Stickelberger's theorem and Herbrand's theorem

January 30, 2014

Let's recall what we had at the end of the last lecture. Let ℓ be an odd prime, and let P be a maximal ideal of $\mathbb{Z}[\zeta_\ell]$ which doesn't lie above ℓ. Thus $\mathbb{F}_P := \mathbb{Z}[\zeta_\ell]/P$ is a finite field of size $q = p^d$, with $p \neq \ell$ and d equal to the order of p in $(\mathbb{Z}/\ell\mathbb{Z})^\times$. Consider the character $\chi : \mathbb{F}_P^\times \to \mathbb{Q}(\zeta_\ell)^\times$ defined by $\chi(a) = \zeta$ if and only if ζ is the unique ℓ^{th} root of unity which is congruent to $a^{-\frac{q-1}{\ell}}$ modulo P. This character gives the Gauss sum

$$G(\chi) = \sum_{a \in \mathbb{F}_P^\times} \chi(a)\zeta_p^{Tr(a)} \in \mathcal{O}_{\mathbb{Q}(\zeta_\ell,\zeta_p)},$$

where $Tr(a) = a + a^p + \ldots + a^{p^{d-1}} \in \mathbb{F}_p$. Recall also that there is a unique maximal ideal $\widetilde{P}$ of $\mathcal{O}_{\mathbb{Q}(\zeta_\ell,\zeta_p)}$ lying above P, namely $\widetilde{P} = (P, 1 - \zeta_p)$.

Then we saw that the unique maximal ideal factorization of $G(\chi)$ is

$$(G(\chi)) = \widetilde{P}^{(p-1)\sum_{a=1}^{\ell-1} \frac{a}{\ell}\widetilde{\sigma_a}^{-1}},$$

where $a \leftrightarrow \sigma_a$ is our notation for the isomorphism $(\mathbb{Z}/\ell\mathbb{Z})^\times \simeq Gal(\mathbb{Q}(\zeta_\ell)/\mathbb{Q})$ (so $\sigma_a(\zeta_\ell) = \zeta_\ell^a$), and $\widetilde{\sigma_a}$ denotes the extension of σ_a to $Gal(\mathbb{Q}(\zeta_\ell, \zeta_p)/\mathbb{Q})$ which fixes ζ_p.

Now, this factorization lives in $\mathcal{O}_{\mathbb{Q}(\zeta_\ell,\zeta_p)}$, but it comes very close to living in just $\mathcal{O}_{\mathbb{Q}(\zeta_\ell)}$. In fact, by applying certain simple linear combinations of Galois automorphisms we can make it live there. This is the observation of Stickelberger:

Theorem 0.1. *Let $x \in \mathbb{N}[Gal(\mathbb{Q}(\zeta_\ell)/\mathbb{Q})]$, say $x = \sum_{b=1}^{\ell-1} x_b \cdot \sigma_b$ with $x_b \in \mathbb{N}$. Then the following conditions are equivalent:*

1. *The element $x \cdot \sum_{a=1}^{\ell-1} \frac{a}{\ell}\sigma_a^{-1}$ lies in $\mathbb{N}[Gal(\mathbb{Q}(\zeta_\ell)/\mathbb{Q})]$ (so, the denominators cancel);*
2. *The sum $\sum_{b=1}^{\ell-1} x_b \cdot b$ is divisible by ℓ;*
3. *We have $\zeta_\ell^x = 1$.*

Furthermore, under these conditions $G(\chi)^{\widetilde{x}}$ lies in $\mathbb{Q}(\zeta_\ell)$ (for all choices of maximal ideal P of $\mathbb{Z}[\zeta_\ell]$ not lying above ℓ), and its maximal ideal factorization is

$$(G(\chi)^{\widetilde{x}}) = P^{x \cdot \sum_{a=1}^{\ell-1} \frac{a}{\ell}\sigma_a^{-1}}.$$

Proof. First let's see that 1 and 2 are equivalent. Certainly, 1 holds if and only if $x \cdot \sum_{a=1}^{\ell-1} a \cdot \sigma_a^{-1}$ is congruent to zero (mod ℓ). But working (mod ℓ), one easily calculates that the coefficient of σ_a^{-1} in that product is equal to $a \cdot \sum_{b=1}^{\ell-1} x_b \cdot b$. Since a is prime to ℓ, we see that all of these coefficients are 0 (mod ℓ) if and only if 2 holds. That was the claim.

Now let's see that 2 and 3 are equivalent. Actually this is clear: since $\zeta_\ell^{\sigma_b} = \zeta_\ell^b$ by definition, we have $\zeta_\ell^x = \zeta_\ell^{\sum_{b=1}^{\ell-1} x_b \cdot b}$. On the other hand $\zeta_\ell^n = \zeta_\ell^m$ if and only if n is congruent to m (mod ℓ).

Thus all the conditions are equivalent. Now let's assume 3, and see that $G(\chi)^{\widetilde{x}} \in \mathbb{Q}(\zeta_\ell)$. Since we know that $G(\chi)^{\widetilde{x}} \in \mathbb{Q}(\zeta_\ell, \zeta_p)$, we have $G(\chi)^{\widetilde{x}} \in \mathbb{Q}(\zeta_\ell)$ if and only if $G(\chi)^{\widetilde{x}}$ is fixed by the Galois group of $\mathbb{Q}(\zeta_\ell, \zeta_p)/\mathbb{Q}(\zeta_\ell)$. But this Galois group is $(\mathbb{Z}/p\mathbb{Z})^\times$, acting as usual on ζ_p and fixing ζ_ℓ. Let's write $c \leftrightarrow \tau_c$ for this identification $(\mathbb{Z}/p\mathbb{Z})^\times = Gal(\mathbb{Q}(\zeta_\ell, \zeta_p)/\mathbb{Q}(\zeta_\ell))$. Then we have

$$\tau_c G(\chi) = \sum_{a \in \mathbb{F}_P} \chi(a) \zeta_p^{c \cdot Tr(a)}.$$

But Tr is $\mathbb{F}_p$-linear, so $c \cdot Tr(a) = Tr(ca)$. But then $a \mapsto ca$ gives a bijection from $\mathbb{F}_P$ to itself, so reindexing the sum we find that

$$\tau_c G(\chi) = \chi(c)^{-1} G(\chi).$$

Since each τ_c evidently commutes with each $\widetilde{\sigma_a}$, it follows that

$$\tau_c G(\chi)^{\widetilde{x}} = \chi(c)^{-x} G(\chi)^{\widetilde{x}}.$$

Thus $G(\chi)^{\widetilde{x}}$ lies in $\mathbb{Q}(\zeta_\ell)$ if and only if $\chi(c)^{-x} = 1$ for all $c \in (\mathbb{Z}/p\mathbb{Z})^\times$. Since χ takes values in the ℓ^{th} roots of unity, this condition is indeed implied by 3.

The last thing to check is that the maximal ideal factorization of $G(\chi)^{\widetilde{x}}$ is as claimed. We can use the fact that two ideals of a number field are equal if and only if they become equal in some larger extension field. (This follows from the fact that extending followed by intersecting back down is the identity operation). Thus it suffices to verify the claim working inside $\mathbb{Q}(\zeta_\ell, \zeta_p)$. But there we see it as a consequence of what we recalled above, together with the fact $\widetilde{P}^{p-1} = P$. □

This theorem is basically just the culmination of a bunch of calculations with Gauss sums. But it has a really interesting corollary:

Corollary 0.2. *Let $x \in \mathbb{N}[Gal]$ satisfy the equivalent conditions of the theorem, and set $y = x \cdot \sum_{a=1}^{\ell-1} \frac{a}{\ell} \sigma_a^{-1}$. Then for every nonzero ideal I of $\mathbb{Z}[\zeta_\ell]$, the ideal I^y is principal ideal.*

Proof. By existence of maximal ideal factorizations, we can assume I is maximal. If I lies above ℓ, then $I = (1 - \zeta_\ell)$ is principal already, hence certainly so is I^y. Otherwise the theorem shows that I^y is generated by $G(\chi)^{\widetilde{x}}$, and hence is principal. □

Corollary 0.3. *Same notation as in the previous corollary, and let C denote the class group of $\mathbb{Q}(\zeta_\ell)$. Then $y \cdot C = 0$.*

(We use additive notation for the class group.)

Thus we have a universal way of "killing" ideal classes. This corollary bears an interesting relationship to the formula

$$h^- = 2 \cdot \ell \cdot \prod_{\chi \text{ odd}} -\frac{1}{2} \sum_{a=1}^{\ell-1} \frac{a}{\ell} \chi(a)^{-1}$$

which we obtained earlier in the course by analytic means. Namely, the corollary and the formula have substantial overlap, but neither exactly covers the others' territory. Let's explore this a bit, by focusing on the part of C where the maximal amount of information can be extracted: this is the *odd ℓ-primary part* of C, namely the subgroup

$$C_\ell^- = \{c \in C \mid \ell^N \cdot c = 0 \text{ for some } N \in \mathbb{N}, \sigma_{-1}(c) = -c\}.$$

This is the intersection of the subgroups C_ℓ and C^-, defined by the first and second conditions respectively.

First let's see what the analytic formula says about this subgroup C_ℓ^-. The left-hand side h^- was by definition the quotient of the class number of $\mathbb{Q}(\zeta_\ell)$ by the class number of $\mathbb{Q}(\zeta_\ell + \zeta_\ell^{-1})$. But we saw that the latter class group injects into the former, and consists of the classes $c \in C$ such that $\sigma_{-1}c = c$. From this it follows that, except possibly for factors of 2 (which, it turns out, are not actually a problem), the number h^- can also be interpreted as the order of the group C^-. Thus, by the structure theory of finite abelian groups, the order of C_ℓ^- is just the largest power of ℓ which divides h^-. Now, all those 2's are relatively prime to ℓ and so can be ignored for these purposes. The upshot is that the analytic formula tells us that the order of C_ℓ^- is equal to the largest power of ℓ which divides the product

$$\ell \cdot \prod_{\chi \text{ odd}} \sum_{a=1}^{\ell-1} \frac{a}{\ell} \chi(a)^{-1}.$$

In particular, despite naive appearances the ℓ's must cancel.

Now let's switch tracks and see what the above corollary to Stickelberger's theorem says about C_ℓ^-. First let's just think about C_ℓ. Choose an N so that $\ell^N \cdot C_\ell = 0$. By your problem set, if $\omega : (\mathbb{Z}/\ell\mathbb{Z})^\times \to (\mathbb{Z}/\ell^N\mathbb{Z})^\times$ denotes the Teichmueller character, then we can decompose

$$C_\ell = \oplus_{i=0}^{\ell-2} C_\ell^i,$$

where $C_\ell^i = \{c \in C_\ell \mid \sigma_a(c) = \omega(a)^i \cdot c\}$.

Now let's apply the corollary to see what kills C_ℓ^i. Let x satisfy the equivalent conditions of the theorem, and let $y = x \cdot \sum_{a=1}^{\ell-1} \frac{a}{\ell} \sigma_a^{-1}$ as before. Then for $c \in C_\ell^i$, we have

$$0 = y \cdot c = x \cdot \sum_{a=1}^{\ell-1} \frac{a}{\ell} \omega(a)^{-i} \cdot c = \sum_{a=1}^{\ell-1} \frac{a}{\ell} \omega(a)^{-i} x \cdot c.$$

Now, I claim that for $i \neq 1$, one can find an x so that $x \cdot c = u \cdot c$, where $u \in \mathbb{Z}$ is relatively prime to ℓ. Indeed, if $i \neq 1$, there is an $a \in \mathbb{Z}$ relatively prime to ℓ such that $\omega(a)^i$ is not congruent to a (mod ℓ). We may as well take a negative. Then consider $x = \sigma_a - a$. This clearly satisfies condition 2, so it's a good x. But $x \cdot c = (\omega(a)^i - a)c$, verifying the claim. Since an integer prime to ℓ will act invertibly on any abelian group killed by ℓ^N, we deduce that when $i \neq 1$, the group C_ℓ^i is killed by $\sum_{a=1}^{\ell-1} \frac{a}{\ell} \omega(a)^{-i}$.

Now, when $i = 1$ this doesn't work, but we can certainly take $x = \ell$, and thereby see that C_ℓ^1 is killed by $\ell \cdot \sum_{a=1}^{\ell-1} \frac{a}{\ell} \omega(a)^{-1}$. (Actually, this even tells us that C_ℓ^1 is trivial, but let's forget that.)

Now, note that C_ℓ^- is just the direct sum of those C_ℓ^i with i odd. Thus we see that Stickelberger's theorem implies that C_ℓ^- is killed by

$$\ell \cdot \prod_{i \text{ odd}} \sum_{a=1}^{\ell-1} \frac{a}{\ell} \omega(a)^{-i}.$$

Seemingly by a miracle, this is actually "the same" expression as complex analysis gave us for the order of C_ℓ^-. The only difference is that here we have characters with values in $(\mathbb{Z}/\ell^N\mathbb{Z})^\times$, whereas there we had characters with values in $\mathbb{C}$. But both expressions actually do give the same integer (mod ℓ^N).

So both the analytic formula and Stickelberger's theorem give the same integer killing C_ℓ^-. But the analytic formula tells you more in that it guarantees that that the ℓ-part of that integer actually is the number of elements of the group C_ℓ^-; and the Stickelberger/Herbrand theorem tells you more in that it lets you break the product apart, and see that each factor annihilates a particular part of C_ℓ^-, namely $\sum_{a=1}^{\ell-1} \frac{a}{\ell}\omega(a)^{-i}$ annihilates C_ℓ^i for $i \neq 1$ (and C_ℓ^1 is trivial).

Given this, it's reasonable to conjecture the following common refinement of these two results: for $i \neq 1$ odd, the order of C_ℓ^i is equal to the ℓ-part of $\sum_{a=1}^{\ell-1} \frac{a}{\ell}\omega(a)^{-i}$. This statement is true: it's known as a corollary to the Mazur-Wiles proof of the Iwasawa main conjecture. So that gives a nice common refinement of what we've talked about in this class. But then again this is just at ℓ and just in the odd part, whereas both Stickelberger's theorem and the analytic formula actually gave us full integral informaion... perhaps it's just my ignorance of the state of the art, but it looks like there are still some connections to tease out and try to understand here.

Well, that was supposed to be the culmination of the course, those two different "ins" to the ideal class group of $\mathbb{Z}[\zeta_\ell]$, giving overlapping information. I think it remains a mystery what the relationship is between these two approaches, despite the evident similarities, and despite the perspective offered by Iwasawa theory. At least, it's definitely a mystery to me.