Math 533 Winter 2021, Lecture 7: Rings and ideals $\to$ Modules

website: `https://www.cip.ifi.lmu.de/~grinberg/t/21w/`

1. Rings and ideals (cont'd)

1.1. Unique factorization domains ([DF, §8.3]) (cont'd)

Here are some examples of UFDs:

- The ring $\mathbb{Z}$ is a UFD. This is, of course, a consequence of Euclid's famous theorem that says that any positive integer can be uniquely decomposed into a product of primes. Our definition of an irreducible factorization differs slightly from the classical notion of a prime factorization in arithmetic, since our irreducible elements are allowed to be negative and since we only require $r \sim p_1 p_2 \cdots p_n$ (rather than $r = p_1 p_2 \cdots p_n$); but it is pretty easy to conciliate the two concepts by replacing all negative factors by their absolute values. For example, $(-3, -2, 2)$ is an irreducible factorization of -12, since $-12 \sim (-3) \cdot (-2) \cdot 2$; but of course it corresponds to the classical prime factorization $12 = 3 \cdot 2 \cdot 2$ of the positive integer 12.
- Any field is a UFD, since every nonzero element is a unit and thus has the empty tuple as its only irreducible factorization.
- We shall soon see that every PID is a UFD.
- The rings

$$\begin{aligned}\mathbb{Z}[2i] &= \{a + b \cdot 2i \mid a, b \in \mathbb{Z}\} \\ &= \{\text{Gaussian integers with an even imaginary part}\}\end{aligned}$$

and

$$\mathbb{Z}\left[\sqrt{-5}\right] = \left\{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\right\}$$

are not UFDs.

In the previous lecture, we proved that an element of a PID is prime if and only if it is irreducible. We shall now prove the same result for UFDs (which is stronger, as we will soon see that every PID is a UFD):

Proposition 1.1.1. Let R be a UFD. Let $r \in R$. Then, r is prime if and only if r is irreducible.

Proof. $\Longrightarrow$: If r is prime, then r is irreducible. In fact, we have already proved this in the previous lecture (and not just for a UFD, but for any integral domain).

$\Longleftarrow$: Assume that r is irreducible. We must show that r is prime.

Let $a, b \in R$ satisfy $r \mid ab$. We must prove that $r \mid a$ or $r \mid b$.

Assume the contrary. Thus, neither a nor b is a multiple of r. Hence, in particular, a and b are nonzero (since 0 is a multiple of r). Thus, a and b have irreducible factorizations (since R is a UFD). Let $(p_1, p_2, \ldots, p_n)$ and $(q_1, q_2, \ldots, q_m)$ be irreducible factorizations of a and b. Thus, $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_m$ are irreducible elements of R satisfying

$$a \sim p_1 p_2 \cdots p_n \qquad \text{and} \qquad b \sim q_1 q_2 \cdots q_m.$$

Multiplying $a \sim p_1 p_2 \cdots p_n$ with $b \sim q_1 q_2 \cdots q_m$, we see that

$$ab \sim p_1 p_2 \cdots p_n q_1 q_2 \cdots q_m \tag{1}$$

(since a product of two units is again a unit).

However, $r \mid ab$. Thus, there exists a $q \in R$ such that $ab = rq$. Consider this q. Note that ab is nonzero (since a and b are nonzero, but R is an integral domain). Thus, q is nonzero (since $q = 0$ would imply $ab = r \underbrace{q}_{=0} = 0$, which would contradict the previous sentence). Hence, q has an irreducible factorization (since R is a UFD). Let $(s_1, s_2, \ldots, s_k)$ be an irreducible factorization of q. Thus, $s_1, s_2, \ldots, s_k$ are irreducible elements of R satisfying $q \sim s_1 s_2 \cdots s_k$. From $q \sim s_1 s_2 \cdots s_k$, we obtain $rq \sim r s_1 s_2 \cdots s_k$. Since $ab = rq$, this rewrites as

$$ab \sim r s_1 s_2 \cdots s_k. \tag{2}$$

Now, we conclude that the two tuples

$$(p_1, p_2, \ldots, p_n, q_1, q_2, \ldots, q_m) \text{ and } (r, s_1, s_2, \ldots, s_k)$$

are two irreducible factorizations of ab (since all their entries $p_1, p_2, \ldots, p_n, q_1, q_2, \ldots, q_m$ and $r, s_1, s_2, \ldots, s_k$ are irreducible, and since (1) and (2) hold). Thus, by the uniqueness condition in the definition of a UFD (which says that the irreducible factorization of an element is unique up to associates), these two tuples must be identical up to associates. In particular, every entry of the second tuple must be associate to some entry of the first. Hence, in particular, the entry r of the second factorization must be associate to one of the entries $p_1, p_2, \ldots, p_n, q_1, q_2, \ldots, q_m$ of the first. In other words, we must have

$$r \sim p_i \text{ for some } i \in \{1, 2, \ldots, n\} \tag{3}$$

or

$$r \sim q_j \text{ for some } j \in \{1, 2, \ldots, m\}. \tag{4}$$

However, both of these possibilities lead to contradictions: Indeed, if (3) holds, then we have $r \mid a$ (since[1] $r \sim p_i \mid p_1 p_2 \cdots p_n \sim a$), which contradicts the fact that a is not a multiple of r. Likewise, if (4) holds, then we have $r \mid b$, which contradicts the fact that b is not a multiple of r. Thus, we get a contradiction in either case, and our proof is complete. □

If R is a UFD, and if $r \in R$ is nonzero, then r is associate to a finite product $p_1 p_2 \cdots p_n$ of irreducible elements (by the definition of a UFD). This product can be simplified by collecting associate factors together. For example, in $\mathbb{Z}$, we have

$$-24 = 2 \cdot (-2) \cdot 2 \cdot 3 = -2^3 \cdot 3.$$

Here is what we get in general:

Proposition 1.1.2. Let R be a UFD. Let $r \in R$ be nonzero. Then:

(a) There exists a list $(q_1, q_2, \ldots, q_k)$ of **mutually non-associate** irreducible elements $q_1, q_2, \ldots, q_k \in R$ as well as a list $(e_1, e_2, \ldots, e_k)$ of **positive** integers such that

$$r \sim q_1^{e_1} q_2^{e_2} \cdots q_k^{e_k}.$$

We shall refer to these two lists as the **prime power factorization** of r.

(b) These two lists are unique up to associates and up to simultaneous permutation. (That is, any two prime power factorizations of r can be transformed into one another by replacing the irreducible elements $q_1, q_2, \ldots, q_k$ by associates, and reordering them while carrying the exponents $e_1, e_2, \ldots, e_k$ along with them.)

Proof of Proposition 1.1.2. **(a)** Start with an irreducible factorization of r, and collect associate factors together. For example, if an irreducible factorization of r has the form $(p_1, p_2, p_3, p_4, p_5, p_6)$ with $p_1 \sim p_4$ and $p_2 \sim p_5 \sim p_6$ (and no other associate relations between its entries), then

$$r \sim p_1 p_2 p_3 p_4 p_5 p_6 \sim p_1 p_2 p_3 p_1 p_2 p_2 = p_1^2 p_2^3 p_3,$$

and this is a prime power factorization of r.

(b) This follows from the uniqueness of an irreducible factorization (up to associates). □

Proposition 1.1.3. Let R be a UFD. Let $a, b \in R$ be nonzero. Then, there exists a list $(p_1, p_2, \ldots, p_n)$ of **mutually non-associate** irreducible elements $p_1, p_2, \ldots, p_n \in R$ as well as two lists $(e_1, e_2, \ldots, e_n)$ and $(f_1, f_2, \ldots, f_n)$ of **nonnegative** integers such that

$$a \sim p_1^{e_1} p_2^{e_2} \cdots p_n^{e_n} \qquad \text{and} \qquad b \sim p_1^{f_1} p_2^{f_2} \cdots p_n^{f_n}.$$

[1] We will use the fact that associates divide each other: i.e., if u and v are two elements of R satisfying $u \sim v$, then $u \mid v$.

Proof. Proposition 1.1.2 shows that a and b have prime power factorizations

$$a \sim q_1^{e_1} q_2^{e_2} \cdots q_k^{e_k} \qquad \text{and} \qquad b \sim r_1^{f_1} r_2^{f_2} \cdots r_m^{f_m}.$$

All we need now is to reconcile these prime power factorizations so that they contain the same irreducible elements (albeit possibly with 0 exponents). For this purpose, we do the following steps:

1. If some of the q_i are associate to some of the r_j, then we replace these q_i by the respective r_j.
2. If some of the q_i don't appear among the r_j, then we insert q_i^0 factors into the prime power factorization of b.
3. If some of the r_j don't appear among the q_i, then we insert r_j^0 factors into the prime power factorization of a.

For example, if $R = \mathbb{Z}$ and $a = 12$ and $b = 45$, and if we start with the prime power factorizations $a \sim 2^2 \cdot (-3)^1$ and $b \sim 3^2 \cdot 5^1$, then Step 1 transforms the prime power factorization of a into $a \sim 2^2 \cdot 3^1$ (since the -3 is replaced by the 3 from the prime power factorization of b); Step 2 then inserts a 2^0 factor into the prime power factorization of b (so it becomes $b \sim 2^0 \cdot 3^2 \cdot 5^1$); Step 3 then inserts a 5^0 factor into the prime power factorization of a (so it becomes $a \sim 2^2 \cdot 3^1 \cdot 5^0$). The resulting factorizations are $a \sim 2^2 \cdot 3^1 \cdot 5^0$ and $b \sim 2^0 \cdot 3^2 \cdot 5^1$, just as promised by Proposition 1.1.3. □

Proposition 1.1.4. Let R be a UFD. Let $a, b \in R$ be nonzero. Let $(p_1, p_2, \ldots, p_n)$, $(e_1, e_2, \ldots, e_n)$ and $(f_1, f_2, \ldots, f_n)$ be as in Proposition 1.1.3. Then:

(a) The element

$$p_1^{\min\{e_1, f_1\}} p_2^{\min\{e_2, f_2\}} \cdots p_n^{\min\{e_n, f_n\}}$$

is a gcd of a and b.

(b) The element

$$p_1^{\max\{e_1, f_1\}} p_2^{\max\{e_2, f_2\}} \cdots p_n^{\max\{e_n, f_n\}}$$

is an lcm of a and b.

Proof. This is done just as it is commonly done for integers in elementary number theory. The details are LTTR. (See, e.g., the proof of Proposition 1.11 in `https://www.math.columbia.edu/~rf/factorization1.pdf` for some details on the proof of part **(a)**; the proof of part **(b)** is similar.) □

Corollary 1.1.5. Any two elements in a UFD have a gcd and an lcm.

Proof. Let a and b be two elements of a UFD R. We must show that a and b have a gcd and an lcm.

If $b = 0$, then this is easy (just show that a is a gcd of a and 0, and that 0 is an lcm of a and 0). Thus, we WLOG assume that $b \neq 0$. For a similar reason, we WLOG assume that $a \neq 0$. Hence, Proposition 1.1.3 shows that there exists a list $(p_1, p_2, \ldots, p_n)$ of **mutually non-associate** irreducible elements $p_1, p_2, \ldots, p_n \in R$ as well as two lists $(e_1, e_2, \ldots, e_n)$ and $(f_1, f_2, \ldots, f_n)$ of **nonnegative** integers such that

$$a \sim p_1^{e_1} p_2^{e_2} \cdots p_n^{e_n} \qquad \text{and} \qquad b \sim p_1^{f_1} p_2^{f_2} \cdots p_n^{f_n}.$$

Thus, Proposition 1.1.4 shows that a and b have a gcd and a lcm. □

Finally, as promised, let us prove the following theorem, which provides us many UFDs to apply the above results to:

Theorem 1.1.6. Any PID is a UFD.

I won't prove Theorem 1.1.6 here; a proof can be found in [DF, §8.3, Theorem 14]. The proof of the existence of an irreducible factorization is rather philosophical and non-constructive; it yields no algorithm for actually finding such a factorization. (And indeed, there are UFDs in which finding such a factorization is algorithmically impossible.) The proof of the uniqueness of an irreducible factorization is an analogue of the proof you know from elementary number theory (since we know that irreducible elements are prime).

The following corollary combines several results we have seen above in a convenient hierarchy:

Corollary 1.1.7. We have

$$\begin{aligned}\{\text{fields}\} &\subseteq \{\text{Euclidean domains}\} \subseteq \{\text{PIDs}\} \subseteq \{\text{UFDs}\} \\ &\subseteq \{\text{integral domains}\} \subseteq \{\text{commutative rings}\} \subseteq \{\text{rings}\}.\end{aligned}$$

1.2. Application: Fermat's $p = x^2 + y^2$ theorem ([DF, §8.3])

As an application of some of the above, we will show a result of Fermat:[2]

Theorem 1.2.1 (Fermat's two-squares theorem)**.** Let p be a prime number such that $p \equiv 1 \bmod 4$. Then, p can be written as a sum of two perfect squares.

[2]The word "prime number" is understood as in classical number theory – i.e., a positive integer $p > 1$ whose only positive divisors are 1 and p. In particular, negative numbers are not allowed as prime numbers, even though they are prime elements of $\mathbb{Z}$.

I will give a rough outline of how this can be proved using rings. Some of the steps I will leave to you (they will be problems on homework set #2).

First, a general curious fact about primes:

Theorem 1.2.2 (Wilson's theorem)**.** Let p be a prime. Then, $(p-1)! \equiv -1 \bmod p$.

For example, for $p = 5$, this is saying that $4! \equiv -1 \bmod 5$. And indeed, $4! = 24 \equiv -1 \bmod 5$.

Proof of Theorem 1.2.2. We must show that $(p-1)! \equiv -1 \bmod p$. Equivalently, we must show that

$$\overline{(p-1)!} = \overline{-1} \text{ in } \mathbb{Z}/p. \tag{5}$$

However, $(p-1)! = 1 \cdot 2 \cdot \cdots \cdot (p-1)$, so that

$$\overline{(p-1)!} = 1 \cdot 2 \cdot \cdots \cdot (p-1) = \overline{1} \cdot \overline{2} \cdot \cdots \cdot \overline{p-1}. \tag{6}$$

But $\mathbb{Z}/p$ is a field (as we know) with p elements $\overline{0}, \overline{1}, \ldots, \overline{p-1}$. Its nonzero elements $\overline{1}, \overline{2}, \ldots, \overline{p-1}$ are thus its units. In other words, its group of units $(\mathbb{Z}/p)^\times$ is precisely the set $\left\{\overline{1}, \overline{2}, \ldots, \overline{p-1}\right\}$ (and all the $p-1$ elements $\overline{1}, \overline{2}, \ldots, \overline{p-1}$ are distinct). Hence,

$$\prod_{a \in (\mathbb{Z}/p)^\times} a = \overline{1} \cdot \overline{2} \cdot \cdots \cdot \overline{p-1}. \tag{7}$$

Recall that $(\mathbb{Z}/p)^\times$ is a group. In particular, any unit has an inverse, which is again a unit. The units $\overline{1}$ and $\overline{-1}$ are their own inverses (since $\overline{1} \cdot \overline{1} = \overline{1 \cdot 1} = \overline{1}$ and $\overline{-1} \cdot \overline{-1} = \overline{(-1) \cdot (-1)} = \overline{1}$), and they are the only units that are their own inverses (you will prove this in Exercise 5 **(a)** on homework set #2). The inverse of the inverse of a unit a is a. Hence, in the product $\prod_{a \in (\mathbb{Z}/p)^\times} a$, we can pair up each factor other than $\overline{1}$ and $\overline{-1}$ with its inverse:

$$\begin{aligned}\prod_{a \in (\mathbb{Z}/p)^\times} a &= \underbrace{\left(a_1 \cdot a_1^{-1}\right)}_{=\overline{1}} \cdot \underbrace{\left(a_2 \cdot a_2^{-1}\right)}_{=\overline{1}} \cdot \cdots \cdot \underbrace{\left(a_k \cdot a_k^{-1}\right)}_{=\overline{1}} \cdot \overline{1} \cdot \overline{-1} \\ &= \overline{1} \cdot \overline{1} \cdot \cdots \cdot \overline{1} \cdot \overline{1} \cdot \overline{-1} = \overline{-1}.\end{aligned} \tag{8}$$

Now, (6) becomes

$$\begin{aligned}\overline{(p-1)!} &= \overline{1} \cdot \overline{2} \cdot \cdots \cdot \overline{p-1} = \prod_{a \in (\mathbb{Z}/p)^\times} a \qquad \text{(by (7))} \\ &= \overline{-1} \qquad \text{(by (8))}.\end{aligned}$$

This proves (5) and thus Theorem 1.2.2.

(Caveat: The above was a little bit wrong for $p = 2$; in that case, the factors $\overline{1}$ and $\overline{-1}$ are actually one and the same factor. But our proof can easily be adapted to the above.) $\square$

Corollary 1.2.3. Let p be a prime such that $p \equiv 1 \bmod 4$. Let $u = \dfrac{p-1}{2} \in \mathbb{N}$. Then, $u!^2 \equiv -1 \bmod p$.

Proof. This follows from exercise 5 **(b)** on homework set #2. □

Now, recall the ring $\mathbb{Z}[i]$ of Gaussian integers. Let $N : \mathbb{Z}[i] \to \mathbb{N}$ be the map that sends each Gaussian integer $a + bi$ (with $a, b \in \mathbb{Z}$) to $a^2 + b^2 \in \mathbb{N}$. It is straightforward to see:

Proposition 1.2.4. We have $N(\alpha\beta) = N(\alpha) N(\beta)$ for any $\alpha, \beta \in \mathbb{Z}[i]$.

Proof. One way to do so is by first showing that $N(\gamma) = \gamma\overline{\gamma}$ for each $\gamma \in \mathbb{Z}[i]$ (where $\overline{\gamma}$ denotes the complex conjugate of γ). Another is by direct computation: Writing α and β as $\alpha = a + bi$ and $\beta = c + di$, we have $\alpha\beta = (a + bi)(c + di) = (ac - bd) + (ad + bc) i$ and therefore

$$\begin{aligned} N(\alpha\beta) &= N((ac - bd) + (ad + bc) i) = (ac - bd)^2 + (ad + bc)^2 \\ &= a^2c^2 - 2acbd + b^2d^2 + a^2d^2 + 2adbc + b^2c^2 \\ &= a^2c^2 + b^2d^2 + a^2d^2 + b^2c^2 = \underbrace{\left(a^2 + b^2\right)}_{=N(\alpha)} \underbrace{\left(c^2 + d^2\right)}_{=N(\beta)} = N(\alpha) N(\beta). \end{aligned}$$

□

Using this fact, we can characterize the units of $\mathbb{Z}[i]$:

Corollary 1.2.5. Let $\alpha \in \mathbb{Z}[i]$. Then, we have the following equivalence:

$$(\alpha \text{ is a unit of } \mathbb{Z}[i]) \iff (N(\alpha) = 1) \iff (\alpha \in \{1, i, -1, -i\}).$$

Proof. This is exercise 6 **(d)** on homework set #2. □

The next lemma is also easy to see:

Lemma 1.2.6. Let α and β be Gaussian integers such that $\alpha \neq 0$. Then, $\alpha \mid \beta$ holds in $\mathbb{Z}[i]$ if and only if $\dfrac{\beta}{\alpha}$ is a Gaussian integer.

Proof. This is proved just as the analogous statement for integers is proved. □

Now we can prove Theorem 1.2.1:

Proof of Theorem 1.2.1. Let $u = \dfrac{p-1}{2}$. Then, $u \in \mathbb{N}$ (actually, $p \equiv 1 \bmod 4$ implies that u is even). Corollary 1.2.3 shows that $u!^2 \equiv -1 \bmod p$. That is,

$$p \mid u!^2 - \underbrace{(-1)}_{=i^2} = u!^2 - i^2 = (u! - i)(u! + i).$$

This is a divisibility in $\mathbb{Z}$, thus also in $\mathbb{Z}[i]$.

The number p is a prime number, and thus prime in $\mathbb{Z}$; but this does **not** mean that it is prime in $\mathbb{Z}[i]$. And in fact, we claim that it isn't. Indeed, if p was prime in $\mathbb{Z}[i]$, then the divisibility $p \mid (u! - i)(u! + i)$ would entail that $p \mid u! - i$ or $p \mid u! + i$; however, neither $p \mid u! - i$ nor $p \mid u! + i$ is true[3].

Thus, we know that p is not prime in $\mathbb{Z}[i]$. But $\mathbb{Z}[i]$ is a Euclidean domain (as we proved at the end of lecture 5), and thus a PID (since we have shown in lecture 6 that any Euclidean domain is a PID). Hence, every irreducible element of $\mathbb{Z}[i]$ is a prime element of $\mathbb{Z}[i]$ (by a proposition we proved in lecture 6). Thus, p cannot be irreducible in $\mathbb{Z}[i]$ (since p is not prime in $\mathbb{Z}[i]$).

However, p is nonzero and not a unit of $\mathbb{Z}[i]$ (since $\dfrac{1}{p}$ is not a Gaussian integer). Therefore, since p is not irreducible, there exist two elements $\alpha, \beta \in \mathbb{Z}[i]$ that satisfy $\alpha\beta = p$ but are not units (by the definition of "irreducible"). Consider these α and β.

From $\alpha\beta = p$, we obtain $N(\alpha\beta) = N(p) = N(p + 0i) = p^2 + 0^2 = p^2$. Thus, $p^2 = N(\alpha\beta) = N(\alpha) N(\beta)$ (by Proposition 1.2.4). However, $N(\alpha)$ and $N(\beta)$ are nonnegative integers (since N is a map $\mathbb{Z}[i] \to \mathbb{N}$). Since p is prime, the only ways to write p^2 as a product of two nonnegative integers are $p^2 = 1 \cdot p^2$ and $p^2 = p^2 \cdot 1$ and $p^2 = p \cdot p$ (by the classical prime factorization theorem from number theory). Hence, the equality $p^2 = N(\alpha) N(\beta)$ (with $N(\alpha)$ and $N(\beta)$ being nonnegative integers) entails that we must be in one of the following two cases:

Case 1: One of the two numbers $N(\alpha)$ and $N(\beta)$ is 1, and the other is p^2.

Case 2: Both numbers $N(\alpha)$ and $N(\beta)$ are p.

Let us consider Case 1. In this case, one of the two numbers $N(\alpha)$ and $N(\beta)$ is 1. We WLOG assume that $N(\alpha) = 1$ and $N(\beta) = p^2$ (since the other possibility can be transformed into this one by swapping α with β). Now, recall that $N(\alpha) = 1$ is equivalent to α being a unit (because of Corollary 1.2.5). However, α is not a unit. This is a contradiction. Hence, Case 1 is impossible.

Thus, we must be in Case 2. In other words, $N(\alpha) = p$ and $N(\beta) = p$.

[3] This is easiest to see using Lemma 1.2.6: Indeed, if we had $p \mid u! - i$, then Lemma 1.2.6 would entail that $\dfrac{u! - i}{p}$ is a Gaussian integer; however, $\dfrac{u! - i}{p} = \dfrac{u!}{p} + \dfrac{-1}{p} i$ is not a Gaussian integer (since its imaginary part $\dfrac{-1}{p}$ is not an integer). Thus, we don't have $p \mid u! - i$. For a similar reason, we don't have $p \mid u! + i$.

Now, α is a Gaussian integer, so we can write it as $\alpha = x + yi$ for some $x, y \in \mathbb{Z}$. Therefore, using these x, y, we have $N(\alpha) = x^2 + y^2$. Hence, $x^2 + y^2 = N(\alpha) = p$. Thus, p is a sum of two perfect squares; Theorem 1.2.1 is proven. $\square$

More about decompositions of integers into sums of perfect squares can be found

- in [DF, §8.3];
- in Keith Conrad's `https://kconrad.math.uconn.edu/math5230f12/handouts/Zinotes.pdf` ;
- in §4.2 of my `https://www.cip.ifi.lmu.de/~grinberg/t/19s/notes.pdf` .

In particular, one can describe precisely which integers can be written as sums of two perfect squares, and in how many ways; and most of these results can be neatly proved using Gaussian integers.

Lagrange proved that every nonnegative integer can be written as a sum of four squares. These days, one of the shortest proofs of this fact uses the so-called *Hurwitz quaternions* – a quaternion analogue of Gaussian integers. See `https://en.wikipedia.org/wiki/Lagrange's_four-square_theorem` or `https://www.mathcs.duq.edu/~haensch/411Materials/Quaternions.pdf` or `https://www.math.brown.edu/reschwar/M153/lagrange.pdf` for the proof.

2. Modules ([DF, Chapter 10])

We now move on from studying rings themselves to studying **modules** over rings. In many ways, modules are even more important than rings, as their definition offers more freedom (which is indeed amply used throughout mathematics). Some would argue that the notion of a ring is merely an ancillary character to that of a module.

2.1. Definition and examples ([DF, §10.1])

For every ring R, there are two notions of an "R-module": The "left R-modules" and the "right R-modules". Let us define the left ones:

Definition 2.1.1. Let R be a ring. A **left R-module** (or a **left module over** R) means a set M equipped with

- a binary operation $+$ (that is, a map from $M \times M$ to M) that is called **addition**;
- an element $0_M \in M$ that is called the **zero element** or the **zero vector** or just the **zero**, and is just denoted by 0 when there is no ambiguity;

- a map from $R \times M$ to M that is called the **action of R on M**, and is written as multiplication (i.e., we denote the image of a pair $(r, m) \in R \times M$ under this map by rm or $r \cdot m$)

such that the following properties (the "**module axioms**") hold:

- $(M, +, 0)$ is an abelian group.
- The **right distributivity law** holds: We have $(r + s)\, m = rm + sm$ for all $r, s \in R$ and $m \in M$.
- The **left distributivity law** holds: We have $r\,(m + n) = rm + rn$ for all $r \in R$ and $m, n \in M$.
- The **associativity law** holds: We have $(rs)\, m = r\,(sm)$ for all $r, s \in R$ and $m \in M$.
- We have $0_R m = 0_M$ for every $m \in M$.
- We have $r \cdot 0_M = 0_M$ for every $r \in R$.
- We have $1m = m$ for every $m \in M$.

When M is a left R-module, the elements of M are called **vectors**, and the elements of R are called **scalars**.

As the name "left R-module" suggests, there is an analogous notion of a **right R-module**. In this latter notion, the action is not a map from $R \times M$ to M, but rather a map from $M \times R$ to M, and we accordingly use the notation mr (rather than rm) for the image of a pair (m, r) under this map. The axioms for a right R-module are similar to the above axioms for a left R-module. (For example, the associative law will now be saying $m\,(rs) = (mr)\, s$ for all $r, s \in R$ and $m \in M$.)

When R is commutative, any left R-module becomes a right R-module in a natural way:

Proposition 2.1.2. Let R be a commutative ring. Then, we can make any left R-module M into a right R-module by setting

$$mr = rm \qquad \text{for all } r \in R \text{ and } m \in M. \tag{9}$$

Similarly, we can make any right R-module into a left R-module. These two transformations are mutually inverse, so we shall use them to identify left R-modules with right R-modules. This will allow us to use the words "left R-module" and "right R-module" interchangeably, and just speak of "R-modules" instead (without specifying whether they are left or right). (Note that this is not allowed when R is not commutative!)

When R is a field, the R-modules are also known as the **R-vector spaces**. These are precisely the vector spaces you have seen in a linear algebra class. A left R-module over an arbitrary ring R is just the natural generalization of a vector space. But while vector spaces have a very predictable structure (in particular, a vector space is uniquely determined up to isomorphism by its dimension), modules can be wild (although the "nice" families of modules, like R^n for $n \in \mathbb{N}$, still exist for every ring). The wilder a ring is, the more diverse are its modules.

One more remark about Definition 2.1.1: The "$0_R m = 0_M$" and "$r \cdot 0_M = 0_M$" axioms are actually redundant (i.e., they follow from the other axioms). I leave it to you to check this.

We will soon see some examples of R-modules; but let us first define R-submodules. If you have seen subspaces of a vector space, this definition won't surprise you:

Definition 2.1.3. Let M be a left R-module. An **R-submodule** (or, to be more precise, a **left R-submodule**) of M means a subset N of M such that

- $a + b \in N$ for any $a, b \in N$;
- $ra \in N$ for any $r \in R$ and $a \in N$;
- $0 \in N$ (where 0 means 0_M).

In other words, an R-submodule of M means a subgroup of the additive group $(M, +, 0)$ that is also closed under scaling by all scalars $r \in R$. Here, **scaling** by an $r \in R$ means the map $M \to M$, $m \mapsto rm$. This map is a group endomorphism[4] of $(M, +, 0)$ (check this!).

(All three axioms in Definition 2.1.3 have names: The "$a + b \in N$" axiom is called "N is closed under addition"; the "$ra \in N$" axiom is called "N is closed under scaling"; the "$0 \in N$" axiom is called "N contains the zero vector".)

An R-submodule of a left R-module M becomes a left R-module in its own right (just as a subring of a ring becomes a ring).[5]

Here are some examples of modules:

- Let R be any ring. Then, R itself becomes a left R-module: Just define the action to be the multiplication of R.

  The R-submodules of this left R-module R are the subsets L of R that are closed under addition and contain 0 and satisfy $ra \in L$ for all $r \in R$ and $a \in L$. These subsets are called the **left ideals** of R. When R is

[4] A **group endomorphism** of a group G means a group homomorphism from G to G.

[5] This is not completely obvious! To prove this, you have to check that any R-submodule of a left R-module M is closed under taking additive inverses. This follows from Proposition 1.2.2 in Lecture 8 below. (Or you can prove it on your own; it is not hard.)

commutative, these are precisely the ideals of R. For general R, however, the notion of an ideal is more restrictive than the notion of a left ideal.

For example, if R is the matrix ring $\mathbb{Q}^{2\times 2}$, then the only ideals of R are $\{0_{2\times 2}\}$ and R itself, but R has infinitely many left ideals (for example, the set of all matrices of the form $\begin{pmatrix} 0 & a \\ 0 & b \end{pmatrix}$ is a left ideal).

- Let R be any ring, and let $n \in \mathbb{N}$. Then,

$$R^n = \{(a_1, a_2, \ldots, a_n) \mid \text{all } a_i \text{ belong to } R\}$$

is a left R-module, where addition and action are defined entrywise: e.g., the action is defined by

$$r \cdot (a_1, a_2, \ldots, a_n) = (ra_1, ra_2, \ldots, ra_n) \text{ for all } r \in R \text{ and } a_1, a_2, \ldots, a_n \in R.$$

The zero vector of this R-module R^n is $(0, 0, \ldots, 0)$.

Note that the zero vector of an R-module is uniquely determined by its addition (in fact, this is true for any group); thus, we don't even need to specify it explicitly when we define an R-module.