Units in Number Rings

Guiding Question

It is easy to tell what the units $\mathcal{O}_K^\times$ are in a quadratic number field K. We always have that α is a unit if and only if $N(\alpha) = \pm 1$, and $|N(\alpha)| \geq 1$. In imaginary quadratic fields, it's easy to see that $N(\alpha) \geq 1$ for all α, and that $N(\alpha) = 1$ exactly when α is a root of unity. Since there are only 12 roots of unity quadratic over $\mathbb{Q}$, this means the unit group of $\mathcal{O}_K$ is finite for an imaginary quadratic field. On the other hand, the theory of Pell equations tells us that for K a real quadratic field, $\mathcal{O}_K^\times/\{\pm 1\}$ is always cyclic. Can we determine the structure of $\mathcal{O}_K^\times$ for an arbitrary number field K?

The answer is yes. The result usually bears the name Dirichlet's Unit Theorem, which says that if K is a number field with r real embeddings and $2s$ complex embeddings, then $\mathcal{O}_K^\times$ is a finitely generated group of rank $r + s - 1$.

Preliminary Questions

Question 1. Prove the assertion above that $\alpha \in \mathcal{O}_K$ is a unit if and only if $N(\alpha) = \pm 1$.

Question 2. Our proof below will proceed in two steps: first we will show that $\mathcal{O}_K^\times$ is finitely generated of rank at most $r + s - 1$, and then we will show that we can find $r + s - 1$ independent units. Reflect a bit on how you think the proof might go based on this description.

(a) What ways can you think of to show that a group is finitely generated? To bound the rank?

(b) What ways can you think of to show that units exist?

$\mathcal{O}_K^\times$ is finitely generated

You are already familiar with the Euclidean embedding $E\colon K \hookrightarrow \mathbb{R}^r \times \mathbb{C}^s$. Although $\mathcal{O}_K^\times$ is a group under multiplication, it's often easier to work with groups written additively. Luckily, since high school we've all known a tool which converts multiplication to addition. Consider the map $L\colon \mathcal{O}_K^\times \to \mathbb{R}^{r+s}$, where the ith component of $L(\alpha)$ is $n_i \log|\sigma_i(\alpha)|$, where $n_i = 1$ if σ_i is real and 2 if it is complex. (This weighting isn't technically necessary for the rest of the proof to work, but it's philosophically more correct to count the complex embeddings twice, since they come in pairs. It will also make some formulas more pretty.) This is obviously a homomorphism.

Question 3. The image of L satisfies an "obvious" linear relation. What is it?

We've succeeded in mapping $\mathcal{O}_K^\times$ into a vector space of dimension $r+s-1$ (the hyperplane satisfying that linear relation), which seems promising. Since $\mathcal{O}_K^\times$ is very "integer-y", it seems pretty reasonable to guess that its image under L is a lattice.

Question 4. Suppose you knew that $L(\mathcal{O}_K^\times)$ was a lattice. What final piece of information would you need to know in order to conclude that $\mathcal{O}_K^\times$ is finitely generated?

There seems to be a standard argument for getting this piece of information. I found one source online calling the result "Kronecker's Theorem", but that seems to refer to another, mostly-unrelated theorem about Diophantine approximation. The same source claimed that the following argument could be replaced by "a more elegant proof using linear algebra", but I couldn't find any sources offering any other arguments. (And I couldn't think of a more elegant proof myself.)

Question 5. Suppose $\alpha \in \mathcal{O}_K$ satisfies $|\sigma_i(\alpha)| = 1$ for all real and complex embeddings.

(a) Show that there is a bound B depending only on $[K : \mathbb{Q}]$, and not on K or α, so that the size of the coefficients of the minimal polynomial of α is $\leq B$.

(b) Conclude that there are only finitely many integer polynomials satisfying that condition.

(c) Conclude that there are only finitely many $\alpha \in \mathcal{O}_K$ satisfying $|\sigma_i(\alpha)| = 1$.

(d) Was it important that $|\sigma_i(\alpha)| = 1$ specifically, or could this have worked if we instead said $|\sigma_i(\alpha)| \leq M$ for some M?

Question 6.

(a) Show that $\ker L$ is finite.

(b) Show that if $\alpha \in \ker L$, then α has finite order.

(c) Conclude that if $\alpha \in \ker L$, then α is a root of unity.

Question 7. Our final step in this section is to argue that $\operatorname{im} L$ is a lattice.

(a) We probably want to make use of a norm on $\mathbb{R}^{r+s}$ in order to make topological arguments. What norm seems most appropriate for interacting with our other information? (They're all equivalent, so in a sense it doesn't matter)

(b) Explain why it's sufficient to check that $\operatorname{im} L$ doesn't cluster near 0.

(c) Fix some arbitrary $\delta > 0$, and suppose $\|L(\alpha)\| < \delta$ with respect to your chosen norm. What does this tell you about $|\sigma_i(\alpha)|$? (The right norm might make seeing things easier here.)

(d) Show that only finitely many α can satisfy the condition from the previous part.

(e) Explain why this shows that $\operatorname{im} L$ does not cluster near 0.

$\mathcal{O}_K^\times$ has enough units

Now we want to show that we can find $r + s - 1$ independent units. It will help to think in the very concrete setting of a real quadratic field.

Question 8. Suppose $K = \mathbb{Q}(\sqrt{d})$ where d is a positive, squarefree integer and $d \not\equiv 1 \pmod 4$. Make a rough sketch of $E(\mathcal{O}_K)$ the Euclidean embedding. On the same sketch, include a sketch of the locus where $N(\alpha) = 1$.

We want to produce lattice points on the $N(\alpha) = 1$ locus. Our main tool for producing lattice points is Minkowski's theorem, so let's try using Minkowski's theorem.

Question 9. What is the smallest area of a Minkowski set for $\mathcal{O}_K$ in terms of d? Can you think of a Minkowski set T with that area which minimizes the maximum norm among points in T? (I.e. minimizes $\max_{\mathbf{x}\in T} N(\mathbf{x})$.)

Unfortunately, because the Minkowski set T we need to take is so big, it might capture lattice points which are not units. We could try to get around this by deforming our Minkowski set so that it's more likely to capture units rather than non-units, for example by gradually making it taller and skinner so that there's just a little bit of area poking out of the $N(\alpha) = 1$ locus. For any $\lambda \in \mathbb{R}^+$, we can obtain a new Minkowski set by shrinking T by λ in the x direction and multiplying it by λ in the y direction. This operation preserves the area of T, so we are guaranteed to still be able to capture a lattice point. Call this new region $T_\lambda \coloneqq (\lambda^{-1}, \lambda) \cdot T$.

Question 10.

(a) Is it possible that the same nonzero lattice point lies in T_λ for all λ?

(b) Explain why each lattice point captured has bounded norm.

It is not true that there are only finitely many elements of bounded norm; indeed, once we know there are infinitely many units, that gives us infinitely many elements of norm 1. However, it is true that there are only finitely many *ideals* of bounded norm, and each lattice point γ that we find generates some (possibly unit) ideal. Thus, we have only finitely many γ_k generating different ideals, and once we find a repetition $(\gamma) = (\gamma_k)$, then $\varepsilon = \gamma/\gamma_k$ will be a unit different from ± 1.

Question 11. In the argument above, was it important that T be any particular Minkowski set? Could we allow T to be an arbitrary Minkowski set?

In the general case, we will instead be looking inside of $\mathbb{R}^r \times \mathbb{C}^s$. We don't merely want to find one unit, but $r + s - 1$ independent units. In the real quadratic case, we varied the region T by stretching along one axis and shrinking along the other. Since we have $r + s$ axes in the general case, it seems reasonable to do the same: multiply by some $\mathbf{x} \in \mathbb{R}^r \times \mathbb{C}^s$ with $N(\mathbf{x}) = 1$ in such a way as to stretch T along one axis while shrinking all the rest, then repeat for each axis. Since the units we find this way will be "spiky along different axes", it seems reasonable that they will end up being independent. (Except the last one, obviously, which must be dependent.)

To formalize this idea a bit, specifically we would like units ε_i, $1 \le i \le r+s-1$, so that $|\sigma_i(\varepsilon_i)|^{n_i} > 1$ and $|\sigma_j(\varepsilon_i)|^{n_i} < 1$ for $j \neq i$.

Question 12. Give an informal explanation for why ε_i satisfying the above criteria should be independent.

To accomplish this, we'll need a little knowledge of how much the units we find can differ from the stretch factor we use.

Question 13.

(a) Show that when we find a repetition and hence produce a unit $\varepsilon = \gamma/\gamma_k$, the coordinates $\mathbf{x}^{-1}\varepsilon$ are uniformly bounded independent of $\mathbf{x}$.

(b) Show that we can in fact find ε_i so that $|\sigma_i(\varepsilon_i)|^{n_i} > 1$ and $|\sigma_j(\varepsilon_i)|^{n_i} < 1$ for $j \neq i$, as desired.

The final step is to pass once again along the logarithmic map and consider $L(\varepsilon_i)$. Since the sum of the coordinates is 0, we'll simply drop the last coordinate. Thus, we are now in the setting where the vectors $\mathbf{a}_i = L(\varepsilon_i)$ satisfy $a_{ij} < 0$ for $j \neq i$ and $\sum_j a_{ij} > 0$. (We can argue directly without passing to the logarithmic map, but the following result about matrices is independently neat.)

Question 14. Throw the $\mathbf{a}_i$ as the columns of a matrix A. Show that $\ker A = 0$, and hence that A has independent columns.

Question 15. Did any part of our arguments require that we were working with $\mathcal{O}_K$ specifically? Or would it have worked with any order?

Follow-up questions

Question 16. Consider the setting where A is a matrix satisfying $a_{ij} < 0$ for $j \neq i$ and $\sum_j a_{ij} > 0$. Rescale A so that the diagonal entries are 1. Here's a random connection to the Perron-Frobenius theorem I made.

(a) Write $A = I - tP$, where P is a non-negative matrix and $t \in \mathbb{R}^+$ is an arbitrary scalar. (Think of t being small so that the entries of P are normal-sized.) What is $\det A$ in terms of P?

(b) Interpret the condition $\sum_j a_{ij} > 0$ in terms of P and t.

(c) Now flip things around: fix any positive matrix P and let $A = I - tP$. Use the fact that we already know $\det A \neq 0$ (under certain conditions on t) to find a lower bound on the least positive eigenvalue of P.

Finiteness of the Class Group

Guiding Question

In $\mathbb{Z}$, every integer factors uniquely as a product of prime numbers. In number rings, this need not be the case; famously, in $\mathbb{Z}[\sqrt{-5}]$ we have $6 = 2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5})$, and all of $2, 3, 1+\sqrt{-5}, 1-\sqrt{-5}$ are irreducible. However, in number rings we still have unique factorization *for ideals*. The class group of a number field is a group which tells us how badly this factorization of ideals can fail to translate into factorizations of numbers. Given that we cannot always have unique factorization in a number ring, the next best thing might be to know that the class group is always finite. (Unique factorization cannot fail "infinitely badly".) This is what we aim to prove.

Preliminary Questions

Question 1. The ideal class group is defined as the group of fractional ideals modulo the subgroup of principal fractional ideals.

(a) Restate what it would mean for the class group to be finite.

(b) How do you think the proof might go? Can you think of any general strategies for showing that something is finite? Can you think of any specific tools in the case of number fields that might get used?

Question 2. A couple reminders about fractions and fractional ideals. Recall that in a number ring, all the local rings are DVRs.

(a) Show that any fraction $\alpha/\beta \in K$ can be represented as γ/n for some $\gamma \in \mathcal{O}_K$, $n \in \mathbb{Z}$.

(b) Let $\mathfrak{f}$ be a fractional ideal. Show that $\mathfrak{f}_\mathfrak{p}$ is a free module of rank 1 over $\mathcal{O}_{K,\mathfrak{p}}$.

(c) Let $x \in \mathfrak{f}$. Show that $(\mathfrak{f}/\langle x \rangle)_\mathfrak{p}$ is the zero module except for at finitely many $\mathfrak{p}$.

(d) Show that there is some $y \in \mathfrak{f}$ which generates $\mathfrak{f}_\mathfrak{p}$ at the finitely many $\mathfrak{p}$ from the previous part. Conclude that $\mathfrak{f}$ is generated by x and y as an $\mathcal{O}_K$-module.

Question 3. Norms are how we measure the "size" of elements and ideals in number fields, so let's review some of their properties. Let K be a number field. Recall that the (absolute) norm of $\alpha \in K$, $N(\alpha)$, is any of the following equal quantities:

- The product of α and all its Galois conjugates.
- $(-1)^{[K:\mathbb{Q}]}a_0$ where a_0 is the constant term of the minimal polynomial of α.
- The determinant of the $\mathbb{Q}$-linear transformation $K \to K$ given by multiplication by α.
- The product of $\sigma(\alpha)$, taken over all embeddings σ of K into $\mathbb{R}$ or $\mathbb{C}$.

Given an ideal $\mathfrak{a} \subset \mathcal{O}_K$, the norm of the ideal $\mathfrak{a}$ is $N(\mathfrak{a}) = \#\mathcal{O}_K/\mathfrak{a}$.

(a) Verify in your head that the norm is multiplicative: $N(\alpha\beta) = N(\alpha)N(\beta)$.

(b) Show that $|N(\alpha)| = N((\alpha))$.

(c) For an ideal $\mathfrak{a}$, confirm that $\mathcal{O}_{K,\mathfrak{p}}/\mathfrak{a}_\mathfrak{p}$ is a finite ring for all primes $\mathfrak{p}$, and that this ring is the trivial ring if $\mathfrak{p} \nmid \mathfrak{a}$.

(d) Show that $N(\mathfrak{a}) = \prod_\mathfrak{p} N(\mathfrak{a}_\mathfrak{p})$.

(e) Show that ideal norm is multiplicative: $N(\mathfrak{ab}) = N(\mathfrak{a})N(\mathfrak{b})$.

Question 4. In what follows, it will be helpful to know a way to relate the discriminant of a number field to its covolume in the Euclidean embedding.

(a) Consider $K = \mathbb{Q}(\sqrt{-7}), \mathcal{O}_K = \mathbb{Z}[\frac{1+\sqrt{-7}}{2}]$. Draw a picture of the lattice $\mathcal{O}_K \subset \mathbb{C}$.

(b) Write down a matrix whose determinant gives the covolume of this lattice.

(c) Write down a matrix whose determinant gives the discriminant of $\mathcal{O}_K$.

(d) Do you see a way to turn the latter matrix into the former?

(e) Redo parts (a)-(d) for $K = \mathbb{Q}(\sqrt{2}), \mathcal{O}_K = \mathbb{Z}[\sqrt{2}]$. What differences do you notice?

(f) How might this generalize to other number fields? Can you prove it?

[section title]

Our goal is to show there can only be finitely many distinct ideal classes. One idea is to define a "size" for a given ideal, and show that each ideal class has a representative of bounded size. But first of all, working with fractional ideals is a bit obnoxious. Furthermore, we aren't even principallly interested in fractional ideals, they were just introduced as a way to turn the set of ideals into a group. So, we would like to restrict our task to integral ideals.

Question 5. Explain why for any fractional ideal $\mathfrak{f}$, there exists a principal integral ideal (α) so that $(\alpha)\mathfrak{f} \subset \mathcal{O}_K$. (Indeed, α can be taken to be in $\mathbb{Z}$.)

As alluded to in the Preliminary Questions, a reasonable way to measure the size of an ideal is with its norm. So, following our above philosophy, we would like to show that every ideal class can be represented by an ideal of bounded norm, where the bound may depend on the field K but not on the specific ideal class chosen.

Question 6. Given an integral ideal $\mathfrak{a}$, restate what it means for the ideal class $[\mathfrak{a}]$ to have an integral representative with norm $\leq B$. Is it easy to state while insisting all ideals involved are integral?

Finding ideals of bounded norm seems difficult, but finding *elements* of bounded norm seems doable, so let's start there.

Question 7. Let's return once again to our examples from above.

(a) Return to your picture of the integers in $\mathbb{Q}(\sqrt{-7})$. What does the region of points with norm $\leq B$ look like in this picture?

(b) Turn now to $\mathbb{Q}(\sqrt{2})$. What does the region of points with norm $\leq B$ look like in this picture?

(c) Our main tool for constructing elements with given properties is Minkowski's theorem. Can we apply Minkowski to both of the above regions? If we can't, can you find a large Minkowski region contained within the region?

(d) How might this generalize to number fields with more real/complex embeddings?

Once we apply Minkowski, in order to get an actual bound on the size of the lattice point we obtain we will need to know the actual volume of the Minkowski region you found in Question 7(d).

Question 8. Use calculus to compute the measure. The number you compute is sometimes called the "Minkowski constant".

Now we can apply Minkowski. To what? If we're looking at the ideal class $[\mathfrak{a}]$, then the only lattice we have to hand is $E(\mathfrak{a})$. So let's try that.

Question 9. Let C be the constant from the previous problem. Show that an integral ideal $\mathfrak{a}$ contains an element a of norm at most $C\,\mathrm{covol}(E(\mathcal{O}_K))N(\mathfrak{a})$. Rewrite the covolume in terms of the discriminant to get a bound depending only on the field and $\mathfrak{a}$.

And now there's a small puzzle to figure out how to finish the argument. We have the ideal $\mathfrak{a}$ and we have the element $a \in \mathfrak{a}$ with relatively small norm, but the norm of a still depends on the $\mathfrak{a}$ we started with.

Question 10.

(a) Can we construct an ideal out of (a) and $\mathfrak{a}$ in a way that removes the dependence on $\mathfrak{a}$?

(b) Once you do, can you see how to finish the argument?

Hints

Question 2(d): Let y_P be local generators for each P, and let $y = \prod_P y_P$ (the product takes place in K, since everything is contained in K). Show that $\mathfrak{f} = (x, y)$.

Question 5: I can see a couple ways to approach this. One is to use Question 2. You could also use the fact that every prime ideal $\mathfrak{p}$ lies over, hence divides, some rational prime (p) to "clear the denominators" in $\mathfrak{f}$.

Question 7(d): Think especially hard about the case of two complex embeddings. What is the norm function? Is the $N(\alpha) \leq B$ region convex? How is this like the case of two real embeddings?

Question 10:

(a) It sure would be nice to *divide* $N(a)$ by $N(\mathfrak{a})$.

(b) The twist: the ideal you construct doesn't represent $\mathfrak{a}$ at all! What does it represent? Is that good enough?