Geometric Complexity Theory VI: The flip via positivity

Dedicated to Sri Ramakrishna

Ketan D. Mulmuley *
The University of Chicago

January 31, 2011

Abstract

Geometric complexity theory (GCT) is an approach towards the P vs. NP and related problems. The article [GCTflip] describes its defining strategy, called the *flip*, to resolve the *self referential paradox*, which is the root difficulty in these problems. This article describes an approach based on *positivity hypotheses* in algebraic geometry and representation theory to implement the flip and thereby resolve the self referential paradox in the arithmetic setting wherein the underlying field of computation has characteristic zero. The main result here is the Decomposition Theorem that decomposes the arithmetic P vs. NP and permanent vs. determinant problems into such positivity hypotheses and easier hardness hypotheses, all without the self referential difficulty.

1 Introduction

This article belongs to a series [GCT1]-[GCT8] of articles on the geometric complexity theory (GCT) approach towards the P vs. NP and related problems. Intuitively, the P vs. NP problem is formidable 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 in this problem as per the Flip Theorem in [GCTflip], which *formalizes* this intuitive difficulty. As such, the main conceptual difficulty in any approach towards this problem is to *break the circle* of self reference around it by decomposing the problem and its variants into subproblems without the self referential difficulty. The main result of this article, the Decomposition Theorem 4.9, is such decomposition of the arithmetic P vs. NP and permanent vs. determinant problems based on *positivity hypotheses* in algebraic geometry and representation theory. This result was announced in the GCT overview [GCTexpo].

We now give a brief overview of this result focussing on the arithmetic permanent vs. determinant problem [V], since this illustrates all the basic ideas. In the arithmetic setting the underlying field of computation has characteristic zero. The problem is to show that $\mathrm{perm}(X)$,

*Supported by NSF grant CCF-1017760.

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 $m = \text{poly}(n)$, or more generally, $m = 2^{\log^a n}$, for a fixed constant $a > 0$, and $n \to \infty$. By a linear representation, we mean that the entries of Y are (possibly nonhomogeneous) linear functions of the entries of X. The best known lower bound on m at present is quadratic [MR2].

In [GCT1] and [GCT2], this problem is reduced to the problem of showing existence of *geometric obstructions*, which are representation theoretic objects in geometric invariant theory that serve as proof certificates of hardness of the permanent. Specifically, the geometric obstructions are irreducible polynomial representations (Weyl modules) $V_\lambda(G)$ of $G = GL_l(\mathbb{C})$, $l = m^2$, that occur in the homogeneous coordinate ring $R_{n,m}$ of a certain projective G-variety $Y_{n,m}$ associated with the permanent but not in the homogeneous coordinate ring S_m of another projective G-variety X_m associated with the determinant. Here m is small as above, and $V_\lambda(G)$ denotes the Weyl module of G labelled by the partition λ. The definitions of $Y_{n,m}$ and X_m are given in Section 2. It is conjectured that the problem of proving existence of these geometric obstructions is equivalent to a stronger form of the permanent vs. determinant conjecture (Conjecture 2.6). Thus [GCT1] and [GCT2] basically reformulates the original hardness problem in the setting of geometric invariant theory. See [BLMW] for further investigation of the mathematical issues that arise here. The main advantage of this reformulation is that the geometric obstruction $V_\lambda(G)$ has a natural compact specification (label), namely the partition λ, and this plays a crucial role in the approach.

But a proof technique cannot be considered an approach if it only provides an equivalent reformulation of the original hardness problem in the language of its choice. It also has to show how to break through the circle of equivalences. This essentially amounts to resolving the self referential paradox mentioned above. The article [GCTflip] describes the defining strategy of GCT, called the *flip*, to resolve this paradox. 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. The article [GCTexpo] explains in what sense the flip amounts to an explicit resolution of the self referential paradox. This article describes an approach to implement the flip, and thereby resolve the self referential paradox, based on the *Positivity Hypotheses (PH)* in algebraic geometry and representation theory. The first positivity hypotheses called PH1 basically say that the multiplicities (number of occurrences) of the Weyl module $V_\lambda(G)$ in the homogeneous coordinate rings $R_{n,m}$ and S_m can be expressed as the number of integer points in *explicitly* (cf. Section 4.1) given polytopes, just as the Littlewood-Richardson coefficient can be expressed as the number of integer points in the explicitly given Hive polytope [KT1]. Such an expression is positive because there is no cancellation in it unlike in the classical character formulae for multiplicities. The second hypotheses called SH (Saturation Hypotheses) say that these multiplicities have a generalized and relaxed form of the saturation property [KT1] that the Littlewood-Richardson coefficients have. These are weaker forms of the second positivity hypotheses (PH2) which say that the stretching functions associated with these coefficients can be expressed, after a small ($\text{poly}(n, m)$) shift, as asymptotic quasipolynomials with nonnegative coefficients.

The self referential difficulty is absent in these positivity hypotheses unlike in the original

hardness hypothesis (the permanent vs. determinant problem), because (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): because PH1 and SH (PH2) for the variety $Y_{n,m}$ 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 variety X_m are statements only about the properties of the determinant and do not depend in any way on the permanent or the complexity class $\#P$.

Formulation of these positivity hypotheses is the key step in this paper. Assuming them, the Decomposition Theorem 4.9 reduces the original hardness hypothesis to a fundamentally "easier" hardness hypothesis called OH (Obstruction Hypothesis). Here 'easier" means whether a given geometric obstruction label λ satisfies the condition in OH can be decided in time polynomial in n, m, and the bitlength of the specification of λ. This ease of verification of an obstruction label is a crucial step in the resolution of the self referential paradox. Thus the Decomposition Theorem decomposes the original hardness (lower bound) problem into the positivity hypotheses (PH1 plus SH), without the self referential difficulty, plus an "easier" hardness hypothesis OH, which too does not have the self referential difficulty once PH1 and SH are proved. Pictorially:

$$\text{(Strong) perm. vs. det. conjecture} \stackrel{\cdots>}{\longleftarrow} PH1 + SH + OH. \tag{1}$$

This decomposition *breaks the circle* of self referential difficulty. Here the left hand side is the stronger form of the permanent vs. determinant conjecture (Conjecture 2.1) defined in [GCT1]. The solid arrow $\longleftarrow$ denotes the formal implication as per the Decomposition Theorem 4.9. The dotted arrow $\cdots >$ indicates evidence for the plausible converse (cf. Section 8) based on the Strong Flip Theorem 2.3 in [GCTflip]. This result shows that the stronger permanent vs. determinant conjecture in conjunction with a stronger form of a standard derandomization conjecture in complexity theory implies solution to a formidable explicit construction problem in algebraic geometry akin to (but even more explicit than) the explicit construction problems in the positivity hypotheses. This suggests that these positivity hypotheses may be in essence *implications* of the stronger permanent vs. determinant and derandomization conjectures. How to make the dotted arrow in the decomposition solid, as one would ideally like, is open. There is a similar decomposition for the P vs. NP problem in characteristic zero; cf. Section 10.

The positivity hypotheses turn out to be formidable because, as explained in Section 5, they encompass and go much further than the century-old plethysm problem in algebraic geometry and representation theory. The Strong Flip Theorem mentioned above suggests that problems of comparable difficulty would arise in *any* approach, though their concrete forms may be very different from PH1 and SH here. In this sense positivity is a hidden root difficulty underneath the fundamental hardness conjectures of complexity theory. This may explain why these conjectures have turned out to be so hard though they look so elementary at the surface.

The rest of this article is organized as follows. Section 2 recalls from [GCT1] and [GCT2] the reduction of the permanent vs. determinant problem to the problem of proving existence of geometric obstructions, and the Strong Flip Theorem from [GCTflip] that formalizes the self referential paradox in this context. Section 3 instantiates the flip strategy in [GCTexpo, GCTflip] using the geometric obstructions and states the conjectural flip hypotheses satisfied by them.

Section 4 states the positivity hypotheses, the resulting decomposition (1) based on the flip strategy, and the results underlying the solid left arrow in the decomposition (1). Section 5 explains how these positivity hypotheses encompass and go much further than the classical plethysm problem in algebraic geometry and representation theory. Sections 6 and 7 provide proofs of the results underlying the solid left arrow in the decomposition (1). Section 8 justifies the dotted right arrow in the decomposition (1) on the basis of the Strong Flip Theorem in [GCTflip]. Section 10 addresses the P vs. NP problem in characteristic zero.

It may be helpful (though strictly not necessary) to have some familiarity with the formalization of the self referential paradox provided by the Flip Theorem in [GCTflip] and the defining flip strategy of GCT to resolve this paradox by going for explicit proofs. All this is explained in the first two sections of [GCTexpo].

Notation: Given a quantity x, we denote the bitlength of its specification by $\langle x \rangle$, and poly$(n, m, \ldots)$ means polynomial of a constant degree in $n, m, \ldots$.

2 Geometric obstructions

In this section we recall from [GCT1] and [GCT2] the reduction of the permanent vs. determinant problem in characteristic zero (cf. Section 1) to the problem of proving existence of geometric obstructions (proof certificates of hardness). These obstructions are crucial for breaking the circle of self reference as described in the later sections.

2.1 Characterization by symmetries

The article [GCT1] begins with an observation that the permanent and determinant are *exceptional*, i.e., are completely characterized by their symmetries in the following sense.

Let Y be an $m \times m$ variable matrix. We think of its entries, ordered say rowwise, as coordinates of $\mathcal{Y} = \mathbb{C}^l$, $l = m^2$. 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 $G = GL(\mathcal{Y}) = GL_l(\mathbb{C})$ with the following action. Given any $\sigma \in G$, map a polynomial $g(Y) \in V$ to $g^\sigma(Y) = g(\sigma^{-1}(Y))$:

$$\sigma : g(Y) \longrightarrow g(\sigma^{-1}Y). \tag{2}$$

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

We say that a point $x \in V$ is characterized by its stabilizer $G_x = \{\sigma \in G \mid \sigma x = x\}$ if x is the only point, up to a constant multiple, in V such that $\sigma x = x$ for all $\sigma \in G_x$. Then:

(D) [Fr]: The point $\det(Y) \in V$ is characterized by its stabilizer $G_{det} \subseteq G = GL(\mathcal{Y})$, which consists of linear transformations in G of the form:

$$Y \to AY'B, \quad Y' = Y \text{ or } Y^t, \quad A, B \in GL_m(\mathbb{C}), \tag{3}$$

with $(\det(A)\det(B))^m = 1$. We refer to this characterization by symmetries of the determinant as property (D).

Similarly, let X be an $n \times n$ variable matrix, whose entries we think of as coordinates of $\mathcal{X} = \mathbb{C}^{n^2}$ after ordering them rowwise. Let $W = \mathbb{C}[X]_n$ be the space of forms (homogeneous

polynomials) of degree n in the entries of X. It is a representation of $H = GL(\mathcal{X}) = GL_{n^2}(\mathbb{C})$. Then:

(P): [MM] The point $perm(X) \in W$ is also characterized by its stabilizer $H_{perm} \subseteq H$, which consists of linear transformations in H of the form:

$$X \to \lambda X' \mu, \quad X' = X \text{ or } X^t,$$

where λ and μ are products of either diagonal or permutation matrices, with obvious constraints on the product of the diagonal entries of the diagonal matrices. We refer to this characterization of the permanent by its symmetries as property (P).

The basic idea now is to exploit the exceptional nature of these polynomials–i.e. the properties (P) and (D)–to construct appropriate proof certificates of hardness (obstructions).

2.2 Class varieties

Towards this end, [GCT1] associates with the determinant and permanent certain projective varieties called the *class varieties* as follows.

Let $P(V)$ be the projective space of V consisting of the lines in V through the origin. Let $P(W)$ be the projective space of W. Identify X 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. Define an embedding $\phi : W \hookrightarrow V$ by mapping any polynomial $h(X) \in W$ to $h^\phi(Y) = z^{m-n}h(X)$. This also defines an embedding of $P(W)$ in $P(V)$, which we denote by ϕ again.

Let $g = \det(Y)$, thought as a point in $P(V)$ (strictly speaking the line through $\det(Y)$ is a point in $P(V)$, but we ignore this distinction here). Similarly, let $h = \text{perm}(X) \in P(W)$, and $f = h^\phi = \text{perm}^\phi(Y) \in P(V)$.

Let

$$\begin{array}{lllll} \Delta_V[g,m] & = & \Delta_V[g] & = & \overline{Gg} \subseteq P(V), \\ \Delta_W[h,n] & = & \Delta_W[h] & = & \overline{Hh} \subseteq P(W), \\ \Delta_V[f,n,m] & = & \Delta_V[f] & = & \overline{Gf} \subseteq P(V), \end{array} \tag{4}$$

where $\overline{Gg}$ denotes the projective closure of the orbit Gg of g, and so on. Then $\Delta_V[g,m]$ and $\Delta_V[f,n,m]$ are projective G-varieties–i.e., varieties with a natural action of G induced by the action on the G-orbits–and $\Delta_W[h,n]$ is a projective H-variety. We call $\Delta_V[f,n,m]$ the *class variety* of the complexity class $\#P$ since the permanent is $\#P$-complete [V], and $\Delta_V[g,m]$ the *class variety* of the complexity class NC since the determinant belongs to NC and is almost complete [V]. The varieties $Y_{n,m}$ and X_m in the introduction (Section 1) are the varieties $\Delta_V[f,n,m]$ and $\Delta_V[g,m]$ here, respectively.

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

$$\Delta_V[f] = \Delta_V[f,n,m] \subseteq \Delta_V[g,m] = \Delta_V[g]. \tag{5}$$

Conversely, if $\Delta_V[f,n,m] \subseteq \Delta_V[g,m]$, then f 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. Since the permanent is $\#P$-complete, it is conjectured that:

Conjecture 2.1 *(Stronger form of the permanent vs. determinant problem) [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 $a > 0$.*

It is easy to see that this is equivalent to:

Conjecture 2.2 *(cf. Conjecture 4.3 in [GCT1]) If $m = poly(n)$, or more generally, $m = 2^{\log^a n}$, $a > 0$ fixed, $n \to \infty$, then $\Delta_V[f, n, m] \not\subseteq \Delta_V[g, m]$.*

2.3 Strong flip theorem and the self referential paradox

The following result formalizes the self referential paradox in this context. It says that any proof of Conjecture 2.1 can be transformed into an *extremely explicit proof* [GCTexpo, GCTflip] assuming a stronger form of the standard derandomization conjecture [IW, KI], which is regarded in complexity theory as easier than the target lower bound under consideration. This result is also crucial for the justification of the flip and positivity hypotheses needed later for breaking the circle of self reference.

Theorem 2.3 (Strong Flip) *(cf. [GCTexpo, GCTflip]) Suppose Conjecture 2.1 (or equivalently, Conjecture 2.2) 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, for any $m = poly(n)$, one can compute in $poly(n, m)$ time a global obstruction set $S_{n,m} = \{X_1, \ldots, X_l\}$, $l = poly(n, m)$, of nonnegative integral $n \times n$ matrices with the following property. Fix any homogeneous polynomial $p(Y) \in V$ such that the line in $P(V)$ corresponding to $p(Y)$ belongs to $\Delta_V[g, m]$. 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, for any such $p(Y)$, there exists a counter example $X_i \in S_{n,m}$ such that $p'(X_i) \neq perm(X_i)$. Thus $S_{n,m}$ contains a counterexample against every point in $\Delta_V[g, m]$ which proves that the point is different from $f = perm^{\phi}(Y)$.

More strongly, Conjecture 2.1 has an extremely explicit proof [GCTexpo, GCTflip]. This means there exists a family $\mathcal{O} = \cup_{n,m}\mathcal{O}_{n,m}$ of sets of bit strings called obstructions *(or obstruction labels) satisfying the following Flip properties F0-4 [GCTexpo, GCTflip] and also the property (G) specified below:*

F0 [Short]*: The set $\mathcal{O}_{n,m}$ is nonempty and contains a short obstruction string s if m is small, i.e., $O(poly(n))$. Here short means the bitlength $\langle s \rangle$ of s is $poly(n, m)$.*

F1 [Easy to decode:] *Each bit string $s \in \mathcal{O}_{n,m}$, m small, denotes a global obstruction set $S_{n,m}(s)$ (just like $S_{n,m}$ above) such that given s, n and m, $S_{n,m}(s)$ can be computed in $poly(\langle s \rangle, n, m)$ time.*

F2 [Rich]*: For every n and $m = poly(n)$, $\mathcal{O}_{n,m}$ contains at least $2^{\Omega(m)}$ pairwise disjoint obstructions, each of $poly(n, m)$ 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 $poly(n, \langle s \rangle, m)$ time.*

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

The same result also holds if we replace sequential polynomial time algorithms in all the statements above by efficient parallel (NC) algorithms that work in polylogarithmic time using polynomially many processors.

We will justify the terminology flip later (Section 3).

For any short obstruction string $s \in \mathcal{O}_{n,m}$, $m = O(\text{poly}(n)$, let $S_{n,m}(s) = \{X_1, \ldots, X_l\}$, X_i nonnegative and integral, $l = \text{poly}(n)$, be the global obstruction set as in F1. 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 induced morphism from the projective variety $\Delta_V[g, m]$ 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 $\hat{\psi}(\Delta_V[g, m]) \subseteq P(\mathbb{C}^l)$. 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 $\hat{\psi}_s(f) \notin \hat{\psi}_s(\Delta_V[g, m])$. The property (G) mentioned above is that:

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

The linear map $\hat{\psi}_s$ above is called an *extremely explicit positive separator* between $\Delta_V[g, m]$ and $f = z^{m-n}\text{perm}(X)$. It is called extremely explicit because (assuming the relevant hardness and derandomization conjectures): (1) given s, the set $S_{n,m}(s)$ which specifies $\hat{\psi}_s$ can be computed in $O(\text{poly}(n, m))$ time by Theorem 2.3, and (2) each coefficient of the representation of $\hat{\psi}_s$ in the standard basis [1] of V can also be computed in $\text{poly}(n, m)$ time; this also follows from Theorem 2.3. It is called positive because each such coefficient is nonnegative. We call $l = \text{poly}(n, m)$ the *dimension* of $\hat{\psi}_s$. Thus Theorem 2.3 says that, assuming the strong arithmetic permanent vs. determinant and derandomization conjectures, one can construct a compact specification $S_{n,m}(s)$ of an extremely explicit positive linear separator of small dimension between $\Delta_V[g, m]$ and f in $\text{poly}(n, m)$ time, when m is small.

Theorem 2.3 formalizes the self referential paradox in the following sense. Given X_i and $p(Y)$, to check if X_i is a counterexample against $p(Y)$, we have to check if $p'(X_i) \neq \text{perm}(X_i)$. This cannot be checked efficiently for general X_i, assuming that the permanent is hard to compute. Yet, by F1 and F3, whether $S_{n,m}(s)$ contains a counterexample against *every* $p(Y) \in \Delta_V[g, m]$ can be checked efficiently even in parallel. This seems to contradict the very hardness of the permanent that we are trying to prove. See [GCTexpo] for further discussion of this self referential paradox. Theorem 2.3 says that extremely explicit resolution of this paradox is *forced* by the strong permanent vs. determinant conjecture, modulo derandomization. In view of this result, the main conceptual difficulty in proving this conjecture is to *break the circle* of self reference by decomposing it into subproblems without the self referential difficulty. This is the goal for the rest of this paper.

Theorem 2.3 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 $\Delta_V[g, m] = \overline{Gg}$ and f can

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

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 [MM2]. 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, assuming the underlying hardness and derandomization conjectures, a short specification $S_{n,m}$ of a linear separator between $\Delta_V[g,m]$ and $f = z^{m-n}\text{perm}(X)$ can be computed in $\text{poly}(n,m)$ time exploiting the exceptional nature of the permanent and the determinant. This may seem impossible and out of reach of the existing algebraic geometry, and this may explain why the fundamental hardness conjectures of complexity theory, which seem so elementary at the surface, have turned out to be so hard.

Theorem 2.3 also suggests *the law of conservation of difficulty*: namely, that any proof of the (strong) permanent vs. determinant conjecture would have to overcome problems of difficulty comparable to the explicit construction of such linear separators. The various flip and positivity hypotheses (FH and PH) described in Sections 3 and 4 are such problems encountered in GCT.

Remark: The self referential difficulty and the strong flip Theorem 2.3 are relevant only for lower bound problems harder than derandomization of polynomial or determinant identity testing. The permanent vs. determinant problem is such a problem as per the existing evidence in complexity theory [KI]. The self referential difficulty is not issue in proving the quadratic lower bound for the permanent [LMR, MR2], which has a relatively simple proof. Indeed, the analogue of Theorem 2.3 in this case will be a statement about the difficulty of the additional derandomization conjecture, not the difficulty of proving the quadratic lower bound.

2.4 Obstructions

We now recall the notion of geometric obstructions to the embedding (5) from [GCT2]. This will be crucial for breaking the circle of self reference.

For that let us examine Conjecture 2.2 closely. To prove it and thereby solve the original permanent vs. determinant problem in characteristic zero, we have to prove that the inclusion

$$\Delta_V[f,n,m] = \Delta_V[f] \subseteq \Delta_V[g] = \Delta_V[g,m] \tag{6}$$

is not possible when m is small. Suppose to the contrary. Let $R_V[f] = R_V[f,n,m]$ and $R_V[g] = R_V[g,m]$ denote the homogeneous coordinate rings of $\Delta_V[f,n,m]$ and $\Delta_V[g,m]$, respectively. These rings were denoted by $R_{n,m}$ and S_m, respectively, in the introduction (Section 1). Let $R_V[f]_d = R_V[f,n,m]_d$ and $R_V[g]_d = R_V[g,m]_d$ be their degree d components. These are finite dimensional G-modules since $\Delta_V[f]$ and $\Delta_V[g]$ are G-varieties. If (6) holds then there is a surjective G-homomorphism from $R_V[g]_d$ to $R_V[f]_d$ obtained by restriction. By dualizing, we get an injective G-homomorphism from the dual $R_V[f]_d^*$ of $R_V[f]_d$ to the dual $R_V[g]_d^*$ of $R_V[g]_d$:

$$R_V[f,n,m]_d^* = R_V[f]_d^* \hookrightarrow R_V[g]_d^* = R_V[g,m]_d^*. \tag{7}$$

Let $V_\lambda(G)$ be the Weyl module [FH] (polynomial irreducible representation) of G labelled

by the partition $\lambda = (\lambda_1, \ldots, \lambda_k)$, i.e., a nondecreasing integer sequence $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_k > 0$, with length $k \leq l$. Let $|\lambda| = \sum_i \lambda_i$ denote the size of λ.

Definition 2.4

(1) [GCT2] A Weyl module $S = V_\lambda(G)$, for a given partition λ, is called an incidence-based geometric obstruction *for the inclusion (6) if $V_\lambda(G)$ occurs as a G-submodule in $R_V[f,n,m]^*$ but not in $R_V[g,m]^*$. By occurring in $R_V[f,n,m]^*$, we mean in $R_V[f,n,m]^*_d$ for some d, which, it is easy to see, has to be $|\lambda|/m$. We call λ an* obstruction label*, and sometimes by abuse of notation, an obstruction as well.*

(2) We say that a Weyl module $S = V_\lambda(G)$ is a multiplicity-based geometric obstruction if its multiplicity in $R_V[f,n,m]^*$ (that is, in $R_V[f,n,m]^*_d$ for $d = |\lambda|/m$) exceeds the multiplicity in $R_V[g,m]^*$.

(3) We call $V_\lambda(G)$ a threshold-based geometric obstruction if there exists a small $k = O(poly(n,m)$ such that the multiplicity of $V_\lambda(G)$ in $R_V[g,m]^*$ is bounded by k and the multiplicity in $R_V[f,n,m]^*$ exceeds k.

In [GCT2] (1) is stated in terms of $SL_l(\mathbb{C})$ but keeping track of the grading information. This is the same as using $GL_l(\mathbb{C})$ instead.

If a geometric obstruction exists, for given n and m, then the inclusion (6) is not possible, and hence, perm(X) cannot be linearly represented as a determinant of an $m \times m$ matrix. Thus a geometric obstruction is a proof certificate of hardness of the permanent. It has a natural short specification, namely, the partition λ. This is crucial in what follows.

Conjecture 2.2 or equivalent Conjecture 2.1 is now implied by the following conjectural Geometric Obstruction Hypothesis (GOH).

Hypothesis 2.5 (GOH) *[GCT2] Incidence-based geometric obstructions exist in the permanent vs. determinant problem when $m = poly(n)$, or more generally, $m = 2^{\log^a n}$, $a > 0$ a constant.*

Furthermore:

Conjecture 2.6 (Equivalence) *The stronger form of the permanent vs. determinant problem (cf. Conjecture 2.1 and 2.2) is equivalent to the problem of proving existence of threshold-based geometric obstructions. Specifically, for given n and m, $\Delta_V[f,n,m] \not\subseteq \Delta_V[g,m]$ iff a threshold-based geometric obstruction exists.*

See Section 8.1 for justification of GOH and this conjecture based on the Strong Flip Theorem 2.3. Also see [BI] for nontrivial computer based numerical evidence for the analogue of GOH in GCT of matrix multiplication.

For simplicity, we focus on incidence-based geometric obstructions in what follows, though GCT can be extended to threshold-based geometric obstructions as well by augumenting the various hypotheses suitably. So when we say geometric obstructions, we henceforth mean incidence-based geometric obstructions unless stated otherwise.

3 The flip: going for explicit construction of obstructions

The goal now is to prove GOH by constructing (incidence based) geometric obstructions *explicitly*. This is natural in view of the Strong Flip Theorem 2.3, which forces an explicit proof, modulo derandomization. This strategy is instantiation of the abstract *flip strategy* from [GCTexpo, GCTflip]. It is called a flip, because it converts the lower bound problem under consideration into an upper bound problem of constructing the obstructions efficiently. We will get a more or less explicit proof in GCT whether we care for explicitness or not (cf. Section 4.5), not surprisingly in view of Theorem 2.3.

3.1 Flip hypotheses

We begin by formulating the Flip Hypotheses (FH) that specify what explicit means in this context.

Let $\mathcal{O}^G = \cup_{n,m}\mathcal{O}^G_{n,m}$ be the family of (incidence-based) geometric obstruction labels, i.e., partitions λ's (cf. Definition 2.4), where $\mathcal{O}^G_{n,m}$ consists of all geometric obstruction labels for given n and m.

The following is a conjectural extremely explicit form of Hypothesis 2.5 (GOH). It says that geometric obstructions can be constructed explicitly just like the global obstruction sets $S_{n,m}(s)$ in the Strong Flip Theorem 2.3.

Hypothesis 3.1 (General FH) *The family $\mathcal{O}^G = \cup_{n,m}\mathcal{O}^G_{n,m}$ of geometric obstruction labels is extremely explicit. This means it satisfies the analogues of the properties F0-4 satisfied by the family of obstructions in the Strong Flip Theorem 2.3. Specifically:*

1. **F0 [Short]***: $\mathcal{O}^G_{n,m}$ is nonempty and contains a short obstruction label (partition) λ if m is small, i.e., $O(poly(n))$, or more generally $O(2^{\log^a n})$, $a > 0$ a constant. Here short means the size $|\lambda| = \sum_i \lambda_i$ is $O(poly(n, m))$.*

2. **F1 [Easy to decode]:** *Given $n, m \geq n$, and an obstruction label $\lambda \in \mathcal{O}^G_{n,m}$, we can construct a global obstruction set $S_{n,m}(\lambda)$ (like $S_{n,m}(s)$ in Theorem 2.3) against all forms in $\Delta_V[g, m]$ in $poly(\langle\lambda\rangle, n, m)$ time, where $\langle\lambda\rangle = \sum_i \log_2 \lambda_i$ is the bitlength of λ.*

3. **F2 [Rich]:** *For every n and $m = poly(n)$, $\mathcal{O}^G_{n,m}$ contains at least $2^{\Omega(m)}$ distinct partitions, each of $poly(n, m)$ bitlength.*

4. **F3 [Easy to verify]:** *Given n, m and a partition λ, whether λ is a valid geometric obstruction label for n and m–i.e., whether $\lambda \in \mathcal{O}^G_{n,m}$–can be verified in $poly(n, \langle\lambda\rangle, m)$ time.*

5. **F4 [Easy to construct]:** *Given n and $m = poly(n)$, a valid geometric obstruction label λ in $\mathcal{O}^G_{n,m}$ of $poly(n, m)$ size can be constructed in $poly(n, m) = poly(n)$ time.*

The family of threshold-based geometric obstruction labels is also extremely explicit.

Just like the conclusion of the strong flip Theorem 2.3, this hypothesis too may seem impossible at the surface, since the existing algorithms in algebraic geometric and representation theory take at least double exponential time to construct a geometric obstruction label for given n and small m (for the same reasons as in Section 2.3). See Section 8.1 for the justification of this hypothesis based on the Strong Flip Theorem 2.3, which suggests that this hypothesis may in essence be an *implication* of the hardness and derandomization conjectures mentioned in the statement of Theorem 2.3.

Next we formulate a weaker form of this hypothesis.

Let $G(\lambda, m)$ denote the multiplicity of $V_\lambda(G)$ in $R_V[g, m]^*$, i.e, in $R_V[g, m]^*_d$, with $d = |\lambda|/m$. Here $|\lambda|$ is assumed to be divisible by m. Otherwise the multiplicity is zero. Similarly let $F(\lambda, n, m)$ be the multiplicity of $V_\lambda(G)$ in $R_V[f, n, m]^*$. To decide if $V_\lambda(G)$ is a geometric obstruction, we have to decide if these multiplicities are zero or nonzero. Specifically, $V_\lambda(G)$ is a geometric obstruction iff $G(\lambda, m)$ is zero and $F(\lambda, n, m)$ is nonzero.

Let $\hat{G} = SL_l(\mathbb{C})$, and $\hat{G}_{det} \subseteq \hat{G}$ be the stabilizer of $\det(Y) \in V$ under the action of $\hat{G}$ on V given by eq.(2). Let $G'(\lambda, m)$ denote the multiplicity of the $\hat{G}_{det}$-invariant in the Weyl module $V_\lambda(\hat{G})$ of $\hat{G}$. It is known (cf. Theorem 1.1 in [GCT2]) that $V_\lambda(\hat{G})$ occurs in the coordinate ring $R_V[g, m]$ iff $G'(\lambda, m)$ is nonzero. The multiplicity $G'(\lambda, m)$ is much easier than $G(\lambda, m)$, because its definition is purely representation theoretic, unlike the definition of $G(\lambda, m)$ which involves algebraic geometry in an essential way. We shall also study $G'(\lambda, m)$ concurrently, because this provides in a much simpler setting a glimpse of the difficulties underlying $G(\lambda, m)$.

Hypothesis 3.2 (FH)

Permanent FH *(1): The multiplicity $F(\lambda, n, m)$ belongs to #P, i.e., has a positive #P-formula (with λ specified in binary, and n and m in unary).*

Permanent FH *(2): The problem of deciding if $F(\lambda, n, m)$ is nonzero belongs to the complexity class P; i.e., has an algorithm that works in poly$(\langle\lambda\rangle, n, m)$ time.*

Determinant FH *(1) and (2): The situation for $G(\lambda, m)$ and $G'(\lambda, m)$ is similar.*

For justification of this hypothesis based on the Strong Flip Theorem 2.3, see Section 8.2.

Proposition 3.3 *Assuming Determinant and Permanent FH (2) (cf. Hypothesis 3.2), the problem of verifying a geometric obstruction (label) in the permanent vs. determinant problem belongs to P: i.e., given λ, n and m, whether λ is a valid geometric obstruction label (cf. Definition 2.4) can be decided in poly$(n, m, \langle\lambda\rangle)$ time (as per F3 in Hypothesis 3.1).*

This follows trivially from the definitions.

Permanent FH crucially depends on the characterization by symmetries of the permanent (the property (P)), because this is crucial in the proof of the strong Flip Theorem 2.3, which is the basis for its justification (Section 8.2). Similarly, Determinant FH crucially depends on the characterization by symmetries of the determinant (the property (D)). If we were to replace the determinant and permanent with general functions without symmetries, FH and General FH would almost certainly fail. This is why GCT can work for only exceptional functions such as the permanent and determinant.

3.2 Breaking of the circle and the flip

The self referential difficulty (Section 2.3) is present in the permanent vs. determinant problem because: (1) m is required to be a small function of n in its statement, and (2) the problem is based on the relationship between the permanent and the determinant, or equivalently, between the complexity classes $\#P$ and NC. This difficulty is absent in Determinant and Permanent FH because (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 Permanent FH is a statement only about the properties of the permanent and does not depend in any way on the determinant or the complexity class NC, and Determinant FH is a statement only about the properties of the determinant and does not depend in any way on the permanent or the complexity class $\#P$.

Furthermore, once Determinant and Permanent FH (2) are proved, geometric obstructions are easy to verify (cf. Proposition 3.3). As we saw in Section 2.3, the self referential paradox is the main obstacle in the implementation of the flip condition for verification (F3). The above discussion says that once Determinant and Permanent FH are proved, F3 is satisfied for geometric obstructions. This means GOH then becomes "easy to verify" and does not have the self referential difficulty anymore. Thus we get the decomposition of the original permanent vs.determinant conjecture as Determinant FH (2) plus Permanent FH (2) plus GOH, all without the self referential difficulty. Pictorially,

$$(Strong)\ Perm.\ vs.\ Det. \stackrel{\cdots>}{\longleftarrow} \mathrm{D}eterminant\ FH\ (2) + \mathrm{P}ermanent\ FH\ (2) + GOH. \qquad (8)$$

Though Proposition 3.3 underlying the solid arrow $\longleftarrow$ is trivial, such decomposition into subproblems without self referential difficulty is possible because of the exceptional nature of geometric obstructions. For example, this is not possible using the global obstructions sets $S_{n,m}(s)$ in the strong flip Theorem 2.3. The dotted arrow $\cdots >$ here denotes the evidence for the plausible converse given in Sections 8.1 and 8.2.

The decomposition (8) breaks the circle of self reference and suggests the following flip strategy for showing existence of geometric obstructions (GOH). It is a partial instantiation of the abstract flip strategy [GCTexpo, GCTflip] in this concrete setting:

Step I: Find "easy" algorithms needed in Determinant and Permanent FH (2) to get an "easy" criterion for verifying an obstruction label as in Proposition 3.3.

Step II: Use this "easy" verification criteria to guess and construct a geometric obstruction label λ *explicitly*, thereby solving the permanent vs. determinant problem.

4 How to prove FH?

In this section we state the main results and positivity hypotheses in this paper that provide an approach to prove a relaxed form of Determinant and Permanent FH (2). This will lead to a decomposition that is more refined than (8).

4.1 Definitions

For this, we need several definitions.

Let $P \subseteq \mathbb{R}^n$ be a rational polytope specified by a system of linear inequalities

$$Ax \leq b, \tag{9}$$

where A is an $m \times n$ rational matrix, b a rational m-vector, and x a variable n-vector. Its bitlength $\langle P \rangle$ is defined to be n plus the maximum bitlength of any linear constraint in the system (9). Now let $\{P_{\alpha,\beta,\ldots}\}$ be a family of polytopes where $\alpha, \beta, \cdots$ are the specification parameters on which $A = A_{\alpha,\beta,\ldots}$ and $b = b_{\alpha,\beta,\ldots}$ depend. We say that this family of polytopes is *explicit*–or simply that the polytope $P = P_{\alpha,\beta,\ldots}$ is explicit–if (1) the bitlength $\langle P_{\alpha,\beta,\ldots} \rangle = O(\text{poly}(\langle \alpha \rangle, \langle \beta \rangle, \cdots))$, and (2) there is a separation oracle [GLS] which, given the parameters $\alpha, \beta, \cdots$, and a rational point x, tells whether x belongs to $P = P_{\alpha,\beta,\ldots}$ in $\text{poly}(\langle P \rangle, \langle x \rangle) = \text{poly}(\langle \alpha \rangle, \langle \beta \rangle, \cdots, \langle x \rangle)$ time, and if $x \notin P$, also gives a hyperplane separating x from P in the same time. In particular, this implies that the membership problem for the polytope $P_{\alpha,\beta,\ldots}$ belongs to the complexity class P. Here the number of constraints m in (9) can be exponential in n. For example, the family of perfect matching polytopes of graphs (with a graph as a parameter) is explicit [GLS], though the number of defining constraints of such a polytope can be exponential if the graph is nonbipartite.

The Ehrhart quasipolynomial $f_P(k)$ of P is defined to be the number of integer points in the dilated polytope kP. It is known to be a quasipolynomial [St3]. Here a function $f(k)$ is called a quasi-polynomial if there exist l polynomials $f_j(k)$, $1 \leq j \leq l$, such that $f(k) = f_j(k)$ if $k = j \mod l$. Here l is supposed to be the smallest such integer, and is called the *period* of $f(k)$.

More generally, we say that a function $f(k)$ is an *asymptotic quasipolynomial*, if there exist l polynomials $f_j(k)$, $1 \leq j \leq l$, for some l, such that $f(k) = f_j(k)$ for all nonnegative integral $k = j$ modulo l for $k \geq a(f)$, for some nonnegative integer $a(f)$ depending on f. The minimum $a(f)$ for which this holds is called the *deviation from quasipolynomiality*. Thus $f(k)$ is a (strict) quasipolynomial when this deviation is zero.

A basic example of an asymptotic quasi-polynomial is the following. Let $P(k)$ be a rational polytope parametrized by nonnegative integral k: i.e., defined by a linear system of the form:

$$Ax \leq kb + c, \tag{10}$$

where A is a rational $m \times n$ matrix, x a variable n-vector, and b and c rational m-vectors. We say the polytope is homogeneous if $c = 0$, and nonhomogeneous otherwise. Let $f_P(k)$ be the number of integer points in $P(k)$. It is known to be an asymptotic quasi-polynomial (cf. Theorems 2.3 and 3.2 in [St] and Chapter 4 in [St3]). We call it the asymptotic Ehrhart quasi-polynomial of the polytope $P(k)$. When $c = 0$ (the homogeneous case) $f_P(k)$ is the Ehrhart quasi-polynomial of $P(1)$. Let

$$b(P) = \min\{k \in \mathbb{N} | \dim(P(k)) = \lim_{k \to \infty} \dim(P(k))\} - 1. \tag{11}$$

Let $\delta(P) = max\{a(f_P), b(P)\}$. We call it the defect of $f_P(k)$. It can be shown that in the worst case it is exponential in the bitlength $\langle P \rangle$ of P. In what follows, we denote the polytope $P(k)$ by just P. From the context it should be clear whether P is homogeneous or nonhomogeneous.

4.2 Quasipolynomiality result

Now we state the basic quasipolynomiality result (Theorem 4.1) which forms a basis for the positivity hypotheses (cf. Section 4.3) in this paper.

Let $F(\lambda, n, m)$, $G(\lambda, m)$, and $G'(\lambda, m)$ be the multiplicities in Hypothesis 3.2. Let $F_{\lambda,n,m}(k) = F(k\lambda, n, m)$, $G_{\lambda,m}(k) = G(k\lambda, m)$, and $G'_{\lambda,m}(k) = G'(k\lambda, m)$ be the stretching functions associated with them. Here $k\lambda$ denotes the partition obtained by multiplying each part of λ by k. Then:

Theorem 4.1 *(1) The stretching function $G'_{\lambda,m}(k)$ is a quasi-polynomial.*
(2) The stretching functions $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ are asymptotic quasi-polynomials.

This follows (cf. Section 6) from the works and ideas of Hilbert, Boutot [Bt], Brion (cf. [D]), Kempf, Flenner [F] and others. The crucial tool in the proof of (1)–the work of Boutot [Bt]–is based on the resolution of singularities in characteristic zero [H]. As such, this proof is highly nonconstructive. It gives no effective bound on the period of the quasipolynomials. It only says the period is finite. The functions $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ are not expected to be quasipolynomials since the varieties $\Delta_V[f, n, m]$ and $\Delta_V[g, m]$ are not normal [Ku]. But their deviations from quasipolynomiality are expected to be small in view of (1); cf. Hypothesis 4.2 below for a precise conjecture.

4.3 Positivity hypotheses

Using Theorem 4.1 we can now state the basic positivity and saturation hypotheses PH1, SH and PH2 that were mentioned in the introduction. Justification for why they should hold, and why they may be in essence implications of the hardness and derandomization conjectures in the statement of the Strong Flip Theorem 2.3, is given later in Section 8.3.

The following Positivity Hypothesis (PH1) says that the (asymptotic) quasipolynomials in Theorem 4.1 can be realized as (asymptotic) Ehrhart quasipolynomials of explicit polytopes (with small defect), just as in the case of Littlewood-Richardson coefficients [BZ, DM] (where the defect as well as the deviation from quasipolynomiality are zero).

Hypothesis 4.2 (PH1)

Permanent PH1: *For every $\lambda, n, m \geq n$, there exists an explicit possibly nonhomogeneous polytope $P(k) = P_{\lambda,n,m}(k)$ (with specification parameters n and m in unary, and λ and k in binary) of specification bitlength $\langle P(k)\rangle = poly(n, m, \langle\lambda\rangle, \langle k\rangle)$ such that*

$$F_{\lambda,n,m}(k) = f_P(k), \tag{12}$$

the asymptotic Ehrhart quasipolynomial of $P(k)$, and the defect $\delta(P)$ of $f_P(k)$ is $O(poly(m, \langle\lambda\rangle))$. If such a polytope exists it is guaranteed by the proof of Theorem 4.1 that its dimension is $poly(n)$ regardless of what m is.

Determinant PH1 (a)*: For every m, there exists an explicit (possibly nonhomogeneous) polytope $Q(k) = Q_{\lambda,m}(k)$ (with specification parameters m in unary and λ and k in binary) of*

specification bitlength $\langle Q\rangle = poly(m, \langle\lambda\rangle, \langle k\rangle)$ *such that*

$$G_{\lambda,m}(k) = f_Q(k), \tag{13}$$

the asymptotic Ehrhart quasi-polynomial of Q*, and the defect* $\delta(Q)$ *of* $f_Q(k)$ *is* $O(poly(m, \langle\lambda\rangle))$*. If such a polytope exists it is guaranteed by the proof of Theorem 4.1 that its dimension is* $poly(n)$ *regardless of what* m *is as long as the length of the partition* λ *is* $poly(n)$ *(as it will be in our applications).*

Determinant PH1 (b)*: A similar explicit polytope* $Q'_{\lambda,m}(k)$ *exists for* $G'_{\lambda,m}(k)$*.*

PH1 implies $\#P$-formulae for $F(\lambda, n, m), G(\lambda, m)$ and $G'(\lambda, m)$ as needed in FH (1) (Hypothesis 3.2).

It has to be stressed here that Permanent and Determinant PH1 are expected to hold only because of the exceptional nature of the permanent and the determinant (cf. Section 8.3). If we replace the permanent and the determinant with general functions with no symmetries, they would almost surely fail for the reasons given in Section 8.3.

To state the next positivity and saturation hypotheses we need a few definitions.

Definition 4.3 *We say that a quasi-polynomial (cf. Section 4.1)* $f(k)$ *is* strictly positive*, if all coefficients of* $f_j(k)$*, for all* j*, are nonnegative. In general, we define the* positivity index $p(f)$ *of* f *to be the smallest nonnegative integer such that* $f(k + p(f))$ *is strictly positive.*

Here and below it is assumed that the leading coefficient of each $f_j(k)$ is nonnegative, as it is in the case of an Ehrhart quasipolynomial; otherwise the positivity index is not defined. Clearly $f(k)$ is strictly positive if and only if its positivity index is zero.

Definition 4.4 *We say that* $f(k)$ *is* strictly saturated *if for any* j*:* $f_j(k) > 0$ *for every* $k \geq 1$*,* $k = j \mod l$*, whenever the polynomial* $f_j(k)$ *is not identically zero. The* saturation index $s(f)$ *of* f *is defined to be the smallest nonnegative integer such that* $f(k + s(f))$ *is strictly saturated.*

Thus $f(k)$ is strictly saturated if and only if its saturation index is zero, and if $f(k)$ is strictly positive, it is strictly saturated. Clearly the saturation idex is bounded above by the positivity index.

By the saturation theorem [KT1], the stretching function associated with the Littlewood-Richardson coefficient (which is known to be a polynomial [Rs]) is strictly saturated. It is conjectured in [KTT] on the basis of considerable evidence that it is strictly positive as well.

Next we extend the definitions of saturation and positivity indices to asymptotic Ehrhart quasipolynomials.

Definition 4.5 *Let* $f = f_P(k)$ *be the asymptotic Ehrhart quasi-polynomial of a possibly nonhomogenoeus polytope* $P(k)$*,* $\delta(P)$ *its defect and* l *its period. Then the* positivity index $p(f)$ *is the smallest nonnegative integer* $\geq \delta(P)$ *such that* $f(k + p(f))$ *is strictly positive. The* saturation index $s(f)$ *is the smallest nonnegative integer* $\geq \delta(P)$ *such that* $f(k + s(f))$ *is strictly saturated. Equivalently,* $s(f)$ *is the smallest nonnegative integer* $\geq \delta(P)$ *such that, for any* $k \geq s(f)$*,* $f_j(k) > 0$*,* $j = k \mod l$*, if the polynomial* $f_j(k)$ *is not identically zero.*

By PH1 (Hypothesis 4.2), $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ are asymptotic Ehrhart quasi-polynomials, whose defects and saturation (positivity) indices are thus well defined[2].

With this in mind we can now state for $F(\lambda, n, m), G(\lambda, m)$ and $G'(\lambda, m)$ a conjectural generalized and relaxed form of the saturation property [KT1] for Littlewood-Richardson coefficients.

Hypothesis 4.6 (SH)

[Permanent SH]: *The saturation index of $F_{\lambda,n,m}(k)$ is poly$(m, \langle\lambda\rangle)$.*

[Determinant SH] (1): *The case of $G_{\lambda,m}(k)$ is similar.*

[Determinant SH] (2): *The case of $G'_{\lambda,m}(k)$ is also similar.*

This follows from:

Hypothesis 4.7 (PH2)

[Permanent PH2]: *The positivity index of $F_{\lambda,n,m}(k)$ is poly$(m, \langle\lambda\rangle)$.*

[Determinant PH2] (1): *The case of $G_{\lambda,m}(k)$ is similar.*

[Determinant PH2] (2): *The case of $G'_{\lambda,m}(k)$ is also similar.*

4.4 A relaxed form of FH

The following result proves a relaxed form of Determinant and Permanent FH (2) (cf. Hypothesis 3.2) assuming PH1.

Theorem 4.8 *Assuming Permanent PH1 (Hypotheses 4.2), and given λ, n, m, and k' greater than the saturation index of $F_{\lambda,n,m}(k)$ (polynomially bounded in Permanent SH (Hypothesis 4.6)), whether $F_{\lambda,n,m}(k')$ is nonzero can be decided in poly$(\langle\lambda\rangle, n, m, \langle k'\rangle)$ time. Similar results hold for $G_{\lambda,m}(k)$ and $G'_{\lambda,m}(k)$ assuming Determinant PH1 and SH.*

This is proved in Section 7.3. SH is needed so that this result holds for small k'.

To prove Determinant and Permanent FH (2) in full generality, one would presumably need some strengthening of PH1 and SH.

4.5 Decomposition

The following result leads to the decomposition (1) of the permanent vs. determinant problem in terms of positivity; cf. Section 4.6.

Theorem 4.9 (Decomposition) *There exists an* explicit *family $\mathcal{O} = \mathcal{O}_{n,m}$ of obstruction labels for the permanent vs. determinant problem in characteristic zero (cf. Hypothesis 2.5), for $m = 2^{\log^a n}$, $a > 0$ fixed, $n \to \infty$, assuming,*

1. Permanent PH1 and Determinant PH1 (a) (cf. Hypothesis 4.2), and

[2]Conjecturally, the choice of the polytopes in PH1 does not matter in what follows.

2. *OH (Obstruction Hypothesis):*

   For all $n \to \infty$, $m = 2^{\log^a n}$, $a > 0$ fixed, there exist λ and k greater than the saturation index bound in Permanent SH (Hypothesis 4.6) such that

   (a) *The affine span of the polytope $P = P_{\lambda,n,m}(k)$ (in Permanent PH1 (a)) contains an integer point. Here by an affine span we mean the smallest dimensional affine space containing the polytope.*

   (b) *The affine span of the polytope $Q = Q_{\lambda,m}(k)$ (in Determinant PH1 (a)) does not contain an integer point.*

The set $\mathcal{O}_{n,m}$ here consists of obstruction specifications of the form (λ, k), where λ and k satisfy OH for given n and m. In this case $V_{k\lambda}(G)$ is an incidence-based geometric obstruction for given n and m, and (λ, k) is a specification of this obstruction. The family $\mathcal{O} = \cup_{n,m}\mathcal{O}_{n,m}$ is explicit in the sense that whether a given obstruction specification (λ, k) satisfies OH can be verified in poly$(n, m, \langle\lambda\rangle, \langle k\rangle)$ time. This is the analogue of F3 in Hypothesis 3.1. The analogue of F0 (shortness) also holds if there exist λ and k of poly(m) bitlength satisfying OH. The other flip properties F1,F2, and F4 in Hypothesis 3.1 are not required in this weaker form of explicitness.

This is proved in Section 7.3. Though SH does not appear explicitly in the statement of Theorem 4.9, it is critical for OH to hold; cf. Section 8.4. OH does not test if the polytope P or Q contains an integer point. This test is hard since the integer programming problem is NP-complete. We need k to be larger than the saturation index bound in Permanent SH precisely to avoid this test.

We said in Section 3 that in GCT we would end up constructing obstructions more or less explicitly, whether we care for explicitness or not. This is because proving OH would not be feasible unless we know the polytopes P and Q in PH1 explicitly. And once we know the polytopes P and Q explicitly, the existence of an explicit family of obstructions follows (Theorem 4.9) without much additional difficulty, as a bonus, whether we care for explicitness or not.

4.6 Breaking the circle

Theorem 4.9 decomposes the original hardness hypothesis (Conjecture 2.2) as PH1 plus SH plus OH. We denote this pictorially as:

$$\text{Strong perm. vs. det. conjecture} \longleftarrow PH1 + SH + OH. \tag{14}$$

Here the self referential difficulty is absent in PH1 and SH for the same reasons that it is absent in Determinant and Permanent FH (cf. Section 3.2). Furthermore, once PH1 and SH are proved, by Theorem 4.9 OH is easy to verify in polynomial time for given obstruction specification (λ, k), and hence, the self referential difficulty is absent in OH, just as in GOH (cf. Section 3.2). This decomposition breaking the circle of self reference is more refined than the earlier decomposition (8). The subproblems on the right hand side here are simpler than the

ones in (8) because PH1, SH, and OH here are polyhedral conditions in contrast to the earlier Determinant and Permanent FH or GOH. The ultimate goal is to continue this refinement until we get a decomposition into subproblems simple enough to be solved.

Theorem 4.9 addresses the solid left arrow in the decomposition (1). The dotted right arrow will be addressed in Sections 8.2 and 8.4.

5 Positivity in algebraic geometry and representation theory

In this section we explain in what sense the positivity hypotheses in this paper encompass and go much further than the classical plethysm and related problems in algebraic geometry and representation theory.

5.1 Plethysm problem

The multiplicities $G'(\lambda, m)$ defined in Section 3 are essentially the Kronecker coefficients [FH] in representation theory. These are defined as follows. Let $H = GL_n(\mathbb{C}) \times GL_n(\mathbb{C})$ be embedded naturally in $G = GL(\mathbb{C}^n \otimes \mathbb{C}^n)$. Given partitions λ, μ and π, the Kronecker coefficient $k^{\pi}_{\lambda,\mu}$ is the multiplicity of the irreducible H-module $V_\lambda(GL_n(\mathbb{C})) \otimes V_\mu(GL_n(\mathbb{C}))$ in the G-module $V_\pi(G)$, considered as an H-module via the natural embedding of H in G. Since, by (3), the stabilizer $\hat{G}_{det} \subseteq \hat{G} = SL_l(\mathbb{C})$ of $\det(Y) \in V$ is (modulo the discrete part) $SL(\mathbb{C}^m) \times SL(\mathbb{C}^m)$ embedded naturally in $SL(\mathbb{C}^m \otimes \mathbb{C}^m)$, it can be shown that $G'(\lambda, m)$ is essentially $k^{\lambda}_{\delta,\delta}$, where $\delta = (m, \ldots, m)$–the partition with m parts each of size m.

The Kronecker coefficients, in turn, are known to be special cases of the fundamental plethysm constants in representation theory [FH]. Given partitions λ, μ and π, the plethysm constant $a^{\pi}_{\lambda,\mu}$ is the multiplicity of the irreducible representation (Weyl module) $V_\pi(H)$ of $H = GL_n(\mathbb{C})$ in the irreducible representation $V_\lambda(G)$ of $G = GL(V_\mu)$, where $V_\mu = V_\mu(H)$ is an irreducible representation of H. Here $V_\lambda(G)$ is considered as an H-module via the representation map

$$\rho : H \to G = GL(V_\mu). \tag{15}$$

The coefficients $G'(\lambda, m)$ are special cases of the plethysm constants, whereas the coefficients $F(\lambda, n, m)$ and $G(\lambda, m)$ are akin to the plethysm constants, but much harder. Thus the plethysm constants are basic prototypes of the multiplicities $G'(\lambda, m)$, $F(\lambda, n, m)$ and $G(\lambda, m)$.

Let us define the bitlength of the input specification of $a^{\pi}_{\lambda,\mu}$ as

$$\langle \lambda, \mu, \pi \rangle = \langle \lambda \rangle + \langle \mu \rangle + \langle \pi \rangle + \min\{\dim(V_\mu), |\lambda|\}.$$

It follows from Klimyk's formula (cf. page 428 in [FH]) that $a^{\pi}_{\lambda,\mu}$ can be expressed as a difference between two $\#P$-formulae:

$$a^{\pi}_{\lambda,\mu} = \sum_a \chi_1(a) - \sum_a \chi_2(a),$$

where a ranges over bitstrings of poly$(\langle \lambda, \mu, \pi \rangle)$ bitlength, and χ_1 and χ_2 are poly$(\langle \lambda, \mu, \pi \rangle)$-time computable 0-1 functions. There is a similar formula for the Kronecker coefficient $k^{\pi}_{\lambda,\mu}$ defining the bitlength of its specification to be $\langle \lambda \rangle + \langle \mu \rangle + \langle \pi \rangle$.

The following is the analogue of Determinant and Permanent FH (Hypothesis 3.2) for the plethysm constants.

Hypothesis 5.1 (Plethysm FH) *(1) There is a $\#P$-formula for the plethysm constants $a^{\pi}_{\lambda,\mu}$ (and hence for the Kronecker coefficients).*

(2) The problem of deciding nonvanishing of plethysm constants (and hence, in particular, Kronecker coefficients) belongs to P. This means, given partitions λ, μ and π, whether $a^{\pi}_{\lambda,\mu}$ is nonzero can be decided in poly$(\langle \lambda, \mu, \pi \rangle)$ time.

Let $\tilde{a}^{\pi}_{\lambda,\mu}(k) = a^{k\pi}_{k\lambda,\mu}$ be the stretching function associated with the plethysm constant $a^{\pi}_{\lambda,\mu}$. Note that μ is not stretched here. It was asked in [Ki] if it is a polynomial. It can be shown that this is not so, even in the special case of Kronecker coefficients. But:

Theorem 5.2 *The stretching function $\tilde{a}^{\pi}_{\lambda,\mu}(k)$ is a quasi-polynomial.*

This generalizes Theorem 4.1 (1), since $G'(\lambda, m)$ is essentially a special case of the Kronecker coefficient.

The following is the analogue of Determinant and Permanent PH1 (Hypothesis 4.2) for the plethysm constants. It implies a $\#P$ formula for $a^{\pi}_{\lambda,\mu}$ as per Plethysm FH (1) (Hypothesis 5.1).

Hypothesis 5.3 (Plethysm PH1) *There exists an explicit possibly nonhomogeneous polytope $P(k) = P^{\pi}_{\lambda,\mu}(k)$ with specification bitlength $\langle P \rangle =$ poly$(\langle \lambda, \mu, \pi \rangle, \langle k \rangle)$ such that*

$$\tilde{a}^{\pi}_{\lambda,\mu}(k) = f_P(k), \tag{16}$$

the asymptotic Ehrhart quasi-polynomial of P (with deviation from quasipolynomiality zero), and P has a specification of the form

$$Ax \leq bk + c,$$

where A depends only on μ and n (the rank of the group $H = GL_n(\mathbb{C})$ in the definition of the plethysm constant), but not on λ and π, and b and c are piecewise homogeneous linear functions of λ and π.

In particular,

$$a^{\pi}_{\lambda,\mu} = \#(P(1)), \tag{17}$$

where $\#(P(1))$ denotes the number of integer points in $P(1)$.

It is known [Ki] that $f_P(k)$ need not satisfy the Ehrhart reciprocity [St3] that the Ehrhart quasipolynomials of homogeneous polytopes must satisfy. This is so even for Kronecker coefficients. Hence $P(k)$ need not be homogeneous in general.

PH1 is a complexity theoretic version of the fundamental plethysm problem in representation theory [FH] that has been intensively studied in the last century and is known to be formidable. And now, as we can see, it lies at the heart of this approach towards the P vs. NP problem. In

the classical plethysm problem [FH] the complexity theoretic issue of explicitness in Plethysm PH1 was not addressed. This is crucial here.

The quasi-polynomial $\tilde{a}^{\pi}_{\lambda,\mu}(k)$ need not be strictly saturated or positive, contrary to what was conjectured in the earlier version [GCT6] of this paper, even for the special case of Kronecker coefficients [BOR]. But its positivity and saturation indices are conjecturally small (as was also verified in [BOR] for the Kronecker coefficient $k^{\pi}_{\lambda,\mu}$ when the heights of λ and μ are at most two, and that of π at most three):

Hypothesis 5.4 (Plethysm SH) *The saturation index (Definition 4.4) of $\tilde{a}^{\pi}_{\lambda,\mu}(k)$ is poly$(\langle \lambda, \mu, \pi \rangle)$.*

This is an analogue of Permanent and Determinant SH for plethysm constants. It follows from the following stronger:

Hypothesis 5.5 (Plethysm PH2) *The positivity index (Definition 4.3) of $\tilde{a}^{\pi}_{\lambda,\mu}(k)$ is poly$(\langle \lambda, \mu, \pi \rangle)$.*

The following result says that a relaxed form of Plethysm FH (2) (Hypothesis 5.1) holds assuming Plethysm PH1 and Plethysm SH.

Theorem 5.6 *Assuming Plethysm PH1, and given λ, μ, π, and k' greater than the saturation index of $\tilde{a}^{\pi}_{\lambda,\mu}(k)$ (polynomially bounded as in Plethysm SH), whether $a^{k'\pi}_{k'\lambda,\mu}$ is nonzero can be decided in poly$(\langle \lambda, \mu, \pi \rangle, \langle k' \rangle)$ time.*

This is proved in Section 7.3.

5.2 Subgroup restriction problem

The plethysm constants can be generalized further as follows.

Let H and G be connected reductive groups, and $\rho : H \to G$ a homomorphism. Here H will generally be a subgroup of G, and ρ its embedding. Let $V_{\pi}(H)$ be an irreducible representation of H, and $V_{\lambda}(G)$ an irreducible representation of G. Here π and λ denote dominant weights of H and G. Let m^{π}_{λ} be the multiplicity of $V_{\pi}(H)$ in $V_{\lambda}(G)$, considered as an H-module via ρ. The plethysm constant is its special case obtained by letting $H = GL_n(\mathbb{C})$, $G = GL(V_{\mu}(H))$, and ρ the representation map (15).

We associate with m^{π}_{λ} the stretching function

$$\tilde{m}^{\pi}_{\lambda}(n) = m^{n\pi}_{n\lambda}. \tag{18}$$

The following is a generalization of Theorem 5.2.

Theorem 5.7 *The stretching function $\tilde{m}^{\pi}_{\lambda}(n)$ is a quasi-polynomial function of n.*

This is proved in Section 6.

One can also formulate analogues of PH1, SH, PH2, and FH for m^{π}_{λ} when H and $\rho : H \to G$ are explicitly given–we omit the details.

5.3 Multiplicities associated with G-varieties

The multiplicities m_λ^π can be generalized further as follows. The resulting generalization include all the multiplicities in this paper: $a_{\mu,\pi}^\lambda, m_\lambda^\pi, F(\lambda,n,m), G(\lambda,m)$, and $G'(\lambda,m)$.

Let H be a connected reductive group, X a projective H-variety i.e., a variety with H-action. Let ρ denote this H-action. Let $R = \oplus_d R_d$ be the homogeneous coordinate ring of X. Let $V_\pi(H)$ be an irreducible representation of H, where π denotes a dominant weight of H. Let s_d^π be the multiplicity of $V_\pi(H)$ in R_d, considered as an H-module via the action ρ.

We associate with s_d^π the stretching function:

$$\tilde{s}_d^\pi(n) = s_{nd}^{n\pi}. \tag{19}$$

Then:

Theorem 5.8 *(1) The stretching function $\tilde{s}_d^\pi(n)$ is an asymptotic quasi-polynomial.*
(2) It is a quasipolynomial if spec(R) is normal with rational singularities.

This is proved in Section 6.

Lemma 5.9 *(a) Theorem 5.8 (2) implies Theorem 5.7 (and hence also Theorem 5.2).*
(b) Theorem 5.8 (1) implies Theorem 4.1 (2).

Proof: (a) Observe that the multiplicity m_λ^π in Section 5.2 is a special case s_π^d. To see this, let H, ρ and G be as in Section 5.2, and let X be the closed G-orbit of the point v_λ corresponding to the highest weight vector of $V_\lambda(G)$ in the projective space $P(V_\lambda(G))$. Then

$$X = Gv_\lambda \cong G/P_\lambda, \tag{20}$$

where $P = P_\lambda = G_{v_\lambda}$ is the parabolic stabilizer of v_λ. We have a natural action of H on X via ρ. Let R be the homogeneous coordinate ring of X. By the Borel-Weil theorem [FH], the degree one component R_1 of R is $V_\lambda(G)$. Hence, s_1^π in this special case is precisely m_λ^π in Section 5.2. By [MR, R, S] (e.g. see Theorem 3.1 in [S]), spec(R) is normal and its singularities are rational. Now (a) follows.

(b) Observe that $F(\lambda,n,m)$ and $G(\lambda,m)$ are special cases of s_π^d, $d = |\lambda|/m$, by letting X be $\Delta_V[f,n,m]$ or $\Delta[g,m]$ and H be $G = GL(\mathcal{Y}) = GL_{m^2}(\mathbb{C})$ as in Section 2. Now (b) follows. Q.E.D.

One can also formulate analogues of PH1, SH, PH2, FH when X is the orbit closure of a point that is characterized by an explicitly given stabilizer–we omit the details.

6 Quasipolynomiality

In this section we prove the basic quasi-polynomiality Theorems 4.1, 5.2, 5.7 and 5.8. They all follow from the following general result.

Let $R = \oplus_k R_d$ be a graded $\mathbb{C}$-algebra with an action of a reductive group H. Let H_0 be the connected component of H containing the identity. Let $H_D = H/H_0$ be its discrete component. Given a dominant weight π of H_0, we consider the module $V_\pi = V_\pi(H_0)$, an H-module with trivial action of H_D. Let s_d^π denote the multiplicity of the H-module V_π in R_d. Let $\tilde{s}_d^\pi(n)$ be the multiplicity of the H-module $V_{n\pi}$ in R_{nd}. This is a stretching function associated with the multiplicity s_d^π. Let $S_d^\pi(t) = \sum_{n\geq 0} \tilde{s}_d^\pi(n) t^n$ be the generating function of $\tilde{s}_d^\pi(n)$.

Theorem 6.1 *(a) (Rationality) The generating function $S_d^\pi(t)$ is rational, and more strongly, $\tilde{s}_d^\pi(n)$ is an asymptotic quasipolynomial (cf. Section 4.2).*

Now assume that spec(R) is normal and that its singularities are rational. Then:

(b) (Quasi-polynomiality) The stretching function $\tilde{s}_d^\pi(n)$ is a quasi-polynomial.

(c) (Positivity) The rational function $S_d^\pi(t)$ can be expressed in a positive form:

$$S_d^\pi(t) = \frac{h_0 + h_1 t + \cdots + h_k t^k}{\prod_j (1 - t^{a(j)})^{k(j)}}, \tag{21}$$

where $a(j)$'s and $k(j)$'s are positive integers, $\sum_j k(j) = k + 1$, where k is the degree of the quasi-polynomial $\tilde{s}_d^\pi(n)$, $h_0 = 1$, and h_i's are nonnegative integers.

Theorem 5.8 follows from this result by letting R be the homogeneous coordinate ring of X as in Section 5.3. By Lemma 5.9, Theorems 5.7, 5.2 and 4.1 (2) follow as well. Theorem 4.1 (1) follows similarly, since $G'(\lambda, m)$ is essentially a special case of the plethysm (Kronecker) constant.

6.1 Proof of Theorem 6.1

The proof is an extension of M. Brion's proof (cf. page 520 in [D])) of quasi-polynomiality of the stretching function associated with a Littlewood-Richardson coefficient of any semisimple Lie algebra.

(b): Assume that spec(R) is normal and that its singularities are rational.

Let C_d be the cyclic group generated by the primitive root ζ of unity of order d. It has a natural action on R: $x \in C_d$ maps $z \in R_k$ to $x^k z$. Let $B = R^{C_d} = \sum_{n\geq 0} R_{nd} \subseteq R$ be the subring of C_d-invariants. Since spec(R) is normal and has rational singularities, by Boutot [Bt], B is also a normal $\mathbb{C}$-algebra and spec(B) has rational singularities.

Assume that H_0 is semisimple; extension to the reductive case being easy. Let π^* be the dominant weight of H_0 such that $V_\pi^* = V_{\pi^*}$; here V_π^* denotes the dual of V_π. By the Borel-Weil theorem [FH],

$$C_{\pi^*} := \oplus_{n\geq 0} V_{n\pi}^* = \oplus_{n\geq 0} V_{n\pi^*},$$

is the homogeneous coordinate ring of the H_0-orbit of the point $v_{\pi^*} \in P(V_{\pi^*})$ corresponding to the highest weight vector of V_{π^*}. This H_0-orbit is isomorphic to H_0/P_{π^*}, where $P_{\pi^*} \subseteq H_0$ is the parabolic stabilizer of v_{π^*}. Hence C_{π^*} is normal and spec(C_{π^*}) has rational singularities; cf. [MR, R, S] (e.g. see Theorem 3.1 in [S]). It follows that $B \otimes C_{\pi^*}$ is also normal, and

spec($B \otimes C_{\pi^*}$) has rational singularities. Consider the action of $\mathbb{C}^*$ on $B \otimes C_{\pi^*}$ given by:

$$x(b \otimes c) = (x \cdot b) \otimes (x^{-1} \cdot c),$$

where $x \in \mathbb{C}^*$ maps $b \in B_n$ to $x^n b$, the action on $\mathbb{C}_{\pi^*}$ being similar. Consider the invariant ring

$$S = (B \otimes C_{\pi^*})^{\mathbb{C}^*} = \oplus_n S_n = \otimes_{n \geq 0} R_{nd} \otimes V_{n\pi}^*. \tag{22}$$

By Boutot [Bt], it is a normal, and spec(S) has rational singularities.

Since $V_{n\pi}$ is an H-module, the algebra S has an action of H. Let

$$T = S^H = \oplus_{n \geq 0} T_n \tag{23}$$

be its subring of H-invariants. By Boutot [Bt], it is normal, and spec(T) has rational singularities–this is the crux of the proof. By Schur's lemma, the multiplicity of the trivial H-representation in $S_n = R_{nd} \otimes V_{n\pi}^*$ is precisely the multiplicity $\tilde{s}_d^\pi(n)$ of the H-module $V_{n\pi}$ in R_{nd}. Hence, the Hilbert function of T, i.e., $\dim(T_n)$, is precisely $\tilde{s}_d^\pi(n)$, and the Hilbert series $\sum_{n \geq 0} \dim(T_n) t^n$ is $S_d^\pi(t)$. Quasipolynomiality of $\tilde{s}_d^\pi(n)$ now follows by applying the following lemma:

Lemma 6.2 *(cf. Lemma 5.4 in [D] and also [F]) If $T = \oplus_{n=0}^{\infty} T_n$ is a graded $\mathbb{C}$-algebra, such that spec(T) is normal and has rational singularites, then* $\dim(T_n)$*, the Hilbert function of T, is a quasi-polynomial function of n.*

(c): Since spec(T) has rational singularities, T is Cohen-Macauley. Let $t_1, \ldots, t_u$ be its homogeneous sequence of parameters (h.s.o.p.), where $u = k + 1$ is the Krull dimension of T. By the theory of Cohen-Macauley rings [St], it follows that its Hilbert series $S_d^\pi(t)$ is of the form

$$\frac{h_0 + h_1 t + \cdots + h_k t^k}{\prod_{i=1}^{k+1} (1 - t^{d_i})}, \tag{24}$$

where (1) $h_0 = 1$, (2) d_i is the degree of t_i, and (3) h_i's are nonnegative integers. This proves (c).

(a): A careful examination of the proof of (b) shows that T is a finitely generated ring for an arbitrary graded $\mathbb{C}$-algebra R with the action of a reductive H–this follows from Hilbert's classical result on finite generation of the algebra of invariants of a reductive-group action. (Boutot's result is not required here.) Now rationality of $S_d^\pi(t)$, and more strongly, asymptotic quasi-polynomiality of $\tilde{s}_d^\pi(n)$ as $n \to \infty$, follows from Hilbert's another classical result since $\tilde{s}_d^\pi(n)$ is the Hilbert function of the finitely generated ring T.

This proves Theorem 6.1. Q.E.D.

7 Saturated integer programming

Integer programming problem is NP-complete. In this section we give (cf. Section 7.1) a polynomial time algorithm for its special case, called *saturated integer programming*, and use it (cf. Section 7.3) to prove Theorem 4.9, 4.8 and 5.6. We also prove (Section 7.2) a worst case upper bound for the saturation index of an Ehrhart quasipolynomial of a polytope. These results together say that the saturation index of the Ehrhart quasipolynomial of a polytope is a good measure of the computational complexity of the associated integer programming problem.

7.1 A polynomial time algorithm

In the saturated integer programming problem we are given an explicit parametrized polytope $P(k)$ (possibly nonhomogeneous) specified as a separation oracle (cf. Section 4.2), a nonnegative integer $\mathrm{sie}(P)$, called a *saturation index estimate*, guaranteed to be higher than or equal to the the saturation index $s(f_P)$ (Definition 4.5) of the asymptotic Ehrhart quasipolynomial $f_P(k)$ of $P(k)$, and an integer $k' > \mathrm{sie}(P)$. The problem is to decide if $P(k')$ contains an integer point.

Theorem 7.1 *The saturated integer programming problem above can be solved in poly*$(\langle P(k')\rangle, \langle k'\rangle)$ *time, where* $\langle P(k')\rangle$ *denotes the bitlength of the specification of* $P(k')$ *in the form of a separation oracle (cf. Section 4.2),* $\langle k'\rangle$ *the bitlength of* k'*.*

Proof: Let $f_P(k)$ be the asymptotic Ehrhart quasi-polynomial of $P(k)$. Let $\mathrm{span}(P(k))$ denote the affine span of $P(k)$, i.e., the smallest dimensional affine space containing $P(k)$. It follows from the definitions of the saturation index $s(f_P)$ (Definition 4.5) and the defect $\delta(P)$ that $\dim(P(k))$ remains the same for all $k > s(f_P)$ and the equations of $\mathrm{span}(P(k))$ are stable for $k > s(f_P)$; i.e., there exist an integral matrix C and integral vectors d and e such that $\mathrm{span}(P(k))$, for any $k > s(f_P)$, is defined by

$$Cx = dk + e. \tag{25}$$

Lemma 7.2 *Suppose* $k > s(f_P)$*. Then* $P(k)$ *contains an integer point iff span*$(P(k))$ *contains an integer point.*

Before proving the lemma, let us prove Theorem 7.1 using it.

First we decide if $P(k')$ is nonempty in $\mathrm{poly}(\langle P(k')\rangle, \langle k'\rangle)$ time using the GLS (Grötschel, Lovász and Schrijver) algorithm for linear programming [GLS] over polytopes given in the form of a separation oracle (cf. Theorem 6.4.1 in [GLS]). If $P(k')$ is empty, then $P(k')$ does not contain an integer point, and we can stop. So assume that it is nonempty.

A simple extension of the GLS algorithm also yields specifications of C, d and e in (25) in $\mathrm{poly}(\langle P(k')\rangle, \langle k'\rangle)$ time (cf. Theorems 6.4.9, and 6.5.5 in [GLS]). This final specification of C, d and e is exact, even though the first part of the GLS algorithm in [GLS] uses the ellipsoid method. Indeed, the use of simultaneous diophantine approximation based on basis reduction in lattices in [GLS] is precisely to ensure this exactness in the final answer. This is crucial for the next step of our algorithm.

Since $k' > \mathrm{sie}(P) \geq s(f_P)$, by Lemma 7.2, it suffices to check if $\mathrm{span}(P(k'))$ contains an integer point; i.e., if the linear diophantine system (25) has an integral solution x for $k = k'$. This can be done in $\mathrm{poly}(\langle P(k')\rangle, \langle k'\rangle)$ time using a polynomial time algorithm for solving linear diophantine systems (cf. Corollary 5.4.9 in [GLS]). This proves Theorem 7.1.

Proof of Lemma 7.2: Clearly, if $P(k)$ contains an integer point, then $\mathrm{span}(P(k))$ contains an integer point. So assume that $k > s(f_P)$ and $\mathrm{span}(P(k))$ contains an integer point. We want to show that $P(k)$ contains an integer point.

Consider the system (25) defining $\mathrm{span}(P(k))$, $k > s(f_P)$. Let $\bar{C}$ be the Smith normal form of C; i.e., $\bar{C} = ACB$ for some unimodular matrices A and B, where the leftmost principal submatrix of $\bar{C}$ is a diagonal, integral matrix, and all other columns are zero; the matrices $\bar{C}, A$

and B can be computed in polynomial time using the algorithm in [KB] for computing the Smith normal form. After a unimodular change of coordinates, by letting $z = B^{-1}x$, span$(P(k))$ is specified by the linear system $\bar{C}z = \bar{d}k + \bar{e}$, where $\bar{d} = Ad$ and $\bar{e} = Ae$. The equations in this system are of the form:

$$\bar{c}_i z_i = \bar{d}_i k + \bar{e}_i, \tag{26}$$

$i \leq \text{codim}(P(k))$, for some integers $\bar{c}_i$, $\bar{d}_i$ and $\bar{e}_i$. By removing common factors if necessary, we can assume that $\bar{c}_i$, $\bar{d}_i$ and $\bar{e}_i$ are relatively prime for each i.

By (26), span$(P(k))$ contains an integer point for given $k > s(f_P)$ iff

$$\bar{d}_i k + \bar{e}_i = 0, \text{ mod } \bar{c}_i \tag{27}$$

for all i. If $\bar{c}_i$ and $\bar{d}_i$ have a common prime factor $p_i > 1$ for some i, then, since $\bar{e}_i$ is not divisible by p_i by our assumption, (27) cannot be satisfied for any integral k, and hence span$(P(k))$ cannot contain an integer point, contrary to our assumption. So assume without loss of generality that $\bar{c}_i$ and $\bar{d}_i$ are also relatively prime for all i. Then, for each i, there exist integers a_i and b_i such

$$a_i \bar{d}_i + b_i \bar{c}_i = 1, \tag{28}$$

and (27) implies

$$a_i \bar{d}_i k + a_i \bar{e}_i = (1 - b_i \bar{c}_i)k + a_i \bar{e}_i = 0 \text{ mod } \bar{c}_i. \tag{29}$$

Thus if span$(P(k))$ contains an integer point for $k > s(f_P)$ then

$$k = -a_i \bar{e}_i \text{ mod } \bar{c}_i, \tag{30}$$

for all i.

Let $f_{j,P}(k)$, $1 \leq j \leq l$, be the polynomials such that $f_P(k) = f_{j,P}(k)$ if $k = j \text{ mod } l$ and $k \geq a(f_P)$, the deviation from quasipolynomiality of $f_P(k)$.

Claim 7.3 *The polynomial $f_{j,P}(k)$ is not identically zero if $j = -a_i \bar{e}_i$, mod $\bar{c}_i$, for all i.*

Before proving the claim, let us verify that Lemma 7.2 follows from it. Suppose span$(P(k))$ contains an integer point for some $k > s(f_P)$. We want to show that $P(k)$ contains an integer point.

Since span$(P(k))$ contains an integer point, by (30), $k = -a_i \bar{e}_i$, mod $\bar{c}_i$, for all i. This means, $j = -a_i \bar{e}_i$, mod $\bar{c}_i$, for all i, for $j = k$, mod l, since, as we shall see below, l is divisible by each $\bar{c}_i$. By the claim, the polynomial $f_{j,P}$ is not identically zero for this j. Since $k > s(f_P)$, it follows from Definition 4.5 of the saturation index that $f_P(k) = f_{j,P}(k)$ is not zero for this j and k. That is, $P(k)$ contains an integer point. This proves Lemma 7.2.

Proof of Claim: Fix any r such that

$$r = -a_i \bar{e}_i, \quad \text{mod} \bar{c}_i \tag{31}$$

for all i. If such an r does not exist, then no j as in the claim can exist, and the claim is vacuously true. Let $Q(s) = P(s+r)$. Then, by (26), (28) and (31), the equations of span$(Q(s))$ are

$$\bar{c}_i z'_i = \bar{d}_i s, \tag{32}$$

for all i, where $z'_i = z_i + t_i$, for some fixed integer t_i. Since $\bar{c}_i$ and $\bar{d}_i$ are relatively prime for all i, span$(Q(s))$ contains an integer point iff

$$s = 0 \bmod \bar{c}_i, \tag{33}$$

for all i; i.e., if s is divisible by the l.c.m. $\tilde{c}$ of $\bar{c}_i$'s.

Let $f_Q(s)$ be the asymptotic Ehrhart quasipolynomial of $Q(s)$. It is just a shifted form of $f_P(k)$: $f_{t,Q}(s) = f_{t+r,P}(s+r)$, for $1 \leq t \leq l$–here $t+r$ is taken modulo l. So to show that $f_{j,P}(k)$ is not identically zero, when $j = -a_i\bar{e}_i$, mod $\bar{c}_i$, for all i, it suffices to show, in view of (31), that $f_{t,Q}(s)$ is not identically zero when $t = 0$, mod $\bar{c}_i$, for all i, i.e., when t is divisible by the l.c.m. $\tilde{c}$ of $\bar{c}_i$'s.

Now consider the dilated polytope $\bar{Q}(s) = Q(\tilde{c}s)$. Let $f_{\bar{Q}}(s)$ be its asymptotic Ehrhart quasipolynomial. By (33), $Q(s)$ contains no integer point unless $\tilde{c}$ divides s. Hence $\tilde{c}$ divides its period l. By eq.(32), the equations defining $\bar{Q}(s)$ are:

$$z'_i = \bar{d}_i(\tilde{c}/\bar{c}_i)s. \tag{34}$$

Thus $f_{t,Q}(s) = f_{t/\tilde{c},\bar{Q}}(s/\tilde{c})$, if $\tilde{c}$ divides t, and it is identically zero otherwise. Hence, to show that $f_{t,Q}(s)$ is not identically zero when t is divisible by $\tilde{c}$, it suffices to show that $f_{j,\bar{Q}}(s)$ is not identically zero for any $1 \leq j \leq l(\bar{Q})$, where $l(\bar{Q}) = l/\tilde{c}$ is the period of $f_{\bar{Q}}(s)$.

For a fixed j, this is equivalent to showing that $\bar{Q}(1)$ contains an integer point z' with $z'_i = x_i/b$, for some integers x_i's and b such that $b = j$ modulo $l(\bar{Q})$. Let us call such a point j-admissible. Because of the form that the equations (34) defining span$(\bar{Q}(s))$ take at $s = 1$, we can assume, without loss of generality, that $\bar{Q}(1)$ is full dimensional. This means the system (34) is empty. Then this follows from denseness of the set of j-admissible points–specifically, that any point in the interior of $\bar{Q}(1)$ can be approximated infinitesimally closely by a j-admissible point. This proves the claim. Q.E.D.

7.2 A general estimate for the saturation index

Now we give a general estimate for the saturation index of the Ehrhart quasipolynomial $f_P(k)$ (cf. Section 4.1) of any polytope P with a specification of the form

$$Ax \leq b, \tag{35}$$

where A is an $m \times n$ matrix, m possibly exponential in n. Let $\|P\| = n + \psi$, where ψ is the maximum bitlength of any entry of A. Trivially, $\|P\| \leq \langle P \rangle$. We do not assume that we know the specification (35) of P explicitly. We only assume that it exists, and that we are told $\|P\|$. Then:

Theorem 7.4 *The saturation index of $f_P(k)$ is $O(2^{poly(\|P\|)})$.*

Conjecturally, this also holds for the positivity index and also for the asymptotic Ehrhart quasi-polynomial of a nonhomogeneous polytope (Section 4.2).

In the worst case the saturation index of the Ehrhart quasipolynomial of a polytope is thus exponential and integer programming over a polytope also takes exponential time–nothing better is expected since this problem is NP-complete. On the other hand when the saturation index is small a relaxed form of integer programming can be solved in polynomial time (Theorem 7.1). In this sense the saturation index of a polytope is a good measure of the computational complexity of the associated integer programming problem.

Proof: Given a quasipolynomial $f(k)$, let $F(t)$ be its generating function

$$F(t) = \sum_{k \geq 0} f(k) t^k.$$

In the proof below we will use a basic fact [St3] that $F(t)$ is a rational function, whose numerator has smaller degree than the denominator, and the roots of the denominator are roots of unity.

Fix a polytope P. There exists a triangulation of P into simplices such that every vertex of any simplex is also a vertex of P. Then

$$f_P(k) = \sum_{\Delta} f_\Delta(k),$$

where Δ ranges over all open simplices in this triangulation; a zero-dimensional open simplex is a vertex. The saturation index of $f_P(k)$ is clearly bounded by the maximum of the saturation indices of $f_\Delta(k)$.

Hence, we can assume, without loss of generality, that P is an open simplex. Let $v_0, \ldots, v_n$ be its vertices. Then, by Ehrhart's result (cf. Theorem 1.3 in [St2]),

$$F_P(t) = \frac{\sum_i h_i t^i}{\prod_{j=0}^n (1 - t^{a_j})}, \tag{36}$$

where $h_0 = 1$, h_i's are nonnegative, and a_j is the least positive integer such that $a_j v_j$ is integral. By Cramer's rule, the bit length of each a_j is poly($\|P\|$). Without loss of generality, we can also assume that a_j's are relatively prime. Otherwise, the estimate on the saturation index below has to be multiplied by the g.c.d. of a_j's. Then the result follows by applying the following lemma to $F_P(t)$, since $\langle a_j \rangle = O(\text{poly}(\|P\|))$; i.e., $a_j = O(2^{O(\text{poly}(\|P\|))})$. Q.E.D.

Lemma 7.5 *Let $f(k)$ be a quasipolynomial whose generating function $F(t)$ has a positive form*

$$F(t) = \frac{\sum_i h_i t^i}{\prod_{j=0}^n (1 - t^{a_j})}, \tag{37}$$

where $h_0 = 1$, h_i's are nonnegative, and a_j's are positive and relatively prime. Let $a = \max\{a_j\}$. Then the saturation index $s(f)$ of $f(k)$ is $O(poly(a, n))$.

Proof: Let $g(k)$ be the quasi-polynomial whose generating function $G(t) = \sum g(k) t^k$ is $1/\prod_{j=0}^n (1 - t^{a_j})$. It is known that this is the Ehrhart quasipolynomial of the polytope $N(a_0, \ldots, a_n)$ defined by the linear system

$$\sum a_j x_j = 1, x_j > 0.$$

The saturation index $s(g)$ of $g(k)$ is bounded by the Frobenius number associated with the set of integers $\{a_j\}$–this is the largest positive integer m such that the diophantine equation

$$\sum_j a_j x_j = m$$

has no positive integral solution $(x_0, \ldots, x_n)$. It is known (e.g. [BDR]) that the Frobenius number is bounded by

$$\sum_j a_j + \sqrt{a_0 a_1 a_2 (a_0 + a_1 + a_2)} = O(\text{poly}(a)),$$

assuming that $a_0 \leq a_1 \ldots$. Hence, $s(g) = O(\text{poly}(a))$.

Since $f(k)$ is a quasi-polynomial, the degree of the numerator of $F(t)$ is less than the degree of the denominator [St3]. Thus the maximum value of i that occurs in (37) is an.

Let $g_i(k)$, $i \leq an$, be the quasi-polynomial whose generating function is $t^i / \prod_{j=0}^{n}(1 - t^{a_j})$. Then

$$s(g_i) \leq i + s(g) = O(\text{poly}(a, n)).$$

Since, h_i's in (37) are nonnegative, $s(f) = \max s(g_i)$. The result follows. Q.E.D.

7.3 Applications of saturated integer programming

Proof of Theorem 4.8

We shall only prove the result for $F_{\lambda,n,m}(k)$, the other cases being similar.

By Permanent PH1, there exists an explicit polytope $P(k) = P_{\lambda,n,m}(k)$ such that $F_{\lambda,n,m}(k) = f_P(k)$, and the bitlength $\langle P(k) \rangle$ for given k is $\text{poly}(n, m, \langle \lambda \rangle, \langle k \rangle)$. Fix k' greater than the saturation index of $F_{\lambda,n,m}(k)$. By Lemma 7.2, $P(k')$ contains an integer point iff $\text{span}(P(k'))$ contains an integer point. The latter can be solved in $\text{poly}(\langle P(k') \rangle, \langle k' \rangle) = \text{poly}(\langle \lambda \rangle, n, m, \langle k' \rangle)$ time using a polynomial time algorithm for solving linear diophantine systems (cf. Corollary 5.4.9 in [GLS]). Q.E.D.

Proof of Theorem 5.6

This is similar to that of Theorem 4.8.

Proof of Theorem 4.9

Assume that Permanent PH1 and Determinant PH1 (a) hold. Let k and λ be such that OH holds for given n and m.

Claim: $V_{k\lambda}(G)$ is an incidence-based geometric obstruction for given n and m.

Proof of the claim: By Determinant PH1 (a), $G_{\lambda,m}(k) = f_Q(k)$. Hence $G(k\lambda, m)$ is the number of integer points in $Q_{\lambda,m}(k)$. By OH (b), the affine span of $Q_{\lambda,m}(k)$ does not contain an integer point. That is, $G(k\lambda, m) = 0$. By OH (a), $\text{span}(P_{\lambda,n,m}(k))$ contains an integer point.

Furthermore, by OH, k is larger than the saturation index of the asymptotic Ehrhart quasipolynomial $F_{\lambda,n,m}(k)$ of $P_{\lambda,n,m}(k)$ (polynomially bounded in Permanent SH). Hence, by Lemma 7.2, $P_{\lambda,n,m}(k)$ contains an integer point. By Permanent PH1, $F(k\lambda, n, m)$ is the number of integer points in the polytope $P_{\lambda,n,m}(k)$. Hence $F(k\lambda, n, m) > 0$. Thus $V_{k\lambda}(G)$ is an incidence-based geometric obstruction by Definition 2.4. This proves the claim.

To prove explicitness of the obstruction family $\mathcal{O}$, we have to show that, given k, λ, n and m, whether k and λ satisfy OH for this n and m can be checked in poly$(\langle k\rangle, \langle\lambda\rangle, n, m)$ time. By Permanent PH1, the polytope $P(k) = P_{\lambda,n,m}(k)$ is explicit and its bitlength for given k is poly$(n, m, \langle\lambda\rangle, \langle k\rangle)$. Its affine span span$(P(k))$ can be computed in polynomial time by the GLS algorithm [GLS] for linear programming, and whether span$(P(k))$ contains an integer point can be decided in polynomial time by using the polynomial time algorithm for solving linear diophantine equations (cf. Corollary 5.4.9 in [GLS]). Thus OH (a) can be checked in polynomial time. Similarly, assuming Determinant PH1 (a), we can check in polynomial time if span$(Q_{\lambda,m}(k))$ contains an integer point. Hence OH (b) also can be checked in polynomial time. We assume that the polynomial bound on the saturation index in Permanent SH (Hypothesis 4.6) is explicitly given. Whether k is larger than this explicit bound is also easy to check. Thus OH can be checked in polynomial time for given k, λ, n and m.

Hence $\mathcal{O}$ is explicit. Q.E.D.

8 Evidence for the plausible converse

In this section we justify the dotted right arrow in (1).

8.1 Why should geometric obstructions exist and general FH hold?

We begin with justification of GOH, the Equivalence Conjecture 2.6, and General FH (Hypothesis 3.1) on the basis of the proof of the Strong Flip Theorem 2.3 in [GCTflip].

This proof is based on: (1) hardness of the permanent (Conjecture 2.2), which is the first assumption in Theorem 2.3, (2) the characterization by symmetries of the permanent (the property (P)), (3) easiness of computing the determinant, and (4) the characterization by symmetries of the determinant (the property (D)), which is needed in all efficient computations of the determinant (e.g. the Guassian elimination). The proof shows that if $\Delta_V[f, n, m] \not\subseteq \Delta_V[g, m]$ then, assuming the additional derandomization conjecture in Theorem 2.3, there exist short and easy-to-compute proof certificates of this noninclusion, namely, the global obstruction sets $S_{n,m}(s)$. The dependence of these obstructions on the representation theoretic characterization by symmetries of the permanent and the determinant is only *indirect* via the proof of Theorem 2.3 and *extrinsic*; i.e., it depends on the embeddings of $\Delta_V[f, n, m]$ and $\Delta_V[g, m]$ in $P(V)$. It is a reasonable conjecture that there exist similar short and easy-to-compute representation theoretic obstructions with *direct* dependence on the *intrinsic* representation theoretic structures of $\Delta_V[f, n, m]$ and $\Delta_V[g, m]$. That is, whatever is extrinsic and indirect can conjecturally be made intrinsic and direct.

Now we have to specify what we mean by the intrinsic representation theoretic structure of $\Delta_V[f, n, m]$ and $\Delta_V[g, m]$. By intrinsic, we mean the representation theoretic structures of

their coordinate rings $R_V[f,n,m]$ and $R_V[g,m]$. The most detailed such structure is provided the multiplicative structure of these rings. But this multiplicative structure is hard to compute. This is so even for the multiplicative structure of the much simpler coordinate ring $\mathbb{C}[G]$ of G. The most favourable basis of $\mathbb{C}[G]$ from the representation theoretic perspective is the canonical basis whose multiplicative structure constants are known to be nonnegative; cf. [Lu]. These structure constants are also known [FKK] to be generalizations of the Kazdan-Lusztig polynomials evaluated at $q = 1$. But the problem of computing values of the Kazdan-Lusztig polynomials at $q = 1$ is $\#P$-hard. In fact, even the much easier problem of computing Littlewood-Richardson coefficients (which are [LT] values of very special kinds of Kazdan-Lusztig polynomials at $q = 1$) is known to be $\#P$-complete [N]. Hence, this computation cannot be done in polynomial time, assuming the standard complexity theory conjecture that $P^{\#P} \not\subseteq P$ (or equivalently, that the permanent cannot be computed in polynomial time.) Since the rings $R_V[f,n,m]$ and $R_V[g,m]$ are much harder than the coordinate ring $\mathbb{C}[G]$, their multiplicative structures are even harder. This means the easy-to-compute representation theoretic obstructions predicted by (the proof of) the strong flip Theorem 2.3 cannot depend on the detailed multiplicative structures of the rings $R_V[f,n,m]$ and $R_V[g,m]$, since these structures are hard to compute.

So to locate such obstructions, we ignore the detailed multiplicative structures and consider instead the much coarser representation theoretic data consisting of the multiplicities of $V_\lambda(G)$ in $R_V[f,n,m]^*$ and $R_V[g,m]^*$ for all λ. The multiplicity based geometric obstructions (Definition 2.4) are precisely the representation theoretic obstructions based on this data.

But even these multiplicities are hard to compute, since they are much harder than the Littlewood-Richardson coefficients, which, as remarked above, are $\#P$-complete [N]. Thus the easy-to-compute representation theoretic obstructions predicted by the strong flip Theorem 2.3 cannot depend on the exact values of the multiplicities if both the multiplicities are large.

So to locate them, we consider even coarser representation theoretic data, namely the threshold data, which specifies, for each $V_\lambda(G)$ and small $k = O(\text{poly}(n,m))$, whether its multiplicity in $R_V[f,n,m]^*$ exceeds k (and similarly for $R_V[g,m]^*$). The threshold-based geometric obstructions are precisely the representation theoretic obstructions based on this data. Fortunately, the fundamental obstacle to efficient computation of the multiplicative or multiplicity data–namely, $\#P$-completeness of the Littlewood-Richardson coefficients–is absent for the threshold data. This is because whether a Littlewood-Richardson coefficient $c^\lambda_{\alpha,\beta} > k$ can be decided in polynomial time, specifically in $\text{poly}(\langle\alpha\rangle, \langle\beta\rangle, \langle\lambda\rangle, k)$ time, where $\langle\alpha\rangle$ denotes the bitlength of the specification of α; cf. [GCT3] for an algorithm for $k = 0$ and [BI] for an algorithm which can be extended to work for general k. Thus the strong flip Theorem 2.3, in conjunction with the argument above, leads to the conjecture that the threshold based geometric obstructions are precisely the short and easy-to-compute intrinsic representation theoretic obstructions predicted by (the proof of) the strong flip Theorem 2.3: specifically, (1) threshold-based geometric obstructions exist as per the Equivalence Conjecture 2.6, and (2) the underlying decision problems can be solved in polynomial time, so that these obstructions are also easy to compute, verify, and decode–this is precisely the General FH (Hypothesis 3.1) for threshold-based obstructions. Thus (1) and (2) together constitute an "intrinsic form" of the (conclusion of the) strong flip Theorem 2.3.

The threshold data is close to the incidence data when m is small ($O(\text{poly}(n))$). This leads to GOH for small m and General FH for incidence-based obstructions.

Here it is important that the group G in the definition of obstructions (Definition 2.4) is $GL_l(\mathbb{C})$. Equivalently, if we let G be $SL_l(\mathbb{C})$, then it is important to consider (as in Definition 1.2 in [GCT2]) the finer representation theoretic grading information which gives, for each partition λ, the multiplicity of the Weyl module $V_\lambda(SL_l(\mathbb{C}))$ of $SL_l(\mathbb{C})$ in $R_V[f,n,m]_d$ and $R_V[g,m]_d$, for each d. If we replace $GL_l(\mathbb{C})$ by $SL_l(\mathbb{C})$ and ignore this grading information, the incidence based obstructions need not exist. That is, we cannot simplify the incidence data used in Definition 2.4 any further. This is the main result of [BI2] for GCT of matrix multiplication. The article [BLMW] also points out importance of the grading information.

Remark 1: One may also wonder why we did not consider geometric obstructions based approximate multiplicities of $V_\lambda(G)$ in $R_V[g,m]^*$ and $R_V[f,n,m]^*$. It is known that approximate values of well behaved $\#P$-complete quantities (such as the permanent of a nonnegative integer matrix [JSV]) can be computed efficiently in probabilistic polynomial time. So it is plausible that even obstructions based on approximate multiplicities are easy to compute. But the proof of the strong flip Theorem 2.3 shows that, whenever $\Delta_V[f,n,m] \not\subseteq \Delta_V[g,m]$, there exist short global obstruction sets $S_{n,m}(s)$ that can be computed not just in polynomial time, but also fast in parallel: specifically, in polylogarithmic time using polynomial number of processors. The standard complexity theory conjecture is that approximate values of $\#P$-complete quantities cannot be computed fast in parallel (when they are large). Hence, obstructions based on approximate multiplicities cannot be the representation theoretic obstructions that are easy to compute in parallel as predicted by the proof of the strong flip Theorem 2.3.

Remark 2: The assumption $P^{\#P} \not\subseteq P$ in the justification in this section can be replaced by Conjecture 2.1 replacing easy-to-compute everywhere by easy-to-compute-in-parallel.

8.2 Why should Determinant and Permanent FH hold?

We have already justified General FH (Hypothesis 3.1) above. For efficient verification of geometric obstruction labels as in F3 in General FH, we need efficient criteria for deciding nonvanishing of $F(\lambda,n,m)$ and $G(\lambda,m)$; cf. Proposition 3.3. This leads to Determinant and Permanent FH (2) (Hypothesis 3.2).

We now justify Determinant FH (1) for the multiplicities $G'(\lambda,m)$, assuming that Determinant FH (2) holds as argued above. The argument for the other multiplicities $G(\lambda,m)$ and $F(\lambda,n,m)$ is similar.

It can be shown using the Weyl character formula that $G'(\lambda,m)$ can be expressed as the difference between two $\#P$-quantities. If $G'(\lambda,m)$ does not belong to $\#P$, then to decide if it is zero, one would have to compute these two quantities and hence $G'(\lambda,m)$ exactly. But the problem of computing $G'(\lambda,m)$ is $\#P$-hard, since the much easier problem of computing Littlewood-Richardson coefficients is $\#P$-complete [N]. Hence exact computation of $G'(\lambda,m)$ cannot be done in polynomial time assuming the standard complexity theory conjecture that $P^{\#P} \not\subseteq NP$. But Determinant FH (2) says that the problem of deciding if $G'(\lambda,m)$ is zero belongs to P. This is why Determinant FH (1) for $G'(\lambda,m)$ is conjectured to hold.

Remark: Just as in Section 8.1, the assumption $P^{\#P} \not\subseteq P$ in the justification above and below can be replaced by Conjecture 2.1 replacing easy-to-compute everywhere by easy-to-compute-in-parallel.

8.3 Why should positivity hold?

Now we justify the positivity hypotheses PH1 and SH (Hypotheses 4.2 and 4.6) assuming FH. We only give the argument for $G'(\lambda, m)$, the arguments for $G(\lambda, m)$ and $F(\lambda, n, m)$ being similar.

By Determinant FH (1), $G'(\lambda, m)$ belongs to $\#P$. By the proof of NP-completeness of the integer programming problem, it follows that $G'(\lambda, m)$ can be written as the number of integer points in an explicit polytope $Q'_{\lambda,m}$. PH1 for $G'(\lambda, m)$ (i.e. Determinant PH1 (b)) basically says just this with an additional naturality constraint that the asymptotic Ehrhart quasipolynomial of $Q'_{\lambda,m}(k)$ should coincide with the asymptotic stretching quasipolynomial (Theorem 4.1) of $G'(\lambda, m)$ (as in the case of the Littlewood-Richardson coefficients [BZ, DM]).

Once PH1 holds for $G'(\lambda, m)$, deciding its nonvanishing is an integer programming problem. Since integer programming is NP-complete, there is no polynomial time algorithm for deciding nonvanishing $G'(\lambda, m)$ (assuming $P \neq NP$) unless the polytope $Q'_{\lambda,m}(k)$ in Determinant PH1 (b) is exceptional. By the results in Section 7 the saturation index of the Ehrhart quasipolynomial of a polytope is a good measure of the computational complexity of the associated integer programming problem. Hence, if Determinant FH (2) holds as argued above, that is, if the problem of deciding if $Q'_{\lambda,m}(k)$ has an integer point belongs to P, then it is a reasonable conjecture that the saturation index of the asymptotic Ehrhart quasipolynomial of $Q'_{\lambda,m}(k)$ is small (polynomial). This is what SH for $G'(\lambda, m)$ (i.e., Determinant SH (2)) says.

8.4 Why should OH hold?

We have already justified GOH in Section 8.1. By the following result, OH is close to GOH.

Lemma 8.1 *Assume PH1 (Hypothesis 4.2). If $k\lambda$ is an incidence-based geometric obstruction label for some integer k greater than the saturation indices of $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ (which by SH are small, i.e., $O(poly(n, m))$), then (λ, k) satisfies OH.*

Proof: If $k\lambda$ is a geometric obstruction label, then by Definition 2.4, $G_{\lambda,m}(k)$ is zero and $F_{\lambda,n,m}(k)$ is nonzero. By PH1 (Hypothesis 4.2), $G_{\lambda,m}(k) = f_Q(k)$, where $Q = Q_{\lambda,m}(k)$, and $F_{\lambda,n,m}(k) = f_P(k)$, where $P = P_{\lambda,n,,m}(k)$. Hence, if $k\lambda$ is a geometric obstruction label, the polytope $P = P_{\lambda,n,,m}(k)$ contains an integer point and $Q_{\lambda,m}(k)$ does not. Since k is greater than the saturation index of $G_{\lambda,m}(k)$, by Lemma 7.2, the affine span of $Q_{\lambda,m}(k)$ does not contain an integer point. The affine span P contains an integer point since P does. Hence (λ, k) satisfies OH. Q.E.D.

This finishes the justification for the dotted arrow $\cdots >$ in the decomposition (1).

9 How to prove positivity?

In this section, we formulate additional positivity hypotheses, called PH0, which suggest an approach to prove PH1.

To state PH0, we need a definition.

Definition 9.1 *Let H be a connected reductive algebraic subgroup of connected reductive algebraic G, and $\mathcal{H}$ its Lie algebra with the standard generators (e_i, f_i and h_i) as described in [FH]. A basis B of a representation $V = V_\lambda(G)$ of G, where λ is the highest weight of G, is called* positive *with respect to the H-action if:*

1 *It is H-compatible. This means there exists a filtration of B:*

$$B = B_0 \supset B_1 \supset \cdots$$

such that $\langle B_i\rangle/\langle B_{i+1}\rangle$, where $\langle B_i\rangle$ denotes the linear span of B_i, is isomorphic to an irreducible H-module.

2 *Each $b \in B$ has a combinatorial indexing label whose bitlength $\langle b\rangle$ is polynomial in $\langle\lambda\rangle$ and the rank $r(G)$ of G. Furthermore, each $b \in B$ is a weight vector for the $\mathcal{H}$-action. Thus b is a highest weight vector if $e_i b$ is zero for all i.*

3 *For each standard generator h of $\mathcal{H}$ and each $b \in B$,*

$$hb = \sum_{b' \in B} (-1)^{d^h_{b,b'}} c^h_{b,b'} b, \tag{38}$$

where each $c^h_{b,b'}$ is a nonnegative integer, and $d^h_{b,b'}$ is an integer that can be computed in poly$(\langle b\rangle, \langle b'\rangle, r(G))$ time.

We call B an explicit positive basis *if, in addition,*

a *Each $c^h_{b,b'}$ has a $\#P$-formula, and its nonvanishing can be decided in time poly$(\langle b\rangle, \langle b'\rangle, r(G))$ time.*

b *Whether a given $b \in B$ is a highest weight vector can be decided in poly$(\langle b\rangle, \langle\lambda\rangle, r(G))$ time.*

We call a positive basis strictly positive if $d^h_{b,b'}$ are all zero.

An explicit positive basis of any finite dimensional representation V of G with respect to the H action can be defined similarly as long as V has a compact combinatorial specification, akin to the specification λ of $V_\lambda(G)$. In this case we let the bitlength $\langle V\rangle$ of this combinatorial specification play the role of $\langle\lambda\rangle$ in the above definition.

If H is not connected, we assume that we are given the standard generators of its connected component H_0 containing the identity and an explicit set S of generators of the discrete part (so that H_0 and S together generate H). An explicit positive basis of a representation V of H is then defined similarly by requiring, in addition, that an explicit positive representation of the form (38) also exists for sb, for every $s \in S$.

The well known positive ($\#P$) Littlewood-Richardson rule for decomposing the tensor product of irreducible $GL_n(\mathbb{C})$-representations (which implies the analogue of PH1 for Littlewood-Richardson coefficients) follows from the proof of the following deep positivity result proved in [Lu]. It is a specialization at $q = 1$ of a more general positivity result in the quantum getting.

Littlewood-Richardson PH0: [Lu] Let $H = GL_n(\mathbb{C})$ be embedded in $G = H \times H$ diagonally. Then the irreducible representation $V_\alpha(H) \otimes V_\beta(H)$ of G has a strictly positive (canonical) basis with respect to the H action. It may be conjectured that this basis is explicitly positive. But this is not known at present.

One may similarly try to prove general PH1 (Hypotheses 4.2 and 5.3) by proving the following generalization of the Littlewood-Richardson PH0 first and then deducing PH1 from it.

Hypothesis 9.2 (PH0)

Plethsym PH0: *Let*

$$H = GL_n(\mathbb{C}) \to G = GL(V), V = V_\mu(H), \tag{39}$$

be the plethysm representation map (15). Then each Weyl module $V_\lambda(G)$ has an explicit positive basis B_μ^λ with respect to the H-action.

PH0 for $\mathbb{C}[V]$, $R_V[f,n,m]$ and $R_V[g,m]$: *$\mathbb{C}[V]_d$, $R_V[f,n,m]_d$ and $R_V[g,m]_d$ (specified compactly by just giving n, m and d in unary) have explicit positive bases with respect to the G-action.*

More strongly, $R_V[g,m]$ has an explicit positive monomial basis $\bar{B}_g$ induced by an explicit positive monomial basis B_g of $\mathbb{C}[V]$ that is simultaneously compatible with the action of G and the action of $G_g \subseteq G$, where G_g denotes the stabilizer of $g = \det(Y) \in V$. This means there exists a finite generating set $S_g = \{s_1, \ldots, s_l\} \subseteq B_g$ with the following properties:

1. *Each $s \in S_g$ is homogeneous and has a combinatorial label of bitlength $\langle s \rangle = poly(n,m)$.*
2. *Each basis element $b \in B_g$ is a monomial in the generators in S_g.*
3. *Let $B_g(d) \subseteq B_g$ be the subset of basis elements of degree d. Then $B_g(d)$ is an explicit positive basis (Definition 9.1) of $R_V[g,m]_d$.*
4. *Let $B_g(d) = B_{g,0}(d) \supset B_{g,1}(d) \supset \cdots$ be the G-compatible filtration of $B_g(d)$ as in Definition 9.1. Let $\tilde{B}_{g,i}(d)$ be the basis of the G-module $\langle B_{g,i}(d)\rangle / \langle B_{g,i+1}(d)\rangle$ induced by $B_{g,i}(d) \setminus B_{g,i+1}(d)$. Then $\tilde{B}_{g,i}(d)$ is an explicit positive basis of $\langle B_{g,i}(d)\rangle / \langle B_{g,i+1}(d)\rangle$ with respect to the G_g-action.*
5. *Let $\mathbb{C}[S_g]$ be the free ring generated by S_g, and I_g the ideal so that $R_V[g,m] = \mathbb{C}[S_g]/I_g$. Let GB_g denote the Gröbner basis of I_g (with an appropriate ordering among the elements in S_g). Then $\bar{B}_g$ is the standard monomial basis of $R_V[g,m]$ with respect to GB_g.*
6. *Let $\bar{B}_g(d) \subseteq \bar{B}_g$ denote the standard monomial basis of $R_V[g,m]_d$ formed by the standard monomials in $\bar{B}_g$ of degree d. The combinatorial specification of any $b = s_{i_1}^{j_1} \ldots s_{i_k}^{j_k} \in \bar{B}_g$ specifies the indices j_t's of s_{i_t}'s occurring here with nonzero exponents. Then each $\bar{B}_g(d)$ is an explicit positive basis of $R_V[g,m]_d$ (Definition 9.1) with respect to the action of G, and also with respect to the action of G_g as in 4 above.*
7. *Each element $c \in GB_g$ has an explicit, positive expression of the form:*

$$c = \sum_{b \in B_g} (-1)^{\alpha_b^c} \beta_b^c b, \tag{40}$$

where b's that occur in the support of c have poly(n,m) degree, α_b^c is a poly(n,m)-time computable integer, and β_b^c is a nonnegative integer with a $\#P$-formula whose nonvanishing can be decided in poly(n,m) time.

PH0 for $R_V[f,n,m]$ and $R_V[g,m]$ is conjectured on the basis of the Strong Flip Theorem 2.3, which suggests that the elimination theory of the class varieties $\Delta_V[g,m]$ and $\Delta_V[f,n,m]$ can be made explicit (which is essentially what PH0 says). Here positivity is essentially a prerequisite for explicitness. Indeed, the structure constants $c_{b,b'}^h$ and β_b^c in (38) and (40) are, in general, hard to compute. Hence we require them to have positive $\#P$-formulae (for the same reasons as in Section 8.2) so that their nonvanishing may be decided in polynomial time.

Plethysm PH0 implies a $\#P$-formula for the plethysm constant, a crucial ingradient of Plethysm PH1 (Hypothesis 5.3):

$$a_{\lambda,\mu}^{\pi} = \sum_{b \in B_{\pi,\mu}^{\lambda}} 1, \tag{41}$$

where $B_{\pi,\mu}^{\lambda} \subseteq B_{\mu}^{\lambda}$ consists of all basis elements that are highest weight vectors with weight π. By 3 (b) in Definition 9.1, whether $b \in B_{\pi,\mu}^{\lambda}$ can be checked in polynomial time. Hence this is a $\#P$-formula. PH0 for $R_V[f,n,m]_d$ and $R_V[g,m]_d$ similarly implies $\#P$-formulae for the multiplicities $F(\lambda,n,m)$ and $G(\lambda,m)$, the crucial ingradient of PH1 for these multiplicities.

One may wonder why we should go through PH0 to prove PH1 for these multiplicities since the Littlewood-Richardson rule has an elementary proof, whereas PH0 for Littlewood-Richardson coefficients [Lu] is far deeper. The reason is again the Strong Flip Theorem 2.3, which suggests that problems of difficulty comparable to general PH0 can be expected in any proof of the strong permanent vs. determinant conjecture, modulo derandomization. Indeed, the problem of constructing an extremely explicit positive separator between $\Delta_V[f,n,m]$ and $\Delta_V[g,m]$ addressed in the strong flip theorem seems harder than the problem of constructing explicit positive bases of $R_V[f,n,m]$ and $R_V[g,m]$ because of the higher level of explicitness in the former problem.

The Strong Flip Theorem does not say anything regarding the plethysm constants. Hence it is plausible that Plethysm PH1 has a much a simpler proof than Plethysm PH0. But Plethysm PH0 is a simpler prototype of PH0 for $R_V[f,n,m]$ and $R_V[g,m]$, and hence, deserves to be studied first.

An approach towards Plethysm PH0 is described in the sequels [GCT7] and [GCT8] to this paper. The basic idea is to quantize the embedding $H \subseteq G$ to get an embedding $H_q \subseteq G_q^H$, where H_q is the standard quantum group [Dr] associated with H, and G_q^H is a nonstandard quantum group constructed in [GCT7]. The article [GCT8] constructs a conjectural canonical basis of an appropriate quantization of $V_\lambda(G)$ with respect to the G_q^H-action. This basis conjecturally yields an explicit positive basis of $V_\lambda(G)$ when specialized at $q=1$. For PH0 for $R_V[g,m]$, one has to similarly quantize the triple $G_g \subseteq G \subseteq GL(V)$ to get an explicit positive basis B_g of $\mathbb{C}[V]$ simultaneously compatible with respect to the action of G and G_g. One also has to show that B_g induces an explicit positive basis $\bar{B}_g$ of $R_V[g,m]$. For this it is crucial that g be characterized by its stabilizer.

One can also formulate analogues of PH0 for the orbit closure of a point $x \in P(V)$ char-

acterized by an explicitly given stabilizer, where V is a representation of a connected algebraic reductive group–we omit the details.

10 The arithmetic P vs. NP problem in characteristic zero

In this section we lift the story for the permanent vs. determinant problem in the preceding sections to the arithmetic P vs. NP problem in characteristic zero defined in [GCT1]. Since the story is very similar, we will be brief.

The role of the permanent is played in the arithmetic P vs. NP problem by the following function $E(X)$ (cf. [GCT1]) defined over $\mathbb{Q}$. 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 $\sigma : \{1, \ldots, m\} \to \{1, \ldots, k\}$, 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 $E(X)$ cannot be computed by an arithmetic circuit over $\mathbb{Q}$ of poly(n) size. It 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 role of the determinant function is played in the P vs. NP problem by the following function $H(Y)$ (cf. [GCT1]), which is P-complete. To define it, consider a generic arithmetic circuit of depth k and width m. It consists of $k+1$ levels of nodes, numbered 0 to k, each level containing m nodes, except the root level zero, which contains a single output node. Each node in the level $i < k$ is connected to every node in level $i+1$. Each node u in the input level k is labeled with the variable y^u; the function computed by this node is defined to be y^u. The function $h(u)$ computed by a node u in level $i < k$ is defined to be $\sum_{v,w} y^u_{v,w} h(v)h(w)$, where v and w range over nodes in level $i+1$ and each $y^u_{v,w}$ is an indeterminate. Let Y be the vector of the variables y^u's at the input level k and the variables $y^u_{v,w}$'s. Let $H(Y)$ be the function computed at the root level zero. It is a homogeneous form in Y with total degree exponential in k. Let $m = O(r^2)$ be the size of Y for the generic circuit with depth and width r. The function $H(Y)$ is characterized by symmetries in a weaker sense (cf. [GCT1]) that is good enough for our purposes.

Let V be the space of homogeneous forms in Y with degree equal to that of $H(Y)$, $P(V)$ the corresponding projective space. Then V and hence $P(V)$ has the action of $G = GL_m(\mathbb{C})$. We think of $h = H(Y)$ as a point in $P(V)$. Let W be the space of forms in X with degree equal to that of $E(X)$. We think of $E(X)$ as a point in $P(W)$. Let $\phi : P(W) \to P(V)$ be the embedding, similar to the embedding ϕ in Section 2.2, and let $e = e(Y) = E^\phi(Y)$.

The orbit closure $\Delta_V[h, m] = \overline{Gh}$ is called the *class variety* associated with the complexity class P, and the orbit closure $\Delta_V[e, n, m] = \overline{Ge}$ is called the *class variety* associated with the complexity class NP. (Alternatively, we can also define $\Delta_V[h, m]$ to be the G-orbit closure of $[h]$, the set of all points in $P(V)$ stabilized by the stabilizer $G_h \subseteq G$ of h.)

It is shown in [GCT1] that if $E(X)$ can be computed by an arithmetic circuit of depth and width $\leq r$ then

$$\Delta_V[e,n,m] \subseteq \Delta_V[h,m] \tag{42}$$

where $m = O(r^2)$. Hence, to solve the nonuniform arithmetic P vs. NP problem in characteristic zero, it suffices to show that this is not possible when $r = \text{poly}(n)$.

Let $R_V[e,n,m]$ and $R_V[h,m]$ be the homogeneous coordinate rings of $\Delta_V[e,n,m]$ and $\Delta_V[h,m]$, respectively. Let $H(\lambda,m)$ denote the multiplicity of $V_\lambda(G)$ in $R_V[h,m]^*$ and $E(\lambda,n,m)$ the multiplicity of $V_\lambda(G)$ in $R_V[e,n,m]^*$. The following is the analogue of Definition 2.4 in this context.

Definition 10.1 *A Weyl module $S = V_\lambda(G)$, for a given partition λ, is called an incidence-based* geometric obstruction *for the inclusion (42) if $E(\lambda,n,m) > 0$ and $F(\lambda,m) = 0$.*

Analogues of FH, PH1, SH, PH2, and PH0 (Hypotheses 3.1, 3.2, 4.2, 4.6, 4.7,9.2) can now be conjectured, and analogues of the decompositions (1), (8), and Theorems 4.1, 4.8, and 4.9 then hold.

Acknowledgements

The author is grateful to Peter Burgisser, Shrawan Kumar, Peter Littelmann, Hari Narayanan, and Jerzy Weyman for helful discussions, and to E. Briand, R. Orellana, and M. Rosas for pointing out in [BOR] an error in the statement of the saturation hypothesis in the earlier version of this paper.

References

[BDR] M. Beck, R. Diaz, S. Robins, The Frobenius problem, rational polytopes, and Fourier-Dedekind sums, Journal of number theory, vol. 96, issue 1, 2002.

[BZ] A. Berenstein, A. Zelevinsky, Tensor product multiplicities and convex polytopes in partition space, J. Geom. Phys. 5(3): 453-472, 1988.

[Bt] J. Boutot, Singularit'es rationelles et quotients par les groupes r'eductifs, Invent. Math.88, (1987), 65-68.

[BOR] E. Briand, R. Orellana, M. Rosas, Reduced Kronecker coefficients and counter-examples to Mulmuley's satuation conjecture SH, arXIv:0810.3163v1 [math.CO] 17 Oct, 2008.

[BI] P. Bürgisser, C. Ikenmeyer, A max-flow algorithm for positivity of Littlewood-Richardson coefficients, FPSAC 2009.

[BI2] P. Bürgisser, C. Ikenmeyer, Geometric complexity theory and tensor rank, arXiv:1011.1350, Nov. 2010.

[BLMW] P. Bürgisser, J. Landsberg, L. Manivel, J. Weyman, An overview of mathematical issues arising in the geometric complexity theory approach to $VP \neq VNP$, arXiv: 0907.2850v1 [cs.CC], July 2009.

[D] R. Dehy, Combinatorial results on Demazure modules, J. of Algebra 205, 505-524 (1998).

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

[DM] J. De Loera, T. McAllister, On the computation of Clebsch-Gordon coefficients and the dilation effect, Experiment Math. 15, (2006), no. 1, 7-20.

[F] H. Flenner, Rationale quasi-homogene singularitäten, Arch. Math. 36 (1981), 35-44.

[FKK] I. Frenkel, M. Khovanov, A.A. Kirillov, Jr., Kazhdan-Lusztig polynomials and canonical basis, Transformation groups, vol. 3, No. 4, 1998, pp. 321-336.

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

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

[GCTexpo] K. Mulmuley, On P vs. NP and geometric complexity theory, To appear in JACM. Available at: http://ramakrishnadas.cs.uchicago.edu

[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.

[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, Earlier version: arXiv:0704.0229, Jan, 2009.

[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.

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

[GLS] M. Grötschel, L. Lovász, A. Schrijver, Geometric algorithms and combinatorial optimzation, 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.

[H] H. Hironaka, Resolution of singularities of an algebraic variety over a field of characteristic zero, Ann. of Math (2), 79: 109-273.

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

[JSV] M. Jerrum, A. Sinclair, E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries, JACM, vol. 51, issue 4, 2004.

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

[KB] R. Kannan, A. Bachem, Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix, SIAM J. comput., 8 (1979) 499-507.

[Ki] R. King, Some remarks on characters of symmetric groups, Schur functions, Littlewood-Richardson and Kronecker coefficients, Workshop on Mathematical Foundations of Quantum Information, Seville, Spain, November, 2009.

[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.

[Ki] A. Kirillov, An invitation to the generalized saturation conjecture, math. CO/0404353., 20 Apr. 2004.

[KT1] A. Knutson, T. Tao, The Honeycomb model of $GL_n(\mathbb{C})$ tensor products I: proof of the saturation conjecture, J. Amer. Math. Soc, 12, 1999, pp. 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.

[LT] B. Leclerc, J. Thibon, Littlewood-Richardson coefficients and Kazhdan-Lusztig polynomials, arXiv:math/9809122, Sept. 1998.

[Lu] G. Lusztig, Introduction to quantum groups, Birkhäuser, 1994.

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

[MM2] 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] V. Mehta, A. Ramanathan, Frobeniuns splitting and cohomology vanishing for Schubert varieties, Ann. Math. 122, 1985, 27-40.

[MR2] T. Mignon, N. Ressayre, A quadratic bound for the determinant and permanent problem, International Mathematics Research Notices (2004) 2004: 4241-4253.

[N] H. Narayanan, On the complexity of computing Kostka numbers and Littlewood-Richardson coefficients, J. of Algebraic combinatorics, vol. 24, issue 3, Nov. 2006.

[R] A. Ramanathan, Schubert varieties are arithmetically Cohen-Macaley, Invent. Math 80, No. 2, 283-294 (1985).

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

[S] K. Smith, F-rational rings have rational singularities, Amer. J. Math. 119 (1997).

[St] R. Stanley, Combinatorics and commutative algebra, Birkhäuser, 1983.

[St2] R. Stanley, Decompositions of rational polytopes, Annals of discrete mathematics 6 (1980) 333-342.

[St3] 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, pp 189-201, 1979.