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, \ldots
- describes a regular map $\phi$ from the parameter space (choice of tensors at the nodes) to an ambient space.
- The image of $\phi$ 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 $n$. 
FIG. 19. Left (a) the circuit realization (internal to the triangle) of the function $f_W$... decomposition or factorization of a state into a tensor network is an entirely different problem which we address here.

Pfaffian circuit/kernel counting example

$\begin{array}{cccc}
\text{NAE} & \text{NAE} & \text{NAE} & \text{NAE} \\
\text{NAE} & \text{NAE} & \text{NAE} & \text{NAE} \\
\text{NAE} & \text{NAE} & \text{NAE} & \text{NAE} \\
\text{NAE} & \text{NAE} & \text{NAE} & \text{NAE} \\
\end{array}$

$\text{5} \ldots + \text{y}$

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

$\text{12 } \times 12$ matrix
### Approximate Dictionary?

<table>
<thead>
<tr>
<th>Tensor Networks in Physics</th>
<th>Graphical Models in Stats/ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPS</td>
<td>HMM</td>
</tr>
<tr>
<td>TTN</td>
<td>GMM</td>
</tr>
<tr>
<td>PEPS</td>
<td>CRF/MRF</td>
</tr>
<tr>
<td>MERA</td>
<td>?DBM?</td>
</tr>
<tr>
<td>DMRG</td>
<td>??</td>
</tr>
</tbody>
</table>

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 $A_1, \ldots, A_d$.

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

What are the polynomial relations that hold among the coefficients $\Psi_{i_1, \ldots, i_n} = \text{tr}(A_{i_1} \cdots A_{i_n})$?

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

\[ P_1 \times P_1 \times P_1 \times P_1 \hookrightarrow P^{15} \]

Segre variety defined by 2 \times 2 minors of flattenings of 2 \times 2 \times 2 \times 2 tensor

\[ \sigma_2(P_1 \times P_1 \times P_1 \times P_1) \]

First secant of Segre variety 3 \times 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.


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

Definition 3.14. Given a partition $\mu = (\mu_1, \cdots, \mu_t) \vdash r$, an $n$-partition $\lambda \vdash n$ and a block $M \in U_d^{\mu}$, we associate to the element $c_\lambda \cdot M \in c_\lambda \cdot U_d^{\mu}$ the $n$-tableau $T = (T_1, \cdots, T_n) = T_1 \otimes \cdots \otimes T_n$ of shape $\lambda$, obtained as follows. Suppose that the block $M$ has the set $\alpha_{i,j}$ in its $i$-th row and $j$-th column. Then we set equal to $i$ the entries in the boxes of $T_j$ indexed by elements of $\alpha_{i,j}$ (recall from Section 2.3 that the boxes of a tableau are indexed canonically: from left to right and top to bottom). Note that each tableau $T_j$ has entries $1, \cdots, t$, with $i$ appearing exactly $\mu_i \cdot d_j$ times. Note also that in order to construct the $n$-tableau $T$ we have made a choice of the ordering of the rows of $M$: interchanging rows $i$ and $i'$ when $\mu_i = \mu_{i'}$ should yield the same element $M \in U_d^{\mu}$, therefore we identify the corresponding $n$-tableaux that differ by interchanging the entries equal to $i$ and $i'$.

Example 3.15. We let $n = 2$, $d = (2, 1)$, $r = 4$, $\mu = (2, 2)$ as in Example 3.2, and consider the 2-partition $\lambda = (\lambda_1, \lambda_2)$, with $\lambda_1 = (5, 3)$, $\lambda_2 = (2, 1, 1)$. We have $c_\lambda \cdot 1_2 \otimes 6_1 \otimes 4_1 = 1_1 \otimes 2_1 \otimes 2_1 \otimes 4_2 + 1_2 \otimes 2_1 \otimes 2_2 \otimes 4_2$.

Let's write down the action of the map $\pi_\mu$ on the tableaux pictured above

$$\pi_\mu \left( \begin{array}{c|c|c|c} 1 & 2 & 2 & 3 \\ \hline 1 & 2 & 2 & 3 \\ \hline 1 & 4 & 4 & \end{array} \right) \otimes \begin{array}{c} 1 \\ \hline 3 \\ \hline 4 \\ \end{array} \right) = \begin{array}{c|c|c|c} 1 & 1 & 1 & 2 \\ \hline 1 & 2 & 2 & \end{array} \otimes \begin{array}{c} 1 \\ \hline 2 \\ \hline 1 \\ \end{array} + \begin{array}{c|c|c|c} 1 & 2 & 2 & 1 \\ \hline 1 & 2 & 2 & \end{array} \otimes \begin{array}{c} 1 \\ \hline 2 \\ \hline 1 \\ \end{array} + \begin{array}{c|c|c|c} 1 & 2 & 2 & 2 \\ \hline 1 & 1 & 1 & \end{array} \otimes \begin{array}{c} 1 \\ \hline 2 \\ \end{array}. $$
Representation theory

- Which tensor products $\mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$ have finitely many orbits under $\text{GL}(d_1, \mathbb{C}) \times \cdots \times \text{GL}(d_n, \mathbb{C})$?
- 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

<table>
<thead>
<tr>
<th>Case $(2, m, n)$</th>
<th>The number of orbits of $\text{GL}_2 \times \text{GL}_m \times \text{GL}_n$</th>
<th>deg $f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(2, 2, 2)$</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>$(2, 2, 3)$</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>$(2, 2, 4)$</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>$(2, 2, n), \ n \geq 5$</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>$(2, 3, 3)$</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>$(2, 3, 4)$</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>$(2, 3, 5)$</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>$(2, 3, 6)$</td>
<td>27</td>
<td>6</td>
</tr>
<tr>
<td>$(2, 3, n), \ n \geq 7$</td>
<td>27</td>
<td>0</td>
</tr>
</tbody>
</table>

The main results of the present paper were published (without proofs) in [6]. We use this opportunity to point out that [6] contains two disappointing mistakes, one of which is a consequence of the other. Namely:

1. In Theorem 2, the line $(2, 2, n), \ n \geq 4$, has ten orbits with representatives 1–9, 19, must be replaced by the line $(2, 2, n), \ n \geq 4$, has ten orbits with representatives 1–7, 11, 13, 19;

2. Accordingly, the figure with the abutment graph should contain no arrow from vertex 19 to vertex 9, but there should be an arrow from the vertex 19 to vertex 13 instead.

I would like to express my deep gratitude to my research supervisor, E.B. Vinberg, for setting the problem, crucial advice, and constant attention to this research.
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

- 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

\[ p(v, h) = \frac{1}{Z} \psi(v, h), \quad Z = \sum_{v, h} \psi(v, h). \]
Restricted Boltzmann machines

Unnormalized fully-observed potential is

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

The probability distribution on the visible random variables is

$$p(v) = \frac{1}{Z} \cdot \sum_{h \in \{0,1\}^k} \psi(v, h), \quad Z = \sum_{v,h} \psi(v, h).$$
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 $\Delta_{2^n-1}$. 
Hadamard Products of Varieties

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

$$X \times Y \longrightarrow \mathbb{P}^m, \ (x, y) \mapsto (x_0y_0 : x_1y_1 : \ldots : x_my_m).$$

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

If $M$ is a subset of the simplex $\Delta_{m-1}$ 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

$$V_n^k = (V_n^1)^{[k]} \quad \text{and} \quad M_n^k = (M_n^1)^{[k]}.$$
RBM as Hadamard product of naïve Bayes
Representational power of RBMs

Conjecture

The restricted Boltzmann machine has the expected dimension: $M^k_n$ 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 $k \geq 2^n - \lceil \log_2(n+1) \rceil$.

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
A multilinear operator
\[ T : U \otimes V \rightarrow W \]
is a tensor

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

\[ M : (A^* \otimes B) \times (B^* \otimes C) \rightarrow A^* \otimes C \]
gives the exponent of matrix multiplication.
Satisfiability and \#CSP problems

Given a problem $P$ in conjunctive normal form:
- a collection of Boolean variables $x_1 \ldots x_m$
- subject to clauses $c_1 \ldots c_p$ (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

\[ \text{# of satisfying assignments} = \langle \text{all possible assignments}, \text{all restrictions} \rangle = \alpha \beta \sqrt{\text{det}(x + y)} \]

4096-dimensional space \((\mathbb{C}^2)^{\otimes 12}\) 12 × 12 matrix
Efficient contraction with Pfaffian circuits

\[ A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \]

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

\[ 2^5 \cdot \left( \frac{6}{2^3} \right)^4 \cdot \text{Pfaff}(\tilde{z} + y) = 14 \text{ satisfying assignments}. \]