Geometry, Invariants, and the Elusive Search for Complexity Lower Bounds

Peter Bürgisser
Technische Universität Berlin
Germany

Simons Institute for the Theory of Computing, Berkeley
Program on "Algorithms and Complexity in Algebraic Geometry"

Open Lecture, September 8, 2014

Motivation

Three multilinear polynomials . . .

The following three polynomials in the variables X_i, X_{ij} over a field F are each given by a sum with exponentially many summands in n:

$$\mathrm{esym}_k := \sum_{i_1<i_2<\cdots<i_k} X_{i_1} X_{i_2} \cdots X_{i_k} \quad (1)$$

$$\det_n := \sum_{\pi \in S_n} \mathrm{sgn}(\pi)\, X_{1\pi(1)} \cdots X_{n\pi(n)} \quad (2)$$

$$\mathrm{per}_n := \sum_{\pi \in S_n} X_{1\pi(1)} \cdots X_{n\pi(n)} \quad (3)$$

We want to compute them from the variables and field elements with as few arithmetic operations $+, -, *$ (possibly also $/$)!

... and their efficient computation

(1): Note that

$$F(T,X) := (T+X_1)\cdots(T+X_n) = \sum_{k=0}^{n} \mathrm{esym}_k(X)\, T^{n-k}$$

can be computed with $O(n)$ operations.
Evaluate $F(t_0,X),\ldots,F(t_n,X)$ for different values t_i and compute $\mathrm{esym}_k(X)$ by interpolation. Total of $O(n^3+n^2)$ operations.

(2): The determinant $\det_n$ can be computed with $O(n^3)$ operations using Gaussian elimination.

(3): A trick due to Ryser gives a computation of the permanent per_n with $O(n\,2^n)$ operations.

Optimality: elementary symmetric polynomials

- The complexity $L(f)$ denotes the minimal number of arithmetic operations sufficient to compute f (from variables and field elements).

- (1): By divide and conquer and FFT: $L(\mathrm{esym}_k) = O(n \log^2 n)$.

- This is essentially optimal: we know the lower bound

$$L(\mathrm{esym}_k) = \Omega(n \log n)$$

  (Strassen '73, Baur and Strassen '83).

  The argument is based on algebraic geometry (degree of varieties).

Optimality: determinant

- (2): $L(\det_n) = O(n^{2.81})$: "Gaussian elimination is not optimal", Strassen 1969.
- $\det_n$ has the same "asymptotic complexity" as $n \times n$ matrix multiplication.
- It is known that
$$L(\det_n) = O(n^{\omega}),$$
where the exponent ω of matrix multiplication is known to satisfy
$$2 \leq \omega < 2.373$$
(Coppersmith & Winograd 1987, Vassilevska-Williams 2011).
- It is a fundamental problem to determine ω. The experts on this are involved in this Simons program.
- It is conjectured that ω can be chosen arbitrarily close to 2.

The permanent

- The importance of the permanent is due to a universality property explained later.
- We don't know of any computation of the permanent per_n that takes a number of arithmetic operations subexponential in n.
- Les Valiant conjectured in 1979 that $L(\mathrm{per}_n)$ grows superpolynomial in n.
- But as of today, we cannot even prove a superlinear lower bound on $L(\mathrm{per}_n)$!

  I sincerely hope that this Simons program will help to improve the state of affairs!

Specific polynomials which are hard to compute

Lower bounds for specific polynomials

- Dimension counting argument (à la Shannon):
  For almost all coefficient systems $a = (a_\pi) \in \mathbb{C}^{n!}$,

$$f = \sum_{\pi \in S_n} a_\pi \, X_{1\pi(1)} \cdots X_{n\pi(n)}$$

  has complexity at least $n! = \#\text{coefficients}$.
- Can this bound be extended to specific choices of a_π?

Strassen 1974

Assume the coefficient vector a equals

$$(\sqrt{1}, \sqrt{2}, \dots, \sqrt{n!}).$$

Then $L(f) = \Omega(\frac{n!}{\log n!})$.

Basic idea of proof

- We identify a polynomial $f \in \mathbb{C}[X]$ of degree $\leq n$ in $m = n^2$ variables with its coefficient sequence, interpreted as a point in $\mathbb{C}^N$, where $N = \binom{m+n}{n}$.
- Observation: the set of polynomials f with $L(f) < r$ equals the image of an explicit polynomial "computation map"

$$\Phi \colon \mathbb{C}^q \to \mathbb{C}^N,$$

  with $q := r^2 + 2mr$ "degrees of freedom".
- Reason: in all possible computations combine the linear operations and only count the multiplication steps. They have the form

$$g_{k+1} \ := \ (\sum_{i=-m}^{k} a_i g_i) * (\sum_{j=-m}^{k} b_j g_j), \quad a_i, b_i \in \mathbb{C},$$

  where $g_{-m}, \dots, g_k$ are the previously computed intermediate results, assuming $(g_{-m}, \dots, g_0) = (1, X_1, \dots, X_m)$.

Connection to algebraic geometry

- The (Zariski) closure of the image of $\Phi\colon \mathbb{C}^q \to \mathbb{C}^N$ is an affine algebraic variety $X_{n,r} \subseteq \mathbb{C}^N$ with $\dim X_{n,r} \leq q$.
- So $X_{n,r}$ consist of all polynomials f of complexity$< r$ and their limits.

Basic strategy

Look for a nonzero polynomial function $R\colon \mathbb{C}^N \to \mathbb{C}$ that vanishes on $X_{n,r}$. We shall call such R a "resultant".

$$R(f) \neq 0 \text{ implies that } f \notin X_{n,r}, \text{ hence } L(f) \geq r.$$

Existence of resultants

Basic strategy

Look for a nonzero polynomial function $R\colon \mathbb{C}^N \to \mathbb{C}$ that vanishes on $X_{n,r}$. We shall call such R a "resultant".

$$R(f) \neq 0 \text{ implies that } f \notin X_{n,r}, \text{ hence } L(f) \geq r.$$

- The components of $\Phi\colon \mathbb{C}^q \to \mathbb{C}^N$ are integer polynomials of degree $\leq rn$ (and bitsize $\leq 2^r \log(mr)$).
- From this one can deduce the existence of a resultant R of degree $\leq (rn)^{r^2}$ (with integer coefficients of absolute value ≤ 3).
- This information is sufficient to prove that $R(f) \neq 0$ for the specific f, since the degree of the field extension $\mathbb{Q}(\sqrt{2}, \ldots, \sqrt{n!})$ over $\mathbb{Q}$ is exponential in $n!$.

Lower bounds for p-definable polynomials?

- In the previous example, the coefficients of f were algebraic numbers, producing a field extension of high degree.
- The challenge is to prove lower bounds for specific polynomials f with integer coefficients.
- We call a family (f_n) of multivariate polynomials p-definable if the coefficient function $\pi \mapsto a_\pi$ can be computed in polynomial time.

Valiant 1979

A superpolynomial bound for any family of p-definable polynomials implies a superpolynomial lower bound for the permanents.

This is a consequence of the VNP-completeness of (per_n).

A curious observation

$$p_n(X) := \prod_{j=1}^{n}(X^2 - j) = \underbrace{\prod_{j=1}^{n}(X - \sqrt{j})}_{f_n(X)} \cdot \underbrace{\prod_{j=1}^{n}(X + \sqrt{j})}_{\tilde{f}_n(X)}.$$

- Both $f_n(X)$ and $\tilde{f}_n(X)$ have complexity at least $\Omega(n/\log n)$: proof with the same techniques as before.
- It seems plausible that the product $f_n(X) \cdot \tilde{f}_n(X)$ is hard as well!
- However, proving this turns out to be hard!

B 2009

A lower bound of the form n^ϵ on $p_n(X)$, for any $\epsilon > 0$, implies superpolynomial lower bounds for the complexity of the permanent.

Permanent versus determinant

Arithmetic complexity classes

- Valiant defined complexity classes VP und VNP, whose objects are sequences (f_n) of multivariate polynomials over some fixed field F.
- VP: "Problems" of linear algebra
- VNP: "Problems" from graph theory, combinatorics, statistical physics, quantum mechanics
- A notion of reduction allows to talk about complete (or universal) objects in these classes.

Completeness of det and per

Valiant 1979-81

(det_n) is complete for VP.
(per_n) is complete for VNP if $\mathrm{char} F \neq 2$.

- Valiant's Hypothesis

$$\mathrm{VP} \neq \mathrm{VNP}$$

  can be seen as an arithmetic version of $\mathrm{P} \neq \mathrm{NP}$.
- It means that the complexity of per_n grows superpolynomial in n.
- $\mathrm{P} \neq \mathrm{NP}$ implies $\mathrm{VP} \neq \mathrm{VNP}$ over $\mathbb{C}$.
- Conclusion: The arithmetic version $\mathrm{VP} \neq \mathrm{VNP}$ has to be proven first. It is close to algebra and geometry and appears more amenable to the known mathematical techniques.

An algorithm-free characterization of $\mathrm{VP} \neq \mathrm{VNP}$

Non-obvious fact: Let $n = 2^m$. There affine linear function $a_{ij} = a_{ij}(X)$ in $X_{11}, \ldots, X_{mm}$ such that

$$\mathrm{per}_m(X) = \det \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix} \qquad (*)$$

Can the size n of the determinant be taken substantially smaller?

$\mathrm{VP} \neq \mathrm{VNP}$ is equivalent to the statement that in (*), the size n of the determinant has to grow faster than any polynomial in m.

Unfortunately, the best known lower bound on n only states $n \geq \frac{1}{2} m^2$ (Mignon & Ressayre 2004).

Orbit closure problems

Refining the basic strategy

- Goal: attack the algorithm-free characterization of $\mathrm{VP} \neq \mathrm{VNP}$ by refining the previous proof of the lower bound for specific polynomials.
- Recall basic strategy: $X_{n,r}$ denoted the closure of the set of easy polynomials (n variables, complexity$< r$).
  We look for a "resultant", i.e., nonzero polynomial function $R\colon \mathbb{C}^N \to \mathbb{C}$ vanishing on $X_{n,r}$. Note $R(f) \neq 0 \Rightarrow f \notin X_{n,r}$.
- We need to have more information on the resultants R! Previously, we only used their existence in certain degrees.
- We shall replace X_{nr} by an algebraic variety $\mathcal{D}et_n$, having lots of symmetries.
- These symmetries allow us to restrict our search to resultants having certain invariant properties, called "highest weight vectors" in represention theory.

The orbit of the determinant

- In mathematics, symmetries are described by groups.
- The determinant has lot of symmetries, coming from $\det(A \cdot B) = \det(A) \cdot \det(B)$.
- $\mathrm{Poly}_n(\mathbb{C}^{n^2})$ denotes the vector space of homogeneous polynomials of degree n in n^2 variables. So $\det_n \in \mathrm{Poly}_n(\mathbb{C}^{n^2})$.
- The group $G := \mathrm{GL}_{n^2}$ acts on $\mathrm{Poly}_n(\mathbb{C}^{n^2})$ by variable substitution.
- The orbit $G \det_n$ of $\det_n$ is defined as the set of polynomials that can be obtained from $\det_n$ by applying all possible group elements: "determinants in disguise".
- One can efficiently decide whether $f \in G \det_n$ (Kayal '11).

The orbit closure $\mathcal{D}et_n$

- $\mathcal{D}et_n$ is defined as the closure of the orbit $G\,\det_n$: we add all "limit polynomials".
- The previous relation (*) can be rewritten as

$$Z^{n-m}\mathrm{per}_m(X) = \det \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \qquad (*')$$

where $a_{ij} = a_{ij}(X, Z)$ are now linear in X_{ij}, $1 \leq i, j \leq m$, and a homogenizing variable Z.

- Observation: If (*') holds, then $Z^{n-m}\mathrm{per}_m(X)$ is in $\mathcal{D}et_n$.

Mulmuley & Sohoni 2001

We should prove that $Z^{n-m}\mathrm{per}_m(X) \notin \mathcal{D}et_n$ for $n \leq m^{\mathcal{O}(1)}$.

The orbit closure problem of deciding $f \in \mathcal{D}et_n$ is much more difficult than the orbit problem $f \in G\,\det_n$: geometric invariant theory.

Decomposition of function spaces via symmetries

Representations

- Resultants are polynomial functions

$$R\colon \mathrm{Poly}_n(\mathbb{C}^{n^2}) \to \mathbb{C}$$

  that vanish on $\mathcal{D}et_n$. They form the vanishing ideal $I(\mathcal{D}et_n)$ of $\mathcal{D}et_n$.
- Let $\mathrm{Poly}(\mathrm{Poly}_n(\mathbb{C}^{n^2}))$ denote the vector space of polynomial functions on $\mathrm{Poly}_n(\mathbb{C}^{n^2})$.
- We have an induced linear action of G on $\mathrm{Poly}(\mathrm{Poly}_n(\mathbb{C}^{n^2}))$ that preserves the vanishing ideal $I(\mathcal{D}et_n)$.
- Representation theory is the study of linear actions of groups on vector spaces. It is also of great relevance in quantum mechanics.
- Each representation splits into a direct sum of irreducible subrepresentations (simultaneous block decomposition for all $g \in G$).
- The isomorphy types V_λ of irreducible representations of $G = \mathrm{GL}_{n^2}$ are labeled by integer vectors $\lambda \in \mathbb{Z}^{n^2}$, where $\lambda_1 \geq \cdots \geq \lambda_{n^2}$.

Plethysms

- Our search for resultants is based on the decomposition

$$\mathrm{Poly}_d(\mathrm{Poly}_n(\mathbb{C}^{n^2})) = \bigoplus_\lambda \mathrm{pleth}_\lambda V_\lambda$$

  into irreducible G-invariant linear subspaces V_λ. The plethysm coefficient $\mathrm{pleth}_\lambda \in \mathbb{N}$ is the multiplicity of V_λ.
- The discrete labels λ are partitions $\lambda_1 \geq \cdots \geq \lambda_{n^2}$ such that $\sum_i \lambda_i = dn$, $\lambda_i \in \mathbb{N}$.
- In the special case $\mathrm{Poly}_d(\mathrm{Poly}_n(\mathbb{C}^2))$, the decomposition describes invariants and covariants of degree n binary forms. Intense study in 19th century: (Cayley, Sylvester, Clebsch, Gordan, Hilbert, ...).
- Plethysm coefficients are not well understood.

Kronecker coefficients

- In the analysis of tensors (or trilinear forms), the following decomposition into irreducible $\mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n$-invariant subspaces is crucial:

$$\mathrm{Poly}_d(\mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n) = \bigoplus_{\lambda,\mu,\nu} \mathrm{kron}(\lambda, \mu, \nu)\, V_\lambda \otimes V_\mu \otimes V_\nu.$$

- The multiplicities $\mathrm{kron}(\lambda, \mu, \nu)$ are called Kronecker coefficients.

Kronecker coefficients prominently show up in the resultant based analysis of $\mathrm{VP} \neq \mathrm{VNP}$ as well as in the analysis of the tensor rank problem (complexity of matrix multiplication).

No combinatorial description of Kronecker coefficients is known!

Resultants in tensor setting

- Specific resultants already have been successfully used for lower bounds on tensor rank.
- In a pioneering work, Strassen (1983) found a resultant (invariant) $\mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^3 \to \mathbb{C}$ of type $\lambda = (3 \times n, 3 \times n, n \times 3)$ vanishing on tensors of border rank$\leq 3n/2$.
- Bläser's lower bound for the rank of matrix multiplication (1999) is based on Strassen's resultant.
- Landsberg and Ottaviani recently improved Bläser's bound by extending Strassen's construction (based on representation theory).
- Ikenmeyer, Hauenstein, Landsberg (2013): Resultant based proof that border rank of 2×2 matrix multiplication equals 7. (Using a highest weight vector of degree 20.)

On the vanishing ideal of $\mathcal{D}et_n$

- The vanishing ideal $I(\mathcal{D}et_n)$ consists of the resultants.
- Decompositions into G-invariant linear subspaces:

$$\begin{aligned} \mathrm{Poly}_d(\mathrm{Poly}_n(\mathbb{C}^{n^2})) &= \bigoplus_\lambda \mathrm{pleth}_\lambda V_\lambda \\ I(\mathcal{D}et_n)_d &= \bigoplus_\lambda \mathrm{multdet}_\lambda V_\lambda. \end{aligned}$$

- There are $\mathrm{multdet}_\lambda$ many linearly independent resultants of type λ.

Mulmuley-Sohoni '08, B-Landsberg-Manivel-Weyman '11

$$\mathrm{pleth}_\lambda - \mathrm{kron}_\lambda \leq \mathrm{multdet}_\lambda \leq \mathrm{pleth}_\lambda$$

where $\mathrm{kron}_\lambda := \mathrm{kron}(\lambda, n \times d, n \times d)$ denotes the Kronecker coefficient of λ and twice the rectangular partition $n \times d := (d, \ldots, d)$.

A "small" example: $n = 3$

- Extensive computer computations by C. Ikenmeyer. Let $n = 3$.
- $\mathrm{Poly}_3(\mathbb{C}^9) = \{\text{cubic forms in 9 variables}\} \simeq \mathbb{C}^{165}$
- For degree $d = 12$ there are many λ with $\mathrm{kron}_\lambda < \mathrm{pleth}_\lambda$. The one of shortest length $\ell(\lambda)$ is

$$\lambda = (13, 13, 2, 2, 2, 2, 2) \vdash 36, \quad \ell(\lambda) = 7.$$

- Here: $\mathrm{pleth}_\lambda = 1$ and $\mathrm{kron}_\lambda = 0$. Therefore $\mathrm{multdet}_\lambda = 1$.
- Hence there is, up to scaling, a unique homogenous polynomial $R\colon \mathrm{Poly}_3(\mathbb{C}^9) \to \mathbb{C}$ of degree 12 of type λ.
  R is a resultant: it vanishes on $\mathcal{D}et_3$.
- Note: R was found as an element of $\mathrm{Poly}_{12}(\mathbb{C}^{165})$, which has dimension $\approx 1.3 \cdot 10^{19}$.

Occurrence obstructions

Occurrence obstructions (Mulmuley & Sohoni, 2001)

- A candidate for occurrence obstructions for $\det_n$ is a type λ such that $\mathrm{multdet}_\lambda = \mathrm{pleth}_\lambda$. This means that **all** polynomials R of type λ vanish on $\mathcal{D}et_n$.
- λ is an occurrence obstruction to $Z^{n-m}\mathrm{per}_m$ in $\det_n$ if, additionally, $R(Z^{n-m}\mathrm{per}_m) \neq 0$ for some candidate R.
- If there is an occurrence obstruction, then $Z^{n-m}\mathrm{per}_m \notin \mathcal{D}et_n$.
- By the previous insight:

$$\mathrm{kron}_\lambda = 0 \Longrightarrow \lambda \text{ is candidate for occurrence obstructions.}$$

- The converse is false.

State of the art regarding occurrence obstructions

- So far, we don't have any examples of occurrence obstructions in the determinant setting!!! ☹☹☹
- Due to huge dimensions, experiments are extremely hard to perform!

- However, good news in the tensor setting $\mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n$. ☺
- B & Ikenmeyer ('13) found an explicit family of occurrence obstructions in this setting. We used this to prove (modest) lower bounds on the border rank of the matrix multiplication tensor.

Occurrence obstructions are hard to describe ☹

- Consider the set

$$S(\mathcal{D}et_n) := \{\lambda \mid \mathrm{multdet}_\lambda < \mathrm{pleth}_\lambda\}$$

of types that are **not** candidates for occurrence obstructions.

- General principles: $S(\mathcal{D}et_n)$ is a finitely generated monoid w.r.t. addition.
- The saturation of $S(\mathcal{D}et_n)$ consists of all λ such that $k\lambda \in S(\mathcal{D}et_n)$ for some $k \in \mathbb{N}_{>0}$.

B, Christandl, Ikenmeyer '11, Kumar '12

The saturation of the monoid $S(\mathcal{D}et_n)$ contains all types λ of length $\leq n$.

- This proves that only the "holes" $\lambda \in S(\mathrm{det}_n)_{sat} \setminus S(\mathrm{det}_n)$ can be occurrence obstructions! Those are hard to analyze.

Ongoing work: search for vanishing Kronecker coefficients

B-Ikenmeyer 2013 found an explicit combinatorial counting function t such that

- $\mathrm{kron}(\lambda, \mu, \nu) \leq t(\lambda, \mu, \nu)$,
- testing $t(\lambda, \mu, \nu) > 0$ is NP-complete.

- This makes it unlikely, that $\mathrm{kron}(\lambda, \mu, \nu) > 0$ can be tested in polynomial time!
- While this sounds like bad news, Ketan Mulmuley pointed out the following positive consequence:

- There are superpolynomially many (λ, μ, ν) of length$\leq n$ with $\mathrm{kron}(\lambda, \mu, \nu) = 0$ (and they can be explicitely constructed).
- Unfortunately, it turns out that always $t(\lambda, \mu, \nu) > 0$ if $\mu = \nu = n \times d$ are rectangular partitions (Ikenmeyer).
- So this argument breakes down in the case of interest!
- Hopefully, a refinement of the upper bound function t can lead to success!

Thank you!