MATH 104C NUMBER THEORY: NOTES

Hans Wenzl

1. DUPLICATION FORMULA AND POINTS OF ORDER THREE

We recall a number of useful formulas. If $P_i = (x_i, y_i)$ are the points of intersection of a line with the elliptic curve $E : y^2 = f(x) = x^3 + ax^2 + bx + c$, then we have

$$x_1 + x_2 + x_3 = m^2 - a, \tag{1}$$

where m is the slope of the line. This is the most efficient way to calculate the x coordinate of the point $P_1 + P_2$. Now observe if the line is the tangent line at the point $P = (x, y)$, we get

$$m^2 = \frac{f'(x)^2}{(2y)^2} = \frac{f'(x)^2}{4f(x)},$$

using $y^2 = f(x)$. Hence we get the *duplication formula* for the x-coordinate $x(2P)$ of the point $2P$, by setting $x_1 = x_2 = x$ and solving for $x_3 = x(2P)$

$$x(2P) = \frac{f'(x)^2}{4f(x)} - a - 2x = \frac{x^4 - 2bx^2 - 8cx + b^2 - 4ac}{4x^3 + 4ax^2 + 4bx + 4c}.$$

Points of order three As we have seen in the lecture by geometric considerations, if a point $P = (x, y)$ has order three, then $2P = -P$ and $x(2P) = x$. Plugging this into the duplication formula, multiplying by the denominator we get the polynomial identity

$$0 = \Psi_3(x) = 2f(x)f''(x) - f'(x)^2 = 3x^4 + 4ax^3 + 6bx^2 + 12cx + (4ac - b^2).$$

Hence a point $P = (x, y)$ can have order 3 only if its x-coordinate satisfies $\Psi_3(x) = 0$.

Lemma If the discriminant $\Delta = \Delta(f)$ is not equal to 0, then Ψ_3 has four distinct roots.

By the Repeated Root Theorem (see below), it suffices to show that Ψ_3 and Ψ_3' have no common zeros. Now observe that

$$\Psi_3'(x) = 2f'(x)f''(x) + 2f(x)f'''(x) - 2f'(x)f''(x) = 2f(x)f'''(x) = 12f(x),$$

where the last equality follows from $f'''(x) = 6$ (as f(x) is a polynomial of order 3 with leading coefficient 1). Assume now there exists a point x_o for which $\Psi_3(x_o) = 0 - \Psi_3'(x_o)$. Then the last equality implies that $12f(x_o) = 0$, i.e. $f(x_o) = 0$. Plugging this into the formula for $\Psi_3(x)$, we obtain

$$0 = \Psi_3(x_o) = 2f(x_o)f''(x_o) - f'(x_o)^2 = -f'(x_o)^2.$$

Hence x_o would be a common root of $f(x)$ and $f'(x)$ as well, contradicting the fact that the discriminant of $f(x)$ is nonzero.

Repeated Root Theorem The following statements are equivalent:

(a) The polynomial $f(x)$ of degree n has n distinct roots.

(b) The polynomials $f(x)$ and $f'(x)$ have no common root.

(c) The discriminant of the polynomial $f(x)$ is nonzero

Recall that for a polynomial $f(x) = x^3 + bx + c$, the discriminant is given by $\Delta(f) = -4b^3 - 27c^2$. If $f(x) = x^2 + bx + c$, then $\Delta(f) = b^2 - 4c$.

Theorem Let $E : y^2 = f(x) = x^3 + ax^2 + bx + c$ be an elliptic curve with $\Delta(f) \neq 0$. Then there exist exactly eight points P in $E(\mathbf{C})$, the set of complex solutions of E of order 3. In particular, these points together with ∞ form a group which is isomorphic to $\mathbf{Z}/3 \times \mathbf{Z}/3$.

Proof. We have shown that if $P = (x, y)$ has order 3, then its x-coordinate must be one of the four distinct roots of the polynomial $\Psi_3(x)$. For each of these values, we have two values $\pm y$ which satisfy E. If P_1 and P_2 are two points of order three, then we also have $3(P_1 + P_2) = 3P_1 + 3P_2 = \infty$, i.e. also the order of their sum must divide 3 and hence must again be one of the points of order 3 or the point ∞. Hence we get a group of nine elements all of which have order at most 3. It is shown in algebra that this group must be isomorphic to $\mathbf{Z}/3 \times \mathbf{Z}/3$.

Exercises:

1. Let $E : y^2 = x^3 + 6x$ with $b > 0$. Calculate all *real* points of order 3. Are there any rational points of order 3?
2. Let $E : y^2 = x^3 - bx$, with $b > 0$.
   (a) Calculate all *real* points of order 4. *Hint* : If P has order 4, then $2P$ has order 2.
   (b) Calculate all rational points of order four on E.
3. Let $E : y^2 = f(x) = x^3 + ax^2 + bx = c$ be an elliptic curve with nonzero discriminant. The goal of this exercise is to show that there are at most two *real* points of order 3 on E.
   (a) Show that its real points (i.e. what we draw on paper) have either one or two connected components, depending on whether $f(x)$ has one or three real solutions.
   (b) Assume that we have two components, and that P_1, P_2 are points in the component C_1 which does not contain the point ∞. Show that $P_1 + P_2$ is in C_0, the component which contains ∞. *Hint* : You may want to consider some geometric argument.
   (c) Show that if P has odd order, then it must be in C_0.
   (d) Show that the polynomial $\Psi_3(x)$ has only one real root α for which $f(\alpha) > 0$ (*Hint* : Use that $\Psi_3' = 12f(x)$).
   (e) Prove in general that there are exactly two real points on E which have order three.

2. MODULAR DEFECT FOR BAD PRIMES

Recall that a prime p is called a bad prime for the elliptic curve $E : y^2 = f(x) = x^3 + ax^2 + bx + c$ if $p|\Delta$, the discriminant of E. Considering the curve E mod p, it has zero discriminant (mod p) if p is a bad prime. One can show as in the previous section that in this case $f(x)$ has repeated roots mod p. We have the following theorem:

Theorem Assume that p is a bad prime for the elliptic curve E.

(a) If p is *very bad*, i.e. we have $y^2 = (x-d)^3$ mod p for some d, then $a_p = 0$ where $a_p = p - N_p$ is the modular defect.

(b) If p is *fairly bad*, i.e. $y^2 = (x-d)^2(x-e)$ mod p, then $a_p = 1$ if $d-e$ is a quadratic residue mod p, and $a_p = -1$ if $d-e$ is not a quadratic residue mod p.

Proof. We first observe that the number of solutions mod p does not change if we make the substitution $x \mapsto x + d$. Hence it suffices to prove claim (a) for $y^2 = x^3$. The crucial observation is that a is a quadratic residue mod p if and only if a^3 is a quadratic residue mod p. Indeed, it is obvious to see that if $a = b^2$ mod p is a quadratic residue, then so is $a^3 = (b^3)^2$ mod p. On the other hand, if $a^3 = c^2$ mod p is a quadratic residue, then we also have $(a^{-1}c)^2 = a^{-2}c^2 = a^{-2}a^3 = a$ mod p, hence a is a quadratic residue (This can also be done more elegantly using the Legendre symbol). Hence we get two solutions for $y^2 = x^3$ if $x = a$ is a quadratic residue, one solution if $x = 0$ and no solution if $x = a$ is not a quadratic residue. Adding up we get $2\frac{p-1}{2} + 1 + 0 = p$ solutions mod p, i.e. $N_p = p$ and $a_p = 0$.

By the same transformation as for (a), it suffices to consider the equation $y^2 = x^2(x + d - e)$. We try to use the same strategy as for part (a). Now if $a \neq 0$, the equation $y^2 = a^2(a+d-e)$ has exactly as many solutions as the equation $y^2 = a+d-e$, depending on whether $a = d - e$ is zero or a quadratic (non-)residue (you should prove this for yourself, along the lines of what was done in the previous paragraph). However, if $a = 0$, the equation $y^2 = a^2(a + d - e) = 0$ has exactly one solution, while the equation $y^2 = a + d - e = d - e$ has either two solutions (if $d - e$ is a quadratic residue) or zero solutions (if $d-e$ is not a quadratic residue). Hence the equation $y^2 = x^2(x+d-e)$ has one solution less (or one solution more) than the equation $y^2 = x+d-e$ if $d-e$ is (or is not) a quadratic residue. Hence we have $N_p = p - 1$ or $N_p = p + 1$ in these two cases. The claim follows from this.

3. MODULAR FORMS, LATTICES AND FERMAT'S LAST THEOREM

1. Definitions and examples Let H_+ be the upper half plane of the complex plane, i.e. the set of all complex numbers $\tau = \alpha + i\beta \in \mathbf{C}$ for which $Im(\tau) = \beta > 0$. A differentiable function F on H_+ is called a *modular form* of weight $2k$ if

$$F(\frac{a\tau + b}{c\tau + d}) = (c\tau + d)^{2k} F(\tau)$$

for any integers a, b, c, d satisfying $ad - bc = 1$. It can be shown that any modular form F

can be written as a power series

$$F(\tau) = \sum_{n=-\infty}^{\infty} c_n q^n, \qquad \text{where } q = e^{2\pi i \tau}.$$

You should check for yourself that $Im(\tau) > 0$ implies $|q| < 1$.

Examples 1. Consider the function

$$G_{2k}(\tau) = \sum_{(n,m)\neq(0,0)} \frac{1}{(m\tau + n)^{2k}},$$

where the summation goes over all pairs of integers $(n, m) \neq (0, 0)$, and $k \geq 2$. It can be shown that these series do converge for all $\tau \in H_+$. It was shown in class that G_{2k} is a modular form of weight $2k$. It is known that all modular forms of weight < 12 are of this form (up to a scalar multiple).

2. Another important function, defined on H_+, is given by

$$\eta(\tau) = e^{\pi i \tau/12} \prod_{n=1}^{\infty} (1 - e^{2\pi i n \tau}) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n),$$

where we substituted $q = e^{2\pi i \tau}$. It can be shown that this infinite product does converge for any $\tau \in H_+$, i.e. for which its imaginary part is positive. It can also be shown that η^{24} is a modular form of weight 12 which is not equal to G_{12}.

Exercises

1. Let F be a modular form of rank k. Express $F(\tau + 1)$ and $F(-1/\tau)$ in terms of $F(\tau)$.
2. (optional). Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a matrix with integer coefficients and with $|det(A)| > 1$. Show that the map $\mathbf{x} \mapsto A\mathbf{x}$ maps the set $\mathbf{Z}^2$ of vectors with integer coefficients into $\mathbf{Z}^2$, but that it is not onto. *Hint* : Show first that A^{-1} has some coefficients which are not integers. Then find a vector $\mathbf{y} \in \mathbf{Z}^2$ such that $A^{-1}\mathbf{y}$ is NOT in $\mathbf{Z}^2$.
3. Do problem 43.4 (a) and (c) of the Friendly Introduction. I plan to do 43.4(b) in class. Here are the problems:
   (a) Calculate $a_p = p - N_p$ for $y^2 = x^3 + p$.
   (c) Same for $y^2 = x^3 - x^2 + p$. You should check in both cases that p is a bad prime, i.e. p divides the discriminant of the corresponding elliptic equations.

If you have spare time and energy, you could already start looking into Exercises 44.1 and 44.2 of the Friendly Introduction.

2. Modular forms are functions on lattices We are going to show here how the perhaps obscure looking definition of modular forms comes from the desire to define functions on lattices in a manageable way. Let $L = L(\omega_1, \omega_2) = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$ be a lattice in $\mathbf{C}$. To

define a function on the set of all lattices is complicated. Instead, we associate to a lattice a number by the formula

$$L(\omega_1, \omega_2) \quad \mapsto \ \tau \ = \ \omega_1/\omega_2 \text{ or } \omega_2/\omega_1, \tag{2}$$

where we pick for τ the number such that τ has positive imaginary part. You may want to check for yourself that for any complex number z which is not real we either have $Im(z) > 0$ or $Im(z^{-1}) > 0$. There are two problems using this for defining functions for lattices:

(a) The map in (2) is not injective. Indeed, if you blow up the lattice by a number α, i.e. you replace the lattice $L(\omega_1, \omega_2)$ by $\alpha L(\omega_1, \omega_2) = L(\alpha\omega_1, \alpha\omega_2)$, both lattices will be mapped to the same number τ, as obviously $(\alpha\omega_1)/(\alpha\omega_2) = \omega_1/\omega_2$.

(b) The number τ depends on which basis vectors we use for the lattice L. If

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{with } \det(A) = ad - bc = 1,$$

we get a new basis $\omega_1' = a\omega_1 + b\omega_2$, $\omega_2' = c\omega_1 + d\omega_2$. We then obtain

$$\tau' = \frac{\omega_1'}{\omega_2'} = \frac{a\omega_1 + b\omega_2}{c\omega_1 + c\omega_2} = \frac{a\omega_1/\omega_2 + b}{c\omega_1/\omega_2 + d} = \frac{a\tau + b}{c\tau + d}.$$

To address these two problems, we define the lattice $L(\tau)$ to be equal to the lattice spanned by τ and 1, with τ in the upper half plane H_+ of the complex plane. Then it follows from the discussion above and these definitions that $L(\omega_1, \omega_2) = \omega_2 L(\omega_1/\omega_2)$. With these definitions, one can prove the following theorem:

Theorem Fix $k \in \mathbf{Z},\ k > 0$. There exists a $1-1$ correspondence between
- functions $\tilde{f}$ on lattices $L \subset \mathbf{C}$ satisfying $\tilde{f}(\alpha L) = \alpha^{-k}\tilde{f}(L)$ for any $\alpha \in \mathbf{C}$, and
- modular forms of weight k, i.e. functions f on H_+ satisfying $f(\frac{a\tau+b}{c\tau+d}) = (c\tau+d)^k f(\tau)$.

The correspondence is given by $\tilde{f}(L(\omega_1, \omega_2)) = \omega_2^{-k} f(\omega_1/\omega_2)$, where we assumed ω_1, ω_2 be taken in the order such that ω_1/ω_2 has positive imaginary part.

3. Hecke operators (*This section can be skipped if short on time*) Fix $N \in \mathbf{Z},\ N > 1$, and let L be a lattice in $\mathbf{C}$. For this section, you might as well assume that $L = L(i) = \mathbf{Z}i + \mathbf{Z}$. A sublattice $L' \subset L$ is just a subgroup of the abelian group $(L, +)$. We showed in class that any sublattice L' of index $[L : L'] = N$ contains the sublattice $NL = N\mathbf{Z}i + N\mathbf{Z}$. (*Proof* If L' is a subgroup of L with index N, then, by definition, the factor group L/L' has order N. So if $\omega \in L$, then, by Lagrange's Theorem, the order of the coset $\omega + L'$ in L/L' is divisible by N. Hence $N(\omega + L') = N\omega + L' = L'$, the identity element of L/L'. We conclude that $N\omega \in L'$. As ω was arbitrary, it follows that $NL \subset L'$.) Hence there are only finitely many sublattices $L' \subset L$ with $[L : L'] = N$, which correspond to certain subgroups of order N in the abelian group $L/NL \cong \mathbf{Z}_N \times \mathbf{Z}_N$.

We can now define the Hecke operator $T(N)$ on modular forms f as follows:

$$T(N)(f)(\tau) = \sum_{L', [L(\tau):L']=N} \tilde{f}(L');$$

here $\tilde{f}$ is the function on lattices corresponding to f by the theorem in the last section, and $L(\tau)$ is the lattice spanned by τ and 1. By the remark above, this is a finite sum.

Theorem (a) If f is a modular form of weight k, then so is $T(N)(f)$. We define $T_k(N)$ to be the restriction of $T(N)$ to the space M_k of modular forms of weight k.

(b) Modular operators satisfy $T_k(p)T_k(p^r) = T_k(p^{r+1}) + p^{k-1}T_k(p^{r-1})$, if p is prime, and $T(N)T(M) = T(NM)$ if $gcd(N, M) = 1$.

Recall that any modular form f can be written as a power series $f(\tau) = \sum_n c_n q^n$, where $q = e^{2\pi i\tau}$. We say that f is a *cusp form* if $f(\tau) = \sum_{n=1}^{\infty} c_n q^n$, with the c_ns being integers. Then we can also write $T(N)(f)$ as a power series, say as $\sum_{n=1}^{\infty} b_n q^n$. In the case $N = p$ prime, we have the following simple relationships between these power series:

$$b_n = \begin{cases} c_{pn} & \text{if } p \not| n, \\ c_{pn} + p^{k-1}c_{n/p} & \text{if } p|n. \end{cases}$$

4. Modular forms associated to elliptic curves We now want to connect the material above with the main object of our course, elliptic curves. Let $E : y^2 = x^3 + ax^2 + bx + c$ be an elliptic curve with $a, b, c \in \mathbf{Z}$, and with its discriminant $\Delta \neq 0$. Let, as usual, $a_p = p - N_p$, where N_p is the number of points on E mod p. We now assign to E a power series

$$f_E(\tau) = \sum_{n=1}^{\infty} c_n q^n, \quad \text{where } q = e^{2\pi i\tau},$$

as follows:

$$\begin{array}{rcll} c_1 & = & 1, & \\ c_p & = & a_p & \text{if } p \text{ is prime}, \\ c_{p^r} & = & (c_p)^r & \text{if } p|\Delta, \\ c_{p^{r+1}} & = & c_p c_p^r - p c_{p^{r-1}} & \text{if } p \not|\Delta, \\ c_{nm} & = & c_n c_m & \text{if } (n, m) = 1. \end{array}$$

These are exactly the patterns you have detected in the last homework exercise. It is *comparatively* easy to indeed prove that these patterns hold for our specific homework example. However, in general, one can not write down f_E via an explicit finite formula. Nevertheless, Taniyama (1956) conjectured the following, for which Weil also proved results providing additional evidence:

Taniyama-Weil Conjecture For every non-degenerate elliptic curve E the power series $f_E = \sum_{n=1}^{\infty} c_n q^n$ defines a modular form of weight 2 with respect to the group $\Gamma_0(N)$ for a suitable positive integer $N|\Delta$ (see below for a description of $\Gamma_0(N)$). This means that for $q = e^{2\pi i\tau}$ we have

$$f_E(\frac{a\tau + b}{c\tau + d}) = (c\tau + d)^2 f_E(\tau),$$

whenever $N|c$ and $ad - bc = 1$.

Here the group $\Gamma_0(N)$ consists of all matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ for which $ad - bc = 1$ and $N|c$. You should check for yourself that $\Gamma_0(N)$ is indeed a group, i.e. the inverse as well as the product of two matrices in $\Gamma_0(N)$ are as well in $\Gamma_0(N)$.

5. Connection to Fermat's Last Theorem It has already been proved by Fermat that the equation $x^n + y^n = z^n$ has no solution in nonzero integers if $n = 3, 4$. To prove it in general for all $n > 2$ it suffices to prove it for all primes $p > 2$ (indeed, if we had a counter example $A^n + B^n = C^n$ for some nonprime $n = mp > 2$, we would also get a counter example $(A^m)^p + (B^m)^p = (C^m)^p$ for the prime p; if $n = 2^k$ for some $k > 1$, we could similarly deduce a counter example for the already proven case $n = 4$). In the 80s, Frey suggested to consider the elliptic curve

$$E_{A,B} : y^2 = x(x - A^p)(x + B^p)$$

for a possible counter example $A^p + B^p = C^p$ of Fermat's Last Theorem. One calculates that this curve would have discriminant $A^{2p}B^{2p}(A^p + B^p)^2 = (ABC)^{2p}$. Having a perfect $2p$-th power as discriminant would be so unusual for an elliptic curve defined over the integers that he conjectured that such a curve could not exist. This turned out to be more complicated than thought. The following result provided more evidence for this conjecture:

Theorem(Ribet) The curve $E_{A,B}$ would be a counter example of the Taniyama-Weil conjecture.

Inspired by this, Andrew Wiles worked on proving the Taniyama-Weil conjecture for six years, and announced in 1993 that he could prove it for semisteable elliptic curves (these are curves which only have fairly bad primes, but no very bad primes - see the Theorem on bad primes in the second section). This is important as the curve $E_{A,B}$ would be semistable. A gap was discovered in Wiles' paper by a referee a few months after the paper was submitted. But Wiles managed to circumvent this gap a year later in joint work with Richard Taylor. Hence we have

Theorem (Wiles, Taylor-Wiles) The Taniyama-Weil conjecture holds for all semistable elliptic curves.

Corollary The equation $x^n + y^n = z^n$ has no solution in integers satisfying $xyz \neq 0$ for $n > 2$.