Integral Domains

As always in this course, a ring $R$ is understood to be a commutative ring with unity.

1 First definitions and properties

**Definition 1.1.** Let $R$ be a ring. A divisor of zero or zero divisor in $R$ is an element $r \in R$, such that there exists an $s \in R$ with $s \neq 0$ and $rs = 0$. Thus, for example, 0 is always a zero divisor.

Example: in $\mathbb{Z}/6\mathbb{Z}$, $0 = 2 \cdot 3$, hence both 2 and 3 are divisors of zero.

One way to find divisors of zero is as follows:

**Definition 1.2.** Let $R$ be a ring. A nilpotent element of $R$ is an element $r$, such that there exists an $n \in \mathbb{N}$ such that $r^n = 0$. Note that 0 is allowed to be nilpotent.

**Lemma 1.3.** Let $R$ be a ring and let $r \in R$ be nilpotent. Then $r$ is a zero divisor.

*Proof.* The set of $n \in \mathbb{N}$ such that $r^n = 0$ is nonempty, so let $m$ be the smallest such natural number. Note that, if $m = 1$, $r = 0$ and hence is a divisor of zero. Otherwise, $0 < m - 1 < m$, so by assumption $r^{m-1} \neq 0$. Hence $r \cdot r^{m-1} = r^m = 0$, with $r^{m-1} \neq 0$, so that $r$ is a divisor of zero. \(\square\)

Example: in $\mathbb{Z}/16\mathbb{Z}$, $0 = 2^4 = 2 \cdot 2^3$, hence 2 is a divisor of zero. But in $\mathbb{Z}/6\mathbb{Z}$, neither 2 nor 3 is nilpotent, so there are examples of divisors of zero which are not nilpotent.

**Definition 1.4.** A ring $R$ is an integral domain if $R \neq \{0\}$, or equivalently $1 \neq 0$, and such that $r$ is a zero divisor in $R \iff r = 0$. Equivalently, a nonzero ring $R$ is an integral domain $\iff$ for all $r, s \in R$ with $r \neq 0$, $s \neq 0$, the product $rs \neq 0 \iff$ for all $r, s \in R$, if $rs = 0$, then either $r = 0$ or $s = 0$. 

1
**Definition 1.5.** Let $R$ be a ring. The cancellation law holds in $R$ if, for all $r, s, t \in R$ such that $t \neq 0$, if $tr = ts$, then $r = s$.

**Lemma 1.6.** A ring $R \neq \{0\}$ is an integral domain $\iff$ the cancellation law holds in $R$.

**Proof.** $\implies$: if $tr = ts$ and $t \neq 0$, then $tr - ts = t(r - s) = 0$. Since $t \neq 0$ and $R$ is an integral domain, $r - s = 0$ so that $r = s$.

$\impliedby$: Suppose that $rs = 0$. We must show that either $r$ or $s$ is 0. If $r \neq 0$, then apply cancellation to $rs = 0 = r0$ to conclude that $s = 0$. Hence $s = 0$.

The following are examples of integral domains:

1. A field is an integral domain. In fact, if $F$ is a field, $r, s \in F$ with $r \neq 0$ and $rs = 0$, then $0 = r^{-1}0 = r^{-1}(rs) = (r^{-1}r)s = 1s = s$. Hence $s = 0$. (Recall that $1 \neq 0$ in a field, so the condition that $F \neq 0$ is automatic.) This argument also shows that, in any ring $R \neq 0$, a unit is not a zero divisor.

2. If $S$ is an integral domain and $R \leq S$, then $R$ is an integral domain. In particular, a subring of a field is an integral domain. (Note that, if $R \leq S$ and $1 \neq 0$ in $S$, then $1 \neq 0$ in $R$.) Examples: any subring of $\mathbb{R}$ or $\mathbb{C}$ is an integral domain. Thus for example $\mathbb{Z}[\sqrt{2}]$, $\mathbb{Q}(\sqrt{2})$ are integral domains.

3. For $n \in \mathbb{N}$, the ring $\mathbb{Z}/n\mathbb{Z}$ is an integral domain $\iff$ $n$ is prime. In fact, we have already seen that $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ is a field, hence an integral domain. Conversely, if $n$ is not prime, say $n = ab$ with $a, b \in \mathbb{N}$, then, as elements of $\mathbb{Z}/n\mathbb{Z}$, $a \neq 0$, $b \neq 0$, but $ab = n = 0$. Hence $\mathbb{Z}/n\mathbb{Z}$ is not an integral domain.

4. If $R$ is an integral domain, then, as we shall see in a minute, $R[x]$ is an integral domain. Hence, by induction, $R[x_1, \ldots, x_n]$ is an integral domain if $R$ is an integral domain. In particular, if $F$ is a field, $F[x_1, \ldots, x_n]$ is an integral domain, as is $\mathbb{Z}[x_1, \ldots, x_n]$.

To prove the last statement (4) above, we show in fact:

**Lemma 1.7.** Let $R$ be an integral domain. Then, if $f, g \in R[x]$ are both nonzero, then $fg \neq 0$ and $\deg(fg) = \deg f + \deg g$.

**Proof.** Let $d = \deg f$ and $e = \deg g$. Then $f = \sum_{i=0}^d a_i x^i$ and $g = \sum_{j=0}^e b_j x^j$ with $a_d, b_e \neq 0$. Since $a_d b_e \neq 0$, the leading term of $fg$ is $a_d b_e x^{d+e}$. Hence $fg \neq 0$ and $\deg(fg) = d + e = \deg f + \deg g$. \qed
Corollary 1.8. Let $R$ be an integral domain. Then the group of units $(R[x])^*$ in the polynomial ring $R[x]$ is just the group of units $R^*$ in $R$ (viewed as constant polynomials).

Proof. Clearly, if $u$ is a unit in $R$, then it is a unit in $R[x]$, so that $R^* \subseteq (R[x])^*$. Conversely, if $f \in (R[x])^*$, then there exists a $g \in R[x]$ such that $fg = 1$. Clearly, neither $f$ nor $g$ is the zero polynomial, and hence $0 = \deg 1 = \deg(fg) = \deg f + \deg g$.

Thus, $\deg f = \deg g = 0$, so that $f, g$ are elements of $R$ and clearly they are units in $R$. Hence $f \in R^*$, so that $(R[x])^* \subseteq R^*$. It follows that $(R[x])^* = R^*$. $\square$

The corollary fails if the ring $R$ has nonzero nilpotent elements. For example, in $(\mathbb{Z}/4\mathbb{Z})[x]$,

$$(1 + 2x)(1 + 2x) = (1 + 2x)^2 = 1 + 4x + 4x^2 = 1,$$

so that $1 + 2x$ is a unit in $(\mathbb{Z}/4\mathbb{Z})[x]$.

Finally, we note the following:

Proposition 1.9. A finite integral domain $R$ is a field.

Proof. Suppose $r \in R$ with $r \neq 0$. The elements $1 = r^0, r, r^2, \ldots$ cannot all be different, since otherwise $R$ would be infinite. Hence there exist $0 \leq n < m$ with $r^n = r^m$. Writing $m = n + k$ with $k \geq 1$, we see that $r^n = r^m = r^{n+k} = r^n r^k$. By induction, since $R$ is an integral domain and $r \neq 0$, $r^n \neq 0$ for all $n \geq 0$. Applying cancellation to $r^n = r^n \cdot 1 = r^n r^k$ gives $r^k = 1$. Finally since $r^k = r \cdot r^{k-1}$, we see that $r$ is invertible, with $r^{-1} = r^{k-1}$. $\square$

2 The characteristic of an integral domain

Let $R$ be an integral domain. As we have seen in the homework, the function $f : \mathbb{Z} \rightarrow R$ defined by $f(n) = n \cdot 1$ is a ring homomorphism and its image is $\langle 1 \rangle$, the cyclic subgroup of $(R, +)$ generated by 1. There are two possibilities:

(i) 1 has finite order $n$, in which case $\langle 1 \rangle \cong \mathbb{Z}/n\mathbb{Z}$, or (2) 1 has infinite order, in which case $\langle 1 \rangle \cong \mathbb{Z}$.

Proposition 2.1. With notation as above,

(i) If 1 has finite order $n$, then $n = p$ is a prime number, and every nonzero element of $R$ has order $p$. 

(ii) If $1$ has infinite order, then every nonzero element of $R$ has infinite order.

Proof. (i) By definition, $n$ is the smallest positive integer such that $n \cdot 1 = 0$. If $n = ab$, where $a, b \in \mathbb{N}$, then (using homework) $0 = n \cdot 1 = (a \cdot 1)(b \cdot 1)$. Since $R$ is an integral domain, one of $a \cdot 1, b \cdot 1$ is 0. Say $a \cdot 1 = 0$. Then $a \geq n$, but since $a$ divides $n$, we must have $a = n$. Hence in every factorization of $n$, one of the factors is $n$, so by definition $n$ is a prime $p$. Moreover, for every $r \in R$, $p \cdot r = (p \cdot 1)r = 0$, so that the order of $r$ divides $p$. If $r \neq 0$, then its order is greater than 1, hence must equal $p$.

(ii) Let $r \in R$, and suppose that $r$ has (finite) order $n \in \mathbb{N}$, so that $n \cdot r = 0$. As in the proof of (i), write $n \cdot r = (n \cdot 1)r$. Since 1 has infinite order, $n \cdot 1 \neq 0$, and hence $r = 0$. Thus, if $r \neq 0$, then $n \cdot r \neq 0$ for every $n \in \mathbb{N}$. Thus $r$ has infinite order.

Definition 2.2. Let $R$ be an integral domain. If $1 \in R$ has infinite order, we say that the characteristic of $R$ is zero. If $1 \in R$ has finite order, necessarily a prime $p$, we say that the characteristic of $R$ is $p$. In either case we write char $R$ for the characteristic of $R$, so that char $R$ is either 0 or a prime number.

Examples: Clearly, the characteristic of $\mathbb{Z}$ is 0. Also, if $R$ and $S$ are integral domains with $R \leq S$, then clearly char $R = $ char $S$. Thus char $\mathbb{Q}$, char $\mathbb{R}$, char $\mathbb{C}$, char $\mathbb{Q}(\sqrt{2})$, etc. are all 0. On the other hand, the characteristic of $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ is $p$. Thus, the characteristic of $\mathbb{F}_p[x]$ is also $p$, so that $\mathbb{F}_p[x]$ is an example of an infinite integral domain with characteristic $p \neq 0$, and $\mathbb{F}_p[x]$ is not a field. (Note however that a finite integral domain, which automatically has positive characteristic, is always a field.)

3 The field of quotients of an integral domain

We first begin with some general remarks about fields. If $F$ is a field and $r, s \in F$ with $s \neq 0$, we write (as usual) $rs^{-1} = r/s$. Note that $r/s = t/w$ if and only if $rw = st$, since $rw = (sw)r/s$ and $st = (sw)t/w$, and by cancellation. Then the laws for adding and multiplying fractions are forced by associativity and distributivity in $F$: for example,

$$r/s + t/w = rs^{-1} + tw^{-1} = (rw)(sw)^{-1} + (ts)(sw)^{-1} = (rw + ts)(sw)^{-1} = (rw + ts)/(sw).$$
Now suppose that $R$ is an integral domain. We would like to enlarge $R$ to a field, in much the same way that we enlarge $\mathbb{Z}$ to $\mathbb{Q}$. To this end, we construct a set whose elements are “fractions” $r/s$ with $r, s \in R$ and $s \neq 0$. Two fractions $r/s$ and $t/w$ are identified if, as in the discussion above for fields, $rw = st$. The correct way to say this is via equivalence classes: on the set $R \times (R - \{0\})$, define the relation $\sim$ on pairs $(r, s)$ by: $(r, s) \sim (t, w) \iff rw = st$.

**Lemma 3.1.** $\sim$ is an equivalence relation.

**Proof.** We must show $\sim$ is reflexive, symmetric, and transitive. Reflexive: $(r, s) \sim (r, s) \iff rs = sr$, which holds since $R$ is commutative. Symmetric: $(r, s) \sim (t, w) \iff rw = st$, in which case $ts = wr$, hence $(t, w) \sim (r, s)$. Transitive (it is here that we use the fact that $R$ is an integral domain): suppose that $(r, s) \sim (t, w)$ and that $(t, w) \sim (u, v)$, with $s, w, v \neq 0$. By definition $rw = st$ and $tv = wu$. Then $rsw = stv = swu$, hence $w(rv) = w(su)$. Since $w \neq 0$ and $R$ is an integral domain, $rv = su$, hence $(r, s) \sim (u, v)$. Thus $\sim$ is transitive.

Define $Q(R)$, the field of quotients of $R$, to be the set of equivalence classes $(R \times (R - \{0\}))/\sim$. Next we need operations of addition and multiplication on $Q(R)$. As is usually the case with equivalence relations, we define these operations by defining them on representative of equivalence classes, and then check that the operations are in fact well-defined. Define

$$[(r, s)] + [(t, w)] = [(rw + st, sw)]; \quad [(r, s)] \cdot [(t, w)] = [(rt, sw)].$$

**Lemma 3.2.** Let $\sim$ and $Q(R)$ be as above.

(i) The operations of addition and multiplication are well-defined.

(ii) $(Q(R), +, \cdot)$ is a field.

(iii) The function $\rho: R \to Q(R)$ defined by $\rho(r) = [(r, 1)]$ is an injective homomorphism.

**Proof.** These are all straightforward if sometimes tedious calculations. For example, to see (i), suppose that $(r, s) \sim (r', s')$. We shall show that $(rw + st, sw) \sim (r'w + s't, s'w)$ and that $(r, sw) \sim (r't, s'w)$. By definition, $rs' = sr'$. Then

$$(rw + st)(s'w) = rws'w + sts'w = (rs')(w^2) + (ss')(tw) = (r's)(w^2) + (ss')(tw) = (r'w + s't)(sw).$$
Hence \((rw + st, sw) \sim (r′w + s′t, s′w)\). Moreover,
\[
(rt)(s′w) = (rs′)(tw) = (r′s)(tw) = (r′t)(sw).
\]
Hence \((rt, sw) \sim (r′t, s′w)\). Similarly, if \((t, w) \sim (t′, w′)\), then \((rw + st, sw) \sim (rw′ + st′, sw′)\) and that \((rt, sw) \sim (r′t, s′w)\).

To see (ii), we must show first that \((Q(R), +)\) is an abelian group and that multiplication is associative, commutative, and distributes over addition. These are all completely straightforward if long computations. Note that \(\ [(0, 1)] = [(0, r)]\) is the additive identity, that \(\ [(r, s)] \sim [(0, 1)] \iff r = 0,\) and that \(\ [(1, 1)] = [(r, r)]\) is a multiplicative identity. Finally, if \(\ [(r, s)] \neq [(0, 1)],\) so that \(r \neq 0,\) then \(\ [(s, r)] \in Q(R)\) and \(\ [(r, s)][(s, r)] = [(rs, rs)] = [(1, 1)].\) Thus \(Q(R)\) is a field.

To see (iii), defining \(\rho(r) = [(r, 1)]\), we see that
\[
\rho(r + s) = [(r + s, 1)] = [(r, 1)] + [(s, 1)] = \rho(r) + \rho(s);
\]
\[
\rho(rs) = [(rs, 1)] = [(r, 1)][(s, 1)] = \rho(r)\rho(s).
\]
Thus \(\rho\) is a homomorphism. It is injective since \(\rho(r) = \rho(s) \iff (r, 1) \sim (s, 1) \iff r = s.\)

From now on we write \([(r, s)]\) as \(r/s\) or as \(rs^{-1}\) and identify \(r \in R\) with its image \(r/1 \in Q(R)\). In this way we view \(R\) as a subring of \(Q(R)\).

Example: 1) let \(F\) be a field and \(F[x]\) the polynomial ring with coefficients in \(F\). Then we denote \(Q(F[x])\) by \(F(x)\). By definition, the elements of \(F(x)\) are quotients \(f/g\), where \(f, g\) are polynomials with coefficients in \(F\). We call \(F(x)\) the field of rational functions with coefficients in \(F\). In particular, taking \(F = \mathbb{F}_p\), the field of rational functions \(\mathbb{F}_p(x)\) is an example of an infinite field (since it contains a subring isomorphic to the polynomial ring \(\mathbb{F}_p[x]\), which is infinite), whose characteristic is \(p > 0\).

2) If \(R = F\) is already a field, then \((r, s) \sim (rs^{-1}, 1)\). Thus the injective homomorphism \(\rho\) is also surjective, hence an isomorphism, so that \(Q(F) \cong F\).

Remark: In the field of quotients \(Q = Q(\mathbb{Z})\) of \(\mathbb{Z}\), we can always put a fraction \(n/m\) in lowest terms, i.e. we can assume that \(\gcd(n, m) = 1\). This says that the equivalence class \([(n, m)]\) has a “best” representative, if we require in addition, say, that \(m > 0\). Such a choice depends on results about factorization in \(\mathbb{Z}\), and is not possible in a general integral domain.

Finally, we show that \(Q(R)\) has a very general property with respect to injective homomorphisms from \(R\) to a field:
**Proposition 3.3.** Let $R$ be an integral domain, $F$ a field, and $\phi: R \to F$ be an injective homomorphism. Then there exists a unique injective homomorphism $\tilde{\phi}: Q(R) \to F$ such that $\tilde{\phi}(r/1) = \phi(r)$. Finally, if every element of $F$ is of the form $\phi(r)/\phi(s)$ for some $r, s \in R$ with $s \neq 0$, then $\hat{\phi}: Q(R) \to F$ is an isomorphism, and in particular $Q(R) \cong F$.

**Proof.** Clearly, if $\tilde{\phi}$ exists, then we must have

$$
\tilde{\phi}(r/s) = \tilde{\phi}(rs^{-1}) = \tilde{\phi}(r)\tilde{\phi}(s^{-1}) = \tilde{\phi}(r)\tilde{\phi}(s)^{-1} = \tilde{\phi}(r)/\tilde{\phi}(s) = \phi(r)/\phi(s).
$$

(Here we have used the fact, which is easy to check, that $\tilde{\phi}(s^{-1}) = \tilde{\phi}(s)^{-1}$.) This proves that $\tilde{\phi}$ is unique, if it exists. Conversely, we try to define $\tilde{\phi}$ by the formula

$$
\tilde{\phi}(r/s) = \phi(r)/\phi(s).
$$

Here $r/s$ is shorthand for the equivalence class $[(r, s)] \in Q(R)$, and the fraction $\phi(r)/\phi(s) = \phi(r)/\phi(s)^{-1}$ is well-defined in $F$ since, as $\phi$ is injective and $s \neq 0$, $\phi(s) \neq 0$. We must first show that $\tilde{\phi}$ is well-defined, i.e. independent of the choice of representative $(r, s) \in [(r, s)]$. Choosing another representative $(r', s') \in [(r, s)]$, we have by definition $rs' = r's$. Hence

$$
\phi(rs') = \phi(r)\phi(s') = \phi(r's) = \phi(r')\phi(s). \quad \text{Dividing by } \phi(s)\phi(s') \text{ gives}
$$

$$
\phi(r)/\phi(s) = \phi(r)\phi(s')/\phi(s)\phi(s') = \phi(r')\phi(s)/\phi(s)\phi(s') = \phi(r')/\phi(s').
$$

Hence $\tilde{\phi}(r/s) = \phi(r)/\phi(s)$ is independent of the choice of representative $(r, s) \in [(r, s)]$. It is then straightforward to check that $\tilde{\phi}$ is a (ring) isomorphism. To see that it is injective, suppose that $\tilde{\phi}(r/s) = \tilde{\phi}(r'/s')$. Then

$$
\phi(r)/\phi(s) = \phi(r')/\phi(s'),
$$

and hence

$$
\phi(rs') = \phi(r)\phi(s') = \phi(r')\phi(s) = \phi(r's).
$$

Since $\phi$ is injective, $rs' = r's$, and hence $r/s = r'/s'$. Thus $\tilde{\phi}$ is injective. (In fact, this is a general property of ring homomorphisms where the domain is a field.) Finally, if every element of $F$ is of the form $\phi(r)/\phi(s)$ for some $r, s \in R$ with $s \neq 0$, then $\tilde{\phi}$ is also surjective, hence an isomorphism.

Here is a typical way we might apply the proposition:

**Lemma 3.4.** Let $R$ be an integral domain with field of quotients $Q(R)$. Then $Q(R[x])$, the field of quotients of the integral domain $R[x]$, is isomorphic to $Q(R)(x)$, the field of rational functions with coefficients in $Q(R)$.
Proof. Since $R$ is isomorphic to a subring of $Q(R)$, there is a natural homomorphism from $R[x]$ to $Q(R)[x]$, and since $Q(R)[x]$ is isomorphic to a subring of its field of quotients $Q(R)(x)$, there is an injective homomorphism from $R[x]$ to $Q(R)(x)$, which amounts to viewing a polynomial with coefficients in $R$ as a particular example of a rational function with coefficients in $Q(R)$. Hence, by the proposition, there is an injective homomorphism $Q(R)[x] \to Q(R)(x)$. To see that it is surjective, it suffices to show that every rational function with coefficients in $Q(R)$ is a quotient of two polynomials with coefficients in $R$. Given such a quotient $f/g$, suppose that $f = \sum_{i=0}^{n} a_i x^i$ and $g = \sum_{j=0}^{m} b_j x^j$, with $a_i, b_j \in Q(R)$. Then $a_i = r_i/s_i$ with $r_i, s_i \in R$ and $s_i \neq 0$. Likewise, $b_j = t_j/w_j$ with $t_j, w_j \in R$ and $w_j \neq 0$. We then proceed to “clear denominators” in the coefficients: Let $N = s_0 \cdots s_n \cdot w_0 \cdots \cdot w_m = \prod_{i=0}^{n} s_i \cdot \prod_{j=0}^{m} w_j$. Then $N(r_k/s_k) = r_k \prod_{i\neq k} s_i \cdot \prod_{j=0}^{m} w_j \in R$, and similarly $N(t_j/w_j) \in R$. Clearly $Nf \in R[x]$ and $Ng \in R[x]$. Thus

$$
\frac{f}{g} = \frac{f}{g} \cdot \frac{N}{N} = \frac{Nf}{Ng}.
$$

It then follows that $f/g = Nf/Ng$ is a quotient of two polynomials with coefficients in $R$. Hence $Q(R[x]) \cong Q(R)(x)$. \qed

Another application of Proposition 3.3 is as follows: let $F$ be a field of characteristic 0. As we have seen in the homework, the function $f: \mathbb{Z} \to F$ defined by $f(n) = n \cdot 1$ is a ring homomorphism. If char $F = 0$, the homomorphism $f$ is injective. Hence by Proposition 3.3 there is an induced homomorphism $\tilde{f}: \mathbb{Q} \to F$. Its image is the set of all quotients in $F$ of the form $n \cdot 1/m \cdot 1$, with $m \neq 0$. In particular, the image of $\tilde{f}$ is a subfield of $F$ isomorphic to $\mathbb{Q}$. Thus every field of characteristic 0 contains a subfield isomorphic to $\mathbb{Q}$, called the prime subfield. It is the smallest subfield of $F$, hence unique, and it can be described by starting with 1 and making sure that we can perform the operations of addition and subtraction and then automatically multiplication (to get the subring isomorphic to $\mathbb{Z}$), and finally division to get the subfield isomorphic to $\mathbb{Q}$. Here “prime” has nothing to do with prime numbers but simply means that the field $\mathbb{Q}$ is a basic, indivisible object.

A similar statement holds if $F$ is a field of positive characteristic, say char $F = p$ where $p$ is a prime number. In this case, the function $f: \mathbb{Z} \to F$ defined by $f(n) = n \cdot 1$ is still a ring homomorphism, but its kernel is $\langle p \rangle$ and hence its image, as an abelian group, is isomorphic to $\mathbb{Z}/p\mathbb{Z}$. The fact that $f$ is a ring homomorphism implies that the image of $f$, as a ring, is
isomorphic to $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$. Thus, every field of characteristic $p$ contains a subfield isomorphic to $\mathbb{F}_p$, again called the prime subfield. The fields $\mathbb{Q}$ and $\mathbb{F}_p$ are more generally called prime fields. They contain no proper subfields, and every field $F$ contains a unique subfield isomorphic either to $\mathbb{Q}$, if $\text{char} F = 0$, or to $\mathbb{F}_p$, if $\text{char} F = p$. 