Tensor Networks in Algebraic Geometry and Statistics

Jason Morton

Penn State

May 10, 2012
Centro de ciencias de Benasque Pedro Pascual

Supported by DARPA N66001-10-1-4040 and FA8650-11-1-7145.

What is algebraic geometry?

Study of solutions to systems of polynomial equations

- Multivariate polynomials $f \in \mathbb{C}[x_1, \ldots, x_n]$.
- The zero locus of a set of polynomials $\mathcal{F}$ is a variety $V(\mathcal{F})$.
- Given a set $S \subset \mathbb{C}^n$, the vanishing ideal of S is

$$I(S) = \{f \in \mathbb{C}[x_1, \ldots, x_n] : f(a) = 0 \; \forall a \in S\}.$$

  Such an ideal has a finite generating set. Closure $V(I(S))$.
- Implicitization: if $x = t$, $y = t^2$, $y - x^2 = 0$ cuts out the image.

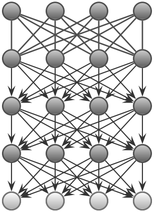
To an algebraic geometer, a tensor network

- appearing in statistics, signal processing, computational complexity, quantum computation, ...
- describes a regular map ϕ from the parameter space (choice of tensors at the nodes) to an ambient space.
- The image of ϕ is an algebraic variety of representable probability distributions, tensor network states, etc.

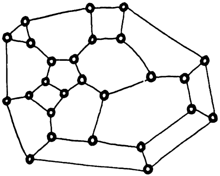
Why are geometers interested?

- Applications (especially tensor networks in statistics and CS) have revived classical viewpoints such as invariant theory.
- Re-climbing the hierarchy of languages and tools (Italian school, Zariski-Serre, Grothendieck) as applied problems are unified and recast in more sophisticated language.
- Applied problems have also revealed gaps in our knowledge of algebraic geometry and driven new theoretical developments
  - Objects which are "large": high-dimensional, many points, but with many symmetries
  - These often stabilize in some sense for large n.

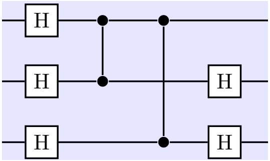
Tensor Networks



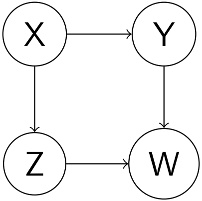
ML and Statistics

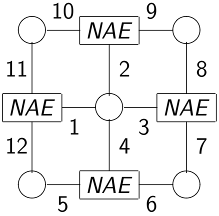


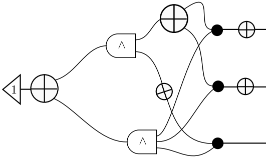
Complexity Theory



Quantum Information







Approximate Dictionary?

Tensor Networks in Physics	Graphical Models in Stats/ML
MPS	HMM
TTN	GMM
PEPS	CRF/MRF
MERA	?DBM?
DMRG	??

In Algebraic Statistics we have been studying the right-hand column

- often determining the ideal / variety / manifold (invariants)
- characteristics of the parameterization map
  - e.g. is it generically injective? Singular locus?
- generally work in complex projective space
  - so pure states are more natural than probabilities
- related optimization, contraction, approximation problems

Algebraic description of MPS

Fix parameter matrices $A_1, \ldots, A_d$.

$$\Psi = \sum_{i_1,\ldots,i_n} \mathrm{tr}(A_{i_1} \cdots A_{i_n}) |i_1 i_2 \cdots i_n\rangle$$

What are the polynomial relations that hold among the coefficients

$$\Psi_{i_1,\ldots i_n} = \mathrm{tr}(A_{i_1} \cdots A_{i_n})?$$

That is, the set of polynomials f in the coefficients such that $f(\Psi_{i_1,\ldots i_n}) = 0$. Organize these invariants into an ideal.

$$I = \{f : f(\Psi_{i_1,\ldots i_n}) = 0\}$$

the space of representable states is the variety $V(I)$ cut out by the invariants. See [Bray M- 2006] for some of them.

Possible applications of invariants of TNS?

- Simplify the computation of quantities of interest
  - e.g. Renyi entropy
- Representability and approximation error
  - which states/systems can be represented and which cannot?
  - bounds on approximation error
- Paths of optimization or time evolution on the manifold of representable states

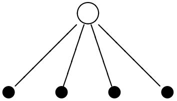
Some of the things we think about

Naïve Bayes / Secant Segre / Tensor Rank

Look at one hiden node in such a network, binary variables

$\mathbb{P}^1$

$\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \hookrightarrow \mathbb{P}^{15}$
Segre variety defined by
2×2 minors of flattenings
of $2 \times 2 \times 2 \times 2$ tensor



$\sigma_2(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$
First secant of Segre variety
3×3 minors of flattenings

Dimension of secant varieties

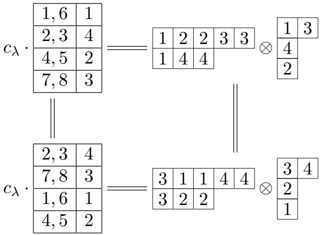
- Recently [Catalisano, Geramita, Gimigliano 2011] showed $\sigma_k(\mathbb{P}^1)^n$ has the expected dimension

$$\min(kn + k - 1, 2^n - 1)$$

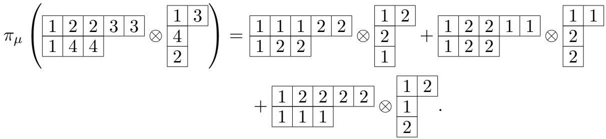
  except $\sigma_3(\mathbb{P}^1)^4$ where it is 13 not 14.
- Progress in Palatini 1909, . . . , Alexander Hirschowitz 1995, 2000, CGG 2002,03,05, Abo Ottaviani Peterson 2006, Draisma 2008, others.
- Classically studied, revived by applications to statistics, quantum information, and complexity; shift to higher secants, solution.
- So a generic tensor of $(\mathbb{C}^2)^{\otimes n}$ can be written as a sum of $\lceil \frac{2^n}{n+1} \rceil$ decomposable tensors, no fewer.

Representation theory of secant varieties

Raicu (2011) proved the ideal-theoretic GSS [Garcia Stillman Sturmfels 05] conjecture using representation theory of ideal of $\sigma_2(\mathbb{P}^{k_1} \times \cdots \times \mathbb{P}^{k_n})$ as a $GL_{k_1} \times \cdots \mathrm{GL}_{k_n}$-module (progress in [Landsberg Manivel 04, Landsberg Weyman 07, Allman Rhodes 08]).



Let's write down the action of the map π_μ on the tableaux pictured above



Representation theory

- Which tensor products $\mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$ have finitely many orbits under $\mathrm{GL}(d_1, \mathbb{C}) \times \cdots \times \mathrm{GL}(d_n, \mathbb{C})$?
- Related to SLOCC-equivalent entanglement classification
- Kac (1980), Parfenov (1998, 2001): up to $\mathbb{C}^2 \otimes \mathbb{C}^3 \otimes \mathbb{C}^6$, orbit representatives and abutment graph

Case $(2, m, n)$	The number of orbits of $\mathrm{GL}_2 \times \mathrm{GL}_m \times \mathrm{GL}_n$	$\deg f$
$(2,2,2)$	7	4
$(2,2,3)$	9	6
$(2,2,4)$	10	4
$(2,2,n), \quad n \geqslant 5$	10	0
$(2,3,3)$	18	12
$(2,3,4)$	24	12
$(2,3,5)$	26	0
$(2,3,6)$	27	6
$(2,3,n), \quad n \geqslant 7$	27	0

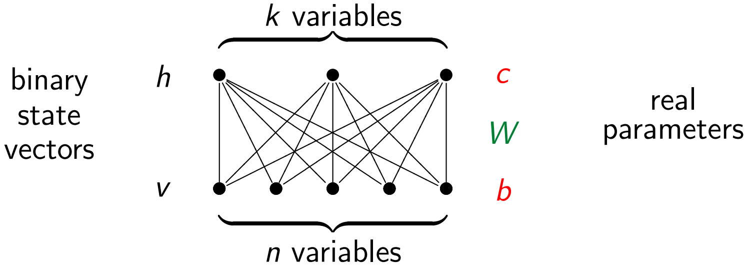
Computational Algebraic Geometry

- There are computational tools for algebraic geometry, and many advances mix computational experiments and theory.
- Gröbner basis methods power general purpose software: Singular, Macaulay 2, CoCoA, (Mathematica, Maple)
  - Symbolic term rewriting
- Numerical Algebraic Geometry: Numerical methods for approximating complex solutions of polynomial systems.
  - Homotopy continuation (numerical path following).
  - Can be used to find isolated solutions or points on each positive-dimensional irreducible component.
  - Can scale to thousands of variables for certain problems.

Identifiability: uniqueness of parameter estimates

- A parameterization of a set of probability distributions is identifiable if it is injective.
- A parameterization of a set of probability distributions is generically identifiable if it is injective except on a proper algebraic subvariety of parameter space.
- Identifiability questions can be answered with algebraic geometry (e.g. many recent results in phylogenetics)
- A weaker question: What conditions guarantee generic identifiability up to known symmetries?
- A still weaker question: is the dimension of the space of representable distributions (states) equal to the expected dimension (number of parameters)? Or are parameters wasted?

Graphical model on a bipartite graph



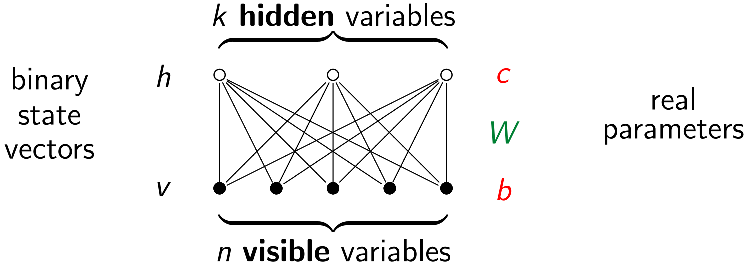
Unnormalized potential is built from node and edge parameters

$$\psi(v, h) \;=\; \exp(h^\top W v + b^\top v + c^\top h).$$

The probability distribution on the binary random variables is

$$p(v, h) \;=\; \frac{1}{Z} \cdot \psi(v, h), \qquad Z = \sum_{v,h} \psi(v, h).$$

Restricted Boltzmann machines



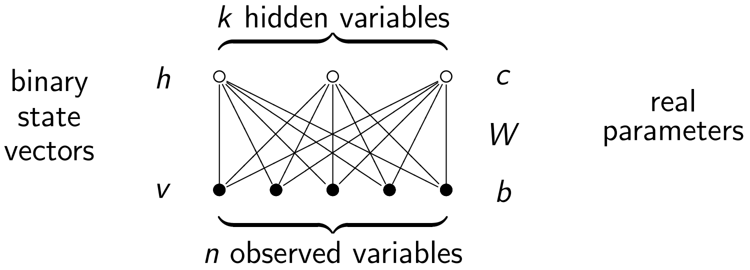
Unnormalized fully-observed potential is

$$\psi(v,h) \;=\; \exp(h^\top W v + b^\top v + c^\top h).$$

The probability distribution on the visible random variables is

$$p(v) \;=\; \frac{1}{Z}\cdot \sum_{h\in\{0,1\}^k} \psi(v,h), \qquad Z = \sum_{v,h} \psi(v,h).$$

Restricted Boltzmann machines



- The *restricted Boltzmann machine* (RBM) is the undirected graphical model for binary random variables thus specified.
- Denote by M_n^k the set of joint distributions as $b \in \mathbb{R}^n, c \in \mathbb{R}^k, W \in \mathbb{R}^{k \times n}$ vary.
- M_n^k is a subset of the probability simplex Δ_{2^n-1}.

Hadamard Products of Varieties

Given two projective varieties X and Y in $\mathbb{P}^m$, their *Hadamard product* $X * Y$ is the closure of the image of

$$X \times Y \dashrightarrow \mathbb{P}^m, \ (x, y) \mapsto (x_0 y_0 : x_1 y_1 : \ldots : x_m y_m).$$

We also define Hadamard powers $X^{[k]} = X * X^{[k-1]}$.

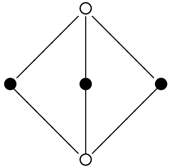
If M is a subset of the simplex Δ_{m-1} then $M^{[k]}$ is also defined by componentwise multiplication followed by rescaling so that the coordinates sum to one. This is compatible with taking Zariski closure: $\overline{M^{[k]}} = \overline{M}^{[k]}$

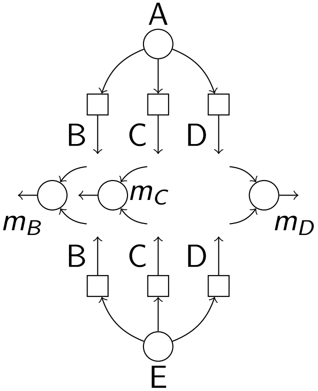
Lemma

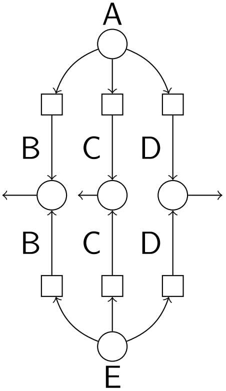
RBM variety and RBM model factor as

$$V_n^k = (V_n^1)^{[k]} \quad \textit{and} \quad M_n^k = (M_n^1)^{[k]}.$$

RBM as Hadamard product of naïve Bayes







Representational power of RBMs

Conjecture

The restricted Boltzmann machine has the expected dimension: M_n^k *is a semialgebraic set of dimension* $\min\{nk + n + k, 2^n - 1\}$ *in* $\Delta_{2^n - 1}$.

We can show many special cases and the following general result:

Theorem (Cueto M- Sturmfels)

The restricted Boltzmann machine has the expected dimension

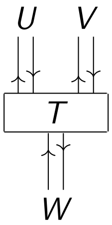
- $nk + n + k$ *when* $k < 2^{n - \lceil \log_2(n+1) \rceil}$
- $\min\{nk + n + k, 2^n - 1\}$ *when* $k = 2^{n - \lceil \log_2(n+1) \rceil}$ *and*
- $2^n - 1$ *when* $k \geq 2^{n - \lfloor \log_2(n+1) \rfloor}$.

- Covers most cases of restricted Boltzmann machines in practice, as those generally satisfy $k \leq 2^{n - \lceil \log_2(n+1) \rceil}$.
- Proof uses tropical geometry, coding theory

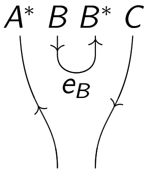
Computational complexity and efficient contraction

Secant varieties in algebraic complexity theory

A multilinear operator
$T : U \otimes V \to W$
is a tensor



The tensor rank $\min\{r : T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i\}$ of



$M : (A^* \otimes B) \times (B^* \otimes C) \to A^* \otimes C$
gives the exponent of matrix multiplication.

Satisfiability and #CSP problems

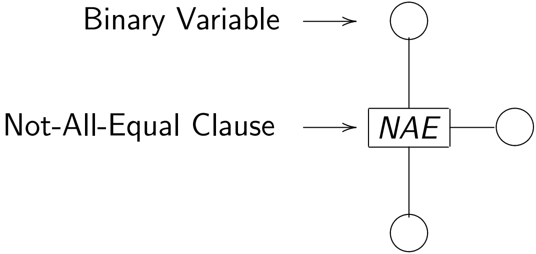
Given a problem P in conjunctive normal form:

- a collection of Boolean variables $x_1 \dots x_m$
- subject to clauses $c_1 \dots c_p$ (all must hold, each true or false), e.g. $OR(i) = 1$ if $i \in \{001, 010, 100, 011, 101, 110, 111\}$

Does there exist a satisfying assignment to the variables?

- Counting the **number** of satisfying assignments is computing a partition function, #P-complete in general.
- In [Landsberg, M-, Norine 2012] and [M- 2010], geometric interpretation and geometrically-motivated generalization of the holographic circuits of Valiant 04.
- Generates new families of efficiently contractable tensor networks
- Beyond noninteracting fermionic linear optics

Binary Variables and NAE clauses

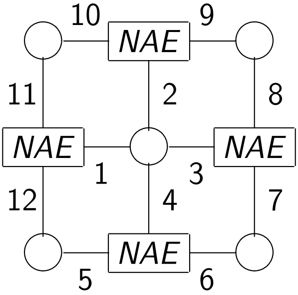


As a tensor, a Boolean predicate is the formal sum of the rows of its truth table as bitstrings.

$$OR_3 = (|0\rangle + |1\rangle)^{\otimes 3} - |000\rangle$$

.

Pfaffian circuit/kernel counting example

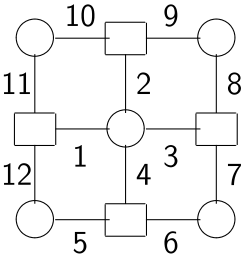


$\#$ of satisfying assignments $=$

$$\langle \text{all possible assignments, all restrictions} \rangle = \alpha\beta\sqrt{\det(x+y)}$$

4096-dimensional space $(\mathbb{C}^2)^{\otimes 12}$ $\qquad$ 12×12 matrix

Efficient contraction with Pfaffian circuits



$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$\begin{pmatrix}
0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1/3 & -1/3 \\
-1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & -1/3 & -1/3 & 0 & 0 \\
1 & -1 & 0 & 1 & 0 & 0 & -1/3 & -1/3 & 0 & 0 & 0 & 0 \\
-1 & 1 & -1 & 0 & -1/3 & -1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1/3 & 0 & -1/3 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1/3 & 1/3 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/3 & 0 & 0 & -1 & 0 & -1/3 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1 & 0 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1/3 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1 & 0 \\
1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1/3 \\
1/3 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0
\end{pmatrix}$$

$2^5 \cdot (\frac{6}{2^3})^4 \cdot \mathrm{Pfaff}(\tilde{z} + y) = 14$ satisfying assignments.

morton@math.psu.edu
www.math.psu.edu/morton/aspsu2012/