BOOK REVIEWS

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 46, Number 1, January 2009, Pages 137–141
S 0273-0979(08)01203-2
Article electronically published on September 15, 2008

Moments, monodromy, and perversity: Diophantine perspective, by Nicholas M. Katz, Annals of Mathematics Studies, 159, Princeton University Press, Princeton, NJ, 2005, viii+475 pp., $59.50, paper, ISBN 13: 978-0-691-12330-1; $99.50 (cloth), ISBN 13: 978-0-691-12329-5

The study of exponential sums of over finite fields goes back to Gauss' computation of

$$\sum_{n=0}^{p-1} \zeta^{n^2}, \tag{1}$$

where p is prime and ζ is a pth root of unity. Approaching (1) naively as a p-step random walk, one might expect it to have absolute value on the order of $\sqrt{p}$. Surprisingly, it turns out that the absolute value is always *exactly* $\sqrt{p}$. (The proof is not difficult, being the finite field analogue of the computation of $\int_{-\infty}^{\infty} e^{-x^2} dx$ via double integrals.)

In general, we would like to understand the behavior of sums of the form

$$\sum_{x \in V(\mathbb{F})} \rho(x),$$

where $\mathbb{F}$ is a finite field, V is an algebraic variety over $\mathbb{F}$, and $\rho(x)$ is a root of unity which depends in some algebraic way on x. Often, $\rho(x)$ is of the form $\chi(f(x))$, where f is a function on X, and χ is a character of the additive or multiplicative group of $\mathbb{F}$. The Kloosterman sums are typical:

$$\sum_{x \in \mathbb{F}_p^\times} \psi(ax + b/x), \tag{2}$$

where $\psi\colon \mathbb{F}_p \to \mathbb{C}^\times$ is the additive character sending $1 \in \mathbb{F}_p$ to $e^{2\pi i/p}$, and $a, b \in \mathbb{F}_p^\times$ are constants. We can regard V in this case as the hyperbola $xy = 1$ (equivalently, the affine line with 0 removed) and f as the function $ax + by$ on this hyperbola. Another typical example is

$$\sum_{x \in \mathbb{F}_p} \chi(x^3 + ax + b), \tag{3}$$

2000 *Mathematics Subject Classification.* Primary 14F20; Secondary 11G05, 11T23, 14G10, 14G15.

where $\chi(c)$ denotes the Legendre symbol $\left(\frac{c}{p}\right)$. As $\chi(c)$ is 1 less than the number of solutions of $y^2 = c$ in $\mathbb{F}_p$, this sum counts the number of points on the affine elliptic curve $y^2 = x^3 + ax + b$, less p.

The sums (2) and (3) no longer have absolute value $\sqrt{p}$, but neither do they behave as one would expect if they were really random walks. In each case, the Riemann hypothesis for curves over finite fields implies that the absolute value is less than $2\sqrt{p}$. To analyze the statistical behavior of such sums, we need to take limits, which means that we need to study infinite families of sums. The obvious way to achieve this is to let p tend to infinity. Until very recently, this seemed beyond reach. There has been some remarkable recent progress [5, 6, 7] on the Sato-Tate conjecture, which predicts the statistical behavior of sums of type (3) as $p \to \infty$, so now, perhaps, one should say that it is merely very difficult. There is an easier alternative, however: we can fix the characteristic of our finite fields, but allow the fields themselves to grow, suitably extending the definition of the summands. (In the additive case, this means replacing $\psi(f(x))$ by $\psi(\mathrm{Trace}(f(x)))$; in the multiplicative case, it means replacing $\chi(f(x))$ by $\chi(\mathrm{Norm}(f(x)))$.) As examples (2) and (3) illustrate, the sums usually have parameters, or more precisely, a parameter variety X, defined over a finite base field $\mathbb{F}_q$. For every positive integer n, every point $x \in X(\mathbb{F}_{q^k})$ determines a sum value $S(k, x)$. We study the value distribution as k and x vary.

Deligne proved in [2] that such families of exponential sums satisfy a generalized Sato-Tate law. In favorable situations, this means that there exists an integer w, a compact Lie group G, and a complex representation (ρ, V) of G with character χ, such that the distribution of values of

$$q^{-kw/2} S(k, x),$$

as k ranges over the positive integers and x ranges over $X(\mathbb{F}_{q^k})$, approaches the distribution of values of $\chi(g)$ for g uniformly distributed on G, as $k \to \infty$. We can express the same thing by saying that an N-tuple of points on the unit circle gives a conjugacy class in $\mathrm{U}(N)$, and the set of N-tuples of eigenvalues of Frobenius elements gives the same distribution on conjugacy classes in $\mathrm{U}(N)$ as the elements $\rho(g) \in \mathrm{U}(N)$ as g ranges over G. The group G can be regarded as the *geometric monodromy group* of the family of exponential sums (or at least the compact real form of geometric monodromy).

The idea behind all of this is that a family of exponential sums is encoded by a geometric object over the parameter space X. In the simplest cases, this object is just a *lisse sheaf* on X, i.e. an ℓ-adic representation of the fundamental group of X. Every point $x \in X(\mathbb{F}_{q^k})$ defines a map $x = \mathrm{Spec}\, \mathbb{F}_{q^k} \to X$ and therefore a map from the fundamental group of x to that of the fundamental group of X. The fundamental group of the spectrum of a field F is the absolute Galois group of F, which in the case $F = \mathbb{F}_{q^k}$ is generated by the Frobenius automorphism $x \mapsto x^{q^k}$. We can therefore think of x as determining a "Frobenius element" (or at least a Frobenius conjugacy class) of $\pi_1(X)$; the sheaf determines the trace of this element. Roughly, G can be thought of as the monodromy group of the sheaf. (We remark parenthetically that while lisse sheaves are the simplest and most natural kind of coefficient systems to consider, they unfortunately have poor stability properties under the cohomological formalism. In [2], Deligne showed that it is technically better to work in a much broader setting, namely complexes

of sheaves of $\mathbb{Z}_\ell$-modules, taken up to quasi-isomorphism. Perverse sheaves, in the sense of Beĭlinson-Bernstein-Deligne [1], based on earlier work of Goresky and Macpherson [3], provide a good compromise, having good stability properties and also fitting well into the weight formalism. In what follows, we shall ignore this subtlety.)

To see how a lisse sheaf can be attached to a family of exponential sums, consider, for example, (3). Let $S(n, a, b)$ denote the value of the sum over $\mathbb{F}_{p^k}$, where a and b are parameters in that field. The Lefschetz trace formula tells us that the number of points on a (projective) elliptic curve E over $\mathbb{F}_{p^k}$ is the alternating sum of traces of the p^k-Frobenius acting on the (étale) cohomology groups of $\bar{E}$ (the same elliptic curve, with scalars extended to $\bar{\mathbb{F}}_{p^k}$). This sum can be written $1 - a_{p^k} + p^k$. The projectivization of the affine curve $y^2 = x^3 + ax + b$ has one point on the line at infinity, so $a_{p^k} = -S(n, a, b)$ gives the trace of Frobenius acting on $H^1(\bar{E})$. We can define a universal elliptic curve over the parameter space (a, b) (at least on the open subvariety $4a^3 + 27b^2 \neq 0$), a family whose fiber at (a, b) is the projectivization of $y^2 = x^3 + ax + b$. The lisse sheaf producing the values of $-S(n, a, b)$ is therefore the relative H^1 of this universal curve.

The key problem is to determine the monodromy groups $G \subset U(N)$ of the sheaves of interest (up to conjugation). In many cases of particular interest, these monodromy groups are known, thanks to earlier work of Katz ([9], [10], [11]). In the great majority of these cases, G turns out to be $\mathrm{SU}(N)$, $\mathrm{SO}(N)$, $\mathrm{O}(N)$, or $\mathrm{Sp}(N)$. The last three possibilities occur when the representation in question is self-dual (i.e., when the exponential sums in the family are real-valued), and in this case, the self-duality, whether symmetric or anti-symmetric, is generally obvious. Setting aside the subtle distinction between $\mathrm{SO}(N)$ and $\mathrm{O}(N)$, one finds that "in nature", monodromy groups are nearly always "as large as possible," given the constraints of dimension and duality.

There are two main sources of information in monodromy calculations. On the one hand, the computation of local monodromy provides conjugacy classes in $\mathrm{GL}(N, \mathbb{C})$ which intersect G (or at least its complexification) nontrivially. Such classes can give a great deal of information; for instance, it is very useful to know that G contains a reflection or that its complexification contains a transvection. On the other hand, if V is the restriction of the standard N-dimensional representation of $\mathrm{U}(N)$ to G, one can often compute the dimension of the space of G-invariants of $V^{\otimes a} \otimes (V^*)^{\otimes b}$ when a and b are sufficiently small non-negative integers. These dimensions can be viewed as *moments* of the measure $\mu_{G,V}$ associated to the pair (G, V):

$$\dim(V^{\otimes a} \otimes (V^*)^{\otimes b})^G = \int_G \chi(g)^a \bar{\chi}(g)^b \, dg = \int_{\mathbb{C}} z^a \bar{z}^b \mu_{G,V}.$$

For example, by Schur's lemma, the $aq = b = 1$ moment equals 1 if and only if V is irreducible. In his earlier books on monodromy, Katz mostly used local monodromy information, with a small admixture of moment data. A major novelty in the book under review is that it places the primary emphasis on moment calculations.

The small moments of "large" monodromy groups are well known from classical invariant theory [13]. For $\mathrm{SU}(N)$ and $a + b < N$, they are given by

$$\dim(V^{\otimes a} \otimes (V^*)^{\otimes b})^{\mathrm{SU}(N)} = \begin{cases} a! & \text{if } a = b, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

For $G \in \{\mathrm{SO}(N), \mathrm{O}(N), \mathrm{Sp}(N)\}$ and $a+b<N$, they are given by

$$\dim(V^{\otimes a} \otimes (V^*)^{\otimes b})^{\mathrm{SU}(N)} = \begin{cases} (a+b-1)!! & \text{if } a+b \text{ is even,} \\ 0 & \text{if } a+b \text{ is odd.} \end{cases} \tag{5}$$

Interestingly, the moments of the measures $\mu_{G,V}$ arising from classical groups *stabilize* to give the moments of Gaussian distributions (complex or real), which implies that, although each individual measure is compactly supported, they tend in the large-N limit to (complex or real) Gaussian measures.

An optimist might ask if (G,V) is determined up to isomorphism, or equivalently, if G is determined up to conjugation in $\mathrm{U}(N)$, by its moments alone. In general, the answer is no. For example, if G is a finite group of order N and V is its regular representation, the only information encoded in the moments of $\mu_{G,V}$ is the order of G. Even for connected semisimple Lie groups, there are examples of nonisomorphic pairs (G,V) (i.e., nonconjugate subgroups of $\mathrm{U}(N)$) with the same Sato-Tate measure [12]. However, in most interesting cases, it appears that (G,V) *is* uniquely determined by $\mu_{G,V}$ and indeed by a small number of moments. In particular, it has been known for at least a decade that any pair (G,V) which gives the same invariants as (4) (resp. (5)) for $a+b \le 4$ must satisfy $|G|<\infty$ or $\mathrm{SU}(N) \subset G \subset \mathrm{U}(N)$ (resp. $G \in \{\mathrm{SO}(N), \mathrm{O}(N), \mathrm{Sp}(N)\}$). Recently, Guralnick and Tiep [4] proved that one can rule out finite groups entirely by considering all moments with $a+b \le 12$, or even $a+b \le 8$ if one knows the value of N. Thus, in most cases that actually arise, the monodromy of an exponential sum can be deduced from moments with $a+b \le 8$, provided that the determinant representation and the sign of the self-duality (if any) is known.

By analogy with the circle method, which becomes easier the more parameters are available, for exponential sums over finite fields, moment computations become easier when the dimension of the parameter variety grows. This marks another difference between this book and previous work of the author, which emphasizes the case in which the parameter variety is a curve. A typical parameter variety here is the space of all polynomials of sufficiently high degree on an underlying vector space or the space of effective divisors on a projective curve linearly equivalent to a divisor of sufficiently high degree.

The goal is to prove that the monodromy groups of the exponential sums in question are as large as possible, given duality and determinantal conditions. The book discusses a number of different kinds of sums, connected with additive and multiplicative characters and with elliptic curves. For an example of the first type, given a prime p and integers $n \ge 1$ and $e \ge 3$, one defines a function on the variety $\mathbb{A}^m/\mathbb{F}_p$ of polynomials on $\mathbb{A}^n/\mathbb{F}_p$ of degree $\le e$ given by

$$P(x) \in \mathbb{A}^m(\mathbb{F}_{p^k}) \mapsto \sum_{x \in \mathbb{F}_{p^k}^n} \psi(\mathrm{Trace}(P(x))).$$

This turns out to be given by a perverse sheaf, but after restriction to a suitable dense open set, it is given by a lisse sheaf of rank N. If $p \ge 7$, then the real compact form of the monodromy group G of this sheaf contains $\mathrm{SU}(N)$. A similar result holds in the setting of multiplicative characters χ: here, the sums in question are

$$\sum_{x \in \mathbb{F}_{p^k}^n} \chi(\mathrm{Norm}(P(x))).$$

A third class of sums extensively considered are those connected with counting points on families of elliptic surfaces defined over a fixed base curve $C/\mathbb{F}_p$. These cases are especially interesting since they make it possible to compute the average analytic rank of various families of elliptic curves over a function field in characteristic p.

Given the weight of machinery necessary just to parse the title *Moments, Monodromy, and Perversity: a Diophantine Perspective*, one might expect the book to be heavy going. On the contrary, it is written in a clear, concrete style, with the emphasis on examples and a wealth of interesting and illuminating digressions. It should appeal strongly to any mathematician interested in the cohomological theory of exponential sums. On the other hand, it is not really an introduction to the subject. The reader in search of such an introduction might profitably consult one of the earlier monographs by the same author, such as [8] or [9].

References

[1] Beĭlinson, A. A.; Bernstein, J.; Deligne, P. Faisceaux pervers. *Analysis and topology on singular spaces, I (Luminy, 1981)*, 5–171, Astérisque, **100**, Soc. Math. France, Paris, 1982. MR751966 (86g:32015)

[2] Deligne, Pierre: La conjecture de Weil. II. *Inst. Hautes Études Sci. Publ. Math.* **52** (1980), 137–252. MR601520 (83c:14017)

[3] Goresky, Mark; MacPherson, Robert. Intersection homology theory. *Topology* **19** (1980), no. 2, 135–162. MR572580 (82b:57010)

[4] Guralnick, Robert M.; Tiep, Pham Huu: Decompositions of small tensor powers and Larsen's conjecture. *Represent. Theory* **9** (2005), 138–208. MR2123127 (2006a:20082)

[5] Clozel, Laurent; Harris, Michael; Taylor, Richard: Automorphy for some ℓ-adic lifts of automorphic mod ℓ representations, preprint.

[6] Harris, Michael; Shepherd-Barron, Nicholas; Taylor, Richard: A family of Calabi–Yau varieties and potential automorphy, preprint.

[7] Taylor, Richard: Automorphy for some ℓ-adic lifts of automorphic mod ℓ representations, II, preprint.

[8] Katz, Nicholas M.: *Sommes exponentielles.* Course taught at the University of Paris, Orsay, Fall 1979. With a preface by Luc Illusie. Notes written by Gérard Laumon. With an English summary. Astérisque, 79. Société Mathématique de France, Paris, 1980. MR617009 (82m:10059)

[9] Katz, Nicholas M.: *Gauss sums, Kloosterman sums, and monodromy groups.* Annals of Mathematics Studies, 116. Princeton University Press, Princeton, NJ, 1988. MR955052 (91a:11028)

[10] Katz, Nicholas M.: *Exponential sums and differential equations.* Annals of Mathematics Studies, 124. Princeton University Press, Princeton, NJ, 1990. MR1081536 (93a:14009)

[11] Katz, Nicholas M.: *Twisted L-functions and monodromy.* Annals of Mathematics Studies, 150. Princeton University Press, Princeton, NJ, 2002. MR1875130 (2003i:11087)

[12] Larsen, M.; Pink, R.: Determining representations from invariant dimensions. *Invent. Math.* **102** (1990), no. 2, 377–398. MR1074479 (92c:22026)

[13] Weyl, Hermann: *The Classical Groups. Their Invariants and Representations.* Princeton University Press, Princeton, N.J., 1939. MR1488158 (98k:01049)

Michael Larsen

Indiana University

E-mail address: larsen@math.indiana.edu