From *SIAM News*, Volume 45, Number 5, June 2012

Elliptic Curve Groups and Their Cryptographic Applications

Elliptic Tales: Curves, Counting, and Number Theory. *By Avner Ash and Robert Gross, Princeton University Press, Princeton, New Jersey, 2012, xxii+253 pages, $29.95.*

BOOK REVIEW

By James Case

The Birch and Swinnerton-Dyer conjecture is perhaps the least well-known of the Clay Institute's seven million-dollar challenge problems. It concerns "rational points"—points whose Cartesian coordinates x and y are both rational numbers—on elliptic curves. With the development of "elliptic curve cryptography" in the last quarter century, such curves have been increasingly important for modern communication systems. Neal Koblitz and Victor S. Miller suggested the application in 1985, when they independently observed that the protocols employed since the mid-1970s in public-key cryptography could be modified to work in arbitrary finite groups, including elliptic curve groups.

Second-generation Public-key Cryptography

First-generation public-key systems, such as RSA and Diffie–Hellman, make use of the multiplicative group of non-zero elements of F_p, the field of residues modulo a large prime number p. Second-generation systems, utilizing elliptic curve groups, offer a variety of advantages: more security per bit of key size than either RSA or DH, lower transmission costs, and more efficient use of electrical power. The degree of security attained with an RSA or DH key of 3072 bits, for instance, is no greater than that for an elliptic curve key of only 256 bits, at a fraction of the monetary cost and electric power consumption. Moreover, the advantages of elliptic curves only increase with enhanced security levels. Power consumption is particularly important for financial transactions, now so routinely conducted with the aid of smartcards and low-power smartcard readers.

The majority of public-key systems in use today employ 1024-bit keys for RSA and DH authentication protocols. In 2005, the National Institute of Standards and Technology declared that such protocols would be adequate through 2010, but recommended their replacement thereafter by more secure alternatives. Since 2005, NIST has published and updated a list of as many as 15 elliptic curves of varying sizes that it deems suitable for current cryptographic use. The National Security Agency is gradually transitioning to elliptic curve-based public-key systems for protecting both classified and unclassified information, as are both the U.S. Department of Defense and NATO.

Various firms and individuals have been granted patents for proprietary implementations of elliptic curve cryptography. The Canadian company Certicom holds more than 130 such patents, and NSA recently purchased a license covering 26 of them as they apply to agency activities. The firm intends to market software toolkits to NSA licensees, and perhaps to others as well.

Theory of Rational Cubic Curves

The theory of rational points on "rational cubic curves," of which elliptic curves are a special case, has been under development since the 19th century. Because there is no known method for deciding in a finite number of steps whether a rational cubic curve contains a rational point, the existing theory is confined to curves known in advance to contain at least one such point,* hereinafter denoted ∞. The corresponding problem for rational quadratics has a complete (if not quite elementary) solution, while that for rational curves of higher order seems to be intractable.

The equation of any rational cubic curve possessing at least one rational solution can be reduced, via well-known transformations, to its Weierstrass *normal form*

$$y^2 = x^3 + Ax + B. \qquad (E)$$

It is occasionally important to distinguish between the equation (E) and the "curve" C consisting, for some field K, of elements (x,y) of $K \times K$ that satisfy (E). Common choices of K are $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}$, and the field F_p mentioned above. Whereas the question of rational points on curves C satisfying (E) makes sense for rational coefficients A and B, the bulk of the existing theory pertains to the case in which both are integers. The possible shapes of such curves are quite diverse, as shown in Figure 1.

The reason it is possible to construct an elaborate theory of rational points on elliptic curves is that such points form finitely generated abelian groups. Although usages differ, most writers call the curve

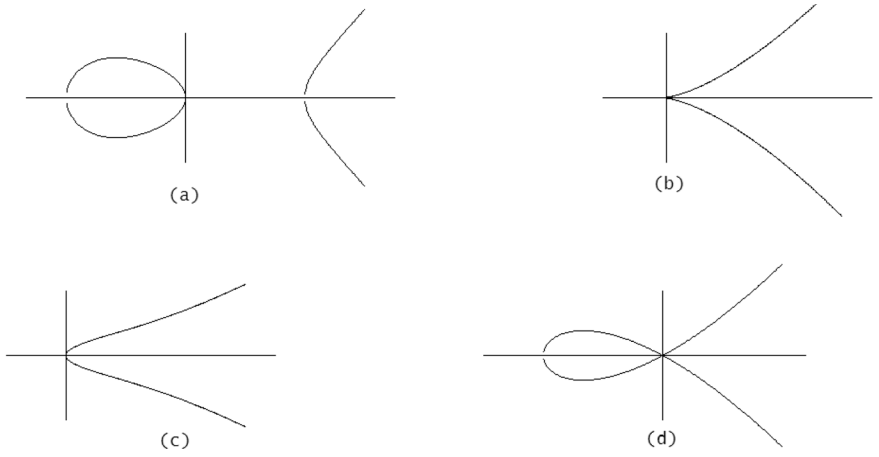

Figure 1. *Solution sets of (E) can assume a variety of shapes. The forms in (a) and (c) are deemed elliptic, those in (b) and (d) are not, due to the singularity (lack of a unique tangent) at the origin of coordinates.*

*A curve is called rational and cubic if it is the null set of a cubic polynomial in two variables with rational coefficients.

C associated with a particular equation (E) "elliptic" only if all three roots of $x^3 + Ax + B$ are distinct, forcing elliptic curves to have unique tangents at every point on the curve. A complete theory, however, requires consideration, as discussed below, of "singular cubics" for which two or more of the roots coincide. The null set of a polynomial $f(x,y)$ can fail to have a unique tangent only at singular points (x,y) at which $f_x = f_y = 0$.

If an elliptic curve C contains two ra-tional points P and Q, it must contain a third—perhaps at infinity—denoted $P * Q = Q * P$. Addition on C can then be defined as $P + Q = \infty * (P * Q)$. Under that operation, the rational points on an elliptic curve C known to contain a rational point ∞ can be shown to form an abelian group G^+, with ∞ as identity element. The English mathematician Louis Mordell showed in 1922 that G^+ is finitely generated.

Given a rational point P on an elliptic curve C, one can form P, $2P = P + P$, $3P = P + P + P$, etc. If any two points in the resulting sequence are equal, say $jP = kP$, then $mP = (k - j)P = \infty$, and P is said to be of "finite order" m. Some but not all elements of G^+ will ordinarily be of finite order, so that the entire group is generated by $P_1, \ldots, P_t, P_{t+1}, \ldots P_{t+r}$, where t is the number of generators of finite order, and r the number of generators of infinite order, known as the "rank" of G^+. A theorem discovered independently by Trygve Nagell (1935) and Elizabeth Lutz (1937) demonstrates that the rational points of finite order are in fact integer points, and makes it possible to calculate their generators from the prime factors of the discriminant $\Delta_E = -16(4A^3 + 27B^2)$ of (E).

Some forty years later, motivated by the fact that 19th-century mathematicians had identified elliptic curves containing rational points of order 2,3,4,5,6,7,8,9,10, and 12, but had never found any containing points of order 11, 13, or higher, Barry Mazur proved a far more intricate theorem: t is either 1 or 2, and the subgroup of G^+ consisting of elements of finite order—traditionally known as the "torsion subgroup" of G^+—is cyclic, of order 2,3,4,5,6,7,8,9,10, or 12, unless it is the direct product of a cyclic subgroup of order 2,4,6, or 8 and another of order 2. In consequence, the torsion subgroup of G^+ cannot contain more than 16 elements, and no longer seems particularly mysterious. It is much harder to say anything of consequence about the subgroup consisting of elements of infinite order. To date, the cryptographic applications involve only the finite elliptic curve groups, for which $r = 0$, but who knows what the future may hold?

From the Weak to the Strong Form of B&SD

The original (weak) form of the Birch and Swinnerton-Dyer conjecture identifies a possible shortcut to the determination of r, while the more recent strong form suggests a way in which the r generators of infinite order can actually be calculated. The technique begins with a remarkable generalization of Euler's famous identity

$$\Pi_p\,(1 - 1/p^s) = \zeta(s) = \Sigma_n\, 1/n^s,$$

in which the product extends over all primes p and the sum over all integers $n > 0$. The required generalization is

$$\Pi_{p\in S} 1/(1 - a_p p^{-s}) \cdot \Pi_{p\in S'}\, 1/(1 - a_p p^{-s} + p^{1-2s}) = L(E,s) = \Sigma_n\, a_n/n^s,$$

where the sum extends over all integers $n > 0$, while the sets S and S' over which the two products extend consist, respectively, of primes p that do and do not divide the discriminant Δ_E. The distinction is important because $E(\mathrm{mod}\ p)$ becomes singular when p divides Δ_E.

Because the coefficients a_n appearing on both sides of the identity depend multiplicatively on the index n, in the sense that $a_{mn} = a_m \cdot a_n$, it suffices to know the values a_p for primes p. If $p \in S'$, then $a_p = p + 1 - N_p$, where N_p is the number of elements of $F_p \times F_p$ that satisfy $E(\mathrm{mod}\ p)$, the restriction of (E) to F_p. If $p \in S$, a_p is either 0, 1, or –1, according to which of three mutually exclusive conditions prevails.

The function $L(E,s)$ so constructed is called the L-function associated with (E). Because its Dirichlet series expansion can be shown to converge absolutely in the right half $Re(s) > 3/2$ of the complex s-plane, $L(E,s)$ is analytic there and can be continued analytically—just like the Riemann zeta function—to the entire complex s-plane. Also like the Riemann zeta function, the function $\Lambda(E,s)$ derived from $L(E,s)$ by the formula $\Lambda(E,s) = (\sqrt{N}/2\pi)^s \Gamma(s) L(E,s)$ can be shown to satisfy the handy functional equation $\Lambda(E,s) = w\Lambda(E,2 - s)$, in which w is either 1 or –1.

Being analytic in the entire complex s-plane, $L(E,s)$ has a Taylor series expansion $c(s - 1)^\rho + d(s - 1)^{\rho+1} + \ldots$ about $s = 1$. The original (weak) form of the Birch and Swinnerton-Dyer conjecture asserts only that $\rho = r$, the rank of G^+, while the more recent strong form proposes a rather complicated algorithm for evaluating first the leading coefficient c and then the r infinite-order generators of G^+. In the final chapter of their book, Ash and Gross point out that the conjecture has been confirmed for $r = 0$ and $r = 1$, and go on to describe some of the computer-aided experiments that led Birch and Swinnerton-Dyer, around 1960, to formulate their conjecture. Those who remember what machine computation was like in the late 1950s will find that part of the story particularly impressive.

If there were nothing more to the story, Ash and Gross would have had no need to write a book. Their stated purpose is to explain the matter in a fashion understandable to a "mathematically inclined high school graduate." To that end, they devote the whole of Part I to explaining the conventions regarding multiple and complex roots of polynomials, homogeneous coordinates, and "points at infinity" needed to justify the conclusion that every cubic curve meets every straight line in exactly three points. Part II explores the anatomy of the group G^+ associated with an arbitrary cubic curve C, be it singular or nonsingular, before delving in Part III into enough complex function theory to make sense of $\zeta(s)$, $L(E,s)$, and $\Lambda(E,s)$, together with their power and Dirichlet series expansions, analytic continuations, and the functional equations they satisfy.

It would take an unusually ambitious high school student to get through the whole of the book in a single summer. Yet a student who received it as a graduation present, and returned to it in subsequent summers, could well have a rewarding experience.

One cannot help being impressed, in reading the book and pursuing a few of the references, by the magnitude of the enterprise it chronicles. From the 19th-century discovery of rational points of orders 2–10 and 12 on elliptic curves, to the 20th-century revelation of the structure of G^+, to the ongoing efforts to resolve the B&SD conjecture, the number, stature, and dedication of those involved are indeed awesome.

James Case writes from Baltimore, Maryland.