Kummer's approach: proof for regular primes

Avi Zeff
Seminar on Fermat's last theorem

Kummer found a way to generalize the approach of working with larger number rings than the integers to find integral solutions to other exponents, along the way discovering some of the fundamentals of modern algebraic number theory (and, ultimately, abstract algebra). Unfortunately, the generalization does not work for all exponents. Instead, Kummer found a class of primes p, regular primes, which abstracts out the key property that we need for the proof to work, and proved Fermat's last theorem for regular prime exponents. Conjecturally, there are infinitely many regular primes, though this is not known; so this at least establishes the theorem for a large class of exponents, but does not get us all the way there.

In order to explain the proof, we'll first need to develop some of the theory of cyclotomic fields and their rings of integers. With this in hand, we can define regular primes and prove Kummer's theorem. The book discusses connections between regular primes and Bernoulli numbers in the last subsection; we'll omit this, and mention only that all primes up to 100 are regular except for 37, 59, and 67.

1 Cyclotomic fields

Fix a prime $p \geq 3$, and let $\zeta = \zeta_p$ be a primitive pth root of unity, i.e. a complex number (up to embedding issues) such that $\zeta^p = 1$ and $\zeta^k \neq 1$ for any $1 \leq k \leq p-1$. If we like, we can imagine $\zeta = e^{2\pi i/p}$, although $e^{2\pi ik/p}$ for any $1 \leq k \leq p-1$ would work just as well. In fact we don't need p to be prime to make this definition; for example ζ_4 could be i or $-i$, since $i^4 = (-1)^4 = 1$ but lower powers are $\pm i$ or -1.

Writing $\mathbb{Q}$ for the field of rational numbers, we let $\mathbb{Q}(\zeta)$ be the smallest field containing all rational numbers as well as ζ, so for example it contains $1-\zeta$, $\frac{2+\zeta^2}{1-\zeta^3}$, etc. This is called the cyclotomic field of order p; it is a degree $p-1$ Galois extension of $\mathbb{Q}$.

Since it has the structure of an abelian group under addition and admits multiplication by rational numbers, it's also a vector space over $\mathbb{Q}$; in fact one can show that it's a finite-dimensional vector space over $\mathbb{Q}$, with basis

$$\{1, \zeta, \zeta^2, \ldots, \zeta^{p-2}\}.$$

(One might expect ζ^{p-1} to be included as well, but since we have

$$1 + \zeta + \zeta^2 + \cdots + \zeta^{p-1} = \frac{\zeta^p - 1}{\zeta - 1} = 0$$

we can write ζ^{p-1} as a linear combination of the lower powers.) The multiplicative structure means that for any $\alpha \in \mathbb{Q}(\zeta)$, we get a linear map on $\mathbb{Q}(\zeta)$ (viewed as a vector space) by multiplication by α. For example, in the basis above multiplication by ζ sends $1 \mapsto \zeta$, $\zeta \mapsto \zeta^2$, etc., up to $\zeta^{p-2} \mapsto \zeta^{p-1} = -(1 + \zeta + \cdots + \zeta^{p-2})$ and so the matrix of multiplication

by ζ is

$$\begin{pmatrix} 0 & 0 & \cdots & \cdots & -1 \\ 1 & 0 & \cdots & \cdots & -1 \\ 0 & 1 & \ddots & \cdots & -1 \\ \vdots & \vdots & \ddots & \ddots & -1 \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix}.$$

We can look at the trace and determinant of this matrix; in this case these are -1 and 1 respectively. More generally, for $\alpha \in \mathbb{Q}(\zeta)$ we write $N(\alpha)$, called the norm of α, for the determinant of the multiplication-by-α matrix, and $\mathrm{Tr}(\alpha)$ for its trace. Both are rational numbers associated to α.

There is an alternative way of thinking about the trace and norm maps: recall that we can always write α as

$$\alpha = a_0 + a_1\zeta + a_2\zeta^2 + \cdots + a_{p-2}\zeta^{p-2}.$$

Replacing ζ with ζ^k for some $1 \leq k \leq p-1$ gives a different generating element for the field $\mathbb{Q}(\zeta)$, but results in the same field; we call the set of possible alternative choices of ζ conjugates of ζ. (Referring back to the example of $p = 4$, ignoring that it's not prime, this is the operation of replacing i with $-i$, its complex conjugate; thus the name.)

If we replaced ζ by a conjugate ζ', we would correspondingly get a new element

$$\alpha' = a_0 + a_1\zeta' + a_2\zeta'^2 + \cdots + a_{p-2}\zeta'^{p-2},$$

which we call a conjugate of α'. Since there are $p-1$ options for ζ', there are likewise $p-1$ options for α', including the original α; these are the conjugates of α. One can show that if we take all of the α' and multiply them together, we get a rational number; similarly if we add them all together. These are exactly the norm and trace:

$$N(\alpha) = \prod_{\alpha'} \alpha', \qquad \mathrm{Tr}(\alpha) = \sum_{\alpha'} \alpha'.^{1}$$

The norm and trace maps give respectively multiplicative and additive maps $\mathbb{Q}(\zeta) \to \mathbb{Q}$.

1.1 Cyclotomic integers

You may be familiar with the notion of algebraic numbers: a complex number α is algebraic (over $\mathbb{Q}$) if there is a nonzero polynomial $f(x) = a_n x^n + \cdots + a_1 x + a_0$ with rational coefficients such that $f(\alpha) = 0$. By clearing denominators, we can in fact equivalently assume that all the coefficients are integers.

We say that a complex number α is *integral* (over $\mathbb{Z}$) if there is a *monic* polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ with integral coefficients, i.e. the leading coefficient is 1. Thus an algebraic integer is an algebraic number, but not necessarily vice versa: a simple example is something like $\alpha = \frac{2}{3}$, which is a zero of the linear polynomial $f(x) = 3x - 2$ but not a zero of any monic polynomial over the integers. More generally, the integral elements of $\mathbb{Q}$ are exactly the integers.

[1] The book inserts factors of the degree, $p-1$; I imagine they have a different convention from the standard one.

Since ζ is a root of $x^p - 1$, it is an algebraic integer. For any finite field extension $K/\mathbb{Q}$, we can ask: which elements of K are integral over $\mathbb{Z}$? This set of elements is called the integral closure of $\mathbb{Z}$ in K, or the ring of integers of K; one can show (with some work) that it is a subring of K.

Since ζ is integral, as are the usual integers, the ring of integers of $\mathbb{Q}(\zeta)$ must contain $\mathbb{Z}[\zeta]$, which consists of linear combinations of ζ^k for $0 \leq k \leq p-2$ with integral coefficients. A priori, it might be larger; but one can show that in fact all the integral elements of $\mathbb{Q}(\zeta)$ are of this form, i.e. $\mathcal{O}_{\mathbb{Q}(\zeta)} = \mathbb{Z}[\zeta]$.

The norm and trace maps preserve integrality, so for $x \in \mathbb{Z}[\zeta]$ we have $\mathrm{Tr}(x)$ and $N(x)$ usual integers. We will be interested in the units of $\mathbb{Z}[\zeta]$, i.e. nonzero elements $x \in \mathbb{Q}(\zeta)$ such that both x and x^{-1} are integral; note that since the norm map is multiplicative, we have $N(x)N(x^{-1}) = N(xx^{-1}) = N(1) = 1$, so $N(x)$ and $N(x^{-1})$ must be integers whose product is 1, i.e. either both 1 or -1. Conversely, if $N(x) = \pm 1$ then x is a unit of $\mathbb{Z}[\zeta]$. Units are preserved under automorphisms of the field, so in particular under conjugation and complex conjugation.

The key fact we will want about the units of $\mathbb{Z}[\zeta]$ is the following.

Proposition 1. *Let e be a unit of $\mathbb{Z}[\zeta]$, and $\overline{e}$ its complex conjugate. Then there exists some $0 \leq k \leq p-1$ such that $e/\overline{e} = \zeta^k$.*

1.2 Ideal theory

For applications to Fermat's last theorem, it would be convenient to extend the fundamental theorem of arithmetic to $\mathbb{Z}[\zeta]$, i.e. the fact that every natural number can be written uniquely (up to order) as a product of prime numbers. That is, we'd like to say that every element of $\mathbb{Z}[\zeta]$ can be written as a product of a unit together with powers of prime elements in an essentially unique way; the earlier attempted proof of Lamé was based on this assumption. Unfortunately, it is not true in general; it first fails for $p = 23$. Kummer realized however that unique factorization still holds for what he termed "ideal numbers," which evolved into the modern notion of ideals in a ring, which we'll briefly review.

The idea is to replace an element x in a (commutative unital) ring R with the set of elements divisible by x, which we write as (x) or xR, i.e. the subset of R consisting of elements of the form xy for $y \in R$; we say that (x) is the ideal generated by x. Note that for a unit $u \in R^\times$, $(ux) = (x)$ since numbers divisible by x are also divisible by ux: $xy = u^{-1}uxy$. In particular, $(u) = (1) = R$, the unit ideal.

These subsets I have two key properties: they are closed under addition and subtraction, and for any element $y \in R$ and $i \in I$ we have $iy \in I$, i.e. multiplication by any element of R preserves I. We define an ideal I of a ring R to be a subset with these properties. (In more abstract language, they are R-submodules of R.)

In a field K, every element is either 0 or a unit, so either an ideal I is just the set $\{0\}$ or it contains a unit and therefore contains the entire field. Thus fields have exactly two ideals, (0) and (1).

More generally, some rings (principal ideal domains[2]) have the property that all of their

[2]Strictly speaking this requires another condition, that of being an integral domain, i.e. having no nontrivial zero divisors or equivalently having an embedding in a field; all the rings with which we'll be

ideals are *principal*, i.e. of the form (x) for some element x, i.e. generated by a single element. More generally though not all ideals will be of this form: for example in the ring $k[x, y]$ for k a field, the ideal (x, y) consisting of all elements of $k[x, y]$ of the form $xf + yg$ for $f, g \in k[x, y]$ cannot be generated by a single element. We say that a ring is Noetherian if every ideal can be generated by finitely many elements.

We are interested in a special kind of Noetherian ring: a Dedekind domain. We won't need the precise definition; we'll just mention that for any finite extension $K/\mathbb{Q}$, the ring of integers $\mathcal{O}_K$ is a Dedekind domain, so in particular $\mathbb{Z}[\zeta]$ is one. These are the kinds of rings that have unique factorization for ideals.

To explain this a little better, we should say what multiplication of ideals means: if I and J are two ideals, then IJ is the ideal consisting of elements of the form ij for $i \in I$ and $j \in J$. In particular $(x)(y) = (xy)$. We also mention that there is a notion of prime ideals, whose definition is completely abstract and ideal-theoretic, a priori having nothing to do with multiplication.

Theorem. *Every nonzero ideal in a Dedekind domain can be written uniquely (up to order) as a product of prime ideals.*

Note that if every ideal was principal, this would boil down to unique factorization on the level of elements; so the Dedekind domains that have unique factorization for elements are exactly the ones which are principal ideal domains. In general, ideals of a Dedekind domain can be generated by at most two elements. There is a construction one can do where we consider principal ideals to be trivial and form a group out of the set of ideals of our Dedekind domain under multiplication modulo trivial ideals, formally adding inverses as needed; this results in a finite abelian group called the class group. One of the key invariants associated to a number field is the class group of its ring of integers, though we won't see it much in this class.

2 Proof of Fermat's last theorem for regular primes

Let $p \geq 5$ be a prime (we handled $p = 3$ in the last talk, as well as the case of exponent 4, so this is what remains). Recall that we can flip signs as needed, as we can rephrase Fermat's last theorem as the following statement: there exist no integer solutions to

$$x^p + y^p + z^p = 0$$

with $xyz \neq 0$. We can safely assume that x, y, and z are relatively prime; we further assume for the moment that in fact $xyz \not\equiv 0 \pmod{p}$, i.e. none of x, y, and z are divisible by p.

Working over $\mathbb{Q}(\zeta)$, we can factor $x^p + y^p$ as

$$(x + y)(x + \zeta y)(x + \zeta^2 y) \cdots (x + \zeta^{p-1} y),$$

so this should be equal to $-z^p$. It is fairly straightforward to show (by Bézout's theorem) that the $x + \zeta^k y$ are pairwise coprime, in the sense that any two would together generate

concerned today embed in $\mathbb{C}$, so we will not worry about this sort of thing.

the unit ideal. Their product is equal to $-z^p$, and so the product of their ideals is equal to $(-z^p) = (-z)^p$, so by unique factorization $(-z)$ factors into a unique product of prime ideals, each of which appears p times in $(-z)^p$, and so each factor $(x + \zeta^k y)$ must be a pth power of some subset of these factors. In particular, $(x + \zeta y) = I^p$ for some ideal I of $\mathbb{Z}[\zeta]$.

Now, we would like to be able to conclude that since $I^p = (x + \zeta y)$ is principal, so is the underlying ideal I. Unfortunately this is not always true, so let's require it as a condition:

Definition. A prime p is regular if for any (prime) ideal I of $\mathbb{Z}[\zeta]$, if I^p is principal than so is I.

This has several equivalent reformulations: for example, p is regular if and only if p does not divide the order of the class group of $\mathbb{Z}[\zeta]$.

More importantly for our applications, it has the following consequence. For $x \in \mathbb{Z}[\zeta]$, writing

$$x = a_0 + a_1\zeta + \cdots + a_{p-2}\zeta^{p-2}$$

we have

$$x^p \equiv a_0^p + (a_1\zeta)^p + \cdots + (a_{p-2}\zeta^{p-2})^p \pmod{p}$$

(by the binomial theorem), which since $\zeta^p = 1$ is just

$$x^p \equiv a_0^p + \cdots + a_{p-2}^p \pmod{p},$$

i.e. x^p is congruent to an integer modulo p. (Note that if we replaced x by a conjugate, since the ζ^k disappear we'd get the same thing, i.e. $x^p \equiv x'^p \pmod{p}$.) If p is regular and x is a unit, the converse is true: if $x \in \mathbb{Z}[\zeta]$ is a unit and is congruent to an integer modulo p, then $x = y^p$ for some unit $y \in \mathbb{Z}[\zeta]$.

We now return to the situation at hand, and assume henceforth that p is regular. We have $(x + \zeta y) = I^p$ for some ideal I, which by regularity must be principal, so we write $I = (t)$. Thus $(x + \zeta y) = (t)^p$, i.e.

$$x + \zeta y = et^p$$

for some unit e of $\mathbb{Z}[\zeta]$. Taking the complex conjugate gives

$$x + \overline{\zeta} y = \overline{e}\overline{t}^p.$$

By Proposition 1, we can write $e/\overline{e} = \zeta^k$ for some $0 \leq k \leq p-1$, i.e. $\overline{e} = \zeta^{-k}e$; and we note that $\overline{\zeta} = \zeta^{-1}$ since $\zeta\overline{\zeta} = |\zeta|^2 = 1$. So we can rewrite the above as

$$x + \zeta^{-1}y = x + \overline{\zeta} y = \overline{e}\overline{t}^p = \zeta^{-k}e\overline{t}^p,$$

and we saw above that $\overline{t}^p \equiv t^p \pmod{p}$. So we have

$$x + \zeta^{-1}y \equiv \zeta^{-k}et^p = \zeta^{-k}(x + \zeta y) \pmod{p},$$

or

$$\zeta^k(x + \zeta^{-1}y) \equiv x + \zeta y \pmod{p}.$$

If $k = 0$, we would have $x + \zeta^{-1}y \equiv x + \zeta y \pmod{p}$, or equivalently

$$(\zeta^2 - 1)y \equiv 0 \pmod{p}.$$

Since $\{1, \zeta^2\}$ is a subset of a basis for $\mathbb{Z}[\zeta]/(p)$, the statement that $(\zeta^2-1)y = -y\cdot 1 + y\cdot\zeta^2 \equiv 0 \pmod{p}$ implies that all of the coefficients must be zero modulo p, i.e. $p|y$, which we have assumed not to be the case.

If $k = 1$, we would have $\zeta x + y \equiv x + \zeta y \pmod{p}$, i.e. $-(x-y)\cdot 1 + (x-y)\cdot\zeta \equiv 0 \pmod{p}$, so similarly each coefficient must be divisible by p so $x \equiv y \pmod{p}$. We set this possibility aside for the moment.

Working similarly for all values of k, we'll always get either $x \equiv 0$, $y \equiv 0$, or $x \equiv y$ modulo p for all k up to $p-1$. The former two possibilities are contrary to our assumption, so we conclude $x \equiv y \pmod{p}$. The same argument applies to x and z or y and z, so

$$x \equiv y \equiv z \pmod{p}.$$

The equation $x^p + y^p + z^p = 0$ then gives

$$3x^p \equiv 0 \pmod{p},$$

so $p|3x^p$ and therefore either $p|3$ or $p|x^p$. Since $p \geq 5$ and $p \nmid x$ by assumption, both are impossible and we get the desired contradiction.

This leaves us with the case where at least one (equivalently exactly one) of x, y, and z is divisible by p. Assume that say z is divisible by p (by symmetry it doesn't matter which variable we pick), so

$$x^p + y^p = -z^p \equiv 0 \pmod{p}.$$

Note that $x+y \equiv x^p+y^p \pmod{p}$, so $x+y \equiv 0 \pmod{p}$. Letting $\pi = \zeta - 1$, the factorization

$$1 + x + x^2 + \cdots + x^{p-1} = \prod_{k=1}^{p-1}(x - \zeta^k)$$

evaluated at $x = 1$ together with the identification $1 - \zeta^k = u_k\pi$ via factorization for some unit u_k gives $p = u\pi^{p-1}$ for some unit u. Therefore $p \in (\pi)$ and so $x + y \equiv 0 \pmod{\pi}$ as well, so

$$x + \zeta y = x + y + \pi y \equiv 0 \pmod{\pi},$$

i.e. π divides $x + \zeta y$. If π divided $x + \zeta y$ more than once, since it divides $x + y$ at least $p - 1$ times it would have to divide π more than once, so $\pi|y$, which since y is an integer would imply $p|y$ contrary to our assumptions; so π divides $x+\zeta y$ exactly once, and the same applies to all the factors $x+\zeta^k y$ for $1 \leq k \leq p-1$. Meanwhile π divides $x+y$ $pm+1$ times, where $m = n(p-1) - 1$ for some $n \geq 1$. Therefore by the same argument as above we can write

$$x + \zeta^k y = \pi e_k t_k^p$$

for units e_k and relatively prime integers $t_k \in \mathbb{Z}[\zeta]$ not divisible by π, so in particular (using $p - 1 \equiv -1 \pmod{p}$)

$$\begin{aligned} x + \zeta y &= \pi e_1 t_1^p, \\ x + \zeta^{-1} y &= \pi e_{-1} t_{-1}^p, \\ x + y &= \pi^{pm+1} e_0 t_0^p. \end{aligned}$$

Solving these equations gives

$$e_1 t_1^p - e_0(1+\zeta)\pi^{pm} t_0^p + \zeta e_{-1} t_{-1}^p = 0.$$

Using the fact that $1+\zeta$ is a unit and reducing modulo p gives

$$t_1^p + \zeta e_{-1}/e_1 t_{-1}^p = e\pi^{pm} t_0^p \equiv 0 \pmod{p}$$

for some unit e. Modulo p, we know that t_1^p and t_{-1}^p are congruent to integers, so $\zeta e_{-1}/e_1$ must be as well; therefore by regularity and the fact that it's a unit we conclude that it is itself a pth power, i.e. we can find elements $x, y, z \in \mathbb{Z}[\zeta]$ relatively prime and not divisible by π with

$$x^p + y^p = e\pi^{pm} z^p$$

(namely $x = t_1$, $y = \sqrt[p]{\zeta e_{-1}/e_1}\, t_{-1}$, and $z = t_0$).

We can then iterate this process to get a new solution to the analogous equation with m replaced by $m-1$ (possibly with a different unit e). In particular, there are no solutions with $m = 0$ (essentially as in the first case), so there are no solutions for any m by infinite descent.