An Implicitization Challenge

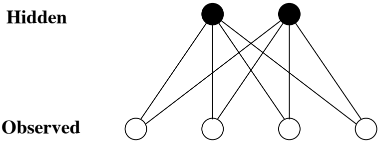
for Binary Factor Analysis

Josephine Yu (Georgia Tech)

joint work with:
María Angélica Cueto and Enrique A. Tobis

Harmony of Gröbner bases and the modern industrial society
Osaka, Japan
July 1, 2010

The Problem: Statistical Model version

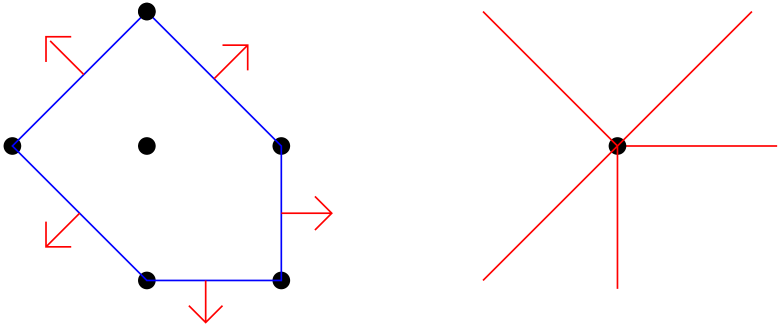


- undirected graphical model of $K_{2,4}$:
  - node $\leftrightarrow$ binary random variable
  - edge $\leftrightarrow$ dependence
- marginalize 2 hidden variables $\rightsquigarrow$ joint distribution of 4 observed variables
- all such $2 \times 2 \times 2 \times 2$ tables are expected to form a hypersurface in Δ_{15}

Implicitization Challenge [Open Problem 7.7 of *Lectures on Algebraic Statistics* by Drton–Sturmfels–Sullivant] :

- Find the defining polynomial.
- What is its multidegree? Newton polytope (vertices and facets)?

Example: Newton polytope and normal fan



$$f = 2x + 3y + 5x^2 + 7xy + 11x^2y + 13xy^2$$

The Problem: Algebraic Formulation

What is the algebraic condition for a $2 \times 2 \times 2 \times 2$ matrix to be the Hadamard (entrywise) product of two $2 \times 2 \times 2 \times 2$ matrices of tensor rank at most two?

Definition: A $2 \times 2 \times 2 \times 2$ matrix has tensor rank at most two if it can be written as a sum of at most two matrices of the form $v_1 \otimes v_2 \otimes v_3 \otimes v_4$ for $v_i \in \mathbb{C}^2$.

This is an implicitization problem. Find the implicit equation(s) (defining equations) of the set given by parameterization:

$$(\mathbb{P}^1 \times \mathbb{P}^1)^8 \to \mathbb{P}^{15}$$

$$p_{ijkl} = (\sum_{s=0}^{1} a_{si} b_{sj} c_{sk} d_{sl}) (\sum_{r=0}^{1} e_{ri} f_{rj} g_{rk} h_{rl}), \quad (i,j,k,l) \in \{0,1\}^4.$$

Equivalently, find the kernel of the map $\mathbb{C}[p] \to \mathbb{C}[a, \ldots, h]$.

Which tools can we use? Gröbner bases? resultants? numerical homotopy continuation? generic points?

Geometry of the Model

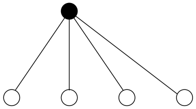
$\mathbb{P}^1$

$\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \hookrightarrow \mathbb{P}^{15}$

Segre variety, defined by

2×2 minors of flattenings of a $2 \times 2 \times 2 \times 2$ matrix

degree 24, dimension 4, toric variety



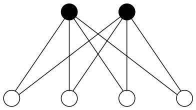
Secant variety of the Segre variety

defined by

3×3 minors of the flattenings

[Landsberg–Manivel]

degree 64, dimension 9



Hadamard product of two copies of secant variety

Defining polynomial? degree?

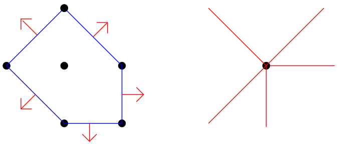
expected to be a hypersurface

has symmetries of the 4-cube

Our Tool: Tropical Geometry

The tropical hypersurface of a polynomial $f \in k[x_1, \ldots, x_n]$ is

$$\begin{aligned} \mathcal{T}(f) &= \text{codim-1 part of normal fan of the Newton polytope of } f \\ &= \{w \in \mathbb{R}^n : \text{in}_w(f) \text{ is not a monomial}\} \end{aligned}$$



The tropical variety of an ideal $I \subset k[x_1, \ldots, x_n]$ is

$$\mathcal{T}(I) = \{w \in \mathbb{R}^n : \text{in}_w(I) \text{ contains no monomial}\}.$$

- $\mathcal{T}(I)$ is a weighted polyhedral fan satisfying balancing condition.
- If I is prime then $\mathcal{T}(I)$ is pure of the same dimension as I and connected in codimension one.
- The tropical Bézout and Bernstein Theorems hold.

References: Bergman, Bieri–Groves, Kapranov–Lind–Einsiedler, Mikhalkin, Bogart–Jensen–Speyer–Sturmfels–Thomas, Maclagan–Sturmfels, ...

Applications to Computational algebra

From the tropical variety $\mathcal{T}(I)$ with multiplicities, we can compute [Dickenstein–Feichtner–Sturmfels (2005)]:

- the (multi)degree of I
- the Chow polytope if I is homogeneous
- the Newton polytope of the generator if I is principal.
  maximal cones in tropical variety $\leftrightarrow$ edges of polytope
  multiplicity $\leftrightarrow$ lattice length of edge

Remarks

- In the hypersurface case, we may be able to recover coefficients by interpolation or other methods (ref: Chris Peterson's talk on Monday).
- Even partial information about tropical varieties can be used to give bounds for invariants, e.g. dimension, degree, ...
- May be helpful for Gröbner bases computations.

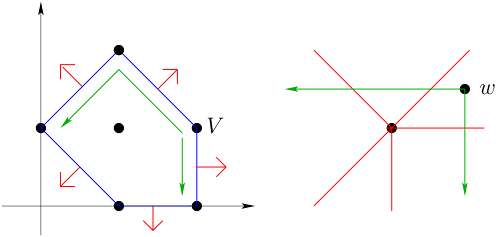
Tropical Hypersurface $\mathcal{T}(f) \rightsquigarrow$ Newton Polytope NP(f)

Theorem [DFS]: Let $w \in \mathbb{R}^n$ be a generic vector and V be the vertex of the polytope NP(f) that attains the maximum of $\{w \cdot x : x \in \text{NP}(f)\}$.

Then the i^{th} coordinate of the vertex V equals

$$\sum_{v \in \mathcal{T}(f) \cap (w - \mathbb{R}_{>0} e_i)} m_v \cdot l_{v,i}$$

where m_v is the multiplicity of v in $\mathcal{T}(f)$, and $l_{v,i}$ is the i^{th} coordinate of primitive integral normal vector to $\mathcal{T}(f)$ at v.



Knowledge of fan structure of $\mathcal{T}(f)$ is unnecessary.

We call this the ray shooting method for computing vertices of the polytope.

Generalizes to orthant shooting method for the Chow polytope.

How to compute tropical varieties?

From the generators of I we can compute $\mathcal{T}(I)$ using Gfan.

Sometimes we can compute $\mathcal{T}(I)$ without knowing generators of I, e.g.

- A-discriminants [DFS]
- implicitization with generic coefficients [Sturmfels–Tevelev–Y.]
- elimination (image under a monomial map of a variety with known tropicalization) [Sturmfels–Tevelev]

In all these cases, we get tropical varieties as sets, not as fans.
Open Problem: How to compute a fan structure of a union of cones?

- Suppose $X \subset \mathbb{C}^m, Y \subset \mathbb{C}^n, X \times Y \subset \mathbb{C}^m \times \mathbb{C}^n$. Then
$$\mathcal{T}(X \times Y) = \mathcal{T}(X) \times \mathcal{T}(Y)$$
as weighted polyhedral complexes, with $m_{\sigma\times\tau} = m_\sigma m_\tau$ for maximal cones $\sigma \subset \mathcal{T}(X), \tau \subset \mathcal{T}(Y)$, and $\sigma \times \tau \subset \mathcal{T}(X \times Y)$.
- Suppose $X, Y \subset \mathbb{C}^m$, and $X \cdot Y \subset \mathbb{C}^m$ is the Hadamard product. Then
$$\mathcal{T}(X \cdot Y) = \mathcal{T}(X) + \mathcal{T}(Y).$$
as sets, with multiplicities given by the Tropical Elimination Theory [ST].

Back to the Challenge: Tropicalizing the Model

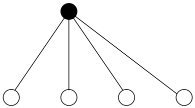
$\mathbb{P}^1$

tropical variety is $\mathbb{TP}^1 = \mathbb{R}^2/(1,1)$

$\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \hookrightarrow \mathbb{P}^{15}$

4-dimensional linear space in

$\mathbb{TP}^{15} = \mathbb{R}^{16}/(1,1,\ldots,1)$

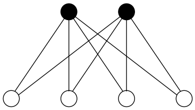


secant variety of the Segre embedding

9-dim fan with 4-dim lineality space in $\mathbb{TP}^{15}$

7680 maximal cones, all with multiplicity 1

computed using Gfan, with symmetry



Hadamard product of two copies of secant variety

Minkowski sum of two copies of the fan

union of 7680^2 cones; $6\,865\,824$ are full dimensional

multiplicities computed with Macaulay 2

unknown fan structure

14-dim fan with 4-dim lineality space in $\mathbb{TP}^{15}$

normal fan of the unknown Newton polytope

Implicitization Challenge: The Degree

With ray-shooting method, we found some vertices of the Newton polytope:

$$(0, 0, 0, 17, 10, 10, 12, 6, 16, 9, 1, 12, 10, 0, 6, 1)$$
$$(0, 0, 1, 17, 13, 6, 17, 1, 17, 1, 6, 13, 1, 17, 0, 0)$$
$$(1, 17, 17, 3, 2, 1, 12, 2, 5, 1, 2, 9, 9, 19, 7, 3)$$

Any one of them gives the multidegree of the polynomial.

Theorem: The defining equation of the model has multidegree (110, 55, 55, 55, 55) with respect to the following grading.

$$\begin{pmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1 \\ 0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1 \\ 0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1 \\ 0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1 \end{pmatrix}$$

LattE: There are 5 529 528 561 944 monomials with this multidegree.

How many of them actually appear in the polynomial?
Can we recover the coefficients?

Computing the polytope

We knew from the tropical variety that the Newton polytope contains 15 788 distinct edge directions (124 up to symmetry).

Finding one vertex using the ray shooting method

- go through 6 865 824 cones, 16 linear programs per cone (reduces to solving linear equations in this case because all cones are simplicial)
- Macaulay 2 took 3 days per vertex
- Python took 10 hours per vertex (3 hours with caching)
- C++ (with GMP) took 45 minutes per vertex
- highly parallelizable, but we did not do this.

Generating more vertices

- walk from chamber to chamber, using data from ray shooting
- use symmetry and parallelize

Computing facets

- compute the facets of tangent cones at found vertices using Polymake
- use knowledge of edge directions (very important!)
- check whether an inequality is a facet inequality of the polytope
- use symmetry and parallelize

We are done when all the facets of found tangent cones are certified as actual facets of the Newton polytope.

Newton Polytope

After a lot of computation ...

Theorem: The Newton polytope of the defining equation of the model is an 11 dimensional polytope in $\mathbb{R}^{16}$ with:

$17\,214\,912$	vertices in	$44\,938$	orbits
$70\,646$	facets in	246	orbits.

Full list at: `http://people.math.gatech.edu/~jyu67/ImpChallenge`

Among the 44 938 orbits of vertices, 215 have size 192 and the rest have size 384.

Orbit sizes of facets:

size	2	8	12	16	24	32	48	64	96	192	384
number of facet orbits	1	2	1	3	1	1	7	3	15	67	145

Coordinate hyperplanes form the "largest" facets, containing 3 907 356 vertices each.

Open Problem: How to compute the facets of the polytope from its normal fan without computing the vertices?

For comparison, the Newton polytope of the $2 \times 2 \times 2 \times 2$ GKZ-hyperdeterminant (projective dual of the Segre variety of $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ in $\mathbb{P}^{15}$) has :

$25\,448$	vertices in	111	orbits
268	facets in	8	orbits

It is a polynomial in the same (or dual) variables, with the same symmetry group and homogeneity space, with degree 24 and 2 894 276 monomials (out of 3 151 812 monomials with the same multi-degree) [Huggins–Sturmfels–Y.–Yuster].

Next steps ?

monomials?	Lattice points in the Newton polytope? (guess: 10^{12})
coefficients?	Linear algebra : integral, numerical, mod p, ... ? Use generic point: partial fractions, LLL, ... ?

Conclusion:

- Tropical geometry is useful for problems in computational algebra.
- For problems that are too large to solve completely, tropical geometry provides partial answers and bounds.

Forward looking:

Harmony of tropical geometry, Gröbner bases, and numerical algebraic geometry can be useful for problems in modern industrial society.

Thank you for your attention.

Rererence María Angélica Cueto, Enrique A. Tobis, and Josephine Yu. "An Implicitization Challenge for Binary Factor Analysis", to appear in *Journal of Symbolic Computation*. arXiv: 1006.1384.