# On P vs. NP, Geometric Complexity Theory, and the Riemann Hypothesis

Dedicated to Sri Ramakrishna

Ketan D. Mulmuley
The University of Chicago

http://ramakrishnadas.cs.uchicago.edu

Technical Report, Computer Science Department,
The University of Chicago

August 20, 2009

**Abstract**

Geometric complexity theory (GCT) is an approach to the $P$ vs. $NP$ and related problems suggested in a series of articles we call GCTlocal [27], GCT1-8 [30]-[35], and GCTflip [28]. A high level overview of this research plan and the results obtained so far was presented in a series of three lectures in the Institute of Advanced study, Princeton, Feb 9-11, 2009. This article contains the material covered in those lectures after some revision, and gives a mathematical overview of GCT. No background in algebraic geometry, representation theory or quantum groups is assumed. For those who are interested in a short mathematical overview, the first lecture (chapter) of this article gives this. The video lectures for this series are available at:

http://video.ias.edu/csdm/pvsnp

They may be a helpful supplement to this article.

# Introduction

This article gives a mathematical overview of geometric complexity theory (GCT), an approach towards the fundamental lower bound problems in complexity theory, such as (Figure 1):

(1) The $P$ vs. $NP$ problem [8, 17, 23]: show that $P \neq NP$;

(2) The $\#P$ vs. $NC$ problem [43]: show that $\#P \neq NC$.

(3) The $P$ vs. $NC$ problem: show that $P \neq NC$.

We focus here on only the nonuniform versions of the above problems in characteristic zero; i.e., when the underlying field of computation is of characteristic zero, say $Q$ or $\mathbb{C}$–what this means will be explained below. The additional problems that need to be addressed when the underlying field of computation is finite would be discussed in GCT11.

The nonuniform characteristic zero version of the $P \neq NC$ conjecture (in fact, something stronger) was already proved in GCTlocal. We shall refer to it as the $P \neq NC$ result without bit operations. It says that the max flow problem cannot be solved in the PRAM model without bit operations in polylog($N$) time using poly($N$) processors where $N$ is the bitlength of the input. This may be considered to be the first unconditional lower bound result of GCT, because, though it can be stated in purely elementary combinatorial terms, being a formal implication of the $P \neq NC$ conjecture, its proof is intrinsically geometric, and no elementary proof is known so far. Furthermore, its proof technique may be considered to be a weaker (local) form of the *flip*, the basic guiding strategy of GCT, which was refined and formalized much later in GCTflip. This was the begining of this geometric approach in complexity theory. The later work in GCT– the subject of this overview–focusses on the other two problems above, namely the $P$ vs. $NP$ and $\#P$ vs. $NC$ problems.

The nonuniform (characteristic zero) version of the $\#P$ vs. $NC$ problem is also known as the permanent vs. determinant problem [43]. It is to show that $\mathrm{perm}(X)$, the permanent of an $n \times n$ variable matrix $X$, cannot be represented linearly as $\det(Y)$, the determinant of an $m \times m$ matrix $Y$, if $m = \mathrm{poly}(n)$, or more generally, $m = 2^{\log^a n}$, for a fixed constant $a > 0$, and $n \to \infty$. By linear representation, we mean the entries of $Y$ are (possibly nonhomogeneous) linear functions of the entries of $X$. There is an analogous characteristic zero version of the $P$ vs. $NP$ problem defined in GCT1, where the role of the permanent is played by an appropriate (co)-NP complete function and the role of the determinant is played by an appropriate $P$-complete function. The main results of GCT for the $\#P$ vs. $NC$ problem in characteristic zero also extend to the $P$ vs. $NP$ problem in characteristic zero. But here we concentrate on only the former problem, since this illustrates all the basic ideas.

The complementary article [29] gives a complexity-theoretic overview of GCT. It describes the main complexity theoretic barrier towards these problems called the *complexity barrier* and the defining strategy of GCT for crossing it called the *flip* [GCT6,GCTflip]: which is to go for *explicit proofs*. By an explicit proof we mean a proof that provides proof certificates of hardness for the hard function under consideration that are short (of polynomial size) and easy to verify (in polynomial time). This barrier turns out to be extremely formidable and is the root cause of all difficulties in these problems. Nonelementary techniques are brought into GCT precisely to cross this barrier. It is not discussed in these lectures. The goal here is to describe the basic ideas of GCT at a concrete mathematical level without getting into such meta issues. But the readers who wish to know the need for the nonelementary techniques in GCT before getting into any mathematics may wish to read that article before this one. On the other hand, the readers who would rather avoid meta issues before getting a concrete mathematical picture may wish to read this article first. We leave the choice to the readers.

The original IAS lectures stated a lower bound called a *weak form of the $\#P$ vs. $NC$ problem.* This is a special case of a much more general result which we shall call *a mathematical form of the $\#P \neq NC$ conjecture*

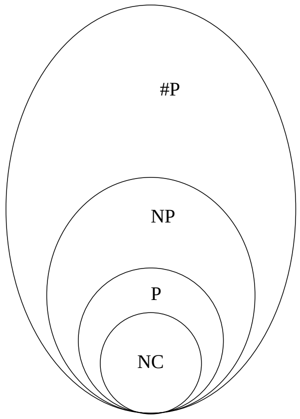


Figure 1: Complexity Classes

(Section 1.2). The article [3] showed that the weak form stated in the IAS lectures is too weak because it has a direct elementary (linear algebraic) proof. Hence in this article it has been replaced with the mathematical form of the $\#P \neq NC$ conjecture mentioned above; cf. Section 1.2. This mathematical form does not have an elementary proof.

The rest of this article is organized in the form of three chapters, one per lecture. The first gives a short mathematical overview of the basic plan of GCT, which is elaborated in the next two lectures.

*Acknowledgement:* The author is grateful to Avi Wigderson for arranging the lectures and the hospitality, to the authors of [3] for pointing out weakness of the lower bound stated during the lectures, and to Shrawan Kumar for bringing the reference [4] to his attention and helpful discussions.

# Chapter 1

# Basic plan

We now outline the basic plan of GCT focussing on the permanent vs. determinant problem in characteristic zero.

## 1.1 Characterization by symmetries

We begin by observing that the permanent and the determinant are *exceptional* polynomial functions. By exceptional we mean that they are completely characterized by the symmetries in the following sense.

Let $Y$ be a variable $m \times m$ matrix. Let $\mathrm{Sym}^m(Y)$ be the space of homogeneous forms of degree $m$ in the $m^2$ variable entries of $Y$. Then, by the classical representation theory, $\det(Y)$ is the only form in $\mathrm{Sym}^m(Y)$ such that, for any $A, B \in GL_m(\mathbb{C})$ with $\det(A)\det(B) = 1$,

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

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

Similarly, $\mathrm{perm}(X)$ is the only form in the space of forms of degree $n$ in the entries of $X$ such that, for any diagonal or permutation matrices $A, B$,

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

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

A basic idea [GCT1] is to a get a handle on the permanent vs. determinant problem by exploiting exceptional nature of these polynomials–i.e., their characteristic properties (P) and (D). Representation theory and algebraic geometry enter inevitably into the study of these properties, because to understand symmetries representation theory (of groups of symmetries) becomes indispensible, and to understand deeper properties of representations algebraic geometry becomes indispensible.

## 1.2 A mathematical form of the $\#P \neq NC$ conjecture

To show how these characteristic properties can be exploited, we now state one application of GCT in the form of a concrete lower bound result–namely a mathematical form of the $\#P \neq NC$ conjecture (Theorem 1.2.4 below)–before going any further.

We begin by observing that the permanent vs. determinant conjecture clearly implies that perm($X$) of any $n \times n$ variable matrix $X$ can not be represented as an $NC$-computable polynomial in the traces of $\bar{X}^j$, $j \geq 0$, $\bar{X} = BXC$ for any (possibly singular) $n \times n$ matrices $B$ and $C$, since $X^j$ can be computed fast in parallel. This can be proved unconditionally. In fact, something stronger.

**Proposition 1.2.1** *There do not exist (possibly singular) $n \times n$ complex matrices $B$ and $C$ and a polynomial $e(w_0, \ldots, w_n)$ such that $perm(X) = g(BXC)$, where $g(X) = e(trace(X^0), trace(X), \ldots, trace(X^n))$.*

This was referred to as the weak form of the $\#P$ vs. $NC$ problem in the original IAS lecture. The article [3] showed that this is too weak by giving an elementary linear algebraic proof [3].

We now state a more general lower bound, which was not stated in the IAS lecture, and which does not have an elementary proof. For that we need a definition.

**Definition 1.2.2** *(a) A polynomial function $p(X_1, \ldots, X_k)$ (of any degree) in the entries of $k$ $n \times n$ variable matrices $X_1, \ldots, X_k$ is called a* generalized permanent *if it has exactly the same symmetries as that of the permanent;*

*i.e., for all nonsingular $n \times n$ matrices $U_i$ and $V_i$, $i \leq k$,*

$$p(U_1X_1V_1, \ldots, U_kX_kV_k) = p(X_1, \ldots, X_k) \text{ iff } perm(U_iXV_i) = perm(X) \quad \forall i.$$

*A precise description of the symmetries of the permanent is given by the property (P). Hence, $U_i$ and $V_i$ above have to be permutation or diagonal matrices (with obvious constraints on the product of their diagonal entries), or product of such matrices.*

*(b) A polynomial function $p(X_1, \ldots, X_k)$ of any degree is called a* hybrid symmetric function *if it has exactly the same symmetries on the left as the determinant and on the right as the permanent. That is, for any $n \times n$ nonsingular matrices $L$ and $R$,*

$$p(LX_1R, \ldots, LX_kR) = p(X_1, \ldots, X_k)$$

*iff $L$ has determinant one, and $R$ can be expressed as a product $\prod_s R_s$, where each $R_s$ is either a diagonal matrix of determinant one or a permutation matrix.*

When $k = 1$ and $n$ is arbitrary, there is just one generalized permanent of degree $n$, namely the usual permanent itself. At the other extreme, when $n = 1$ and $k$ is arbitrary, every function in $k$ variables is a generalized permanent and also a hybrid symmetric function. But in general there are far fewer generalized permanent and hybrid symmetric functions, and their spaces have highly nontrivial structures that are intimately linked to some fundamental problems of representation theory; cf. Section 3.2 and [GCT6].

Now we have the following: Assume that the characteristic of the underlying field is zero.

**Observation 1.2.3 (Implication of the nonuniform $\#P \neq NC$ conjecture)** *(a) Assuming the nonuniform $\#P \neq NC$ conjecture in characteristic zero, no $\#P$-complete generalized permanent $p(X_1, \ldots, X_k)$ of poly$(n, k)$ degree can be expressed as an $NC$-computable polynomial function of the traces of $\bar{X}_i^j$, $1 \leq i \leq k$, $j =$ poly$(n, k)$, where $\bar{X}_i = B_iX_iC_i$, $i \leq k$, for any $n \times n$ complex (possibly singular) matrices $B_i$ and $C_i$.*

*(b) Similarly, assuming the nonuniform $\#P \neq NC$ conjecture in characteristic zero, no hybrid symmetric function $p(X_1, \ldots, X_k)$ can be expressed as an $NC$-computable polynomial function of the traces of monomials of poly$(n, k)$ length in $\bar{X}_i = BX_iC$, for any possibly singular $n \times n$ matrices $B$ and $C$. Here by a monomial in $\bar{X}_i$'s we mean any product of the form $\bar{X}_{i_1} \ldots \bar{X}_{i_s}$, $i_1, \ldots, i_s \leq k$; $s$ is called its length.*

This follows because $\bar{X}_i^j$ or monomials in $\bar{X}_i$'s of polynomial length are clearly $NC$-computable. Note that in (b), unlike in (a), all $X_i$'s are transformed to $\bar{X}_i$ in the same way.

When $n = 1$ and $k$ is arbitrary, this implication is equivalent to the original nonuniform $\#P \neq NC$ conjecture (in characteristic zero), since then any polynomial in $x_1, \ldots, x_k$ is a generalized permanent or a hybrid symmetric function, and a polynomial function of the traces of (monomials in) $x_i$'s means any polynomial in $x_1, \ldots, x_k$. This, i.e., the general $\#P \neq NC$ conjecture in characteristic zero, cannot be proved unconditionally at present. But the next case of this implication, $n > 1$ and $k$ arbitrary, can be:

**Theorem 1.2.4 (A mathematical form of the $\#P \neq NC$ conjecture)**

*The implication above holds unconditionally for any $n > 1$ and arbitrary $k$.*

*In fact, something stronger then holds. Namely, when $n > 1$ and $k$ is arbitrary:*

*(a) No generalized permanent $p(X_1, \ldots, X_k)$ can be expressed as a polynomial function of the traces of $\bar{X}_i^j$, $1 \leq i \leq k$, $j \geq 0$, where $\bar{X}_i = B_i X_i C_i$, $i \leq k$, for any $n \times n$ complex (possibly singular) matrices $B_i$ and $C_i$.*

*(b) No hybrid symmetric function $p(X_1, \ldots, X_k)$ can be expressed as a polynomial function of the traces of monomials in the matrices $\bar{X}_i = B X_i C$, for any possibly singular $n \times n$ matrices $B$ and $C$.*

When $k = 1$ and $p(X_1)$ is the usual permanent, (a) specializes to Proposition 1.2.1.

We are calling this a mathematical form for two reasons. First, it needs no restriction on the computational complexity of $p(X_1, \ldots, X_k)$ or the polynomial in the traces, (though for trivial reasons we can assume without loss of generality that the polynomial in the traces is computable in $2^{\text{poly}(n,k,d)}$ time, where $d$ is the degree of $p$). Thus it is rather in the spirit of the classical result of Galois theory which says that a polynomial whose Galois group is not solvable cannot be solved by any number of radical operations, without any restriction on the number of such operations (though again there is a trivial upper bound on the number of such operations needed if the polynomial is solvable by radicals). Second, observe that the permanent has two characteristic properties: 1) the property P (mathematical), and 2) $\#P$-completeness (complexity-theoretic). The usual complexity theoretic form of the $\#P \neq NC$ conjecture is a lower bound for all polynomial functions with the $\#P$-completeness property, whereas the mathematical form is a lower bound for all polynomial functions with the property (P). In other words, the complexity theoretic form is associated with the $\#P$-completeness property of the permanent and the mathematical form (a) with the mathematical property (P). Mathematical form (b) is associated with a hybrid of (P) and (D).

The result indicates that there is thus a chasm between the two adjacent cases: $n = 1$, $k$ arbitrary (the usual nonuniform complexity theoretic $\#P \neq NC$ conjecture), and $n = 2$, $k$ arbitrary (its mathematical form above). The complexity theoretic form is far far harder than the mathematical form.

For some specific generalized permanents (cf. Section 3.2), this result again has an elementary linear algebraic proof as in [3]. It also has an elementary linear algebraic proof for a generic generalized permanent or hybrid symmetric function. A nontrivial part of this result (which needs nonelementary proof) is that it holds for any generalized permanent or hybrid symmetric function.

The statement (a) of the mathematical form has two proofs. The first uses only basic geometric invariant theory [38]. Basically the proof of the weak form (or rather its nonhomgeneous version) given in [32] works here as well with obvious modifications; Bharat Adsul [2] has also independenly found a similar proof. But this proof is naturalizable; i.e. it cannot cross the natural proof barrier in [40] as pointed out in [32].

The second proof sketched in this article is not naturalizable and has a deeper structure that is crucial for further progress in GCT. Specifically, it uses the same proof strategy as for the general permanent vs. determinant problem and hence serves as a test case of the general proof strategy in a nontrivial special case.

The statement (b) of the mathematical form does not have a proof of the first kind, but only of the second kind; i.e., based on the global flip, which is not naturalizable. (Well, there is a slight (still nonnaturalizable) variant of the second proof that is nonconstructive. This is obtained by carrying out the last (easy) step nonconstructively. In essence it is not really much different from the proof based on the global glip. Hence we shall ignore this variation here). It may be conjectured that (b) cannot be proved by any naturalizable technique.

Since the only proof of (b) that we have is very similar to the second

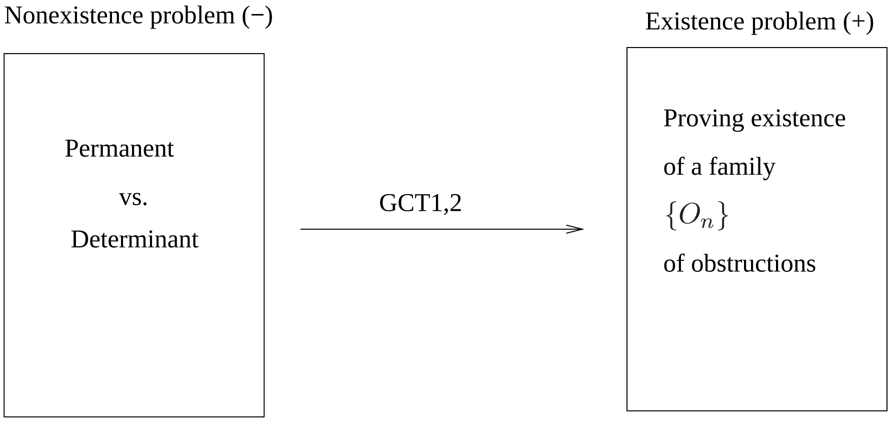


Figure 1.1: Reduction from nonexistence to existence

nonnaturalizable proof of (a), we shall only focuss on the second proof of (a) in this article and leave (b) and its further discussion to [GCT6].

## 1.3 From nonexistence to existence

The rest of this lecture outlines the GCT approach to the general permanent vs. determinant problem in characteristic zero, and then points out the crucial steps in this plan which can be completely executed for the mathematical form of the $\#P \neq NC$ conjecture above, but which are conjectural at present for the general (i.e. complexity theoretic) permanent vs. determinant problem.

The first step (GCT1,2) is to reduce this nonexistence problem–i.e., that there is no small linear representation of the permanent as a determinant–to an existence problem–specifically, to the problem of proving existence of a family $\{O_n\}$ of obstructions; cf. Figure 1.1. Here an *obstruction* $O_n$ is a proof-certificate of hardness of $\mathrm{perm}(X)$, $X$ an $n \times n$ variable matrix, just as the Kurotowski minor is a proof-certificate of the nonplanarity of a graph. Specifically, it is some algebro-geometric-representation-theoretic gadget whose existence for every $n$ serves as a guarantee that $\mathrm{perm}(X)$ cannot be represented linearly as $\det(Y)$, when $m = 2^{\log^a n}$, $a > 0$ fixed, $n \to \infty$ (i.e., for $n$ greater than a large enough constant depending on $a$).

This reduction to existence is carried out as follows (cf. lecture 2 for

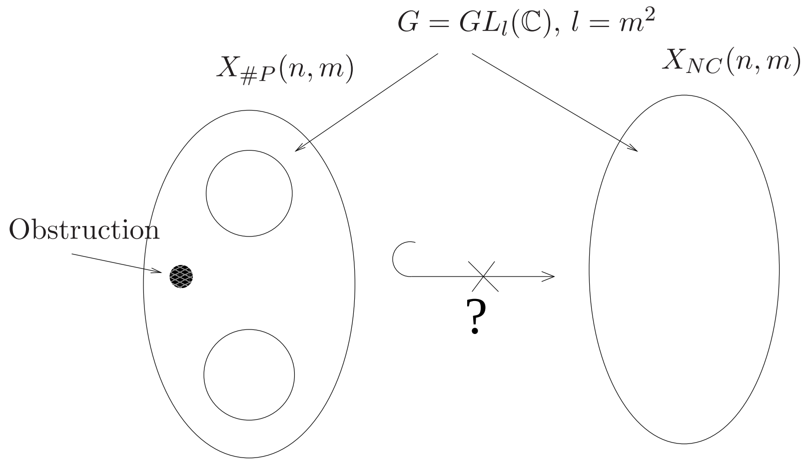


Figure 1.2: Class varieties

details).

First, we associate (GCT1) with the complexity class $\#P$ a family $\{X_{\#P}(n,m)\}$ of (group-theoretic) class varieties (what this means is explained below), and with the complexity class $NC$ a family of $\{X_{NC}(n,m)\}$ of (group-theoretic) class varieties such that: if perm$(X)$, $\dim(X) = n$, can be represented linearly as $\det(Y)$, $\dim(Y) = m > n$, then

$$X_{\#P}(n,m) \subseteq X_{NC}(n,m). \tag{1.1}$$

Each class variety is a (projective) algebraic variety, by which we mean that it is the zero set of a system of multivariate homogeneous polynomials with coefficients in $\mathbb{C}$ (akin to the usual curves and surfaces). It is group-theoretic in the sense that it is constructed using group-theoretic operations and the general linear group $G = GL_l(\mathbb{C})$, $l = m^2$, of $l \times l$ invertible complex matrices acts on it, and furthermore the groups of symmetries of the permanent and the determinant, which we shall refer to as $G_{perm}$ and $G_{det}$, are embedded in this group $G$ as its subgroups in some way. Here action means moving the points of the class variety around, just as $G$ moves the points of $\mathbb{C}^l$ around by the standard action via invertible linear transformations. The goal is to show that the inclusion (1.1) is impossible (obstructions are meant to ensure this); cf. Figure 1.2.

Since each class variety has a $G$-action, the space of polynomial functions on each class variety has a representation-theoretic structure, which puts constraints on which representations of $G$ can live on that variety (i.e., in the space of polynomial functions on that variety). Informally, an obstruction is an irreducible (minimal) representation of $G$ that can live on $X_{\#P}(n,m)$ but not on $X_{NC}(n,m)$; cf. Figure 1.2. Existence of an obstruction $O_n$, for every $n$, assuming $m = 2^{\log^a n}$, $a > 1$ fixed, implies that the inclusion (1.1) is not possible, since $O_n$ cannot live on $X_{NC}(n,m)$. Thus an obstruction blocks the inclusion (1.1).

To define an obstruction formally, we need to recall some basic representation theory. By a classical result of Weyl, the irreducible (polynomial) representations of $G = GL_l(\mathbb{C})$ are in one-to-one correspondence with the partitions $\lambda$ of length at most $l$, by which we mean integral sequences $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_k > 0$, $k \leq l$, where $k$ is called the length of $\lambda$. The irreducible representation of $G$ in correspondence with $\lambda$ is denoted by $V_\lambda(G)$, and is called the *Weyl module* of $G$ indexed by $\lambda$. Symbolically:

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

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

Weyl also proved that every finite dimensional representation of $G$ can be decomposed into irreducible representations–i.e., can be written as a direct sum of Weyl modules. Thus Weyl modules are the basic building blocks of the representation theory of $G$, and every finite dimensional representation of $G$ can be thought of as a complex building made out of these blocks.

Now suppose $m = 2^{\log^a n}$, $a > 1$ fixed, $n \to \infty$. Suppose to the contrary that

$$X_{\#P}(n,m) \subseteq X_{NC}(n,m). \tag{1.2}$$

Let $R_{\#P}(n,m)$ denote the homogeneous coordinate ring of $X_{\#P}(n,m)$; i.e., the ring of polynomial functions [1] on $X_{\#P}(n.m)$. Let $R_{\#P}(n,m)_d$ be the degree-$d$-component of $R_{\#P}(n,m)$ consisting of functions of degree $d$. We define $R_{NC}(n,m)$ and $R_{NC}(n,m)_d$ similarly. Since $X_{\#P}(n,m)$ has the action of $G$, $R_{\#P}(n,m)$ also has an action of $G$; i.e., it is a representation of $G$. Hence, $R_{\#P}(n,m)_d$ is a finite dimensional representation of $G$. Similarly,

[1]Though the functions here are not functions in usual sense; but let us not worry about this

$R_{NC}(n,m)$ is a representation of $G$, and $R_{NC}(n,m)_d$ a finite dimensional representation of $G$.

If (1.2) holds, then we get a natural map from $R_{NC}(n,m)$ to $R_{\#P}(n,m)$ obtained by restricting a function on $X_{NC}(n,m)$ to $X_{\#P}(n,m)$. By basic algebraic geometry, this map is surjective and is a $G$-homomorphism. Furthermore, it is degree-preserving. This means there is a surjective $G$-homomorphism from $R_{NC}(n,m)_d$ to $R_{\#P}(n,m)_d$. Symbolically:

$$R_{\#}(n,m)_d \leftarrow R_{NC}(n,m)_d. \tag{1.3}$$

Let $R_{\#P}(n,m)_d^*$ denote the dual of $R_{\#P}(n,m)_d$; i.e., the set of linear maps from $R_{\#P}(n,m)_d$ to $\mathbb{C}$. Then (1.3) implies that there is an injective $G$-homomorphism from $R_{\#P}(n,m)_d^*$ to $R_{NC}(n,m)_d^*$. Symbolically:

$$R_{\#}(n,m)_d^* \hookrightarrow NC(n,m)_d^*. \tag{1.4}$$

**Definition 1.3.1** *(GCT2) An obstruction $O_n$ is a Weyl module $V_\lambda(G)$ that occurs as a subrepresentation in $R_{\#P}(n,m)_d^*$ but not in $R_{NC}(n,m)_d^*$, for some $d$. We call $\lambda$ an* obstruction label, *and sometimes, by abuse of notation, an obstruction as well.*

*A strong obstruction $O_n$ is a Weyl module $V_\lambda(G)$ that occurs as a subrepresentation in $R_{\#P}(n,m)_d^*$ but does not contain a nonzero invariant (fix point) of the subgroup $G_{det} \subset G$ of the symmetries of the determinant. It can be shown [31] that a strong obstruction is an obstruction in the above sense.*

Here by an invariant we mean a point in $V_\lambda(G)$ which is fixed (does not move) with respect to the action of the subgroup $G_{det} \subset G$.

**Proposition 1.3.2** *(GCT2) Existence of an obstruction $O_n$, for all $n \to \infty$, with $m = 2^{\log^a n}$, $a > 1$ fixed, implies perm$(X)$, $\dim(X) = n$, cannot be represented linearly as* $\det(Y)$, $\dim(Y) = m$.

This follows just from the definition of an obstruction, and leads to:

**Goal 1.3.3** *(GCT2) Prove existence of a (strong) obstruction family $\{O_n = V_{\lambda_n}(G)\}$ using the exceptional nature of perm$(X)$ and* $\det(Y)$, *i.e., using the properties (P) and (D) in Section 1.1.*

## 1.4 Obstructions for the mathematical form

The following result achieves this goal for the mathematical form (Theorem 1.2.4).

**Theorem 1.4.1** *There exists a (strong) obstruction family $\{O_{n,k}\}$ for the mathematical form (a) of the $\#P \neq NC$ conjecture, and also for the mathematical form (b).*

This implies Theorem 1.2.4. The notion of obstructions here is similar to the one in the general case.

The proof of this result based on the results of GCT1 and 2 in geometric invariant theory [38] is outlined in the third lecture. It produces a family $\{O_{n,k} = V_\lambda(G)\}$ of (strong) obstructions, with a different $G$ than in the general complexity theoretic case. Furthermore this family is *strongly explicit* in the sense that the specification $\lambda_{n,k}$ of each $O_{n,k}$ has polynomial, in fact, $O(nk)$ bitlength, and can be constructed in polynomial, in fact, $O(nk)$ time (regardless of the complexity of the polynomial in the traces in the statement of Theorem 1.2.4).

## 1.5 Towards existence of obstructions in general via positivity

We now proceed to describe the main results of GCT for the general (complexity-theoretic) permanent vs. determinant problem in the context of Goal 1.3.3.

Towards that end, we define certain representation-theoretic *stretching* functions. Let $F_{\lambda,n,m}(k)$ denote the number of occurences (multiplicity) of the Weyl module $V_{k\lambda}(G)$ in $R_{\#P}(n,m)^*$ (i.e. $R_{\#P}(n,m)^*_d$, for some $d$) as a subrepresentation. Let $G_{\lambda,m}(k)$ denote the multiplicity of the trivial one dimensional representation (invariant) of $G_{det} \subset G$ in $V_{k\lambda}(G)$. In other words, $G_{\lambda,m}(k)$ is the dimension of the subspace of invariants of $G_{det}$ in $V_{k\lambda}(G)$. These are statistical functions associated with the class variety $X_{\#P}(n,m)$ and the subgroup embedding $G_{det} \hookrightarrow G$. In the first case, the statistics associates with every number (stretching parameter) $k$ the multiplicity of the corresponding Weyl module $V_{k\lambda}(G)$ in $R_{\#P}(n,m)^*$ and in the second case the dimension of the subspace of invariants of the symmetries of the determinant in $V_{k\lambda}(G)$; cf. Figure 1.3.

Let us call a function $f(k)$ *quasipolynomial*, if there exist $l$ polynomials

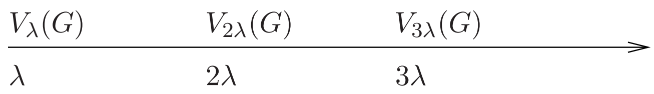


Figure 1.3: statistics

$f_i(k)$, $1 \leq i \leq l$, for some $l$, such that $f(k) = f_i(k)$ for all nonnegative integral $k = i$ modulo $l$; here $l$ is called the period of the quasi-polynomial. Thus quasi-polynomials are hybrids of polynomial and periodic functions. We say that $f(k)$ is an *asymptotic quasipolynomial* if there exist $l$ polynomials $f_i(k)$, $1 \leq i \leq l$, for some $l$, such that $f(k) = f_i(k)$ for all nonnegative integral $k = i$ modulo $l$ for $k \geq a(f)$, for some nonnegative integer 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 fundamental example of a quasi-polynomial is the Ehrhart quasi-polynomial $f_P(k)$ of a polytope $P$ with rational vertices. It is defined to be the number of integer points in the dilated polytope $kP$. By the classical result of Ehrhart, it is known to be a quasi-polynomial. More generally, let $P(k)$ be a polytope parametrized by nonnegative integral $k$: i.e., defined by a linear system of the form:

$$Ax \leq kb + c, \tag{1.5}$$

where $A$ is an $m \times n$ matrix, $x$ a variable $n$-vector, and $b$ and $c$ some constant $m$-vectors. Let $f_P(k)$ be the number of integer points in $P(k)$. It is known to be an asymptotic quasi-polynomial. We shall call it the asymptotic Ehrhart quasi-polynomial of the parametrized polytope $P(k)$. In what follows, we denote a parametrized polytope $P(k)$ by just $P$. From the context it should be clear whether $P$ is a usual nonparametrized polytope or a parametrized polytope.

**Theorem 1.5.1** *(GCT6)*

*(a) The function $G_{\lambda,m}(k)$ is a quasi-polynomial.*

*(b) The function $F_{\lambda,n,m}(k)$ is an asymptotic quasi polynomial.*

*Analogous result holds for the $P$ vs. $NP$ problem in characteristic zero.*

The proof of Theorem 1.5.1 is based on:

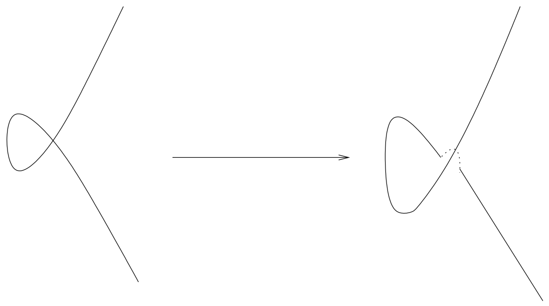

Figure 1.4: Resolution

1. The classical work of Hilbert in invariant theory,
2. The resolutions of singularities in characteristic zero [16]: this roughly says that the singularities of any algebraic variety in characteristic zero can be untangled (resolved) nicely in a systematic fashion; cf. Figure 1.4.
3. Cohomological works of Boutot, Brion, Flenner, Kempf and others based on this resolution; cf. [6, 9] and GCT6 for the history and other references.

As such, this proof is highly nonconstructive. It gives no effective bound on the period $l$–it just says that $l$ is finite.

*Remark:* The original IAS lecture stated a conditional form of (b), which said that $F_{\lambda,n,m}(k)$ is a quasi-polynomial if the singularities of the class variety $X_{\#P}(n,m)$ are rational and normal (in some algebro-geometric sense). Recently, Shrawan Kumar [37] has shown that the singularities of $X_{\#P}(n,m)$ are not normal if $m > n$. This means $F_{\lambda,n,m}(k)$ need not be a quasi-polynomial and asymptotic quasi-polynomiality as in (b) is all that we can expect. This is fine as long as the singularities of $X_{\#P}(n,m)$ are not too bad in the sense described in Remark 3 after Hypothesis 1.6.1 below.

The following hypothesis says that $G_{\lambda,m}(k)$ can be realized as the Ehrhart quasi-polynomial of a polytope, and $F_{\lambda,n,m}(k)$ can be realized as the asymptotic Ehrhart quasi-polynomial of a parametrized polytope.

**Hypothesis 1.5.2 (PH) [Positivity Hypothesis]** *(GCT6)*

*(a) For every $\lambda, n, m \geq n$, there exists a parametrized polytope $P = P_{\lambda,n,m}(k)$*

*such that*

$$F_{\lambda,n,m}(k) = f_P(k). \tag{1.6}$$

*It is also assumed here that there exists for $P_{\lambda,n,m}$ a specification of the form (1.5), where $A$ is independent of $\lambda$, and $b$ and $c$ are piecewise homogeneous linear functions of $\lambda$.*

*(b) For every $m$, there exists a (usual nonparametrized) polytope $Q = Q_{\lambda,m}$ such that*

$$G_{\lambda,m}(k) = f_Q(k). \tag{1.7}$$

*It is assumed here that there exists for $Q_{\lambda,m}$ a specification of the form (1.5) with $k = 1$ and $c = 0$ where $A$ is independent of $\lambda$, and $b$ is is a piecewise homogeneous linear function of $\lambda$.*

*Analogous positivity hypothesis also holds for the $P$ vs. $NP$ problem in characteristic zero.*

If such $P$ and $Q$ exist, their dimensions are guaranteed to be small by the proof of Theorem 1.5.1: specifically, the dimension $P$ is guaranteed to be bounded by a polynomial in $n$, and the dimension of $Q$ by a polynomial in $n$ (but independent of $m$), if the length of $\lambda$ is poly($n$) (as it would be in our applications).

When PH holds, we say that $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ have *positive convex representations*. Here positivity refers to the fact that the Ehhart function $f_P(k)$ is a positive expression:

$$f_P(k) = \sum_v 1,$$

where $v$ ranges over all integer points in $P(k)$–there are no alternating signs in this expression. Convexity refers to the convexity of the polytopes $P(k)$ and $Q$.

But, a priori, it is not at all clear why PH should even hold. Many numerical functions in mathematics are quasi-polynomials or asymptotic quasi-polynomials (e.g., the Hilbert function [2] of any projective variety), but they rarely have positive convex representations. PH is expected to hold because of the exceptional nature of the determinant and the permanent. For concrete mathematical evidence and justification, see GCT6,7, and 8.

[2] Hilbert function $h_Z(k)$ of a projective algebraic variety $Z$ is defined to be $\dim(R(Z)_k)$, where $R(Z)$ is the homogeneous coordinate ring of $Z$ and $R(Z)_k$ its degree $k$-component.

The hypothesis PH alone is not sufficient to prove the existence of obstructions. The precise statement of a sufficient condition is given in the theorem below.

**Theorem 1.5.3** *(GCT6) There exists a family $\{O_n\}$ of (strong) obstructions for the $\#P$ vs. $NC$ problem in characteristic zero, for $m = 2^{\log^a n}$, $a > 1$ fixed, $n \to \infty$, assuming,*

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

   *For all $n \to \infty$, there exists $\lambda$ such that $P_{\lambda,n,m}(k) \neq \emptyset$ for all large enough $k$ and $Q_{\lambda,m} = \emptyset$.*

*Analogous result holds for the $P$ vs. $NP$ problem in characteristic zero.*

Mathematical evidence and arguments in support of OH are given in GCT6. The analogous OH that arises in the context of the mathematical form of the $\#P \neq NC$ conjecture can be proven unconditionally.

We call $\lambda$ a *polyhedral obstruction* (or rather, polyhedral obstruction-label) if it satisfies OH. In this case, $k\lambda$, for some integer $k \geq 1$ is a (strong) obstruction–we just have to choose $k$ large enough so that $P_{\lambda,n,m}(k)$ contains an integer point. Henceforth, whenever we say obstruction, we actually mean polyhedral obstruction.

There is a fundamental difference between the nature of PH and OH. PH is a *mathematical hypothesis*, because there is no constraint on what $m$ should be in comparison to $n$ in its statement. In contrast, OH is a *complexity theoretic hypothesis*, because $m$ needs to be small in comparison to $n$ for it to hold.

## 1.6 The flip: Explicit construction of obstructions

In principle PH may have a nonconstructive proof (like that of Theorem 1.5.1) which only tells that such polytopes exist without explicitly constructing them. But proving OH may not be feasible unless the polytopes $P$ and $Q$ in PH are given explicitly. This suggests the following strategy for proving existence of obstructions proposed in GCT6 and GCTflip.

(1) Prove the following stronger explicit form of PH, which is reasonable since the polytopes $P$ and $Q$, if they exist, are already guaranteed to be of small (polynomial) dimension:

**Hypothesis 1.6.1 (PH1)** *(GCT6)*

*(a) There exists an* explicit *parametrized polytope $P_{\lambda,n,m} = P_{\lambda,n,m}(k)$ as in PH (a). Explicit means:*

1. *The polytope is specified by an explicit system of linear constraints, where the bitlength of (the description of) each constraint is poly$(n, \langle\lambda\rangle, \langle m\rangle)$. Here and in what follows, $\langle z\rangle$ denotes the bitlength of the description of $z$.*
2. *The membership problem for the polytope $P_{\lambda,n,m}(k)$ also belongs to the complexity class $P$. That is, given a point $x$, whether it belongs to $P_{\lambda,n,m}(k)$ can also be decided in poly$(\langle x\rangle, \langle\lambda\rangle, n, \langle m\rangle)$ time. Furthermore, we assume that if $x$ does not belong to the polytope, then the membership algorithm also gives a hyperplane separating $x$ from the polytope in the spirit of [15].*

*(b) There exists a similar explicit (nonparametrized) polytope $Q_{\lambda,m}$ satisfying PH (b) with the polynomial bounds that depend on $n$ and the bitlength of $\lambda$, but not on $m$.*

*Remark 1:* Note the occurrence of $\langle m\rangle$ instead of $m$ in the polynomial bounds in (a) (which implies that the bounds here become poly$(n)$ when $m < 2^n$ and $\langle\lambda\rangle = \text{poly}(n)$), and the absence of $m$ in the polynomial bounds in (b), which means they again are poly$(n)$ when $\langle\lambda\rangle = \text{poly}(n)$). The reasons for this will be explained in Lecture 3 (cf. remarks before Hypothesis 3.7.5 and after Hypothesis 3.7.9).

*Remark 2:* In particular, PH1 implies that the functions $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ belong to the complexity class $\#P$.

*Remark 3:* PH1 also implies that the deviation from quasipolynomiality of $F_{\lambda,n,m}(k)$ is small, specifically, $2^{O(\text{poly}(\langle\lambda\rangle,n,\langle m\rangle))}$, i.e., the bitlength of the deviation is polynomial. As remarked after Theorem 1.5.1, this deviation would have been zero, i.e., $F_{\lambda,n,m}(k)$ would have been a (strict) quasi-polynomial, if the singularities of the class variety $X_{\#P}(n,m)$ were all normal and rational, which, as we know now, is not the case [37]. So small deviation from quasi-polynomiality implied by PH1 basically means that the deviation from rationality and normality of the singularities of the class variety $X_{\#P}(n, m)$ is small (cf. Theorem 1.5.1). This is the basic minimum that is required by PH1.

Like PH, PH1 is also a mathematical hypothesis in the sense that it puts no constraint on what $m$ should be in comparison to $n$. Of course, unlike PH, there is some complexity theoretic aspect to it, but it is secondary in comparison to the complexity-theoretic aspect of OH, since smallness of $m$ with respect to $n$ is the crux of the lower bound problems under consideration.

(2a) **[The flip]**

Let $m = 2^{\log^a n}$, for a fixed $a > 1$. Using the explicit forms of the polytopes $P$ and $Q$ in PH1, show existence of an *explicit* family $\{O_n = V_{\lambda_n}(G)\}$ of (polyhedral) obstructions satisfying OH. We say that an obstruction (proof-certificate) $\lambda_n$ is *explicit* if is "short" and "easy to verify":

1. Short: This means its bitlength $\langle \lambda_n \rangle$ is poly($n$), regardless what $m$ is, as long as it is $\leq 2^{\log^a n}$, for some fixed $a > 1$.
2. Easy to verify: given $n, m \leq 2^n$ and $\lambda_n$, whether $\lambda_n$ is a valid polyhedral obstruction can be verified in poly($n, \langle \lambda \rangle$) time. In particular, this is poly($n$) when $\langle \lambda \rangle =$ poly($n$).

Existence of an explicit family of polyhedral obstructions is equivalent to saying that the problem of deciding existence of polyhedral obstructions for given $n$ in unary and $m$ in binary belongs to $NP$–we shall refer to this decision problem as $DP(OH)$. This definition of explicitness is quite natural since the class $NP$ is a class of problems whose proof-certificates (witnesses) are short and easy to verify. As such, the flip–going for explicit obstructions–is a proof-strategy that is literally given to us on a platter by the $P$ vs. $NP$ problem itself. Why it is called flip will be explained later.

It should be stressed that we are primarily interested in only proving existence of obstructions. Whether they are explicit or not does not really matter in the original statement of the problem. But we need to know the polytopes $P$ and $Q$ explicitly (as in PH1) so that proving OH is feasible. But once PH1 is proved, existence of an explicit family follows automatically, as a bonus, whether we care for it or not.

To see why, let us observe that the second condition above (ease of verification) follows directly from PH1 and the polynomial time algorithm for linear programming on polytopes given by separation oracles [15]. Shortness also follows from PH1.

Thus it is as if the $P$ vs. $NP$ problem is forcing us to go for explicit obstructions.

(2b) **[The strong flip (optional)]**

Using the explicit forms of the polytopes $P$ and $Q$ in PH1, construct (rather than just show existence of) a *strongly explicit* family $\{O_n = V_{\lambda_n}(G)\}$ of obstructions satisfying OH. We say that an explicit family of obstructions is strongly explicit if, for each $n$, a valid obstruction-label $\lambda_n$ can be constructed in poly($n$) time. In particular, the height and the bitlength of $\lambda$ is poly($n$) (short) regardless what $m$ is, as long as it is $\leq 2^{\log^a n}$, for some fixed $a > 1$.

For the purposes of the lower bound problems that we are interested in, the flip (just explicit existence) would suffice and the stronger flip (explicit construction) is optional. But the stronger flip can give us deeper insight into these lower bound problems; cf. [29] and GCTflip for more on this.

Now we turn to a few obvious questions.

## 1.7 What has been achieved by all this?: The meaning of the flip

Let us now see what has been achieved so far in the context of the $\#P \neq NC$ conjecture in the nonuniform setting (characteristic zero), the argument for the $P$ vs. $NP$ problem in characteristic zero being similar. At first glance, it may seem that all that GCT has achieved is to exchange a known difficult problem of complexity theory with a new very difficult problem of algebraic geometry. In order to see that something is gained in exchange let us reexamine the original question.

The goal of the original conjecture is to prove that perm($X$), dim($X$) = $n$, cannot be computed by an arithmetic circuit $C$ of size $m = \text{poly}(n)$, or more generally, $m \leq 2^{\log^a n}$, for some fixed $a > 1$, and depth $O(\log^a n)$. Symbolically, let $f_C(X)$ denote the function computed by $C$. Then we want to prove that

$$(\textbf{IOH}) : \forall n \geq n_0 \forall C \exists X : \text{perm}(X) \neq f_C(X), \tag{1.8}$$

where $n_0$ is a sufficiently large constant and $C$ ranges over circuits of size $m = \text{poly}(n)$. For given $X$ and $C$, the problem of deciding if perm($X$) $\neq$

$f_C(X)$ belongs to $P^{\#P}$. Let $DP(IOH)$ denote the decision problem of deciding for given $n$ and $m$ (in unary) whether (1.8) holds with $C$ ranging over circuits of size $m$. Since there are two alternating layers of quantifiers in (1.8), it belongs to $\Pi_2^{\#P}$, which is very high in the complexity hierarchy (cf. Figure 1.5). Hence, we refer to the original hypothesis (1.8) to be proven as IOH (Infeasible Obstruction Hypothesis). Of course, IOH is expected to be a tautology, and hence (1.8) is expected to be verifiable for small $m = \text{poly}(n)$ in $O(1)$ time–but we do not know that as yet.

Equivalently, the goal of IOH is to prove existence of a *trivial obstruction*, which is a table that lists for each small $C$ as above a counterexample $X$ so that $\text{perm}(X) \neq f_C(X)$; cf. Figure 1.6. The number of rows of this table is equal to the number of circuits $C$'s of size $m = \text{poly}(n)$ and depth $O(\log^a n)$. Thus the size of this table is exponential; i.e., $2^{O(\text{poly}(n))}$. (Well, only if the underlying field of computation is finite. For infinite fields, such as $Q$ or $\mathbb{C}$ in this paper, there is another notion of a trivial obstruction (cf. GCT6). But let us imagine that the underlying field is finite for this argument.) The time to verify whether a given table is a trivial obstruction is also exponential, and so also the time to decide if such a table exists and construct one (optional) for given $n$ and $m$. From the complexity theoretic viewpoint, this is an infeasible (inefficient) task. That is why we call this trivial, brute force strategy of proving IOH, based on existence of trivial obstructions, an *infeasible strategy*.

In contrast, assuming PH1, $DP(OH)$, the decision problem of deciding if a new polyhedral obstruction exists, belongs to $NP$ (in a stronger sense assuming that $n$ is given in unary but $m$ is given in binary instead of unary) as we have already observed. Thus, assuming PH1, we have transformed the original decision problem for trivial obstructions $DP(IOH) \in \Pi_2^{\#P}$ to the decision problem for the new polyhedral obstructions $DP(OH) \in NP$, in the process bringing down the time to verify an obstruction from exponential (for the original trivial obstruction) to polynomial (for the new polyhedral obstruction); cf. Figure 1.5. It is crucial here that PH1, the main tool for this reduction, is a mathematical hypothesis, not complexity-theoretic (cf. the remarks after Theorem 1.5.3 and Hypothesis 1.6.1). The task of verifying an obstruction has also been transformed from the infeasible (exponential-time) to the feasible (polynomial-time). Hence the name of this strategy: the flip, from the infeasible to the feasible. Positivity (PH1) is used in the flip as a means to eliminate the quantifying variables in IOH and bring down the complexity of the decision problem associated with the obstruction hypothesis; cf. Figure 1.5.

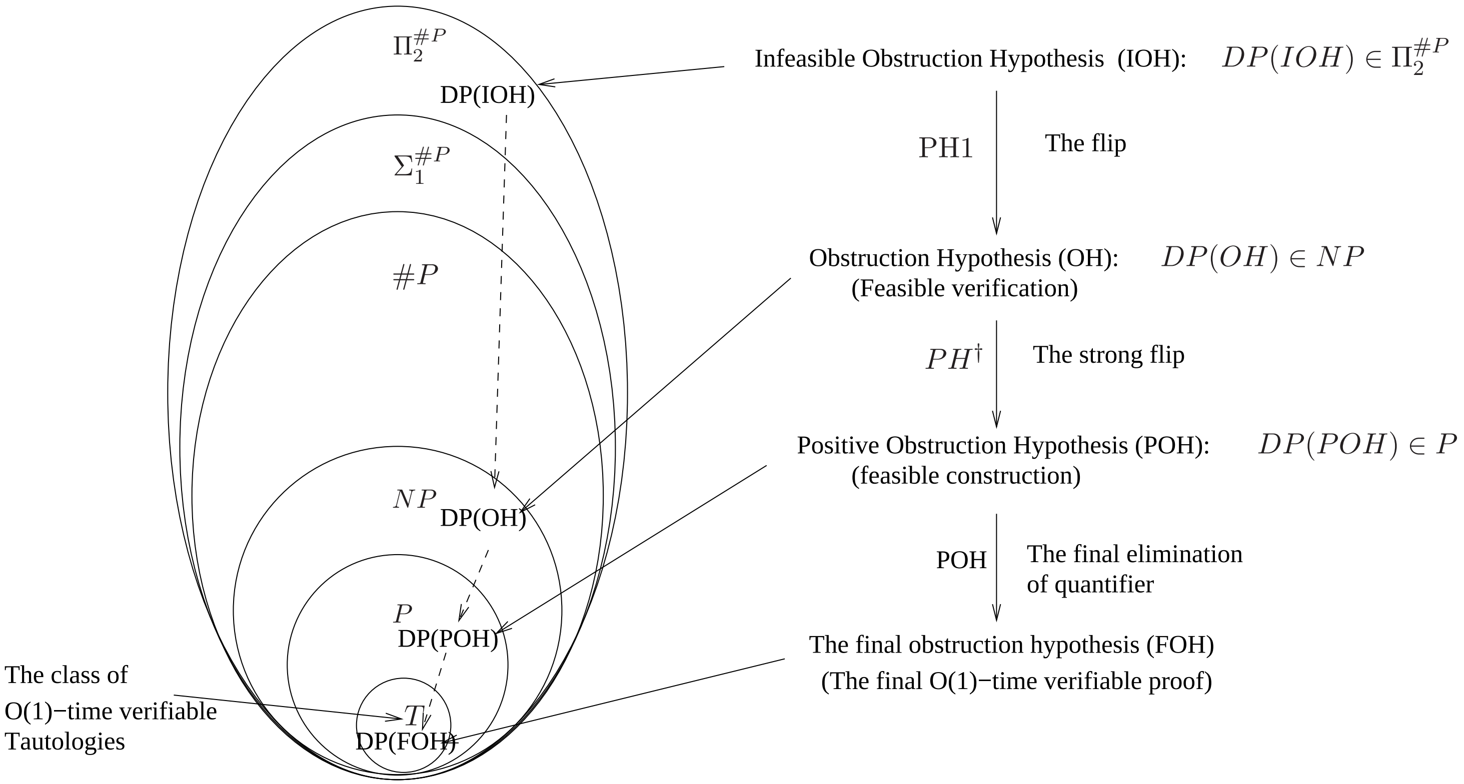


Figure 1.5: Positivity as a means to eliminate the quantifiers and reduce the complexity of the decision problem associated with the obstruction hypothesis

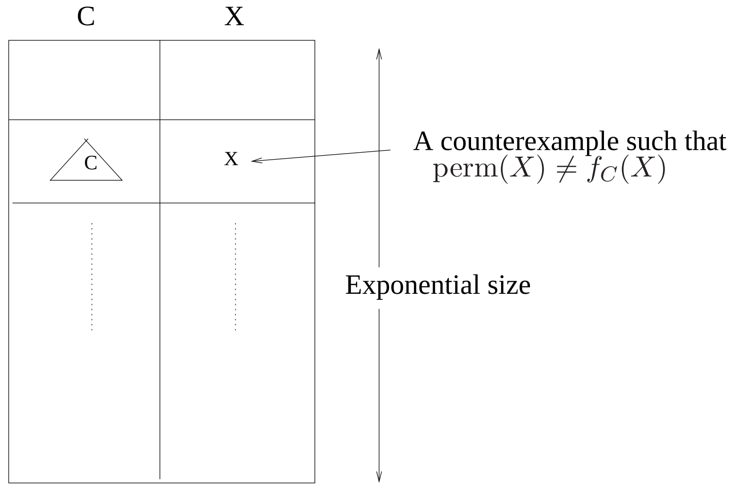


Figure 1.6: A trivial obstruction

This process can be extended further. Assuming an additional positivity hypothesis PH4 specified below, $OH$, whose associated decision problem $DP(OH)$ belongs to $NP$, can be transformed to POH (Positivity Obstruction Hypothesis), whose associated decision problem $DP(POH) \in P$; i.e., whether a new obstruction exists for given $n$ in unary and $m$ in binary can then be decided in polynomial time, and if so, it can also be constructed in polynomial time; cf. Figure 1.5. (The hypothesis is called PH4 instead of PH2, because PH2 and PH3 are some other positivity hypotheses in GCT6 that complement PH1). Once this final positivity hypothesis POH is proven, the obstruction hypothesis is reduced to a tautology (FOH: Final Obstruction Hypothesis), which can be verified and constructed in $O(1)$ time–i.e., the associated decision problem $DP(FOH)$ is $O(1)$-time solvable. This would then give us the final $O(1)$-size proof.

Thus the basic idea of the flip is to use positivity systematically as a means to eliminate the quantifiers and reduce the complexity of the decision problem associated with the obstruction hypothesis until the obstruction hypothesis is finally reduced to an $O(1)$-time verifiable tautology; cf. Figure 1.5.

Now let us specify PH4 and POH. Towards that end, let $k \leq l = m^2$ be the length of $\lambda$. Define

$$
\begin{aligned}
\bar{P}_{n,m} &= \{\lambda : P_{\lambda,n,m} \neq \emptyset\} \subseteq \mathbb{C}^k, \\
\bar{Q}_m &= \{\lambda : Q_{\lambda,m} \neq \emptyset\} \subseteq \mathbb{C}^k.
\end{aligned} \tag{1.9}
$$

The following is a consequence of a fundamental result [7] in geometric invariant theory [38].

**Theorem 1.7.1** *The sets $\bar{P}_{n,m}$ and $\bar{Q}_m$ are convex polytopes in $\mathbb{C}^k$.*

Then:

**Hypothesis 1.7.2 PH4** *(GCT6) The membership problems for these polytopes belong to P and so also the problem of deciding if vol($\bar{P}_{n,m} \setminus \bar{Q}_m$), the volume of the relative complement $\bar{P}_{n,m} \setminus \bar{Q}_m$, is nonzero (positive); i.e. if $\bar{P}_{n,m} \not\subseteq \bar{Q}_m$. By polynomial time, we mean poly($n, l, \langle m \rangle$) time. This is poly($n$), if $m \leq 2^{o(n)}$ and $l =$ poly($n$).*

**Hypothesis 1.7.3 (POH)**

*For all $n \to \infty$, assuming $m = 2^{\log^a n}$, $a > 1$ fixed,*

$$
vol(\bar{P}_{n,m} \setminus \bar{Q}_m) > 0,
$$

*for $k = (n+1)^2$.*

## 1.8 How to prove PH?

There is a basic prototype of PH in representation theory, which will be described in detail towards the end of Lecture 3. We shall refer to it as Plethysm PH. It says that the stretching functions akin to $F_{\lambda,n,m}(k)$ and $G_{\lambda,m}(k)$ associated with fundamental multiplicities in representation theory called plethysm constants also have analogous positive convex representations.

This is known for a very special case of the plethysm constant called the Littlewood-Richardson (LR) coefficient $c^{\lambda}_{\alpha,\beta}$. It is defined to be the number of occurences of the Weyl module $V_\lambda(G)$ in the tensor product of $V_\alpha(G)$ and $V_\beta(G)$, considered as a $G$-module by letting $G$ act on each factor of the tensor product independently. The classical Littlewood-Richardson rule, which we shall refer to as LR PH, implies that the stretching function $\tilde{c}^{\lambda}_{\alpha,\beta}(k) = c^{k\lambda}_{k\alpha,k\beta}$

associated with the Littlewood-Richardson coefficient has a positive convex representation.

Plethysm PH happens to be a fundamental open problem of representation theory, older than the $P$ vs. $NP$ problem itself. It has been studied intensively in the last century, and is known to be formidable. And now, as explained the third lecture, it also turns out to be the heart of this approach towards the $P$ vs $NP$, the $\#P$ vs. $NC$ problems.

A basic plan to prove Plethysm PH is given in GCT6. It is partially implemented in GCT7 and 8. See Figure 1.7 for a pictorial depiction of the plan. It strives to extend the proof of LR PH based on the theory of the standard quantum group [12, 18, 24]. There it comes out as a consequence of a (proof of a) deep positivity result [19, 24], which we shall refer to as LR PH0. It says that the tensor product of two representations of the standard quantum group has a canonical basis [18, 24] whose structure coefficients are all positive [24] polynomials (i.e., polynomials with nonnegative coefficients). The only known proof of this result is based on the Riemann Hypothesis over finite fields proved in [10], and the related works [5]. This Riemann Hypothesis over finite fields is itself a deep positivity statement in mathematics, from which LR PH can thus be deduced, as shown on the bottom row of Figure 1.7.

If one were only interested in LR PH, one does not need this powerful machinery, because it has a much simpler algebraic-combinatorial proof. But the plan to extend the proof of LR PH to Plethysm PH in GCT6,7,8 is like a huge inductive spiral. To make it work, one needs a stronger inductive hypothesis than Plethysm PH–this is precisely Plethysm PH0 (which will be described in the third lecture; cf. Hypothesis 3.9.4). Thus what is needed now is a systematic lifting of the bottom arrow in Figure 1.7 to the top, so as to complete the commutative diagram, so to speak.

Initial steps in this direction have been taken in GCT7,8. First, GCT7 constructs a *nonstandard quantum group*, which generalizes the notion of a standard quantum group [12], and plays the same role in the plethysm setting that the standard quantum group plays in the LR setting. Second, GCT8 gives an algorithm to construct a canonical basis for a representation of the nonstandard quantum group that is conjecturally correct and has the property Plethysm PH0, which is a generalization of LR PH0 supported by experimental evidence. Now what is needed to complete the commutative diagram in Figure 1.7 is an appropriate nonstandard extension of the Riemann hypothesis over finite fields and the related works [5, 10, 19, 24] from which Plethysm PH0 can be deduced. This–the top-right corner of the diagram–is the main open problem at the heart of this approach.

## 1.9 How to prove OH or POH?

We do not know, since the proof of OH and POH would really depend on the explicit forms of the polytopes that arise in PH1/PH4, and we have no idea about them at this point. GCT does suggest that proving OH/POH should be feasible "in theory", i.e., *theoretically feasible*, assuming PH1/4, since then DP(OH)/DP(POH) belong to NP/P and polynomial-time is complexity theory stands for feasible "in theory". In other words, GCT gives a reason to believe now that proving the $P \neq NP$ conjecture should be theoretically feasible. Even this was always questioned in the field of complexity theory so far, because the $P$ vs. $NP$ problem is a universal statement regarding all of mathematics (that says theorems cannot be proven automatically). But, as we also know by now, there is a huge gap between theory and practice–e.g., just because some problem is in $P$ does not necessarily mean that it is feasible in practice. Similarly, the actual implementation of the GCT flip via positivity is expected to be immensely difficult "in practice", as Figure 1.7 suggests.

## 1.10 Is positivity necessary?

Finally, if the positivity problems are so hard, one may ask if they can not be avoided somehow. Unfortunately, there is a formidable barrier towards the $P$ vs. $NP$ and related problems, called the *complexity barrier* [29, 28] which is universal in the sense that any approach towards these problems would have to tackle it, not just GCT. The flip, i.e., explicit construction of obstructions, is the most natural and obvious way to cross this barrier, and the natural way may well be among the most effective. The existing mathematical evidence suggests that any such natural approach to cross this barrier would have to say something deep regarding positivity (Plethysm PH/PH0) either explicitly or *implicitly*, even if the approach does not utter a word about algebraic geometry or representation theory. That is, Plethysm PH and PH0 may indeed be the heart of the fundamental lower bound problems in complexity theory; a detailed story and a precise meaning of the key phrase *implicit* would appear in the revised version of GCTflip.

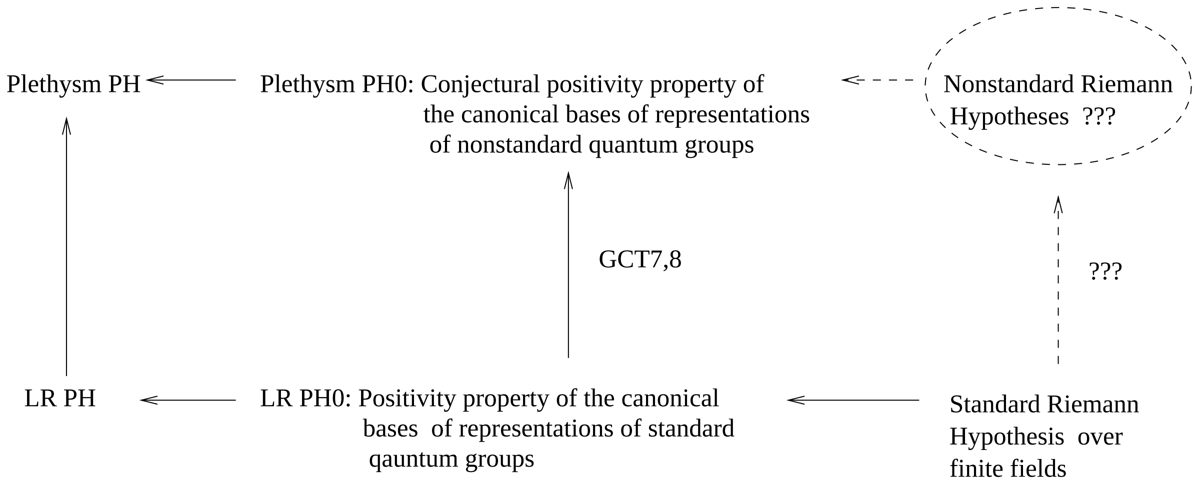


Figure 1.7: A commutative diagram

# Chapter 2

# Class varieties and obstructions

Let us begin by restating the permanent vs. determinant problem (characteristic zero) in a form that will be convenient here. Let $X$ be an $n \times n$ variable matrix. Let $Y$ be an $m \times m$ variable matrix, $m \geq n$. We assume that $X$ is the, say, bottom-right minor of $Y$, and $z$ is some entry of $Y$ outside $X$, which will be used as a homogenizing variable; cf. Figure 2.1. Let $M_{m^2}(\mathbb{C})$ denote the space of complex $m^2 \times m^2$ matrices. Suppose $m = 2^{\log^a n}$, $a > 1$ fixed, and $n \to \infty$. Then the problem is to show that there does not exist a matrix $A \in M_{m^2}(\mathbb{C})$ such that

$$\mathrm{perm}(X) z^{m-n} = \det(AY), \tag{2.1}$$

where, in the computation of $AY$, $Y$ is thought of as an $m^2$-vector after straightening it, say, columnwise, and the result is brought back to the matrix form to compute its determinant. It is easy to see that this problem is equivalent to the homogeneous restatement of the permanent vs. determinant problem in the last lecture. The best known lower bound on $m$ at present is quadratic [26].

The goal of this lecture:

**Goal 2.0.1** *(GCT1,2) Reduce the permanent vs. determinant problem to a problem in geometric invariant theory (GIT) so that we can then start applying the machinery of algebraic geometry and representation theory.*

Specifically,

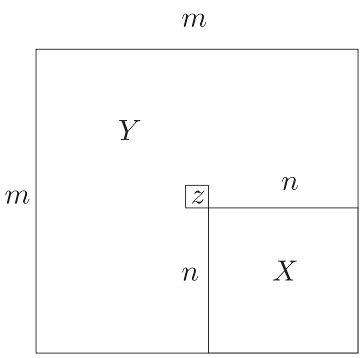


Figure 2.1: Variable matrix $Y$ and its submatrix $X$

1. Define the class varieties $X_{\#P}(n,m)$ and $X_{NC}(n,m)$ associated with the complexity classes $\#P$ and $NC$.
2. Define obstructions.
3. Reduce the permanent vs. determinant problem to the problem of showing existence of obstructions.

For this, we need to review some basic representation theory, algebraic geometry, and geometric invariant theory. The base field throughout is $\mathbb{C}$.

## 2.1 Basic representation theory

Let $G$ be a group. By a representation of $G$, we mean a vector space $W$ with a homomorphism from $G$ to $GL(W)$, the space of invertible linear transformations of $W$. It is called irreducible if it contains no nontrivial proper subrepresentation.

**Definition 2.1.1** *We say that $G$ is* reductive *if every finite dimensional* [1] *representation of $G$ is completely reducible; i.e., can be written as a direct sum of irreducible representations.*

All finite groups are reductive–a classical fact [14]. For example, let $S_2$ be the symmetric group on two symbols, and $\mathbb{C}^2$ its standard representation

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

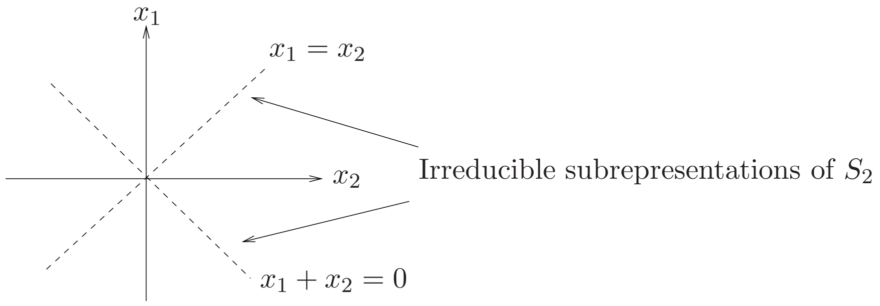


Figure 2.2: Decomposition of the standard representation of the symmetric group $S_2$

(permutation of the coordinates). Then $\mathbb{C}^2$ is a direct sum of two irreducible subrepresentations given by the lines $x_1 = x_2$ and $x_1 + x_2 = 0$; cf. Figure 2.2.

Weyl proved [14] that $G = GL_n(\mathbb{C})$, the general linear group of invertible $n \times n$ matrices, is reductive, so also $SL_n(\mathbb{C})$, the special linear group of invertible $n \times n$ matrices with determinant one.

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

$$W = \oplus_i m_i W_i, \tag{2.2}$$

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

Weyl also classified these building blocks. Specifically, he showed that the (polynomial[2]) irreducible representations of $G$ are in one-to-one correspondence with the partitions (integral sequences) $\lambda : \lambda_1 \geq \lambda_2 \geq \lambda_k > 0$ of length $k \leq n$; we denote this partition by $\lambda = (\lambda_1, \ldots, \lambda_k)$. It can be pictorially depicted by the corresponding Young diagram consisting of $\lambda_i$ boxes in the $i$-th row (Figure 2.3). An irreducible representation of $G$ in correspondence with a partition $\lambda$ is denoted by $V_\lambda(G)$, and is called a Weyl module.

For example, if $\lambda = (r)$, i.e., when the Young diagram consists of just one row of $r$ boxes, then $V_\lambda(G)$ is simply the space $\mathrm{Sym}^r(X)$ of all homogeneous forms of degree $r$ in the variables $x_1, \ldots, x_n$ with the following action of $G$.

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

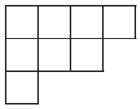

Figure 2.3: A Young diagram for the partition $(4,3,1)$

Given $f(X) \in \mathrm{Sym}^r(X)$ and $\sigma \in G$, map $f(X)$ to

$$f^\sigma(X) = f(X\sigma), \tag{2.3}$$

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

## 2.2 Basic algebraic geometry

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

## 2.3 Basic geometric invariant theory

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

$$f^\sigma(v) = f(\sigma^{-1}v). \tag{2.4}$$

(This is dual of the action in (2.3)). Here $\sigma^{-1}v$ denotes $\rho(\sigma^{-1})(v)$, where $\rho : G \to GL(V)$ is the representation map.

**Definition 2.3.1** *A projective variety $Y \subseteq P(V)$ is called a $G$-variety if the ideal $I(Y)$ is a $G$-submodule (i.e., a $G$-subrepresentation) of $\mathbb{C}[V]$.*

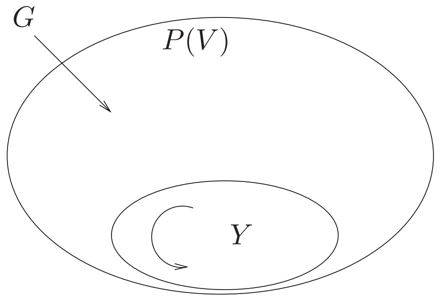


Figure 2.4: A $G$-subvariety of $P(V)$

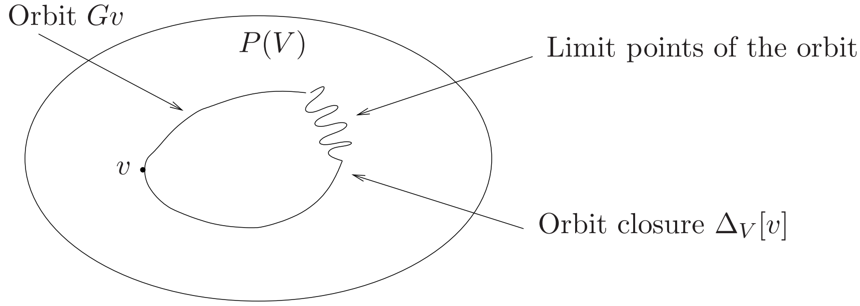


Figure 2.5: Orbit closure $\Delta_V[v]$

This means, under the action of $G$, the points of $Y$ are moved to the points within $Y$, i.e., each $\sigma \in G$ induces an automorphism of $Y$; cf. Figure 2.4.

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

$$Gv = \{gv \mid g \in G\}. \tag{2.5}$$

The orbit closure of $v$ is:

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

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

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

The algebraic geometry of the orbit closure $\Delta_V[v]$ for general $v$ is hopeless. It can be tractable only if $v$ is exceptional.

## 2.4 Class varieties and obstructions

We now construct the class varieties associated with the complexity classes $\#P$ and $NC$ as orbit closures of suitable exceptional points (the permanent and the determinant).

Let $X, Y, z$ be as in the beginning of this lecture; cf. Figure 2.1. Let $V = \mathrm{Sym}^m(Y)$ be the space of homogeneous forms of degree $m$ in the entries of $Y$. It is a representation of $G = GL(Y) = GL_l(\mathbb{C})$, $l = m^2$, with the following action. Given any $\sigma \in G$, map $g(Y)$ to $g^\sigma(Y) = g(\sigma^{-1}(Y))$:

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

Here $Y$ is thought of as an $m^2$-vector by straightening it, just as in (2.1).

Similarly, let $W = \mathrm{Sym}^n(X)$ be the space of forms of degree $n$ in the entries of $X$. It is a representation of $H = GL(X) = GL_{n^2}(\mathbb{C})$. We define an embedding $\phi : W \hookrightarrow V$ by mapping any $h(X) \in W$ to $h^\phi(Y) = z^{m-n}h(X)$. This also defines an embedding of $P(W)$ in $P(V)$, which we denote by $\phi$ again.

Let $g = \det(Y)$. We think of it as a point in $P(V)$. Let $h = \mathrm{perm}(X) \in P(W)$. Let $f = h^\phi = \mathrm{perm}^\phi(Y) \in P(V)$. Let

$$\begin{array}{lclcl} \Delta_V[g,m] & = & \Delta_V[g] & = & \overline{Gg} \subseteq P(V), \\ \Delta_W[h,n] & = & \Delta_W[h] & = & \overline{Hh} \subseteq P(W), \\ \Delta_V[f,m,n] & = & \Delta_V[f] & = & \overline{Gf} \subseteq P(V). \end{array} \tag{2.6}$$

We call $\Delta_V[g,m]$ the *class variety* associated with $NC$, since $\det(Y) \in NC$ and is $NC$-complete. It was denoted by $X_{NC}(n,m)$ in the previous lecture; notice that it actually depends only on $m$, and not on $n$ (the notation was chosen to make it look symmetric like what follows). We call $\Delta_V[f,n,m]$ the class variety associated with $\#P$. It was denoted by $X_{\#P}(n,m)$ in the previous lecture. We call $\Delta_W[h,n]$ the base class variety associated with $\#P$.

**Proposition 2.4.1** *(GCT1) If $h = perm(X)$, $X$ an $n \times n$ matrix, can be expressed linearly as a determinant of an $m \times m$ matrix, $m > n$, then*

$$f \in \Delta_V[g,m] = \Delta_V[g], \tag{2.7}$$

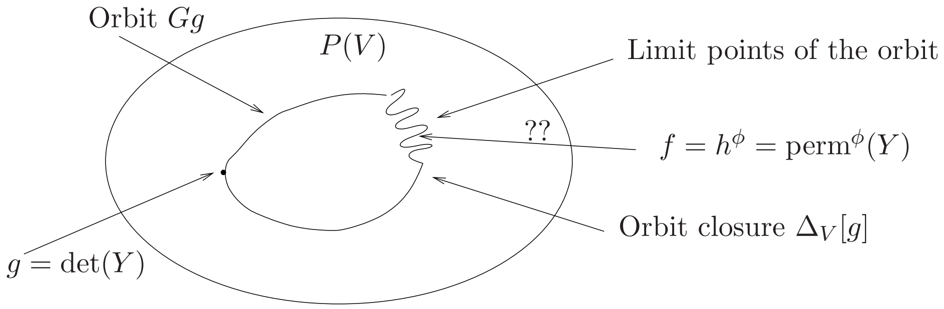


Figure 2.6: Does $f = h^\phi = \text{perm}^\phi(Y) \in \Delta_V[g]$?

*or equivalently,*

$$\Delta_V[f] = \Delta_V[f, n, m] \subseteq \Delta_V[g.m] = \Delta_V[g]. \tag{2.8}$$

*Conversely, if* $f \in \Delta_V[g, m]$*, then* $f$ *can be approximated infinitely closely by an expression of the form* $\det(AY)$*,* $A \in G$*, thinking of* $Y$ *as an* $m^2$*-vector.*

The first statement follows because $G = GL_{m^2}(\mathbb{C})$ is dense in $M_{m^2}(\mathbb{C})$, and the second because the $G$-orbit of $g$ is dense in $\Delta_V[g, m]$.

**Conjecture 2.4.2** *(GCT1) If* $m = 2^{\log n}$*,* $a > 1$ *fixed,* $n \to \infty$*, then* $\Delta_V[f, n, m] \not\subseteq \Delta_V[g, m]$.

By Proposition 2.4.1, this would solve the permanent vs. determinant problem in characteristic zero.

## How to prove the conjecture?

Suppose to the contrary:

$$\Delta_V[f, n, m] = \Delta_V[f] \subseteq \Delta_V[g] = \Delta_V[g, m]. \tag{2.9}$$

Then, by basic algebraic geometry, there is a surjective homomorphism from the homogeneous coordinate ring $R_V[g]$ of $\Delta_V[g]$ to the homogeneous coordinate ring $R_V[f]$ of $\Delta_V[f]$ obtained by restriction (Figure 2.7). Pictorially:

$$R_V[f, n, m] = R_V[f] \leftarrow R_V[g] = R_V[g, m]. \tag{2.10}$$

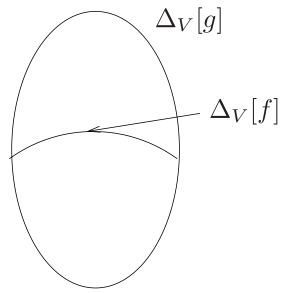


Figure 2.7: Restriction map

Furthermore, since the surjection is degree preserving, we get a similar surjection among the degree-$d$ components:

$$R_V[f,n,m]_d = R_V[f]_d \leftarrow R_V[g]_d = R_V[g,m]_d. \tag{2.11}$$

Since $\Delta_V[f]$ and $\Delta_V[g]$ are $G$-varieties, these are (finite-dimensional) $G$-modules. Furthermore, the homomorphism is a $G$-homomorphism, again by basic algebraic geometry. By dualizing, we get an injective $G$-homomorphism from the dual $R_V[f]_d^*$ of $R_V[f]_d$ to that of $R_V[g]_d$:

$$R_V[f,n,m]_d^* = R_V[f]_d^* \hookrightarrow R_V[g]_d^* = R_V[g,m]_d^*. \tag{2.12}$$

**Definition 2.4.3** *(GCT2) A Weyl module $S = V_\lambda(G)$ is called an* obstruction *for the inclusion (2.9), or an obstruction for the pair $(f,g)$, if $V_\lambda(G)$ occurs as a $G$-submodule in $R_V[f,n,m]_d^*$ but not in $R_V[g,m]_d^*$, for some $d$. We call $\lambda$ an* obstruction label*, and sometimes, simply an obstruction as well.*

Here occurs means the multiplicity of $V_\lambda(G)$ in $R_V[f,n,m]_d^*$ is nonzero (cf. eq. (2.2)).

If an obstruction exists for the pair $(f,g)$, for given $n$ and $m$, then the inclusion (2.9) is not possible. So the strategy to prove Conjecture 2.4.2 is to prove existence of such obstructions when $m$ is not too large.

## 2.5 Why should obstructions exist?

But, a priori, it is not at all clear why such obstructions should even exist. They are expected to exist only because the class varieties $\Delta_V[f]$, $f = h^\phi$, and $\Delta_V[g]$ are exceptional, since $h = \mathrm{perm}(X)$ and $g = \det(Y)$ are exceptional (cf. Section 1.1). Next, we wish to describe in what sense the class varieties are exceptional.

### 2.5.1 Exceptional orbit closures (group-theoretic varieties)

For that, we need to introduce the general notion of exceptional orbit closures.

Let $V$ be any finite dimensional representation of $G$, $v \in P(V)$ a point, and $\hat{v}$ any nonzero point on the line in $V$ corresponding to $v$. Let $H = G_{\hat{v}}$ be the stabilizer of $\hat{v}$:

$$H = G_{\hat{v}} := \{\sigma \in G \mid \sigma\hat{v} = \hat{v}\}.$$

**Definition 2.5.1** *(GCT1) We say that $v$ is characterized by its stabilizer $H$ if $\hat{v}$ is the only point (fix point) in $V$ such that $hv = v$ for all $h \in H$.*

**Observation 2.5.2** *If $\hat{v}$ is completely characterized by its stabilizer then the orbit closure $\Delta_V[v]$ is completely determined by the associated group triple:*

$$H = G_{\hat{v}} \hookrightarrow G \to K = GL(V), \tag{2.13}$$

*where the second arrow corresponds to the representation of $G$ on $V$.*

Because, once we know $K$, we know $V$ (upto dual). And once we know the embeddings $G \to K$ and $H \to G$, we know $\hat{v} \in V$, it being the only fix point of $H$ in $V$. We call (2.13) the *group-triple associated with* $\Delta_V[g]$. We also call $H \to G$ the *associated primary couple*, and $G \to K$ the *associated secondary couple*.

**Definition 2.5.3** *The orbit closure $\Delta_V[v]$, when $\hat{v}$ is completely characterized by its stabilizer, is called a* group-theoretic *variety.*

Coming back to the class varieties:

**Proposition 2.5.4** *(GCT1)*

*(1) The determinant $\hat{g} = \det(Y) \in V$, $V = Sym^m(Y)$, is completely characterized by its stabilizer $G_{\hat{g}} \subseteq G = GL(Y) = GL_{m^2}(\mathbb{C})$. Hence the class variety $\Delta_V[g]$ is group-theoretic.*

*Similarly, $\Delta_W[h]$, $h = perm(X) \in P(W)$, $W = Sym^n(X)$, and $\Delta_V[f]$, $f = h^\phi$, are group-theoretic.*

*Proof:* Based on classical invariant (representation) theory.

(1) It is known that the stabilizer of $\hat{g} = \det(Y)$ in $G = GL(Y)$ is the subgroup $G_{\hat{g}}$ generated by linear transformations of the form:

$$Y \to AY^*B, \quad Y^* = Y \text{ or } Y^t, \quad A, B \in GL_m(\mathbb{C}),$$

with $\det(A)\det(B) = 1$. Ignoring this restriction, the continuous part $G^0_{\hat{g}}$ of $G_{\hat{g}}$ is essentially $GL_m(\mathbb{C}) \times GL_m(\mathbb{C})$ embedded in $G = GL_{m^2}(\mathbb{C})$ naturally:

$$G^0_{\hat{g}} = GL_m(\mathbb{C}) \times GL_m(\mathbb{C}) \hookrightarrow GL(\mathbb{C}^m \otimes \mathbb{C}^m) = GL_{m^2}(\mathbb{C}).$$

It follows from classical representation theory that $\hat{g}$ is the only fix point of $G_{\hat{g}}$ in $V$.

(2) The stabilizer of $\hat{h} = \text{perm}(X) \in \text{Sym}^n(X)$ in $H = GL(X)$ is the subgroup $H_{\hat{h}}$ generated by linear transformations of the form:

$$X \to \lambda X^* \mu, \quad X^* = X \text{ or } X^t,$$

where $\lambda$ and $\mu$ are either diagonal or permutation matrices, with obvious constraints on the the product of the diagonal entries when they are diagonal.

The discrete part $H^d_{\hat{h}}$ of $H_{\hat{h}}$ is isomporphic to $S_n \times S_n$, $S_n$ the symmetric group, embedded in $H = GL_{n^2}(\mathbb{C})$ naturally:

$$H^d_{\hat{h}} = S_n \times S_n \hookrightarrow GL(\mathbb{C}^n \otimes \mathbb{C}^n) = GL_{n^2}(\mathbb{C}).$$

Again, by classical representation theory, $\hat{h}$ is the only fix point of $H_{\hat{h}}$ in $W$.

(3) The stabilizer $G_{\hat{f}}$ of $\hat{f} = \hat{h}^\phi \in V$ in $G = GL(Y)$ consists of linear transformations of the form $Y \to AY$, thinking of $Y$ is an $m^2$-vector in

which the $n^2$ entries of its submatrix $X$ come last, preceded by the entry $z \in Y \setminus X$, and $A \in GL_{m^2}(\mathbb{C})$ is a matrix of the form

$$\begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & 0 & a \end{bmatrix}$$

with $a \in H_{\hat{h}} \subseteq GL(X)$ (upto a constant multiple), and $det(A)$ suitably restricted. The middle $*$ here acts on $z$, $a$ on the $X$-part of $Y$, and the $*$'s in the first column on the $Y \setminus (X \cup \{z\})$ part of $Y$. Again $\hat{f}$ is the only fix point of $G_{\hat{f}}$. Q.E.D.

### 2.5.2 On existence of obstructions

The main point of Proposition 2.5.4 is that the information in the class varieties is completely captured by the associated group triples. Pictorially:

$$\begin{array}{ll} \Delta_V[g] & \cong G_{\hat{g}} \hookrightarrow G \hookrightarrow K = GL(V), \\ \Delta_V[f] & \cong G_{\hat{f}} \hookrightarrow G \hookrightarrow K, \\ \Delta_W[h] & \cong H_{\hat{h}} \hookrightarrow H \hookrightarrow L = GL(W), \end{array} \tag{2.14}$$

where $\cong$ denotes equivalence at the level of information; i.e., there is no-information-loss.

Furthermore, by Tannakian duality [11], (algebraic) groups are determined by their representations; pictorially:

$$\text{Tannakian duality: Groups} \longleftrightarrow \text{Representations} \tag{2.15}$$

Thus the determinant and the permanent are encoded by the associated group triples with no information loss, and the triples, in turn, are encoded by the associated representation theories again with no information loss. This means the algebraic geometry of the class varieties is, in principle, completely determined by the geometric representation theory of the associated group triples. Hence the difference between the class varieties–which is what Conjecture 2.4.2 is all about–should be reflected as a difference between the representation theories of the associated group triples. This is why obstructions, which can be thought as representation-theoretic “differences”, should exist. See GCT2 for precise mathematical results and conjectures in the Tannakian spirit supporting this intuition.

This leads to:

**Conjecture 2.5.5** *(GCT2) An obstruction (label) $\lambda_n$ exists for all $n \to \infty$, if $m = 2^{\log^a n}$, $a > 1$ fixed.*

This implies Conjecture 2.4.2.

The basic plan of GCT now is:

1. Understand geometric representation theory of the group triples associated with the class varieties in depth using (nonstandard) quantum groups.
2. Translate this understanding to understand the algebraic geometry of the class varieties.
3. Use this understanding to find obstructions as in Conjecture 2.5.5.

## 2.6 The flip

The following is a stronger form of Conjecture 2.5.5:

**Conjecture 2.6.1 [PHflip]** *(cf. GCT6 and GCTflip) There exists an explicit family $\{\lambda_n\}$ of obstructions (labels), if $m = 2^{\log^a n}$, $a > 1$ fixed, $n \to \infty$.*

Here explicit means feasible: i.e., short and easy to verify:

1. Short: the bitlength $\langle \lambda_n \rangle$ of $\lambda_n$ is poly$(n) = n^b$, for some fixed $b$, regardless of what $m$ is, as long as it is not too large as above.
2. Easy to verify: The problem of verifying obstruction-labels belongs to $P$. That is, given $n, m$ and $\lambda_n$, whether $\lambda_n$ is a valid obstruction-label that can belong to the above family can be decided in poly$(\langle \lambda_n \rangle, n)$ time, again regardless of what $m$ is, as long as it is not too large.

Here one may only consider a restricted class of obstructions (labels), and the verification algorithm may only verify if the given label $\lambda$ belongs to that restricted class in polynomial time. This is fine as long as such restricted $\lambda_n$ exists for every $n$.

The conjecture suggests the following basic strategy, called the *flip* (cf. GCT6, GCTflip), for proving existence of obstructions:

1. Find an "easy" criterion for verifying (recognizing) an obstruction (possibly restricted). Here easy means:
   (a) Easy in theory: polynomial-time, and
   (b) Easy in practice: usable in the next step.
2. Use this criterion to show existence of an explicit family of obstructions.
3. More strongly (optional), show how to construct an explicit $\lambda_n$ for each $n$ in poly($n$) time; we call such a family $\{\lambda_n\}$ a strongly explicit family of obstructions.

Thus the flip reduces the hard nonexistence problem to the "easy" existence problem for obstructions.

## 2.7 The $P$-barrier

By divine justice, finding such "easy" criterion for verification is an extremely hard problem.

To see why, let us examine the basic decision problems that arise in the context of verification of obstructions.

**Problem 2.7.1 (Basic decision problems)**

*(a) Given $\lambda, n, m$, does $V_\lambda(G)$ occur in $R_V[f, n, m]$?*

*(b) Given $\lambda, m$, does $V_\lambda(G)$ occur in $R_V[g, m]$?*

Actually, the following relaxed forms of these would suffice for our purpose:

**Problem 2.7.2 (Relaxed basic decision problems)**

*(a)' Given $\lambda, n, m$, does $V_{k\lambda}(G)$, for some integer $k \geq 1$, occur in $R_V[f, n, m]$? If so, find one such $k$.*

*(b)' Given $\lambda, m$, does $V_{k\lambda}(G)$, for some integer $k \geq 1$ occur in $R_V[g, m]$? If so, find one such $k$.*

We need efficient polynomial-time algorithms for these relaxed decision problems.

To see the main difficulty here, observe that the dimension of the ambient space $P(V)$ is

$$M = \dim(P(V)) = \binom{m^2+m-1}{m-1} = \exp(m^2), \tag{2.16}$$

when $V = \mathrm{Sym}^m(Y)$, and $Y$ is $m^2 \times m^2$ variable matrix. Thus $M$ is the number of monomials in $m^2$ variables of degree $m$ (minus one actually). Furthermore, by a classical formula of Weyl [14],

$$\dim(V_\lambda(G)) = O(\exp(m, \langle\lambda\rangle)) = 2^{O(m+\langle\lambda\rangle}. \tag{2.17}$$

Currently the best unconditional algorithms for (a), (b), (a)', or (b)', based on general-purpose algorithms in algebraic geometry and representation theory take $O(\dim(\mathbb{C}[V]_s))$ space, $s = |\lambda| = \sum_i \lambda_i$ (the size of $\lambda$). This is roughly $s^M$, i.e., exponential in $M$ and hence double exponential in $m$. The time taken is exponential in space, and hence, triple exponential in $m$.

We cannot expect much better using such general-purpose algorithms, since they all use Grobner basis algorithms, and the problem of constructing Grobner bases is EXPSPACE-complete [25]; here EXP means exponential in the dimension $M$.

Thus to get polynomial time algorithms for (a)' and (b)', we have to address:

**Problem 2.7.3 [The $P$-barrier]** *(cf. GCT6, GCTflip)*

*Bring this running time down from triple exponential in $m$ to polynomial in $n$.*

This is a massive task. For general $g$ and $h$, it is impossible–i.e., the problems (a)' and (b)' are hopeless–for the reasons given above. We refer to this as the GIT chaos (GIT=Geometric Invariant Theory); cf. Figure 2.8. Conjecture 2.6.1 says, against such odds, that this task should still be possible for the exceptional $g = \det(Y)$ and $h = \mathrm{perm}(X)$ that arise in GCT, and also for similar functions characterized by their symmetries that arise in the context of the $P$ vs. $NP$ problem.

Thus the main question here is:

**Question 2.7.4** *How to cross this $P$-barrier?*

GCT6 gives a plan for crossing this barrier assuming certain mathematical positivity hypotheses. This will be the subject of the next lecture.

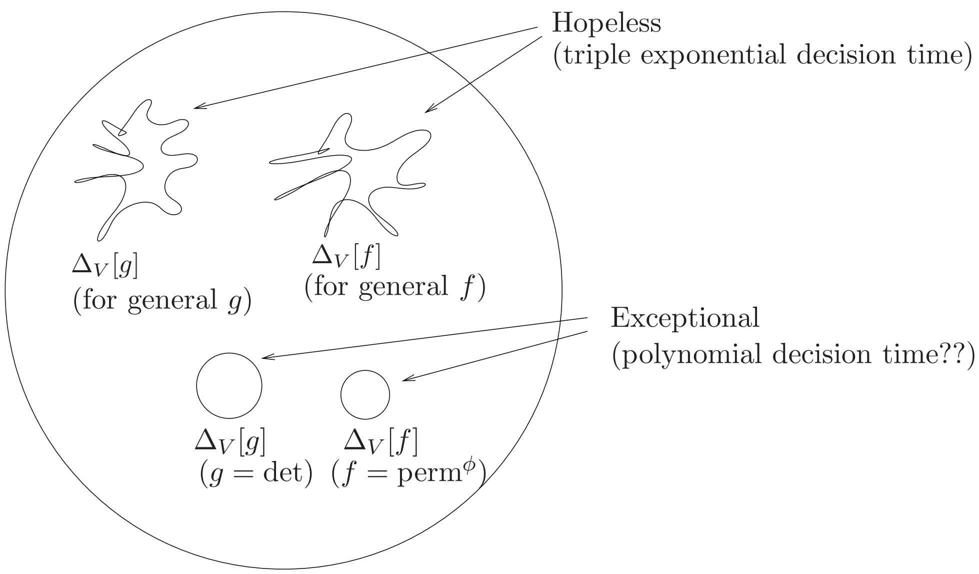


Figure 2.8: GIT chaos and the $P$-barrier

# Chapter 3

# Positivity

In this lecture, we study positivity hypotheses in mathematics in the context of the problem of showing existence of obstructions (Conjecture 2.5.5) and the $P$-barrier (Section 2.7).

Henceforth, we let $G = SL(Y)$ instead of $GL(Y)$ and $H = SL(X)$ instead of $GL(X)$. This makes no essential difference since our ambient space is $P(V)$, and two points in $V$ differing by a nonzero scalar correspond to the same point in $P(V)$. Thus everything discussed in the first two lectures goes through for this $G$ and $H$ as well. This lecture assumes more familiarity with representation theory that in the previous lectures; Appendix covers the additional concepts needed here.

## 3.1 On the $G$-module structure of the homogeneous coordinate rings of the class varieties

Given $v \in P(V)$, let $\hat{v}$ denote any nonzero point on the line corresponding to $v$ in $P(V)$. Let $\hat{g} = \det(Y) \in V$ (not $P(V)$), and $\hat{h} = \mathrm{perm}(X) \in W$ (not $P(W)$). Let $G_{\hat{g}} \subseteq G$ and $H_{\hat{h}}$ be their stabilizers.

**Theorem 3.1.1** *(GCT2)*

*(1) $V_\lambda(G)$ occurs in $R_V[g]^*$ (i.e., in $R_V[g]^*_d$, for some $d$) iff it contains a $G_{\hat{g}}$-invariant (i.e, a trivial subrepresentation–a fix point).*

*(2) $V_\pi(H)$ occurs in $R_W[h]^*$ iff $V_\pi(H)$ contains an $H_{\hat{h}}$-invariant.*

This reduces some questions concerning algebraic geometry of the class varieties to those concerning representation theory of the associated group triples (cf. Observation 2.5.2 and the remarks after it)–or rather, the associated primary couples in this case–in keeping with the basic plan discussed in Section 2.5.2.

It is easy to see that $G_{\hat{g}}$ is reductive (Definition 2.1.1) from its description in the proof of Proposition 2.5.4. Hence $V_\lambda(G)$ contains a $G_{\hat{g}}$-invariant iff the dual $V_\lambda(G)^*$ does. Thus, this theorem also holds if we replace $V_\lambda(G)$ and $V_\pi(H)$ by $V_\lambda(G)^*$ and $V_\pi(H)^*$, respectively.

*Proof:* We will only prove (1), (2) being similar. Let $\hat{\Delta}_V[g,m] \subseteq V$ denote the affine cone of $\Delta_V[g,m] \subseteq P(V)$. This is the union of all lines in $V$ corresponding to the points in $\Delta_V[g,m]$. Thus $R_V[g,m]$, the homogeneous coordinate ring of $\Delta_V[g,m]$, can also be thought of as the coordinate ring of $\hat{\Delta}_V[g,m]$.

(A) [The trivial part]: Suppose $V_\lambda(G)$ occurs in $R_V[g,m]$. The goal is to show that $V_\lambda(G)$ contains a $G_{\hat{g}}$-invariant.

Fix any copy $S$ of $V_\lambda(G)$ in $R_V[g,m]$.

**Claim 3.1.2** *Not all functions in $S$ can vanish at $\hat{g}$.*

Suppose to the contrary. Then, since $S$ is a $G$-module, all functions in $S$ vanish on the orbit $G\hat{g} \subseteq V$ as well. By homogeneity of the functions in $S$, then vanish on the cone of $G\hat{g}$ in $V$. But this cone is dense in $\hat{\Delta}_V[g,m]$, since $Gg$ is dense in $\Delta_V[g,m]$. Thus all functions in $S$ vanish on $\hat{\Delta}_V[g,m]$, and hence, $S$ cannot occur in $R_V[g,m]$; a contradiction. This proves the claim.

Now consider the evaluation map at $\hat{g}$:

$$\psi : S \to \mathbb{C},$$

which maps every function in $S$ to its value at $\hat{g}$. It belongs to $S^*$, the dual of $S$. It is $G_{\hat{g}}$-invariant since $\hat{g}$ is fixed by $G_{\hat{g}}$. Thus $S^*$, and hence $S$, contains a nonzero $G_{\hat{g}}$-invariant. This proves (A).

(B) [The nontrivial part]: Suppose $V_\lambda(G)$ contains a $G_{\hat{g}}$-invariant. The goal is to show that it occurs in $R_V[g,m]$.

For this we need the notion of stability in geometric invariant theory [38], which we now recall.

Let $Z$ be a finite dimensional $G$-representation, $G = SL_l(\mathbb{C})$.

**Definition 3.1.3** *[38] A point $z \in Z$ is called stable with respect to the $G$-action if the orbit $Gz$ is closed in $Z$ in the complex (equivalently, Zariski) topology on $Z$.*

Example: Let $Z = M_l(\mathbb{C})$, the space of $l \times l$ complex matrices, with the adjoint action of $G$ given by:

$$z \rightarrow \sigma z \sigma^{-1},$$

for any $z \in Z$ and $\sigma \in G$. Then it can be shown that $z \in Z$ is stable iff $z$ is diagonalizable. For example, under the action of

$$\sigma = \begin{bmatrix} t & 0 \\ 0 & t^{-1} \end{bmatrix},$$

we have:

$$z = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \underset{\sigma}{\longrightarrow} \begin{bmatrix} 1 & at^2 \\ 0 & 1 \end{bmatrix} \underset{t \rightarrow 0}{\longrightarrow} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus the orbit of the nondiagonalizable $z$ contains a diagonalizable limit point, which cannot be contained in the orbit. Hence $z$ is not stable.

Most points in any representation $Z$ of $G$ are stable [38]. The nontrivial problem is to show that a specific $z \in Z$ is stable. For this, there is a very useful Hilbert-Mumford-Kempf criterion of stability [38], using which can be proved:

**Theorem 3.1.4** *(GCT1) The point $\hat{g} = \det(Y) \in V = Sym^m(Y)$ is stable with respect to the action of $G = SL(Y)$. That is, the orbit $G\hat{g} \subseteq V$ is closed in $V$.*

Now let us get back to (B). Since $\hat{g}$ is stable, the orbit $G\hat{g}$ is closed in $V$, and hence in $\hat{\Delta}_V[g] = \hat{\Delta}_V[g,m] \subseteq V$. That is, the orbit $G\hat{g}$ is a closed affine subvariety of $\hat{\Delta}_V[g,m]$. Hence, there is a surjective $G$-homomorphism from the coordinate ring $R_V[g,m]$ of $\hat{\Delta}_V[g,m]$ to the coordinate ring $\mathbb{C}[G\hat{g}]$ of $G\hat{g}$.

It suffices to show that $S = V_\lambda(G)$ occurs in $\mathbb{C}[G\hat{g}]$. Now $G\hat{g} \cong G/L$, where $L = G_{\hat{g}}$ is stabilizer of $\hat{g}$. By the algebraic form of the Peter-Weyl theorem [14], the coordinate ring $\mathbb{C}[G]$ of $G$ (considered as an affine variety) decomposes as a $G$-module:

$$\mathbb{C}[G] \cong \oplus_\alpha V_\alpha(G) \otimes V_\alpha(G)^*.$$

Now

$$\mathbb{C}[G/L] = \mathbb{C}[G]^L,$$

the ring of $L$-invariants in $\mathbb{C}[G]$. Thus,

$$\mathbb{C}[G\hat{g}] = \mathbb{C}[G/L] = \mathbb{C}[G]^L = \oplus_\alpha V_\alpha(G) \otimes [V_\alpha(G)^*]^L.$$

Therefore $V_\alpha(G)$ occurs in $\mathbb{C}[G\hat{g}]$ iff $V_\alpha(G)^*$ contains an $L$-variant. Since $L$ is reductive, this is so iff $V_\alpha(G)$ contains an $L$-invariant. Thus $S = V_\lambda(G)$ occurs in $\mathbb{C}[G\hat{g}]$.

This implies (B), and proves Theorem 3.1.1.

We now wish to state a similar result for the coordinate ring $R_V[f, n, m]$, $f = h^\phi = \text{perm}^\phi$. For that, we need a few definitions. Let $W = \text{Sym}^n(X)$ and $V = \text{Sym}^m(Y)$ be as above. Furthermore, let $\bar{W} = \text{Sym}^m(\bar{X})$, where $\bar{X}$ is the $(n+1) \times (n+1)$ bottom-right submatrix of $Y$ containing $X$ and $z$ in Figure 2.1. Let $\bar{H} = SL(\bar{X})$. Thus we have

$$H = SL(X) \subseteq \bar{H} = SL(\bar{X}) \subseteq G = SL(Y).$$

Let $h(X) = \text{perm}(X)$, and $\bar{h}(\bar{X}) = z^{m-n} h(X) \in P(\bar{W})$. Let

$$\Delta_{\bar{W}}[\bar{h}] = \Delta_{\bar{W}}[\bar{h}, n, m] \subseteq P(\bar{W})$$

be the closure of the orbit $\bar{H}\bar{h}$. Let $R_{\bar{W}}[\bar{h}] = R_{\bar{W}}[\bar{h}, n, m]$ be its homogeneous coordinate ring.

**Theorem 3.1.5** *(GCT2)*

*(a) $V_\lambda(G)$ occurs in $R_V[f, n, m]^*$ iff the length of $\lambda$ is at most $(n+1)^2$ and $V_\lambda(\bar{H})$ occurs in $R_{\bar{W}}[\bar{h}, n, m]^*$.*

*(b) If $V_\lambda(\bar{H})$ occurs in $R_{\bar{W}}[\bar{h}, n, m]^*$, then it contains as a subrepresentation an $H$-module $V_\alpha(H)$ containing an $H_{\hat{h}}$-invariant, where $H_{\hat{h}} \subseteq H$ is the stabilizer of $\hat{h}$.*

*(c) Conversely, if $V_\alpha(H)$ contains an $H_{\hat{h}}$-invariant, there exists a $\lambda$ lying over $\alpha$ such that $V_\lambda(\bar{H})$ occurs in $R_{\bar{W}}[\bar{h}, n, m]^*$, and hence, $V_\lambda(G)$ occurs in $R_V[f, n, m]^*$. Here lying over means (a) the length of $\lambda$ is $\leq (n+1)^2$, and (b) $V_\alpha(H)$ occurs in $V_\lambda(\bar{H})$, considered as an $H$-module via the natural embedding $H = SL(X) \subseteq \bar{H} = SL(\bar{X})$.*

## 3.2 A mathematical form of the $\#P \neq NC$ conjecture

We now apply GCT to prove Theorem 1.4.1 (a) stated in the first lecture; the proof of Theorem 1.4.1 (b) is very similar, and is omitted here. The same proof technique actually yields a more general result.

To state it, we need a few definitions. Let $H = SL(X) = SL_{n^2}(\mathbb{C})$ as before, and let $\tilde{H} = SL_n(\mathbb{C}) \times SL_n(\mathbb{C})$ be embedded naturally in $SL(X) = SL(\mathbb{C}^n \otimes \mathbb{C}^n)$ (each $SL_n$ factor acts on the corresponding $\mathbb{C}^n$). Thus the representation $\mathrm{Sym}^n(X)$ for $H$ can also be considered to be a representation of $\tilde{H}$.

Let $\hat{t} = \mathrm{trace}(X^n) \in \mathrm{Sym}^n(X)$ and $\tilde{H}_{\hat{t}} \subseteq \tilde{H}$ its stabilizer. It consists of all linear transformations of the form:

$$X \to AXA^{-1}, \tag{3.1}$$

for all $A \in SL_n$. Thus $\tilde{H}_{\hat{t}} = SL_n$, embedded in $SL_n \times SL_n$ naturally:

$$\sigma \to (\sigma, (\sigma^{-1})^t),$$

for all $\sigma \in SL_n$. Let $h = \mathrm{perm} \in P(\mathrm{Sym}^n(X))$ as before, $\hat{h}$ the corresponding point in $\mathrm{Sym}^n(X)$, and $\tilde{H}_{\hat{h}} \subseteq \tilde{H}$ its stabilizer; it is esentially the stabilizer described in the proof of Proposition 2.5.4.

Let $W$ be any polynomial representation of $\tilde{H}^k = \tilde{H} \times \cdots \times \tilde{H}$ ($k$ copies of $\tilde{H}$). Let $w \in P(W)$ be a point, and $\hat{w} \in W$ any nonzero point on the line corresponding to $w$. We say that $w$ is a *generalized trace-like* point if $\hat{w}$ is an invariant of $\tilde{H}^k_{\hat{t}} = \tilde{H}_{\hat{t}} \times \cdots \times \tilde{H}_{\hat{t}}$ ($k$ copies of $\tilde{H}_{\hat{t}}$); i.e., $\tilde{H}^k_{\hat{t}} \subseteq \tilde{H}^k_{\hat{w}}$. We say it is a *generalized permanent-like* point if similarly $\tilde{H}^k_{\hat{w}} \subseteq \tilde{H}^k_{\hat{h}}$. We say that it is a *generalized permanent* if $\tilde{H}^k_{\hat{w}} = \tilde{H}^k_{\hat{h}}$.

As an example, let $\mathbb{C}[X]$ be the ring of polynomial functions in the entries $x_{ij}$ of $X$, with the natural action of $\tilde{H} = SL_n \times SL_n$ (one factor acting on the left and the other on the right), the case of the more general ring $\mathbb{C}[X_1, \ldots, X_k]$ being similar. Let $\mathbb{C}[X]^{\tilde{H}_{\hat{t}}} \subseteq \mathbb{C}[X]$ be the subring of the invariants of $\tilde{H}_{\hat{t}}$; i.e., the subring of generalized trace-like points in $\mathbb{C}[X]$. It is generated by $\mathrm{trace}(X^j)$, $j \geq 0$, by (a variant of) the first fundamental theorem of invariant theory [14]. A generalized permanent in $\mathbb{C}[X]$ is essentially the same as a generalized permanent in Definition 1.2.2 (for $k = 1$). There is a slight difference between two definitions. In Definition 1.2.2 we let $U_i$ and $V_i$ be any matrices in $GL_n(\mathbb{C})$, whereas here we are taking them to be in $SL_n(\mathbb{C})$. Thus as per the definition in this section $\det(X)$ is a generalized permanent-like function, but not a generalized permanent. Everything in this section holds for a generalized permanent in Definition 1.2.2 as well. In what follows, we shall assume that a generalized permanent is as defined in this section.

Let $\mathbb{C}[X]^{\tilde{H}_{\hat{h}}} \subseteq \mathbb{C}[X]$ be the subring of invariants of $\tilde{H}_{\hat{h}}$; i.e., the subring of generalized permanent-like functions. By the classical result of Hilbert, it is finitely generated. No finite explicit set of generators for this ring is known (unlike for the ring of generalized-trace like functions). But an explicit basis for this ring is known. It is as follows. To every $n \times n$ magic square $A$ of weight $r$–i.e. a matrix of nonnegative integers whose each row and column sums to $r$–assign a *basic generalized permanent-like function*

$$p_A(X) = \sum_{A'} x_{A'},$$

where $A'$ ranges over all matrices obtained by permuting the rows and/or columns of $A$, and $x_{A'} = \prod_{ij} x_{ij}^{a'_{ij}}$, $a'_{ij}$ the entries of $A'$, denotes the monomial associated with $A'$. This is a $\#P$-computable and $\#P$-complete function of $A$ and $X$. Furthermore, the basic generalized permanent-like functions form a basis of $\mathbb{C}[X]^{\tilde{H}_{\hat{h}}}$. Not all generalized permanent-like functions are generalized permanents. For example, $p_A(X)$, when every entry of $A$ is one, is not a generalized permanent, since it has more symmetries than that of the permanent. But most generalized permanent-like functions would be generalized permanents.

Now let $W$ be any polynomial representation of $\tilde{H}^k$, $\tilde{H} = SL_n(\mathbb{C}) \times SL_n(\mathbb{C})$. Given any $\sigma \in \tilde{H}^k$ and $w \in P(W)$, let $w^\sigma = \rho(\sigma)(w)$, where $\rho : \tilde{H} \to GL(W)$ is the representation map. Since this map is polynomial, $w^\sigma$ is well defined for any $\sigma \in (M_n(\mathbb{C}) \times M_n(\mathbb{C}))^k$. Let $\tilde{\Delta}_W[w] \subseteq P(W)$ denote the orbit closure of $w$ with respect to the $\tilde{H}^k$ action; i.e., the closure of the orbit $\tilde{H}^k w$.

**Theorem 3.2.1** *Let $w$ be any generalized trace-like point in $W$, and and $h$ any generalized permanent in $W$. Then, for any $\sigma \in (M_n(\mathbb{C}) \times M_n(\mathbb{C}))^k$, $w^\sigma \neq h$. More generally, $\tilde{\Delta}_W[w]$ does not contain $h$.*

This reduces to Theorem 1.2.4 (a) when $W = \mathbb{C}[X_1, \ldots, X_k]$. The naturalizable proof in [32] based on basic geometric invariant theory also works in this case. There is a similar generalization of Theorem 1.2.4 (b); for this the naturalizable proof technique in [32] does not work. It is omitted here.

We now sketch the proof of Theorem 3.2.1 based on obstructions only for $W = \mathbb{C}[X]$. The details for the general case are similar and are left to the reader. This proof (unlike the naturalizable one in [32]) can be easily adapted to prove Theorem 1.2.4 (b); we will leave this to the reader.

For $W = \mathbb{C}[X]$, Theorem 3.2.1 follows from:

**Theorem 3.2.2** *There exists a family $\{O_n\}$ of obstructions in this case.*

*Proof:* Let $h \in \mathbb{C}[X]$ be any generalized permanent, and $w \in \mathbb{C}[X]$ any generalized trace-like point. The class varieties $\tilde{\Delta}_W[w], \tilde{\Delta}_W[h] \subseteq P(W)$ are now defined with respect to the $\tilde{H}$-action and the obstructions are $\tilde{H}$-Weyl-modules defined similarly. Let $\tilde{R}_W[h]$ and $\tilde{R}_W[w]$ be the homogeneous coordinate rings of $\tilde{\Delta}_W[w]$ and $\tilde{\Delta}_W[h]$.

It can be using shown using Kempf's criterion of stability [38] that $\hat{h}$ is stable with respect to the $\tilde{H}$ action–the proof of this fact is similar to the stability related proofs in GCT1. Specifically, in this setting Kempf's criterion in a concrete form says that $\hat{h}$ is stable if the standard irreducible representation $\mathbb{C}^n \otimes \mathbb{C}^n$ of $\tilde{H} = SL_n(\mathbb{C}) \times SL_m(\mathbb{C})$ is also an irreducible representation of its subgroup $\tilde{H}_{\hat{h}}$, which is easy to check. The crucial point here is that this proof needs to know only about the stabilizer of $\hat{h}$ and nothing else. Using stability of $\hat{h}$ it then follows from the general results in GCT2 that the analogue of Theorem 3.1.1 (2) holds for this $h$.

The stabilizer $\tilde{H}_{\hat{w}}$ contains the stabilizer $\tilde{H}_{\hat{t}} \subseteq \tilde{H}$ of $\hat{t} = \mathrm{trace}(X^n) \in \mathrm{Sym}^n(X)$ as described in (3.1).

We need the following two facts.

(a) Any irreducible $\tilde{H}$-module is of the form $V_\alpha(SL_n) \otimes V_\beta(SL_n)$. By the classical Schur's lemma it contains a $\tilde{H}_{\hat{t}}$-invariant iff $\alpha = \beta$;. Hence, it does not contain a $\tilde{H}_{\hat{w}}$-invariant if $\alpha \neq \beta$.

(b) (Cf. [4]) An irreducible representation of $\tilde{H}$ of the form $1 \otimes V_\gamma(SL_n)$, where 1 stands for the trivial representation of $SL_n$ and $|\gamma| = 2n$, contains a $\tilde{H}_{\hat{h}}$-invariant iff $\gamma$ is even–if $\gamma = (\gamma_1, \gamma_2, \cdots)$, this means every $\gamma_i$ is divisible by 2. Here $|\gamma| = \sum_i \gamma_i$ denotes the size of $\gamma$.

Let $\gamma$ be any even partition with $|\gamma| = 2n$. By (b) and Theorem 3.1.1 (2) (or rather its analogue in this case mentioned above), $1 \otimes V_\gamma(SL_n)$ occurs in $\tilde{R}_W[h]^*$. By (a), it does not contain a $\tilde{H}_{\hat{w}}$-invariant. By Theorem 3.1.1 (2) again (or rather its analogue in this case), it cannot occur in $\tilde{R}_W[w]^*$. Therefore, $1 \otimes V_\gamma(SL_n)$ is an obstruction. Q.E.D.

The proof above shows that any $1 \otimes V_\gamma(SL_n)$, $|\gamma| > 0$, which contains an $\tilde{H}_{\hat{h}}$-invariant is an obstruction. One can show nonconstructively, i.e., without using [4], that there is such $\gamma$ for every $n$. This then yields a nonconstructive proof of this result (whose major part is the same as in the explicit proof); a similar nonconstructive proof also works for Theorem 1.2.4 (b).

## 3.3 From the mathematical towards the general complexity theoretic form

We now discuss what is needed to lift the proof of the mathematical form of the $\#P \neq NC$ conjecture to the general complexity theoretic form. There are two issues.

(1) There is a serious leak in Theorem 3.1.5, because there can be several $\lambda$ lying over $\alpha$, and that result does not tell us exactly which one of them would occur in $R_{\bar{W}}[\bar{h}, n, m]$ or $R_V[f, n, m]^*$, nor does it tell us which $V_\lambda(G)$'s occur in $R_{\bar{W}}[\bar{h}, n, m]_d$ or $R_V[f, n, m]^*_d$, for a fixed $d$. Such refined information can be obtained from a general positivity hypothesis (PH: Hypothesis 1.5.2) for $R_V[f, n, m]$ (which was denoted by $R_{\#P}(n, m)$ in Lecture 1). We will discuss this issue in Section 3.7.

(2) To use Theorem 3.1.1 and Theorem 3.1.5 we need an effective criteria for:

**Problem 3.3.1** *(a) Does $V_\lambda(G)$ contain a $G_{\hat{g}}$-invariant?*
*(b) Does $V_\pi(H)$ contain an $H_{\hat{h}}$-invariant?*

These are special cases of the general subgroup restriction problem which we discuss next in the following section.

## 3.4 The subgroup restriction problem

Let $H$ be a reductive subgroup of $G = GL(V)$, where $V$ is an explicitly given finite dimensional representation of $H$. Symbolically:

$$H \overset{\rho}{\hookrightarrow} G = GL(V), \tag{3.2}$$

where $\rho$ denotes the representation map. For example, we can have $H = GL_n(\mathbb{C})$, and $V = V_\mu(H)$, the Weyl module of $H$. Then $\mu$ specifies the representation map $\rho$ completely, and hence, we shall also use $\mu$ in place of $\rho$ in this case–called the *plethysm* case. Symbolically:

$$H \overset{\mu}{\hookrightarrow} G = GL(V), \quad V = V_\mu(H). \tag{3.3}$$

Given any partition $\lambda$, the Weyl module $V_\lambda(G)$ of of $G$ can be considered an $H$-module via the representation map $\rho$. Since $H$ is reductive, it is completely reducible as an $H$-module:

$$V_\lambda(G) = \bigoplus_\pi a^\lambda_{\pi,\rho} V_\pi(H), \tag{3.4}$$

where $a^\lambda_{\pi,\rho}$ denotes the multiplicity of $V_\pi(H)$ in $V_\lambda(G)$. In the plethysm case, we also denote $a^\lambda_{\pi,\rho}$ by $a^\lambda_{\pi,\mu}$, and call it the *plethysm constant*.

**Problem 3.4.1 (Subgroup restriction problem)** *(1) Given partitions $\lambda, \pi$ and $\rho$, does $V_\pi(H)$ occur as a subrepresentation of $V_\lambda(G)$? That is, is $a^\lambda_{\pi,\rho}$ positive?*

*(2) Find a good positive formula for $a^\lambda_{\pi,\rho}$ akin to the usual positive formula for the permanent which does not have any alternating signs. What good and positive means would be elaborated later (cf. Hypothesis 3.6.2).*

**Problem 3.4.2 (Plethysm problem)** *The special case of the subgroup restriction problem for the representation map (3.3), obtained by replacing $\rho$ by $\mu$.*

The two special cases that arise in the context of Problem 3.3.1 are:

(1) Let $\hat{g} = \det(Y) \in \mathrm{Sym}^m(Y)$, $G = GL(Y) = GL_{m^2}(\mathbb{C})$, and $H = G_{\hat{g}} \subseteq G$, the stabilizer of $\hat{g}$; cf. the proof of Proposition 2.5.4 for its description. If we ignore the discrete (and torus) part of the stabilizer, then the subgroup restriction problem here is for the embedding:

$$GL_m \times GL_m \hookrightarrow GL(\mathbb{C}^m \otimes \mathbb{C}^m).$$

(2) Let $\hat{h} = \mathrm{perm}(X) \in \mathrm{Sym}^n(X)$, and $H = GL(X)$, and $H_{\hat{h}}$ the stabilizer of $\hat{h}$; cf. the proof of Proposition 2.5.4 for its description. If we ignore the

continuous part of the stabilizer, then the subgroup restriction problem here is for the embedding:

$$S_n \times S_n \hookrightarrow GL(\mathbb{C}^n \otimes \mathbb{C}^n),$$

where $S_n$ is the symmetric group on $n$ letters.

It is a classical result of representation theory that (1) can be reduced to the plethysm problem. By [4], (2) can also be reduced to the plethysm problem. So the plethysm problem is the fundamental special case of the subgroup restriction problem that we will be interested in (though the following results also hold for the general subgroup restriction problem).

## 3.5 Littlewood-Richardson problem

One completely understood special case of the subgroup restriction problem is the Littlewood-Richardson (LR) problem. This arises when $H = GL_n(\mathbb{C})$ embedded in $G = H \times H$ diagonally:

$$\begin{array}{ccl} H & \rightarrow & G = H \times H \\ \sigma & \rightarrow & (\sigma, \sigma). \end{array} \tag{3.5}$$

Then every irreducible representation of $G$ is of the form $V_\alpha(H) \otimes V_\beta(H)$. Considered as an $H$-module via the above diagonal embedding, it decomposes:

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

The multiplicities $c^\lambda_{\alpha,\beta}$ are called Littlewood-Richardson coefficients. Let $\tilde{c}^\lambda_{\alpha,\beta}(k) = c^{k\lambda}_{k\alpha,k\beta}$ be the associated stretching functions.

**Theorem 3.5.1**

1. **[LR PH1]** *There exists a polytope of $P^\lambda_{\alpha,\beta}$ with description of poly$(\langle\alpha\rangle, \langle\beta\rangle, \langle\lambda\rangle)$ bitlength such that:*
$$c^\lambda_{\alpha,\beta} = \#(P^\lambda_{\alpha,\beta}),$$
*the number of integer points in $P^\lambda_{\alpha,\beta}$, and*
$$\tilde{c}^\lambda_{\alpha,\beta}(k) = c^{k\lambda}_{k\alpha,k\beta} = \#(kP^\lambda_{\alpha,\beta}) = f_{P^\lambda_{\alpha,\beta}}(k),$$

*the Ehrhart quasipolynomial of $P^{\lambda}_{\alpha,\beta}$. This provides a good positive formula for the Littlewood-Richardson coefficients.*

2. **[Saturation Theorem]** *[22]: $c^{\lambda}_{\alpha,\beta} \neq 0$ iff $P^{\lambda}_{\alpha,\beta} \neq \emptyset$.*

3. *(GCT3,[22]) Given $\alpha, \beta, \lambda$, whether $c^{\lambda}_{\alpha,\beta}$ is nonzero (i.e. positive) can be decided in poly($\langle\alpha\rangle, \langle\beta\rangle, \langle\lambda\rangle$) time.*

Here the third statement follows from the first two by a polynomial time algorithm for linear programming [15].

## 3.6 Plethysm problem

Let us now focus on the plethysm problem (Problem 3.4.2). Let $\tilde{a}^{\lambda}_{\pi,\mu}(k) = a^{k\lambda}_{k\pi,\mu}$ be the stretching function associated with the plethysm constant $a^{\lambda}_{\pi,\mu}$. Let

$$A^{\lambda}_{\pi,\mu}(t) = \sum_{k \geq 0} \tilde{a}^{\lambda}_{\pi,\mu}(k)$$

be the associated generating function. It was asked in [21] if it is a rational function. The following result shows something stronger:

**Theorem 3.6.1** *(GCT6) The stretching function $\tilde{a}^{\lambda}_{\pi,\mu}(k)$ is a quasi-polynomial.*

This implies, in particular, that $A^{\lambda}_{\pi,\mu}$ is rational by a standard result of enumerative combinatorics [42].

The proof below is motivated by Brion's proof [9] of quasipolynomiality of the stretching functions associated with the Littlewood-Richardson coefficients (of arbitrary type).

*Proof:*

Let $H = GL_n(\mathbb{C})$, and $V = V_{\pi}(H)$. Let $U \subseteq H$ be the subgroup of lower triangular matrices with 1's on the diagonal. Then it is known (cf. Appendix) that there is a unique (up to constant multiple) nonzero point $\hat{v} = \hat{v}_{\pi} \in V$ that is stabilized by $U$; i.e., such that $u\hat{v} = \hat{v}$ for all $u \in U$. The point $\hat{v}$ is called the highest weight vector of $V_{\pi}(H)$. Let $v = v_{\pi}$ be the corresponding point in $P(V) = P(V_{\pi}(H))$. Then it is known that the orbit $Hv \subseteq P(V)$ is already closed. That is, the orbit closure $\Delta_V[v]$ (with respect to the $H$ action) is just the orbit $Hv$ itself. Furthermore, by Borel-Weil [14], the homogeneous coordinate ring $R_V[v]$ of $\Delta_V[v] = Hv$ has the following decomposition as an $H$-module:

$$R_V[v] = \oplus_k V_{k\pi}(H)^*, \tag{3.6}$$

where the superscript $*$ denotes the dual. We can also think of $R_V[v]$ as the coordinate ring of $\hat{\Delta}_V[v]$, the affine cone of $\Delta_V[v]$. It is known that the singularities of $\hat{\Delta}_V[v]$ are rational and normal; e.g., see [41].

*Remark:* By normal, we mean that for each $x \in X = \hat{\Delta}_V[v]$, there exists a (classical) neighbourhood $U \subseteq X$ of $x$, such that $U \setminus (U \cap \mathrm{sing}(X))$ is connected; where $\mathrm{sing}(X)$ is the subvariety of $X$ consisting of all its singular points. Rational is much more difficult to define. Roughly it means the following. By Hironaka [16], all singularities of $X$ can be resolved (untangled)– cf. Figure 1.4. With each singularity of $X$, one can associate a cohomological object that measures the difficulty of this resolution. A singularity is called rational if this cohomological object vanishes. This means the singularity is sufficiently nice.

By abuse of terminology, we say that the ring $R_V[v]$ in (3.6) is normal and rational. Similarly, it can be shown that the ring

$$S = \oplus_\pi V_{k\pi}(H)^* \otimes V_{k\lambda}(G) \tag{3.7}$$

is normal and rational. (Formally, this means the singularities of the variety, or rather the scheme, which can be associated with this ring, are rational and normal.)

Let $S^H$ denote the ring of $H$-invariants in $S$:

$$S^H = \{s \in S \mid hs = s, \quad \forall h \in H\}.$$

By (3.7),

$$S^H = \oplus_k [V_{k\pi}(H)^* \otimes V_{k\lambda}(G)]^H, \tag{3.8}$$

where the superscript $H$ on the right hand side again denotes the operation of taking $H$-invariants.

By a classical result of Hilbert [39], $S^H$ is a finitely generated ring (since $S$ is finitely generated). Furthermore, since $S$ is normal and rational, it follows by Boutot [6] that $S^H$ is normal and rational (this is the crux of the argument).

Let $h_{S^H}(k) = \dim(S^H_k)$ denote the Hilbert function of $S^H$, where $S^H_k$ denotes the degree-$k$ component of $S^H$.

By Schur's lemma [14],

$$\dim([V_{k\pi}(H)^* \otimes V_{k\lambda}(G)]^H) = a^{k\lambda}_{k\pi,\mu} = \tilde{a}^{\lambda}_{\pi,\mu}(k),$$

the multiplicity of $V_{k\pi}(H)$ in $V_{k\lambda}(G)$. Hence,

$$h_{S^H}(k) = \tilde{a}^{\lambda}_{\pi,\mu}(k).$$

By Flenner [13], $h_{S^H}(k)$ is a quasi-polynomial, since $S^H$ is rational and normal. Thus it follows that $\tilde{a}^{\lambda}_{\pi,\mu}(k)$ is also a quasi-polynomial. Q.E.D.

**Hypothesis 3.6.2 (Plethysm PH)** *(GCT6)*

*There exists a polytope of $P = P^{\lambda}_{\pi,\mu}$ with description of poly($\langle\lambda\rangle, \langle\pi\rangle, \langle\mu\rangle$) bitlength such that*

$$\tilde{a}^{\lambda}_{\pi,\mu}(k) = f_P(k), \tag{3.9}$$

*the Ehrhart quasi-polynomial of $P$. In particular,*

$$a^{\lambda}_{\pi,\mu}(k) = \#(P), \tag{3.10}$$

*the number of integer points in $P$.*

*This would provide the sought good positive formula for the plethysm constant $a^{\lambda}_{\pi,\mu}$ (cf. Problem 3.4.2)*

Here it is assumed that the polytope is presented by a separation oracle as in [15], and the bitlength $\langle P\rangle$ of the description of $P$ is defined to be $l+s$, where $l$ is the dimension of the ambient space in which $P$ is defined by linear constraints, and $s$ the maximum bitlength of any defining constraint. Notice that the polytope $P$ here depends only on $\lambda, \pi$ and $\mu$ but not on $H$, just like the plethysm constant $a^{\lambda}_{\pi,\mu}$ itself.

**Theorem 3.6.3** *(GCT6) Assuming Plethysm PH, whether $a^{k\lambda}_{k\pi,\mu} > 0$ for some $k \geq 1$ can be decided in poly($\langle\lambda\rangle, \langle\pi\rangle, \langle\mu\rangle$) time. If so, one such $k$ can also be found in polynomial time.*

*Proof:* By linear programming [15]. One has to just decide if $P^{\lambda}_{\pi,\mu}$ is nonempty, and if so, find a vertex $v$ of $P$ and choose $k$ such that $kv$ has integral coordinates. Q.E.D.

## 3.7 Positivity and the existence of obstructions in the general case

Now we describe how positivity can help in proving the existence of obstructions in the general case of the $\#P$ vs. $NC$ problem.

Towards that end, first we introduce a stronger notion of obstructions. We follow the same notation as in Section 3.1. Thus $G = SL(Y) = SL_{m^2}(\mathbb{C})$ as there.

**Definition 3.7.1** *(GCT2) A Weyl module $V_\lambda(G)$ is called a* strong obstruction *for the pair $(f, g)$, if $V_\lambda(G)$ occurs in $R_V[f, n, m]^*$, i.e. in $R_V[f, n, m]_d^*$ for some $d$, but does not contain a nonzero $G_{\hat{g}}$-invariant.*

It follows from Theorem 3.1.1 (1) that a strong obstruction is also an obstruction as per Definition 2.4.3. Furthermore, by Theorem 3.1.5, we have:

**Proposition 3.7.2** *A Weyl module $V_\lambda(G)$ is a* strong obstruction *for the pair $(f, g)$, iff*

1. *The length of $\lambda$ is at most $(n+1)^2$,*
2. *$V_\lambda(\bar{H})$ occurs in $R_{\bar{W}}[\bar{h}, n, m]^*$, i.e., in $R_{\bar{W}}[\bar{h}, n, m]_d^*$ for some $d$ (which has to be $|\lambda|/m$).*
3. *$V_\lambda(G)$ does not contain a nonzero $G_{\hat{g}}$-invariant.*

Now let $G_\lambda(k) = G_{\lambda,m}(k)$ denote the multiplicity of the trivial representation of $G_{\hat{g}}$ in $V_{k\lambda}(G)$.

**Theorem 3.7.3** *(GCT6) The stretching function $G_{\lambda,m}(k)$ is a quasi-polynomial.*

This is proved like Theorem 3.6.1; in fact, this is essentially its special case.

The following is a precise form Hypothesis 1.5.2 (b). It is essentially a special case of Plethysm PH:

**Hypothesis 3.7.4 (PH)** *(GCT6)*

*There exists a polytope of $Q_\lambda$ such that*

$$G_{\lambda,m}(k) = f_{Q_\lambda}(k), \tag{3.11}$$

*for every $m$.*

Here the polytope $Q_\lambda$ does not depend on $G$ or its dimension $m = \dim(G)$, for the same reasons that the polytope $P$ in the Plethysm PH does not depend on $H$ there; cf. the remark after the Plethysm PH. Furthermore, if $Q_\lambda$ exists, its dimension is guaranteed to be polynomial in the length of $\lambda$ by the proof of Theorem 3.7.3.

The following is a precise form of Hypothesis 1.6.1 (b).

**Hypothesis 3.7.5 (PH1)** *(GCT6)*

*There exists an explicit polytope $Q_\lambda$ satisfying PH in Hypothesis 3.7.4. Here explicit means:*

1. *The polytope is specified by an explicit system of linear constraints, each constraint of bitlength poly($\langle\lambda\rangle$) (note no dependence on $m$).*
2. *The membership problem for the polytope $Q_\lambda$ also belongs to the complexity class $P$. That is, given a point $x$, whether it belongs to $Q_\lambda$ can also be decided in poly($\langle x\rangle, \langle\lambda\rangle$) time. Furthermore, we assume that if $x$ does not belong to the polytope, then the membership algorithm also gives a hyperplane separating $x$ from the polytope in the spirit of [15].*

The following addresses a relaxed form of Problem 3.3.1, which is enough for our purposes:

**Theorem 3.7.6** *(1) Assuming PH1 above (Hypothesis 3.7.5), whether $V_{k\lambda}(G)$ contains a $G_{\hat{g}}$-invariant, for some $k \geq 1$, can be decided in poly($\langle\lambda\rangle$) time. By Theorem 3.1.1, this is equivalent to deciding whether $V_{k\lambda}(G)$ occurs in $R_V[g,m]^*$ for some $k \geq 1$.*

*(2) Assuming an analogous PH1 for the subgroup restriction problem for $H_{\hat{h}} \subseteq H$, whether $V_{k\pi}(H)$ contains an $H_{\hat{h}}$-invariant, for some $k \geq 1$, can also be decided in poly($\langle\pi\rangle, n$) time. By Theorem 3.1.1, this is equivalent to deciding whether $V_{k\pi}(H)$ occurs in $R_W[h,n]^*$ for some $k \geq 1$.*

*Proof:* Similar to that of Theorem 3.6.3. Q.E.D.

A similar result for $R_V[f,n,m]$ or $R_{\bar{W}}[\bar{h},n,m]$ would not follow from the Plethysm PH (or more generally, the subgroup restriction PH) because of the serious leak in Theorem3.1.5 that we discussed in Section 3.3. One needs a more general PH for this. We turn to this issue next.

For any $\lambda$ of length $\leq (n+1)^2$, let $F_{\lambda,n,m}(k)$ be the multiplicity of $V_{k\lambda}(\bar{H})$ in $R_{\bar{W}}[\bar{h}, n, m]^*$, which by Theorem 3.1.5, coincides with the multiplicity of $V_{k\lambda}(G)$ in $R_V[f, n, m]^*$. Thus $F_{\lambda,n,m}(k)$ is the same as the function with the same notation in Theorem 1.5.1 (b). The following is its restatement.

**Theorem 3.7.7** *The function $F_{\lambda,n,m}(k)$ is an asymptotic quasi-polynomial.*

The singularities of the class variety $\Delta_{\bar{W}}[\bar{h}, n, m]$ here are not normal [37] when $m > n$. But in view of the exceptional nature of the class variety and Theorems 3.6.1 and 3.7.3, it may be conjectured that the deviation from rationality and normality is small; cf. the remarks after Hypothesis 3.7.9 below.

The following is a restatement of Hypothesis 1.5.2 (a).

**Hypothesis 3.7.8 (PH) [Positivity Hypothesis]** *For every $\lambda, n, m \geq n$, there exists a parametrized (cf. (1.5)) polytope $P = P_{\lambda,n,m} = P_{\lambda,n,m}(k)$ such that*

$$F_{\lambda,n,m}(k) = f_P(k) \tag{3.12}$$

*It is also assumed that $P_{\lambda,n,m}$ has a specification as in Hypothesis 1.5.2 (a).*

If such $P$ exists, its dimension is guaranteed to be polynomial in $n$ (by the proof of Theorem 3.7.7) essentially because the dimension of $\bar{H}$ is $O(n^2)$ and does not depend on $m$.

The following is a restatement of Hypothesis 1.6.1 (a).

**Hypothesis 3.7.9 (PH1)**

*There exists an* explicit *parametrized polytope $P_{\lambda,n,m} = P_{\lambda,n,m}(k)$ as in Hypothesis 3.7.8.*

The meaning of explicit here is as in Hypothesis 1.6.1. In particular, the polynomial bounds are meant to be polynomial in $\langle \lambda \rangle, n$ and $\langle m \rangle$, instead of $m$. Because $F_{\lambda,n,m}(k)$ is the multiplicity of $V_{k\lambda}(\bar{H})$ in $R_{\bar{W}}[\bar{h}, n, m]^*$, $\dim(\bar{H}) = (n+1)^2$, which does not depend on $m$, and $m$ occurs only in the definition of $\bar{h}(\bar{X}) = z^{m-n} h(X)$ as a numeric parameter akin to the numeric parameters $\lambda_i$'s.

Furthermore, PH1 above implies that the deviation from quasipolynomiality of $F_{\lambda,n,m}(k)$ is small, specifically, $O(2^{O(\text{poly}(\langle\lambda\rangle,n,\langle m\rangle))})$, so that the bitlength of the deviation is polynomial. This would mean that the deviation from rationality and normality of the singularities of the class variety $\Delta_{\bar{W}}[\bar{h},n,m]$ is also small; cf. Theorem 3.7.7 and the remarks after it.

**Theorem 3.7.10** *Assuming general PH1 (Hypothesis 3.7.9), whether $V_{k\lambda}(G)$ occurs in $R_V[f,n,m]^*$–or equivalently, whether $V_{k\lambda}(\bar{H})$ occurs in $R_{\bar{W}}[\bar{h},n,m]^*$–for some $k \geq 1$, can be decided in poly($\langle\lambda\rangle, n, \langle m\rangle$) time.*

The following is a refined version of Theorem 1.5.3:

**Theorem 3.7.11** *(GCT6) There exists a family $\{O_n = V_{\lambda_n}(G)\}$ of strong obstructions for the general #P vs. NC problem in characteristic zero, for $m = 2^{\log^a n}$, $a > 1$ fixed, $n \to \infty$, assuming,*

1. *PH above (Hypotheses 3.7.4 and 3.7.8) , and*
2. *OH (Obstruction Hypothesis):*

   *For all $n \to \infty$, there exists $\lambda$ such that $P_{\lambda,n,m}(k) \neq \emptyset$ for all sufficiently large $k$ and $Q_\lambda = \emptyset$.*

*For an analogous result for the P vs. NP problem, see GCT6.*

We define $\lambda$ to be a polyhedral obstruction (label) if it satisfies OH here. In this case it is easy to see that $k\lambda$, for some $k \geq 1$, is a strong obstruction.

## 3.8 Positivity and the $P$-barrier

Theorem 3.7.6 and 3.7.10 imply:

**Theorem 3.8.1** *Given $\lambda, n, m$, whether $\lambda$ is a polyhedral obstruction (label) can be decided in poly($\langle\lambda\rangle, n, \langle m\rangle$) time assuming PH1 (Hypotheses 3.7.5 and 3.7.9). In other words, the $P$-barrier for verification of obstructions (Section 2.7) can be crossed assuming PH1.*

In conjunction with Theorem 3.7.11, this implies its stronger form:

**Theorem 3.8.2** *(GCT6) There exists an explicit (cf. Section 1.6) family* $\{\lambda_n\}$ *of polyhedral obstructions, for the general* $\#P$ *vs.* $NC$ *problem in characteristic zero, for* $m = 2^{\log^a n}$, $a > 1$ *fixed,* $n \to \infty$, *assuming PH1 and OH above.*

*Analogous result holds for the* $P$ *vs.* $NP$ *problem in characteristic zero.*

The strategy now is to prove PH, or rather PH1, first, and then prove OH using the explicit forms of the polytopes in PH1. GCT2,6,7,8 together give an evidence for why PH/PH1 and OH should hold. As far as proving OH is concerned, there is nothing that we can say at this point since it depends on the explicit forms of the polytopes in PH/PH1. The remaining question is the following.

## 3.9 How to prove PH1 and why should it hold?

We now briefly describe the plan in GCT6 to prove PH1 by generalizing the proof of LR PH1 (Theorem 3.5.1) based on the theory of standard quantum groups [12, 18, 24].

For that we need a definition.

**Definition 3.9.1** *Let* $H$ *be a connected reductive subgroup of a connected reductive* $G$*. A basis* $B$ *of a representation* $V$ *of* $G$ *is called* positive *with respect to the* $H$*-action if:*

1. *If it* $H$*-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* $\langle B_i \rangle$ *denote the linear span of* $B_i$*, is isomorphic to an irreducible* $H$*-module. In other words, this filtration gives a Jordan-Holder series of* $V$*.*

2. *For each standard generator* $h$ *of (the Lie algebra of)* $H$ *and each* $b \in B$*,*

$$hb = \sum_{b' \in B} c^h_{b,b'} b,$$

   *where each* $c^h_{b,b'}$ *is a nonnegative rational.*

LR PH1 is a consequence of the proof of a much deeper positivity result:

**Theorem 3.9.2 (LR PH0)** *[24, 5] Let $H = GL_n(\mathbb{C})$ embedded in $G = H \times H$ diagonally as in (3.5). Then each irreducible representation of $G$ has a positive basis with respect to the $H$ action.*

The proof of this result goes via the theory of the standard quantum group. Specifically, the diagonal embedding (3.5) is first quantized [12] in the form

$$H_q \to H_q \times H_q, \tag{3.13}$$

where $H_q$ is the standard quantum group, a quantization of $H$ that plays the same role in quantum mechanics that the standard group $H$ plays in classical mechanics. (Well, (3.13) is not really accurate, because what is quantized in [12] is not $H$ but rather its universal enveloping algebra. We shall ignore this technicality here.) It is then shown that the irreducible representations of $H_q$ and $H_q \times H_q$ have extremely rigid *canonical* bases [18, 24], which are positive [24], and have many other remarkable properties. The only known proof of this positivity [24] is based on the Riemann hypothesis over finite fields and the related works [10, 5]

**Goal 3.9.3** *Lift this LR story to the plethysm problem (and the more general subgroup restriction problem).*

In this context:

**Hypothesis 3.9.4 (Plethysm PH0)** *Let*

$$H = GL_n(\mathbb{C}) \to GL(V), V = V_\mu(H), \tag{3.14}$$

*be the plethysm homomorphism (3.3). The each Weyl module $V_\lambda(G)$ has a positive basis with respect to the $H$-action.*

**Theorem 3.9.5** *(GCT7) The plethysm map (3.14) can be quantized in the form*

$$H_q \to G_q^H, \tag{3.15}$$

*where $H_q$ is the standard quantum group associated with $H$ and $G_q^H$ is a new nonstandard quantization of $G$.*

*A similar result holds for general connected reductive $H$ as well.*

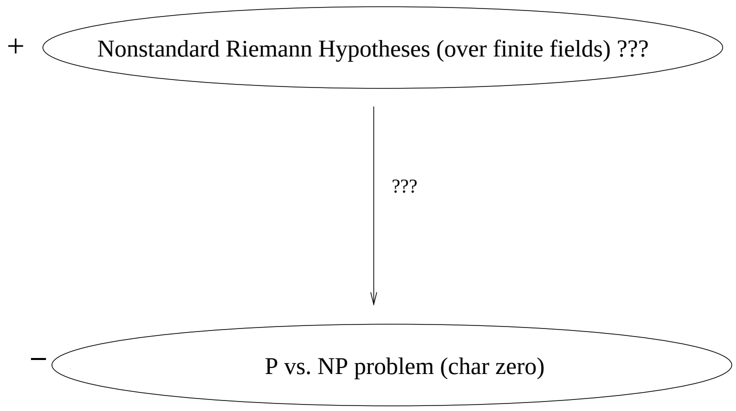


Figure 3.1: The ultimate goal of GCT

Furthermore, GCT8 gives a conjecturally correct algorithm to construct canonical bases of irreducible representations $G_q^H$ with conjectural positivity and other properties from which Plethysm PH0 would follow. Experimental evidence for positivity of the conjectural canonical bases in GCT8 constitutes the main evidence for Plethysm PH0, and hence Plethysm PH1/PH.

The general PH1 (Hypothesis 3.7.9) can be regarded as a generalization of the Plethysm PH1 for the triple (cf. Observation 2.5.2)

$$\bar{H}_{\hat{h}} \to \bar{H} = SL(\bar{X}) \to L = GL(\bar{W}) \tag{3.16}$$

associated with the class variety $\Delta_{\bar{W}}[\bar{h}]$, rather than the plethysm couple (3.14). To go from the Plethysm PH1 to the general PH1, one has to similarly quantize the triple (3.16) and develop an analogous theory of canonical bases for this quantized triple. But first, we have to understand the couples. Hence the Plethysm PH0/PH1 can be regarded as the heart of GCT. To prove the nonstandard quantum group conjectures in GCT7,8 that arise in this context, a substantial nonstandard extension of the work [10, 5, 19, 24] surrounding the standard Riemann hypothesis over finite fields may be necessary; cf. Figure 1.7. Thus the ultimate goal of GCT would be to deduce the ultimate negative hypothesis of mathematics, $P \neq NP$ conjecture (in characteristic zero), from the ultimate positive hypotheses–namely, as yet unknown, nonstandard Riemann hypotheses (over finite fields); cf. Figure 3.1.

# Bibliography

[1] S. Aaronson, Is P versus NP formally independent? Bulletin of the EATCS 81: 109-136 (2003).

[2] B. Adsul, private communication.

[3] B. Alexeev, J. Tsimerman, A direct proof of Mulmuley's weak $\#P$ versus $NC$ result, manuscript.

[4] S. Ariki, J. Matsuzawa, I. Terada, Representations of Weyl groups on zero weight spaces of }-modules, Algebraic and topological theories, pp. 546-568 (1985).

[5] A. Beilinson, J. Bernstein, P. Deligne, Faisceaux pervers, Astérisque 100, (1982), Soc. Math. France.

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

[7] M. Brion, On the general faces of the moment polytope, IMRN International Mathematics Research Notices, No. 4, 1999.

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

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

[10] P. Deligne, La conjecture de Weil II, Publ. Math. Inst. Haut. Étud. Sci. 52, (1980) 137-252.

[11] P. Deligne, Categories tannakiennes, in The Grothendieck Festschrift, Volume 2, 11-195, Birkhauser, 19990.

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

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

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

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

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

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

[18] M. Kashiwara, On crystal bases of the $q$-analogue of universal enveloping algebras, Duke Math. J. 63 (1991), 465-516.

[19] D. Kazhdan, G. Lusztig, Schubert varieties and Poincare duality, Proc. Symp. Pure Math., AMS, 36 (1980), 185-203.

[20] G. Kempf: Instability in invariant theory, Annals of Mathematics, 108 (1978), 299-316.

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

[22] 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.

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

[24] G. Lusztig, Canonical bases arising from quantized enveloping algebras, J. Amer. Math. Soc. 3, (1990), 447-498.

[25] 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.

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

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

[28] K. Mulmuley, On P vs. NP, geometric complexity theory, and the flip I, Technical report TR-2007-16, computer science department, The university of Chicago, September 2007; revised version under preparation. This and the following GCT papers are available at http://ramakrishnadas.cs.uchicago.edu.

[29] K. Mulmuley, On P vs. NP, Geometric Complexity Theory, Explicit proofs, and the Complexity Barrier, under preparation, to be available at the above website soon.

[30] 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.

[31] 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.

[32] K. Mulmuley, An appendix to Geometric Complexity Theory II, technical report, computer science department, the university of chicago, February, 2009.

[33] K. Mulmuley, Geometric complexity theory VI: The flip, Revised version under preparation, Earlier version: Technical Report TR-2007-04, computer science department, The University of Chicago, September, 2007.

[34] 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.

[35] 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.

[36] K. Mulmuley, Geometric complexity theory XI: on the flip over fields of positive characteristic, under preparation.

[37] Shrawan Kumar, private communication.

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

[39] V. Popov, E. Vinberg, Invariant theory, in Encyclopaedia of Mathematical Sciences, Algebraic Geometry IV, Eds. A. Parshin, I. Shafarevich, Springer-Verlag, 1989.

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

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

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

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

# Appendix: A bit more of representation theory

Here we go into the basic representation theory a bit more than in Section 2.1; in particular, we describe an explicit construction of Weyl modules.

Let $G$ be a group. We say that a vector space $V$ is a *representation* of $G$, or a $G$-*module*, if there is a homomorphism

$$\rho : G \to GL(V), \tag{3.17}$$

where $GL(V)$ is the general linear group of invertible transformations of $V$. We denote $\rho(g)(v)$ by $g \cdot v$–the result of the action of $g$ on $v$. A $G$-subrepresentation $W \subseteq V$ is a subspace that is invariant under $G$; i.e., $g \cdot w \in W$ for every $w \in W$. If $G$ is clear from the context, we just call it subrepresentation. We say that $V$ is *irreducible* if it does not contain a proper nontrivial subrepresentation. A $G$-*homomorphism* from a $G$-module $U$ to a $G$-module $V$ is map $\psi : U \to V$ such that $\psi(g \cdot u) = g \cdot (\psi(u))$ for all $u \in U$.

We say that $G$ is *reductive* if every finite dimensional representation $V$ of $G$ is *completely reducible.* This means it can be expressed as a direct sum of irreducible representations in the form

$$V = \bigoplus_{\lambda} m_\lambda V_\lambda(G) \tag{3.18}$$

where $\lambda$ ranges over all indices (labels) of irreducible representations of $G$, $V_\lambda(G)$ denotes the irreducible representation of $G$ with label $\lambda$, and $m_\lambda V_\lambda(G)$ denotes a direct sum of $m_\lambda$ copies of $V_\lambda(G)$. Here $m_\lambda$ is called the *multiplicity* of $V_\lambda(G)$ in $V$. It is a basic fact of representation theory that for reductive groups, the decomposition (3.18) is unique; i.e., $m_\lambda$'s are uniquely defined. If $m_\lambda > 0$, we say that $V_\lambda(G)$ *occurs* in $V$.

An example of a nonreductive group is a solvable group that is not abelian. In this case a subrepresentation $W \subseteq V$ need not have a complement $W^\perp$ such that $V = W \oplus W^\perp$.

Every finite group is reductive. Thus $S_n$, the symmetric group on $n$ letters, is reductive. A prime example of a continuous reductive group is the general linear group $GL_n(\mathbb{C}) = GL(\mathbb{C}^n)$, the group of nonsingular $n \times n$ matrices, and its subgroup the special linear group $SL_n(\mathbb{C}) = SL(\mathbb{C}^n)$ of matrices with determinant one. Any product of reductive groups is also reductive. These are the only kinds of reductive groups that we need to know in this article. So whenever we say reductive, the reader may wish to assume that the group is a general or special linear group or a symmetric group or a product thereof.

We say that the representation (3.17) of $GL_n(\mathbb{C})$ or $SL_n(\mathbb{C})$ is polynomial if for every $g \in G$, every entry in the matrix form of $\rho(g)$ is a polynomial in the entries of $g$.

Complete reducibility as in eq.(3.18) means every finite dimensional representation of a reductive group is composed of irreducible representations. These can be thought of as the building blocks in the representation theory of reductive groups, and it is important to know what these building blocks are.

For $G = GL_n(\mathbb{C})$ and $SL_n(\mathbb{C})$ this was done by Weyl [14]. The polynomial irreducible representations of $GL_n(\mathbb{C})$ are in one-to-one correspondence with the tuples $\lambda = (\lambda_1, \ldots, \lambda_k)$ of integers, where $k \leq n$ and $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_k > 0$. Here $\lambda$ is called a *partition* of length $k$ and size $|\lambda| = \sum_i \lambda_i$. Its bitlength $\langle \lambda \rangle$ is defined to be the total bitlength of all $\lambda_i$'s.

Thus the polynomial irreducible representations of $GL_n(\mathbb{C})$ are labelled by partitions $\lambda$ of length at most $n$, but any size. The irreducible representation corresponding to a partition $\lambda = (\lambda_1, \lambda_2, \ldots)$ is denoted by $V_\lambda(GL_n(\mathbb{C}))$, and is called a *Weyl module* of $GL_n(\mathbb{C})$. When $GL_n(\mathbb{C})$ is clear from the context, we shall denote it by simply $V_\lambda$.

Each partition $\lambda$ corresponds to a Young diagram, which consists of $k$ rows of boxes, with $\lambda_i$ boxes in the $i$-th row. For example, the Young diagram corresponding to $(4, 2, 1)$ is shown below:

$$\ydiagram{4,2,1}$$

When thinking of a partition, it is helpful to think of the corresponding Young diagram. Thus each Weyl module is labelled by a Young diagram of height at most $n$. This is a useful combinatorial tool for studying the Weyl modules.

A Weyl module $V_\lambda$ is explicitly constructed as follows. This construction of Deyruts as well as Weyl's original construction are given in [14]. Let $Z$ be an $n \times n$ variable matrix. Let $\mathbb{C}[Z]$ be the ring of polynomials in the entries of $Z$. It is a representation of $GL_n(\mathbb{C})$. Action of a matrix $\sigma \in GL_n(\mathbb{C})$ on a polynomial $f \in \mathbb{C}[Z]$ is given by

$$(\sigma \cdot f)(Z) = f(Z\sigma). \tag{3.19}$$

By a numbering (filling), we mean filling of the boxes of a Young diagram by numbers in $[n]$; for example:

$$\begin{ytableau} 1 & 2 & 4 & 3 \\ 2 & 3 \\ 1 \end{ytableau}$$

We call such a numbering a *(semistandard) tableau* if the numbers are strictly increasing in each column and weakly increasing in all rows; e.g.

$$\begin{ytableau} 1 & 2 & 3 & 3 \\ 2 & 3 \\ 4 \end{ytableau}$$

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

With every numbering $T$, we associate a polynomial $e_T \in \mathbb{C}[Z]$, which is a product of minors for each column of $T$. The $l \times l$ minor $e_c$ for a column $c$

of length $l$ is formed by the first $l$ rows of $Z$ and the columns indexed by the entries $c_j$, $1 \leq j \leq l$, of $c$. Thus $e_T = \prod_c e_c$, where $c$ ranges over all columns in $T$. The Weyl module $V_\lambda$ is the subrepresentation of $\mathbb{C}[Z]$ spanned by $e_T$, where $T$ ranges over all numberings of shape $\lambda$ over $[n]$. Its one possible basis is given by $\{e_T\}$, where $T$ ranges over semistandard tableau of shape $\lambda$ over $[n]$.

Let $B \subseteq GL_n(\mathbb{C})$ be the subgroup of upper triangular matrices. It is called the *Borel subgroup* of $GL_n(\mathbb{C})$. An element $v_\lambda \in V_\lambda$ is called a *highest weight vector* if it is an eigenvector for the action of each $b \in B$. It is easy to show that $V_\lambda$ has a unique highest weight vector, upto a constant multiple: it is $e_{T_0}$, where $T_0$ is the canonical tableau whose $i$-th row contains only $i$'s, for each $i$; e.g.

$$\begin{array}{|c|c|c|c|}\hline 1 & 1 & 1 & 1 \\ \hline 2 & 2 \\ \cline{1-2} 3 \\ \cline{1-1}\end{array}$$

Let $P \subseteq GL_n(\mathbb{C})$ be the subgroup of upper block triangular matrices, where the sizes of the blocks are fixed. For example:

$$\begin{bmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ 0 & 0 & * & * & * & * \\ 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * & * \end{bmatrix}$$

Such subgroups are called *parabolic*. Let $P_\lambda$ be the (projective) stabilizer of the highest weight vector $v_\lambda = e_{T_0}$; i.e., the set of all $\sigma \in GL_n(\mathbb{C})$ such that $\sigma \cdot v_\lambda = c(\sigma) v_\sigma$, for some complex number $c(\sigma)$. Then it is easy to show that $P_\lambda$ is parabolic, where the sizes of the blocks are completely determined by $\lambda$.

The irreducible representation of $GL_n(\mathbb{C})$ corresponding to the Young diagram that consists of just one column of length $n$ is the determinant representation: $g \to \det(g)$. When restricted to the subgroup $SL_n(\mathbb{C}) \subseteq GL_n(\mathbb{C})$ this becomes trivial. More generally, $V_\lambda(GL_n)$ and $V_{\lambda'}(GL_n)$ give the same representation of $SL_n(\mathbb{C})$ if $\lambda'$ is obtained from $\lambda$ by removing columns of length $n$. Hence, irreducible polynomial representations of $SL_n(\mathbb{C})$ are in one to one correspondence with partitions of length less than $n$, and are obtained from the ones of $GL_n(\mathbb{C})$ by restriction.

¡