Research Statement

Yifeng Huang

My research covers a range of problems in algebraic geometry, topology, combinatorics and number theory. A recurring theme in my research described in Section 1 and 2 is to investigate combinatorial patterns that appear and reappear in multiple contexts, and to explain their connection from a geometric point of view. The introduction, the motivation and the content of my work are detailed in each of the sections below.

- Varieties of modules, and matrix enumeration problems over finite fields (Section 1, [24, 23]);
- Cohomology of configuration spaces (Section 2, [8, 22]);
- Unit equations in the quaternion algebra (Section 3, [25]);
- Random matrices over a p-adic ring (Section 4, [9]).

1 Counting points on varieties of modules over finite fields

A central problem in algebra and representation theory is to classify finite-dimensional modules over an algebra over a field up to isomorphism. As this problem is out of reach in general, we often consider a geometric alternative, namely, to consider the geometric object that parametrizes the modules. This geometric object is the *variety of modules.*

A prototypical example of the variety of modules is the variety of commuting matrices, also called the *commuting variety.* Let $R = \mathbb{C}[x, y]$ be the polynomial ring in two variables, then n-dimensional modules over R can be parametrized by the variety

$$C_n = \{(A, B) \in \mathrm{Mat}_n(\mathbb{C}) \times \mathrm{Mat}_n(\mathbb{C}) : AB = BA\}, \tag{1.1}$$

so we say that C_n is the variety of n-dimensional modules over R.

It was proved by Motzkin and Taussky [29] and Gerstenhaber [19] that C_n is irreducible. Feit and Fine [14] found the point count of C_n over a finite field (namely, the number of commuting pairs of matrices over a finite field), and Bryan and Morrison [5] reproved and refined the above result from the point of view of motivic classes. The motivic class of a variety in the Grothendieck ring is the combinatorics of how the variety is built up from other varieties via “cut-and-paste”, and it refines the knowledge of point counting over finite fields.

Extensions and variants of the variety of modules and the commuting variety have been introduced and studied; see [2, 6, 7, 11, 33]. Most of the study is focused on geometric problems, such as irreducibility, description of irreducible components, and description of a certain GIT quotient. However, little had been known about the point counts or other motivic problems about these varieties, except for a few cases [5, 18]. The focus of my research is to fill in this gap.

1.1 Variety of modules over a nodal singular curve

1.1.1 Background

Let R be an algebra over a finite field $\mathbb{F}_q$ such that all quotient fields of R are finite. We define the *Cohen–Lenstra series* of R as

$$\widehat{Z}_R(x) := \sum_M \frac{1}{|\mathrm{Aut}\, M|} x^{\dim_{\mathbb{F}_q} M}, \tag{1.2}$$

where M ranges over all isomorphism classes of finite-cardinality R-modules. The coefficients of $\widehat{Z}_R(x)$ have two interpretations: they are closely related to the $\mathbb{F}_q$-point count of the variety of modules over R, and they represent the statistics of classification of modules over R. For instance,

$$\widehat{Z}_{\mathbb{F}_q[[t]]}(x) = \frac{1}{(1-xq^{-1})(1-xq^{-2})\dots} = \sum_{n=1}^{\infty} \frac{q^{n^2-n}}{|\mathrm{GL}_n(\mathbb{F}_q)|} x^n. \tag{1.3}$$

This means, on one hand, that the number of n by n nilpotent matrices over $\mathbb{F}_q$ is q^{n^2-n}, a result of Fine and Herstein [15]. We remark that the variety of modules over $\mathbb{F}_q[[t]]$ is precisely the variety of nilpotent matrices over $\mathbb{F}_q$. This formula also means that the number of isomorphism classes of finite-cardinality $\mathbb{F}_q[[t]]$-modules, weighted inversely by the size of the automorphism group, is given by $1/((1-q^{-1})(1-q^{-2})\dots)$, a number obtained from evaluating $\widehat{Z}_{\mathbb{F}_q[[t]]}(x)$ at $x=1$; see the work of Cohen and Lenstra [10]. Moreover, the distribution of the dimension (over $\mathbb{F}_q$) of a "random" $\mathbb{F}_q[[t]]$-module is encoded in the coefficients of $\widehat{Z}_{\mathbb{F}_q[[t]]}(x)$.

We notice that $\widehat{Z}_{\mathbb{F}_q[[t]]}(x)$ can be expressed as an infinite product of simple factors. The same can be said about $\widehat{Z}_{\mathbb{F}_q[[u,v]]}$ [18]:

$$\widehat{Z}_{\mathbb{F}_q[[u,v]]}(x) = \prod_{i,j\geq 1} (1-q^{-j}x^i)^{-1}. \tag{1.4}$$

We can view these series from a geometric point of view. Let X be a variety over $\mathbb{F}_q$ and p be a closed point, and we define the *local Cohen–Lenstra series* of the pair (X,p) to be $\widehat{Z}_{X,p}(x) := \widehat{Z}_{\widehat{\mathcal{O}}_{X,p}}(x)$, where $\widehat{\mathcal{O}}_{X,p}$ is the completed local ring of X at p. The series $\widehat{Z}_{X,p}(x)$ has a natural geometric meaning, as it represents the statistics of classification of finite-length coherent sheaves on X supported at p. From this point of view, $\widehat{Z}_{\mathbb{F}_q[[t]]}(x)$ is a series attached to a smooth $\mathbb{F}_q$-point on a curve, while $\widehat{Z}_{\mathbb{F}_q[[u,v]]}(x)$ is attached to a smooth $\mathbb{F}_q$-point on a surface. If p is a singular point of X, we wonder how the mysterious series $\widehat{Z}_{X,p}(x)$ would be different from the smooth case, and how the difference possibly encodes the geometry of the singularity (X,p).

1.1.2 Contribution

My work [24] generalizes the computation of the local Cohen–Lenstra series to a singularity case. Throughout this section, let $R=\mathbb{F}_q[[u,v]]/(uv)$. An $\mathbb{F}_q$-point p on an $\mathbb{F}_q$-variety X is said to be a *nodal singularity* if the completed local ring $\widehat{\mathcal{O}}_{X,p}$ is isomorphic to R. The modeling example of a node is when X is the union of two axes on a plane and p is the origin.

Theorem 1.1 (H. [24])**.**

(a) *The series $\widehat{Z}_R(x)$ has a meromorphic continuation to all of $\mathbb{C}$. Moreover,*

$$\widehat{Z}_R(x) = \left(\frac{1}{(1-xq^{-1})(1-xq^{-2})\dots}\right)^2 H_q(x), \tag{1.5}$$

where $H_q(x)$ is an explicit series that is entire (convergent for all x).

(b) $\widehat{Z}_R(1) = \left(\frac{1}{(1-q^{-1})(1-q^{-2})\dots}\right)^2.$

Comparing with the formula for $\widehat{Z}_{\mathbb{F}_q[[t]]}(x)$, we notice that (1.5)(a) says $\widehat{Z}_R(x) = \widehat{Z}_{\mathbb{F}_q[[t]]}(x)^2 H_q(x)$ for some entire series $H_q(x)$. The result (together with some nonvanishing properties of $H_q(x)$ based on its explicit expression) implies that the poles of $\widehat{Z}_R(x)$ are double poles located at the poles of $\widehat{Z}_{\mathbb{F}_q[[t]]}(x)$. So even though the local Cohen–Lenstra series of a node is not a nice infinite product like the case of a smooth point, it is "not too far" from the smooth case. In fact, Theorem 1.1(a) implies that if X is any reduced curve with only nodal singularities, and $\widetilde{X}$ is a resolution of singularity of X, then $\widehat{Z}_X(x)/\widehat{Z}_{\widetilde{X}}(x)$ is entire. This means that the (global) Cohen–Lenstra series of a nodal curve X is not too far from the "smooth version" of X.

Theorem 1.1(b) computes the weighted number of finite-cardinality R-modules up to isomorphism. Note that this is the square of the similar count for $\mathbb{F}_q[[t]]$-modules. The resulting number is surprisingly clean, given that the classification of R-modules (see [4] and its references) is much more complicated than the classically well-known classification of $\mathbb{F}_q[[t]]$-modules.

The proof of Theorem 1.1 requires counting pairs of mutually annihilating matrices and finding an explicit form of $H_q(x)$ in the factorization. I conjecture that the existence of a meromorphic continuation should hold for the local Cohen–Lenstra series of other curve singularities, and moreover, the series should have a meaningful factorization involving geometric invariants of the singularity. This problem appears to be very challenging because the corresponding matrix enumeration problems are not as approachable. Nevertheless, I expect that the investigation of this problem, or even a geometric proof of Theorem 1.1, would lead to more geometric understandings about the variety of modules over the local ring at a curve singularity. Verifying the conjecture concretely for other examples of singularities is part of my ongoing research.

1.2 Variety of modules over the quantum plane

1.2.1 Background

Quantum deformation is another way to generalize the commuting variety. More precisely, given a nonzero element ζ in $\mathbb{C}$, we define the *ζ-commuting variety* to be

$$K_{\zeta,n} := \{(A,B) \in \mathrm{Mat}_n(\mathbb{C}) \times \mathrm{Mat}_n(\mathbb{C}) : AB = \zeta BA\}. \tag{1.6}$$

If $\zeta = 1$, then we get the usual commuting variety. While the commuting variety parametrizes modules over the polynomial algebra $\mathbb{C}[x,y]$, the ζ-commuting variety parametrizes modules over a "deformed" version of $\mathbb{C}[x,y]$, which is called the *quantum plane*. Specifically, the quantum plane is the noncommutative $\mathbb{C}$-algebra in variables x and y such that $xy = \zeta yx$. When $\zeta = 1$, the quantum plane reduces to $\mathbb{C}[x,y]$. When $\zeta \neq 1$, the quantum plane is not commutative and can be viewed as a deformation from the $\zeta = 1$ case.

Geometric properties of the ζ-commuting variety have been studied by Chen and Lu [6], and there has been work about the point count of other generalizations of the commuting variety, see for instance [18].

1.2.2 Contribution

My work [23] generalizes the result of Feit and Fine [14] about the point count of the commuting variety. Fix a nonzero element ζ in $\mathbb{F}_q$, and let

$$K_{\zeta,n}(\mathbb{F}_q) := \{(A,B) \in \mathrm{Mat}_n(\mathbb{F}_q) \times \mathrm{Mat}_n(\mathbb{F}_q) : AB = \zeta BA\} \tag{1.7}$$

be the set of $\mathbb{F}_q$-points of the ζ-commuting variety. The following result gives its count.

Theorem 1.2 (H. [23]). *Let m be the smallest positive integer such that $\zeta^m = 1$; in other words, ζ is a primitive m-th root of unity of $\mathbb{F}_q$. We have the following identity of power series in x:*

$$\sum_{n=0}^{\infty} \frac{|K_{\zeta,n}(\mathbb{F}_q)|}{(q^n-1)(q^n-q)\dots(q^n-q^{n-1})} x^n = \prod_{i=1}^{\infty} F_m(x^i;q), \tag{1.8}$$

where

$$F_m(x;q) := \frac{1-x^m}{(1-x)(1-x^m q)} \cdot \frac{1}{(1-x)(1-xq^{-1})(1-xq^{-2})\dots}. \tag{1.9}$$

If $\zeta = 1$, then $m = 1$, so we get

$$\sum_{n=0}^{\infty} \frac{|K_{1,n}(\mathbb{F}_q)|}{(q^n-1)(q^n-q)\dots(q^n-q^{n-1})} x^n = \prod_{i=1}^{\infty} \frac{1}{(1-x^i q)(1-x^i)(1-x^i q^{-1})\dots}, \tag{1.10}$$

which recovers the result of [14]. As the first nontrivial case of Theorem 1.2, if q is odd and $\zeta = -1$, then $m = 2$, and we get

$$\sum_{n=0}^{\infty} \frac{|K_{-1,n}(\mathbb{F}_q)|}{(q^n-1)(q^n-q)\dots(q^n-q^{n-1})} x^n = \prod_{i=1}^{\infty} \frac{1+x^i}{1-x^{2i}q} \frac{1}{(1-x^i)(1-x^i q^{-1})\dots}. \tag{1.11}$$

This gives the point count of the anti-commuting variety $AB + BA = 0$ in odd characteristic.

The count of $K_{\zeta,n}(\mathbb{F}_q)$ has a refinement that exhibits a curious phenomenon. Let

$$U_{\zeta,n}(\mathbb{F}_q) := \{(A, B) \in \mathrm{Mat}_n(\mathbb{F}_q) \times \mathrm{Mat}_n(\mathbb{F}_q) : AB = \zeta BA,\ A \text{ nonsingular}\}, \tag{1.12}$$

and

$$N_{\zeta,n}(\mathbb{F}_q) := \{(A, B) \in \mathrm{Mat}_n(\mathbb{F}_q) \times \mathrm{Mat}_n(\mathbb{F}_q) : AB = \zeta BA,\ A \text{ nilpotent}\}. \tag{1.13}$$

Theorem 1.3 (H. [23])**.** *Let m, ζ be as before. We have the following identities of power series in x:*

(a)

$$\sum_{n=0}^{\infty} \frac{|K_{\zeta,n}(\mathbb{F}_q)|}{|\mathrm{GL}_n(\mathbb{F}_q)|} x^n = \left(\sum_{n=0}^{\infty} \frac{|U_{\zeta,n}(\mathbb{F}_q)|}{|\mathrm{GL}_n(\mathbb{F}_q)|} x^n \right) \left(\sum_{n=0}^{\infty} \frac{|N_{\zeta,n}(\mathbb{F}_q)|}{|\mathrm{GL}_n(\mathbb{F}_q)|} x^n \right), \tag{1.14}$$

where we recall that $|\mathrm{GL}_n(\mathbb{F}_q)| = (q^n - 1)(q^n - q)\dots(q^n - q^{n-1})$.

(b)

$$\sum_{n=0}^{\infty} \frac{|U_{\zeta,n}(\mathbb{F}_q)|}{|\mathrm{GL}_n(\mathbb{F}_q)|} x^n = \prod_{i=1}^{\infty} G_m(x^i; q), \tag{1.15}$$

where

$$G_m(x; q) := \frac{1 - x^m}{(1 - x)(1 - x^m q)}. \tag{1.16}$$

(c)

$$\sum_{n=0}^{\infty} \frac{|N_{\zeta,n}(\mathbb{F}_q)|}{|\mathrm{GL}_n(\mathbb{F}_q)|} x^n = \prod_{i=1}^{\infty} H(x^i; q), \tag{1.17}$$

where

$$H(x; q) := \frac{1}{(1 - x)(1 - xq^{-1})(1 - xq^{-2})\dots}. \tag{1.18}$$

Theorem 1.3 refines Theorem 1.2 by computing $|U_{\zeta,n}(\mathbb{F}_q)|$ and $|N_{\zeta,n}(\mathbb{F}_q)|$, which are the building blocks of $|K_{\zeta,n}(\mathbb{F}_q)|$ in light of Theorem 1.3(a). If $\zeta = 1$, then Theorem 1.3 recovers the computation of $|U_{1,n}(\mathbb{F}_q)|$ and $|N_{1,n}(\mathbb{F}_q)|$ by Feit and Fine.

We now explain the curious phenomenon. Bryan and Morrison [5] noted that $|U_{1,n}(\mathbb{F}_q)|$ and $|N_{1,n}(\mathbb{F}_q)|$ are not separate and independent building blocks of $|K_{1,n}(\mathbb{F}_q)|$, but they "determine" each other using ingredients from algebraic geometry (over commutative rings). However, for general ζ, $|N_{\zeta,n}(\mathbb{F}_q)|$ does not seem to be able to recover $|U_{\zeta,n}(\mathbb{F}_q)|$ because $|N_{\zeta,n}(\mathbb{F}_q)|$ does not depend on m (the order of ζ as a root of unity) but $|U_{\zeta,n}(\mathbb{F}_q)|$ does, as can be observed from Theorem 1.3. One could argue that it is still possible to recover $|U_{\zeta,n}(\mathbb{F}_q)|$ from $|N_{\zeta,n}(\mathbb{F}_q)|$ together with the geometry of the quantum plane $xy = \zeta yx$, since the geometry of $xy = \zeta yx$ would depend on m. Investigating this possibility may lead us to interesting observations about the geometry of the quantum plane.

2 Cohomology of configuration spaces

The study of configuration spaces is an area where topology, number theory and combinatorics interplay frequently. For a topological space X, the *(unordered) configuration space* of X is defined as

$$\mathrm{Conf}^n(X) := \{(x_1, \dots, x_n) \in X^n : x_i \neq x_j \text{ for } i \neq j\}/S_n, \tag{2.1}$$

where the symmetric group S_n acts on X^n by permuting coordinates. When X is a quasi-projective variety over a field, then $\mathrm{Conf}^n(X)$ is also a variety. Consider the configuration space of the affine line $\mathbb{A}^1$, both over $\mathbb{C}$ and $\mathbb{F}_q$. Over $\mathbb{C}$, the set of $\mathbb{C}$-points of $\mathrm{Conf}^n(\mathbb{A}^1)$ has an analytic topology that can be identified with the topological space $\mathrm{Conf}^n(\mathbb{C})$, whose i-th Betti numbers $h^i(\mathrm{Conf}^n(\mathbb{C}))$ were computed by Arnol'd [1]:

$$\big(h^0(\mathrm{Conf}^n(\mathbb{C})), h^1(\mathrm{Conf}^n(\mathbb{C})), \dots\big) = \begin{cases} (1,0,0,\dots,0,\dots), & n = 0,1; \\ (1,1,0,\dots,0,\dots), & n \geq 2. \end{cases} \tag{2.2}$$

Over $\mathbb{F}_q$, the set of $\mathbb{F}_q$-points of $\mathrm{Conf}^n(\mathbb{A}^1)$ can be identified with the set of monic square-free polynomials in $\mathbb{F}_q[t]$ of degree n. We have

$$|\mathrm{Conf}^n(\mathbb{A}^1)(\mathbb{F}_q)| = \begin{cases} q^n, & n = 0,1; \\ q^n - q^{n-1}, & n \geq 2. \end{cases} \tag{2.3}$$

We note the apparent connection between (2.2) and (2.3):

$$|\mathrm{Conf}^n(\mathbb{A}^1)(\mathbb{F}_q)| = \sum_{i \geq 0} h^i(\mathrm{Conf}^n(\mathbb{C}))(-1)^i q^{n-i}. \tag{2.4}$$

This connection turns out to be meaningful: the results (2.2) and (2.3) imply each other due to a result of Kim [27] that the mixed Hodge structure on $H^i(\mathrm{Conf}^n(\mathbb{C}))$ is pure of weight $2i$. Using the purity result of the mixed Hodge structure, one can reach a formula about the *mixed Hodge numbers* $h^i_{p,q}(\mathrm{Conf}^n(\mathbb{C}))$ of $\mathrm{Conf}^n(\mathbb{C})$, a notion due to Deligne [12] that can be almost viewed as a common refinement[1] of the Betti numbers and the $\mathbb{F}_q$-point count:

$$\sum_{i,p,q \geq 0} h^i_{p,q}(\mathrm{Conf}^n(\mathbb{C})) x^{n-p} y^{n-q} (-u)^{2n-i} = \begin{cases} (xyu)^n, & n = 0,1; \\ (xyu)^n - (xyu)^{n-1}, & n \geq 2. \end{cases} \tag{2.5}$$

Substituting $x = y = 1$, we recover (2.2). Substituting $u = 1$ and $xy = q$, we recover the right-hand side of (2.3).

It is definitely not the case in general that the point count and the Betti numbers of the configuration space recover each other. However, there do exist other phenomena that occur in parallel in Betti numbers and in point count of configuration spaces, to be detailed in the following sections. The focus of my research is to explore such phenomena and explain the analogy by providing a common refinement of both sides.

To make the analogy more transparent, we use the notion of *virtual Poincaré polynomial* in place of the point count. The virtual Poincaré polynomial of a smooth compact complex variety encodes its Betti numbers in the coefficients. The notion of virtual Poincaré polynomial is then extended to all complex varieties so that it behaves well with respect to cut-and-paste and cartesian product, just like the point count. The virtual Poincaré polynomial of a complex variety is determined by the point count of this variety over "enough" finite fields, a result of Nicholas Katz [20, §6]. If X is a smooth complex variety of complex dimension n, the virtual Poincaré polynomial $P^{\mathrm{vir}}(X) = P^{\mathrm{vir}}(X; u) \in \mathbb{Z}[u]$ can be computed in terms of the mixed Hodge numbers:

$$P^{\mathrm{vir}}(X; u) = \sum_{p,q,i \geq 0} (-1)^i h^i_{p,q}(X) u^{2n-p-q}, \tag{2.6}$$

so the mixed Hodges numbers give a refinement of the virtual Poincaré polynomial.

[1] The mixed Hodge numbers *do* refine the Betti numbers because $h^i = \sum_{p,q \geq 0} h^i_{p,q}$. However, the mixed Hodge numbers only refine the virtual Poincaré polynomial, a slightly coarser notion than the point count, though they are closely related.

2.1 Punctured elliptic curve

2.1.1 Background

A punctured elliptic curve E^* is a natural generalization of the affine line in the following sense. Note that $\mathbb{A}^1 = \mathbb{P}^1 \setminus \mathrm{pt}$, the projective line with one point punctured. So E^* is the genus-1 analogue of the genus-0 case $\mathbb{A}^1$. Over $\mathbb{C}$, a punctured elliptic curve has the topology of a punctured torus.

Vakil and Wood [37] found a *motivic* formula of $\mathrm{Conf}^n(X)$ for *any* quasi-projective variety X. By a motivic formula, we mean a formula like $[\mathbb{P}^n] = [\mathrm{pt}] + [\mathbb{A}^1] + \cdots + [\mathbb{A}^n]$ that states how a variety is built up from other varieties via cut-and-paste. In particular, the motivic formula for $\mathrm{Conf}^n(X)$ implies the following formulas for the point count and the virtual Poincaré polynomial of $\mathrm{Conf}^n(E^*)$ of a punctured elliptic curve:

- $$\sum_{n=0}^{\infty} |\mathrm{Conf}^n(E^*)(\mathbb{F}_q)| t^n = \frac{(1-\alpha t)(1-\beta t)(1-qt^2)}{(1-qt)(1-\alpha t^2)(1-\beta t^2)}, \tag{2.7}$$

  where α, β are a pair of conjugate complex numbers satisfying $\alpha\beta = q$ that appear in the Hasse–Weil zeta function of the underlying elliptic curve E;

- $$\sum_{n=0}^{\infty} P^{\mathrm{vir}}(\mathrm{Conf}^n(E^*)(\mathbb{C}); u) t^n = \frac{(1-ut)^2(1-u^2t^2)}{(1-u^2t)(1-ut^2)^2}. \tag{2.8}$$

If we replace α and β in (2.7) by u, and $q = \alpha\beta$ by u^2, then we get the right-hand side of (2.8). Therefore, the virtual Poincaré polynomial in (2.8) encodes essentially the same information as the point count.

2.1.2 Contribution

Using the data of $h^i(\mathrm{Conf}^n(E^*)(\mathbb{C}))$ computed by Drummond-Cole and Knudsen [13], Gilyoung Cheong and I [8] discovered a striking resemblance between the Betti numbers and the virtual Poincaré polynomial:

$$\begin{aligned}\sum_{n,i\geq 0} h^i(\mathrm{Conf}^n(E^*)(\mathbb{C}))(-u)^{2n-i} t^n &= 1 + (u^2-2u)t + (u^4 - 2u^3 + 2u^2)t^2 \\ &\quad + (u^6 - 2u^5 + 4u^4 - 4u^3)t^3 \\ &\quad + (u^8 - 2u^7 + 4u^6 - 5u^5 + 3u^4)t^4 + \dots\end{aligned} \tag{2.9}$$

$$\begin{aligned}\sum_{n=0}^{\infty} P^{\mathrm{vir}}(\mathrm{Conf}^n(E^*)(\mathbb{C}); u) t^n &= \frac{(1-ut)^2(1-u^2t^2)}{(1-u^2t)(1-ut^2)^2} \\ &= 1 + (u^2-2u)t + (u^4 - 2u^3 + 2u)t^2 \\ &\quad + (u^6 - 2u^5 + 4u^3 - 4u^2)t^3 \\ &\quad + (u^8 - 2u^7 + 4u^5 - 5u^4 + 3u^2)t^4 + \dots\end{aligned} \tag{2.10}$$

The two power series in u, t are the same except that the powers of u skip occasionally in (2.10), while there is no skip in (2.9). In particular, while the power series in (2.9) that encodes the Betti numbers of $\mathrm{Conf}^n(E^*)$ is unlikely to be rational, it can be read off of a rational series in such a subtle way.

We stated the precise pattern of the resemblance between the two power series, and gave and proved a common refinement of (2.9) and (2.10).

Theorem 2.1 (Cheong and H. [8])**.** *Keeping the notation above, we have*

(a) *The Betti numbers of* $\mathrm{Conf}^n(E^*)(\mathbb{C})$ *satisfy*

$$\sum_{n,i\geq 0}(-1)^i h^i(\mathrm{Conf}^n(E^*)(\mathbb{C}))u^{2n-w(i)}t^n = \frac{(1-ut)^2(1-u^2t^2)}{(1-u^2t)(1-ut^2)^2}, \tag{2.11}$$

where

$$w(i) := \begin{cases} 3i/2, & i \text{ even;} \\ (3i-1)/2, & i \text{ odd.} \end{cases} \tag{2.12}$$

(b) *The mixed Hodge numbers of* $\mathrm{Conf}^n(E^*)(\mathbb{C})$ *satisfy*

$$\sum_{n,i,p,q\geq 0}(-1)^i h^i_{p,q}(\mathrm{Conf}^n(E^*)(\mathbb{C}))x^{n-p}y^{n-q}u^{2n-w(i)}t^n = \frac{(1-xut)(1-yut)(1-xyu^2t^2)}{(1-xyu^2t)(1-xut^2)(1-yut^2)}, \tag{2.13}$$

where $w(i)$ *is the same as before.*

The particular formula for $w(i)$ says that the power skipping in (2.10) occurs periodically, following the pattern already shown. Theorem 2.1(b) implies Theorem 2.1(a) by letting $x = y = 1$. If we replace u by 1 and then replace x and y by u, then (2.13) becomes (2.10).

The key geometric input behind the proof of Theorem 2.1 is a purity statement about the mixed Hodge structure, where the significance of $w(i)$ becomes evident.

Theorem 2.2 (Cheong and H. [8])**.** *The mixed Hodge structure on the rational cohomology group* $H^i(\mathrm{Conf}^n(E^*)(\mathbb{C});\mathbb{Q})$ *is pure of weight* $w(i)$*, where* $w(i)$ *is the same as in Theorem 2.1. In other words,* $h^i_{p,q}(\mathrm{Conf}^n(E^*)(\mathbb{C})) = 0$ *unless* $p + q = w(i)$*.*

Theorem 2.2 allows the known motivic formula (2.10) to recover (2.9), since the purity statement pinpoints the precise mixed Hodge numbers that contribute to each coefficient of (2.10).

The proof of Theorem 2.2 uses the Leray spectral sequence for the inclusion $F(E^*, n) \hookrightarrow (E^*)^n$ described by Totaro [36], where $F(E^*, n)$ is the *ordered configuration space*, which is closely related to $\mathrm{Conf}^n(E^*)$ because $\mathrm{Conf}^n(E^*)$ is a quotient of $F(E^*, n)$ by the symmetric group S_n. Apart from this spectral sequence, the proof also relies on unusually favorable properties that are specific to one-punctured varietie.

Gilyoung Cheong and I are working on generalizing our work to higher genus cases, where the purity statement analogous to Theorem 2.2 turns out to be false.

2.2 Puncturing an already noncompact variety

2.2.1 Background

From the motivic formula of the configuration space due to Vakil and Wood [37], removing a point has an easily described effect on the point count and the virtual Poincaré polynomial of the configuration space. Let X be any quasi-projective variety over $\mathbb{C}$ or $\mathbb{F}_q$, and let p be a $\mathbb{C}$-point ($\mathbb{F}_q$-point) of X. Consider the open subvariety $X^* = X \setminus p$, then we have

$$\sum_{n=0}^{\infty} P^{\mathrm{vir}}(\mathrm{Conf}^n(X^*)(\mathbb{C}))t^n = \frac{1}{1+t}\sum_{n=0}^{\infty} P^{\mathrm{vir}}(\mathrm{Conf}^n(X)(\mathbb{C}))t^n \tag{2.14}$$

and

$$\sum_{n=0}^{\infty}|\mathrm{Conf}^n(X^*)(\mathbb{F}_q)|t^n = \frac{1}{1+t}\sum_{n=0}^{\infty}|\mathrm{Conf}^n(X)(\mathbb{F}_q)|t^n. \tag{2.15}$$

The analogous story for Betti numbers is much more sutble and case-by-case. In fact, it is surprising that a pattern exists at all. Napolitano [30] showed that if X is any connected noncompact topological surface, and p is a point of X, then for $X^* = X \setminus p$,

$$H^i(\mathrm{Conf}^n(X^*), \mathbb{Z}) \cong \bigoplus_{t=0}^{\infty} H^{i-t}(\mathrm{Conf}^{n-t}(X), \mathbb{Z}). \tag{2.16}$$

He noted that this splitting relation does not hold if X is compact. Kallel [26] extended this splitting with field coefficients to higher dimensions: if X is a closed connected oriented manifold of real dimension $2d$ with $r \geq 1$ points punctured (in particular, X is never compact), and p is a point of X, then for $X^* = X \setminus p$,

$$H^i(\mathrm{Conf}^n(X^*), \mathbb{F}) \cong \bigoplus_{t=0}^{\infty} H^{i-(2d-1)t}(\mathrm{Conf}^{n-t}(X), \mathbb{F}), \tag{2.17}$$

where $\mathbb{F}$ is any field. If we take $\mathbb{F} = \mathbb{Q}$, then (2.17) implies

$$\sum_{i,n\geq 0} h^i(\mathrm{Conf}^n(X^*))(-u)^i t^n = \frac{1}{1+u^{2d-1}t} \sum_{i,n\geq 0} h^i(\mathrm{Conf}^n(X))(-u)^i t^n, \tag{2.18}$$

whose resemblance with (2.14) can be seen more directly.

2.2.2 Contribution

My work [22] gives a common refinement of the motivic formula (2.14) and the splitting theorem in terms of Betti numbers (2.18).

Theorem 2.3 (H. [22])**.** *Let X be a conncected compact smooth complex variety of dimension d with $r \geq 1$ points punctured (in particular, X is never compact). Let p be a point of X, then for $X^* = X \setminus p$, we have*

$$\begin{aligned}\sum_{p,q,i,n\geq 0} & h^i_{p,q}(\mathrm{Conf}^n(X^*)) x^p y^q (-u)^i t^n \\ &= \frac{1}{1+(xy)^d u^{2d-1} t} \sum_{p,q,i,n\geq 0} h^i_{p,q}(\mathrm{Conf}^n(X)) x^p y^q (-u)^i t^n.\end{aligned} \tag{2.19}$$

If $x = y = 1$, then (2.19) becomes (2.18). In (2.19), if we follow the substitution $u \mapsto 1, x \mapsto u^{-1}, y \mapsto u^{-1}, t \mapsto tu^{2d}$, then we recover (2.14), where we note that

$$\frac{1}{1+(xy)^d u^{2d-1} t} \mapsto \frac{1}{1+(u^{-1}u^{-1})^d (1^{2d-1})(tu^{2d})} = \frac{1}{1+t}. \tag{2.20}$$

It is worth noting that the very formula (2.19) implies that the mixed Hodge structure on $H^i(\mathrm{Conf}^n(X^*))$ is not pure in general, even in the case where $H^i(\mathrm{Conf}^n(X))$ is pure.

Theorem 2.3 is further generalized and refined in the same work. As an application, (2.17) holds for more classes of higher dimensional noncompact smooth varieties.

We say the rational cohomology of a complex manifold X to be *pure of slope* λ if the mixed Hodge structure on $H^i(X; \mathbb{Q})$ is pure of weight λi for all i. In particular, for any i such that λi is not an integer, this condition forces $H^i(X; \mathbb{Q})$ to be zero.

Theorem 2.4 (H. [22])**.**

(a) *Theorem 2.3 holds more generally if X is a connected noncompact smooth complex variety whose rational cohomology is pure of slope λ for some rational number $1 \leq \lambda \leq 2$.*

(b) *Theorem 2.3 also holds if X is such a variety with $r \geq 0$ points punctured.*

(c) *If X is as in* (b)*, then we have an S_n-equivariant formula for the rational cohomology of the ordered configuration space $F(X^*, n)$ of X^*:*

$$H^i(F(X^*, n), \mathbb{Q}) \cong \bigoplus_{t=0}^{\infty} \operatorname{Ind}_{S_{n-t}}^{S_n} H^{i-(2d-1)t}(F(X, n-t), \mathbb{Q})(-d \cdot t), \tag{2.21}$$

where the isomorphism is understood both as S_n-representations, and as mixed Hodge structures up to semisimplification. The notation $(-d \cdot t)$ denotes the Tate twist, which does nothing to the underlying vector space but shifts the degrees in the mixed Hodge structure.

Theorem 2.3 is a special case of Theorem 2.4 because a one-punctured compact smooth complex variety is pure of slope 1, namely, the i-th rational cohomology is pure of weight i. Other cases that satisfy the hypothesis of Theorem 2.4 include:

- X is the affine space $\mathbb{C}^d$, or the complement of a hyperplane arrangement therein. In this case, X is pure of slope 2.
- X is the torus $(\mathbb{C}^*)^d$, or the complement of a union of 1-codimensional subtori therein. In this case, X is pure of slope 2.
- X is the complement of a smooth plane curve in $\mathbb{P}^2$. In this case, X is pure of slope 3/2.

Theorem 2.4(c) recovers (2.19) by taking S_n-invariants on both sides, so it is a refinement of Theorem 2.3.

The proof of these results uses the Leray spectral sequence arising from an arrangement of subvarieties, developed by [3, 32, 35]. This spectral sequence is a generalization of the spectral sequence of Totaro [36], and it allows computing the cohomology of $\operatorname{Conf}^n(X)$ from the cohomology of the variety Y that contains X as an open subvariety. The crux of my proof is finding the suitable choice of Y (as there are abundant choices present). For example, if X is a smooth compact complex variety with $r \geq 1$ points removed, it is important that we let Y be the same compact variety with exactly one of those points removed.

I am working on generalizing Theorem 2.4 further, aimed at unifying the special cases where this theorem holds.

3 Unit equations on quaternions and other division algebras

3.1 Background

A unit equation is an equation of the form $x_1 + \cdots + x_n = 1$, where x_i ranges from some multiplicative group of "units". The fundamental theorem in the study of unit equations is, if $\Gamma_1, \ldots, \Gamma_n$ are finitely generated multiplicative subgroups of $\mathbb{C}^\times$, then the unit equation $x_1 + \cdots + x_n = 1$ only has finitely many *nondegenerate* solutions with $x_i \in \Gamma_i$, where a solution is nondegenerate if no subsum equals to 1. Moreover, there are effective bounds for the solutions. This theorem has applications [21, 28] to the study of *Diophantine equations* (polynomial equations asking for integer solutions) and *arithmetic dynamics* (the behavior of iterations of algebro-geometric maps, such as polynomials and rational functions).

One way to apply the unit equations is to study the dynamics of a geometric object X by studying the set of endomorphisms of X using the unit equations whenever applicable. For example, Odesky [31, Theorem 1.2.3] shows that if f and g are non-injective regular self-maps on an elliptic curve E over $\mathbb{C}$, then they must have a common iterate whenever there exist a forward orbit of f and a forward orbit of g that intersect at infinitely many points. The key is reducing the problem to a unit equation on the endomorphism algebra $\operatorname{End}(E)$, where the finiteness theorem applies. The same technique works if E

is replaced by a high dimensional analogue, namely, an abelian variety A, as long as a version of the finiteness theorem holds for unit equations on $\mathrm{End}(A)$, which may not be commutative. This motivates the study of unit equations on noncommutative algebras. Little has been known about unit equations in the noncommutative setting, where many important techniques (such as p-adic completion) do not exist.

3.2 Contribution

I proved an effective finiteness result [25] for unit equations on the algebra $\mathbb{H}_a$ of quaternions $a+bi+cj+dk$ with a, b, c, d being real algebraic numbers. The *norm* of a quaternion $a+bi+cj+dk$ is $\sqrt{a^2+b^2+c^2+d^2}$.

Theorem 3.1 (H. [25])**.** *Let Γ_1, Γ_2 be semigroups of $\mathbb{H}_a^\times$ generated by finitely many elements of norms greater than 1, and fix $a, a', b, b' \in \mathbb{H}_a^\times$. If Γ_1 is commutative, then the equation*

$$afa' + bgb' = 1 \tag{3.1}$$

has only finitely many solutions with $f \in \Gamma_1$ and $g \in \Gamma_2$. Moreover, there are effective bounds for the solutions (f, g).

We point out that Γ_2 need not be commutative, and even though Γ_1 is commutative, the product afa' need not be commutative. Theorem 3.1 implies an analogue of Odesky's orbit intersection result, where E is replaced by an elliptic curve in positive characteristic. I conjectured that a generalization of Theorem 3.1 holds for any finite dimensional division algebra over $\mathbb{Q}$ and all choices of finitely generated semigroups. It turns out that the generalized statement holds if all the semigroups are commutative [25, Propositions 1.5 and A.2], which implies an analogue of Odesky's orbit intersection result for any simple abelian variety in any characteristic [25, Corollary 1.7], since the endomorphism algebra of a simple abelian variety is a division algebra in general. The current method does not work if both Γ_1 and Γ_2 are noncommutative, and tackling the conjecture even for the simplest examples of noncommutative Γ_1 and Γ_2 require new techniques to be introduced.

4 Random matrices over a p-adic ring

4.1 Background

For a prime p, consider the ring $\mathbb{Z}_p$ of p-adic integers and equip the additive group $\mathrm{Mat}_n(\mathbb{Z}_p)$ of n by n matrices with the Haar measure with total measure 1. Questions can be asked about the distribution of a certain quantity associated to a random matrix. Friedman and Washington [16] proved the following formulas about the distribution of the *cokernel*: Let H be a finite abelian p-group, then

$$\lim_{n\to\infty} \operatorname*{Prob}_{A\in \mathrm{Mat}_n(\mathbb{Z}_p)} (\operatorname{coker} A \cong H) = |\operatorname{Aut} H|^{-1} \prod_{i=1}^{\infty} (1-p^{-i}), \tag{4.1}$$

$$\lim_{n\to\infty} \operatorname*{Prob}_{A\in \mathrm{GL}_n(\mathbb{Z}_p)} (\operatorname{coker}(A - 1_n) \cong H) = |\operatorname{Aut} H|^{-1} \prod_{i=1}^{\infty} (1-p^{-i}), \tag{4.2}$$

where 1_n is the $n \times n$ identity matrix and the probability measure on $\mathrm{GL}_n(\mathbb{Z}_p)$ is the restriction of the measure on $\mathrm{Mat}_n(\mathbb{Z}_p)$ but with the total measure normalized to 1. The formulas mean that the cokernel of a random $\mathbb{Z}_p$-matrix (or a random invertible $\mathbb{Z}_p$-matrix minus the identity) distribute according to the *Cohen–Lenstra distribution*, which assigns to a p-group a probability inversely proportional to the size of automorphism group.

Random matrices over a finite field also produce the Cohen–Lenstra distribution, but in greater generality. Fulman's result [17] implies that the distribution of the nilpotent parts of the rational canonical forms of polynomials of a random matrix in $\mathrm{GL}_n(\mathbb{F}_q)$: let $P_1(t), \dots, P_r(t)$ be distinct monic irreducible polynomials in $\mathbb{F}_q[t] \setminus \{t\}$ of degrees $d_1, \dots, d_r$, and $H_1, \dots, H_r$ be finite-cardinality $\mathbb{F}_q[[t]]$-modules, then

$$\begin{aligned}&\lim_{n\to\infty} \underset{A\in \mathrm{GL}_n(\mathbb{F}_q)}{\mathrm{Prob}} (P_j(A)[t^\infty] \cong H_j, 1 \le j \le r)\\ &= \begin{cases} \prod_{j=1}^r |\mathrm{Aut}_{\kappa_j[[t]]}(H_j' \otimes_{\mathbb{F}_q} \kappa_j)|^{-1} \prod_{i=1}^\infty (1-q^{-id_j})^{-1}, & H_j \text{ is of the form } (H_j')^{\oplus d_j};\\ 0, & \text{otherwise,}\end{cases}\end{aligned} \tag{4.3}$$

where κ_j is the field extension of $\mathbb{F}_q$ of degree d_j, and $P_j(A)[t^\infty]$ denotes the $\mathbb{F}_q[[t]]$-module associated to the nilpotent part of the rational canonical form of $P_j(A)$. This result means that the nilpotent part of $P_j(A)$ follows a variant of the Cohen–Lenstra distribution, and the nilpotent part of $P_1(A), \dots, P_r(A)$ distribute independently as $n \to \infty$. The nontrivial generality of (4.3) lies in that the polynomials $P_j(t)$ may not be of degree 1.

The motivation of my research is to understand the relationship between these results by investigating the cokernel distribution of a polynomial of a random $\mathbb{Z}_p$-matrix.

4.2 Contribution

In [9], Gilyoung Cheong and I found a common generalization of (4.1) and (4.2).

Theorem 4.1 (Cheong and H. [9])**.** *Let $P_1(t), \dots, P_r(t) \in \mathbb{Z}_p[t]$ be monic polynomials whose reductions mod p are distinct irreducible polynomials in $\mathbb{F}_p[t]$. Suppose* $\deg P_r(t) = 1$*, then given a finite abelian p-group H, we have*

$$\lim_{n\to\infty} \underset{A\in \mathrm{Mat}_n(\mathbb{Z}_p)}{\mathrm{Prob}} \begin{pmatrix} \operatorname{coker} P_1(A) = \cdots = \operatorname{coker} P_{r-1}(A) = 0 \\ \operatorname{coker} P_r(A) \cong H \end{pmatrix} = |\mathrm{Aut}(H)|^{-1} \prod_{j=1}^{r} \prod_{i=1}^{\infty} (1 - p^{-i \deg P_j}) \tag{4.4}$$

This formula specializes to (4.1) if we take $r = 1$ and $P_1(t) = t$, and it specializes to (4.2) if we take $r = 2$, $P_1(t) = t$ and $P_2(t) = t - 1$. Since the degrees of $P_1, \dots, P_{r-1}$ can be higher than 1, Theorem 4.1 can be viewed as an analogue of (4.3) in the context of p-adic matrices.

The proof makes use of the combinatorial technique of *cycle index* used by Stong [34] as well as the Cohen–Lenstra series (in the sense of Section 1.1) of the variety $\mathbb{A}^1_{\mathbb{F}_q} \setminus \{P_1, \dots, P_r\}$, with an explanation available in [8, §8.2]. Cheong and I made a conjectural generalization of Theorem 4.1 that involves p-groups $H_1, \dots, H_r$, where $H_1, \dots, H_{r-1}$ are no longer required to be zero. Investigating this conjecture will certainly involve new techniques, and it would be an exciting project in the future.

References

[1] V. I. Arnol'd. The cohomology ring of the group of dyed braids. *Mat. Zametki*, 5:227–231, 1969.

[2] V. Baranovsky. The variety of pairs of commuting nilpotent matrices is irreducible. *Transform. Groups*, 6(1):3–8, 2001.

[3] C. Bibby. Cohomology of abelian arrangements. *Proc. Amer. Math. Soc.*, 144(7):3093–3104, 2016.

[4] V. M. Bondarenko, T. G. Gerasimova, and V. V. Sergeichuk. Pairs of mutually annihilating operators. *Linear Algebra Appl.*, 430(1):86–105, 2009.

[5] J. Bryan and A. Morrison. Motivic classes of commuting varieties via power structures. *J. Algebraic Geom.*, 24(1):183–199, 2015.

[6] X. Chen and M. Lu. Varieties of modules over the quantum plane. 2019. arXiv: 1910.02198.

[7] X. Chen and W. Wang. Anti-commuting matrices. *Trans. Amer. Math. Soc.*, 373(3):1597–1617, 2020.

[8] G. Cheong and Y. Huang. Rationality for the Betti numbers of the unordered configuration spaces of a punctured torus. Submitted. arXiv: 2009.07976.

[9] G. Cheong and Y. Huang. Cohen–Lenstra distributions via random matrices over complete discrete valuation rings with finite residue fields. *Illinois Journal of Mathematics*, 65(2):385 – 415, 2021.

[10] H. Cohen and H. W. Lenstra, Jr. Heuristics on class groups of number fields. In *Number theory, Noordwijkerhout 1983 (Noordwijkerhout, 1983)*, volume 1068 of *Lecture Notes in Math.*, pages 33–62. Springer, Berlin, 1984.

[11] W. Crawley-Boevey and J. Schröer. Irreducible components of varieties of modules. *J. Reine Angew. Math.*, 553:201–220, 2002.

[12] P. Deligne. Théorie de Hodge. II. *Inst. Hautes Études Sci. Publ. Math.*, (40):5–57, 1971.

[13] G. C. Drummond-Cole and B. Knudsen. Betti numbers of configuration spaces of surfaces. *J. Lond. Math. Soc. (2)*, 96(2):367–393, 2017.

[14] W. Feit and N. J. Fine. Pairs of commuting matrices over a finite field. *Duke Math. J.*, 27:91–94, 1960.

[15] N. J. Fine and I. N. Herstein. The probability that a matrix be nilpotent. *Illinois J. Math.*, 2:499–504, 1958.

[16] E. Friedman and L. C. Washington. On the distribution of divisor class groups of curves over a finite field. In *Théorie des nombres (Quebec, PQ, 1987)*, pages 227–239. de Gruyter, Berlin, 1989.

[17] J. Fulman. Random matrix theory over finite fields. *Bull. New. Ser. Am. Math. Soc.*, 39(1):51–85, 2001.

[18] J. Fulman and R. Guralnick. Enumeration of commuting pairs in Lie algebras over finite fields. *Ann. Comb.*, 22(2):295–316, 2018.

[19] M. Gerstenhaber. On dominance and varieties of commuting matrices. *Ann. of Math. (2)*, 73:324–348, 1961.

[20] T. Hausel and F. Rodriguez-Villegas. Mixed Hodge polynomials of character varieties. *Invent. Math.*, 174(3):555–624, 2008. With an appendix by Nicholas M. Katz.

[21] M. Hindry and J. H. Silverman. *Diophantine Geometry: an Introduction.* Springer-Verlag, New York, 2013.

[22] Y. Huang. Cohomology of configuration spaces on punctured varieties. arXiv: 2011.07153.

[23] Y. Huang. Counting on the variety of modules over the quantum plane. Submitted. arXiv: 2110.15570.

[24] Y. Huang. Mutually annhilating matrices, and a Cohen–Lenstra series for the nodal singularity. Submitted. arXiv: 2110.15566.

[25] Y. Huang. Unit equations on quaternions. *Q. J. Math.*, 71(4):1521–1534, 2020.

[26] S. Kallel. Symmetric products, duality and homological dimension of configuration spaces. In *Groups, homotopy and configuration spaces*, volume 13 of *Geom. Topol. Monogr.*, pages 499–527. Geom. Topol. Publ., Coventry, 2008.

[27] M. Kim. Weights in cohomology groups arising from hyperplane arrangements. *Proc. Amer. Math. Soc.*, 120(3):697–703, 1994.

[28] S. Lang. Integral points on curves. *Publications Mathématiques de l'IHÉS*, 6:27–43, 1960.

[29] T. S. Motzkin and O. Taussky. Pairs of matrices with property L. II. *Trans. Amer. Math. Soc.*, 80:387–401, 1955.

[30] F. Napolitano. On the cohomology of configuration spaces on surfaces. *J. London Math. Soc. (2)*, 68(2):477–492, 2003.

[31] A. Odesky. *The Orbit Intersection Problem and Polynomial Functions.* PhD Thesis, the University of Michigan–Ann Arbor, 2020.

[32] D. Petersen. A spectral sequence for stratified spaces and configuration spaces of points. *Geom. Topol.*, 21(4):2527–2555, 2017.

[33] R. W. Richardson. Commuting varieties of semisimple Lie algebras and algebraic groups. *Compositio Math.*, 38(3):311–327, 1979.

[34] R. Stong. Some asymptotic results on finite vector spaces. *Adv. in Appl. Math.*, 9(2):167–199, 1988.

[35] P. Tosteson. Lattice spectral sequences and cohomology of configuration spaces. 2016. arXiv: 1612.06034.

[36] B. Totaro. Configuration spaces of algebraic varieties. *Topology*, 35(4):1057–1067, 1996.

[37] R. Vakil and M. M. Wood. Discriminants in the Grothendieck ring. *Duke Math. J.*, 164(6):1139–1185, 2015.