Review of

Elliptic curves, by Anthony W. Knapp

This book is about elliptic curves and modular functions, two topics that are intimately related in both accidental and essential ways. As emphasized by André Weil in his magisterial historical introduction to contemporary number theory [W], the arithmetic study of elliptic curves is, in spite of the clear reference to the integral calculus in the adjective *elliptic*, in many respects antecedent to the geometric and analytic study, and at least as important. There is still today hardly a domain of mathematics in which elliptic integrals or elliptic functions do not appear, and although we have far more questions about the arithmetic of elliptic curves than Fermat, we do not have many more answers.* Speculation about links between them and other subjects, appealing though it be, has nevertheless to be examined with circumspection, even when it is taken seriously.

Some of the more widely accepted speculation links diophantine analysis, in particular the diophantine analysis of elliptic curves, to the representation theory of reductive groups. Those of us who are attracted by such speculation run the risk of forgetting that number theory has to do ultimately with numbers. We are not alone, but that is no consolation. Thus the present author, an analyst and algebraist who has written an important introduction to the representation theory of reductive groups [K], is to be congratulated on his success with a text in which the number theory is primary and the representation theory completely suppressed.

As a complex-analytic object an elliptic curve is obtained by dividing the additive group $\mathbf{C}$ of the complex numbers by a discrete lattice L of rank two. The result is a compact Riemann surface S as well as a topological group. The $\mathfrak{p}$ functions of Weierstrass yield an isomorphism

$$z(\bmod L) \to (x, y) = (\mathfrak{p}(z), \mathfrak{p}'(z))$$

of S with the complex points $E(\mathbf{C})$ of a projective planar curve E defined by an equation of the form

$$y^2 = 4x^3 - g_2x - g_3, \tag{A}$$

* Hours after I finished writing this review the news that Andrew Wiles had proved Fermat's Last Theorem hit the front pages of the world's newspapers. I have resisted the very strong temptation to modify assertions and emphases that are now inappropriate. Since the Taniyama(-Weil) conjecture is a major element in his proof, there promises to be an enormous demand for discussions of it suitable to mathematicians with no great knowledge of the advanced theory of modular forms. The present book serves that purpose. The relation between Fermat's Theorem and the conjecture is briefly explained in its final paragraphs.

Appeared in Bulletin (New Series) of the A.M.S. 30, No. 1, (1994), pp. 96–100.

in which g_2 and g_3 are constants that depend on L. These curves as well as those obtained from them by rational transformations of the variables are the elliptic curves of the title.

For certain L the coefficients g_2 and g_3 or the coefficients of the transformed equation will be rational numbers, and it is such equations for which it is appropriate to ask whether rational solutions exist and to which the classical arithmetical theory applies. It is due to Fermat, and is referred to as *infinite descent*.

Since $E(\mathbf{C})$ is realized as $\mathbf{C}/L$, it is a group. The group law can be implemented by rational transformations with coefficients in the field generated by the coefficients of the equations defining E. Thus for the curves of greatest arithmetic interest, those with rational coefficients, the set $E(\mathbf{Q})$ of rational solutions of the equations is a group. For a curve in the form (A) such a solution is of the form

$$p = (x, y) = (X/Z, Y/Z),$$

where X, Y, and Z are integers whose greatest common divisor is 1. Clearly the height of p defined as $\max\{|X|, |Y|, |Z|\}$ is a fair measure of the size of (x, y) from the viewpoint of anyone trying to find rational solutions.

If the point p is doubled in the sense of the group law to obtain

$$q = 2 \circ p = (x', y') = (X'/Z', Y'/Z'),$$

then X', Y', and Z' will be quotients of homogeneous functions of degree larger than one in X, Y, and Z, so that intuition suggests that the height of q will be substantially larger than that of p. Division by 2 is a more delicate matter than multiplication because it may entail the introduction of irrationalities, but will have the advantage that it tends to decrease the height, permitting (upon iteration) the search for rational solutions to be confined within a limited range. This is the essence of the method of descent, summarized more clearly on the first page of Knapp's Chapter 4 or in [W, Appendix IV to Chapter II].

The method finds its most formal expression in the celebrated theorem of Mordell which asserts that the group $E(\mathbf{Q})$ is finitely generated, but nothing more, so that the details of the structure of $E(\mathbf{Q})$ remain open. According to a theorem of B. Mazur there are only a few possibilities for the torsion subgroup. Although Knapp discusses at some length the techniques used to analyze it in particular cases, his principal concern is with the rank of $E(\mathbf{Q})$.

The rank is understood only conjecturally and only in terms of objects of a much less elementary nature than the algebraic operations on elliptic curves: the L-functions attached to the curves. The notion of an L-function, of which the Riemann zeta function is the first example, has two different, but closely related sources: the distribution of primes and the study of congruences. As Artin observed, the similarity between function fields over finite fields and number fields suggests the introduction of congruence zeta functions, attached to curves over finite fields. Artin was prudent; he confined himself to planar curves defined by an equation quadratic in the ordinate, observing that, in contrast to the classical functions, the congruence zeta function was rational, but that the analogue of the Riemann hypothesis appeared to be true. That this was so was confirmed in 1936 by Hasse for elliptic curves, and more generally by Weil a few years later.

The proof for elliptic curves, although not easy, is elementary. Knapp, although he draws up short before proofs that are technically too advanced, skimps neither on examples nor on accessible proofs. There is also ample background and transitional material, presented in a way especially useful to readers accustomed to supplementing their mathematical reading with independent computational asides.

The major transition is from the method of descent, which takes up the first third of the book, to L-functions and the conjecture of Birch-Swinnerton-Dyer for the rank, as well as the conjecture attached to the names of Taniyama and Weil that relates elliptic curves to modular functions.

The author begins his discussion of L-functions close to their beginning, with Dirichlet L-functions and their applications to the study of primes in arithmetic progressions, but his primary interest is with the L-functions associated by Hecke to modular forms. Hecke has influenced the theory of L-functions in two quite different ways. First of all, he provided the proof of the analytic continuation of the L-functions associated to general number fields, as well as the first general definition of such functions. Secondly, he also showed how to attach to modular forms L-functions that were defined by Euler products and that possessed an analytic continuation and a functional equation. He proved this by a method that appears to be quite different from that used for the L-functions associated to number fields. The earlier method (modified by Tate and Godement-Jacquet) is, however, now also the customary method for dealing with the second type of function. There is some value in stressing this, since it reveals the very important structural similarities between Dirichlet characters and modular forms; but to do so would have changed the nature of the book, which, in spite of the wealth of material, does without a great deal of formal baggage.

The congruence zeta function of an elliptic curve has the form

$$\frac{(1-\alpha_1 t)(1-\alpha_2 t)}{(1-t)(1-pt)}.$$

The Riemann hypothesis for such functions that was proved by Hasse asserts that $|\alpha_1| = |\alpha_2| = p^{1/2}$. If the E curve is defined over **Q** then, as Knapp explains, a global L-function can be formed by taking the product of the local numerators,

$$L(s,E) = \prod_p{}' \frac{1}{(1-\alpha_1 p^{-s})(1-\alpha_2 p^{-s})}.$$

It is natural to suppose that this function, which converges in a half plane, can be analytically continued to the entire plane. The first, crude form of the conjecture of Birch-Swinnerton-Dyer — and this is the only form that is pertinent to Knapp's book — is that the order of vanishing of $L(s,E)$ at $s=1$ is the rank of $E(\mathbf{Q})$. This is a marvelous but unproved conjecture that renders the analytic continuation of $L(s,E)$ even more important. The conjecture of Taniyama-Weil replaces the still conjectural analytic continuation by another affirmation that has the disadvantage of being yet more difficult but the advantage that it is susceptible of thorough numerical testing.

The theory of modular functions and modular forms, defined on the upper half-plane $\mathcal{H}$ and subject to appropriate tranformation laws with respect to the group $\Gamma = SL(2,\mathbf{Z})$ of fractional linear transformations, is closely related to the theory of elliptic curves, because the family of all isomorphism classes of elliptic curves over **C** can be parametrized by the quotient $\Gamma\backslash\mathcal{H}$. This is an important, although formal, relation that assures that this and related quotients have a natural structure as algebraic curves X over **Q**. The relation between these curves and elliptic curves predicted by the Taniyama-Weil conjecture is, on the other hand, far from formal.

Since these curves can be defined over **Q**, it is also possible to attach to them both congruence zeta-functions and the analogue $L(s,X)$ of the global L-function $L(s,E)$. Thanks to the contributions of Eichler and Shimura, the functions $L(s,X)$ are, in contrast to $L(s,E)$, well understood. They have the form

$$L(s,X) = \prod_i L(s,f_i),$$

where each of the functions $L(s,f_i)$ is one of the Euler products attached by Hecke to a modular form, and thus can be analytically continued. In particular the order of vanishing of $L(s,X)$ at $s=1$ is well defined.

The Taniyama-Weil conjecture predicts a similar relation for elliptic curves over **Q**, namely $L(s,E) = L(s,f)$, for a felicitous choice of f. Even in this form the conjecture is of great appeal, for it permits the function $L(s,E)$ to be analytically continued. There is a similar conjecture for the Artin L-functions associated to tetrahedral, octahedral, and icosahedral representations. It is also very important, and in part established, but it does not have such concrete arithmetical consequences as that of Taniyama-Weil, nor is it part of a theory with such an ancient tradition. It can be also be tested numerically [B], but not yet so readily [C]. Moreover the theory of Eichler-Shimura and of the Hecke operators acting on the curves X, and on their integrals, to which the last third of the book is devoted provides a rich, and relatively concrete, conceptual and computational context in which the Taniyama-Weil conjecture can be better formulated and more easily understood and appreciated by a broad spectrum of mathematicians.

Knapp's *Elliptic Curves* is not the book from which to learn everything about elliptic curves. The deeper parts of the arithmetic theory, involving complex multiplication and cohomology, are absent; so is the more elaborate analytic part, involving theta functions or Jacobi elliptic functions. There is, nonetheless, a great deal of material that is presented carefully and is fun to read, and most of the basic techniques and open problems are there. Occasionally a word or two of further explanation would have made it easier for the reader to find his way through an argument, but such omissions are rare, and the author has promised to rectify them. The book can be recommended to students and to experienced mathematicians. There are few of us, even in closely related fields, who will not learn something from it.

References

[A] E. Artin, *Quadratische Körper im Gebiet der höheren Kongruenzen,* I, II, Math. Z. 19 (1924) 153–246.

[B] J. Buhler, *Icosahedral Galois representations,* Springer Lecture Notes in Math., vol. 654, Springer-Verlag, New York, 1978.

[C] J.E. Cremona, *Algorithms for modular elliptic curves,* Cambridge Univ. Press, London and New York, 1992.

[K] Anthony W. Knapp, *Representation theory of semisimple groups: An overview based on examples,* Princeton Univ. Press, Princeton, NJ, 1986.

[W] André Weil, *Number Theory, an approach through history,* Birkhäuser, Basel, 1984.