# Notes on Ring Theory

*by Avinash Sathaye, Professor of Mathematics*
February 1, 2007

# Contents

# 1 Ring axioms and definitions.

**Definition: Ring** We define a ring to be a non empty set $R$ together with **two binary operations** $f, g : R \times R \Rightarrow R$ such that:

1. $R$ is **an abelian group** under the operation $f$.
2. The operation $g$ is associative, i.e. $g(g(x, y), z) = g(x, g(y, z))$ for all $x, y, z \in R$.
3. The operation $g$ is distributive over $f$. This means:

$$g(f(x, y), z) = f(g(x, z), g(y, z))$$

and

$$g(x, f(y, z)) = f(g(x, y), g(x, z))$$

for all $x, y, z \in R$.

Further we define the following natural concepts.

1. **Definition: Commutative ring.** If the operation $g$ is also commutative, then we say that $R$ is a commutative ring.
2. **Definition: Ring with identity.** If the operation $g$ has a two sided identity then we call it the identity of the ring. If it exists, the ring is said to have an identity.
3. **The zero ring.** A trivial example of a ring consists of a single element $x$ with both operations trivial. Such a ring leads to pathologies in many of the concepts discussed below and it is prudent to assume that our ring is not such a singleton ring. It is called the "zero ring", since the unique element is denoted by 0 as per convention below.

   **Warning:** We shall always assume that our ring under discussion is not a zero ring. However, when we construct a new ring from a given ring, we need to make sure that we have not created the zero ring.
4. **Conventions.** To simplify notation, we write $x + y$ in place of $f(x, y)$. The corresponding identity element is naturally denoted by 0. The operation is simply called the addition operation in the ring.

   We shall also replace $g(x, y)$ by $x \cdot y$ or simply $xy$, if there is no confusion. We naturally call the resulting operation the multiplication in $R$.

   The corresponding identity of multiplication, if it exists, is naturally denoted by 1.[1]

[1] Some textbooks make a term "rng" to denote a ring possibly without the unity, whereas the term "ring" is reserved for rings with unity. The idea is that the "i" in the spelling stands for the unity! While cute, this is useless since it is hard to say "rng"!

5. Note that there is a good reason not to make a fuss about the additive identity 0, since it always exists. One sometimes distinguishes the element 1 by calling it the multiplicative identity.

6. It can be shown that for rings with identity, the distributive law forces the operation $f$ to be commutative and hence our assumption of "abelianness" is a natural one. [2]

7. **Definition: Zero divisor** An element $x \in R$ is said to be a zero divisor if $x \neq 0$ and there is some nonzero $y \in R$ such that $xy = 0$ or $yx = 0$. Sometimes people name these two possible cases ($xy = 0$ or $yx = 0$) as conditions for a left or right zero divisors. We shall not emphasize this.

   It is important to **note** that a zero divisor is never zero! [3]

8. A related concept for the identity 1 is:

   **Definition: Unit in a ring.** An element $x \in R$ is said to be a unit if $xy = yx = 1$ for some $y \in R$.

   The set of units of a ring $R$ is denoted by $R^{\times}$.

   Note that in contrast with the zero divisor concept, the element 1 is counted as a unit.

   It is easily seen the the set $R^{\times}$ is a group under multiplication.

9. **Definition: divisibility in a ring.** We say that an element $x$ of a ring divides $y$ if $y = xz$ for some $z$ in the ring. In symbols we write $x|y$ and we may also say $y$ is divisible by $x$ or that $x$ is a factor of $y$.

   Thus units clearly divide every element of the ring. In a commutative ring, it is easy to show that every factor of a unit is a unit.

10. Note that the set of units and the set of zero divisors are disjoint.

    To see this, let $x$ be a unit with $xy = yx = 1$. If $xz = 0$ then $yxz = y0 = 0$ but at the same time, $yxz = (1)z = z$. Thus $xz = 0$ implies $z = 0$ and this proves that $x$ is not a zero divisor. The case when $zx = 0$ is similar.

[2] To see this, simply expand $(1 + 1)(x + y)$ in two different ways to get:

$$x + y + x + y = x + x + y + y$$

and deduce $y + x = x + y$ by cancelling $x$ from left and $y$ from right.

[3] It is tempting to define a zero divisor as any divisor (or factor) of zero. But then 0 will always be such a zero-divisor and we will need a term "proper zero divisor" to discuss the interesting non zero factors of zero. We have chosen the current definition to avoid adding the word proper every time!

On the other hand, assume that $z$ is a zero divisor, so that $z \neq 0$ and $zw = 0$ for some non zero $w$. We prove that $z$ is not a unit. If $xz = 1$ for some $x$, then $xzw = (1)w = w \neq 0$, but $xzw = x(0) = 0$, a contradiction. The case when $wz = 0$ is similar.

11. One of the most useful type of a ring is defined next.

    **Definition: Integral domain.**

    A ring $R$ is said to be an integral domain if it is **commutative**, **contains $\mathbf{1 \neq 0}$** and has **no zero divisors.** In an integral domain, we have cancellation for multiplication, namely $xy = xz$ implies $x = 0$ of $y = z$. To see this, simply rewrite the equation as $x(y - z) = 0$ and use the condition on zero divisors.

    **Definition: Field.**

    An integral domain $R$ is said to be a field if its non zero elements are units, i.e. $R^\times = R \setminus 0$. Later on we will see how any integral domain can be enlarged to a field called its quotient field.

    Many examples of fields are well known, $\mathbb{Q}$ the field of rational numbers, $\Re$ the field of real numbers, $\mathbb{C}$ the field of complex numbers and $\mathbb{Z}_p$ the finite field with $p$ elements when $p$ is a prime.

    The field $\mathbb{Q}$ is the so-called quotient field of $\mathbb{Z}$.

    **Definition: Division ring.**

    A ring $R$ is said to be a "division ring" if the condition $R^\times = R \setminus 0$ holds. Thus, we can define a field as a commutative division ring.

    One of the best examples of a division ring is the ring of real Hamilton Quaternions:

    $$\mathbb{H} = \{a + bi + cj + dk | a, b, c, d \in \Re\}$$

    where the products are defined by

    $$i^2 = j^2 = k^2 = -1 \text{ and } ij = k = -ji, jk = i = -kj, ki = j = -ik.$$

    Verify that

    $$(a + bi + cj + dk)(a - bi - cj - dk) = a^2 + b^2 + c^2 + d^2$$

    and deduce that we indeed have a division ring!

12. **Definition: Subring.** If $R$ is a ring and $S \subset R$, then $S$ is said to be a subring of $R$ if $S$ is a ring under the operations induced from $R$. It

will follow that the 0 from $R$ belongs to $S$, but the identity 1 may or may not belong to $S$.

Clearly, a subring $S$ forms an additive subgroup of $R$ and has to be closed under multiplication.

Thus the set $3\mathbb{Z} = \{3n|n \in \mathbb{Z}\}$ is a subring of $\mathbb{Z}$ which does not contain the identity. In fact, if a subring of $\mathbb{Z}$ contains 1, then it is obvious that it coincides with $\mathbb{Z}$.

13. For groups, we could use certain subgroups to create quotient groups, namely the normal subgroups. For rings, we can make quotient rings if our subrings are "ideal".!

    **Definition: Ideal.** A subset $I$ of a ring $R$ is said to be a left ideal if

    - $I$ is an additive subgroup of $R$ and
    - for every $a \in I$ and $x \in R$ we have $xa \in I$. This can be shortened as $RI \subset I$.

    A right ideal is defined similarly, where we require $IR \subset I$.

    The subset $I$ is said to be an ideal if it is both a left and a right ideal (or the so-called two-sided ideal). Clearly, for commutative rings, we only need to use the left ideal conditions above.

14. Let $R$ be a ring and $I$ an ideal in $R$. We define the quotient ring $R/I$ as follows.

    **Definition: Quotient ring.**

    Consider the set of additive cosets

    $$S = \{x + I|x \in R\}.$$

    Since $I$ is clearly a normal subgroup of the additive group $R$, we already know that $S$ is a well defined (abelian) quotient group. We define the multiplicative structure on $S$ by:

    $$(x + I)(y + I) = xy + I.$$

    It is easy to check that this is well defined and defines a ring. [4]

[4] The hardest part is to check that this is well defined. Thus, let $x + I = p + I$ and $y + I = q + I$, i.e. $x = p + p_1$ and $y = q + q_1$ where $p_1, q_1 \in I$. Then

$$xy = pq + pq_1 + qp_1 + p_1q_1$$

where we note that the last three terms are in $I$ by definition of an ideal. Hence $xy + I = pq + I$. Note how both left and right conditions on an ideal are used.

The rest of the check is easy.

It is customary to denote this quotient ring as simply $R/I$ without introducing a new symbol $S$.

15. As in groups, we define a homomorphism or a map which respects the ring structures. All definitions are similar except the the multiplicative structure is also invoked.

    Let $\phi : R \to S$ be a map of rings. We say that $\phi$ is a (ring-)homomorphism if for all $x, y \in R$:

    - $\phi(x \pm y) = \phi(x) \pm \phi(y)$ and
    - $\phi(xy) = \phi(x)\phi(y)$.

    The image of the homomorphism is the total image $\phi(R) = \{\phi(x)|x \in R\}$ and it is easily seen to be a subring of $S$.

    The kernel of the homomorphism is the set of all elements mapping to 0, i.e.
    $$Ker(\phi) = \{x \in R|\phi(x) = 0\}.$$

    It is easy to check that $Ker(\phi)$ is a (two-sided) ideal of $R$.

    Following the terminology in groups, we see that the homomorphism $\phi$ is injective if and only if $Ker(\phi) = \{0\}$ and it is surjective if an only if $S = \phi(R)$. As before, an isomorphism is a homomorphism which is injective and surjective.

    If we set $I = Ker(\phi)$, then the map $\psi : R/I \to \phi(R)$ defined by $\psi(x + I) = \phi(x)$ is easily seen to be an isomorphism.

    Thus, we have the basic identity
    $$\phi(R) \cong R/Ker(\phi).$$

# 2 Examples of rings.

We now list several important examples of rings which will be studied in greater details later.

1. **Rings derived from integers.** A lot of insight in the rings comes from the basic ring of integers $\mathbb{Z}$. It is indeed an integral domain with many special properties.

   The finite rings $\mathbb{Z}_n$ derived from $\mathbb{Z}$ give basic examples of finite commutative rings. In fact, in $\mathbb{Z}_n$ the identity 1 is written as $[1]_n$ in our convention and has the property that $r \cdot [1]_n$ or the sum of $r$ terms $[1]_n + [1]_n + \cdots$ is $0 = [0]_n$ exactly when $r$ is a multiple of $n$.

We make a:

**Definition: Characteristic of a ring.** A ring $R$ with 1 has characteristic $n$ if $n$ is the first positive integer for which $1+1+\cdots$ $n$ terms $= 0$. In case, such an $n$ does not exist, the characteristic is said to be 0.

A simpler description of the characteristic is as follows. Define a homomorphism $\phi : \mathbb{Z} \to R$ by $\phi(1) = 1_R$ where we have made up the notation $1_R$ for the identity in $R$. The $Ker(\phi)$ is then an ideal in $\mathbb{Z}$ and it is easy to show that any ideal in $\mathbb{Z}$ is simply of the form $d\mathbb{Z}$ where $d$ is some non negative integer. Indeed, $d$ is simply the GCD of all members of the ideal!

The characteristic of the ring $R$ is $n$ if $Ker(\phi) = n\mathbb{Z}$. It is also clear that the subring $\phi(\mathbb{Z})$ of $R$ is isomorphic to $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$.

2. **Polynomial rings and related costructions.** We shall give a rather general way of constructing rings from base rings which includes the construction of polynomial and power series rings.

   Let $G$ be an abelian group. We say that $G$ is an ordered abelian group if $G$ is a disjoint union of sets denoted as $G_+, G_-, 0$ such that:

   (a) An element $x \in G$ is in $G_+$ if and only if $-x \in G_-$. In particular, $G$ has no elements $x$ with $x + x = 0$, except 0.

   (b) If $x, y \in G_+$, then $x + y \in G_+$ and similarly, if $x, y \in G_-$ then $x + y \in G_-$. In fact, the condition on $G_-$ can be deduced from the condition on $G_+$.

   (c) This defines an order on $G$, namely we define $x < y$ if an only if $y - x \in G_+$.

   **An abstract construction.**

   Let $R$ be a commutative ring with 1 and consider the set of all functions from $G$ to $R$ denoted by $R^G$. For any function $f \in R^G$ we shall denote its support by $Supp(f) = \{x \in G | f(x) \neq 0\}$.

   Define $WO(R, G)$ to be the set of all functions in $R^G$ whose support is well ordered (under the given order on $G$).

   We make $WO(R, G)$ into a ring by defining componentwise addition and multiplication by the Cauchy product defined thus:

   If $f, g \in WO(R, G)$, then the product $fg \in WO(R, G)$ is defined by

$$fg(x) = \sum_{y \in Supp(f), z \in Supp(g), x = y + z} f(y)f(z).$$

Some serious work is involved in proving that the sum has only finitely many terms and the resulting product also has well ordered support.

We shall leave further details of these concepts for private contemplation for now. We only illustrate a few special cases which are fundamental to the ring theory.

- **Polynomial ring.**

  If we take the ordered abelian group as $\mathbb{Z}$ and let $S$ be the set of non negative integers, then we can define the polynomial ring (in one variable over $R$) as

  $$\{f \in R^{\mathbb{Z}} | Supp(f) \text{ is a finite subset of } S\}.$$

  If we consider a function $h \in R^{\mathbb{Z}}$ defined by $h(1) = 1$ and $h(n) = 0$ if $n \neq 1$ and for convenience denote it by $X$, then it is easy to see that our ring contains all powers $X, X^2, X^3, \cdots$ and we simply get the set

  $$\{\sum_{i=0}^{m} a_i X^i | \text{ where some } m \in S \text{ and } a_i \in R\}.$$

  This is usually denoted as $R[X]$ where $X$ is identified with the special function. It is called the polynomial ring in one variable over $R$ (or with coefficients in $R$).

  Having defined a polynomial ring in one variable, we may use it as a base ring and create a polynomial ring in two variables as $R[X][Y]$ which is briefly written as $R[X, Y]$. In general, we can define and use a polynomial ring in several variables $R[X_1, \cdots, X_n]$ or even in infinitely many variables if convenient. [5]

- **Power series.**

  If we use the same setup as above, but drop the condition that the supports of functions is finite, then we get the ring of formal power series with elements of the form

  $$\sum_{i=0}^{\infty} a_i X^i.$$

  The resulting ring is denoted by $R[[X]]$. As before, we can add several variables at a time to make $R[[X_1, \cdots, X_n]]$ or even infinitely many variables.

[5] If we allow infinitely many variables, then we have a choice to allow only finitely many variables at a time in a given element, or allow infinitely many variables to appear in a single element. We get different rings and both can be useful!

- **Generalized polynomial or power series rings.** We may allow the support to be a finite subset of $\mathbb{Z}$ and get a ring denoted as $R[X, X^{-1}]$ where we allow finitely many positive or negative exponents for $X$. This is called the ring of Laurent polynomials over $R$.

  A similar construction for power series requires assuming that the support is well ordered and the ring is then denoted as $R((X))$. [6]

3. **Group rings.**

   If $G$ is a finite group (not necessarily abelian) then we can repeat the above abstract construction to make $R^G$ into a ring. This time, we don't worry about the order, since the order is only needed to make sure that the product has only finitely many terms to collect at a time. Since our whole set $G$ is finite, this is not a problem at all.

   The function $f \in R^G$ may be conveniently written as

   $$\sum f(g) \cdot g$$

   where the symbols "$g$" are simply place holders. This is called the group ring of $G$ over $R$ and is simply denoted as $RG$.

4. **Matrix rings.** Let $n$ be a natural number and let $M_n(R)$ be the set of $n \times n$ matrices with entries in $R$. The usual addition and multiplication of matrices makes it into a ring and a rich source of examples of non commutative rings. Note that we need to fix a positive integer $n$ to work with such rings. Sometimes, one defines a ring of infinite matrices whose elements have the shape:

   $$\begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix}$$

   where $A$ is an $n \times n$ matrix for some finite $n$ and $I$ stands for an infinite diagonal matrix with 1's down its diagonal. Here the two 0's are supposed to represent zero matrices of appropriate sizes to fill up an infinite square matrix.

   This ring may be denoted with $M(R)$. If we are working with finitely many matrices of this ring, we can find an $n$ such that all of them can be thought to have a shape with an $n \times n$ part in the top left corner, extended with $I$ and 0's. Then their matrix operations can be essentially worked out in $M_n(R)$ and filled out with $I$ and 0's.

[6] Why do we not write $R[X, X^{-1}]$ as $R(X)$? The reason is that this particular symbol is used for the so-called quotient field (or the total quotient ring) to be defined later.

# 3 Ideal operations.

Ideals are to be thought of as a natural generalization of integers. In $\mathbb{Z}$ the ring of integers, we know how to work with individual integers, how to factor and divide and otherwise combine them. In arbitrary rings, elements may not have unique or finite factorization and to restore some of the properties of integers, one defined what should ideally behave like numbers, hence the word ideal.

Given an element $r$ in a ring $R$, let $a, b \in R$ and note that any (two-sided) ideal containing $r$ must also contain $arb$. If $1 \in R$, then $a, b$ are allowed to be any integers (or really their images in the ring). If $1 \notin R$ then we can still make perfect sense out of $nr$ as $r + \cdots + r$ - a sum of $n$ terms if $n > 0$ and $-r - \cdots - r$ - a sum of $-n$ terms if $n < 0$. For convenience, we interpret $an$ as $na$. Thus we shall consider integers multiples of elements of $R$ even though integers themselves may not be in $R$.

Given a ring $R$ and any subset $S$, it is easy to see that the set:

$$\{a_1 s_1 b_1 + \cdots + a_m s_m b_m | a_i, b_i \in R\ ,\ s_i \in S \text{ and } m \in \mathbb{Z}_+\}$$

is easily seen to be a two-sided ideal in $R$. It is called the ideal generated by $S$ and is denoted as $(S)$. [7]

**Remark.**

If the set $S$ is finite, say $S = \{s_1, \cdots, s_m\}$, then it is tempting to write

$$(S) = \{\sum_1^m a_i s_i | a_i \in R\}$$

but this is **wrong** if $R$ does not contain 1. We need to modify it to:

$$(S) = \{\sum_1^m (a_i + d_i) s_i | a_i \in R, d_i \in \mathbb{Z}\}$$

where $d_i s_i$ is defined thus:

- If $d_i > 0$ then $d_i s_i = s_i + s_i + \cdots$ a sum of $d_i$ terms.
- If $d_i < 0$ then $d_i s_i = -(-d_i s_i)$, calculated using the above definition for $-d_i$.
- If $d_i = 0$ then $d_i s_i = 0$.

[7] This is at variance with the "subgroup generated by" notation $< S >$, but is common for ideals. Also, if $S$ is a finite set, then we drop the set notation; for example, we write $(s, t)$ for $(\{s, t\})$.

If the ring is $\mathbb{Z}$ and $S$ is any subset of $\mathbb{Z}$, then we know that $(S) = (d)$ where $d$ is the GCD of all the elements of $S$. This is the main consequence of the Euclidean algorithm in $\mathbb{Z}$ which says that the GCD of any two integers is a linear combintion of the two with integer coefficients. Thus, every ideal in integers is generated by one element, or principle. In general, we shall study rings in which every ideal is finitely generated and these are called Noetherian rings.

We now define a set of operations on ideals and illustrate each by their counterpart in integers. We shall **assume that the ring is commutative,** but the definitions can be suitably interpreted in non commutative rings also, provided extra care is taken and suitable conditions imposed.

1. Given two ideals $I, J$ in $R$, their sum $I + J$ is defined as the ideal generated by $\{a + b | a \in I, b \in J\}$. It is easy to show that if $I = (S)$ and $J = (T)$, then $I + J = (S \bigcup T)$.

   Since ideals in $\mathbb{Z}$ are principle, we can say that $I = (n), J = (m)$ for some $n, m \geq 0$ and then it is known that $I + J = (n, m) = (d)$, where $d = \text{GCD}(n, m)$.

   More generally, the idea of sum can be extended to even infinite collection of ideals, where we consider elements which belong to the sum of finitely many ideals at a time.

2. Given two ideals $I, J$ in $R$, their intersection is defined as the usual set theoretic intersection and is easily seen to be an ideal. As before, this extends to an infinite set as well.

   In $\mathbb{Z}$ the ideal $(n) \bigcap (m)$ simply evaluates to $(\text{LCM}(n, m))$. Here an infinite intersection will reduce to the zero ideal unless it is essentially a finite intersection.

3. Given two ideals $I, J$ in $R$, their product is defined as the ideal generated by $\{rs | r \in I, s \in J\}$. It is denoted as $IJ$ and is really equal to $(IJ)$ in our usual convention for products of sets.

   In $\mathbb{Z}$ we get that $(n)(m) = (nm)$.

4. Given ideals $I, J$ in $R$ we define their quotient $I : J$ to be the set $\{r \in R | rJ \subset I\}$.

   In $\mathbb{Z}$, this gives $(n) : (m) = (n/d)$ where $d = \text{GCD}(n, m)$. Thus, for example, $(6) : (4) = (3)$. For proof, note that $3(4) \subset (12) \subset (6)$ and if $4x$ is divisible by 6, then clearly 3 must divide $x$.

5. Given an ideal $I$ in $R$, by its radical, we mean the ideal

$$\{x | x^n \in I \text{ for some positive integer } n\}.$$

The notation for this radical ideal is $\sqrt{I}$.

In $\mathbb{Z}$, the $(\sqrt{n})$ is the ideal $(m)$ where $m$ is obtained from the product of all distinct prime factors of $n$. Thus $(\sqrt{108}) = (2 \cdot 3) = (6)$.

**Remark.**

Avoid the confusion with the usual meaning of square root. In ideals the symbol covers all $n$-th roots!

# 4 Extending rings.

We now discuss some useful ways of creating new rings out of old ones.

**Convention.**

For convenience, we shall assume that we have a commutative ring $R$ with 1. The reader may try to generalize the constructions by dropping these conditions.

## 4.1 Adjoining one element to a ring.

Let $R$ be a ring (with the convention in force, of course) and let $S = R[X]$ a polynomial ring over $R$. Note that $R$ can be identified as the subring of $R[X]$ consisting of polynomials of degree 0.

*We shall do this identification without comment in the future.*

### 4.1.1 Adjoining an integral element.

**Definition: Integral element.**

An element $x \in A$ is said to be **algebraic** over a subring $B$ if it is the root of some non zero polynomial with coefficients in $B$. This means:

$$b_0 x^n + b_1 x^{n-1} + \cdots b_n = 0$$

for some $b_0, b_1, \cdots, b_n \in B$ with $b_0 \neq 0$. The element $X$ is then a root of the polynomial $f(X) = b_0 X^n + b_1 X^{n-1} + \cdots + b_n \in B[X]$.

An element $x \in A$ is said to be **transcendental** over a subring $B$ if it is not algebraic. In this case, it is easy to see that the set

$$B[x] = \{p_0 x^m + \cdots + p_m | p_i \in B, m \text{ is a non negative integer } \}$$

is actually a ring isomorphic to the polynomial ring $B[X]$ under the isomorphism mapping $X$ to $x$.

A ring $A$ is said to be algebraic over a subring $B$ if every element of $A$ is algebraic over $B$.

An element $x \in A$ is said to be integral over a subring $B$ if it is a root of a monic polynomial with coefficients in $B$, i.e.

$$x^n + b_1 x^{n-1} + \cdots + b_n = 0$$

for some $b_1, \cdots, b_n$ in $B$.

A ring $A$ is said to be integral over $B$ if every element of $A$ is integral over $B$.

**Example.** Recall the Newton's Rational Root Theorem: If a rational number $r$ is a root of a polynomial $f(X) = a_0 X^n + \cdots a_n$ of degree $n$ with integer coefficients and if $r = \frac{p}{q}$ in reduced form, then $p$ divides $a_n$ and $q$ divides $a_0$.

This theorem applied to the case when $f(X)$ is monic (or $a_0 = 1$) gives that $r$ must be an integer factor of $a_n$.

In particular, it says that a rational root of a monic polynomial is an integer! In other words, a rational number integral over the ring of integers is actually an integer. This explains the term "integral".

A similar argument can be used for $k[X]$, a polynomial ring over a field $k$. This gives a theorem which says that a rational function (ratio of two polynomials in $k[X]$) is integral over $k[X]$ is integral over $k[X]$ iff it is in $k[X]$.

Let

$$f(X) = X^n + a_1 X^{n-1} + \cdots + a_n$$

be a monic polynomial of degree $n$ in $S$ so that $a_i \in R$.

Let $I = (f(X))$ the ideal generated by $f(X)$ in $S$ and let $T = S/I$ and let $\phi : S \to T$ be the canonical residue class map $\phi(z) = z + I$.

We claim that

1. $\phi$ restricted to $R$ is an isomorphism on to $\phi(R)$.

2. $w = \phi(X)$ satisfies $w^n + \phi(a_1) w^{n-1} + \cdots + \phi(a_n) = 0$.

3. **Now we identify** $R$ with $\phi(R)$, so for $r \in R$, we replace $\phi(r)$ by $r$.

   Then we can simply say that $T$ is a ring containing $R$ and

   $$f(w) = w^n + a_1 w^{n-1} + \cdots + a_n = 0.$$

   In other words, $T$ is an extension of $R$ containing a root $w$ of $f(X)$.

4. The ring $T$ can be described thus:

   $$T = \{r_0 w^{n-1} + r_1 w^{n-2} + \cdots + r_{n-2} w + r_{n-1} | r_i \in R\}.$$

In other words $T$ consists of expressions $g(w)$ as $g(X)$ varies over all polynomials in $X$ of degree at most $n-1$ over $R$. To compute the product $g_1(w)g_2(w)$ we write

$$g_1(X)g_2(X) = q(X)f(X) + r(X)$$

where $r(X)$ is the remainder of $g_1(X)g_2(X)$ mod $f(X)$ and noting that $f(w) = 0$ we get $g_1(w)g_2(w) = r(w)$.

5. The ring $T$ can be thought of as a ring obtained by adjoining a root of $f(X)$ to $R$.

**PROOF.** The first claim is easily checked. As for the second claim, note that $w^n + \phi(a_1)w^{n-1} + \cdots + \phi(a_n) = \phi(f(X))$ and this is zero since $f(X) \in I = Ker(\phi)$.

Given any element $g(X) \in R[X]$ we write $g(X) = q(X)f(X) + r(X)$ by the usual division algorithm and since $\phi(f(X)) = 0$ we see that

$$\phi(g(X)) = g(w) = \phi(r(X)) = r(w).$$

The rest of the claims are now easy to verify.

**Example.** Let $R = \mathbb{Z}$ and $f(X) = X^2 - D$ for some integer $D$. The ring $T = \mathbb{Z}[X]/(X^2 - D)$ can then be written as:

$$T = \{aw + b | a, b \in \mathbb{Z}\}$$

where $w = \phi(X)$ satisfies $w^2 - D = 0$, i.e. $w = \sqrt{(}D)$. Thus $T$ can be thought of as $\mathbb{Z}[w]$ an extension of $\mathbb{Z}$ obtained by adjoining $w$, a square root of $D$. It follows that $-w$ is also a square root of $D$.

Several particular cases are noteworthy.

- If $D = 2$, then we get the ring $\mathbb{Z}[\sqrt{2}]$, the ring obtained by adjoining the square root of 2.

- If we take $D = -1$ so that $f(X) = X^2 + 1$, then we get the usual imaginary square root of $-1$. Indeed, if we take $R = \Re$ the field of reals and make $\Re[X]/(X^2 + 1)$ then we get the field of complex numbers $\mathbb{C}$ .

- **Food for thought.** Notice that we have put no restriction on $D$. If we take $D = 1$, then we already have two square roots $1, -1$ of $D$ in $\mathbb{Z}$. Our construction produces two more, namely $w, -w$. How can we have four square roots? The problem is that our ring $T$ has zero divisors. Thus we have
$$0 = w^2 - 1 = (w - 1)(w + 1)$$

yet neither $w-1$ nor $w+1$ are zero! Thus we need to be careful and restrain our intuition when using this algebraic device!

We run into such problems only when we allow our $f(X)$ to be reducible.

- Consider $D = 5$. Let $g(X) = X^2 + X - 1$ and let $S = \mathbb{Z}[\omega]$ the ring obtained by adjoining a root $\omega$ of $g(X)$.

  Use $\omega^2 + \omega = 1$ and we can see that for $w = 2\omega + 1$ we have

  $$w^2 = (2\omega + 1)^2 = 4\omega^2 + 4\omega + 1 = 4(\omega^2 + \omega) + 1 = 5.$$

  Thus our ring $\mathbb{Z}[\sqrt{5}]$ can be identified with the subring $\mathbb{Z}[2\omega + 1]$ of $S$.

  It can be shown that whenever $D = 1+4n$ we can use $g(X) = X^2+X-n$ and get

  $$\mathbb{Z}[\sqrt{D}] = \mathbb{Z}[2\omega + 1] \subset \mathbb{Z}[\omega]$$

  where $g(\omega) = 0$.

  The book defines $\mathcal{O}(D)$ to be $\mathbb{Z}[\sqrt{D}]$ if $D$ is a squarefree integer congruent to 2 or 3 mod 4 and $\mathcal{O}(D) = \mathbb{Z}[\omega]$ if $D$ is congruent to 1 mod 4.

  This will be discussed in more details later.

### 4.1.2 Adjoining a general algebraic element.

If we work as above but now allow

$$f(X) = a_0X^n + a_1X^{n-1} + \cdots + a_n$$

where $a_i \in R$ and $a_0 \neq 0$ so $f(X)$ has degree $n$.

The ring $T = R[X]/(f(X))$ is defined as above, but because of the leading coefficient $a_0$ not being a unit, we may not be able to carry out the usual division algorithm completely.

Letting $\phi$ denote the usual canonical homomorphism and $w = \phi(X)$ we still get that $w$ is a root of $f(X)$ after we identify $R$ with $\phi(R)$.

The elements of $T$ can be still expressed as polynomials in $w$, but we can no longer restrict their degree to less than $n$.

**Example.** Let $R = \mathbb{Z}$ and $f(X) = 2X - 1$. Then $T$ is identified with $\mathbb{Z}[\frac{1}{2}]$ since $\frac{1}{2}$ is a unique root of $f(X)$.

With $R = \mathbb{Z}$ we can also consider $f(X) = 4X^2 - 5$ and note that our ring $T$ may be thought of as $\mathbb{Z}[\frac{\sqrt{5}}{2}]$. Note that $T$ contains a subring $\mathbb{Z}[\sqrt{5}]$ already known to us and this subring is integral over $\mathbb{Z}$ but the whole ring $T$ is not.

## 4.2 Inverting a set.

As we saw above, the result of adjoining a root of $2X - 1$ was to allow the fraction $\frac{1}{2}$ into the ring $\mathbb{Z}$. We now generalize this process for a general ring $R$, with the continued

**Convention:** The ring $R$ is assumed to be commutative with 1.

Let $S$ be a subset of $R$ which is multiplicatively closed, which means the product of any two elements of $S$ is again in $S$.

As **a temporary convenience,** we shall also assume that $S$ contains no zero divisors or zero. We shall drop this convention later.

We now construct a ring $A$ containing $R$ in which every element of $S$ becomes a unit. In principle, we could add the inverse of each element of $S$ but we we can do the whole set in in operation as follows.

1. Start with a set
$$RR = \{(a, s) | a \in R, s \in S\}$$
and define an equivalence relation on $RR$ by $(a, s) \cong (b, t)$ iff there exists $s_1 \in S$ such that
$$s_1(at - bs) = 0.$$
We leave for the reader to check that this is an equivalence relation.

   Now, under our temporary convenient assumption, we it follows that $s_1$ is not needed, since every element of $S$ is non zero and a non zero divisor, so the condition reduces to $at - bs = 0$. We have included it so the construction will work in general.

2. Let the equivalence class of a pair $(a, s)$ be denoted by $[a, s]$. The intention is that this class $[a, s]$ shall represent the intended fraction $\frac{a}{s}$ in our resulting ring $A$.

   Define
$$A = \{[a, s] | a \in R, s \in S\}.$$
Define ring operations on $A$ by
$$[a, s] + [b, t] = [at + bs, st] \text{ and } [a, s][b, t] = [ab, st].$$
It is straightforward but tedious to check that these are well defined and make a valid ring.

   The set
$$\{[a, 1] | a \in R\}$$

is seen to form a subring isomorphic to $R$ under the correspondence $[a, 1] \leftrightarrow a$.

The corresponding elements of $S$ become $[s, 1]$ and have the inverse $[1, s] \in A$, thus we have accomplished our task.

3. The ring $A$ is suggestively denoted by $S^{-1}R$ and its elements $[a, s]$ may be denoted as $\frac{a}{s}$ for convenience in actual use.

**Examples.** We give several illustrations of this construction.

1. Let $R$ be an integral domain and let $S = \{s \in R | s \neq 0\}$. Then $S^{-1}R$ is a field called the quotient field of $R$ or the field of fractions of $R$.

   We may use the notation $qt(R)$ to denote it.

   Thus $qt(\mathbb{Z}) = \mathbb{Q}$ .

2. Let $R$ be still a commutative ring with 1 and take $S$ to be the set of non zero divisors of $R$. Then the ring $S^{-1}R$ need not be a field but is called the "Total Quotient Ring" of $R$.

   We will still use the suggestive notation $qt(R)$ for it.

   Consider the ring $R = \mathbb{Q}\ [X, Y]/(XY)$ and for convenience denote by $x, y$ the natural images of $X, Y$ respectively in $R$.

   It is easy to see that elements of $R$ can be described as $a+xp(x)+yq(y)$ where $p(x)$ and $q(y)$ are polynomial expressions in $x, y$ over $R$.

   It is not hard to see that $a + xp(x) + yq(y) = 0$ iff $a = 0 = p(x) = q(y)$ and that $a + xp(x) + yq(y)$ is a non zero divisor iff $a \neq 0$. Let

   $$S = \{a + xp(x) + yq(y) \in R | a \neq 0\}$$

   and verify that $S$ is a multiplicatively closed set of non zero divisors as required.

   The ring $S^{-1}R$ is the desired total quotient ring $qt(R)$. It is not a field but its non units are zero divisors, which cannot be units any way!

3. We shall now use the same example $R$ above to illustrate inverting multiplicative sets which contain zero divisors. Thus,**we now drop the convenient convention of avoiding zero divisors in our multiplicative sets.**

   Thus, let

   $$M = \{a + xp(x) \in R | a \neq 0 or p(x) \neq 0\}.$$

   The set is multiplicatively closed but does contain the zero divisor $x$.

Form the ring $A = M^{-1}R$ and consider the map $\psi : R \to A$ defined by $\psi(r) = [r, 1]$. Note that $[y, 1] = [xy, x] = [0, x] = [0, 1]$ from known equivalences of the classes.

Thus, the image of the whole ideal $(y)$ is zero! It is easy to deduce that $Ker(\psi) = (y)$ and hence $\psi(R) \cong R/(y)$.

Now $R/(y)$ can be thought of as $\mathbb{Q}\,[X, Y]/(XY, Y)$ and since the ideal $(XY, Y) = (Y)$ we see that $R/(y) = \mathbb{Q}\,[X, Y]/(Y) \cong \mathbb{Q}\,[X]$.

Thus the ring $A$ contains an isomorphic copy of the ring $\mathbb{Q}\,[X]$ and further calculations show it to be nothing but $qt(\mathbb{Q}\,[X])$. It is customary to denote this field as $\mathbb{Q}\,(X)$.

# 5 Important domains.

We now begin the study of some important integral domains which appear as fundamental objects for many ring theoretic investigations.

## 5.1 Euclidean Domains.

The motivation for these domains is the pair of very familiar rings, the ring of integers $\mathbb{Z}$ and the ring of polynomials in one variable $k[X]$.

Both of these domains have a norm and a division algorithm. We explain the meaning of this and give a definition at the same time.

**Definition: Norm.** Let $R$ be a domain. A norm on $R$ is a map $N : R \to \mathbb{Z}_{\geq 0}$ where $N(0) = 0$. The norm is said to be positive, if $a \neq 0$ implies $N(a) \neq 0$.

**Definition: Euclidean Domain.** A domain $R$ is said to a "euclidean domain" with respect to a norm $N$, if given any two elements $a, b \in R$ with $b \neq 0$ we can write $a = qb + r$ for some $q, r \in R$ such that $r = 0$ or $N(r) < N(b)$.

**Examples.**

- Define a norm in $\mathbb{Z}$ by $N(x) = |x|$. This is a positive norm. Then $\mathbb{Z}$ becomes a euclidean domain since the usual division algorithm works.

- In $k[X]$, let a norm be defined by $N(f(X)) = \deg_X(f(X))$ if $f(X) \neq 0$ and for this discussion, we agree that $N(0) = 0$.

  This is not a positive norm. It can be changed to a positive norm by the simple trick of defining a different function:

$$N^*(f(X)) = 2^{\deg_X(f(X))} \text{ if } f(X) \neq 0 \text{ and } N^*(0) = 0.$$

The division algorithm is defined for the polynomials as well and yields a euclidean domain.

- Another useful example of norm comes from the ring of algebraic integers which we now discuss.

  Define $\mathcal{O}(D)$ to be $\mathbb{Z}[\sqrt{D}]$ if $D$ is a squarefree integer congruent to 2 or 3 mod 4 and $\mathcal{O}(D) = \mathbb{Z}[\omega]$ if $D$ is congruent to 1 mod 4. Here the $\omega$ is defined to be $\frac{1+\sqrt{D}}{2}$ and satisfies the quadratic equation $X^2 - X - s = 0$ where $D = 1 + 4s$.

  Thus we have two cases:

Case 1 $\mathcal{O}(D) = \mathbb{Z}[X]/(X^2 - D)$ if $D$ is squarefree and 2 or 3 modulo 4, while

Case 2 $\mathcal{O}(D) = \mathbb{Z}[X]/(X^2 - X - s)$ when $D$ is squarefree of the form $D = 1 + 4s$.

In each case, define $\alpha$ to be the canonical image of $X$ in $\mathcal{O}(D)$ and note that every element of the ring is then of the form $a + b\alpha$ where we are identifying $a, b \in \mathbb{Z}$ with their images in $\mathcal{O}(D)$.

We define the norm when $D$ is squarefree negative.

We define the norm by $N(a+b\alpha) = a^2 - b^2 D$ in Case 1. Note that this is simply the product of $a + b\alpha$ with its conjugate $a - b\alpha$.

We define the norm to be $N(a+b\alpha) = a^2 + ab - sb^2$ in the second case. Note that this is also a product of the conjugates, but this time, the conjugate of $a + b\alpha$ is $a + b(1 - \alpha)$.

Note that in either case, we need to show that the map is into $\mathbb{Z}_{\geq 0}$ and this is easy to see by the negative discriminant of the quadratic expression.

### 5.1.1 Properties of euclidean domains.

The main properties of euclidean domains are the following. Let $R$ be a euclidean with norm $N$.

1. Given any two elements $a, b \in R$ with at least one non zero, then there is a $d \in (a, b)$ such that $a, b \in (d)$. In particular $(d) = (a, b)$.

   **Proof.** Take $d \neq 0$ to be an element of $(a, b)$ such that $N(d)$ is minimum among all elements of $(a, b)$. Then the division algorithm guarantees that when any element of $(a, b)$ is divided by $d$ then the remainder must be zero. Thus $(a, b) \subset (d)$.

In usual terminology, we can say that $d$ divides both $a, b$ and any common divisor of $a, b$ divides $d$. Hence $d = \mathrm{GCD}(a, b)$. Moreover, if $d^* \in R$ is any other element satisfying the same conditions, then $(d) = (d^*)$ guarantees that either is a multiple of the other by a unit. For convenience we make a:

**Definition: associate elements** Two elements of a ring are said to be associates of each other if one of them is a multiple of the other by a unit. Note that associate elements generate the same ideal, but in a ring with zero divisors, two elements generating the same ideal need not be associates.

**Remark.** Note that in $\mathbb{Z}$ it is customary to make the GCD unique by requiring it to be positive. In integers $\pm 1$ are the only units and so this is possible.

2. The above proof also generalizes to any ideal $I \subset R$, since we can still take $d$ as the minimal norm non zero element. Thus every ideal of the ring is generate by one element or is principal.

   We thus make a:

   **Definition: Principal Ideal Domain P.I.D.** An integral domain is said to be a principal ideal domain (P.I.D. for short) if every ideal in it is principal.

   Note that there is no reason to require the ring to be an integral domain, so we also make a:

   **Definition: Principal Ideal Ring. P.I.R.** A commutative ring is said to be a principal ideal ring (P.I.R. for short) if every ideal in it is principal.

   **Remark.** As we note below a P.I.D. is not necessarily a euclidean domain. It is true that a P.I.D. can be shown to be equipped with something called a Dedekind-Hasse norm which we briefly discuss.

   A norm $N$ on a domain $S$ is said to be a Dedekind-Hasse norm if it is positive and has the following property: Given any two elements $a, b \in S$, either $a \in (b)$ or there is some element $0 \neq z \in (a, b)$ such that $0 < N(z) < N(b)$.

   Note that, $z = sa + tb$ for some $s, t \in S$ and this is almost like the division algorithm, except in the usual division algorithm, $s$ is required to be 1.

   **Remark.** It is known that existence of a Dedekind-Hasse norm is equivalent to the P.I.D. property. It is easy to see why the Dedekind-Hasse norm would imply the P.I.D. property; simply imitate the euclidean proof by taking a non zero least norm elemnt in the ideal and argue that it must generate the ideal. The converse needs properties of unique factorization domains (U.F.D.) developed later.

**Remark.** In spite of the norm being available, we do not necessarily get a euclidean domain in general. We discuss some special cases below.

1. **The case of $\mathbf{D = -5}$.**

   We shall show that this ring has non principal ideals and hence is not euclidean.

   Here $\mathcal{O} = \mathbb{Z}[X]/(X^2+5)$ and we set $\alpha$ as the canonical image of $X$. Suppose $0 \neq w = a(b+c\alpha) \in \mathcal{O}$ where $a, b, c \in \mathbb{Z}$ with $b, c$ coprime.

   First assume that $c \neq 0$.

   Let $I = (w)$ the principal ideal. What is $I \bigcap \mathbb{Z}$, when we identify $\mathbb{Z}$ with its canonical image in $\mathbb{Z}$?

   Here is the calculation. Let $u = w(p+q\alpha) \in \mathbb{Z}$. Then we get

   $$u = a(b+c\alpha)(p+q\alpha) = a(bp-5cq) + a(bq+cp)\alpha \in \mathbb{Z}.$$

   It follows that $bq + cp = 0$ and because of coprimeness of $b, c$ we see that $(p, q) = \lambda(b, -c)$ for some $\lambda \in \mathbb{Z}$.

   It follows that $u = a(b^2+5c^2)\lambda$ and thus $I \bigcap \mathbb{Z} = (a(b+5c^2))\mathbb{Z}$.

   If $c = 0$ then the calculation gets easier and we see that $u = abp$ so $I \bigcap \mathbb{Z} = (ab)\mathbb{Z}$.

   Thus, in either case, $I \bigcap \mathbb{Z} = (a(b+5c^2))$.

   We now **claim** that the ideal $(3, 2+\alpha)$ cannot be principal.

   **Proof.** Suppose it is generated by some $w$ as above. Then $3 \in (a(b^2 + 5c^2))$. But $|a(b^2+5c^2)| \geq 5$, unless $c = 0$. Hence we must have $c = 0$ and $w = ab$, with $3 \in (ab)$. Clearly $ab = \pm 1$ is not an option, since then $I = (1)$, but $\mathcal{O}/I$ is easily seen to be a non zero ring.

   Then we may assume $w = ab = \pm 3$. But then $2 + \alpha \notin (w)$,since 3 does not divide 2, a contradiction.

   **Exercise.** The reader should work out the general case of intersection of a principal ideal with $\mathbb{Z}$ in the ring $\mathcal{O}$.

2. **Some examples of Euclidean rings.** We describe some $\mathcal{O}(D)$ which are euclidean.

(a) Case of $D = -1, -2$. $\mathcal{O}(D) = \mathbb{Z}[X]/(X^2 - D)$ and we shall write $\alpha$ for the canonical image of $X$ as before. We shall use the norm $N(a + b\alpha) = a^2 - Db^2$.

We now show how the division algorithm works. Given $p + q\alpha \in \mathcal{O}(D)$ and $0 \neq a + b\alpha \in \mathcal{O}(D)$, we let

$$u + v\alpha = \frac{p + q\alpha}{a + b\alpha}$$

where $u, v \in \mathbb{Q}$ . Find $r, s \in DZ$ such that

$$|u-r| \leq \frac{1}{2} \geq |v-s| \text{ and hence } N((u-r)+(v-s)\alpha) \leq \frac{1}{4}+\frac{1}{4}(-D).$$

For convenience set $\theta = (u-r)+(v-s)\alpha$ and note that $N(\theta) < 1$ when $D = -1$ or $-2$.

We have

$$p + q\alpha = (a + b\alpha)(r + s\alpha) + (a + b\alpha)\theta$$

and by multiplicativity of this norm function, the norm of the last term is less than $N(a+b\alpha)$, so $(a+b\alpha)\theta$ has the desired property of the remainder.

(b) Case of $D = -3, -7, -11$. In these cases $\mathcal{O}(D)$ is $\mathbb{Z}[X]/(X^2 - X - s)$ where $s = -1, -2, -3$ respectively. Setting $\alpha$ to the the image of $X$ as before, we have the norm equal to

$$N(a + b\alpha) = a^2 + ab - sb^2 = (a + \frac{b}{2})^2 - (s + \frac{1}{4})b^2.$$

Noting that $D = 1 + 4s$ we note that the last term simplifies to $-(\frac{D}{4})b^2$.

We imitate the above proof by writing

$$u + v\alpha = \frac{p + q\alpha}{a + b\alpha}$$

where $u, v \in \mathbb{Q}$ .

Find $r, s \in \mathbb{Z}$ such that

$$|v - s| \leq \frac{1}{2} \geq |(u + \frac{v}{2}) - (r + \frac{s}{2})| = |(u - r) + \frac{v - s}{2}|.$$

Let as before $\theta = (u - r) + (v - s)\alpha$ and note that

$$N(\theta) = \left((u - r) + \frac{v - s}{2}\right)^2 - \frac{D}{4}(v - s)^2.$$

Thus when $D = -3, -7$ or $-11$

$$N(\theta) \le \frac{1}{4} + \frac{|D|}{16} = \frac{4+|D|}{16} < 1.$$

The rest of the proof follows as before giving $(a+b\alpha)\theta$ as a valid remainder.

3. **A domain which is a P.I.D. but not euclidean.** The ring $\mathcal{O}(-19)$ is very special. It is shown to be non euclidean under **any possible** norm and yet unlike the non euclidean ring $\mathcal{O}(-5)$ it is a P.I.D. - i.e. a domain in which every ideal is principal!

   Even though elementary, the proof is somewhat long and we leave it as an assignment to look it up and digest.

# 6 Principal Ideal Domains. P.I.D.

We now discuss some useful properties of PID or more generally of PIR (principal ideal ring).

1. **Stability of an increasing sequence of principal ideals.**

   Let $R$ be a principal ideal ring (PIR) and if $(x_1, x_2, \cdots, x_n, \cdots)$ is a sequence of elements in $R$, such that $x_{n+1}$ divides $x_n$ for all $n \ge 1$. (In other words, $x_n \in (x_{n+1})$ = the ideal generated by $x_{n+1}$.)

   Then the sequence of ideals $(I_n)$ stabilizes for $n >> 0$, which means there is some sufficiently large $N$ such that $I_N = I_n$ for all $n \ge N$.

   **Proof.** The ideal generated by all the $x_i$ is generated by some $x$, since $R$ is a PIR. Thus $x = \sum_1^N a_i x_i$ for some $N$ where $a_N \neq 0$. From the given divisibility, we see that $x \in x_N$. Since $x_N \in (x)$ we see that $(x) = (x_N)$. It follows that for any $n > N$, we have $I_N = (x_N) \subset (x_n) = I_n$ from hypothesis, while $I_n \subset (x) = (x_N) = I_N$ from construction.

   Hence $I_n = I_N$ for all $n \ge N$.

2. **Definition: Irreducible/reducible element** An element $x$ in a domain $R$ is said to be reducible if $x = yz$ for some $y, z \in R$ such that $x$ is not an associate of either $y$ or $z$. Equivalently, we could also state the condition as neither $y$ nor $z$ is a unit.

   Yet another way of stating the condition is to write $x = yz$ for some $y, z \in R$ such that $(x) \neq (y)$ and $(x) \neq (z)$.

   An element $x$ in a domain is said to be irreducible if it is not reducible.

Note that units are irreducible under this definition and so is the zero element.

**Remark.** If the ring is not an integral domain, then the different formulations are not necessarily equivalent and we leave the generalizations to the reader's imagination at this point.

3. **Factorization in a PID.** We now prove that every element of a PID can be written as a product of finitely many irreducible elements.

   **Proof.** Suppose if possible we have a non empty collection of elements which cannot be written as a product of finitely many irreducible elements. For convenience, let us call them "bad" elements. Any element which is a product of finitely many irreducibles will be called good. Note that product of finitely many good elements is obviously a product of finitely many irreducibles and hence good.

   Consider the set

   $$S = \{(x) | x \in R \text{ and } x \text{ is bad.}\}$$

   Note that a bad element $x$ must be factorable as $x = yz$ such that at least one of $y, z$ is again bad and $x$ is not an associate of either $y$ or $z$.

   For proof of this, note that $x$ being bad, must be reducible, otherwise it is a singleton product of irreducibles. Also, we already know that if $y, z$ were good then so would be $x$, a contradiction.

   Thus, if $y$ is bad then we get $(x) \subsetneq (y)$ where $(x), (y)$ both belong to $S$. Repeating this process, we can get an infinite sequence of principal ideals generated by bad elements:

   $$(x_1) \subsetneq (x_2) \subsetneq \cdots \subsetneq (x_n) \cdots \text{ with } (x_n) \in S \text{ for all } n = 1, 2, \cdots .$$

   This is a contradiction to the stability of an increasing sequence of ideals that we established.

4. Uniqueness of expression. Now that we have established existence of factorization of elements as products of irreducibles, we investigate the uniqueness of such an expression.

   First we need a Lemma.

   **Lemma. Irreducible is prime in a PID.** If $x$ is an irreducible element in a PID $R$, then the ideal $(x)$ is a prime ideal.

   **Proof.** Suppose that $yz \in (x)$ where $y, z \in R$ and $y \notin (x)$. We wish to show that $z \in (x)$.

Set the ideal $I = (x, y)$ and assume that $I = (w)$ for some $w$. Clearly, we can write $w = ax + by$ for some $a, b \in R$. Note that $x, y \in (w)$. Since $x$ is irreducible, either $w$ is a unit or $(w) = (x)$. Clearly, if $(w) = (x)$ then we get $y \in (w) = (x)$ a contradiction.

Hence $w$ is a unit. Now $wz = axz + byz \in (x)$, since $yz \in (x)$. Since $w$ is a unit, $z \in (x)$ as we needed.

■

Now we shall prove the uniqueness of the factorizations. Here is the statement:

Given any non zero element $x \in R$ where $R$ is a PID, there is a finite set of prime ideals $P(x) = \{(p_1), \cdots, (p_n)\}$, and a set of non negative integers $\{a_1, \cdots, a_n\}$ such that $x$ is an associate of $\prod_1^n p_i^{a_i}$.

The set of prime ideals $P(x)$ is uniquely determined by $x$ and for each $(p_i) \in P(x)$, the corresponding exponent $a_i$ is also uniquely determined by $x$.

**Proof.** First we shall establish the uniqueness of $P(x)$. So suppose, if possible

$$x = \epsilon \prod_{i=1}^{n} p_i^{a_i} = \tau \prod_{j=1}^{m} q_i^{b_i}$$

where $\epsilon, \tau$. Consider the corresonding sets of primes $S = \{p_i\}_1^n$ and $T = \{q_j\}_1^m$ where we are naturally assuming that $\{(p_i)\}$ are distinct primes and similarly $\{q_j\}$ are distinct primes.

Since $q_j$ divides $x$, it is clear that it divides one of $p_i$ and since $p_i$ is irreducible, $(q_j) = (p_i)$. Thus we see that every member of $T$ is in $S$. Similarly every member of $S$ is in $T$ and thus $S = T$ and this common set of prime ideals is the $P(x)$ which is thus uniquely determined by $x$.

The only remaining thing to prove is the uniqueness of exponents. Suppose if possible we have an example of non uniqueness

$$x = \epsilon \prod_{i=1}^{n} p_i^{a_i} = \tau \prod_{i=1}^{n} p_i^{b_i}$$

where $\epsilon, \tau$ are units. Moreover, all $a_i$ are non negative and some $a_i \neq b_i$.

Without loss of generality we may assume that $0 < a_1 < b_1$. Clearly $x \in (p_1^{a_1})$ and write $x = yp_1^{a_1}$. Then we see that

$$y = \epsilon \prod_{i=2}^{n} p_i^{a_i} = \tau p_1^{b_1 - a_1} \prod_{i=2}^{n} p_i^{b_i}.$$

Here the first expression for $y$ is not divisible by $p_1$, while the second one is divisible by $p_1$ since $b_1 - a_1 > 0$. This is a contradiction and we have the uniqueness established.

To be continued ...

# Index