Two proofs of the infinitude of primes

Ben Chastek
REU summer 2001 by Professor Garrett

This paper is the product of an eight week research program that was headed by Professor Paul Garrett. It was designed to give undergraduates the opportunity to experience research in mathematics. Originally I had wanted to learn about topology and algebraic number theory, but due to time constraints I found that I only could learn about the latter.

At the beginning of summer I wanted to be able to relate the math that I had already been taught to the math I was learning. I then noticed that I could relate the two through a proof of the infinitude of primes. This seemed ideal to me, as this was one of the many ideas i had learned earlier in the year. Although most of the ideas are explained, If I were to continue this in the future I would concentrate on a proof that i could not do; relating prime ideals to prime integers. With that I will begin with a proof that Euclid gave using aspects described in number theory.

Suppose that there are only finitely many primes, $p_1 \dots p_n$ where n is a finite number. Then let $N = p_1 p_2 \dots p_n + 1$. By unique factorization, N has a prime divisor, call it p. Then p cannot be any of $p_1 \dots p_n$, otherwise p would divide the difference $N - p_1 p_2 \dots p_n = 1$, which is impossible. Therefore p is still another prime and we conclude that there are infinitely many primes.

I will proceed with the following background information that is necessary in order to understand the basics about ring theory and field theory. It can be found in any book about Abstract Algebra. First is the definition of a group. A group G is defined to be a set with a binary operation $*$ having the following properties:
• a, b in G implies that $a * b$ is in G (closure)
• there exists an $e \in G$ such that $a * e = a = e * a$ for all $a \in G$ (identity)
• $(a * b) * c = a * (b * c)$ (associativity)
• for all $a \in G$ there exists $a^{-1} \in G$ such that $a * a^{-1} = e = a^{-1}a$. (Inverse)

In general a group does not needs to be commutative. A group that is commutative is said to be *abelian*, the defining property being
• For all $a, b \in G$, $a * b = b * a$.

The term 'operation' does not always refer to a familiar tangible operation. It can mean the regular operations of multiplication, addition ..., but it can also mean an abstract form. They may seem odd and without basis, but many times they can be helpful in group theory.

Some clear examples of groups (abelian in this case) are the usual integers, $\mathbf{Z}$, under addition, the rational numbers $\mathbf{Q}$ under addition, and the positive real numbers, R^+ under multiplication. It is important to note that while $\mathbf{Z}$ forms a group under addition it does not form a group under multiplication. This is shown by the lack of an inverse for some elements. For example, $2 \times ? = 1$.

Some abstract entities have further operations and properties. A ring is a non-empty set R, with two operations $+$ and $\cdot$, satisfying
• closure under $+$
• existence of an identity for $+$
• associativity of $+$
• existence of an inverse for $+$
• commutativity of $+$
Note that these five conditions require that the set is an abelian group under addition. In addition, the set R must have
• closure under multiplication
• associativity of multiplication
• distributivity: $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$ for all $a, b, c \in R$.

As with groups, a ring's multiplication does not need to be commutative. A ring for which $a \cdot b = b \cdot a$ is said to be a *commutative* ring (not an *abelian* ring).

A commutative ring R is an *integral domain* if for $a, b \in R$ $a \cdot b = 0$ implies that either $a = 0$ or $b = 0$. Although this may seem to be obviously always true, there are many examples (matrices, for example) where this fails.

A ring is a *division ring* if every non-zero element has a multiplicative inverse. A ring R is a *field* if it is a commutative division ring. A good example of a field is the set of rational numbers.

For a ring R, and a non-empty subset I of R, I is an *ideal* of R if :
• I is an additive subgroup of R.
• for $r \in R$, $a \in I$, $ra \in I$ and $ar \in I$.

It is important to note that in commutative rings $ar = ra$, and therefore there is no distinction between left, right, and two-sided ideals. In non-commutative rings this is not generally true and therefore ideals are said to be either left, right, or two-sided.

An integral domain R is a *principal ideal domain* if every ideal I is principle, that is, if $I = aR$ for some $a \in R$.

Now that we have a background of basic group, ring, and field theory, we can move on to algebraic number theory. Here we will learn about Dedekind domains, and how they relate to principal ideal domains. First is the theory of the *field of fractions*, which as its name may suggest is a field made up of fractions. The formal definition is stated thusly:

Suppose that R is a commutative integral domain. The field of fractions K of R is a field containing R so that every non-zero element $r \in R$ has an inverse (denoted $\frac{1}{r}$) in K. Every element of K can be written in the form $\frac{r}{s}$ for $r, s \in R$. This definition should seem particular familiar as it is the exact construction of $\mathbf{Q}$ from $\mathbf{Z}$. (Some things need to be checked to know that such K exists, but this is not hard, and is not terribly interesting.)

One of the basic ideas of algebraic number theory is the idea of an algebraic number. Algebraic numbers are defined to be roots of polynomial equations with rational coefficients. For example, $\sqrt{2}$ is an algebraic number because it satisfies the equation $x^2 - 2 = 0$. For that matter, all numbers expressible in terms of radicals are algebraic numbers. It is easy to see that every integer a is an algebraic number since it satisfies the equation $x - a = 0$.

Although some numbers are obviously algebraic just because of their form, in general it is not easy to express in simpler terms the roots of high-degree polynomial equations, and conversely it is often difficult to determine for which polynomial a complicated radical expression is a root. For example, it may be intuitively clear that $\sqrt[5]{3} + \sqrt{7}$ is an algebraic number, but it is not obvious to which polynomial equation it is a root. A number that is not expressible as a root of a polynomial (with rational coefficients) is called *transcendental*. Some common examples are π and e (the base of the natural logarithms).

An algebraic number that is a root to a *monic* equation with *integer* coefficients is called an *algebraic integer* or simply an *integer* for short. A *number field* K is a finite field extension of $\mathbf{Q}$.

Next is the definition of *module*. Like ideals, modules are left and right sensitive in general. For a ring R, a module, in this case a right module, is a set M which has addition and also has scalar multiplication by elements of R. That is to say that if m, n are in M and $r \in R$, then there are elements $m + n$ and $m \cdot r$ in M. M is an abelian group under addition, and the scalar multiplication must have the following properties.
• $m(rs) = (mr)s$ for all m in M and r, s in R (associativity)
• $(m + n)r = mr + nr$ and $m(r + s) = mr + ms$ for all m, n in M and r, s in R (distributivity)
• $m \cdot e = m$ for all m in M, where e is the identity element in R (property of the identity)

Not too surprisingly a *left module* is defined similarly, where in place of right scalar multiplication there is left multiplication. When dealing with modules it is often necessary to work with *submodules* and *generators*. For a given module M over a fixed ring R, a submodule of M is a subset M' of M such that:
• $0 \in M'$
• if $m, n \in M'$ then $m + n \in M'$
• if $m \in M'$ and $r \in R$, then $m \cdot r \in M'$.

Two obvious cases of sub-modules are are the 0-module, $0 = \{0\}$, and also M itself. When a submodule M' of M is neither 0 nor the whole module M it is called a *proper* submodule of M. Clearly a submodule is a module itself. Also, since R is a module over itself, then it is clear that an ideal $\mathbf{a}$ in R is an R-submodule of R. Sub-modules and modules are an important idea of algebraic number theory. They are also needed to explain *Noetherian-ness* and *integrality* which shall be discussed below.

In order to understand Dedekind domains it is important to give an explanation of integrality and Noetherian-ness. These two concepts deserve an extensive discussion in their own right, but a definition and an explanation will suffice here. First is integrality which is defined thusly: for $\mathbf{o}$ a ring and $\mathbf{O}$ a ring extension of $\mathbf{o}$, $a \in \mathbf{O}$ is *integral over* $\mathbf{o}$ if and only if there is a monic polynomial f in $\mathbf{o}[X]$ such that $f(a) = 0$. This definition is very similar to that of algebraic number. In fact when $\mathbf{o}$ is a field then 'algebraic over' and 'integral over' coincide.

An often useful theorem relating to integrality is that: An element a in $\mathbf{O}$ is integral over $\mathbf{o}$ if and only if the ring $\mathbf{o}$[a] is finitely generated as an $\mathbf{o}$ module. The proof of this can be found in Frohlich and Taylor on page 27.

A definition and a theorem may not be the best way to understand integrality. Therefore here is an example of how integrally closed-ness fails: $\mathbf{Z}[\sqrt{-3}]$. This is not integrally closed since the non-trivial root of unity $\frac{\omega=(-1+\sqrt{-3})}{2}$ lies in the field of fractions $\mathbf{Q}(\sqrt{-3})$ and is integral over $\mathbf{o}$ because it satisfies the polynomial $X^2 + X + 1 = 0$.

Now that integrality has been established is is necessary to define Noetherian-ness. An $\mathbf{o}$ 0-module M is called noetherian if all of its submodules are finitely generated over $\mathbf{o}$. The ring $\mathbf{o}$ is a noetherian ring if $\mathbf{o}$ is a noetherian $\mathbf{o}$-module, i.e. if all the ideals of $\mathbf{o}$ are finitely generated over $\mathbf{o}$. A great example of noetherian-ness is that a principle ideal domain is automatically a noetherian ring. In addition, any finite ring is a noetherian ring, and any finite module is a noetherian module.

Making use of the earlier definition of a module, I will now define noetherian-ness in the context of modules, in particular in terms of ascending chains of modules. This condition puts a 'limit' on the modules of a ring. Let $M_1 \subset M_2 \subset \ldots$, denote an ascending sequence of $\mathbf{o}$-modules. The chain is said to stabilize if there is an integer k such that $M_j = M_k$ for all $j \geq k$. In addition to the ascending chain condition, a module M is said to satisfy the maximal condition if any non-empty partially ordered set of submodules of M has a maximum.

A theorem that goes with the conditions of noetherian are: Let M be a module. then the following are equivalent:
• M is noetherian
• M satisfies the ascending chain condition
• M satisfies the maximal condition

Proof: 1 implies 2: Let $\{M_i\}$ be an ascending chain of modules as above. Then $\bigcup M_i$ is a submodule of M, so it is finitely generated. Choose k large enough so that the generators of $\bigcup M_i$ are in M_k, so the chain ends there.

2 implies 3: By contradiction assume that X is a partially ordered set of submodules that does not have a maximum. Assume that for some $i \geq 1$ there is a chain of modules $M_1 \subset \ldots \subset M_i$ of members of X, with

proper inclusions. Since M_i is not maximal there is a further member and that then does not stabilize, and there is an infinite ascending chain in M. This is a contradiction and therefore 2 implies 3.

3 implies 1: Given a submodule N of M, let X be the set of finitely generated submodules of N. Clearly, X is not empty, so it has a maximum L. if $L \neq N$, there is some n in N with n not in L but then $L + nR \in X$, but this contradicts the maximality of L so $L = N$, and then N is finitely generated and therefore it is noetherian.

Before proceeding with Dedekind domains it is important to understand unique factorization. For example we often think of the ordinary integers as having unique factorization. In fact this has been proven for a long time. In the integers a number like 6 is factored into to prime elements 2 and 3. In this case 2,3 are prime elements because they have no non trivial factorization. For a more formal definition we say that an element v is irreducible if there is no factorization $v = wx$ of v in R except for the trivial factors, meaning that either w or x is a unit (has a multiplicative inverse) in R.

Unique factorization does not always hold, and in fact in a Dedekind domain that is many times the case. In Dedekind domains the idea of ideals allows the factorization of ideals into prime elements. Therefore a definition is called for. A Dedekind domain, named after the mathematician who first studied them, is an integral domain, $\mathbf{o}$, for which the following three conditions are true:

1) $\mathbf{o}$ is a noetherian Ring 2) $\mathbf{o}$ is integrally closed in its field of fractions K 3) all non-zero prime elements of $\mathbf{o}$ are maximal ideals

For an ideal to be maximal means that the ideal is proper, and if contained inside of another proper ideal, then the two ideals are equal. Otherwise stated formulaic-ally: For $\mathbf{a} \neq 0$ and $\mathbf{a} \subset \mathbf{b} \subset \mathbf{o}$ for an ideal $\mathbf{b}$ implies $\mathbf{a} = \mathbf{b}$.

It is also important to note that Dedekind domains have many of the same desired properties as do the integers $\mathbf{Z}$.

A theorem that will come in very handy is

Theorem: Any principle ideal domain (PID) is a Dedekind domain.

Proof: We must show that a PID has all three of the characteristics mentioned above. As has already been stated any PID is automatically noetherian. Therefore number 1 is trivial. Next we will show that all of the non-zero prime ideals are maximal. Let $\mathbf{a} = a \cdot \mathbf{o}$ denote a prime ideal of $\mathbf{o}$, and also let $\mathbf{b} = b \cdot \mathbf{o}$ be a maximal ideal which contains $\mathbf{a}$. Therefore we have that $a = bc$ for some c in $\mathbf{o}$. Since $\mathbf{a}$ is a prime ideal, either $b \in \mathbf{a}$ and so $\mathbf{a} = \mathbf{b}$ as was needed to be shown, or else $c = ad$ for some d in $\mathbf{o}$. If the second is true then we have that $a = abd$, so that b in $\mathbf{o}^*$ and also b^{-1} in $\mathbf{o}$ which is a contradiction and therefore all prime ideals are maximal. Finally let $\frac{a}{b}$ be integral over $\mathbf{o}$, with $a, b \in \mathbf{o}$, $b \neq 0$, and $\gcd(a, b) = 1$. Then $\frac{a}{b}$ is a root of a monic polynomial with some coefficients $c_i \in \mathbf{o}$, by definition. That is,

$$\left(\frac{a}{b}\right)^n + c_{n-1}\left(\frac{a}{b}\right)^{n-1} + \ldots + c_1\left(\frac{a}{b}\right) + c_0 = 0$$

and by multiplying out the denominators

$$a^n + bc_{n-1}a^{n-1} + \ldots + b^{n-1}c_1 a + b^n c_0 = 0$$

this tells us that b divides a^n. Since $\gcd(a, b) = 1$, it follows that b is in $\mathbf{o}^*$, and therefore $\frac{a}{b} \in \mathbf{o}$. Thus $\mathbf{o}$ is integrally closed in its field of fractions. It is therefore shown that all PID's are Dedekind domains.

Theorem: : In every number field of finite degree the ring $\mathbf{o}$ of algebraic numbers is a Dedekind domain. In addition to this every non-zero ideal of $\mathbf{o}$ can be uniquely factored into prime ideals.

A proof of the first part can be found in any book on algebraic number theory. I will give the proof of the second part. It is very similar to the proof that Euclid gave to prove the unique factorization of the integers into prime numbers.

Proof: Note that every non-zero ideal $\mathbf{a}$ is a product of prime ideals. Indeed, if this were false, then there is a maximal ideal $\mathbf{a}$ not such a product, and cannot be prime. Thus $\mathbf{a} \subset \mathbf{p}$ and $a \neq \mathbf{p}$ for some prime $\mathbf{p}$. Then $\mathbf{ap}^{-1} \subset \mathbf{o}$ and $\mathbf{ap}^{-1} \neq \mathbf{a}$ but contains $\mathbf{a}$. Hence $\mathbf{ap}-1$ has a factorization, which when multiplied by $\mathbf{p}$ gives a factorization of $\mathbf{a}$.

Given two fractional ideals $\mathbf{a}$, $\mathbf{b}$ we say that $\mathbf{a}$ divides $\mathbf{b}$ if and only if there exists an ideal $\mathbf{c}$ such that $\mathbf{ac} = \mathbf{b}$. This amounts to $\mathbf{a} \subset \mathbf{b}$, because in that case, we take $\mathbf{ca}^{-1}\mathbf{b}$.

From the definition of prime ideal, we see that whenever $\mathbf{a}, \mathbf{b}$ are two ideals and $\mathbf{a}$ divides $\mathbf{ab}$ then $\mathbf{p}$ divides $\mathbf{a}$ or $\mathbf{p}$ divides $\mathbf{b}$. Then given two factorizations

$$\mathbf{p}_1\mathbf{p}_2 \dots \mathbf{p}_r = \mathbf{q}_1\mathbf{q}_2 \dots \mathbf{q}_s$$

into prime ideals, we conclude that $\mathbf{p}_1$ divides the product on the right, hence divides some $\mathbf{q}_i$, that is to say equal to some $\mathbf{q}_i$. Multiplying by $\mathbf{p}^{-1}$ on both sides of the equality, proceed with induction to prove that $r = s$ and that the factors on both sides are equal up to a permutation.

If $\mathbf{a}$ is a fractional ideal $\neq 0$ and $c \in \mathbf{o}$ is such that $c \neq 0$ and $c\mathbf{a} \subset \mathbf{o}$ then $c = \mathbf{p}_1 \dots \mathbf{p}_r$ and $c\mathbf{a} = \mathbf{q}_1 \dots \mathbf{q}_s$. Therefore $\mathbf{a}$ has the factorization

$$\mathbf{a} = \frac{\mathbf{q}_1 \dots \mathbf{q}_s}{\mathbf{p}_1 \dots \mathbf{p}_r}$$

If we cancel any prime appearing both in the numerator and in the denominator, then it is clear that the factorization is unique.

Theorem: In every number field of finite degree there are only finitely many prime ideals that divide any given (ordinary) prime number p.

This is clearly an important theorem as it relates the prime ideals to prime numbers, that up until now had been separated. However, the proof of this is rather difficult and will be left out. Instead I proceed to

Theorem: A Dedekind domain with only finitely many prime ideals is a principle ideal domain, and as such, every non-zero element is, up to units, the product of prime elements in a unique way.

Proof: Let $\mathbf{p}_1 \dots \mathbf{p}_s$ be the prime ideals. Given any ideal

$$\mathbf{a} = \mathbf{p}_1^{R_1} \dots \mathbf{p}_s^{r_s} \neq 0$$

select an element x in $\mathbf{p}_i$ but not in $\mathbf{p}_i^2$ and find an element α of $\mathbf{o}$ such that

$$\alpha = {x_i}^{r_i} mod({p_i}^{r_i+1})$$

If

$$\alpha = \mathbf{p}_1^{e_1} \dots \mathbf{p}_s^{e_s}$$

is a factorization of the ideal generated by α, then it is obvious that $e_i = r_i$, for all i, and that $\mathbf{a} = \alpha$.

It is now time to give Larry Washington's proof of the infinitude of prime numbers. This proof relates back to the theorems of unique factorization, principle ideal domains, and the number of prime ideals that divide a prime number.

Consider the field of numbers $a + b\sqrt{-5}$, such that a, b are in $\mathbf{Q}$. The ring $\mathbf{o}$ of algebraic integers in this field consist of the numbers of the above form, with a, b in $\mathbf{Z}$. It is easy to verify that $2, 3, 1 + \sqrt{-5}, 1 - \sqrt{-5}$

are prime elements of this ring, since they cannot be decomposed into factors that are algebraic integers, unless one of the factors is 1 or -1. Also note that 6 can be written in the form $6 = 2 \cdot 3$ and also $6 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5})$. This means that factorization is not unique (up to units). Therefore by a previous theorem $\mathbf{o}$ is not a principle ideal domain. It therefore must have infinitely many prime ideals. Since each (ordinary) prime number is divisible by a finite number of prime ideals in $\mathbf{o}$, there must be infinitely many prime numbers.

We have now reached the conclusion that there are infinitely many primes, and it has been done so in two different ways.

Bibliography:

Berrick, A.J.; Keating M.E.; An Introduction to Rings and Modules. Cambridge University Press, 2000

Frohlich, A; Taylor, M.J. Algebraic Number Theory. Cambridge University Press, New York, 1991.

Janusz, Gerald J. Algebraic Number Fields second edition. American Mathematical Society, 1996.

Lang, Serge. Algebraic Number Theory. Springer-Verlag, New York, 1986

Ribenboim, Paulo. The Little Book of Big Primes. Springer-Verlag, New York, 1991.