Heidi Lie Williams

14 November 2002

"An application of elliptic curves to the problem of factorization."

The notes for this lecture have been compiled from the following sources:

- I. Blake, G. Seroussi, and N. Smart, *Elliptic Curves in Cryptography*, Cambridge University Press, 2002.
- J. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, 1986.
- J. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, 1992.
- J. Silverman, *A Friendly Introduction to Number Theory*, Prentice Hall, 1996.
- A. Wiles, "The Birch and Swinnerton-Dyer Conjecture," *Collected Papers*.

First, to finish covering our final topic from last week's lecture:

**An overview of the Birch-Swinnerton-Dyer conjecture.**

- Deemed by the Clay Mathematics Institute to be one of the seven most important open problems in mathematics, the Birch-Swinnerton-Dyer conjecture can serve as a nice way to end our introduction to the theory of elliptic curves.
- This problem hinges on the mathematical properties of elliptic curves and, in particular, the set of points that are solutions to a given elliptic curve. Formulated in the 1960's, it concerns the set of rational points on an elliptic curve. Associated with each elliptic curve is a mathematical object called an *L*-function, which is a formula that encodes information about the curve in a different form. Informally, the Birch-Swinnerton-Dyer conjecture states that there exist an infinite number of rational points if and only if the curve's *L*-function equals zero at a certain value.
- Since the 1960's, the theory of *L*-functions of elliptic curves has been developed by a number of authors, but the Birch-Swinnerton-Dyer conjecture was the first link between the *L*-function and the set of rational points. The conjecture was developed from experimental evidence found computationally at Cambridge.
- Let us examine the Birch-Swinnerton-Dyer a bit more formally. Given any curve $C_0$, define the set of rational points of the curve to be $C_0(\mathbb{Q})$. Mordell conjectured, and in 1983 Faltings proved the following deep result.
- **<u>Theorem</u>**. If the genus of $C_0$ is $\geq 2$, then $C_0(\mathbb{Q})$ is finite.

- The most elusive case of investigating $C_0(\mathbb{Q})$ is for curves of genus 1. There may or may not be rational solutions and no method is known for determining which is the case for any given elliptic curve. Moreover, when there exist rational solutions there may or not be infinitely many.
- Recall from last time that Mordell's Theorem told us that the group of all rational points on an elliptic curve over $\mathbb{Q}$ is finitely generated.
- A principle fact of elementary group theory is that any finitely generated Abelian group is the direct sum of a finite group and a finite number of cyclic groups (each isomorphic to $\mathbb{Z}$, the set of integers). If we denote the set of rational points of the curve to be $C_0(\mathbb{Q})$, then

  $$C_0(\mathbb{Q}) = \mathbb{Z}^r \oplus [C_0(\mathbb{Q})]^t,$$

  where the number of copies of $\mathbb{Z}$ is equal to the rank of $C_0(\mathbb{Q})$ (although we will not discuss it in-depth in this lecture, it is worth noting that the rank of an elliptic curve is a vary important invariant of the curve) and the subgroup $[C_0(\mathbb{Q})]^t$ is the set of torsion points of the curve (note that this is the same torsion subgroup that we discussed during last week's lecture, when we presented the results of Nagell, Lutz, and Mazur). It is interesting to note that this torsion subgroup is relatively easy to compute both in theory and in practice. The rank, on the other hand, is very mysterious. There are numerous important open questions about $C_0(\mathbb{Q})$, such as how large the rank *r* of an elliptic curve can be, or whether there exists an effective algorithm for computing *r* for a given elliptic curve. There exist numerous fascinating conjectures concerning the rank of elliptic curves.
- The rank of a "randomly chosen" elliptic curve over $\mathbb{Q}$ tends to be fairly small, and it is quite difficult to produce such curves of even moderately high rank. Nevertheless, there exists the following "folklore" conjecture:
- **Conjecture.** There exist elliptic curves over $\mathbb{Q}$ with arbitrarily large rank.
- The principle evidence for this conjecture comes from the work of Tate, who showed the analogous result to be true for function fields (i.e. $\mathbb{Q}$ replaced by the field of rational functions).
- Neron has constructed an infinite family of elliptic curves over $\mathbb{Q}$ with rank at least 10.
- Mestre has produced examples with rank at least 12 over $\mathbb{Q}$, and his ideas can be used to produce curves of even higher rank (although his ideas, unfortunately, do not seem well-suited to producing infinite families of such curves).

- Using our definitions defined previously: Let *C* be an elliptic curve over $\mathbb{Q}$ in Weierstrass form $C: y^2 = x^3 + ax^2 + bx + c$ such that $a, b, c \in \mathbb{Q}$, $\Delta$ the discriminant, $N_p$ the number of solutions (mod *p*), and $a_p = p - N_p$. Then we can define the <u>incomplete *L*-series</u> of *C* to be

$$L(C,s) := \prod_{p \nmid 2\Delta} \left(1 - a_p p^{-s} + p^{1-2s}\right)^{-1}$$

  where *L* is a function of the complex variable *s*.
- It has been proven that *L*(*C*, *s*) has a holomorphic continuation as a function of *s* to the whole complex plane.
- The *L*-function of an elliptic curve has a definition based on details about a series of groups connected with the curve, namely those by arise by considering the curve over various finite fields. In other words, we can describe the *L* – series as a generating function which records information about the reduction of an elliptic curve modulo every prime. It is interesting to note that there exist intimate relations, both known and conjectural, between *L* – series and the theory of modular forms.
- We can now state the Millennium prize problem as follows.
- <u>**Birch-Swinnerton-Dyer Conjecture.**</u> The Taylor expansion of *L*(*C*, *s*) at *s* =1 has the form

$$L(C,s) = c(s-1)^r + \text{ higher order terms,}$$

  where $c \neq 0$ and *r* = rank ($C(\mathbb{Q})$). In particular, *L*(*C*, 1) = 0 if and only if $C(\mathbb{Q})$ is infinite.
- Note that a proof of this conjecture in a stronger form (using what are called <u>completed *L*-series</u>) would give an effective means of finding generators for the group of rational points.
- In 1976, Coates and Wiles showed a special case of the Birch-Swinnerton-Dyer conjecture by proving that elliptic curves with complex multiplication having an infinite number of solutions have *L* – series which are zero at the relevant fixed point, but they were unable to prove the converse.
- It is hoped that the proof of the Birch-Swinnerton-Dyer conjecture will offer insight into the general problem of investigating rational points on elliptic curves.

**An application of elliptic curves to the problem of factorization.**

- Elliptic curves have for a long time formed a central topic in several branches in theoretical mathematics; now the arithmetic of elliptic curves has practical applications as well.
- Applications of elliptic curves range from a proof of Fermat's last theorem to the design of secure cryptosystems.

- As an interesting side note: H. W. Lenstra of Berkeley recently presented a novel application of elliptic curves: a mathematical analysis of Escher's lithography 'Print Gallery.' Lenstra was reportedly struck by the blank spot in Escher's print. Although Escher only had a high-school mathematics background, he drew the picture by brilliant and methodical intuition, and the mathematical machinery underlying the image turned out to be elliptic curves. Escher's goal is reported to have been to "create a cyclic budge having neither beginning nor end." Lenstra searched for an exact mathematical formula for the repetitive pattern in order to find a "recipe" for filling the missing spot. Using conformal maps, the goal was achieved: Lenstra and his colleagues were able to generate several breathtaking possible completions for the missing space. The New York Times quoted Lenstra saying that he finds it impossible to put himself inside Escher's mind, but rather that he "finds it useful to identify Escher with nature." He claims to be a "physicist who tries to model nature."
- In cryptography, elliptic curves have three main applications: primality testing, cryptosystems, and factorization methods. Why are there these applications in cryptography? Because elliptic curves over finite fields provide an inexhaustible supply of finite Abelian groups which, even when large, are amenable to computation because of their rich structure. Such applications are the first use of elliptic curves, which are among the most richly structured and intensively studied objects in modern number theory and algebraic geometry.
- A key spark that set off the trend of increasing interest in elliptic curves on the part of cryptographers was the ingenious use of elliptic curves by H.W. Lenstra of Berkeley to obtain a new factorization method that in many respects is better than earlier known ones. The main advantages of Lenstra's method are that it is useful in combination with other factorization methods and that it (unlike its competitors) has a small storage requirement.
- We will begin by describing a factorization method due to Pollard. Pollard's so-called "p-1" method will not work for all integers *n*, but when it does in fact work it will be fairly efficient. Further, Pollard's method will serve as a prototype for the elliptic curve method we will later discuss.
- The idea underlying Pollard's algorithm is not difficult. Suppose that *n* happens to have a prime factor *p* such that $p-1$ is a product of "small" primes. Recall that from Fermat's Little Theorem, we know that if *p* does not divide *a*, then

$$\begin{aligned} & a^{p-1} \equiv 1 \pmod{p} \\ & \Rightarrow p \,\Big|\, \gcd\left(a^{p-1}-1, n\right) \end{aligned}.$$

Of course, at the start we do not know what *p* is, so we cannot compute $a^{p-1}-1$. Instead, we choose an integer

$$k = 2^{e_2}\cdot 3^{e_3}\cdot 5^{e_5}\cdot\ldots\cdot r^{e_r}$$

where *2…r* are the first few primes and $e_1,\ldots,e_r$ are "small" positive integers. Then we can compute $\gcd\left(a^k-1,n\right)$. If we are lucky in that *n* has a prime factor such that

$$p-1\,|\,k$$

then we have

$$p\,|\,a^k-1.$$

In this case, $\gcd\left(a^k-1,n\right)$ is greater than or equal to *p*, which is greater than 1. If $\gcd\left(a^k-1,n\right)$ is not equal to *n*, then this gcd provides a nontrivial factor of *n*. Then can we continue this process by factoring *n* into two pieces, factoring each of these, etc. On the other hand, if $\gcd\left(a^k-1,n\right)$ equals *n*, then we can choose a new *a* and restart. If $\gcd\left(a^k-1,n\right)$ is equal to 1, we can choose a larger *k*. Thus, the main idea is to compute $\gcd\left(a^k-1,n\right)$.

- **<u>Example.</u>** To illustrate Pollard's *p* – 1 method, let us try to factor *n* = 246082373. First, we must check to see if *n* is prime. Some computational work can show that

$$2^{n-1}\left(\operatorname{mod} n\right)\neq 1$$

therefore *n* must be composite. Let us choose

$$\begin{aligned}&a=2\\&k=2^2\cdot 3^2\cdot 5=180=2^2+2^4+2^5+2^7\end{aligned}.$$

Therefore we need to compute $2^{2^i}\left(\operatorname{mod} n\right), 0\leq i\leq 7$. We can summarize our results in the following table:

| ***i*** | $2^{2^i}\left(\operatorname{mod} 246082373\right)$ |
|---|---|
| 0 | 2 |
| 1 | 4 |
| 2 | 16 |
| 3 | 256 |
| 4 | 65536 |
| 5 | 111566955 |

| | |
|---|---|
| 6 | 166204404 |
| 7 | 214344997 |

Using this table, we can compute the following:

$$2^{180} = 2^{2^2+2^4+2^5+2^7} = 16 \cdot 65536 \cdot 111566955 \cdot 214344997 = 121299227 \left(\operatorname{mod} n\right)$$

Then a calculation using the Euclidean Algorithm yields

$$\gcd\left(2^{180} - 1, n\right) = \gcd\left(121299227, 246082373\right) = 1$$

Therefore our method has failed, which means *n* has no prime factors *p* such that *p*-1 divides 180. To remedy, we can choose a larger *k*. Let us choose

$$k = 2^2 \square 3^2 \square 5 \square 7 = 2520 = 2^3 + 2^4 + 2^6 + 2^7 + 2^8 + 2^{11}.$$

Thus, we can extend our table as follows:

| ***i*** | $2^{2^i} \left(\operatorname{mod} 246082373\right)$ |
|---|---|
| 8 | 111354998 |
| 9 | 82087367 |
| 10 | 7262369 |
| 11 | 104815687 |

Now we are able to compute

$$2^{2560} = 2^{2^3+2^4+2^6+2^7+2^8+2^{11}} \equiv 101220672 \left(\operatorname{mod} n\right).$$

Again, the Euclidean Algorithm yields

$$\gcd\left(2^{2520} - 1, n\right) = 2521$$

and thus we have found a non-trivial factor of *n*. In fact, it turns out that we have completely factored *n*, as

$$n = 2521 \cdot 97613 .$$

- We can now state Pollard's *p* – 1 algorithm more formally.
- **Pollard's p – 1 Factorization Algorithm.** Let $n \geq 2$ be a composite integer for which we are able to find a factor.

**STEP 1**: Choose a small number *k* which is a product of "small" primes to "small" powers.

**STEP 2**: Choose an arbitrary integer *a* such that 1 < *a* < *n*.

**STEP 3:** Calculate $\gcd(a,n)$. If this gcd is greater than 1, then *a* is a nontrivial factor of *n*, which means we are done! If this gcd is equal to 1, go to **STEP 4**.

**STEP 4**: Calculate $D = \gcd\left(a^k - 1, n\right)$. If $1 < D < n$, then *D* is a non-trivial factor of *n*, which means we are done! If $D = 1$, go back to **STEP 1** and choose a larger *k*. If $D = n$, go back to **STEP 2** and choose a larger *a*.

- Note that this algorithm only works in a "reasonable" amount of time if *n* has a prime divisor *p* such that *p*-1 is a product of "small" primes to "small" powers.
- Now we are ready to describe Lenstra's idea for using elliptic curves to create an algorithm which (conjecturally) will not have this defect.
- Pollard's algorithm is based on the fact that the non-zero elements in $\mathbb{Z}/p\mathbb{Z}$ form a group $\left(\mathbb{Z}/p\mathbb{Z}\right)^{\times}$ of order *p*-1. Thus, if *p*-1 divides *k*, then $a^k = 1$ in this group.
- Lenstra's idea is to replace the group $\left(\mathbb{Z}/p\mathbb{Z}\right)^{\times}$ by the group of points on an elliptic curve *C* over $\mathbb{F}_p$, and to replace the integer *a* by a point $P \in C\left(\mathbb{F}_p\right)$.
- As in Pollard's algorithm, we choose an integer *k* composed of a product of small primes. Then, if the number of elements in $C\left(\mathbb{F}_p\right)$ divides *k*, we will have $kP = O$, the point at infinity on $C\left(\mathbb{F}_p\right)$. The fact that $kP = O$ will allow us to find a non-trivial factor of *n*.
- What is the advantage of Lenstra's algorithm? If we choose only one curve *C* with integer coefficients and consider its reductions modulo various primes, then there is no advantage. For a single curve *C*, we win if there exists some *p* which divides *n* such that the number of elements in $C\left(\mathbb{F}_p\right)$ is a product of small primes. Similarly, we win using Pollard's algorithm if there exists some *p* which divides *n* such that the *p*-1 is a product of small primes. But, suppose we don't "win." Using Pollard's *p*-1 algorithm, we have hit a dead end. But with Lenstra's algorithm, we can simply choose a new elliptic curve and start over, and it turns out that the odds of eventually having success with this method are quite good.
- We can now state these vague ideas more explicitly.
- **<u>Lenstra's Elliptic Curve Factorization Algorithm.</u>** Let $n \geq 2$ be a composite integer for which we are able to find a factor.

**STEP 1**: Check that $\gcd(n,6)=1$ and that $n \neq m^r$ for some $r \geq 2$ (i.e. check that *n* is not a perfect power).

**STEP 2:** Choose random integers $b, x_1, y_1$ between 1 and *n*.

**STEP 3:** Let $c = y_1^2 - x_1^3 - bx_1 \pmod{n}$, and let *C* be the curve

$$C : y^2 = x^3 + bx + c$$
$$P = (x_1, y_1) \in C$$

**STEP 4:** Check $\gcd\left(4b^3 + 27c^2, n\right) = 1$. If this gcd = *n*, choose a new *b*. If this gcd is greater than 1 but less than *n*, then this is a non-trivial factor of *n*, which means we are done!

**STEP 5:** Choose a number *k* which is a product of small primes to small powers.

**STEP 6:** Compute

$$kP = \left( \frac{a_k}{d_k^2}, \frac{b_k}{d_k^3} \right).$$

**STEP 7:** Calculate $D = \gcd(d_k, n)$. If 1 < *D* < *n*, then *D* is a non-trivial factor of *n*, which means we are done! If *D*=1, go back to **STEP 5** and increase *k* or choose a new curve. If *D*=*n*, go back to **STEP 5** and decrease *k*.

- Why does Lenstra's algorithm work? In the limited time we have it is difficult to explain how the actual factorization is discovered.
- Suppose we are lucky enough to choose a curve *C* and a number *k* such that for some prime *p* such that *p* divides *n*, we have that the number of elements in $C\left(\mathbb{F}_p\right)$ divides *k*. Then every element in $C\left(\mathbb{F}_p\right)$ has order dividing *k*, so in particular if we take the point *P* in $C\left(\mathbb{F}_p\right)$ and reduce it module *p*, we will know $\overset{\square}{k}P = k\overset{\square}{P} = \mathrm{O}$, the point at infinity, i.e. the reduction of *kP* modulo *p* is the point at infinity. Therefore *p* divides $d_k$, so *p* divides $\gcd\left(d_k, n\right)$. The fundamental relevance of elliptic curves for factorization is the fact that if we have a number *n* to be factored, we can work with an elliptic curve over $\square_n$, even though $\square_n$ is not a field and treating it as such might be considered "illegal." When an illegal curve operation is encountered, it is exploited to find a factor of *n*.
- Lenstra's Elliptic Curve Method (ECM) is the most successful factoring algorithm. Why? For numbers of the form *N*=*pq*, where *p, q* are primes roughly of the same size, the elliptic curve method is inefficient compared to the quadratic or number field sieve methods. However, relatively few numbers of that form, so for a random integer of around one hundred digits ECM should find the

prime factors before other factoring methods. However, in cryptography one is usually interested in the types of numbers for which *p, q* are roughly the same size. Hence it could appear that the uses of ECM in cryptography are very limited; this, however, is not the case. The large prime variations of both the quadratic and number field sieve algorithms require the factorization of numbers as a sub-procedure, and it is at this stage that the ECM method can be applied most successfully.

- An interesting problem one may address after having found a factor via an ECM scheme is: what is the actual group order that allowed the factor discovery? One approach, which has been used by Richard Brent at Oxford, is to simply "backtrack" in the algorithm until the precise largest- and second-largest primes are found, and so on until the curve is completely factored. However, there may be a simpler, more elegant method to obtain via an algorithmic method the actual order of the curve. To this end, I will be working to utilize algorithmic methods and point-counting algorithms such as Schoof's to attempt to answer this question.