The Geometric Complexity Theory Approach to Algebraic Complexity

Mark Bun

June 1, 2012

Contents

1 Introduction

In [9] and a sequence of seven subsequent papers, K. Mulmuley and M. Sohoni advance a unified approach to separating complexity classes called the *Geometric Complexity Theory* (GCT). The approach relates complexity classes to projective orbit closures in certain spaces of polynomials. It effectively recasts complexity theoretic conjectures as algebro-geometric and representation-theoretic questions.

The primary goal of this paper is to give a complete and relatively self-contained account of the $\overline{\mathbf{VP}}_{\mathrm{WS}}$ vs. **VNP** conjecture and the GCT approach to its resolution. This conjecture is an analog of **P** vs. **NP** for arithmetic circuits, and is fundamentally a question of whether the matrix permanent is inherently harder to compute than the determinant. The main ideas are from the preprint [5] of Bürgisser et al. I wish to acknowledge my advisor Jim Morrow for the impact of our discussions – and his healthy skepticism – on both the mathematics and the exposition of this paper.

1.1 Notation

Let k be a field. Write $k[x]$ and $k[[x]]$ for the rings of polynomials and formal power series, respectively, in k over the variable x. Similarly, let $k(x)$ and $k((x))$ be the fields of rational functions and formal Laurent series over x.

If V is a vector space over k, let $\mathbb{P}V$ be the space of lines through the origin in V. For $x \in V$, we denote by $[x]$ the corresponding line in $\mathbb{P}V$.

Let S^nV denote the vector space of homogeneous polynomials of degree n over V^*. That is, S^nV consists of combinations of the form

$$\sum_{j=1}^{N} a_j \ell_j^n,$$

where $N \in \mathbb{N}, a_j \in k$ and $\ell_j \in V^*$. We call an element of this space an *n-form* over V. If $V = k^m$, we can identify the elements of the dual basis on $(k^m)^*$ with the indeterminates $x_1, \ldots, x_m$ to get a subspace of $k[x_1, \ldots, x_m]$.

Example 1. The polynomials $x_1^4 + ix_1^2x_2^2$ and $x_1^2x_2x_3 - x_2x_3^3$ are elements of $S^4\mathbb{C}^3$. This illustrates the natural inclusion $S^nk^m \subset S^nk^{m'}$ for $m' > m$. We will primarily be concerned with the permanent and determinant, per_n and $\det_n$, which are contained in $S^n\mathbb{C}^{n^2}$.

The general linear group $GL(V)$ of automorphisms of V has a natural left action on S^nV. For $\sigma \in GL(V)$, $f \in S^nV$ and $x \in V$, define

$$(\sigma \cdot f)(x) = f(\sigma^{-1}x).$$

This action extends to $\mathbb{P}(S^nV)$ by taking $\sigma \cdot [f] = [\sigma \cdot f]$. If G is a subgroup of $GL(V)$, we write $G \cdot f$ for the orbit of f under G, and $\overline{G \cdot f}$ for its Zariski closure. When the field is understood, we generally write GL_n for $GL(k^n)$.

Example 2. Consider $\det_2 \in S^2\mathbb{C}^4$ as the determinant

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

If we define $\sigma \in GL_4$ by $\sigma(x) = x/2$, then $\sigma \cdot \det_2(a, b, c, d) = 4(ad - bc)$.

1.2 Topological Facts

If our goal is to convert complexity theoretic conjectures into robust algebro-geometric ones, we should aim to make statements in terms of the Zariski topology. However, much of our analysis of sequences of polynomials over $\mathbb{C}$ is simplified if we instead work in the Euclidean topology. The following results allow us to switch between the two rather seamlessly.

Definition 1. Let X be a topological space. A subset $S \subset X$ is *constructible* if it is a finite union of locally closed sets. That is, we can write S as

$$S = \bigcup_{i=1}^{n} U_i \cap F_i$$

where the U_i's are open and the F_i's are closed.

Theorem 1 (Chevalley)**.** *Let X and Y be algebraic varieties, and let $f : X \to Y$ be a morphism. Then the image of f is a Zariski constructible subset of Y.*

The key theorem is the following result from [10]. It generalizes to algebraically closed topological fields, but the statement over $\mathbb{C}$ will suffice for our purposes.

Theorem 2. *Let X be a variety over $\mathbb{C}$, and let S be a Zariski constructible subset. Then S has the same closure in the Zariski topology as it does in the Euclidean topology.*

Some of our results are easiest to state with respect to a topology on a polynomial ring $A_n = k[x_1, \ldots, x_n]$. Each subset $A_n^d = \{f \in k[x_1, \ldots, x_n] : \deg f \leq d\}$ is a finite dimensional vector space over k, with a basis consisting of the monomials

$$\bigcup_{0 \leq d' \leq d} \{x_1^{d_1} x_2^{d_2} \ldots x_n^{d_n} : d_1 + \cdots + d_n = d'\},$$

and can thus be given the Zariski topology. The obvious inclusions $A_n^d \hookrightarrow A_n^{d'}$ for $d \leq d'$ make $\{A_n^d : d \in \mathbb{N}\}$ into a directed system, whose direct limit induces a topology on A_n. If $k = \mathbb{C}$, we can apply the same procedure using the Euclidean topology.

2 Algebraic Complexity

In this section, we describe Valiant's model of algebraic complexity. This will give a the framework for discussing Valiant's hypothesis and the main separation problems considered by the GCT program.

2.1 Arithmetic Circuits

Our first definition will be that of an *arithmetic circuit.* We work over a fixed field k, which we will usually take to be $\mathbb{C}$. An arithmetic circuit's mission in life is to compute a polynomial in $k[x_1, \dots, x_n]$. It consists of a finite acyclic directed graph whose vertices have in-degree 0 or 2. If $u \to v$ is an edge in the graph, we call u a parent of v and v a child of u. The leaves (vertices with in-degree 0) are called *inputs* and are labeled by either a variable x_i or by a constant from k. We define the *value* at such a vertex to be its label. The internal vertices, called *computation gates*, are labeled by $\times$ or $+$. The value at a gate is defined recursively as the polynomial obtained by applying its operation to the values of its parents. An arithmetic circuit must have exactly one vertex of out-degree 0, which we call its *output.* The polynomial computed by a circuit is just its value at this vertex.

Example 3. An arithmetic circuit over $k = \mathbb{F}_2$ is known as a Boolean circuit. This case has been extensively studied, but lower bounds for Boolean circuits are notoriously hard.

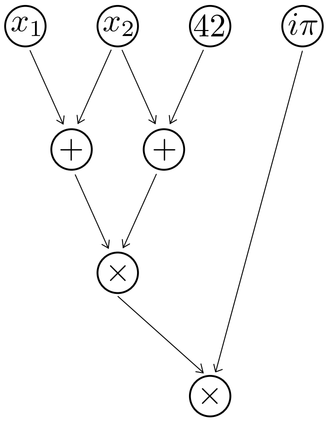
By taking a topological ordering of its underlying graph, we can equivalently view an arithmetic circuit as a *straight-line program.* Such a program is a finite sequence of instructions, each of which computes a sum or a product of inputs and results of prior instructions. The term "straight-line" refers to the absence of any branching or looping instructions in these programs.

Example 4. The circuit and straight-line program in Figure 1 compute the polynomial $i\pi(42 + x_2)(x_1 + x_2)$.

Recall that our objective is to study the difficulty of computing the determinant and permanent as the number of input variables increases. But one of the salient features of the arithmetic circuit is that it is only capable of computing a polynomial over a fixed number of variables. It makes no sense to think of a single circuit as computing "the determinant." So instead, we consider *families* of polynomials parametrized by input size, and compute them using a different circuit for each number of variables. This inherent need to partition according to input size is why circuit computation is sometimes referred to as *non-uniform computation.*

The *size* of a circuit is its number of vertices, and its *depth* is the length of its longest directed path. We define the *complexity* $L(f)$ of a polynomial $f \in k[x_1, \dots, x_n]$ to be the size of the smallest circuit computing f. The complexity class **VP** (we suppress the dependence on k) comprises the sequences of polynomials $\{f_n\}$ such that $L(f_n)$ and $\deg f_n$ are polynomially bounded in

Figure 1: Equivalence between circuits and straight-line programs

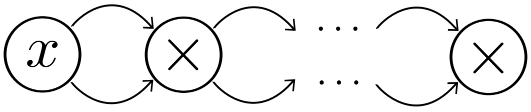


$$\begin{aligned}
\texttt{R}_{-4} &\leftarrow \texttt{x}_1\\
\texttt{R}_{-3} &\leftarrow \texttt{x}_2\\
\texttt{R}_{-2} &\leftarrow \texttt{42}\\
\texttt{R}_{-1} &\leftarrow \texttt{i}\pi\\
\texttt{R}_{0} &\leftarrow \texttt{R}_{-4} + \texttt{R}_{-3}\\
\texttt{R}_{1} &\leftarrow \texttt{R}_{-3} + \texttt{R}_{-2}\\
\texttt{R}_{2} &\leftarrow \texttt{R}_{0} \times \texttt{R}_{1}\\
\texttt{R}_{3} &\leftarrow \texttt{R}_{1} \times \texttt{R}_{2}
\end{aligned}$$

n. Equivalently, there exist polynomials $p(n)$ and $q(n)$ such that $\deg f_n \leq p(n)$ and for each n, there is a circuit of size at most $q(n)$ computing f_n.

Example 5. The degree condition in the definition of **VP** is not redundant. Consider the sequence of polynomials $f_n = x^{2^n}$. The circuit in Figure 2 of size $n+1$ computes f_n.

Figure 2: A circuit computing x^{2^n}



One can view **VP** as an algebraic analog of **P**, the complexity class of languages decidable in polynomial-time by a Turing machine. The other complexity class in Valiant's hypothesis is **VNP**, an analog of **NP**, the class of languages decidable in polynomial-time by a nondeterministic machine. A sequence $\{f_n\}$ is called *p-definable* if there exists a sequence $\{g_n\} \in \mathbf{VP}, g_n \in k[x_1, \ldots, x_{u(n)}]$ where

$$f_n(x_1, \ldots, x_{v(n)}) = \sum_{e \in \{0,1\}^{u(n)-v(n)}} g_n(x_1, \ldots, x_{u(n)}, e_1, \ldots, e_{u(n)-v(n)}).$$

In other words, f_n is a sum over (perhaps exponentially many) partial evaluations of g_n on 0s and 1s. The sum is analogous to the multiple transition rules of a nondeterministic Turing machine. The class **VNP** is the set of p-definable sequences. It is easy to see that $\mathbf{VP} \subset \mathbf{VNP}$. That this inclusion is strict is the content of Valiant's hypothesis, an analog of the **P** vs. **NP** question.

Conjecture 1 (Valiant's hypothesis)**.** $\mathbf{VP} \neq \mathbf{VNP}$*.*

There is strong evidence that Valiant's hypothesis is true. For instance, assuming the generalized Riemann hypothesis and that Valiant's hypothesis is false (over some field), the polynomial hierarchy collapses to the second level [2].

2.2 Reducibility and Completeness

To prove Valiant's hypothesis, it would suffice to find a sequence of polynomials in **VNP** that is not in **VP**. How would one go about looking for such a sequence? By analogy to **P** and **NP**, our starting point is to examine the computationally hardest sequences in **VNP**, i.e., those that are complete with respect to some notion of reducibility. We make this precise as follows.

Definition 2. A polynomial $f \in k[x_1, \ldots, x_n]$ is a *projection* of a polynomial $g \in k[x_1, \ldots, x_m]$ if there exist $a_1, \ldots, a_m \in k \cup \{x_1, \ldots, x_n\}$ such that

$$f(x_1, \ldots, x_n) = g(a_1, \ldots, a_m).$$

In other words, f is obtained from g by substituting its variables with variables or constants from k. If this is the case, we write $f \leq g$.

We say a sequence $\{f_n\}$ is a *p-projection* of $\{g_n\}$, written $\{f_n\} \leq_p \{g_n\}$, if there is a polynomially bounded function $t : \mathbb{N} \to \mathbb{N}$ such that $f_n \leq g_{t(n)}$.

In the framework of algebraic complexity, p-projection plays a role analogous to the Karp reduction.

Lemma 1. *The classes* **VP** *and* **VNP** *are closed under p-projection.*

Proof. This is easy, and we will just show it for **VP** to illustrate the idea. Suppose $\{g_n\} \in \mathbf{VP}$ with polynomials $p(n)$ and $q(n)$ such that $\deg g_n \leq p(n)$ and there is a circuit of size at most $q(n)$ computing g_n. Let $\{f_n\}$ be a p-projection of $\{g_n\}$ with $f_n \leq g_{t(n)}$. Then $\deg f_n \leq \deg g_{t(n)} \leq p(t(n))$ which is polynomially bounded. Take the circuit of size at most $q(t(n))$ that computes $g_{t(n)}$ and simply replace its input variables with the a_i's that make $f(x_1, \ldots, x_n) = g(a_1, \ldots, a_{t(n)})$. This gives us a circuit of polynomially bounded size that computes f_n. □

Definition 3. A p-definable sequence $\{f_n\}$ is **VNP**-*complete* if every p-definable sequence is a p-projection of $\{f_n\}$. More generally, if $\mathcal{C}$ is a complexity class of sequences of polynomials, $\{f_n\} \in \mathcal{C}$ is $\mathcal{C}$-*complete* if every sequence in $\mathcal{C}$ is a p-projection of $\{f_n\}$.

The motivation for considering complete problems is as follows. If there were a **VNP**-complete sequence $\{f_n\}$ with $\{f_n\} \in \mathbf{VP}$, then $\mathbf{VNP} \subset \mathbf{VP}$ by the closure of **VP** under p-projection. So assuming the existence of a **VNP**-complete sequence, the separation of **VP** and **VNP** is actually equivalent to exclusion of this sequence from **VP**.

Fortunately, such a sequence not only exists, but is a familiar object. Consider a matrix of variables $X = \{x_{i,j} : 1 \leq i, j \leq n\}$. Valiant showed that the sequence of matrix permanents, defined by

$$\mathrm{per}_n(X) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i,\sigma(i)}$$

is **VNP**-complete.

Lemma 2. *The sequence* per_n *is p-definable.*

Proof. This is Valiant's argument from [11]. For $n \times n$ matrices X and Y, let

$$g(X,Y) = \left(\prod_{i=1}^{n} \sum_{j=1}^{n} y_{ij}\right) \left(\prod_{i=k \text{ xor } j=m} (1 - y_{ij} y_{km})\right) \left(\prod_{i=1}^{n} \sum_{j=1}^{n} x_{ij} y_{ij}\right)$$

First note that g can be computed by a circuit with size $O(n^3)$. Suppose Y is a 0-1 matrix. Then the first factor is nonzero iff each row has at least one 1. Similarly, the second factor is nonzero iff each row and column has at most one 1. Thus $g(X,Y)$ is nonzero iff Y is a permutation matrix representing some permutation σ, in which case its value is $\prod_{i=1}^{n} x_{i,\sigma(i)}$. So

$$\mathrm{per}_n(X) = \sum_{e \in \{0,1\}^{n \times n}} g(X, e)$$

and per_n is p-definable. □

Lemma 3. *If the characteristic of* k *is not* 2*, then every p-definable sequence is a p-projection of* per_n.

We remark that in characteristic 2, the permanent is equal to the determinant, so this line of inquiry is uninteresting regardless.

Valiant's proof of Lemma 3 in [11] goes by way of an important graph interpretation of the permanent. A square matrix can be viewed as the adjacency matrix of a weighted complete directed graph. A *cycle cover* of this graph is a partition of its vertices into directed cycles. We define the weight of a cycle cover to be the product of the edges in it. Cycle covers are in correspondence with permutations of the vertices, so the permanent is just the sum of the weights of all cycle covers of this graph.

The idea of the argument is to consider an arbitrary sequence $\{f_n\}$ that is p-definable in terms of the sequence $\{g_n\}$. Then there is a graph G_n whose adjacency matrix has permanent g_n. We then form a new graph G'_n by adding cycles to G_n so that the weights of the cycle covers of G'_n correspond to evaluations of g_n. The permanent of G'_n then yields f_n. For the sake of completeness, we remark that this proof applies to **VNP** defined for formulas rather than arithmetic circuits, but Valiant showed that the two classes are actually equivalent.

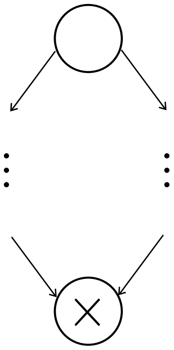
For an algebraic proof of this result, see [4, Thm. 21.29].

2.3 Weakly Skew Circuits

To state the variant of Valiant's conjecture that the GCT program aims to resolve, we have to introduce a few more complexity classes. An arithmetic circuit is said to be *weakly skew* if for every multiplication gate α, one of α's parents is computed by a *separate subcircuit.* A separate subcircuit is one that is not connected to the rest of the circuit by any other edges. In other words, we require that removing one of the edges pointing to α disconnects the graph.

For $f \in k[x_1, \dots, k_n]$, let $L_{\text{ws}}(f)$ be the size of the smallest weakly skew circuit computing f. A sequence $\{f_n\}$ is in the complexity class $\mathbf{VP}_{\text{ws}}$ if $L_{\text{ws}}(f_n)$ is polynomially bounded.

Figure 3: A situation disallowed in a weakly skew circuit



Lemma 4. *A weakly skew circuit with m variable inputs computes a polynomial with degree at most m.*

Proof. We induct on m. The case $m = 1$ is handled by induction on the multiplication gates. Let α be a multiplication gate, and suppose the degree at both of its parents is at most 1. Let β be its parent that is computed by a separate subcircuit. If the degree at β is 1, then α's other parent cannot depend on the variable input, and thus has degree zero. So the degree at α is also at most 1.

For $m > 1$, consider a multiplication gate α with parents β and γ. By weak skewness, β and γ depend on disjoint subsets of the variable inputs (see Figure 3). If one of these subsets is empty, the degree at its corresponding gate is zero, and we recurse via the other parent until we get both subsets to be nonempty. The degree at α is at most the sum of the degrees at β and γ, which is by induction at most m. □

Thus we see that weak skewness eliminates the pathology of Example 5.

Corollary 1. $\mathbf{VP}_{\mathrm{ws}} \subset \mathbf{VP}$.

The class $\mathbf{VP}_{\mathrm{ws}}$ is of interest because of its connection to the determinant. Just as we saw that the permanent is **VNP**-complete, the determinant

$$\det_n(X) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} x_{i,\sigma(i)}$$

turns out to be $\mathbf{VP}_{\mathrm{ws}}$-complete. It is not known whether the determinant is also **VP**-complete. The following two lemmas give the specifics.

Lemma 5. $L_{\mathrm{ws}}(\det_n) = O(n^5)$.

This is the best-known upper bound, and the algorithm is due to Berkowitz [1].

Lemma 6. *If* $L_{\mathrm{ws}}(f) \leq m$*, then* f *is a projection of* $\det_{m+1}$.

In his original paper, Valiant used the same idea as in Lemma 3 to prove a version of this result for formulas. This result from [8] uses a similar construction.

2.4 Approximate Complexity

One can view the GCT program as attack on the relaxation of Valiant's hypothesis to the question $\mathbf{VP}_{\mathrm{ws}} \neq \mathbf{VNP}$. However, if our goal is to study GL_{n^2}-orbit closures, it is actually more natural to examine *approximate complexity.* Our source is the framework developed in [3]. We will state everything in terms of the standard circuit complexity L, but everything works equally well with the complexity L_{ws} treated above.

Let $K = k(\epsilon)$ denote the field of rational functions in the variable ϵ over k, and for $y \in k$, let R_y be the subring of functions defined at $\epsilon = y$. To ease notation, let $R = R_0$. For $F \in R_y[x_1, \ldots, x_n]$, let $F_{\epsilon=y}$ be the image of F under the morphism induced by $\epsilon \mapsto y$.

Definition 4. Let $f \in k[x_1, \ldots, x_n]$. The approximate complexity $\underline{L}(f)$ is the smallest natural number r such that there exists $F \in R[x_1, \ldots, x_n]$ with $F_{\epsilon=0} = f$ and $L(F) \leq r$. The complexity $L(F)$ is defined with respect to constants taken from K.

We define the complexity class $\overline{\mathbf{VP}}$ (resp. $\overline{\mathbf{VP}_{\mathrm{ws}}}$) to be the set of sequences $\{f_n\}$ with $\underline{L}(f_n)$ (resp. $\underline{L}_{\mathrm{ws}}(f_n)$) polynomially bounded in n.

Bürgisser [3] explains the motivation for this definition as follows. Let $k = \mathbb{R}$ and suppose $\underline{L}(f) \leq r$. Then there exists a polynomial $F(x; \epsilon) \in R[x]$ such that $F(x; \epsilon) = f(x) + \epsilon\rho(x; \epsilon)$ where $L(F) \leq r$. Let M be the supremum of ρ over $\epsilon \in [0, 1]$ and $\|x\|_\infty \leq 1$. Then for all but finitely many ϵ near zero, we can use this small circuit to compute $F(x; \epsilon)$. Moreover, $|F(x; \epsilon) - f(x)| \leq M\epsilon$, so this computation has absolute error bounded by $M\epsilon$. Thus f can be approximated uniformly to arbitrary precision using at most r gates.

While our algebraic definition aligns with our intuition about approximate algorithms, it is easier to work with the following equivalent topological characterization from [3].

Theorem 3. *Let $A_n = \mathbb{C}[x_1, \ldots, x_n]$. The set $\{f \in A_n : \underline{L}(f) \leq r\}$ is the closure of the set $\{f \in A_n : L(f) \leq r\}$ in either the Zariski or Euclidean topologies.*

This allows us to write polynomials with small approximate complexity as limits of polynomials with small circuit complexity. The proof is based on two lemmas from Lehmkuhl and Lickteig [7].

Lemma 7. *Let $V \subset k^n$ and $W \subset k^m$ be affine varieties, $\varphi : V \to W$ a dominant morphism (i.e., $\overline{\varphi(V)}$ is dense in W), and $t \in W \setminus \varphi(V)$. Then there is a curve $C \subset V$ such that $t \in \overline{\varphi(C)}$, where the closure is taken in the Zariski topology.*

Proof sketch. We first inductively construct a decreasing sequence of subvarieties $V \supset V_{m-1} \supset \cdots \supset V_1$ where $t \in \overline{\varphi(V_j)}$ and $\dim \overline{\varphi(V_j)} = j$. At each step, we set $W_j = \overline{\varphi(V_{j+1})} \cap H$ where H is a hyperplane passing through t and intersecting W_{j+1} properly. We then choose V_j to be a component of $\varphi^{-1}(W_j)$ such that $\overline{\varphi(V_j)} = W_j$.

Once V_1 is constructed, we use a similar procedure to cut down its dimension until we obtain a curve. Pick $x \in W_1$ so that each component of $\varphi^{-1}(x)$ has dimension $\dim V_1 - 1$, and let H be a hyperplane whose intersection with φ^{-1} is proper and nonempty. Then we choose a component $V_1' \subset V_1 \cap H$ such that $\overline{\varphi(V_1')} = W_1$. □

While the next lemma is crucial in establishing Theorem 3, its proof is almost entirely abstract nonsense.

Lemma 8. *Let k be algebraically closed, $C \subset k^n$ be an affine curve, and $\varphi : C \to k^m$ be a morphism. For $t \in \overline{\varphi(C)}$, there exist formal Laurent series $p_1, \ldots, p_n \in k((\epsilon))$ such that the order of each p_i is at least $-\deg C$, and $F = \varphi(p_1, \ldots, p_n)$ is defined over $k[[\epsilon]]$ and satisfies $F_{\epsilon=0} = t$.*

Proof (of Theorem 3). We first verify that the Zariski and Euclidean closures are the same. By Theorem 2, it suffices to check that $\{f \in A_n : L(f) \leq r\}$ is constructible. Define an equivalence relation on arithmetic circuits by $\alpha \sim \beta$ if the circuits are identical up to the choices of their constants, and call the resulting equivalence classes *circuit configurations.* If Γ is a circuit configuration, let $P(\Gamma)$ be the set of polynomials computed by the circuits in Γ. Then $\{f \in A_n : L(f) \leq r\}$ is the finite union of the sets $P(\Gamma)$ where Γ is a configuration with size at most r. Moreover, if the circuits in Γ have m constant inputs, then evaluation yields a morphism $\varphi_\Gamma : \mathbb{C}^m \to \{f \in A_n : \deg f \leq 2^r\}$ with image $P(\Gamma)$, so each $P(\Gamma)$ (and hence their union) is constructible by Chevalley's Theorem.

One direction of the characterization is easy. If $f \in A_n$ and $\underline{L}(f) \leq r$, then there is a polynomial $F \in R[x_1, \ldots, x_n]$ with $F_{\epsilon=0}$ and a circuit α of size at most r computing F. For all but finitely many $y \in \mathbb{C}$, replacing ϵ with y in α yields a circuit with constants in $\mathbb{C}$ that computes $F_{\epsilon=y}$. Thus we can find a sequence $y_n \to 0$ such that $\lim_{n\to\infty} F_{\epsilon=y_n} = F_{\epsilon=0} = f$, so f is in the Euclidean closure of $\{f \in A_n : L(f) \leq r\}$.

For the other direction, suppose f is in the Zariski closure of $\{f \in A_n : L(f) \leq r\}$. Recall that we can write $\{f \in A_n : L(f) \leq r\}$ as a finite union of images of the morphisms φ_Γ. Thus there is some circuit configuration Γ with size at most r such that $f \in \overline{\varphi_\Gamma(\mathbb{C}^m)}$. By the first lemma, we can find a curve $C \subset \mathbb{C}^m$ such that $f \in \overline{\varphi_\Gamma(C)}$. So by the second lemma, there are Laurent series $p_1, \ldots, p_m$ having degree at least $-\deg C$ such that $F = \varphi_\Gamma(p_1, \ldots, p_m) \in \mathbb{C}[[\epsilon]]$ and $F_{\epsilon=0} = f$. Since C has at most r multiplication gates, we can replace $p_1, \ldots, p_m$ with their truncations to order $2^r \deg C$, yielding a polynomial $\tilde{F} \in \mathbb{C}[\epsilon]$ with $L(\tilde{F}) \leq r$ and $\tilde{F}_{\epsilon=0}$. □

As we mentioned before, the upshot of this characterization is that it allows us to explicitly write polynomials with low approximate complexity as limits of polynomials computed by small circuits. In the case where the target polynomial is a d-form, we can assume that the approximating polynomials are themselves d-forms.

Corollary 2. *Let $f \in A_n$ be a d-form with $\underline{L}(f) \leq r$. Then there is a sequence $\{f_n\}$ of d-forms with $L(f_n) \leq poly(r)$ such that $f_n \to f$ in the Euclidean topology.*

Proof. It suffices to show that given a polynomial $g = g_0 + g_1 + \cdots + g_n$ with complexity r and g_ℓ homogeneous of degree ℓ, we can find a circuit with size $poly(r, n)$ that computes g_d. The key identity that allows us to construct such a circuit is

$$g_0(x) + g_k(x) + g_{2k}(x) + \cdots + g_{\lfloor n/k \rfloor k}(x) = \frac{1}{k} \sum_{j=0}^{k-1} g(\omega^j x)$$

where ω is the principal k-th root of unity. To verify the identity, we expand the sum on the right hand side as

$$\begin{aligned}\sum_{j=0}^{k-1} g(\omega^j x) &= \sum_{j=0}^{k-1} \sum_{\ell=0}^{n} g_\ell(\omega^j x)\\ &= \sum_{\ell=0}^{n} g_\ell(x) \sum_{j=0}^{k-1} \omega^{j\ell},\end{aligned}$$

where the inner sum is k if ℓ is a multiple of k and 0 otherwise. Our goal now is to pick off the g_d term. Define

$$m_k(x) = g_k(x) + g_{2k}(x) + \cdots + g_{\lfloor n/k \rfloor k}(x).$$

The argument above shows that there is a $O(nr)$ circuit that computes each m_k. It is easy to check that we can write $g_1(x) = \sum_{k=1}^{n} c_k m_{kd}(x)$ where c_k is defined as follows. First, let $c_1 = 1$. If k is prime, let $c_k = -1$. Otherwise, recursively define $c_k = -1 - \sum c_j$ where the sum is taken over the proper factors of k. This construction thus gives an $O(n^2 r)$ circuit computing g_1. □

Remark: In the case where we consider weakly skew circuits, the result of this construction is also weakly skew and has size $O(r^3)$ by Lemma 4.

A result to this effect is implicitly assumed in [5, Prop. 9.3.2], so it is likely that this construction can be simplified – perhaps even with a smaller size blowup.

3 Algebro-Geometric Characterization

The unified approach of the GCT program is to formulate complexity theoretic conjectures in terms of the projective orbit closure problem:

> Let k be a field and let $V = S^n k^m$. Fix $f, g \in \mathbb{P}V$. If G is a subgroup of GL_m, is f contained in $\overline{G \cdot g}$?

Our goal is to characterize a conjecture along the lines of "the permanent is not a p-projection of the determinant" as a projective orbit closure problem. Recall that projection works by replacing a polynomial's variables with variables and constants, thus reducing the number of variables it depends on. So to examine whether the the permanent is p-projection of the determinant, we would need to compare per_m to $\det_{p(m)}$ for polynomially bounded p – but these polynomials live in different spaces!

To remedy this, we introduce a new variable ℓ to bump up the degree of the permanent. More precisely, suppose per_m depends on the variables $x_1, \ldots, x_{m^2}$ and $\det_n$ on $x_1, \ldots, x_{n^2}$ for $n > m$. Let ℓ be any one of the variables $x_{m^2+1}, \ldots, x_{n^2}$. Then by ignoring the rest of the variables, $\ell^{n-m}\,\mathrm{per}_m$ is a homogeneous polynomial of degree n over $x_1, \ldots, x_{n^2}$, and hence $\ell^{n-m}\,\mathrm{per}_m, \det_n \in S^n\mathbb{C}^{n^2}$.

Conjecture 2. *There does not exist a constant $c \geq 1$ such that*

$$\overline{GL_{m^{2c}} \cdot [\ell^{m^c - m} \operatorname{per}_m]} \subset \overline{GL_{m^{2c}} \cdot [\det_{m^c}]}$$

for sufficiently large m.

The group action here is the one described in Section 1.1, and the closures are taken in the Zariski topology. This conjecture is the primary object of study in [5] for the following reason.

Theorem 4. *Conjecture 2 is equivalent to $\{\operatorname{per}_m\} \notin \overline{\mathbf{VP}}_{\mathrm{ws}}$.*

First, a few preliminaries to simplify the proof.

Lemma 9. *Let V be an n-dimensional affine variety over $\mathbb{C}$, and let S be a nonempty subset with the property that if $x \in S$, then $\lambda x \in S$ for all nonzero $\lambda \in \mathbb{C}$. If $x \in S$ is nonzero, then $x \in \overline{S}$ if and only if $[x] \in \overline{[S]} \subset \mathbb{P}V$.*

Proof. Let $I([S])$ be the ideal of homogeneous polynomials vanishing on $[S]$, whose zero set $Z(I([S])) = \overline{[S]}$. Every polynomial in $I([S])$ also vanishes on S, so if $x \in \overline{S} = Z(I(S))$, then $[x] \in \overline{[S]}$.

Now suppose $x \in \overline{[S]}$, and let $f \in I(S)$. We need to show that $f(x) = 0$. Write $f = \sum_{i=0}^{d} f_i$ where f_i is homogeneous of degree i. Let s be any point in S. Then for every $\lambda \neq 0$,

$$0 = f(\lambda s) = \sum_{i=0}^{d} \lambda^i f_i(s),$$

so $f_i(s) = 0$ for every i. This shows that $f_i \in I([S])$, so $f_i(x) = 0$ and thus

$$f(x) = \sum_{i=0}^{d} f_i(x) = 0,$$

so $x \in Z(I(S)) = \overline{S}$.

□

Lemma 10. *Let $f \in S^n\mathbb{C}^{n^2}$. The closure of $GL_{n^2} \cdot f$ in $S^n\mathbb{C}^{n^2}$ is the same in the Euclidean topology as it is in the Zariski topology.*

Proof. Observe that GL_{n^2} is the complement of the determinantal hypersurface $\{\det_{n^2} = 0\} \subset \operatorname{Mat}_{n^2 \times n^2}$, so it is an open subset of an affine variety, hence a variety. The result follows from Chevalley's theorem applied to the morphism $\sigma \mapsto f \circ \sigma^{-1}$ from GL_{n^2} into $S^n\mathbb{C}^{n^2}$. □

Lemma 11. *Let G be a group acting by homeomorphisms on a first-countable topological space V. Then for $v, w \in V$, $w \in \overline{G \cdot v}$ if and only if $\overline{G \cdot w} \subset \overline{G \cdot v}$.*

Proof. The "if" direction is obvious. Suppose $w \in \overline{G \cdot v}$. Then there is a sequence $g_k \cdot v \to w$ where $g_k \in G$. For each $g \in G$, $g \cdot w = \lim_{k \to \infty} gg_k \cdot v$ by continuity, so $g \cdot w \in \overline{G \cdot v}$. We hence conclude that $\overline{G \cdot w} \subset \overline{G \cdot v}$. □

Proof (of Theorem 4). By Lemma 10, we can read Conjecture 2 in terms of the Euclidean topology. It suffices to show that there is no constant c such that

$$\ell^{m^c-m}\operatorname{per}_m \in \overline{GL_{m^{2c}} \cdot \det_{m^c}} \tag{1}$$

for sufficiently large m if and only if $\{\operatorname{per}_m\} \notin \overline{\mathbf{VP}}_{\mathrm{ws}}$. For the "if" direction, suppose there are constants c and m_0 such that (1) holds for all $m \geq m_0$. Then there is a sequence $\{\sigma_k\}$ in $GL_{m^{2c}}$ such that $\lim_{k\to\infty} \sigma_k \cdot \det_{m^c} = \ell^{m^c-m}\operatorname{per}_m$. By Lemma 5, there is a weakly-skew circuit of size polynomial in m^c computing $\det_{m^c}(x)$. Replacing each input from x with a circuit computing the relevant entry of $\sigma_k^{-1}x$ gives a weakly-skew circuit for $\sigma_k \cdot \det_{m^c}$ with size bounded by a polynomial $Cm^{c'}$. Now replace ℓ in this circuit with the constant 1, so it computes some polynomial f_k. Then $\lim_{k\to\infty} f_k = \operatorname{per}_m$ and $L_{\mathrm{ws}}(f_k) \leq Cm^{c'}$, so by Theorem 3's topological characterization of approximate complexity, $\{\operatorname{per}_m\} \in \overline{\mathbf{VP}}_{\mathrm{ws}}$.

Now suppose $\{\operatorname{per}_m\} \in \overline{\mathbf{VP}}_{\mathrm{ws}}$. Then there exists m_0 such that $\underline{L}_{\mathrm{ws}}(\operatorname{per}_m)$ is polynomially bounded for $m \geq m_0$. By Corollary 2, there is a constant c such that there is a sequence $\{f_k\}$ of m-forms with $L_{\mathrm{ws}}(f_k) < m^c$ such that $\lim_{k\to\infty} f_k = \operatorname{per}_m$. The completeness of the determinant (Lemma 6) tells us that each f_k is a p-projection of $\det_{m^c}$, so $f_k(x) = \det_{m^c}(M_k)$ where M_k is a $m^c \times m^c$ matrix with the x_i's and constants as its entries. We now homogenize with a new variable ℓ. This gives us $\ell^{m^c-m} f_k(x) = \ell^{m^c} f_k(x/\ell) = \det_{m^c}(M_k')$, where M_k' is obtained from M_k by replacing each constant entry b with $b\ell$ and leaving the variables x_i unchanged. Each entry of M_k' is now a linear form in $x_1, \ldots, x_{m^2}, \ell$, so because $GL_{m^{2c}}$ is dense in $\operatorname{Mat}_{m^c \times m^c}$, we have $\ell^{m^c-m} f_k \in \overline{GL_{m^{2c}} \cdot \det_{m^c}}$. □

In this equivalence, we only used two essential properties of $GL_{m^{2c}}$. First, in Lemma 9 we used its homogeneity over $\mathbb{C}$ (i.e. that if $\sigma \in GL_{m^{2c}}$, then $\lambda\sigma$ is too for $\lambda \in \mathbb{C} \setminus \{0\}$). Second, in Lemma 10 we used the fact that GL_{n^2} is variety. If we relax the first condition, we can easily extend the "only if" direction to a more general situation.

Theorem 5. *If G is an algebraic subgroup of $GL_{m^{2c}}$ and $\{\operatorname{per}_m\} \in \overline{\mathbf{VP}}_{\mathrm{ws}}$, then there is a constant $c \geq 1$ such that*

$$\overline{G \cdot [\ell^{m^c-m}\operatorname{per}_m]} \subset \overline{G \cdot [\det_{m^c}]}$$

for sufficiently large m.

That is, if one could prove non-containment of the projective orbit closures, then $\overline{\mathbf{VP}}_{ws}$ would be different from **VNP**. The upshot of looking at this generalization occurs when we take $G = SL_{m^{2c}}$ – indeed, these are the orbits that Mulmuley and Sohoni examine in their original paper – since several important SL-orbits turn out to be closed.

Definition 5. Let V be a representation of a reductive group G, and let $x \in V$. Then $[x]$ is *G-stable* if $G \cdot x$ is closed in V.

Note that $[\det_{m^c}]$ and $[\ell^{m^c-m}\operatorname{per}_m]$ are not stable with respect to the GL-actions discussed above, since their orbit closures contain zero.

Kempf [6, Cor. 5.1] showed that if G has no nontrivial central one-parameter subgroup, and if the stabilizer of $[x]$ in G is not contained in any proper parabolic subgroup of G, then $[x]$ is G-stable. In [9, Thm. 4.6, 4.7], Kempf's criterion is used to show that $[\det_m]$ and $[\operatorname{per}_m]$ are stable with respect to the action of SL_{m^2} on $S^m\mathbb{C}^{m^2}$. On the other hand, while $[\ell^{m^c-m}\operatorname{per}_m]$ is not stable under the action of $SL_{m^{2c}}$, we can start to understand it through the theory of *partial stability* discussed in §4.

4 Resolution Approaches

Here we present some of the highlights in [5] for using results from representation theory to actually resolve the projective orbit closure problems in the previous section. There is much more to say on this topic, so our goal is to just provide a bit of context for the relevant mathematical issues.

4.1 $\mathbf{VP_{ws}}$ vs. $\overline{\mathbf{VP_{ws}}}$

As a brief digression, we present the approach in [5] for examining whether our use of approximate complexity is necessary in Theorem 4. Recall from §2.4 that R is the subring of $k(\epsilon)$ of functions defined at $\epsilon = 0$.

Definition 6. Let $g \in S^n k^m$ and f in the GL_m-orbit closure of g. If there exist $q \in \mathbb{N}$, R-linear forms $y_1, \ldots, y_m$ in x, and $F \in S^n R^m$ such that

$$g(y_1, \ldots, y_m) = \epsilon^q f(x) + \epsilon^{q+1} F(x),$$

then we say that f can be approximated with order at most q along a curve in the orbit of g.

The following is a claim from [5]

> Suppose there is a polynomial $q(n)$ such that whenever $f \in \overline{GL_{n^2}\det_n} \subset S^n k^{n^2}$, we can approximate f with order at most $q(n)$ along a curve in the orbit of $\det_n$. Then $\mathbf{VP}_{\text{ws}} = \overline{\mathbf{VP}}_{\text{ws}}$.

The argument relies on showing that if g is a polynomial with $L_{\text{ws}}(g) < n$, then g is contained in the GL_{n^2}-orbit closure of $\det_n$. It is not clear why this makes sense; for instance, if g is a nonzero constant and $n > 1$, then g is not even in $S^n k^{n^2} \supset \overline{GL_{n^2} \cdot \det_n}$.

4.2 Coordinate Rings of Orbits

Let $X \subset \mathbb{P}V$ be a closed projective variety. If S is the space of polynomials over V^* and $I(X)$ is the ideal of homogeneous polynomials vanishing on X, we write $k[X] = S/I(X)$ for the homogeneous coordinate ring of X. In order to resolve

Conjecture 2, the GCT program attempts to show that for every $c \geq 1$, there are infinitely many m for which there is an irreducible $GL_{m^{2c}}$-module appearing in $\mathbb{C}[\overline{GL_{m^{2c}} \cdot [\ell^{m^c-m} \operatorname{per}_m]}]$, but not appearing in $\mathbb{C}[\overline{GL_{m^{2c}} \cdot [\det_{m^c}]}]$. The goal then is to understand these coordinate rings well enough to construct such a sequence of $GL_{m^{2c}}$-modules.

Let W be a vector space, V a $GL(W)$-module, and $v \in V$ nonzero. The coordinate ring $\mathbb{C}[\overline{GL(W) \cdot v}]$ has a natural grading by degrees. Namely, for $\delta \in \mathbb{N}$, we let $\mathbb{C}[\overline{GL(W) \cdot v}]_\delta$ consist of the equivalence classes of polynomials with degree δ. This is a grading because the proof of Lemma 9 shows that $I(\overline{GL(W) \cdot v})$ is a homogeneous ideal. The following proposition from [5] relates the coordinate rings discussed above.

Theorem 6. *Let V be a $GL(W)$-module and let $v \in V$ be $SL(W)$-stable. An irreducible $SL(W)$-module appears in $\mathbb{C}[SL(W) \cdot v]$ if and only if it appears in $\mathbb{C}[\overline{GL(W) \cdot v}]_\delta$ for some δ.*

It turns out that the $SL(W)$-stability of v allows us to describe $\mathbb{C}[SL(W) \cdot v]$ with just representation theory. Let G be a reductive group and index its irreducible G-modules by the set Λ_G^+ of dominant integral weights. For $\lambda \in \Lambda_G^+$, let $V_\lambda = V_\lambda(G)$ be the irreducible G-module with highest weight λ. If H is a subgroup of G and V is a G-module, let V^H be the subspace of V fixed by every element of H. If we further suppose H is a closed subgroup, we can decompose the coordinate ring of G/H as

$$\mathbb{C}[G/H] = \bigoplus_{\lambda \in \Lambda_G^+} (V_\lambda^*)^{\oplus \dim V_\lambda^H}.$$

Specializing to the case at hand, we let H be the stabilizer of v in $SL(W)$, so that $SL(W) \cdot v = SL(W)/H$.

4.3 Partial Stability

Theorem 6 gives the fundamental relationship between the coordinate rings of $\overline{GL(W) \cdot v}$ and $SL(W) \cdot v$, but only holds when v is $SL(W)$-stable. While $[\det_n]$ and $[\operatorname{per}_n]$ are SL_{n^2}-stable, $[\ell^{n-m} \operatorname{per}_m]$ is not. However, this point turns out out to be partially stable in the following sense, which will still allow us to relate the two coordinate rings.

Definition 7. Let G be a reductive group and V a G-module. Let P be a parabolic subgroup of G with Levi decomposition KU. Let R be a reductive subgroup of K. Let $G([v])$ denote the stabilizer of $[v]$ in G. Then $[v] \in \mathbb{P}V$ is *(R, P)-stable* if $U \subset G([v]) \subset P$ and $[v]$ is stable under the restricted action of R on V.

Let $W = A \oplus A' \oplus B$ where $A \cong \operatorname{Mat}_{m \times m}$ and $\dim A' = 1$. Let $\ell \in A'$ be nonzero. Then $[\ell^{n-m} \operatorname{per}_m]$ is (R, P)-stable, taking $R = SL(A)$, P the parabolic subgroup of G preserving $A \oplus A'$, and $K = GL(A \oplus A') \times GL(B)$.

Given partitions π and π', we write $\pi \mapsto \pi'$ if $\pi_1 \geq \pi'_1 \geq \pi_2 \geq \pi'_2 \geq \ldots$. Let $\lambda(\pi)$ be the highest weight of the Schur module $S_\pi \mathbb{C}^N$ considered as an SL_N-module, so $S_\pi \mathbb{C}^N = V_{\lambda(\pi)}$. Here is the main result from [5].

Theorem 7. *Let $v = \ell^{n-m} \operatorname{per}_m$.*

1. *A module $S_\nu W^*$ appears in $\mathbb{C}[\overline{GL(W) \cdot v}]_\delta$ if and only if $S_\nu(A \oplus A')^*$ appears in $\mathbb{C}[\overline{GL(A \oplus A') \cdot v}]_\delta$. If this is the case, then there is a partition ν' such that $\nu \mapsto \nu'$ and $V_\lambda(\nu')(SL(A)) \subset \mathbb{C}[\overline{SL(A) \cdot [v]}]_\delta$.*

2. *If $V_\nu(SL(A)) \subset \mathbb{C}[\overline{SL(A) \cdot [v]}]_\delta$, then there are partitions π, π' such that $\pi \mapsto \pi'$, $\lambda(\pi') = \nu$ and $S_\pi W^* \subset \mathbb{C}[\overline{GL(W) \cdot [v]}]_\delta$.*

3. *A module $V_\lambda(SL(A))$ appears in $\mathbb{C}[\overline{SL(A) \cdot [v]}]$ if and only if it occurs in $\mathbb{C}[SL(A) \cdot v]$.*

References

[1] Stuart Berkowitz, *On computing the determinant in small parallel time using a small number of processors*, Information Processing Letters **18** (1984), 147–150.

[2] Peter Bürgisser, *Cook's versus Valiant's hypothesis*, Theoretical Computer Science **235** (2000), 71–88.

[3] ______, *The complexity of factors of multivariate polynomials*, FOCS '01 Proceedings of the 42nd IEEE symposium on Foundations of Computer Science, 2001.

[4] Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi, *Algebraic complexity theory*, Springer, 1997.

[5] Peter Bürgisser, J. M. Landsberg, Laurent Manivel, and Jerzy Weyman, *An overview of mathematical issues arising in the geometric complexity approach to* **VP** $\neq$ **VNP**, arXiv:0907.2850v2, 2011.

[6] George Kempf, *Instability in invariant theory*, Annals of Mathematics **108** (1978), 299–316.

[7] Thomas Lehmkuhl and Thomas Lickteig, *On the order of approximation in approximative triadic decompositions of tensors*, Theoretical Computer Science **66** (1989), 1–14.

[8] Guillaume Malod and Natacha Portier, *Characterizing Valiant's algebraic complexity classes*, Journal of Complexity **24** (2008), no. 1, 16–38.

[9] Ketan Mulmuley and Milind Sohoni, *Geometric complexity theory I. An approach to the* **P** *vs.* **NP** *and related problems*, SIAM Journal on Computing **31** (2001), no. 2, 496–526.

[10] David Mumford, *The red book on varieties and schemes*, 2 ed., Springer, 1999.

[11] L. G. Valiant, *Completeness classes in algebra*, STOC '79 Proceedings of the eleventh annual ACM symposium on Theory of Computing, 1979.