On P vs. NP and Geometric Complexity Theory

Dedicated to Sri Ramakrishna

Ketan D. Mulmuley *
The University of Chicago

(Appears in JACM, vol. 58, issue 2, April 2011)

April 20, 2011

Abstract

This article gives an overview of the geometric complexity theory (GCT) approach towards the P vs. NP and related problems focussing on its main complexity theoretic results. These are: (1) two concrete lower bounds, which are currently the best known lower bounds in the context of the P vs. NC and permanent vs. determinant problems, (2) the Flip Theorem, which formalizes the self referential paradox in the P vs. NP problem, and (3) the Decomposition Theorem, which decomposes the arithmetic P vs. NP and permanent vs. determinant problems into subproblems without self referential difficulty, consisting of positivity hypotheses in algebraic geometry and representation theory and easier hardness hypotheses.

1 Introduction

Geometric complexity theory (GCT) is an approach towards the P vs. NP and related problems [C, Kp, Le, V] initiated in [GCTpram] with a proof of a special case of the $P \neq NC$ conjecture and developed in a series of articles [GCT1]-[GCT8] and [GCTflip], with further developments in [Bu, BLMW, Ku, LMR]. Intuitively, the P vs. NP problem is formidable,

*Supported by NSF grant CCF-1017760.

because being a universal statement about mathematics which says that discovery is hard, it can potentially preclude its own proof and be independent of the axioms of set theory. Resolution of this *self referential paradox* is the root difficulty underneath this problem as per the Flip Theorem in [GCTflip] which *formalizes* this paradox. As such, the main conceptual difficulty in any approach towards this problem is to *break the circle* of self reference by decomposing the problem and its variants into subproblems without self reference. The Decomposition Theorem in [GCT6] provides such decompositions for the arithmetic P vs. NP [GCT1] and permanent vs. determinant [V] problems based on positivity hypotheses in algebraic geometry and representation theory. This article gives an overview of these main complexity theoretic results in GCT.

In more detail, [GCT1] and [GCT2] define *geometric obstructions* (proof certificates of hardness) in these problems based on geometric invariant theory [MFK]. It is conjectured that showing existence of such geometric obstructions is equivalent to slightly stronger forms of the original hardness conjectures. Thus [GCT1] and [GCT2] provide conjecturally equivalent reformulations of the original arithmetic hardness conjectures in the setting of geometric invariant theory. The main advantage of this reformulation is that geometric obstructions have natural compact specifications (based on the classical results of Weyl) and this plays a crucial role in the subsequent story. But such an equivalent reformulation of the original hardness conjectures does not address the self referential paradox. To resolve it, an approach has to break the circle of equivalences.

Accordingly, the defining *flip strategy* of GCT to resolve the self referential paradox was subsequently formulated in [GCTflip]. The strategy is to go for an *explicit proof* of hardness. By this we essentially mean a proof that provides proof certificates of hardness, called *obstructions*, that are easy to verify and construct (in polynomial time). The strategy is called a flip because it reduces the lower bound problems to upper bound problems: showing that verification and construction of proof certificates belong to P. Section 2 explains in what sense the flip amounts to explicit resolution of the self referential paradox.

The articles [GCT3, GCT4, GCT5] investigate basic problems in representation theory suggested by this flip strategy. The article [GCT6] extends the investigation in [GCT1]-[GCT5] to provide an approach to implement the flip in characteristic zero, and thereby resolve the self referential paradox, assuming certain *positivity hypotheses* in algebraic geometry and representation theory. Specifically, the Decomposition Theorem (Theorem 4.15)

in [GCT6] shows how the original hardness conjectures can be *decomposed* into these positivity hypotheses (which are not self referential) plus easier hardness hypotheses which do not have the self referential difficulty once the positivity hypotheses are proved. All these hypotheses are supported by the strong Flip Theorem in [GCTflip] (cf. Theorem 4.6). This result shows that certain stronger versions of the arithmetic hardness and derandomization conjectures in complexity theory imply solutions to formidable explicit construction problems in algebraic geometry akin to (but even more explicit than) the explicit construction problems in these positivity hypotheses. This suggests that the positivity hypotheses here may be in essence *implications* of these stronger hardness and derandomization conjectures.

But the positivity hypotheses turn out to be formidable. They encompass and go far beyond nontrivial special cases of the classical plethysm problem [Fu] in algebraic geometry and representation theory. In view of the Strong Flip Theorem 4.6, problems comparable in difficulty to the explicit construction problems that arise in this theorem or in these positivity hypotheses may be expected in *any* approach towards the P vs. NP and related problems. This *law of conservation of difficulty* may explain why the fundamental hardness conjectures in complexity theory, which look so elementary at the surface, have turned out to be so formidable.

This article focuses on the arithmetic setting, wherein the underlying field of computation has characteristic zero. This setting captures the self referential difficulty in the boolean setting. Once this is resolved in the arithmetic setting, it is no longer an issue. But additional problems need to be resolved to change the base field of computation from Q or $\mathbb{C}$, as in the arithmetic setting, to a finite field, as in the boolean setting. These additional problems and the resulting decomposition of the boolean nonuniform P vs. NP problem will described in a later paper.

We also describe in this article two concrete lower bounds in GCT. These are currently the best known lower bounds in the context of the P vs. NC and permanent vs. determinant problems. The first lower bound is a special case of the $P \neq NC$ conjecture proved in [GCTpram] using a weaker form of the flip much before the stronger flip was formalized in [GCTflip] (cf. Section 3). The second lower bound in [LMR] for the permanent vs. determinant problem implies (but is stronger than) the earlier quadratic lower bound for the permanent in [MR] (cf. Section 4.4).

One may ask if GCT can be used to reprove the earlier known lower bounds in complexity theory such as the ones for constant depth [BS] or

monotone [Rz] circuits. Unfortunately, the answer is no. The explicit construction problems in GCT for constant depth circuits are not fundamentally different from the ones in GCT for polylogarithmic depth circuits. GCT is geared towards proving lower bounds in natural and realistic models of computation that are powerful enough to allow efficient computation of the determinant and related algebraic problems. This is because the determinant lies at the foundation of algebraic geometry and representation theory. The explicit construction (upper bound) problems in these fields that arise in the flip strategy of GCT cannot be solved in a model in which the determinant cannot be computed efficiently. As such, GCT does not help in the world without determinants. By this, we mean the models of computation, such as constant depth, monotone, or quadratic size circuits, in which efficient computation of the determinant is not possible. Not surprisingly, both the lower bounds in GCT mentioned above are in the models in which efficient computation of the determinant is possible.

The rest of this article is organized as follows. Section 2 describes the flip strategy to resolve the self referential paradox, and the Flip Theorem in [GCTflip] that formalizes this paradox. Section 3 describes the special case of the $P \neq NC$ conjecture proved in [GCTpram]. Section 4 gives an overview of the GCT approach to implement the flip for the permanent vs. determinant problem [V] in characteristic zero. It describes the Decomposition Theorem in [GCT6] that decomposes this problem into subproblems without self referential difficulty based on the positivity hypotheses in algebraic geometry and representation theory. The story for the arithmetic P vs. NP problem defined in [GCT1] is similar; cf. Decomposition Theorem 4.15. The lower bound in [LMR] based on GCT is also described in Section 4. Section 5 gives concluding remarks.

The first half of this paper does not assume any familiarity with algebraic geometry or representation theory. The second half (from Section 4 onwards) assumes familiarity with basic notions in these fields. They are reviewed in the Appendix for the readers not familiar with them. It may also be helpful to go through the video [GCTtutorial] of the FOCS 2010 tutorial based on this overview.

Acknowledgement: The author is grateful to Janos Simon and Josh Grochow for helpful discussions, and to the referees for helpful comments.

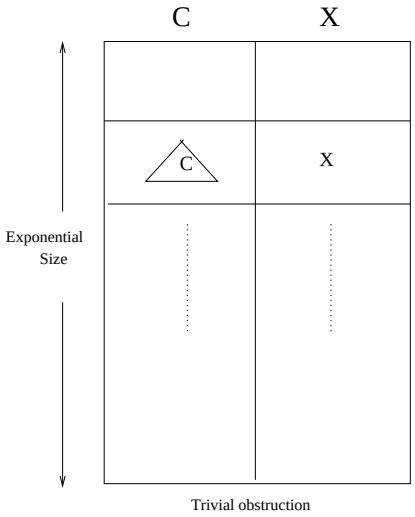

Figure 1: Trivial obstruction

2 The flip and the self referential paradox

In this section we describe the defining *flip strategy* of GCT to resolve the self referential paradox and the Flip Theorem in [GCTflip] that formalizes this paradox.

Towards this end, let us fix an NP-complete function $f(X) = f(x_1, \ldots, x_n)$, say SAT. The goal of the nonuniform P vs. NP problem is to show that there does not exist a circuit C of size $m = \text{poly}(n)$ that computes $f(X)$, $n \to \infty$. Symbolically, let $C(X)$ denote the function computed by C. Then we want to prove that

$$\forall n, m = \text{poly}(n)\ \forall C \exists X : f(X) \neq C(X). \tag{1}$$

Equivalently, the goal is to prove:

Hard Obstruction Hypothesis (HOH): For every large enough n, and $m = \text{poly}(n)$, there exists a *trivial obstruction (i.e. a "proof-certificate" of hardness)* to efficient computation of $f(X)$. Here by a trivial obstruction we mean a table (cf. Figure 1) that lists for every small C a counterexample X such that $f(X) \neq C(X)$.

The number of rows of this table is equal to the number of circuits of size $m = \text{poly}(n)$. Thus the size of this table is exponential. The time to verify

whether a given table is a trivial obstruction is also exponential, and so is the time of the obvious algorithm to decide if such a table exists for given n and m. From the complexity theoretic viewpoint, this is a hard task. So we call this trivial brute force strategy for proving the nonuniform P vs. NP conjecture, based on existence of trivial obstructions, a *hard strategy*. Hence, the terminology Hard Obstruction Hypothesis. It is really just a restatement of the original problem.

Any proof strategy for the P vs. NP problem has to answer the following question:

Question 2.1 *(a) In what sense is the strategy fundamentally different from the trivial brute force strategy above and not just a restatement of the original problem?*

(b) Why should the strategy be even feasible, i.e., lead to a proof of $O(1)$ (finite) size?

Question (b) is basically the one we discussed in the introduction: namely, why should the $P \neq NP$ conjecture even have a proof? Why can it not be independent of the axioms of the set theory? The natural proof barrier [RR] says that the proof technique used for proving lower bounds for constant depth circuits cannot be used to prove $P \neq NP$ essentially because the conjecture itself stands in its way. Why cannot this conjecture stand in the way of every proof technique like this? As we shall see, (a) and (b) are closely related. Hence, let us address (a) first. We will address (b) later in Section 2.4.

2.1 The flip

In the context of (a), the most natural abstract strategy that is fundamentally better than the trivial strategy is suggested by the P vs. NP problem itself. Before we formally define it, let us first see what is wrong with the trivial obstruction from the complexity-theoretic perspective. That is quite clear. First, it is long, i.e., its description takes exponential space. Second, it is hard to verify (and also construct); i.e., this takes exponential time. Since NP is the class of problems with proof-certificates that are short (of polynomial-size) and easy to verify (in polynomial-time), this then leads to the following strategy for proving the nonuniform $P \neq NP$ conjecture, based on proof certificates (obstructions) that are short, and easy to verify (and also easy to construct). We call this strategy *the flip*. This refers to the

transition from the "hard" (exponential time verifiable/constructible) trivial obstructions to the "easy" (polynomial time verifiable/constructible) new obstructions. It also refers to the transition from the lower bound problem to the upper bound problem—specifically, the problem of finding an efficient algorithm to verify and construct an obstruction.

Formally, we say that a technique for proving the nonuniform $P \neq NP$ conjecture (using the function $f(X)$) is a *flip* if there exists a family $\mathcal{O} = \cup_{m,n}\mathcal{O}_{n,m}$ of sets of bit strings called *obstructions* (or obstruction labels), which serve as proof certificates of hardness of $f(X)$, having the following Flip properties F0-F4.

F0 [Short]: The set $\mathcal{O}_{n,m}$ is nonempty and contains a short obstruction string s if m is small, i.e., when $m = O(\text{poly}(n))$, or more generally, $m = O(2^{\log^a n})$, $a > 1$ a fixed constant. Here short means the bit length $\langle s \rangle$ of s is $\text{poly}(n, m)$.

To state F1, we define a *small global obstruction set* $S_{n,m}$ to efficient computation of $f(X)$, for given n and m, to be a small set $\{X_1, \ldots, X_l\}$, $l = \text{poly}(n, m)$, of inputs such that, for any circuit C of size $\leq m$, $S_{n,m}$ contains a counterexample $X_C = X_j$, for some $j \leq l$, such that $f(X_C) \neq C(X_C)$. Then:

F1 [Easy to decode:] Each bit string $s \in \mathcal{O}_{n,m}$, m small and s short, denotes a small global obstruction set $S_{n,m}(s)$ to efficient computation of $f(X)$ such that: (a) given s, n and m, $S_{n,m}(s)$ can be computed in $\text{poly}(\langle s \rangle, n, m)$ time–in particular, if s is short, the size of $S_{n,m}(s)$ is $\text{poly}(n, m)$–and, (b) given s, n, m and any circuit C of size $\leq m$, a set $S_{n,m,C}(s) \subseteq S_{n,m}(s)$ of $O(1)$ size can be computed in $\text{poly}(\langle s \rangle, n, m)$ time such that $S_{n,m,C}(s)$ contains some counterexample X_C such that $f(X_C) \neq C(X_C)$. A stronger form of (b) is (b'): given s, n, m and C, a counterexample $X_C \in S_{n,m}(s)$ as above can be computed in $\text{poly}(\langle s \rangle, n, m)$ time (we do not explicitly study this variant in this paper).

F2 [Rich]: For every n and $m = \text{poly}(n)$, $\mathcal{O}_{n,m}$ contains at least $2^{\Omega(m)}$ pairwise disjoint obstructions, each of $\text{poly}(n, m)$ bitlength. Here we say that two obstructions $s, s' \in \mathcal{O}_{n,m}$ are disjoint if $S_{n,m}(s)$ and $S_{n,m}(s')$ are disjoint.

F3 [Easy to verify]: Given n, m and a string s, whether s is a valid obstruction string for n and m–i.e., whether $s \in \mathcal{O}_{n,m}$–can be verified in $\text{poly}(n, \langle s \rangle, m)$ time.

F4 [Easy to construct]: For each n and $m = \text{poly}(n)$, a valid obstruction string $s_{n,m} \in \mathcal{O}_{n,m}$ can be constructed in $\text{poly}(n, m) = \text{poly}(n)$ time.

This finishes the description of F0-4 defining a flip.

We say that a proof of the $NP \not\subseteq P/poly$ conjecture (using $f(X)$) is *extremely explicit* if it proves existence of an obstruction family $\mathcal{O}$ satisfying F0-4. It is called explicit (or strongly explicit) if only F0, F2 and F3 (resp. F0, F2,3, and 4) are formally proved. By GCT, we henceforth mean any approach that is *geometric and explicit*.

We can similarly define the flip and explicit proofs in the context of other lower bound problems in complexity theory, such as the $\#P$ vs. NC problem or the P vs. NC problem. An "easy" algorithm means an NC-algorithm in this context. We can also define these notions for the arithmetic permanent vs. determinant problem [V] or the arithmetic (nonuniform) version of the P vs. NP conjecture [GCT1] described below (Section 2.3). In the arithmetic setting, the underlying field of computation in the circuit is a field of characteristic zero, such as Q or $\mathbb{C}$.

2.2 Self-referential paradox

We now explain in what sense implementation of the the flip amounts to extremely explicit resolution of the self referential paradox.

Towards this end, let us examine the properties F in Section 2.1 more closely. For an obstruction $s \in \mathcal{O}_{n,m}$, let $S_{n,m}(s)$ denote the corresponding global obstruction set in F1 (a) (for decoding) that can be computed in polynomial time. To simplify the argument, let us replace F1 (b) by (b)'. The decoding algorithm in (b)' gives in polynomial time a counterexample $X_C \in S_{n,m}(s)$ for every small circuit C of size $\leq m$. Let $\tilde{S}_{n,m}(s)$ denote the trivial obstruction of exponential size that lists for every small C this X_C. Then the new obstruction $S_{n,m}(s)$ can be thought of as a polynomial size encryption (information compression) of the trivial obstruction $\tilde{S}_{n,m}(s)$; cf. Figure 2. To verify a given row of $\tilde{S}_{n,m}(s)$, we have to check if $f(X_C) \neq C(X_C)$ for the C corresponding to that row. For general X_C, this cannot be done in polynomial time, assuming $P \neq NP$, since f is NP-complete. Yet F3 (for verification) says that whether s is a valid obstruction, i.e., whether each of the *exponentially many* rows of $\tilde{S}_{n,m}(s)$ specifies a counterexample, can be verified in polynomial time. At the surface, it may seem that to prove $P \neq NP$, this requires proving $P = NP$. Implementation of the flip thus amounts to extremely explicit resolution of this self referential paradox. At the surface it seems impossible.

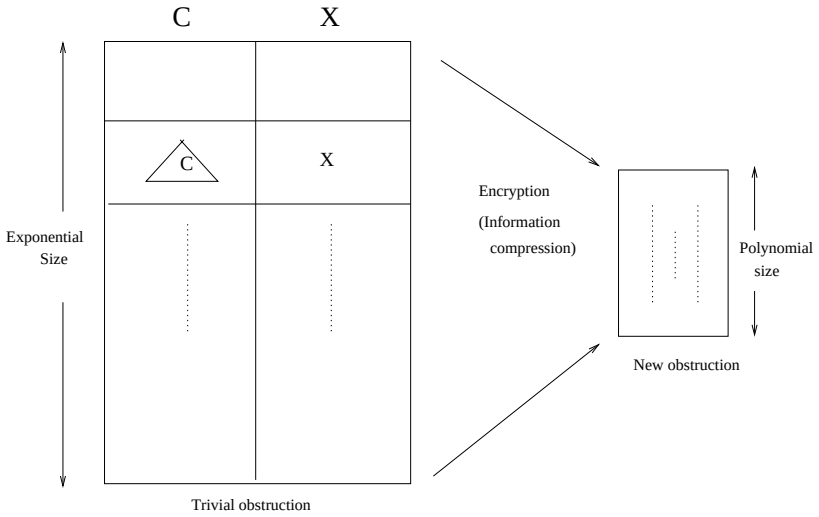

Figure 2: The new obstruction $S_{n,m}(s)$ as an encryption of the trivial obstruction $\tilde{S}_{n,m}(s)$

2.3 The flip theorem

If so, why are we going for explicit proofs, especially when proving existence of obstructions even nonconstructively suffices in principle? The reason is provided by the following Flip Theorem 2.3. It says that any proof of the arithmetic version of the P vs. NP conjecture can be converted into an extremely explicit proof assuming that circuit identity testing [KI] can be derandomized in a blackbox fashion. This standard derandomization assumption is generally regarded as easier than the target lower bound. Thus the arithmetic P vs. NP conjecture *forces* an extremely explicit resolution of the self referential paradox, modulo derandomization. This formalizes the self referential paradox in the arithmetic setting. There is also a similar result for an invariant theoretic average case version of the boolean $NP \not\subseteq P/poly$ conjecture [GCTflip]. We only consider the nonuniform setting here. See [GCTflip] for analogous results in the uniform setting.

We begin with a preliminary lemma.

Lemma 2.2 (Flip) *(cf. GCTflip) Suppose the permanent of an $n \times n$ integer matrix X cannot be computed by any arithmetic circuit (over Q) of $m = poly(n)$ total bit size. Suppose also that the complexity class E (con-*

sisting of the problems that can be solved in exponential time) does not have subexponential size circuits, or, less stringently, that black box polynomial identity testing [Ag, KI] can be derandomized (cf. GCTflip for a precise statement of what this means). Then:

(1) For every n and $m = poly(n)$, it is possible to compute in $poly(n, m) = poly(n)$ time a small set $S_{n,m} = \{X_1, \ldots, X_l\}$, $l = poly(n, m) = poly(n)$, of $n \times n$ integer matrices such that for every arithmetic circuit C of total bit size $\leq m$, $S_{n,m}$ contains a matrix X_C which is a counter example against C; i.e, such that $perm(X_C)$ is not equal to the value $C(X_C)$ computed by the circuit. The set $S_{n,m}$ is thus a small global obstruction set of $poly(n, m) = poly(n)$ size against all small circuits of total bit size $\leq m$.

(2): Furthermore, assuming a slight strengthening (given in GCTflip) of the assumption that E does not have subexponential size circuits , or less stringently, that black box polynomial identity testing can be derandomized, arithmetic hardness of the permanent has an extremely explicit proof. Specifically, there exists, for every n and $m = poly(n)$, a set $\tilde{\mathcal{O}}_{n,m}$ of obstructions (bit strings) satisfying F0-F4. (F0-F4 for arithmetic hardness of the permanent are very similar to F0-F4 in Section 2.1. Hence we do not state them here).

This result (except for F1 (b)) follows easily by derandomizing [NW, IW] the co-RP algorithm in [KI] for testing if a given arithmetic circuit computes the permanent using its downward self-reducibility. But we cannot prove an analogous result for the P vs. NP problem using self reducibility alone. Using downward self reducibility, the article [At] gives, assuming $NP \not\subseteq P/poly$, a probabilistic polynomial time algorithm for finding, given any small circuit C, a counterexample on which it differs from SAT; but this algorithm cannot efficiently produce a small *global* obstruction set against *all* small circuits. The best earlier result in the context of the P vs. NP problem was proved in [FPS]. It gave a probabilistic polynomial time algorithm with an access to the SAT oracle for computing a small set of satisfiable formulae that contains a counterexample against every small circuit claiming to compute SAT. The main difficulty in the context of the P vs. NP problem is to accomplish the same task in polynomial time under reasonable complexity theoretic assumptions without any access to the SAT oracle. This difficulty is overcome in the arithmetic setting in the following result (Theorem 2.3).

Before stating it, let us review the arithmetic nonuniform version of the P vs. NP problem from [GCT1]. The role of the permanent is played in this

problem by the following function $E(X)$ defined over Q. Take a set $\{X_i^j | 1 \leq j \leq k,\ 1 \leq i \leq m\}$ of m-dimensional vector variables, for a fixed constant $k \geq 3$. Here each X_i^j is an m-vector. So there are km vector variables overall. Let X be the $m \times km$ variable matrix whose columns consist of these km variable vectors. For any function $\sigma : \{1, \ldots, m\} \rightarrow \{1, \ldots, k\}$, let $\det_\sigma(X)$ denote the determinant of the matrix X_σ whose i-th column is $X_i^{\sigma(i)}$. Define $E(X) = \prod_\sigma \det_\sigma(X)$ where σ ranges over all such functions. The function $E(X)$ is also characterized by symmetries (cf. GCTflip) just like the permanent. Let $n = km^2$ be the total number entries in X. By the (nonuniform) arithmetic P vs. NP problem in characteristic zero we mean the problem of showing that $E(X)$ is hard in the arithmetic setting. Specifically, the problem is to show that $E(X)$ cannot be computed by an arithmetic circuit over Q or $\mathbb{Z}$ of poly(n) size. This is a formal implication of the usual nonuniform P vs. NP problem (i.e., $NP \not\subseteq P/poly$ conjecture) since deciding if $E(X)$ is zero is NP-complete [Gu].

The following is an analogue of Lemma 2.2 in this setting.

Theorem 2.3 (Flip) *(cf. GCTflip) The arithmetic nonuniform P vs. NP conjecture above has an extremely explicit proof, assuming that it holds and that black box polynomial identity testing can be derandomized.*

This is proved using the fact that $E(X)$ (like perm(X)) is characterized by its symmetries (cf. Section 4.1) in conjunction with the hardness vs. randomness principle [NW, IW, KI]. See [GCTflip] for an analogue of Theorem 2.3 for a stronger average case form of the boolean $NP \not\subseteq P/poly$ conjecture based on the above characterization of $E(X)$ by its symmetries. This result implies that any NP-complete function (e.g. SAT) has an extremely explicit proof assuming this average case conjecture and that BPP can be derandomized in a black box fashion.

The article [GCTflip] also proves stronger versions of Lemma 2.2 and Theorem 2.3 by combining the hardness vs. randomness principle [IW, KI] and characterization by symmetries with classical algebraic geometry. These stronger versions say that certain stronger forms of the arithmetic hardness and derandomization conjectures under consideration imply polynomial time algorithms for formidable explicit construction problems in algebraic geometry. We shall state such a result (cf. Theorem 4.6) in the context of the permanent vs. determinant problem later.

2.4 Breaking the circle

In view of the flip theorems, the main conceptual difficulty in any approach towards the P vs. NP and related problems is to *break the circle* of self referential difficulty around them by decomposing each hardness problem into subproblems of the form

$$hardness \cong subproblem1 + subproblem2 + \cdots \tag{2}$$

with a reasonable justification as to why each subproblem on the right hand side does not have the self referential difficulty. This can then be taken as a concrete evidence that the approach can resolve the self referential paradox and lead to proofs of finite size of the fundamental hardness conjectures, thereby answering Question 2.1 (b).

This self referential difficulty is an issue only if the lower bound problem under consideration is at least as hard as derandomization of polynomial or determinant identity testing. Otherwise, an analogous flip theorem for such a lower bound will really be a statement about the difficulty of the additional derandomization hypothesis. Not surprisingly, the known lower bounds that are easier than derandomization, such as a quadratic lower bound for the permanent [MR] or lower bounds for monotone [Rz] or constant depth [BS] circuits, have proofs that are far from explicit. For example, $PARITY \notin AC_0$ [BS] is proved by decomposing this problem into two subproblems: (1) Show that AC_0 is easy to approximate by low degree polynomials, and (2) Show that PARITY is hard to approximate by such polynomials. Here (1) is an easiness hypothesis that depends solely on AC_0, and (2) is an hardness hypothesis that is easier to prove than the original hardness hypothesis. Thus this decomposition *decouples* $PARITY$ (and NC_1) from AC_0. Though the proof of this decomposition is trivial, it is the main high-level conceptual step in the proof of this lower bound. We want similar decompositions of the fundamental hardness conjectures in complexity theory that decouple the underlying complexity classes. But this task is far harder for the conjectures harder than derandomization, because this decoupling amounts to breaking the circle of self referential difficulty.

Indeed, there is a fundamental difference between the P vs. NP problem and the lower bound problems in the restricted models of computation such as constant depth circuits. The latter are statements about the *weakness* of the restricted models. In contrast, by the flip Theorem 2.3, the P vs. NP problem is a statement about the *strength* of the complexity class P. It does not say that P is small and weak, but rather that P is big and strong–

strong enough to assert that "I am different from NP". This assertion is self referential and also paradoxical, being a statement of strength rather than weakness. To resolve this self referential paradox, one essentially has to show (modulo derandomization) that P contains formidable explicit construction problems akin to the ones which arise in the flip Theorem 2.3 and the strong flip Theorem 2.3 stated later. The same is also true for other fundamental hardness conjectures harder than derandomization. This strength of P and related complexity classes makes decomposition leading to decoupling in the fundamental hardness conjectures in complexity theory a challenge.

The Decomposition Theorem described in Section 4 gives the sought concrete decompositions of the type (2) of the arithmetic P vs. NP and permanent vs. determinant problems into subproblems without self referential difficulty. This decomposition is fundamentally different from the decomposition of the $PARITY$ vs. AC_0 problem in two ways. First, it is based on the flip strategy going towards explicit proofs. This is natural, though not necessary, in view of the flip Theorem 2.3. It is also consistent with the natural proof barrier [RR] by which decompositions based on approximation cannot work. Second, the decomposition is based on nonelementary constructions in algebraic geometry and representation theory.

3 The $P \neq NC$ result without bit operations

Now one may ask why we are going via algebraic geometry to get such decompositions when the lower bound problems under consideration have elementary statements that make no mention of algebraic geometry. To see why, we describe in this section the special case of the $P \neq NC$ conjecture proved in [GCTpram] via algebraic geometry before we turn to the Decomposition Theorem. We shall call this special case the $P \neq NC$ result without bit operations. This can be considered to be the first concrete lower bound result of GCT. It was proved using a weaker form of the flip much before the stronger form was formalized in [GCTflip]. It says that:

Theorem 3.1 *(cf. [GCTpram]) The P-complete max-flow problem cannot be solved in polylog(N) parallel time using poly(N) processors in the PRAM model without bit operations, where N denotes the total bit length of the input, not just the number of input parameters.*

The model here is the usual PRAM model with arithmetic $+, -, *$, comparison and branching operations, but no bit operations. It includes virtu-

ally all known parallel algorithms for algebraic and weighted optimization problems. Most importantly, it contains efficient algorithms [KV] for computing the determinant and roots of polynomials. Hence it is quite realistic and natural in contrast to the the constant depth [BS] or monotone [Rz] circuit models used for proving lower bounds earlier. Theorem 3.1 seems to be the only known superpolynomial lower bound that is a nontrivial implication of a fundamental separation conjecture (comparable to the $P \neq NC$ conjecture) and holds unconditionally in a natural and realistic model of computation powerful enough to allow efficient computation of the determinant and roots of polynomials.

The proof of Theorem 3.1 is based on classical algebraic and diophantine geometry, though the result itself can be stated in purely elementary and combinatorial terms. No elmentary proof of this result is known so far. This should explain why we are going via algebraic geometry in GCT. After all, Theorem 3.1 is a much weaker implication of the $P \neq NC$ conjecture since it does not involve the self referential difficulty (being easier than derandomization). Until we can prove this result or a comparable implication of a fundamental hardness conjecture without algebraic geometry, it is unrealistic to expect that we will be able to prove the $P \neq NC$ conjecture or anything comparable without it.

The proof of Theorem 3.1 is *quasi-explicit* in the sense that (1) it produces a global obstruction set S_N of $2^{\log^{O(a)} N}$ inputs that contains a counterexample against every branching program C (i.e. a circuit with arithmetic, branching and comparison operations) of depth $\log^a N$ and size $2^{\log^a N}$, for any positive constant a, and (2) S_N can be constructed in $O(\log^{O(a)} N)$ time using $2^{\log^{O(a)} N}$ processors in the PRAM model without bit operations. Specifically, for given N, it produces an explicit parametrized graph $g_N(z_1, z_2)$ with $2^{\log^{O(a)} N}$ nodes and edges labelled with capacities that are linear forms in z_1 and z_2 such that any branching program C of depth $\log^a N$ and size $2^{\log^a N}$ fails to compute the max flow correctly on $g_N(z_1, z_2)$ for some integral z_1 and z_2 of bitlengh $O(\log^{O(a)} N)$. The set S_N consists of instantiations of $g_N(z_1, z_2)$ for all integral z_1 and z_2 of $O(\log^{O(a)} N)$ bitlength.

But this proof technique cannot prove the unrestricted $P \neq NC$ conjecture for two reasons. First, it is quasi-explicit instead of explicit. This is too weak to resolve the self referential paradox in the P vs. NC conjecture. Second, it associates with a computation an algebraic object (a semialgebraic set) and then reasons solely on the basis of the degree of this object. It was pointed out in [GCTpram] (cf. Chapter 7 therein) that any such purely

degree based proof technique can not work in the context of the fundamental separation conjectures such as the P vs. NC conjecture. The recent article [AW] also says something similar.

The article [GCTpram] also suggested an idea for overcoming this algebraic degree barrier: namely, associate with the fundamental complexity classes algebraic varieties with group actions that capture the symmetries of computation and then reason on the basis of the deeper representation theoretic structure of these varieties rather than just their degrees. This was the starting point for the investigation in [GCT1]-[GCT8] via geometric invariant theory.

4 GCT approach to implement the flip

In this section we outline the resulting GCT approach to implement the flip via geometric invariant theory focussing on the permanent vs. determinant problem [V] in characteristic zero (over $\mathbb{C}$), the story of the arithmetic P vs. NP problem being similar. The goal is to decompose this problem (cf. Decomposition Theorem 4.15) into positivity hypotheses in algebraic geometry and representation theory and an easier hardness hypothesis, all without self referential difficulty. We also describe the concrete lower bound in [LMR] based on a weaker form of the flip. The basic notions of representation theory and algebraic geometry needed in this section are reviewed in the appendix. The reader not familiar with algebraic geometry and representation theory may refer to it whenever necessary.

The permanent vs. determinant problem (in characteristic zero, over $\mathbb{C}$) is to show that $\mathrm{perm}(X)$, the permanent of an $n \times n$ variable matrix X, cannot be represented linearly as $\det(Y)$, the determinant of an $m \times m$ matrix Y, if $m = \mathrm{poly}(n)$, or more generally, $m = 2^{\log^a n}$, for a fixed constant $a > 0$, and $n \to \infty$. The best known lower bound on m at present is quadratic [MR]. Here, by a linear representation, we mean that the entries of Y are (possibly nonhomogeneous) linear functions (over $\mathbb{C}$, Q, or Z) of the entries of X.

The goal now is to make enough progress towards the flip properties F until we get a concrete decomposition of the permanent vs. determinant problem of the form (2).

4.1 Characterization by symmetries

The starting point is an observation [GCT1] that the permanent and determinant are *exceptional* polynomial functions, where by exceptional we mean they are completely characterized by symmetries in the following sense.

Let Y be a variable $m \times m$ matrix. Let $\mathbb{C}[Y]_m$ be the space of homogeneous forms of degree m in the m^2 variable entries of Y. Then, by classical representation theory [Fr], $\det(Y)$ is the unique nonzero form in $\mathbb{C}[Y]_m$, up to a constant multiple, such that, for any $m \times m$ invertible matrices A and B with $\det(A)\det(B) = 1$,

(D): $\det(Y) = \det(AY^*B)$,

where $Y^* = Y$ or Y^t. Thus $\det(Y)$ is completely characterized by its symmetries, and hence, is exceptional. We refer to this characteristic property of the determinant as property (D) henceforth.

Similarly, by classical representation theory [MM], $\mathrm{perm}(X)$ is the unique form, up to a constant multiple, in the space $\mathbb{C}[X]_n$ of homogeneous forms of degree n in the entries of X such that, for any diagonal or permutation matrices A and B,

(P): $\mathrm{perm}(X) = \mathrm{perm}(AX^*B)$,

where $X^* = X$ or X^t, and the product of the entries of A is one, when A is diagonal, and similarly for B. Thus $\mathrm{perm}(X)$ is also completely characterized by its symmetries, and hence, is exceptional. We shall refer to this characteristic property of the permanent as property (P) henceforth.

The proofs of the flip Lemma 2.2 and Theorem 2.3 are based on the characterization by symmetries. Hence a natural strategy for implementing the flip is to exploit this characterization by symmetries.

4.2 Geometric obstructions

Towards that end, we first recall from [GCT1, GCT2] the notion of *geometric obstructions* which would let us exploit these symmetries of the permanent and the determinant.

This is done in two steps. First, [GCT1] constructs certain projective algebraic varieties $\Delta[\mathrm{perm}, n, m]$ and $\Delta[\det, m]$ such that if $\mathrm{perm}(X)$, X an $n \times n$ variable matrix, can be represented linearly as $\det(Y)$, with $\dim(Y) = m > n$, then

$$\Delta[\text{perm}, n, m] \subseteq \Delta[\det, m]. \tag{3}$$

The goal is to show, using algebraic geometry and representation theory, that this inclusion (3) is impossible when $m = poly(n)$, or more generally, when $m = 2^{\log^a n}$, $a > 0$ a constant.

Here by a projective algebraic variety we mean the zero set of a system of homogeneous multivariate polynomials with coefficients in $\mathbb{C}$; cf. Appendix. The formal definition of $\Delta[\text{perm}, n, m]$ and $\Delta[det, m]$ is as follows.

Let Y be an $m \times m$ variable matrix. We think of its entries, ordered say rowwise, as coordinates of $\mathcal{Y} = \mathbb{C}^r$, $r = m^2$. Let X be an $n \times n$ variable matrix. We identify it with an $n \times n$ submatrix of Y, say, the bottom-right minor of Y, and let z be any variable entry of Y outside X. We use it as a homogenizing variable.

Let $V = \mathbb{C}[Y]_m$ be the space of homogeneous polynomials of degree m in the variable entries of Y. It is a representation of $G = GL(\mathcal{Y}) = GL_r(\mathbb{C})$ with the following action. Given any $\sigma \in G$, map a polynomial $g(Y) \in V$ to $g^\sigma(Y) = g(\sigma^{-1}(Y))$:

$$\sigma : g(Y) \longrightarrow g(\sigma^{-1}Y). \tag{4}$$

Here Y is thought of as an m^2-vector by straightening it rowwise.

Let $P(V)$ be the projective space of V consisting of the lines in V through the origin. Let $g = \det(Y)$, thought of as a point in $P(V)$, and let $f = z^{m-n}\text{perm}(X)$, again thought of as a point in $P(V)$.

Let

$$\begin{array}{lcl} \Delta[\det, m] & = & \overline{Gg} \subseteq P(V), \\ \Delta[\text{perm}, n, m] & = & \overline{Gf} \subseteq P(V), \end{array} \tag{5}$$

where $\overline{Gg}$ denotes the projective closure (cf. Appendix) of the orbit Gg of g. Then it follows from classical algebraic geometry that $\Delta[\det, m]$ and $\Delta[\text{perm}, n, m]$ are projective varieties. Furthermore, it can be shown that they are projective G-varieties, i.e., varieties with a natural action of G (that moves the points in the varieties around) induced by the action on the G-orbits. We call $\Delta[\text{perm}, n, m]$ a *class variety* of the complexity class $\#P$ based on the permanent function which is $\#P$-complete [V], and $\Delta[\det, m]$ a *class variety* of the complexity class NC based on the determinant which belongs to NC and is almost complete [V].

It is easy to show (cf. Propositions 4.1 and 4.4 in [GCT1]) that if $\text{perm}(X)$ can be expressed linearly as the determinant of an $m \times m$ ma-

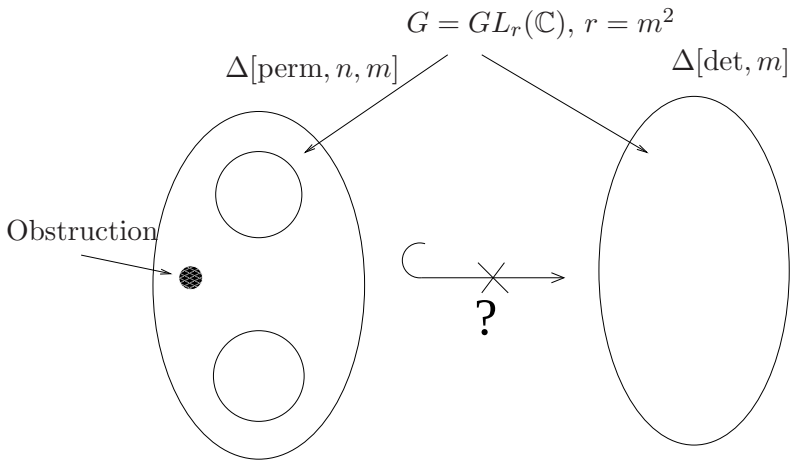

Figure 3: Class varieties

trix, $m > n$, then

$$\Delta[\mathrm{perm}, n, m] \subseteq \Delta[\det, m], \tag{6}$$

and conversely, if $\Delta[\mathrm{perm}, n, m] \subseteq \Delta[\det, m]$, then $f = z^{m-n}\mathrm{perm}(X)$ as a point in $P(V)$ can be approximated infinitesimally closely by a point in $P(V)$ of the form $\det(AY)$, $A \in G$, thinking of Y as an m^2-vector. The following conjecture is thus a stronger form of the arithmetic permanent vs. determinant conjecture in [V] over $\mathbb{C}$.

Conjecture 4.1 *(Strong arithmetic form of the permanent vs. determinant conjecture) [GCT1] The point $f \in P(V)$ cannot be approximated infinitesimally closely as above if $m = poly(n)$, and more generally, $m = 2^{\log^a n}$ for any constant $a > 0$.*

Equivalently, if $m = poly(n)$, or more generally, $m = 2^{\log^a n}$, $a > 0$ fixed, $n \to \infty$, then

$$\Delta[perm, n, m] \not\subseteq \Delta[\det, m]. \tag{7}$$

Geometric obstructions are meant to be proof certificates of (7). Intuitively, they are representation theoretic objects that live on $\Delta[perm, n, m]$ but not on $\Delta[\det, m]$ (for $m = \mathrm{poly}(n)$); cf. Figure 3. Their very existence then serves as a guarantee that the embedding as in (6) is not possible, because otherwise they would be living on $\Delta[\det, m]$ as well.

To define them formally, we need to recall some basic representation theory; cf. Appendix. By a classical result of Weyl [Fu], the irreducible (poly-

nomial) representations of $G = GL_r(\mathbb{C})$ are in one-to-one correspondence with the partitions λ of length at most r. By a partition $\lambda = (\lambda_1, \lambda_2, \ldots)$ we mean an integral sequence $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_k > 0$, $k \leq r$, where k is called the length of λ. The irreducible representation of G in correspondence with λ is denoted by $V_\lambda(G)$. It is called the *Weyl module* of G indexed by λ; cf. Appendix for its explicit description. Symbolically:

$$\text{Irreducible representations of } G \stackrel{\text{Weyl}}{\Longleftrightarrow} \quad \text{partitions } \lambda.$$

$$\text{Weyl module } V_\lambda(G) \longleftrightarrow \lambda.$$

Weyl also proved that every (polynomial) finite dimensional representation of G can be decomposed into irreducible representations. This means it can be written as a direct sum of Weyl modules. Thus Weyl modules are the basic building blocks of the representation theory of G, and every finite dimensional representation of G can be thought of as a complex building made out of these blocks.

Definition 4.2 *[GCT2] A* geometric obstruction $O_{n,m}$ *is a Weyl module* $V_\lambda(G)$ *that lives on* $\Delta[perm, n, m]$ *but not on* $\Delta[\det, m]$ *(Figure 3). Formally, this means* $V_\lambda(G)$ *occurs as a subrepresentation of* G *in the (dual of the) homogeneous coordinate ring* $R[perm, n, m]$ *of* $\Delta[perm, n, m]$ *but not in the (dual of the) homogeneous coordinate ring* $R[\det, m]$ *of* $\Delta[\det, m]$*. We call* λ *the* obstruction label.

A relaxed geometric obstruction is a Weyl module $V_\lambda(G)$ *that occurs as a subrepresentation of* G *in* $I[\det, m]/(I[\det, m] \cap I[perm, n, m])$*, where* $I[\det, m]$ *denotes the ideal of* $\Delta[\det, m]$ *and* $I[perm, n, m]$ *the ideal of* $\Delta[perm, n, m]$*. This is equivalent to saying that* $V_\lambda(G)$ *has a copy in* $I[\det, m]$ *that does not vanish identically at* $f = z^{m-n} perm(X)$*.*

Here the homogeneous coordinate ring of a projective variety $W \subseteq P(V)$ means the ring of all polynomial functions on its affine cone in V consisting of all lines in V corresponding to the points in W. By the ideal of W, we mean the ring of polynomial functions which vanish on it. See Appendix for further details.

It is easy to show that a geometric obstruction is also a relaxed geometric obstruction. Existence of a geometric obstruction $O_{n,m}$, for given n and m, implies that the inclusion (3) is not possible, since $O_{n,m}$ cannot live on

$\Delta[\det, m]$. Existence of a relaxed geometric obstruction also implies the same.

Thus:

Proposition 4.3 *[GCT2] Existence of a geometric obstruction (or a relaxed geometric obstruction) $O_{n,m}$, for given n and m, implies perm(X), with* $\dim(X) = n$, *cannot be represented linearly as* $\det(Y)$, *with* $\dim(Y) = m$.

Hence Conjecture 4.1 is implied by the following conjecture.

Hypothesis 4.4 (Geometric Obstruction Hypothesis (GOH)) *(cf. [GCT2]) A geometric obstruction exists if $m =$ poly(n), or more generally, if $m = 2^{\log^a n}$, $a > 0$ fixed, $n \to \infty$.*

See [GCT2] for supporting results and [GCT6] for evidence, based on the Strong Flip Theorem (Theorem 4.6) below, leading to conjectural *equivalence* between a relaxed form of this hypothesis and the strong permanent vs. determinant Conjecture 4.1. Also see [Bu] for computer-based numerical evidence for the analogous hypothesis for the lower bound problem for matrix multiplication.

4.3 A high level plan

A high level of plan of GCT in the context of the permanent vs. determinant problem can now be described as follows.

Let $G_g \subseteq G$ be the group of symmetries of $g = \det(Y) \in V$. Since $\det(Y)$ is characterized by its symmetries, it is completely determined as a point in V by the triple:

$$G_g \hookrightarrow G \hookrightarrow K = GL(V). \tag{8}$$

Thus the G-module structure of the homogeneous coordinate ring $R[\det, m]$ is also completely determined by this triple. Since algebraic groups are determined by their representations [DM], all information about the G-module structure of $R[\det, m]$ is thus contained in the representation theory of the group triple (8). Now:

1. Study the representation theory of the group triple (8) in depth and use it to understand the G-module structure of the homogeneous coordinate ring $R[\det, m]$. Understand the G-module structure of $R[f, n, m]$ similarly using the characterization of f by its symmetries.

2. Compare these G-module structures to locate a geometric obstruction, when $m = \text{poly}(n)$.

The goal is to carry out these steps *explicitly*, exploiting the characterization by symmetries of the permanent and determinant. We shall specify what explicit means later; cf. Hypothesis 4.8. This approach is extremely rigid in the sense that it only works for extremely rare hard functions that are characterized by their symmetries. This extreme rigidity is much more than what is needed to bypass the natural proof barrier [RR].

4.4 A quadratic lower bound

The following result provides a concrete lower bound application of a relaxed form of GCT in the context of the permanent vs. determinant problem.

Theorem 4.5 *(cf. Theorem 1.0.1 and 1.0.2 in [LMR]) A strongly explicit relaxed geometric obstruction $V_\lambda(G)$ exists when $m \leq n^2/2$. By strongly explicit, we mean that a relaxed obstruction label λ can be constructed in $O(n)$ time*[1]*.*

Thus $\Delta[perm, n, m] \not\subseteq \Delta[\det, m]$, when $m \leq n^2/2$.

This proves Conjecture 4.1 for $m \leq n^2/2$ and implies the earlier quadratic lower bound for the permanent in [MR].

4.5 Strong flip theorem

However, it is still far from the final goal because it does not address the self referential difficulty in the permanent vs. determinant problem. We have already remarked in Section 2.4 that this difficulty is not an issue in the context of the quadratic lower bound problem above since it is easier than derandomization. Explicit constructions to overcome this difficulty would have to be far more difficult. This is the content of the following stronger form of the flip theorem in this setting.

[1]This obstruction is the explicit Weyl module in the ideal $I[\det, m]$ described in Theorem 1.0.2 in [LMR]. It is not shown in this paper that this Weyl module does not vanish identically at $f = z^{m-n}\text{perm}(X)$. But this follows easily from the results in this paper and in [MR].

Theorem 4.6 (Strong Flip) *Suppose Conjecture 4.1 holds and that black box determinant identity testing [KI] can be derandomized in a stronger form as specified in [GCTflip] (cf. Section 8.1 therein). Then Conjecture 4.1 has an extremely explicit proof satisfying analogous F0-4 and an additional property (G) described below.*

In particular, this means there exists an obstruction family $\mathcal{O} = \cup_{n,m}\mathcal{O}_{n,m}$, with $\mathcal{O}_{n,m}$ nonempty when $m = \text{poly}(n)$, such that, for any obstruction label $s \in \mathcal{O}_{n,m}$, $m = \text{poly}(n)$, one can compute in $\text{poly}(n, m)$ time a global obstruction set $S_{n,m}(s) = \{X_1, \ldots, X_l\}$ of inputs, $l = \text{poly}(n, m)$, with the following property. Fix any homogeneous polynomial $p(Y)$ in V that belongs to $\Delta[\det, m]$ (thinking of a homogeneous polynomial in V, by an abuse of notation, as a point in $P(V)$). Let $p'(X)$ denote the polynomial obtained from $p(Y)$ by substituting zero for all variables in Y other than z and X, and 1 for z. Then there exists a counter example $X_i \in S_{n,m}(s)$ such that $p'(X_i) \neq \text{perm}(X_i)$. This specifies F1 (decoding) in this setting. Other flip properties are analogous; cf. GCTflip for their details.

Let $\psi = \psi_s : V \to \mathbb{C}^l$ be the homogeneous linear map that maps any homogeneous form $p(Y) \in V$ to the point $(p'(X_1,), \ldots, p'(X_l)) \in C^l$. Let $\hat{\psi} = \hat{\psi}_s$ denote the corresponding morphism from the projective variety $\Delta[\det, m]$ to the projective variety $P(\mathbb{C}^l)$. It is not defined when the tuple $(p'(X_1,), \ldots, p'(X_l))$ is identically zero. Its image is $\hat{\psi}(\Delta[\det, m]) \subseteq P(\mathbb{C}^l)$. It can be ensured that that $\psi(f) \in \mathbb{C}^l$, $f = z^{m-n}\text{perm}(X)$, is not an identically zero tuple. Hence it defines a point in $P(\mathbb{C}^l)$, which we denote by $\hat{\psi}(f)$. Then $S_{n,m}(s)$ is a global obstruction set iff $\hat{\psi}_s(f) \notin \hat{\psi}_s(\Delta[\det, m])$.

The property (G) mentioned above is that:

(G): The point $\hat{\psi}_s(f)$ does not belong to the projective closure of $\hat{\psi}_s(\Delta[\det, m]) \subseteq P(\mathbb{C}^l)$, when $m = \text{poly}(n)$.

This follows from F0-F4 and classical algebraic geometry. It is crucial because in GCT we are finally interested in constructing an obstruction by geometric techniques. If $\hat{\psi}_s(f)$ belongs to the closure of $\hat{\psi}_s(\Delta[\det, m])$, every polynomial function that vanishes on $\hat{\psi}_f(\Delta[\det, m])$ will also vanish on $\hat{\psi}_s(f)$. No algebro-geometric technique can construct such $\hat{\psi}_s$. The property (G) rules out this pathology and says that $\hat{\psi}_s$ is well behaved geometrically.

The linear map $\hat{\psi}_s$ above is called an *extremely explicit separator* between $\Delta[\det, m]$ and $f = z^{m-n}\text{perm}(X)$. It is called extremely explicit because (assuming the relevant hardness and derandomization conjectures) (1) given s, $S_{n,m}(s)$ can be computed in $O(\text{poly}(n, m))$ time by Theorem 4.6,

and (2) each coefficient of $\hat{\psi}_s$ in the standard basis of V [2] can also be computed in poly(n, m) time; this also easily follows from Theorem 4.6. We call $l = \text{poly}(n, m)$ the *dimension* of $\hat{\psi}_s$. Thus Theorem 4.6 says that, assuming the strong arithmetic permanent vs. determinant and derandomization conjectures, one can construct an extremely explicit family of linear separators of small dimension between $\Delta[\det, m]$ and f.

Theorem 4.6 critically depends on the exceptional nature of f and $g = \det(Y)$. It will almost never hold for general f and g in place of the permanent and determinant. For general f and g, a global obstruction set $S_{n,m}$ that gives a linear separator ψ between $\Delta[g, m] = \overline{Gg}$ and f can be constructed (if it exists) using general purpose algorithms for elimination theory in algebraic geometry for computing multivariate resultants and Gröbner bases of the ideals of algebraic varieties. But these algorithms take $\Omega(2^{\dim(V)})$ time. Since $\dim(V)$ is exponential in n and m, the time taken is at least double exponential in n and m. Nothing better can be expected for general f and g, because elimination theory is intractable in general. For example, the problem of computing the Gröbner basis is EXPSPACE-complete [MMr]. This means it takes in general space that is exponential in the dimension of the ambient space, which is $P(V)$ here. In contrast, Theorem 2.3 says that a short specification $S_{n,m}$ of an extremely explicit linear separator between $\Delta[\det, m]$ and $f = z^{m-n}\text{perm}(X)$ can be computed in poly(n, m) time exploiting the exceptional nature of the permanent and the determinant. This may seem impossible on the basis of the existing algebraic geometry.

At present, such extremely explicit separators of small dimension can be constructed in algebraic geometry only between very special kinds of algebraic varieties, such as the Grassmanian or the flag varieties [Fu], and very special kinds of points. This can be done using the second fundamental theorem of invariant theory [Fu], which gives a very nice explicit set of generators for the ideals of these varieties. But these varieties have very low complexity in comparison to $\Delta[\det, m]$. Specifically, their complexity according to a certain measure of complexity of orbit closures defined in [LV] is zero, whereas that of $\Delta[\det, m]$ is $O(m^2)$. The problem of explicit construction of linear separators when the complexity of the underlying variety is so high seems very formidable and far beyond the reach of the existing machinery in algebraic geometry.

[2] The standard basis representation of any form $f \in V = \mathbb{C}[Y]_m$ is given by its coefficients.

By Theorem 4.6, problems comparable in difficulty to the formidable explicit construction problems in algebraic geometry occurring in the conclusion of this theorem can be expected in *any* approach towards the fundamental hardness and derandomization conjectures in complexity theory. We call this the *law of conservation of difficulty*. It may explain why these conjectures, which seem so elementary at the surface, have turned out to be so hard.

4.6 Flip hypothesis

The Strong Flip Theorem 4.6 and Theorem 4.5 suggest the following strengthening of Hypothesis 4.4.

Hypothesis 4.7 (Flip) *(cf. [GCT6]) A strongly explicit geometric obstruction exists if $m = poly(n)$. By strongly explicit we mean that the partition λ specifying this obstruction $V_\lambda(G)$ can be constructed in $poly(n, m)$ time for any $m = poly(n)$. Furthermore, it can be assumed that the size $|\lambda| = \sum_i \lambda_i$ is $O(poly(n, m))$.*

The following is the more elaborate form of this hypothesis suggested by the Strong Flip Theorem 4.6.

Hypothesis 4.8 (Flip Hypothesis, FH) *(cf. [GCT6])*

FH[General]: *The family of geometric obstructions (cf. Definition 4.2) is extremely explicit, satisfying analogues of F0-F4 for global obstruction sets $S_{n,m}(s)$ in Theorem 4.6. In particular, this means:*

1. **Verification:** *Given n, m and λ, whether λ is a geometric obstruction label can be verified in $poly(n, m, \langle\lambda\rangle)$ time, where $\langle\lambda\rangle = \sum_i \log_2 \lambda_i$ denotes the bitlength of the specification of the partition $\lambda = (\lambda_1, \lambda_2, \ldots)$.*

2. **Construction:** *For given n and $m = poly(n)$, a geometric obstruction label (partition) λ can be constructed in $poly(n, m)$ time. Furthermore it can be assumed that the size $|\lambda|$ is $poly(n, m)$.*

3. **Decoding:** *Given a geometric obstruction label λ for given n and small m, a small global obstruction set $S_{n,m}(\lambda) = \{X_1, \ldots, X_l\}$, $l = poly(n)$, of inputs can be computed in $poly(n, m)$ time such that, for any form $p(Y) \in \Delta[\det, m]$, there exists X_i, $i \leq l$, such that $p'(X_i) \neq perm(X_i)$.*

FH[Determinant]:

1. **Verification:** *Given m and λ, whether $V_\lambda(G)$ lives on $\Delta[\det, m]$ (i.e., whether $V_\lambda(G)$ is a G-subrepresentation of the dual of the homogeneous coordinate ring $R[\det, m]$) can be decided in poly$(m, \langle\lambda\rangle)$ time.*

2. **Construction and decoding:** *Given m, one can compute in poly(m) time a label λ of poly(m) size such that $V_\lambda(G)$ does not live on $\Delta[\det, m]$. Furthermore, given $n \leq m$ and any such λ, one can compute in poly(m) time a small hitting set $H_{n,m,\lambda} = \{X_1, \ldots, X_l\}$, $l =$ poly(m), of inputs such that for any form $p(Y) \in \Delta[\det, m]$, with $p'(X)$ not identically zero, there exists X_i, $i \leq l$, such that $p'(X_i) \neq 0$.*

FH[Permanent] [Verification]: *Given n, m, and λ, whether $V_\lambda(G)$ lives on $\Delta[perm, n, m]$ can be decided in poly$(n, m, \langle\lambda\rangle)$ time.*

See [GCT6] for a detailed justification of these flip hypotheses based on the Strong Flip Theorem 4.6, which suggests that these hypotheses may be in essence *implications* of the hardness and derandomization conjectures in the statement of Theorem 4.6.

FH[Determinant] for construction and decoding implies derandomization [KI] of determinant identity testing. This provides a GCT approach to derandomization. There is an analogous hypothesis for the arithmetic form of the P vs. NP problem and derandomization [KI] of general polynomial identity testing; cf. [GCT6].

Like the Strong Flip Theorem 4.6, Hypothesis 4.8 crucially depends on the exceptional nature of the permanent and the determinant. For general functions, not characterized by their symmetries, this hypothesis will almost always fail. Hence the approach to implement the flip based on this hypothesis is extremely rigid.

FH[Determinant] does not have the self referential difficulty in the sense that (1) m is not required to be a small function of n in its statement, and (2) it only depends on the properties of the determinant, and not on the relationship between the permanent and the determinant (or equivalently, between the complexity classes $\#P$ and NC). The case of FH[Permanent] is similar. FH[Determinant](verification) and FH[Permanent] (verification) together imply FH[general] (verification), which says that geometric obstructions in GOH are easy to verify. We saw in Section 2.2 that the self referential difficulty is the main obstacle to ease of verification of obstructions (F3). Hence, once FH[Determinant](verification) and FH[Permanent]

(verification) are proved, GOH (Hypothesis 4.4) does not have the self referential difficulty any more. This decomposes the strong permanent vs. determinant Conjecture 4.1 into three subproblems without self referential difficulty, namely, FH[Determinant](verification), FH[Permanent] (verification), and GOH. Pictorially,

$$\text{Strong perm. vs. det.} \cong FH[Determinant](verification) + FH[Permanent](verification) + GOH. \quad (9)$$

This decomposition breaks the circle of self referential difficulty. Here the exceptional nature of geometric obstructions is crucial. For example, such a break is not possible using a relaxed geometric obstruction, in general, or the global obstruction set $S_{n,m}(s)$ in the strong flip Theorem 4.6.

The goal now is to prove FH[Determinant] and FH[permanent] for verification to get an efficient criterion for verifying a geometric obstruction as in FH[general](verification), and then use this criterion to guess and construct a geometric obstruction explicitly, when $m = \text{poly}(n)$.

4.7 How to prove FH?

We now describe the approach in [GCT6] to prove somewhat weaker forms of FH[determinant] (verification) and FH[permanent] (verification), assuming certain *positivity hypotheses* (cf. Section 4.7.3) in algebraic geometry and representation theory. This will lead to a more refined decomposition than (9) in terms of these positivity hypotheses.

We begin by recalling what is known about the analogue of FH and positivity in the context of the simplest and best understood multiplicities in representation theory–namely, the Littlewood-Richardson coefficients [Fu]. Then we describe what is needed to lift the Littlewood-Richardson story to general FH (Hypothesis 4.8).

4.7.1 Littlewood-Richardson coefficients

Given partitions α, β, and λ, the Littlewood-Richardson coefficient $c^{\lambda}_{\alpha,\beta}$ is the multiplicity (the number of copies) of the Weyl module $V_\lambda(G)$ in the tensor product $V_\alpha(G) \otimes V_\beta(G)$, considered as a G-module by letting G act on both factors independently. Thus $V_\alpha(G) \otimes V_\beta(G)$ has the following complete

decomposition (cf. Appendix) as a G-module:

$$V_\alpha(G) \otimes V_\beta(G) = \oplus_\lambda c^\lambda_{\alpha,\beta} V_\lambda(G).$$

Let $\tilde{c}^\lambda_{\alpha,\beta}(k) = c^{k\lambda}_{k\alpha,k\beta}$ be the associated stretching function [Rs]. The following results are known.

Polynomiality: The stretching function $\tilde{c}^\lambda_{\alpha,\beta}(k)$ is a polynomial in k [Rs].

Polyhedral LR (Littlewood-Richardson) rule]: There exists an explicit polytope (e.g. Hive polytope [KT]) $P = P^\lambda_{\alpha,\beta}$ such that (1) $c^\lambda_{\alpha,\beta}$ is equal to the number integer points in P, and more generally, (2) $\tilde{c}^\lambda_{\alpha,\beta}(k) = f_P(k)$, where $f_P(k)$ is the Ehrhart function of P, i.e., the number of integer points in the dilated polytope kP. It is a polynomial by the preceding polynomiality property. For general rational P, $f_P(k)$ is a quasipolynomial; cf. [St] and Section 4.7.2 below. By an explicit polytope, we mean, given a rational point x, whether $x \in P^\lambda_{\alpha,\beta}$ can be decided in poly$(\langle x\rangle, \langle\alpha\rangle, \langle\beta\rangle, \langle\lambda\rangle)$ time, where $\langle x\rangle$ denotes the bitlength of x.

The preceding polyhedral LR rule is basically a consequence of the classical Littlewood-Richardson rule. See [Fu2] for its full description and an elementary proof, and [GCT3] for a full elementary description of the polyhedral LR rule. This rule implies a $\#P$-formula for $c^\lambda_{\alpha,\beta}$.

LR Saturation Theorem [KT]: The polynomial $\tilde{c}^\lambda_{\alpha,\beta}(k)$ is saturated. This means if $\tilde{c}^\lambda_{\alpha,\beta}(k)$ is nonzero for some $k \geq 1$, then $c^\lambda_{\alpha,\beta}$ is also nonzero.

LR saturation theorem is a consequence of the following conjecture in [KTT] supported by substantial experimental evidence. We refer to it as the second positivity hypothesis PH2.

LR PH2 [KTT]: All coefficients of the polynomial $\tilde{c}^\lambda_{\alpha,\beta}(k)$ are nonnegative.

Nonvanishing of LR coefficients: The problem of deciding nonvanishing of Littlewood-Richardson coefficients belongs to the complexity class P (cf. [GCT3] and [KT2]). That is, given α, β, and λ, whether $c^\lambda_{\alpha,\beta}$ is nonzero can be decided in poly$(\langle\alpha\rangle, \langle\beta\rangle, \langle\lambda\rangle)$ time. This easily follows from the polyhedral LR rule, LR saturation theorem, and a polynomial time algorithm [GLS] for linear programming. It proves the analogue of FH[Determinant and Permanent][verification] (Hypothesis 4.8) for Littlewood-Richardson coefficients.

4.7.2 Definitions

Now we wish to lift this Littlewood-Richardson story to general FH (Hypothesis 4.8). Towards that end, we need some definitions.

Let $F_{\lambda,n,m}(k)$ denote the number of copies of the Weyl module $V_{k\lambda}(G)$ that live on $\Delta[\mathrm{perm}, n, m]$. Here $k\lambda$ denotes the partition obtained by multiplying each number in the integral sequence (partition) λ by k. Let $G_{\lambda,m}(k)$ denote the number of copies of the Weyl module $V_{k\lambda}(G)$ that live on $\Delta[\det, m]$. Thus $V_\lambda(G)$ is a geometric obstruction (cf. Definition 4.2) iff $F(\lambda, n, m) = F_{\lambda,n,m}(1)$ is nonzero and $G(\lambda, m) = G_{\lambda,m}(1)$ is zero.

Given a polytope Q, let $f_Q(k)$ denote the number of integer points in the dilated polytope kQ. More generally, given a parametrized polytope $P(k)$ defined by a linear system of the form:

$$P(k): \qquad Ax \leq kb + c, \tag{10}$$

where A is an $s \times t$ integer matrix, x a variable t-vector, and b and c some integral constant s-vectors, let $f_P(k)$ denote the number of integer points in $P(k)$. It is a classical result of Ehrhart [St] that $f_P(k)$ is asymptotically a quasi-polynomial for general c and a quasipolynomial when $c = 0$. We call it the (asymptotic) Ehrhart quasipolynomial of the polytope $P(k)$. Here we call a function $f(k)$ an asymptotic quasipolynomial if there exist a nonnegative integer $a(f)$, which we call the asymptotic defect of $f(k)$, a positive integer l, which we call the period of $f(k)$, and polynomials $f_1(k), \ldots, f_l(k)$ such that, for every integer $k > a(f)$, $f(k) = f_i(k)$ if $k = i$ modulo l. We call $f(k)$ a quasipolynomial if $a(f) = 0$.

We define the *positivity index* $p(f_P)$ of $f(k) = f_P(k)$ to be the smallest nonnegative integer such that: (1) for every j, the coefficients of $f_j(k+p(f_P))$ are all nonnegative, and (2) $p(f_P) \geq \min\{a(f_P), b(f_P)\}$, where $b(f_P) = \min\{k+1 | P(k) \neq \emptyset\}$; it is $-\infty$ if $P(k)$ is always nonempty. We define the *saturation index* $s(f_P)$ of $f(k) = f_P(k)$ to be the smallest nonnegative integer such that (1) for any j, $f_j(k) > 0$ for every $k \geq s(f_P)$, $k = j \bmod l$, whenever $f_j(k)$ is not an identically zero polynomial, and (2) $s(f_P) \geq \min\{a(f_P), b(f_P)\}$. Clearly, $s(f_P) \geq a(f_P)$.

4.7.3 Positivity hypotheses

In this section, we describe the positivity hypotheses in algebraic geometry and representation theory which provide an approach to prove FH.

The following result generalizes in a relaxed form the polynomiality property of the stretching function associated with LR-coefficients (cf. Section 4.7.1).

Lemma 4.9 *[GCT6] The functions $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ are asymptotic quasi-polynomials.*

This follows from the classical work of Hilbert.

The following positivity hypothesis says that the asymptotic quasipolynomials in Lemma 4.9 can be realized as asymptotic Ehrhart quasi-polynomials of *explicit* polytopes. This generalizes the polyhedral LR rule (cf. Section 4.7.1).

Hypothesis 4.10 (PH1) [Positivity Hypothesis] *(cf. [GCT6])*
(a) For every $\lambda, n, m \geq n$, there exists an explicit parametrized polytope $P(k) = P_{\lambda,n,m}(k)$ such that

$$F_{\lambda,n,m}(k) = f_P(k). \tag{11}$$

If such a polytope exists it is guaranteed by the proof of Lemma 4.9 that its dimension is poly(n) regardless of what m is. By explicit we mean the polytope is given by a separation oracle [GLS] that, given any rational point x, decides if $x \in P(k)$ and gives a separating hyperplane if it does not in poly$(n, m, \langle x\rangle, \langle k\rangle, \langle\lambda\rangle)$ time, where $\langle\ \rangle$ denotes the bitlength of specification.
(b) For every m and λ, there exists an explicit parametrized polytope $Q(k) = Q_{\lambda,m}(k)$ such that

$$G_{\lambda,m}(k) = f_Q(k). \tag{12}$$

If such a polytope exists it is guaranteed by the proof of Lemma 4.9 that its dimension is poly(n) regardless of what m is as long as the length of λ is poly(n) (as it will be in our applications). Explicitness is defined similarly.

The following hypothesis says that $G_{\lambda,m}(k)$ and $F_{\lambda,n,m}(k)$ have a saturation property that generalizes in a relaxed form the LR saturation theorem (cf. Section 4.7.1).

Hypothesis 4.11 (Saturation Hypothesis (SH)) *The saturation indices of $G_{\lambda,m}(k)$ and $F_{\lambda,n,m}(k)$ are poly$(n, m, \langle\lambda\rangle)$.*

This follows from the following generalization in a relaxed form of LR PH2 (cf. Section 4.7.1).

Hypothesis 4.12 (Positivity Hypothesis (PH2)) *The positivity indices of $G_{\lambda,m}(k)$ and $F_{\lambda,n,m}(k)$ are poly$(n, m, \langle\lambda\rangle)$.*

See [GCT6] for a detailed justification of PH1 and SH based on the Strong Flip Theorem 4.6, which suggests that these positivity hypotheses may be in essence *implications* of the hardness and derandomization conjectures in the statement of Theorem 4.6.

PH1, in particular, implies that $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ have $\#P$ formulae. Positivity refers to the positive form of a $\#P$-formula, i.e., the absence of any negative sign as in the usual formula for the permanent. PH1,SH, and PH2 critically depend on the fact that the permanent and determinant are characterized by their symmetries (cf. Section 4.1). If we replace these functions with general functions without symmetries, these hypotheses would almost certainly fail (just like the Strong Flip Theorem 4.6, which is the key ingradient in the justification of these hypotheses in [GCT6]).

These positivity hypotheses are fundamentally different from the original hardness hypothesis (the permanent vs. determinant conjecture) because the self-referential difficulty is absent in them for two reasons: (1) m is not required to be a small function of n in their statements, and (2) they do not depend on the relationship between the permanent and the determinant (or equivalently, between the complexity classes $\#P$ and NC). This is because PH1 and SH (PH2) for the class variety $\Delta[\text{perm}, n, m]$ are statements only about the properties of the permanent and do not depend in any way on the determinant or the complexity class NC, and similarly PH1 and SH (PH2) for the class variety $\Delta[\det, m]$ are statements only about the properties of the determinant and do not depend in any way on the permanent or the complexity class $\#P$.

The following result proves weaker forms of FH[Determinant] (verification) and FH [Permanent] (verification) assuming these positivity hypotheses.

Theorem 4.13 *[GCT6] Assume that PH1 and SH for $F_{\lambda,n,m}(k)$ (Hypotheses 4.10,4.11) hold. Then given λ, n, m, and k' greater than the saturation index of $F_{\lambda,n,m}(k)$, whether $F_{\lambda,n,m}(k')$ is nonzero can be decided in poly$(\langle\lambda\rangle, n, m, \langle k'\rangle)$ time. Similar result holds for $G_{\lambda,m}(k)$ assuming PH1 and SH for $G_{\lambda,m}(k)$.*

This follows from the following result that provides a polynomial time algorithm for a special case of integer programming, called *saturated integer programming*. In contrast, integer programming is NP-complete in general.

Theorem 4.14 *(cf. [GCT6]) Let $P(k)$ be a parametrized polytope specified as a separation oracle [GLS] with bitlength of specifcation $\langle P(k)\rangle$. Let k' be a nonnegative integer guaranteed to be greater than the saturation index $s(F_p)$ (cf. Section 4.7.2) of the asymptotic Ehrhart quasipolynomial $f_P(k)$ of P. Then whether $P(k')$ contains an integer point can be decided in poly$(\langle P(k')\rangle, \langle k'\rangle)$ time.*

This implies Theorem 4.13 since $F_{\lambda,n,m}(k')$ is nonzero iff the polytope $P_{\lambda,n,m}(k')$ in PH1 contains an integer point.

4.8 Decomposition theorem

Now we describe how these positivity hypotheses can be used for decomposing the permanent vs. determinant problem, the story for the arithmetic P vs. NP problem being similar.

Theorem 4.15 (Decomposition) *(cf. [GCT6])*
(A): *There exists an* explicit *family $\mathcal{O} = \cup_{n,m}\{O_{n,m}\}$ of geometric obstruction labels for the permanent vs. determinant problem for $m = 2^{\log^a n}$, $a > 1$ fixed, $n \to \infty$, assuming,*

1. *PH1, and*
2. *OH (Obstruction Hypothesis):*

   For all $n \to \infty$, there exists λ, and k of poly$(n, m, \langle\lambda\rangle)$ bitlength greater than the saturation index of $F_{\lambda,n,m}(k)$ (polynomially bounded in SH), such that:

   (a) *$P_{\lambda,n,m}(k) \neq \emptyset$ and the affine span of $P_{\lambda,n,m}(k)$ contains an integer point. Here by an affine span we mean the smallest dimensional affine space containing the polytope.*

   (b) *The affine span of $Q_{\lambda,m}(k)$ does not contain an integer point.*

In this case, $V_{k\lambda}(G)$ is a geometric obstruction for given n and m. The set $\mathcal{O}_{n,m}$ consists of the specification labels (k, λ) of such $V_{k\lambda}(G)$ for given n amd m. Explicitness of $\mathcal{O}$ means (cf. Section 2.1) the obstruction-labels (k, λ) are easy to verify: specifically, given n, m, and (k, λ), whether OH is satisfied can be checked in poly$(n, m, \langle\lambda\rangle, \langle k\rangle)$ time.

(B): *Analogous result holds for the arithmetic P vs. NP problem (Section 2.3), letting a class variety for NP based on $E(X)$ play the role of $\Delta[perm, n, m]$, and a class variety for P based on a function $H(X)$ defined in [GCT1] play the role of $\Delta[\det, m]$.*

This follows from the proof of Theorem 4.14. See [GCT6] for justification of OH based on the strong flip Theorem 4.6. Though SH is not explicitly needed in the statement of this result, it is crucial for OH to hold; cf. [GCT6].

We have already seen that the self-referential difficulty is absent in PH1 and SH (cf. Section 4.7.3). Once these positivity hypotheses are proved, there is no self referential difficulty in OH too, because it then becomes "easy to verify". Specifically, by Theorem 4.15, the problem of verifying whether a given obstruction label (k, λ) satisfies the condition in OH then belongs to the complexity class P. Recall (Section 2.2) that the self referential difficulty is the main obstacle to ease of verification of obstructions (F3). This is resolved once PH1 and SH are proved. Thus Theorem 4.15 decomposes the strong permanent vs. determinant Conjecture 4.1 into the positivity hypotheses PH1 and SH, and an easier hardness hypothesis OH, all without the self referential difficulty. Pictorially,

$$\text{Strong perm. vs. det. problem} \cong PH1 + SH + OH. \qquad (13)$$

This decomposition breaking the circle of self referential difficulty is more refined than (9). The problems on the right hand side here are polyhedral and linearized, and hence, simpler than the ones on the right hand side of (9). The decomposition for the arithmetic P vs. NP problem is similar.

The goal now is to prove PH1 and SH, and use the explicit polytopes $P(k) = P_{\lambda,n,m}(k)$ and $Q(k) = Q_{\lambda,m}(k)$ in PH1 to find a geometric obstruction satisfying OH explicitly. For this $P(k)$ and $Q(k)$ need to be explicit not just in a theoretical sense but also in a practical sense so that they can be used to prove OH. In this strategy, an explicit proof–specifically, F3 for verification–is essentially forced on us. We need to know the polytopes $P(k)$ and $Q(k)$ in PH1 explicitly to prove OH, and once this is done, Theorem 4.15 gives a polynomial time criterion for verifying obstructions in $\mathcal{O}$ whether we care for explicitness or not.

4.9 How to prove positivity?

The multiplicities $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ in PH1 are akin to but much harder than the so-called plethysm constants in representation theory [Fu]. The classical plethysm problem in algebraic geometry and representation theory asks for positive formulae (without any alternating signs) for the plethysm constants, but without raising any complexity theoretic issues, such as explicitness. As such, the positivity hypothesis PH1 encompasses and goes far beyond this plethysm problem. Its statement brings together algebraic geometry, representation theory, and complexity theory, the last because explicitness is its crucial ingradient. The articles [GCT4, GCT7, GCT8] suggest an approach to prove PH1 via generalizations of quantum groups [Dr] called nonstandard quantum groups; we refer the reader to these articles for this story.

4.10 What is P?

The modest lower bounds in Theorems 3.1 and 4.5 are separated from the strong lower bound problems harder than derandomization (such as the permanent vs. determinant or the arithmetic P vs. NP problems) by the circle of self referential difficulty; cf. Figure 4. To break into this circle, we have to show that P contains formidable explicit construction problems in algebraic geometry and representation theory, such as the ones that arise in the strong flip Theorem 4.6 or the various flip and positivity hypotheses in Sections 4.6 and 4.7.3. By the law of conservation of difficulty (cf. Section 4.5) based on the Strong Flip Theorem 4.6, comparable understanding of P would be needed in *any* approach. Unfortunately, our current understanding of P is too modest. Until we understand P and the theory of upper bounds at the required strong level, we may not expect any further lower bounds that are fundamentally different from the known modest bounds.

5 Summary

We have outlined in this article the GCT approach to resolve the self referential paradox in the context of the arithmetic P vs. NP and related problems. GCT has revealed formidable explicit construction and positivity problems at the crossroads of algebraic geometry, representation theory, and complexity theory hidden underneath the fundamental hardness and derandomization conjectures. It is hoped that these three fields will now come

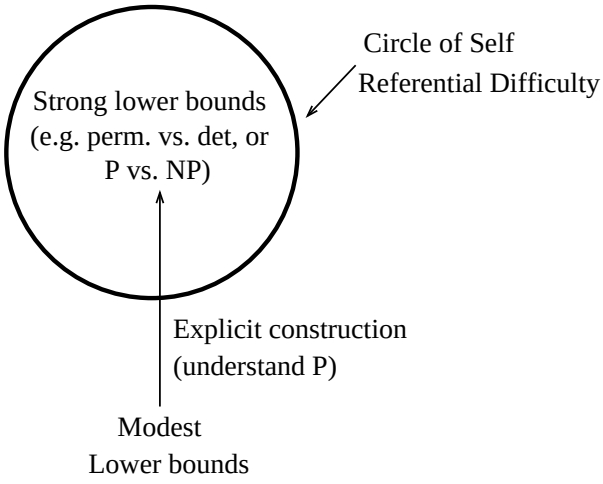

Figure 4: What is P?

together to work on this approach towards the P vs. NP problem.

References

[AW] S. Aaronson, A. Wigderson, Algebrization: a new barrier in complexity theory, ACM transactions on computing theory, 1(1), 2009.

[Ag] M. Agrawal, Proving lower bounds via pseudo-random generators, Proceedings of the FSTTCS, pages 92-105, 2005.

[At] A. Atserias, Non-uniform hardness for NP via black-box adversaries, ECCC, report 154 (2005).

[BS] R. Boppana, M. Sipser: The complexity of finite functions, Handbook of Theoretical Computer Science, vol. A, Edited by J. van Leeuwen, North Holland, Amsterdam, 1990, 757–804.

[Bu] P. Bürgisser, C. Ikenmeyer, Geometric complexity theory and tensor rank, arXiv:1011.1350, November, 2010.

[BLMW] P. Bürgisser, J. Landsberg, L. Manivel, J. Weyman, An overview of mathematical issues arising in the geomet-

ric complexity theory approach to $VP \neq VNP$, arXiv: 0907.2850v1 [cs.CC], July 2009.

[C] S. Cook: The complexity of theorem-proving procedures. Proceedings of the third annual ACM Symposium on Theory of Computing. 151-158. (1971).

[DM] P. Deligne, J. Milne, Tannakian categories, Hodge cycles, motives and Shimura varieties, Lecture notes in mathematics, 1981, vol. 900, pp. 101-228.

[Dr] V. Drinfeld, Quantum groups, Proc. Int. Congr. Math. Berkeley, 1986, vol. 1, Amer. Math. Soc. 1988, 798-820.

[FPS] L. Fortnow, A. Pavan, S. Sengupta, Proving SAT does not have small circuits with an application to the two queries problem, JCSS 74 (2008), 358-363.

[Fr] G. Frobenius, Uber die Darstellung der endlichen Gruppen durch lineare Substitutionen, Sitzungsber Deutsch. Akad. Wiss. Berlin (1897), 994-1015.

[Fu] W. Fulton, J. Harris, Representation theory, A first course, Springer, 1991.

[Fu2] W. Fulton, Young tableau, Cambridge university press, 1997.

[GCTpram] K. Mulmuley: Lower bounds in a parallel model without bit operations, The SIAM Journal On Computing, vol. 28, no. 4, 1999, pages 1460-1509.

[GCTabs] K. Mulmuley, M. Sohoni, Geometric complexity theory, P vs. NP and explicit obstructions, in "Advances in Algebra and Geometry", Edited by C. Musili, the proceedings of the International Conference on Algebra and Geometry, Hyderabad, 2001.

[GCTtutorial] K. Mulmuley, Geometric complexity theory, Video of FOCS 2010 tutorial available at: http://techtalks.tv/focs/2010.

[GCTflip] K. Mulmuley, Explicit proofs and the flip, arXiv:1009.0246, September 2010.

[GCT1] K. Mulmuley, M. Sohoni, Geometric complexity theory I: an approach to the P vs. NP and related problems, SIAM J. Comput., vol 31, no 2, pp 496-526, 2001.

[GCT2] K. Mulmuley, M. Sohoni, Geometric complexity theory II: towards explicit obstructions for embeddings among class varieties, SIAM J. Comput., Vol. 38, Issue 3, June 2008, pp 1175-1206.

[GCT3] K. Mulmuley, M. Sohoni, Geometric complexity theory III: on deciding positivity of Littlewood-Richardson coefficients, cs. ArXiv preprint cs. CC/0501076 v1 26 Jan 2005.

[GCT4] K. Mulmuley, M. Sohoni, Geometric complexity theory IV: quantum group for the Kronecker problem, cs. ArXiv preprint cs. CC/0703110, March, 2007.

[GCT5] K. Mulmuley, H. Narayanan, Geometric complexity theory V: on deciding nonvanishing of a generalized Littlewood-Richardson coefficient, Technical Report TR-2007-05, computer science department, The University of Chicago, May, 2007. Available at: http://ramakrishnadas.cs.uchicago.edu

[GCT6] K. Mulmuley, Geometric complexity theory VI: The flip via positivity, technical report, computer science department, the university of Chicago, July 2010. Revised version under preparation. Available at: http://ramakrishnadas.cs.uchicago.edu.

[GCT7] K. Mulmuley, Geometric complexity theory VII: Nonstandard quantum group for the plethysm problem, Technical Report TR-2007-14, computer science department, The University of Chicago, September, 2007. Available at: http://ramakrishnadas.cs.uchicago.edu.

[GCT8] K. Mulmuley, Geometric complexity theory VIII: On canonical bases for the nonstandard quantum groups, Technical Report TR 2007-15, computer science department, The university of Chicago, September 2007. Available at: http://ramakrishnadas.cs.uchicago.edu.

[GLS] M. Grötschel, L. Lovász, A. Schrijver, Geometric algorithms and combinatorial optimization, Springer-Verlag, 1993.

[Gu] L. Gurvits, On the complexity of mixed determinants and related problems, pp. 447-458, Lecture notes in computer science, Springer Verlag, September, 2005.

[IW] R. Impagliazzo, A. Wigderson, $P = BPP$ unless E has sub-exponential circuits: Derandomizing the XOR lemma, Proceedings of the 29th STOC, 1997.

[KI] V. Kabanets, R. Impagliazzo, Derandomizing polynomial identity tests means proving circuit lower bounds, Computational Complexity, 13(1-2), pages 1-46, 2004.

[Kp] R. Karp: Reducibility among combinatorial problems. R. E. Miller and J. W. Thatcher (eds.) Complexity of computer computations, Plenum Press, New York, 1972, 85-103.

[KV] R. Karp, V. Ramachandran, Parallel algorithms for shared memory machines, in J. van Leeuwen (ed.), Handbook of theoretical computer science (vol. A), Elsevier, 1990, pp. 869-941.

[KTT] R. King, C. Tollu, F. Toumazet Stretched Littlewood-Richardson coefficients and Kostka coefficients. In, Winternitz, P., Harnard, J., Lam, C.S. and Patera, J. (eds.) Symmetry in Physics: In Memory of Robert T. Sharp. Providence, USA, AMS OUP, 99-112., CRM Proceedings and Lecture Notes 34, 2004.

[KT] A. Knutson, T. Tao: The honeycomb model of $GL_n(C)$ tensor products I: proof of the saturation conjecture, J. Amer. Math. Soc. 12 (1999) 1055-1090.

[KT2] A. Knutson, T. Tao, Honeycombs and sums of Hermitian matrices, Notices Amer. Math. Soc. 48, (2001) No. 2, 175-186.

[Ku] S. Kumar, Geometry of orbits of permanents and determinants, arXiv:1007.1695, July 2010.

[LMR] J. Landsberg, L. Manivel, N. Ressayre, Hypersurfaces with degenerate duals and the Geometric Complexity Theory Program, arXiv:1004.4802, April, 2010.

[Le] A. Levin: Universal sequential search problems. Problems of information transmission (translated from Problemy Peredachi Informatsii (Russian)) 9 (1973).

[LV] D. Luna, Th. Vust: Plongements d'espaces homogenes. Comment. Math. Helv. 58, 186 (1983).

[MM] M. Marcus, F. May, The permanent function, Canad. J. math., 14 (1962), 177-189.

[MMr] E. Mayr, and A. Meyer, The complexity of the word problems for commutative semigroups and polynomial ideals, Advances in mathematics, 46 (3): 305-329, 1982.

[MR] T. Mignon, N. Ressayre, A quadratic bound for the determinant and permanent problem, International Mathematics Research Notices (2004) 2004: 4241-4253.

[Mu] D. Mumford, Algebraic geometry I: complex projective varieties, Springer, 1976.

[MFK] D. Mumford, J. Fogarty, F. Kirwan: Geometric invariant theory. Springer-Verlag, 1994.

[NW] N. Nisan, A. Wigderson, Hardness vs. randomness, J. Comput. Sys. Sci., 49 (2): 149-167, 1994.

[Rs] E. Rassart, A polynomiality property for Littlewood-Richardson coefficients, arXiv:math.CO/0308101, 16 Aug. 2003.

[Rz] A. Razborov, Lower bounds on the monotone complexity of some boolean functions, Doklady Akademii Nauk SSSR 281 (1985), 798-801.

[RR] A. Razborov, S. Rudich, Natural proofs, J. Comput. System Sci., 55 (1997), pp. 24-35.

[St] R. Stanley, Enumerative combinatorics, vol. 1, Wadsworth and Brooks/Cole, Advanced Books and Software, 1986.

[V] L. Valiant: The complexity of computing the permanent. Theoretical Computer Science 8 , 189-201 (1979).

Appendix: Basic representation theory and algebraic geometry

In this appendix we review the basic notions of representation theory and algebraic geometry used in this paper; cf. [Fu, Fu2, Mu] for more details.

Basic representation theory

Let G be a group. By a representation of G, we mean a vector space W with a homomorphism from G to $GL(W)$, the space of invertible linear transformations of W. It is called irreducible if it contains no nontrivial proper subrepresentation. We say that G is *reductive* if every finite dimensional [3] representation of G is completely reducible; i.e., can be written as a direct sum of irreducible representations.

All finite groups are reductive–a classical fact [Fu]. Weyl proved [Fu] that $G = GL_n(\mathbb{C})$, the general linear group of invertible $n \times n$ matrices, is reductive, so also $SL_n(\mathbb{C})$, the special linear group of invertible $n \times n$ matrices with determinant one.

This means every finite dimensional representation W of G can be written as a direct sum:

$$W = \oplus_i m_i W_i, \tag{14}$$

where W_i ranges over all finite dimensional irreducible representations of G and m_i denotes the multiplicity of W_i in W. Thus the irreducible representations are the building blocks of any finite dimensional representation.

Weyl also classified these building blocks. Specifically, he showed that the (polynomial[4]) irreducible representations of G are in one-to-one correspondence with the partitions (integral sequences) $\lambda : \lambda_1 \geq \lambda_2 \geq \lambda_k > 0$ of length $k \leq n$; we denote this partition by $\lambda = (\lambda_1, \ldots, \lambda_k)$. It can be pictorially depicted by the corresponding Young diagram consisting of λ_i boxes in the i-th row (Figure 5). An irreducible representation of G in correspondence with a partition λ is denoted by $V_\lambda(G)$, and is called a Weyl module.

If $\lambda = (r)$, i.e., when the Young diagram consists of just one row of r boxes, then $V_\lambda(G)$ is simply the space $\mathrm{Sym}^r(X)$ of all homogeneous forms

[3] There are some technical restrictions on what types of finite dimensional representations can be considered here (e.g. rational), which we ignore here.

[4] We say that a representation $\rho : G \to GL(W)$ is polynomial if the entries of $\rho(g)$, $g \in G$, are polynomial functions of the entries of g.

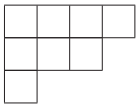

Figure 5: A Young diagram for the partition $(4, 3, 1)$

of degree r in the variables $x_1, \ldots, x_n$ with the following action of G. Given $f(X) \in \mathrm{Sym}^r(X)$ and $\sigma \in G$, map $f(X)$ to

$$f^\sigma(X) = f(X\sigma), \tag{15}$$

thinking of $X = (x_1, \ldots, x_n)$ as a row vector. This construction can be generalized to arbitrary λ as follows.

Let Z be an $n \times n$ variable matrix. Let $\mathbb{C}[Z]$ be the ring of polynomials in the entries of Z. It is a representation of $GL_n(\mathbb{C})$. Action of a matrix $\sigma \in GL_n(\mathbb{C})$ on a polynomial $f \in \mathbb{C}[Z]$ is given by

$$(\sigma \cdot f)(Z) = f(Z\sigma). \tag{16}$$

By a numbering (filling), we mean filling of the boxes of a Young diagram by numbers in $[n]$. We call such a numbering a *(semistandard) tableau* if the numbers are strictly increasing in each column and weakly increasing in all rows.

The partition corresponding to the Young diagram of a numbering is called the *shape* of the numbering.

With every numbering T, we associate a polynomial $e_T \in \mathbb{C}[Z]$, which is a product of minors for each column of T. The $l \times l$ minor e_c for a column c of length l is formed by the first l rows of Z and the columns indexed by the entries c_j, $1 \leq j \leq l$, of c. Thus $e_T = \prod_c e_c$, where c ranges over all columns in T. The Weyl module V_λ is the subrepresentation of $\mathbb{C}[Z]$ spanned by e_T, where T ranges over all numberings of shape λ over $[n]$. Its one possible basis is given by $\{e_T\}$, where T ranges over semistandard tableau of shape λ over $[n]$.

Basic algebraic geometry

Let $V = \mathbb{C}^m$, $P(V)$ the associated projective space consisting of lines in V through the origin, $\mathbb{C}[V]$ the coordinate ring of V, which can also be

thought of as the homogeneous coordinate ring of $P(V)$. Let $x_1, \ldots, x_m$ be the coordinates of V. A projective algebraic variety Y in $P(V)$ is defined to be the zero set of a set of homogeneous forms in $x_1, \ldots, x_m$ (it is also assumed that this zero set is irreducible; i.e., cannot be written as the union of two similar nonempty zero sets). The ideal $I(Y)$ of Y is defined to be the space of all forms in $\mathbb{C}[V]$ that vanish on Y. The homogeneous coordinate ring $R(Y)$ of Y is defined to be $\mathbb{C}[V]/I(Y)$.

Basic geometric invariant theory

Now let V be a finite dimensional representation of $G = GL_n(\mathbb{C})$. Then $\mathbb{C}[V]$ is a G-module (i.e., a representation) with the action that, for any $\sigma \in G$, maps $f(v) \in \mathbb{C}[V]$ to

$$f^\sigma(v) = f(\sigma^{-1}v). \tag{17}$$

(This is dual of the action in (15)). Here $\sigma^{-1}v$ denotes $\rho(\sigma^{-1})(v)$, where $\rho : G \to GL(V)$ is the representation map.

Definition 5.1 *A projective variety $Y \subseteq P(V)$ is called a G-variety if the ideal $I(Y)$ is a G-submodule (i.e., a G-subrepresentation) of $\mathbb{C}[V]$.*

This means, under the action of G, the points of Y are moved to the points within Y, i.e., each $\sigma \in G$ induces an automorphism of Y.

Let $v \in P(V)$ be a point, and Gv the orbit of v:

$$Gv = \{gv \mid g \in G\}. \tag{18}$$

The orbit closure of v is:

$$\Delta_V[v] = \overline{Gv} \subseteq P(V).$$

The closure is taken in the complex topology on $P(V)$ by adding all limit points of the orbit.

Basic fact of algebraic geometry: $\Delta_V[v]$ is a projective G-variety.

The algebraic geometry of the orbit closure $\Delta_V[v]$ for general v is hopeless. It can be tractable only if v is exceptional. The class varieties in GCT (cf. Section 4.2) are orbit closures of exceptional functions–namely the determinant and the permanent.