise are Gaussian, hence the higher order moment and cumulant tensors of the observed variables are generally nonzero. This motivates the notion of kth-order factor analysis model.

Definition 1. *Let $k \geq 2$ be an integer. A* kth-order factor analysis model *is a family of random vectors X of observed variables that are correlated to another vector Y of hidden variables (called factors) via equation* (1)*, where ε is a noise component. The model relies on the following assumptions:*

(1) *All moment and cumulant tensors of Y and ε exist and are finite up to order k.*
(2) *The vectors Y and ε are independent each other.*
(3) *The components of Y are mutually independent, and similarly for ε.*
(4) *The vectors Y and ε have mean equal to zero.*

Definition 1 has a natural counterpart in terms of cumulant and moment tensors.

Definition 2. *Let p, m and k be nonnegative integers with $k \geq 2$.*

(*a*) *The* kth-order cumulant factor analysis model *is the subset of tuples $(\mathcal{C}^{(2)}, \ldots, \mathcal{C}^{(k)})$ of symmetric tensors $\mathcal{C}^{(r)} \in \mathrm{Sym}^r(\mathbb{R}^p)$ that are the cumulant tensors for some random vector $X \in \mathbb{R}^p$ in the kth-order factor analysis model. We denote this subset by $\mathcal{C}_{p,m}^{(\leq k)}$.*
(*b*) *The* kth-order moment factor analysis model *is the subset of tuples $(\mathcal{M}^{(2)}, \ldots, \mathcal{M}^{(k)})$ of symmetric tensors $\mathcal{M}^{(r)} \in \mathrm{Sym}^r(\mathbb{R}^p)$ that are the moment tensors for some random vector $X \in \mathbb{R}^p$ in the kth-order factor analysis model. We denote this subset by $\mathcal{M}_{p,m}^{(\leq k)}$.*

These subsets may be described as the image of a polynomial map onto a Cartesian product of symmetric tensor spaces. Our goal is to compute its dimension and we provide conditions under which the image has positive codimension. The main result is presented in the following theorem.

Theorem 3. *For $k \geq 3$, the dimension and the codimension of the kth-order factor analysis model are*

(2)

$$\dim(\mathcal{M}_{p,m}^{(\leq k)}) = \dim(\mathcal{C}_{p,m}^{(\leq k)}) = \min\left\{(k-1)(p+m)+pm-\binom{m}{2}, \binom{p+k}{k}-p-1\right\}$$

(3)

$$\operatorname{codim}(\mathcal{M}_{p,m}^{(\leq k)}) = \operatorname{codim}(\mathcal{C}_{p,m}^{(\leq k)}) = \frac{1}{k!}\max\left\{h_m^{(k)}(p), 0\right\}$$

where

$$h_m^{(k)}(p) = k!\binom{p+k}{k} - k!(k+m)p + k!\left[\binom{m}{2} - (k-1)m - 1\right] \tag{4}$$

In particular, the codimension of the kth-order factor analysis model is positive if

1. *$m \in [2k-1]$ and $p \geq p_0$, where p_0 is the unique positive root of the polynomial $h_m^{(k)}(p)$.*
2. *$m \geq 2k$ and $p \geq p_0$, where p_0 is the largest positive root of the polynomial $h_m^{(k)}(p)$.*
3. *$m \geq 2k$ is sufficiently large and $p \geq 1$.*

Furthermore, if we assume $p \geq m+2$, then the previous assumptions are automatically satisfied.

References

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

[2] C. Améndola, M. Drton, A. Grosdos, R. Homs, and E. Robeva. Third-order moment varieties of linear non-gaussian graphical models. *arXiv preprint* `https://arxiv.org/abs/2112.10875`, 2021.

[3] E. Robeva and J.-B. Seby. Multi-trek separation in linear structural equation models. *SIAM Journal on Applied Algebra and Geometry*, 5(2):278–303, jan 2021.

[4] C. Spearman. General intelligence, objectively determined and measured. *American Journal of Psychology*, 15:201–293, 1904.

[5] C. Spearman. *The Abilities of Man*. Macmillan, New York, 1927.

Structured log-concave density estimation

Elina Robeva

(joint work with Kaie Kubjas, Olga Kuznetsova, Pardis Semnani, Luca Sodomaco, Sharvaj Kubal, and Christian Campbell)

In this talk I discussed two different families of log-concave densities that have additional structure – log-concave undirected graphical models, and log-concave densities of random vectors whose coordinates are independent after an orthogonal transformation.

The first project on log-concave undirected graphical models is joint work with Kaie Kubjas, Olga Kuznetsova, Pardis Semnani, and Luca Sodomaco. We study

the problem of maximum likelihood estimation of densities that are log-concave and lie in the graphical model corresponding to a given undirected graph G. More precisely, we assume that each density in our family factorizes according to the graph G and all factors are log-concave. We show that the maximum likelihood estimate (MLE) is the product of the exponentials of several tent functions, one for each maximal clique of G. While the set of log-concave densities in a graphical model is infinite-dimensional, our results imply that the MLE can be found by solving a finite-dimensional convex optimization problem. We provide an implementation and a few examples. Furthermore, we show that the MLE exists and is unique with probability 1 as long as the number of sample points is larger than the size of the largest clique of G when G is chordal. We show that the MLE is consistent when the graph G is a disjoint union of cliques. Finally, we discuss the conditions under which a log-concave density in the graphical model of G has a log-concave factorization according to G.

The second project on Log-concave Density Estimation with Orthogonal Independent Components is joint work with Sharvaj Kubal and Christian Campbell. We study the problem of estimating a log-concave density on $\mathbb{R}^d$ under the assumption that there exists an orthogonal transformation that makes the coordinates of the random vector independent. While log-concave density estimation is hard both computationally and statistically, we show that the independence assumption alleviates both issues, while still maintaining a large non-parametric class. We show that under mild assumptions one needs $(\frac{1}{\epsilon})^2$ (up to constants and log factors) samples for our proposed estimator to be within ϵ of the original density in squared Hellinger distance. While finding the log-concave maximum likelihood estimate can be done via a finite-dimensional convex optimization program, it is slow to compute and impractical in high dimensions. Our estimator can be computed efficiently, making it more practical to use.

The complexity of Gaussian conditional independence inference

Tobias Boege

1. Gaussian conditional independence

Given a vector of jointly distributed random variables $(\xi_i : i \in N)$ indexed by a finite set N and disjoint subsets $I, J, K \subseteq N$, the *conditional independence (CI)* statement $[I \perp\!\!\!\perp J \mid K]$ asserts: whenever the outcome of the subvector $\xi_K = x_K$ is known, the conditional distributions $\xi_I \mid \xi_K = x_K$ and $\xi_J \mid \xi_K = x_K$ are independent.

If ξ follows a multivariate normal (i.e., Gaussian) distribution with mean $\mu \in \mathbb{R}^N$ and covariance matrix $\Sigma \in \mathrm{PD}_N$, then conditional independence is an algebraic condition on Σ:

Lemma 1 ([7, Proposition 4.1.9]). *For $\Sigma \in \mathrm{PD}_N$ and any $\mu \in \mathbb{R}^N$ defining a Gaussian random vector ξ we have $[X_I \perp\!\!\!\perp X_J \mid X_K]$ if and only if $\operatorname{rk} \Sigma_{IK,JK} = |K|$.*

The shorthand notation IK for $I \cup K$ is customary in the theory of CI structure; thus $\Sigma_{IK,JK}$ in the above definition is the submatrix of Σ with rows indexed by $I \cup K$ and columns by $J \cup K$. We will also liberally identify $i \in N$ with the singleton subset $\{ i \} \subseteq N$ because the marginal vector (ξ_i) does not differ form the component ξ_i for any statistical or algebraic purposes.

It follows from Lemma 1 (and in fact even weaker properties of Gaussian conditional independence known as the semigraphoid, intersection and composition properties) that

$$[\xi_I \perp\!\!\!\perp \xi_J \mid \xi_K] \;\Leftrightarrow\; \bigwedge_{\substack{i \in I,\\ j \in J}} [\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K].$$

Hence, the conditional independences of a Gaussian are uniquely determined by the subset where $|I| = |J| = 1$. We will concentrate on them in the remainder of this article. The rank condition in Lemma 1 is then equivalent to the vanishing of a single polynomial $\Sigma[i \perp\!\!\!\perp j \mid K] := \det \Sigma_{iK,jK}$, which is called an *almost-principal minor* of Σ.

In this talk we consider statistical models of Gaussians defined by conditional independece and dependence assumptions and with positive definite covariance matrix. Since, according to Lemma 1, the mean μ plays no role in this setting, we identify a Gaussian distribution with its covariance matrices.

Definition 2. *A* Gaussian CI model *is a subset of* PD_N *defined by conditional independence and conditional dependence assumptions.*

Each conditional independence assumption $[\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K]$ imposes an equation $\Sigma[i \perp\!\!\!\perp j \mid K] = 0$ on the points in the statistical model; the conditional dependence assumption $\neg[\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K]$ the corresponding inequation $\Sigma[i \perp\!\!\!\perp j \mid K] \neq 0$.

2. The inference problem

A basic algorithmic problem about CI models is to decide when their specification is *consistent*, i.e., when does there exist a covariance matrix at all satisfying all constraints? This problem is equivalent to the conditional independence inference problem for Gaussians, in the following sense.

Let $\mathcal{P}$ and $\mathcal{Q}$ be sets of CI statements. In the inference problem one wishes to decide if the boolean formula

$$(\Rightarrow) \qquad \bigwedge \mathcal{P} \;\Rightarrow\; \bigvee \mathcal{Q}$$

is true for all Gaussian distributions. This implication asserts that every Gaussian satisfying *all* the CI assumptions in $\mathcal{P}$ must satisfy *at least one* of the CI conclusions in $\mathcal{Q}$. The model defined by conditional independence assumptions $\mathcal{P}$ and conditional dependence assumptions $\mathcal{Q}$ consists exactly of the counterexamples to the validity of $(\Rightarrow)$.

Lemma 3. *Deciding if the definition of a CI model is consistent is polynomial-time equivalent to the problem of deciding if an inference formula is valid.*

This reduces a seemingly logical problem to a geometric one of proving or refuting the existence of a point in a semialgebraic set defined by determinants. Hence, an upper bound on the algorithmic complexity of this task is the *existential theory of the reals*, a complexity class commonly encountered in polynomial optimization and computational geometry. Tarski's transfer principle and the Positivstellensatz in real algebraic geometry [1] prove the following "theorem of the alternative":

Theorem 4. *If a CI model is non-empty, then it contains a covariance matrix with real algebraic entries. If the CI model is empty, there exists a* final polynomial *witnessing this.*

For the concept of final polynomials, see [2, Section 3.6] and its references. The bottom line of this theorem is that both possible answers to the consistency problem for CI models (and hence to the inference problem) have algebraic *certificates* which can be stored exactly on a computer and allow verification of the claim in off-the-shelf computer algebra software.

3. Universality theorems

The reduction of the inference problem to the existential theory of the reals implies upper bounds on its complexity — in an algorithmic sense as well as in an algebraic sense (no transcendental numbers are required to witness consistency of a CI model). Unfortunately, these upper bounds are attained. This is the content of two "universality theorems" for Gaussian CI models proved in [2, Chapter 5]:

Theorem 5. *For every finite real extension $\mathbb{K}$ of $\mathbb{Q}$ there exists a CI model $\mathcal{M}$ such that $\mathcal{M} \neq \emptyset$ but $\mathcal{M} \cap \mathrm{PD}_N(\mathbb{K}) = \emptyset$.*

Theorem 6. *The problem of deciding consistency for CI models is complete for the existential theory of the reals under polynomial-time many-one reductions.*

This means that deciding whether the very special semialgebraic sets that are Gaussian CI models are empty or not is just as difficult as deciding this for a general semialgebraic set. Moreover, if the set is non-empty, there is no a priori upper bound on the complexity of a certificate in the sense of algebraic extension degree over the rational numbers. This resolves negatively a conjecture by Petr Šimeček [5] about the existence of rational points in Gaussian CI models.

Proof outline. Both results follow from an encoding of arbitrary (primary basic) semialgebraic sets into CI models. First notice that an almost-principal minor $[i \perp\!\!\!\perp j \mid xyz]$ may be used to the effect of storing an inner product with respect to the inverse of the Σ_{xyz}-submatrix into the σ_{ij} entry, by Schur complement expansion:

$$0 = \Sigma[i \perp\!\!\!\perp j \mid xyz] = \sigma_{ij} - \Sigma_{i,xyz}\Sigma_{xyz}^{-1}\Sigma_{xyz,j}.$$

Afterwards, the almost-principal minor $[i \perp\!\!\!\perp j]$ can be used to set this entry to zero, hence to make the vectors $\Sigma_{i,xyz}$ and $\Sigma_{j,xyz}$ orthogonal with respect to the inner product given by Σ_{xyz}^{-1}. Interpreting one of these vectors as the homogeneous coordinates of a point and the other as coordinates of a line in the projective plane

(transformed by Σ_{xyz}^{-1}), this allows the modeling of incidence geometry through Gaussian conditional independence constraints.

Having a model of the projective plane, the encoding of arbitrary polynomial constraints is an application of the *von Staudt* constructions described in detail in [4, Section 5.6]. □

An analogous conjecture to the one by Šimeček, but for rational points on discrete as opposed to Gaussian CI models, was proposed by Matúš in [3] and is still open. Unlike the Gaussian case, discrete CI models are infinite-dimensional because the number of states of each random variable is not prescribed. It would be interesting to investigate whether universality theorems can be proved for *binary state spaces*. As has been demonstrated by Šimeček in his thesis [6], binary CI equations can be bent under additional assumptions to look very close to the almost-principal minors of Gaussian CI.

References

[1] J. Bochnak, M. Coste, and M.-F. Roy: *Real algebraic geometry*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, vol. 36, Springer, 1998.

[2] T. Boege, *The Gaussian conditional independence inference problem*, PhD thesis, Otto-von-Guericke-Universiät Magdeburg (2022). DOI: 10.25673/86275.

[3] F. Matúš. Conditional independences among four random variables. III. Final conclusion. *Combin. Probab. Comput.*, 8(3):269–276, 1999.

[4] J. Richter-Gebert: *Perspectives on projective geometry. A guided tour through real and complex geometry*. Springer, 2011.

[5] P. Šimeček. Gaussian representation of independence models over four random variables. In *COMPSTAT conference*, 2006.

[6] P. Šimeček. *Nezávislostní modely*. Dissertation (in Czech), Charles University, Prague, Czech Republic, 2007.

[7] S. Sullivant: *Algebraic Statistics*. Graduate Studies in Mathematics, vol. 194, American Mathematical Society (AMS), 2018.

Linear Causal Disentanglement

Anna Seigal

(joint work with Chandler Squires, and Caroline Uhler)

Causal disentanglement seeks a representation of data involving latent variables that relate to one another via a causal model. We consider linear causal distanglement: observed variables that are a linear transformation of a linear latent causal model. The setup is identifiable if the linear transformation and the latent causal model are unique. We show that one intervention on each latent variable is sufficient and, in the worst case, necessary for identifiability.

1. Setup

We consider p latent variables $Z = (Z_1, \ldots, Z_p)$ observed in contexts $k \in \{0, \ldots, K\}$. Context $k = 0$ is thought of as *observational* data, while contexts $k \in [K]$ are *interventional* data. We make the following assumptions.

(a) **Linear latent model**: Let $\mathcal{G}$ be a DAG with nodes ordered so that an edge $j \to i$ implies $j > i$. The variables Z follow a linear structural equation model: in context k, they satisfy

$$Z = A_k Z + \Omega_k^{1/2}\varepsilon, \qquad \mathrm{Cov}(\varepsilon) = \mathrm{I_p},$$

where $I_p \in \mathbb{R}^{p\times p}$ is the identity matrix, $\Omega_k \in \mathbb{R}^{p\times p}$ is diagonal with positive entries, and $A_k \in \mathbb{R}^{p\times p}$ has $(A_k)_{ij} \neq 0$ if and only if there is an edge $j \to i$ in $\mathcal{G}$. In context k we have $Z = B_k^{-1}\varepsilon$, where $B_k = \Omega_k^{-1/2}(I_p - A_k)$.

(b) **Single-node perfect interventions:** For each $k \in \{1, \ldots, K\}$, there exists $i_k \in \{1, \ldots, p\}$ such that

$$B_k = B_0 + \mathbf{e}_{i_k}\mathbf{c}_k^\top, \quad \text{where } \mathbf{c}_k = \lambda_k \mathbf{e}_{i_k} - B_0^\top \mathbf{e}_{i_k} \text{ for some } \lambda_k > 0.$$

The intervention zeros the weight on all edges with target i_k and changes the variance at node i_k.

(c) **Linear observations:** There is an invertible $G \in \mathbb{R}^{p\times p}$ such that $X = GZ$ in every context k. Let $H := G^{-1}$. Without loss of generality, we set the entry of largest absolute value in each row of H to 1. If multiple entries in a row have same absolute value we set the leftmost entry to be positive.

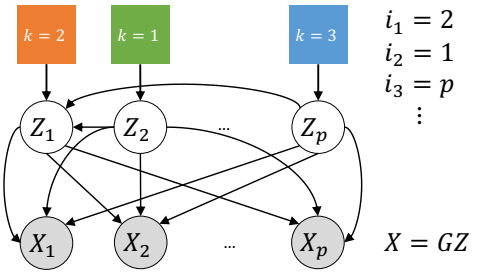

FIGURE 1. The setup

The precision (inverse covariance) matrix on X in context k is

$$\Theta_k := H^\top B_k^\top B_k H. \tag{1}$$

We consider an unknown latent DAG $\mathcal{G}$ with unknown weights on its edges, unknown variances on its nodes, unknown new variances under each intervention, and an unknown mixing map to the observed variables. That is, our goal is to decompose the precision matrices $\{\Theta_k\}_{k=0}^K$ to recover $\mathcal{G}$, H, and $\{B_k\}_{k=0}^K$.

2. Main result

We define $S(\mathcal{G})$ to be the permutations on p letters such that $\sigma(j) > \sigma(i)$ for all edges $j \to i$. For example, if $\mathcal{G}$ is a complete graph then $S(\mathcal{G})$ contains only the identity. If $\mathcal{G}$ has no edges then $S(\mathcal{G})$ is the group of permutations on p letters. Our main result is the following.

Theorem 1. *Assume the setup above, with one intervention on each latent node. Then the graph $\mathcal{G}$, the intervention targets, and the parameters are identifiable up to $S(\mathcal{G})$: given a solution $(B_0, \dots, B_K, H)$, the set of solutions is*

$$\{(P_\sigma B_0 P_\sigma^\mathsf{T}, \dots, P_\sigma B_K P_\sigma^\mathsf{T}, P_\sigma H) : \sigma \in S(\mathcal{G})\}.$$

For the proof, we introduce a matrix decomposition defined on a partial order. Recall that the RQ decomposition writes $H \in \mathbb{R}^{p\times p}$ as $H = RQ$ for an upper triangular $R \in \mathbb{R}^{p\times p}$ and orthogonal $Q \in \mathbb{R}^{p\times p}$. We generalize the RQ decomposition.

Definition 2 (The partial order RQ decomposition)**.** *Given a partial order $\prec$, the partial order RQ decomposition writes $H \in \mathbb{R}^{p\times p}$ as $H = RQ$, where $R \in \mathbb{R}^{p\times p}$ satisfies $R_{ii} \geq 0$ and $R_{ij} = 0$ for $i \not\preceq j$, and where $\mathbf{q}_i$, the i-th row of $Q \in \mathbb{R}^{p\times p}$, is norm one and orthogonal to $\mathbf{q}_j$ whenever $i \prec j$.*

Note that this specialises to the usual RQ decomposition if $\prec$ is the total order $1 < 2 < \cdots < p$.

The proof of Theorem 1 proceeds by applying the partial order RQ decomposition to the partial order $\prec_\mathcal{G}$ coming from the graph $\mathcal{G}$, where $i \prec_\mathcal{G} j$ if and only if there is a directed path $j \to \cdots \to i$ in $\mathcal{G}$.

References

[1] A. Seigal, Ch. Squires, and C. Uhler. Linear Causal Disentanglement via Interventions. *arXiv preprint* `https://arxiv.org/abs/2211.16467`, 2022.

Maximizing Divergence to Toric Models

Serkan Hoşten
(joint work with Yulia Alexandr)

Introduction

Let M_A be a toric (discrete exponential) model in the probability simplex

$$\Delta_{n-1} = \left\{ (p_1, \dots, p_n) \,:\, p_i \geq 0, i = 1, \dots, n, \;\; \sum_{i=1}^{n} p_i = 1 \right\}$$

associated to a $d \times n$ matrix

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

where $a_i \in \mathbb{N}^{d-1}$ for $i = 1, \dots, n$. For p and q in Δ_{n-1}, the Kullback-Leibler (KL) divergence is defined as

$$D(p \parallel q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}.$$

Given any model $M \subset \Delta_{n-1}$ and a point $p \in \Delta_{n-1}$ the minimizer of $D(p \parallel q)$ for $q \in M$ is the maximum likelihood estimate of p. We denote the divergence from p

to M by $D(p \parallel M)$. The main focus of the talk is to understand and compute all points $p \in \Delta_{n-1}$ which maximize $D(p \parallel M_A)$ to a toric model M_A.

Review of Prior Work

The problem was studied first by Ay and Knauf [1] where they gave upper bounds on the maximum divergence to complete independence models. They characterized the instances when the bound is attained in this case. Matúš has worked out necessary and sufficient conditions for the maximizers [2]. He has also considered the problem for hierarchical loglinear models [3]. Rauh's PhD thesis [4] is an in-depth summary and besides other results provides an algorithm to compute the maximizers based on concepts from oriented matroids.

Contribution

We report on very preliminary joint work with Yulia Alexandr. First, we show that when M is a linear model, i.e., when M is the intersection of Δ_{n-1} with an affine linear space we show that the maximizers are among the vertices log-Voronoi polytopes corresponding to the vertices of M. Then we present an algorithm to compute the maximizers of divergence to toric models M_A based on Matúš' characterization of these maximizers and using the chamber complex of A. We show simple conditions that help to cut down the work in this algorithm. Finally, we point to further directions to pursue.

References

[1] Nihat Ay and Andreas Knauf, *Maximizing multi-information*, Kybernetika **42** (2006), no. 5, 517–538.

[2] František Matúš, *Optimality conditions for maximizers of the information divergence from an exponential family*, Kybernetika **43** (2007), no. 5, 731–746.

[3] ———, *Divergence from factorizable distributions and matroid representations by partitions*, IEEE Trans. Inform. Theory **55** (2009), no. 12, 5375–5381.

[4] Johannes Rauh, *Finding the maximizers of the information divergence from an exponential family*, 2011, Thesis (Ph.D.)– University of Leipzig.

Reporters: Tobias Boege and Roser Homs Pons

Participants

Prof. Dr. Carlos Enrique Améndola Cerón
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
GERMANY

Muhammad Ardiyansyah
Dept. of Mathematics & Systems Analysis
Aalto University
Suurpellon Puistokatu 12 B 37 Espoo
02250 Aalto
FINLAND

Dr. Tobias Boege
Dept. of Mathematics & Systems Analysis
Aalto University
Suurpellon Puistokatu 12 B 37 Espoo
02250 Aalto
FINLAND

Prof. Dr. Mathias Drton
Technische Universität München
Lehrstuhl für Mathematische Statistik
Boltzmannstr. 3
85748 Garching bei München
GERMANY

Dr. Eliana M. Duarte Gelvez
Centro de Matematica Universidade do Porto,
Faculdade de Ciências, Universidade do Porto
Rua do Campo Alegre 687
4169-007 Porto
PORTUGAL

Prof. Dr. Robin Evans
Department of Statistics
University of Oxford
24-29 St. Giles'
Oxford OX1 3LB
UNITED KINGDOM

Dr. Alexandros Grosdos
Technische Universität München
Lehrstuhl für Mathematische Statistik
Boltzmannstr. 3
85748 Garching bei München
GERMANY

Prof. Dr. Elizabeth Gross
Dept. of Mathematics
University of Hawaii
2565 McCarthy Mall
Honolulu 96822
UNITED STATES

Dr. Ben Hollering
Department of Mathematics
North Carolina State University
Box 8205
Raleigh NC 27695-8205
UNITED STATES

Dr. Roser Homs Pons
Centre de Recerca Matemàtica - CRM
Campus de Bellaterra
Edifici C
08193 Bellaterra, Barcelona
SPAIN

Prof. Dr. Serkan Hosten
Department of Mathematics
San Francisco State University
1600 Holloway Avenue
San Francisco 94132
UNITED STATES

Prof. Dr. Thomas Kahle
FMA - IAG
Otto-von-Guericke-Universität
Magdeburg
Universitätsplatz 2
39106 Magdeburg
GERMANY

Prof. Dr. Kaie Kubjas
Department of Mathematics and
Systems Analysis
Aalto University
P.O. Box 11100
00076 Aalto
FINLAND

Prof. Dr. Steffen Lauritzen
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
2100 København
DENMARK

Dr. Pratik Misra
Department of Mathematics
KTH
10044 Stockholm
SWEDEN

Dr. Irem Portakal
Zentrum Mathematik
TU München
Boltzmannstr. 3
85748 Garching bei München
GERMANY

Dr. Elina Robeva
Department of Mathematics
University of British Columbia
121-1984 Mathematics Road
Vancouver BC V6T 1Z2
CANADA

Prof. Dr. Jose Israel Rodriguez
Department of Mathematics
University of Wisconsin-Madison
480 Lincoln Drive
53706 Madison
UNITED STATES

Dr. Anna Seigal
Department of Mathematics
Harvard University
Science Center
One Oxford Street
Cambridge MA 02138-2901
UNITED STATES

Dr. Liam Solus
Department of Mathematics
KTH
Lindstedtsvägen 25
10044 Stockholm
SWEDEN

Dr. Milan Studený
Institute of Information Theory and
Automation
Academy of Sciences of the Czech
Republic
Pod Vodárenskou věží 4
180 00 Praha
CZECH REPUBLIC

Prof. Dr. Seth Sullivant
Department of Mathematics
North Carolina State University
Box 8205
Raleigh NC 27695-8205
UNITED STATES

Prof. Dr. Caroline Uhler
EECS Department / IDSS
Massachusetts Institute of Technology
32-D634
77 Massachusetts Avenue
Cambridge, MA 02139
UNITED STATES

Dr. Piotr Zwiernik
9th Floor, Ontario Power Building
Department of Statistical Sciences
University of Toronto
700 University Ave.
Toronto ON M5S 2E4
CANADA