Geometry of Numbers

December 4, 2013

Recall that if $F/\mathbb{Q}$ is a finite extension (F is a "number field"), then we defined a subring $\mathcal{O}_F \subset F$ consisting of the *integral* elements, i.e. those $\alpha \in F$ which satisfy the following equivalent properties:

1. $\mathbb{Z}[\alpha]$ is finitely generated as an abelian group;
2. $\mathbb{Z}[\alpha]$ is a finite free $\mathbb{Z}$-module with basis $1, \alpha, \ldots, \alpha^{d-1}$, where $d = deg(\alpha) = [\mathbb{Q}(\alpha) : \mathbb{Q}]$;
3. there exists some monic polynomial with $\mathbb{Z}$-coefficients of which α is a root;
4. the minimal polynomial of α has $\mathbb{Z}$-coefficients.

But, we said it was not clear a priori that $\mathcal{O}_F$ itself satisfies the analog of the first two points, i.e., that $\mathcal{O}_F$ is a free $\mathbb{Z}$-module of rank $[F : \mathbb{Q}]$ (= free abelian group isomorphic to $\mathbb{Z}^{[F:\mathbb{Q}]}$). Our goal today is to prove that fact, and to set up a nice picture for $\mathcal{O}_F$ which will be useful in the future.

The idea is that a good way to picture a finite free $\mathbb{Z}$-module is to imagine it sitting discretely inside a finite dimensional $\mathbb{R}$-vector space which it spans, as in the usual picture of $\mathbb{Z} \subset \mathbb{R}$, or, to be slightly more exotic, $\mathbb{Z}[i] \subset \mathbb{C}$ (but *not* like $\mathbb{Z}[\sqrt{2}] \subset \mathbb{R}$, which is everywhere dense). Here is the main definition.

Definition 0.1. *Let V be a finite-dimensional $\mathbb{R}$-vector space, and $L \subset V$ an additive subgroup. We say that L is a* lattice *if L is discrete as a subset of V. We say that L is a* full lattice *if it is a lattice, and the $\mathbb{R}$-span of L is equal to V.*

Some remarks are in order. First, by an "additive subgroup", we mean a subgroup of the underlying abelian group of V. So L is just a subset of V which is closed under vector addition and subtraction. When we say that L is discrete in V, we mean relative to the canonical topology on V. There are two (equivalent) ways to describe this topology. First, by choosing a basis we can identify $V \simeq \mathbb{R}^d$; then we can just transport the usual topology on $\mathbb{R}^d$ along this isomorphism. The resulting topology on V does not depend on the isomorphism, because a change-of-basis on $\mathbb{R}^d$ induces a homeomorphism of $\mathbb{R}^d$. Second, we can arbitrarily choose a nondegenerate inner product on V (think of the usual dot product on $\mathbb{R}^d$). This induces a norm, hence a topology, in the usual way. Again, the topology does not depend on the inner product chosen. In practice we'll take this second option, and in fact we'll sometimes refer to distances in V, balls in V, and so in. These are supposed to be taken with respect to some arbitrarily fixed inner product on V. Really, you can just picture $\mathbb{R}^d$ with its standard inner product if you like.

Also, recall that $L \subset V$ being discrete means, by definition, that every point $x \in L$ has an open neighborhood U in V such that the only point of $L \cap U$ is x ("x is an isolated point"). But in our case this is equivalent to requiring the same thing only for $x = 0$. Indeed, if U is such a neighborhood for 0, then $x + U$ is such a neighborhood for an arbitrary $x \in L$. So to check something is a lattice, we only

need to see that 0 is an isolated point. This is like how to check a homomorphism is injective, you only need to see the kernel is trivial.

Finally, we remark that one can often easily reduce consideration of arbitrary lattices to consideration of full lattices, by replacing V by the $\mathbb{R}$-span of L. Just cut down the ambient space until the lattice is full.

OK, now here is the main result.

Theorem 0.2. *Let $L \subset V$ be a full lattice in a real vector space of dimension d. Then L is isomorphic to $\mathbb{Z}^d$ as an abelian group, and a $\mathbb{Z}$-basis for L is necessarily also an $\mathbb{R}$-basis for V.*

Another way of phrasing this is to say that the *pair* $L \subset V$ is isomorphic to the standard pair $\mathbb{Z}^d \subset \mathbb{R}^d$. Every full lattice looks like you'd think.

Here is the crucial lemma for the proof:

Lemma 0.3. *Let $L \subset V$ be a nonzero lattice in a finite dimensional real vector space V. Choose arbitrarily an inner product on V. Then there exists a nonzero point of L which is of minimal distance to 0.*

Proof. Since L is nonzero, there exists a nonzero $x \in L$. Let D denote the distance from x to 0, and consider the closed ball $B(0, D)$ of radius D centered at 0 in V. Note that $L \cap B(0, D)$ is both discrete and compact, so it must be a finite set. Thus the set of nonzero points of $L \cap B(0, D)$ is finite and non-empty. So there must be a point of minimal distance to the origin among all the points in $L \cap B(0, D)$. But everything outside $B(0, D)$ has distance $> D$ from the origin, so such a point will actually be minimal out of all points in L. □

Now we give the proof of the theorem. It's a nice proof, in that it actually gives a procedure for constructing a $\mathbb{Z}$-basis of L.

Proof. We work by induction on d. When $d = 0$ there is nothing to prove. So suppose $d > 0$. Since $L \subset V$ is full, it must be that L is nonzero. By the lemma, then, there exists a point $x_0 \in L$ of minimal distance to the origin. The idea is that x_0 should extend to a $\mathbb{Z}$-basis of L. With an eye towards using the inductive hypothesis, consider the homomorphism on quotients

$$L/(\mathbb{Z} \cdot x_0) \to V/(\mathbb{R} \cdot x_0)$$

induced by the inclusion $L \to V$. I claim that this homomorphism is injective, and its image is a full lattice.

The "full" claim is pretty clear: if everything in V is in the $\mathbb{R}$-span of L, then certainly the same holds after passing to quotients. So we just need to check that the homomorphism is injective and the image is a lattice. We can actually do both at once if we show that there is an open neighborhood U of 0 in $V/(\mathbb{R} \cdot x_0)$ such that the only point of $L/(\mathbb{Z} \cdot x_0)$ whose image lies in U is the zero element of $L/(\mathbb{Z} \cdot x_0)$. Unwinding the quotients, what this means is that we need to find an $\epsilon > 0$ such that the tube of radius ϵ around the subspace $\mathbb{R} \cdot x_0$ contains no points of L other than the integer multiples of x_0.

But let's think about regions where we know there can be no points of L other than the $\mathbb{Z} \cdot x_0$. Let d_0 be the distance from x_0 to 0. Then one such region is the open ball of radius d_0 around the origin, because x_0 was of minimal distance to the origin. But then by the usual translation argument, we automatically get more such regions, namely the open balls of radius d_0 centered at any point of L,

in particular at any point of $\mathbb{Z} \cdot x_0$. It's easy to see that the union of the open balls of radius d_0 centered at the $\mathbb{Z} \cdot x_0$ contains a tube of radius $\epsilon = \frac{\sqrt{3}}{2} d$ around $\mathbb{R} \cdot x_0$ (I drew the picture in class), so that does the job for us.

Thus we've shown that $L/(\mathbb{Z} \cdot x_0)$ is a full lattice in $V/(\mathbb{R} \cdot x_0)$. But now the dimensions are one less. So by the inductive hypothesis there exists a $\mathbb{Z}$-basis $\overline{x_1}, \ldots \overline{x_{d-1}}$ of $L/(\mathbb{Z} \cdot x_0)$ which is simultaneously an $\mathbb{R}$-basis of $V/(\mathbb{R} \cdot x_0)$. But then a simple algebraic lemma shows that, if $x_1, \ldots, x_{d-1}$ is any lift of the $\overline{x_1}, \ldots \overline{x_{d-1}}$ to L, then $x_0, x_1, \ldots, x_{d-1}$ is a $\mathbb{Z}$-basis of L which is simultaneously an $\mathbb{R}$-basis of V, as desired. (One way of phrasing this algebraic lemma, which avoids explicitly thinking about elements and linear combinations, is this: on the one hand, every "short exact sequence" with free quotient splits; but on the other hand, every splitting induces a direct sum decomposition.) $\square$

The proof turned out to be long, but I hope you'll agree that the idea was simple: to produce a $\mathbb{Z}$-basis of L, start by finding the closest vector to the origin, then take the closest vector to the origin not in the span of the first vector, then take the closest vector to the origin not in the span of the previous two, etcetera, and then you'll be done after d steps.

Okay, now let's apply this to number theory! Let F be a number field with ring of integers $\mathcal{O}_F$. We want to find a nice ambient Euclidean space in which to embed $\mathcal{O}_F$. This will be the *Minkowski space*. First let me give two illustrative examples.

Example 1. Let F be an imaginary quadratic field, $F = \mathbb{Q}(\sqrt{-d})$ with d a positive squarefree integer. Then the Minkowski space is just $\mathbb{C}$, viewed as a two-dimensional vector space over $\mathbb{R}$. Certainly F sits inside $\mathbb{C}$, thus so does $\mathcal{O}_F$. It's not hard to see that $\mathcal{O}_F \subset \mathbb{C}$ is a lattice. For example, when $d = -1$, this is a square lattice (generated by 1 and ζ_4), and when $d = -3$ it is a hexagonal lattice (generated by 1 and ζ_3). For other d you get a similar lattice, but it's not as symmetric. (This is related to the fact that for any other d the only units in $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ are ± 1).

Example 2. Let F be a real quadratic field, so now $F = \mathbb{Q}(\sqrt{d})$ with d a positive squarefree integer. We don't want to take the usual embedding $\mathbb{Q}(\sqrt{d}) \subset \mathbb{C}$ anymore. Even though the dimensions are right, the image of $\mathcal{O}_F$ is not a lattice: it's actually a dense subset of the line $\mathbb{R} \subset \mathbb{C}$. Somehow we need to see $\mathcal{O}_F$ "pop out" into its natural two dimensions. The canonical way to do this is to instead take the Minkowski space to be $\mathbb{R} \oplus \mathbb{R}$, and map $F \to \mathbb{R} \oplus \mathbb{R}$ by having the first coordinate be the usual embedding $\mathbb{Q}(\sqrt{d}) \subset \mathbb{R}$, but the second should be the *conjugate* of this embedding, i.e. first compose with the nontrivial automorphism of $\mathbb{Q}(\sqrt{d})$ (meaning $a + b\sqrt{d} \mapsto a - b\sqrt{d}$) and then embed into $\mathbb{R}$. For example, when d is not $1 \pmod 4$ so that $\mathcal{O}_F = \mathbb{Z}[\sqrt{d}]$, then the image is the lattice generated by $(1, 1)$ and $(\sqrt{d}, -\sqrt{d})$, these being the images of the $\mathbb{Z}$-basis vectors 1 and $\sqrt{d}$ in $\mathbb{Z}[\sqrt{d}]$.

The Minkowski space for a general F will be a mix of these two examples: it will be a direct sum of the form $\mathbb{R}^r \oplus \mathbb{C}^s$, with $r + 2s = n$, the degree of $F/\mathbb{Q}$.

Now, there's a completely formal way to produce this Minkowski space from F: just take the tensor product $F \otimes \mathbb{R}$. Generally, for a $\mathbb{Q}$-vector space V, the tensor product $V \otimes \mathbb{R}$ is a canonically determined $\mathbb{R}$-vector space which "has the same basis" as V. Morally, you just formally allow yourself to scalar multiply not just by elements of $\mathbb{Q}$, but also by elements of $\mathbb{R}$. It's a purely algebraic construction. In class I spent some time talking about this, but now I've decided it's best to skip that whole story, and just use the ad hoc, more explicit description of Minkowski space, which is the following.

Definition 0.4. *Let F be a number field. Then the Minkowski space $F_{\mathbb{R}}$ is defined to be the real vector space*

$$\mathbb{R}^r \oplus \mathbb{C}^s,$$

where there is a factor of $\mathbb{R}$ for every field embedding $F \to \mathbb{R}$, and a factor of $\mathbb{C}$ for every complex-conjugate pair of field embeddings $F \to \mathbb{C}$ which do not *land inside $\mathbb{R}$.*

Note that this definition does recover the quadratic examples we discussed. This is me stopping while you note that.

The background to this definition is the following. We know from field theory that for any separable extension of fields E/L and any algebraically closed field L' containing L, there are exactly $n = deg(E/L)$ homomorphisms $E \to L'$ which restrict to the identity on L. (Idea: a generator for the extension E/L can go to any of the n roots of its minimal polynomial in L'.) Recall also that a field homomorphism is automatically injective, meaning it's an embedding. Specializing to our case, we find that there are exactly $n = deg(F/\mathbb{Q})$ field embeddings $F \to \mathbb{C}$.

But note that complex conjugation gives a $\mathbb{Z}/2$-action on this set of embeddings. The fixed points are exactly the *real embeddings* of F, i.e. the field homomorphisms $F \to \mathbb{R}$. The number of these is the r above. The non-fixed points then necessarily come in complex-conjugate pairs; the number of such pairs is the s above. In total we see that $r + 2s = n$, so that our above-defined Minkowski space has the "right dimension" as a real vector space.

The next thing to specify is how $\mathcal{O}_F$ embeds into $F_{\mathbb{R}} = \mathbb{R}^s \oplus \mathbb{C}^s$. In fact, all of F embeds there, via a map $F \to \mathbb{R}^s \oplus \mathbb{C}^s$ defined as follows. The real factors are labelled by the embeddings $F \to \mathbb{R}$, so on the real factors we can just use the corresponding embedding $F \to \mathbb{R}$ to map to that real factor. To handle the complex factors, we have to arbitrarily chose one out of each pair of complex conjugate non-real complex embeddings $F \to \mathbb{C}$, and then we can use those chosen embeddings to map to the corresponding $\mathbb{C}$-factors.

This business of having to arbitrarily choose one out of each pair of conjugate embeddings should seem a little funny. I mean, at the beginning I said that Minkowski space could be canonically defined in terms of F as $F \otimes \mathbb{R}$, and there the embedding $F \to F \otimes \mathbb{R}$ is tautological. The explanation is that when you unwind the tensor product $F \otimes \mathbb{R}$, you find that it doesn't quite identify with the Minkowski space $F_{\mathbb{R}}$ defined above. Instead it identifies with a variant where you replace each factor of $\mathbb{C}$ by a two-dimensional real vector space whose elements are pairs of complex conjugate complex numbers. That's the more canonical version of Minkowski space; for example it receives a canonical map from F, without choices. The choices only come if you identify this more enlightened Minkowski space with the naive concrete definition given above. But we'll generally stick with the concrete version, so we have to choose one out of every pair of conjugate complex embeddings. (Note that we did this implicitly in the imaginary quadratic example!)

In any case, we have this canonical map $F \to F_{\mathbb{R}}$ from F to the Minkowski space, a real vector space of dimension $n = [F : \mathbb{Q}]$. Basically, to define it you just take all n of the complex embeddings $F \to \mathbb{C}$, then cut the real ones down to size and throw out half the complex ones as being redundant. As a bit of shorthand terminology, we'll call the individual factors of the image of F under this map $F \to F_{\mathbb{R}} = \mathbb{R}^r \oplus \mathbb{C}^s$ the *Minkowski coordinates* of F. Now that that's out of the way, we can prove what we wanted.

Theorem 0.5. *The above map $\mathcal{O}_F \to F_{\mathbb{R}}$ embeds the ring of integers $\mathcal{O}_F$ as a full lattice in Minkowski space $F_{\mathbb{R}}$.*

Proof. First we will just show that $\mathcal{O}_F$ embeds as a lattice. So we need to find an open neighborhood U of 0 in $F_{\mathbb{R}}$, such that 0 the only element of $\mathcal{O}_F$ whose image lies in U.

To tease this out, let $\alpha \in \mathcal{O}_F$. Consider the polynomial $p(X) = \prod_\sigma (X - \sigma\alpha)$, where σ ranges over all complex embeddings $F \to \mathbb{C}$. This polynomial has integer coefficients. Indeed, it certainly has rational coefficients, since the Galois group of the normal closure of F just permutes the factors; hence fixes the polynomial; hence fixes the coefficients. But also all of these $\sigma\alpha$ are algebraic integers, and the coefficients are polynomials in them. Thus, since the ring of algebraic integers is a ring, it follows that the coefficients are algebraic integers. But an algebraic integer which is also a rational number has to be an integer (we checked $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$). Thus $p(X)$ has integer coefficients, as claimed. (Alternate proof: the collection of all the $\sigma\alpha$'s is just the set of conjugates of α, taken with some uniform multiplicity. Thus $p(X)$ is just a power of the minimal polynomial of α, and hence has integer coefficients.)

But note, on the other hand, that the $\sigma\alpha$ are all determined by the Minkowski coordinates: they are either Minkowski coordinates themselves, or else complex conjugates of Minkowski coordinates. In particular, they are continuous functions of the Minkowski coordinates. Since polynomials are continuous, it follows that the coefficients of $p(X)$ are also continuous functions of the Minkowski coordinates. Or we could say, the polynomial $p(X)$ itself is a continuous function of the Minkowski coordinates. But at the point 0 this polynomial is X^n. Thus, in a neighborhood of 0 the polynomial is close to X^n, say each of its coefficients is at most distance $1/2$ from the corresponding coefficients of X^n. But two integers can't be so close unless they're equal, so in a neighborhood of 0 the only possibility for $p(X)$ is X^n, whose only root is 0. That proves that 0 is an isolated point of $\mathcal{O}_F$ in $F_{\mathbb{R}}$, so that $\mathcal{O}_F$ is a lattice in $F_{\mathbb{R}}$.

To finish, we have to see that the $\mathbb{R}$-span of $\mathcal{O}_F$ is all of $F_{\mathbb{R}}$. Well in any case $\mathcal{O}_F$ is a full lattice in its $\mathbb{R}$-span, so $\mathcal{O}_F$ is free of rank $\leq n$. We just want to see that the rank can't be strictly less than n. But we know this already, since we can certainly find lots of rank-n submodules of $\mathcal{O}_F$, namely all the $\mathbb{Z}[\alpha]$'s where α is any algebraic integer which generates F. (To get such an α, start with any generator of F, then multiply by a suitable integer to clear denominators in its minimal polynomial. It will still generate, but now it's also an algebraic integer.) □

In the linear-algebraic "tensor product" description of Minkowski space, contrasted with the "Minkowski coordinates" approach we've adopted, this proof looks as follows. To every α in the ring $F \otimes \mathbb{R}$ we can assign the characteristic polynomial $p_\alpha(X)$ of α acting on $F \otimes \mathbb{R}$, viewed as a finite-dimensional real vector space. This assignment $\alpha \mapsto p_\alpha(X)$ is then a continuous functions. But on the other hand, when α lies in $\mathcal{O}_F$, then the polynomial $p_\alpha(X)$ has integer coefficients. Indeed, by choosing arbitrarily a basis of $F/\mathbb{Q}(\alpha)$ we see that $p_\alpha(X)$ is a power of the characteristic polynomial of α acting on $\mathbb{Q}(\alpha)$, which is the minimal polynomial of α. Then we finish the proof as above: by continuity, when α is near 0 the polynomial $p_\alpha(X)$ must be near $p_0(X) = X^n$. But integers can't be near each other without being equal, so any $\alpha \in \mathcal{O}_K$ which is near 0 must itself be 0.

In any case, we get the following corollary.

Corollary 0.6. *$\mathcal{O}_F$ is a free abelian group of rank $n = [F : \mathbb{Q}]$.*

At the end of class, I outlined how the rest of this course will go. We're aiming for the theorem of Kummer quoted two lectures ago — you know, the weird one relating the class number of $\mathbb{Q}(\zeta_p)$ to numerators of Bernoulli numbers. The key to this theorem is the following formula (I hope I did my simplifications right...):

$$h_p^- = -(2 \cdot p^{-1})^{\frac{p-3}{2}} \cdot \prod_{\chi} \left(\sum_{k=1}^{p-1} \chi(k) \cdot k \right).$$

Here the number on the left is a variant of the class number h_p of $\mathbb{Q}(\zeta_p)$, and on the right we have χ ranging over all characters $(\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{C}^\times$ which are *odd*, i.e. $\chi(-1) = -1$. (This oddness is related to the "$-$" in h_p^-. But there's no analogous statement if you remove the $-$ and remove the word "odd".)

Now, the basic idea behind this formula is simple, but the details are amazing. The idea is that the above formula is somehow obtained by "analytic continuation" of problem 2 on our first problem set! In more fancy terms, problem 2 of our problem set describes how primes decompose in the ring of integers $\mathbb{Z}[\zeta_p]$. We can package all of this information into a generating function known as the *zeta function* of $\mathbb{Q}(\zeta_p)$, and what emerges from problem 2 is a description of this zeta function in terms of similar functions ("L-functions") which live on $\mathbb{Q}$ rather than on $\mathbb{Q}(\zeta_p)$. This is where those characters χ come in. Then the magic happens. You stop thinking of these zeta functions as formal generating functions, and actually treat them as complex analytic functions. In particular, you analytically continue them to the point $s = 0$, where the original series didn't converge. Then, after messing around a bit (and inputting some nice geometry of numbers!), you get the h_p^- out of the zeta functions, and you get the right-hand side of the above formula out of the L-functions. And that's the proof.

So we'll try to do that. Then, at the end, we'll look into some refinement of the above formula. The left hand side of the equation is the size of a naturally defined abelian group, and one asks what the significance of the right hand side is in terms of abelian groups. This is part of a deep and difficult story (related to the "main conjecture of Iwasawa theory", proved by Mazur and Wiles), but we can at least investigate a part of it, due to Herbrand in the 19th century. The approach will be to try to get a handle on the class group of $\mathbb{Q}(\zeta_p)$ by seeing how certain explicit elements of $\mathbb{Q}(\zeta_p)$ decompose into prime factors. These explicit elements are the Gauss sums which came up in Gauss's orginal study of cyclotomy. So we'll finally get to go back to that.