A numerical toolkit for the likelihood correspondence

- **Jose Israel Rodriguez**, University of Wisconsin --- Madison
  - jRodriguez43@wisc.edu
  - https://sites.google.com/wisc.edu/jose/

- **Workshop on Computational and Applied Enumerative Geometry** (The Fields Institute)
  - June 3 - 7, 2024
- **Organizers**
  - Taylor Brysiewicz - Western University
  - Luis Garcia Puente - Colorado College
  - Nickolas Hein - Benedictine College
  - Jordy Lopez Garcia - Texas A&M University
  - Kevin Purbhoo - University of Waterloo
- **Goal:** Discuss our main character (likelihood correspondence) as a multiprojective variety with a (long term) view towards understanding biology (long branch attraction)
- **Outline**
  - Algebraic statistics: Invitation, ML degrees, likelihood geometry and maximum likelihood estimation
  - Numerical toolkit for a biprojective variety (bidegree geometrically)
  - Applications

Who are you, what do you do?

Little did we realize …

MR4297926 - Solving decomposable sparse systems
Brysiewicz, Taylor; Rodriguez, Jose Israel; Sottile, Frank; Yahl, Thomas
Numer. Algorithms **88** (2021), no. 1, 453–474.

MR4285764 - Decomposable sparse polynomial systems
Brysiewicz, Taylor; Rodriguez, Jose Israel; Sottile, Frank; Yahl, Thomas
J. Softw. Algebra Geom. **11** (2021), no. 1, 53–59.

MR3810571 - Trace test
Leykin, Anton; Rodriguez, Jose Israel; Sottile, Frank
Arnold Math. J. **4** (2018), no. 1, 113–125.

MR4166467 - A numerical toolkit for multiprojective varieties
Hauenstein, Jonathan D.; Leykin, Anton; Rodriguez, Jose Israel; Sottile, Frank
Math. Comp. **90** (2021), no. 327, 413–440.
(Reviewer: Dayton, Barry H.)

MR3833644 - Numerical computation of Galois groups
Hauenstein, Jonathan D.; Rodriguez, Jose Israel; Sottile, Frank
Found. Comput. Math. **18** (2018), no. 4, 867–890.
(Reviewer: Dayton, Barry H.)

Algebraic Statistics

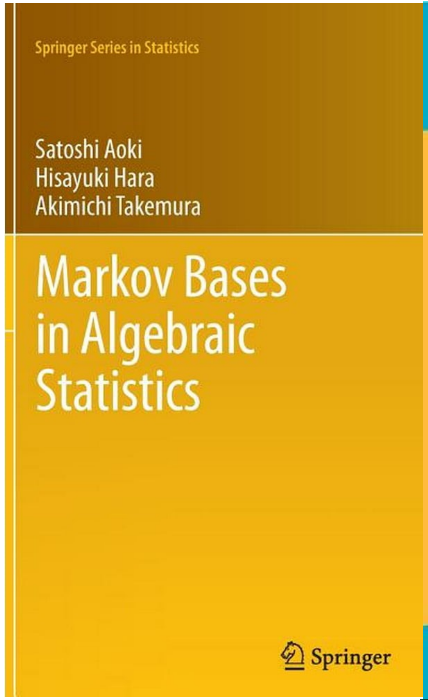
An invitation

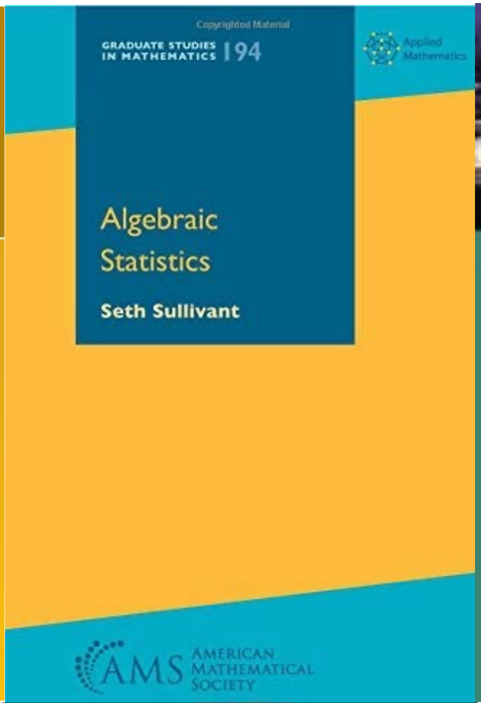
Algebraic Statistics

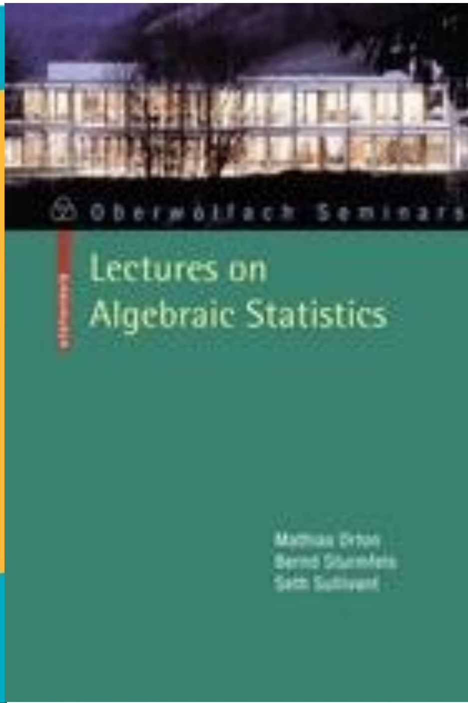
Interdisciplinary field with applications

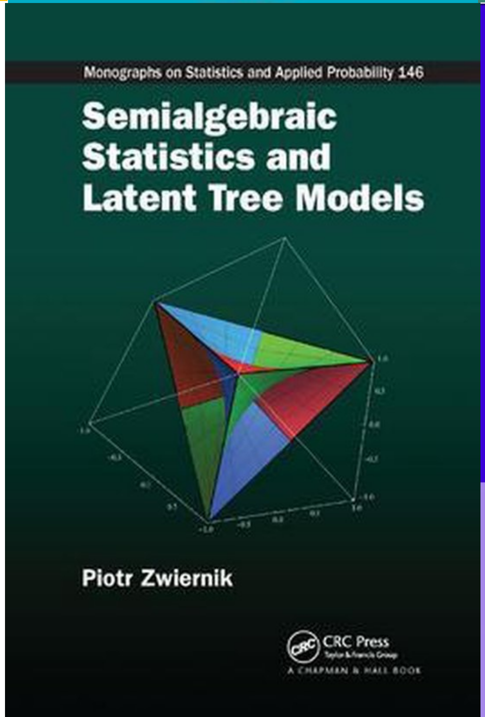
- **Textbooks**: multiple to choose from
  - Algebraic geometry
  - Combinatorics
  - Commutative algebra
  - More geometry
- **Dictionary** between algebra and statistics
  - Computation drives these connections
    - Symbolic and numerical

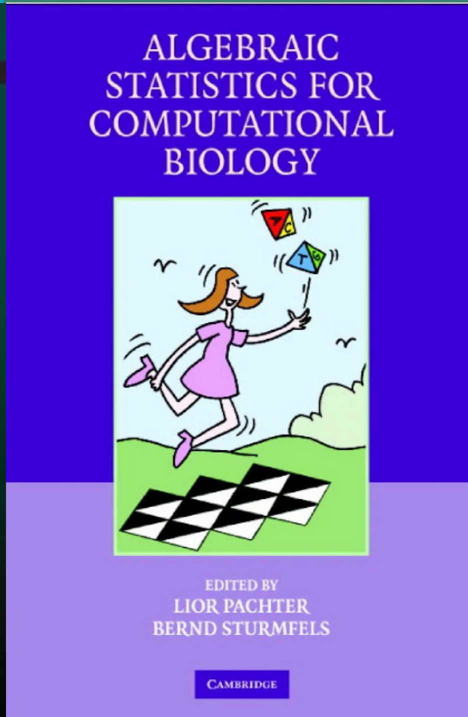
Probability/Statistics	Algebra/Geometry
Probability distribution	Point
Statistical model	(Semi)Algebraic set
Discrete exponential family	Toric variety
Conditional inference	Lattice points in polytopes
Maximum likelihood estimation	Polynomial optimization
Model selection	Geometry of singularities
Multivariate Gaussian model	Spectrahedral geometry
Phylogenetic model	Tensor networks
MAP estimates	Tropical geometry







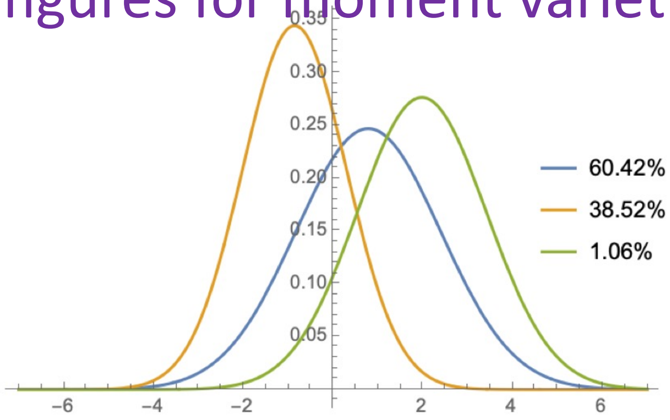


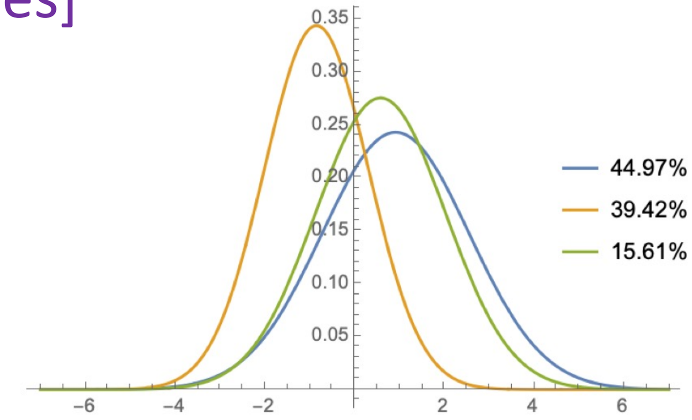


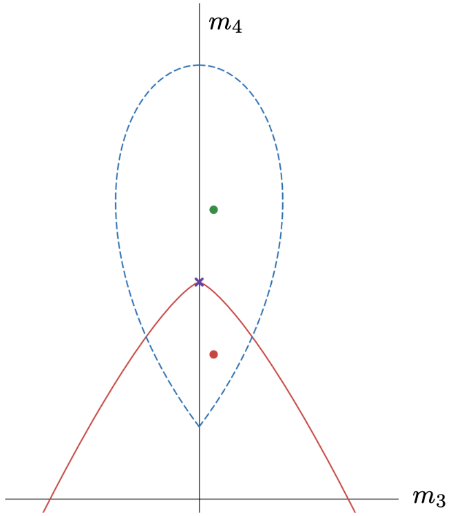
Algebraic Statistical Model (Discrete models)

A variety restricted to a probability simplex

- Probability simplex Δ_n
  - Set of points with nonnegative coordinates summing to one
  - i-th coordinate is the probability of observing event i
- From a variety **construct** an algebraic statistical model:
  - Restrict an affine variety to Δ_n
  - For a projective variety, the hyperplane at infinity is given by the sum of coordinates
- **Often**: the models have a parametric description
  - Finding an implicit description is phrased as an elimination problem
- **Remark**: we take a narrow notion of an algebraic statistical model today
  - Wide range of models I will not discuss
  - E.g., [Lindberg's figures for moment varieties]







An Example: Hardy-Weinberg Curve

Algebraic statistical model: Restricting a curve X to the probability simplex using elimination.

- Algebraic curve X parameterized by

$$\mathbb{R} \to \mathbb{R}^3, \quad \theta \mapsto (p_0, p_1, p_2) = ((1-\theta)^2, 2\theta(1-\theta), \theta^2)$$

- **Intuition:** If $\theta = \frac{1}{2}$ then after flipping a coin twice:
  p_0 is the probability of observing 0 heads, …. .
  https://www.desmos.com/3d/lk7lmsmvg5
- **Implicit relations:**
  Find algebraic constraints
  on (p_0, p_1, p_2)
- **Inspection:** [board]

iMSi Institute for Mathematical and Statistical Innovation

- (Sept 18-22, 2023): "Invitation to algebraic statistics and applications"
  - Jose Rodriguez, Serkan Hosten, Kaie Kubjas, Thomas Kahle, Fatemeh Mohammadi, Guillaume A. Pouliot.
- (Oct 9-13, 2023) Alg Stat for Ecological and Biological Systems
  - **Elizabeth Gross,** Elina Robeva, Seth Sullivant , Eliana Duarte.
- (Nov 6-10, 2023) Algebraic Economics
  - **Sonja Petrović,** Piotr Zwiernik, Ngoc Tran, Mladen Kolar.
- (Dec 11-15) New directions in algebraic statistics: Statistical Learning, Neural networks, Bayesian statistics
  - Mathias Drton, Lek-Heng Lim, Jonathan Hauenstein, Debdeep Pati.

Data & Information • Uncertainty Quantification

Algebraic Statistics and Our Changing World

New Methods for New Challenges

September 18 — December 15, 2023

MD **Mathias Drton** Technical University of Munich

EG **Elizabeth Gross** University of Hawaiʻi at Mānoa

LL **Lek-Heng Lim** University of Chicago

SP **Sonja Petrović** Illinois Institute of Technology

ER **Elina Robeva** University of British Columbia

JR **Jose Rodriguez** University of Wisconsin – Madison

PZ **Piotr Zwiernik** University of Toronto & Pompeu Fabra University

BS **Bernd Sturmfels** MPI Leipzig & UC Berkeley

BACKGROUND:
Invitation To Alg Stat recordings:
https://www.imsi.institute/activities/invitation-to-algebraic-statistics-and-applications/

✓ Maximum likelihood degrees

Models + Data = Likelihood correspondence

An Example: Hardy-Weinberg Curve

Continuing on with the example: maximum likelihood estimation

- Algebraic curve X parameterized by

$$\mathbb{R} \to \mathbb{R}^3, \quad \theta \mapsto (p_0, p_1, p_2) = ((1-\theta)^2, 2\theta(1-\theta), \theta^2)$$

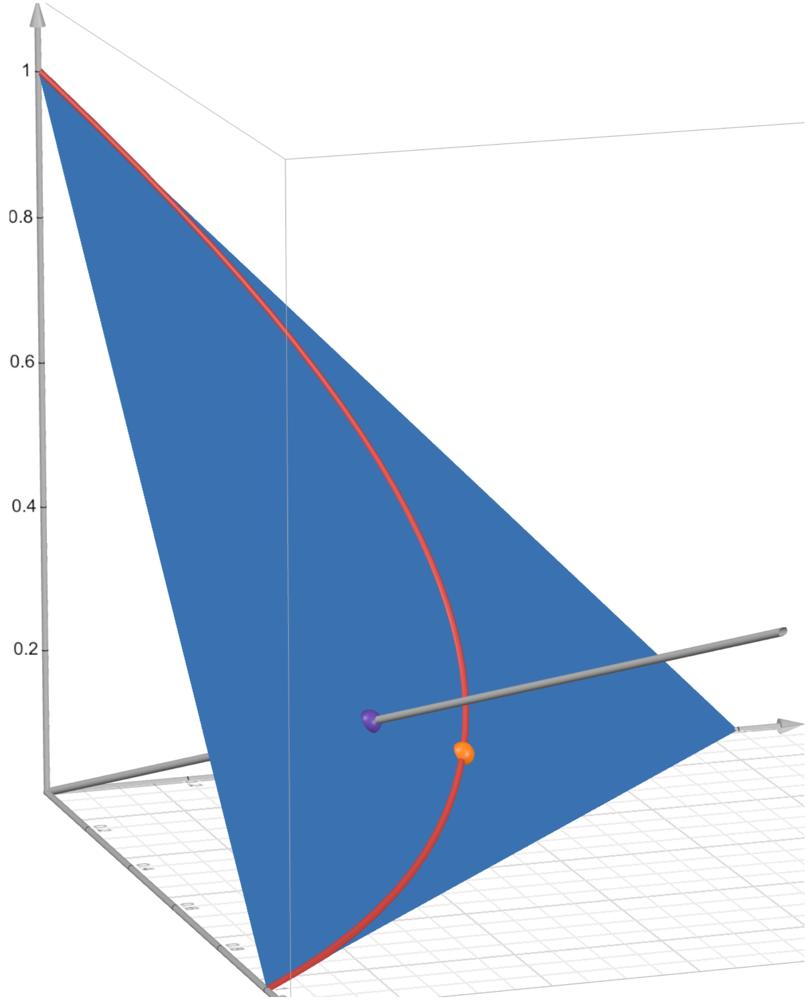
- **Intuition:** If $\theta = \frac{1}{2}$ then after flipping a coin twice: p_0 is the probability of observing 0 heads, …. . https://www.desmos.com/3d/lk7lmsmvg5
- **Implicit relations:** Find algebraic constraints on (p_0, p_1, p_2)
- **Inspection:** [board]
- Biased coin is flipped two times
  - Unknown: the probability of a heads θ.
- **Aim:** Recover (1) the parameter θ and (2) probability of observing *k=0,1,2* heads
- **Approach:** *Maximum likelihood estimation*

- **Data** is iid: (Sample size *N*=10) HH, HH, HT, TH, TT, TT, HT, TT, TT, TT
- Summarize as the data vector: (5,3,2)
- For generic data, this likelihood function [board] has a unique critical point
- **ML degree of the model**
  - counts critical points
  - # of solutions to likelihood eqs
  - Upper bound on the algebraic complexity of MLE
- ML degree one example https://www.desmos.com/3d/cjidlbag7r

Likelihood correspondence

Capture the geometry of maximum likelihood estimation

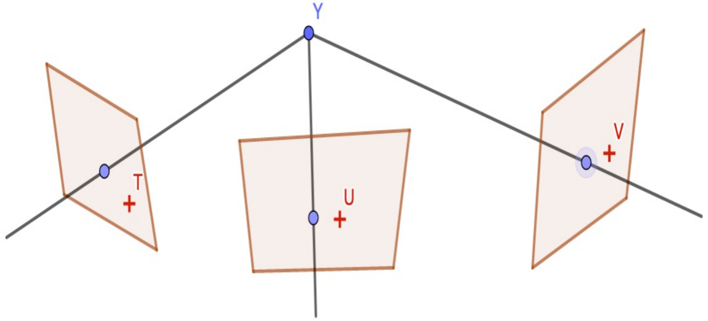
- **Likelihood correspondence**: encapsulates the geometry of mle (optimization problem)
  - **Remark**: although subtleties for boundary components and singular locus
- The likelihood correspondence is a biprojective variety
  - I.e., a multiprojective variety
- In equations: [see blackboard]

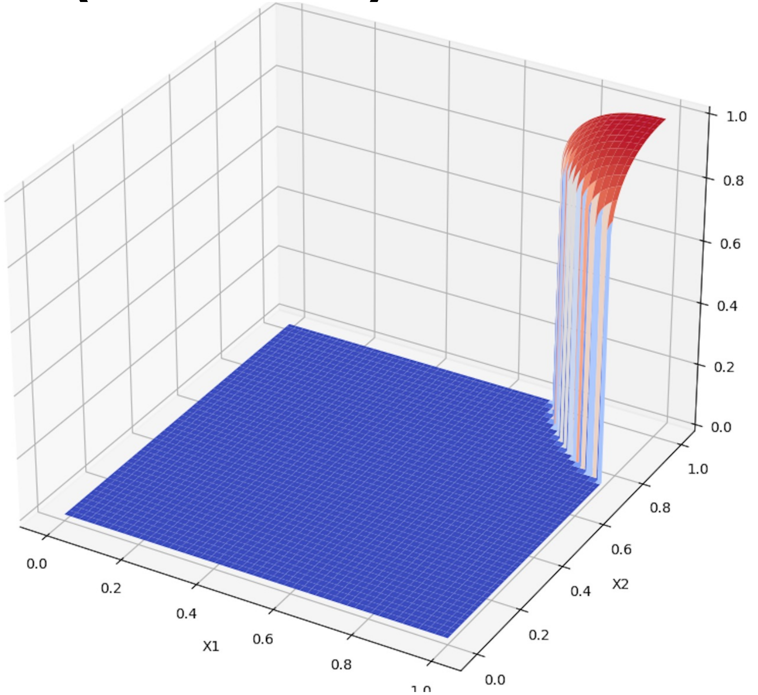


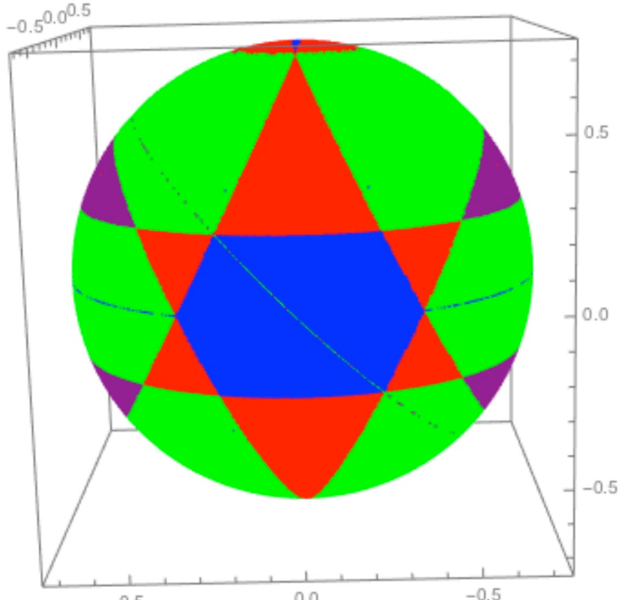
Multiprojective varieties

Multidegrees through a geometric lens

Multiprojective applications + projective (starter) toolkit







- **Broader Impactful Applications**
  - Computer vision
  - Power systems
  - Economics
- **Numerical Toolkit** (Projective starter pack) [Blackboard]
  - **Idea** (Bertini's Thm): intersect with a general linear space of complimentary dimension
    - Read off invariants like degree and dimension
    - Sampling
    - Component membership test
- **Trust building**: a trace test for completeness
  - Irreducible curve in the plane (on an affine chart) [Blackboard]
- **Trace Test Thm**: The sums are a linear function in t.

If we only use continuation on a proper subset of intersection points, the resulting sums are **not** a linear function in t.

  - Irreducibility of the curve is crucial, as is genericity of linear's.

Multidegrees

- **Recall projective setting**:
  - **Input**: integer *c*
  - Construct a general *c* dimensional linear space
  - **Output**: the number of isolated points in intersection
  - **Idea**: The degree of the union of irreducible codimension *c* components
- **Biprojective setting:**
  - **Input** *(a,b)* integers such that *a+b=c*
  - Construct *(A,B)* general linear spaces with dimension *(a,b)=(dim(A) , dim(B))*
  - **Output**: the number of isolated points in intersection
  - **Convenient multidegree notation**:
    - polynomial with integer coefficients counting intersection points

$$d_{(c,0)} \cdot \mathrm{p}^c + d_{(c-1,1)} \cdot \mathrm{p}^{c-1}\mathrm{u} + \cdots + d_{(1,c-1)} \cdot \mathrm{p}\mathrm{u}^{c-1} + d_{(0,c)} \cdot \mathrm{u}^c$$

- **Example:** Irreducible curve [Blackboard]

$$\mathbb{V}(xu^2 - 36)$$

A trace test

Curves in an affine chart

A trace test by example

- Consider $\mathbb{C} \times \mathbb{C}$ as an affine chart of $\mathbb{P}^1 \times \mathbb{P}^1$
- For t=0, three points of intersection:

$$\mathbb{V}(xu^2 - 36) \cap \mathbb{V}\left((x-1)(u-3) + 3t \right)$$

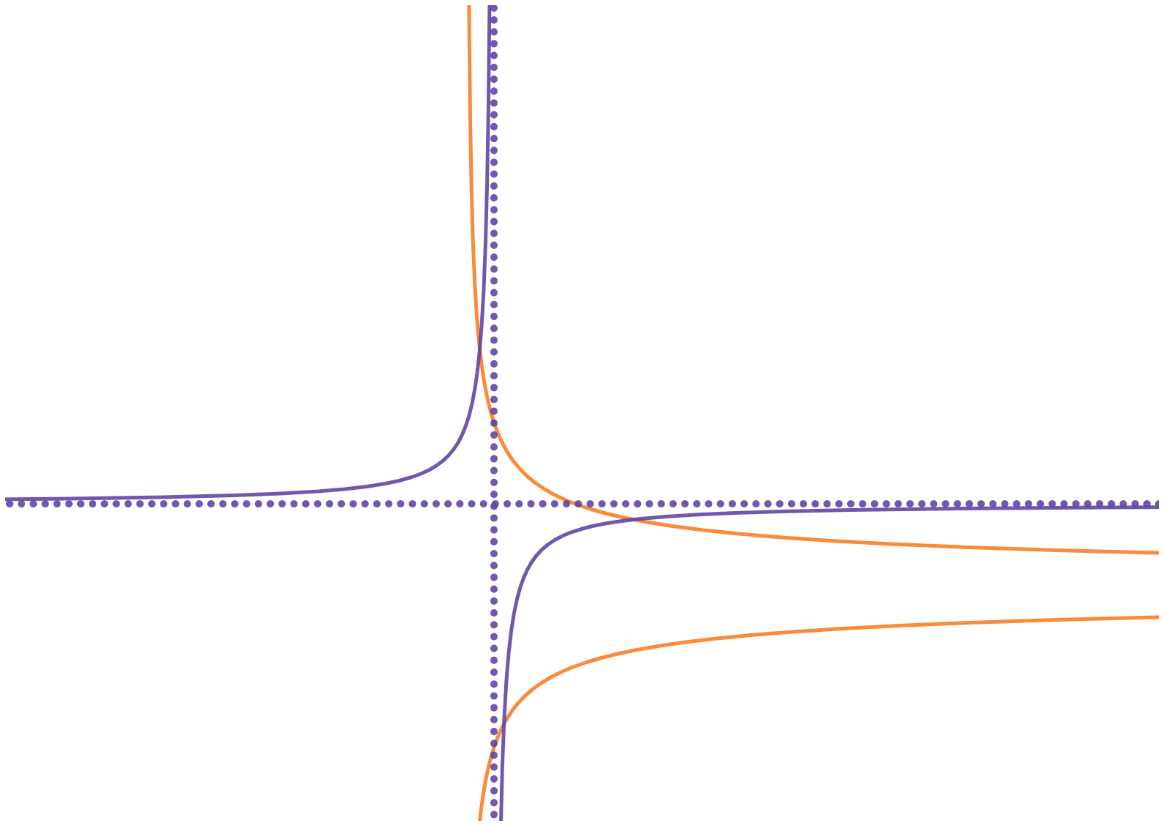
- Irreducible curve $\mathbb{V}(xu^2 - 36)$
- Union of linear's: $\mathbb{V}(x-1) \cup \mathbb{V}(u-3)$
- For $(x,u) \in \mathbb{C} \times \mathbb{C}$ construct the

  rank 1 matrix $$\begin{bmatrix} x \\ 2+3x \end{bmatrix} \begin{bmatrix} u \\ 5+7u \end{bmatrix}^T \in \mathbb{P}^3$$

- The points give matrices that we sum:

$$\begin{bmatrix} 1 \\ 5 \end{bmatrix} \begin{bmatrix} 6 \\ 47 \end{bmatrix}^T + \begin{bmatrix} 1 \\ 5 \end{bmatrix} \begin{bmatrix} -6 \\ -37 \end{bmatrix}^T + \begin{bmatrix} 4 \\ 14 \end{bmatrix} \begin{bmatrix} 3 \\ 26 \end{bmatrix}^T = \begin{bmatrix} 12 & 114 \\ 42 & 414 \end{bmatrix}$$

- **Continuation**: vary t from 0 to an integer.
  - Get three new points;
    Get three new matrices;
    Get their sum.



$$t=0 \quad \begin{bmatrix} 12 & 114 \\ 42 & 414 \end{bmatrix} \qquad t=1 \quad \begin{bmatrix} 12 & 124 \\ 48 & 486 \end{bmatrix} \qquad t=2 \quad \begin{bmatrix} 12 & 134 \\ 54 & 558 \end{bmatrix} \qquad t=3 \quad \begin{bmatrix} ? & ? \\ ? & 630 \end{bmatrix}$$

A trace test by example

- Consider $\mathbb{C} \times \mathbb{C}$ as an affine chart of $\mathbb{P}^1 \times \mathbb{P}^1$
- For t=0, three points of intersection:

$$\mathbb{V}(xu^2 - 36) \cap \mathbb{V}\left((x-1)(u-3) + 3t \right)$$

- Irreducible curve $\mathbb{V}(xu^2 - 36)$
- Union of linear's: $\mathbb{V}(x-1) \cup \mathbb{V}(u-3)$
- For $(x,u) \in \mathbb{C} \times \mathbb{C}$ construct the

rank 1 matrix $$\begin{bmatrix} x \\ 2+3x \end{bmatrix} \begin{bmatrix} u \\ 5+7u \end{bmatrix}^T \in \mathbb{P}^3$$

- The points give matrices that we sum:

$$\begin{bmatrix} 1 \\ 5 \end{bmatrix} \begin{bmatrix} 6 \\ 47 \end{bmatrix}^T + \begin{bmatrix} 1 \\ 5 \end{bmatrix} \begin{bmatrix} -6 \\ -37 \end{bmatrix}^T + \begin{bmatrix} 4 \\ 14 \end{bmatrix} \begin{bmatrix} 3 \\ 26 \end{bmatrix}^T = \begin{bmatrix} 12 & 114 \\ 42 & 414 \end{bmatrix}$$

- **Continuation**: vary t from 0 to an integer.
  - Get three new points;
    Get three new matrices;
    Get their sum.

$$\overset{t=0}{\begin{bmatrix} 12 & 114 \\ 42 & 414 \end{bmatrix}} \quad \overset{t=1}{\begin{bmatrix} 12 & 124 \\ 48 & 486 \end{bmatrix}} \quad \overset{t=2}{\begin{bmatrix} 12 & 134 \\ 54 & 558 \end{bmatrix}} \quad \overset{t=3}{\begin{bmatrix} ? & ? \\ ? & 630 \end{bmatrix}}$$

A trace test by example

- Consider $\mathbb{C} \times \mathbb{C}$ as an affine chart of $\mathbb{P}^1 \times \mathbb{P}^1$
- For t=0, three points of intersection:

$$\mathbb{V}(xu^2 - 36) \cap \mathbb{V}\left((x-1)(u-3) + 3t \right)$$

- Irreducible curve $\mathbb{V}(xu^2 - 36)$
- Union of linear's: $\mathbb{V}(x-1) \cup \mathbb{V}(u-3)$
- For $(x,u) \in \mathbb{C} \times \mathbb{C}$ construct the

rank 1 matrix $$\begin{bmatrix} x \\ 2+3x \end{bmatrix} \begin{bmatrix} u \\ 5+7u \end{bmatrix}^T \in \mathbb{P}^3$$

- The points give matrices that we sum:

$$\begin{bmatrix} 1 \\ 5 \end{bmatrix} \begin{bmatrix} 6 \\ 47 \end{bmatrix}^T + \begin{bmatrix} 1 \\ 5 \end{bmatrix} \begin{bmatrix} -6 \\ -37 \end{bmatrix}^T + \begin{bmatrix} 4 \\ 14 \end{bmatrix} \begin{bmatrix} 3 \\ 26 \end{bmatrix}^T = \begin{bmatrix} 12 & 114 \\ 42 & 414 \end{bmatrix}$$

- **Continuation**: vary t from 0 to an integer.
  - Get three new points;
    Get three new matrices;
    Get their sum. $\underset{t=0}{\begin{bmatrix} 12 & 114 \\ 42 & 414 \end{bmatrix}}$ $\underset{t=1}{\begin{bmatrix} 12 & 124 \\ 48 & 486 \end{bmatrix}}$ $\underset{t=2}{\begin{bmatrix} 12 & 134 \\ 54 & 558 \end{bmatrix}}$ $\underset{t=3}{\begin{bmatrix} ? & ? \\ ? & 630 \end{bmatrix}}$

- **Thm**: The sums are a linear function in t.
- If we only use continuation on a proper subset of intersection points, the resulting sums are **not** a linear function in t.
  - Irreducibility of the curve is crucial, as is genericity of linear's.
- **Takeaway:** verifies completeness of a pair of multidegrees.
  - Generalizes so long as you can restrict your positive dimensional variety to the curve case
    - Issue: Cartesian products of curves

Likelihood correspondence

Bidegrees for polynomial optimization

Likelihood correspondence (Continued)

Capture the geometry of maximum likelihood estimation

- **Likelihood correspondence**: encapsulates the geometry of mle (optimization problem)
  - **Remark**: although subtleties for boundary components and singular locus
- The likelihood correspondence is a biprojective variety
  - I.e., a multiprojective variety
- In equations: [see blackboard]
- What is the multidegree of our red curve example?
  - ML degree is 1
  - The model's degree is 2.

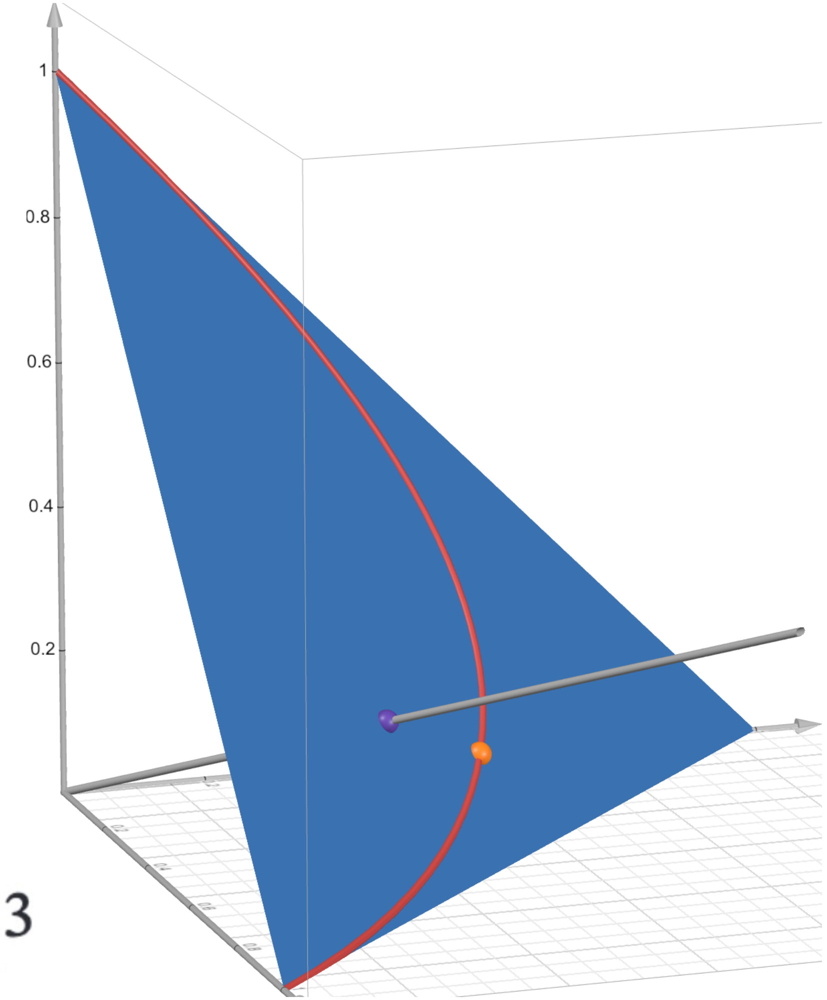
$$p^2 + 2pu$$

- **2x2x2 rank one tensors**: ML degree one; Degree of model is 6
  - Bidegree of likelihood correspondence

$$p^7 + 3p^6u + 6p^5u^2 + 6p^4u^3$$

- Restricted to a general hyperplane: ML degree is 15
  - Further restricted to another hyperplane: ML degree is 18
- **Sectional ML degree**

$$p^7 + 15p^6u + 18p^5u^2 + 6p^4u^3$$



Huh-Sturmfels Involution

- [HS] Assuming the model is smooth, given the bidegree of the likelihood correspondence **produce** the sectional ML degrees and vice versa

$$B_Y(\mathrm{p},\mathrm{u}) = \frac{\mathrm{u}\cdot S_Y(\mathrm{p},\mathrm{u}-\mathrm{p}) - \mathrm{p}\cdot S_Y(\mathrm{p},0)}{\mathrm{u}-\mathrm{p}}$$

$$S_Y(\mathrm{p},\mathrm{u}) = \frac{\mathrm{u}\cdot B_Y(\mathrm{p},\mathrm{u}+\mathrm{p}) + \mathrm{p}\cdot B_Y(\mathrm{p},0)}{\mathrm{u}+\mathrm{p}}$$

- **Our contribution**: extend to singular situation

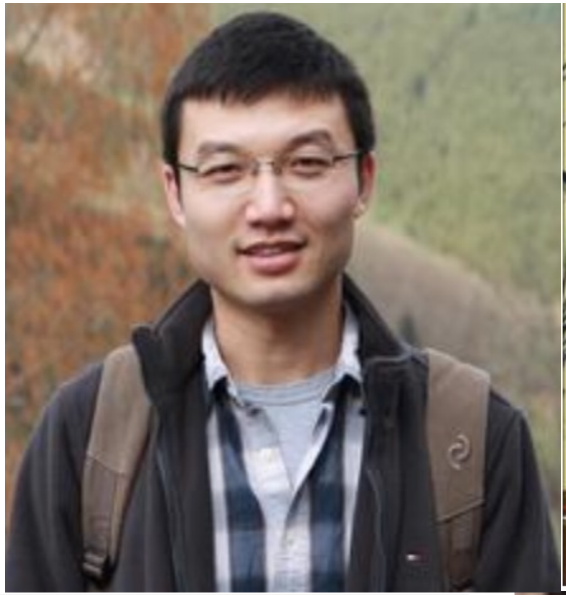


TABLE 2 ML Bidegree for rank one tensors.

k =dim(Y_k)	$n = 2^k - 1$	B_{Y_k} and S_{Y_k}
2	3	$p^3 + 2p^2u + 2pu^2$
		$p^3 + 4p^2u + 2pu^2$
3	7	$p^7 + 3p^6u + 6p^5u^2 + 6p^4u^3$
		$p^7 + 15p^6u + 18p^5u^2 + 6p^4u^3$
4	15	$p^{15} + 4p^{14}u + 12p^{13}u^2 + 24p^{12}u^3 + 24p^{11}u^4$
		$p^{15} + 64p^{14}u + 132p^{13}u^2 + 96p^{12}u^3 + 24p^{11}u^4$

Society for Industrial and Applied Mathematics

Program Co-Chairs:
Sonja Petrović
Carina Curto

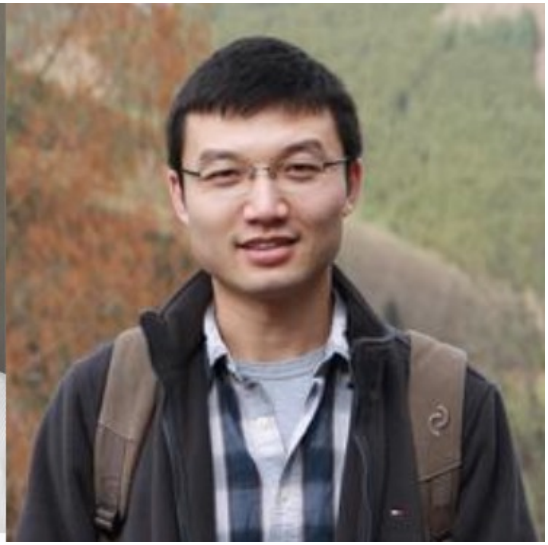
SIAM Activity Group on Algebraic Geometry

This activity group brings together researchers who use algebraic geometry in industrial and applied mathematics. We welcome participation from theoretical mathematical areas and those areas falling under the broadly interpreted notion of algebraic geometry and its applications.

SIAM AG25 at University of Wisconsin --- Madison

July 7 – 11, 2025

Local Organizers:
Gheorghe Craciun,
Laurentiu Maxim,
Jose Israel Rodriguez,
Botong Wang



3x3 Matrices with rank at most two

These correspond to a mixture model from statistics

- The ML degree is 10
  - Given a generic 3x3 data matrix, the respective likelihood function has 10 critical points
- The ML bidegree is

$$(10\mathrm{p}^8 + 24\mathrm{p}^7\mathrm{u} + \cdots) = (6\mathrm{p}^8 + 12\mathrm{p}^7\mathrm{u} + \cdots) + (4\mathrm{p}^8 + 12\mathrm{p}^7\mathrm{u} + \cdots)$$

- Symmetrize the data matrix to pass to the right hand side.
- **Idea:** Use coefficient parameter homotopy to test
  - Take data to a generic subvariety and see how many (if any!) critical points follow suit
- **Elizabeth Gross** and taking data to zero

ML table of 4×4 rank 2 matrices

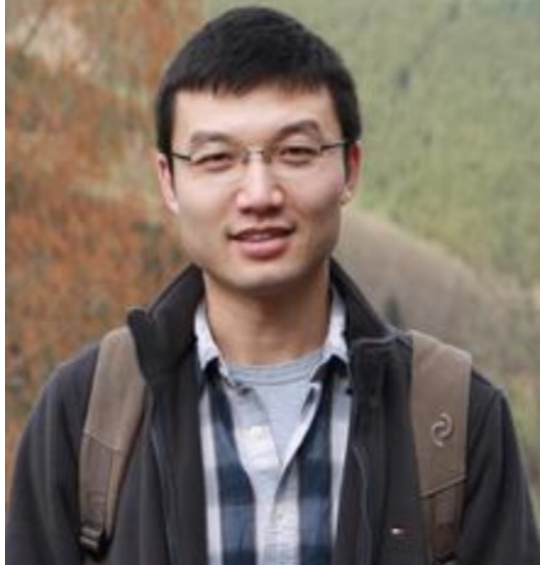
$R\backslash\backslash S$	$\{\}$	$\{11\}$	$\{11,44\}$	$\{11,22,44\}$
$\{\}$	191	118	76	51
$\{11\}$		73	42	25
$\{22\}$				25
$\{44\}$			42	25
$\{11,22\}$				17
$\{11,44\}$			31	17
$\{22,44\}$				17
$\{11,22,44\}$				14

ML table of 4×4 rank 3 matrices

$R\backslash\backslash S$	$\{\}$	$\{11\}$	$\{11,44\}$	$\{11,22,44\}$
$\{\}$	191	73	31	14
$\{11\}$		118	42	17
$\{22\}$				17
$\{44\}$			42	17
$\{11,22\}$				25
$\{11,44\}$			76	25
$\{22,44\}$				25
$\{11,22,44\}$				51

Offshoots and new directions

- Putting into practice
  - U-generation (Tim Duff and Anton Leykin)
    - Implementation: Equation by equation methods for solving polynomial systems
- Polynomial rings in infinitely many variables
- Conormal varieties are dimension *n*
  - Equation by equation method?
- Applications:
  - Long branch attraction (**Max Hill**, Wisconsin -> UC Riverside -> ICERM -> U of Hawaii)
    - Coefficient parameter homotopy and semi-algebraic descriptions
    - Software to solve these problems
    - Given generic data we can recover a phylogenetic tree(s) with four leaves
    - What does non-generic data mean?
  - Economic fragility in supply chains










Thank you!
+Coauthors!