# Factorization in Integral Domains

Throughout these notes, $R$ denotes an **integral domain**.

## 1 Unique factorization domains and principal ideal domains

**Definition:** For $r, s \in R$, we say that $r$ *divides* $s$ (written $r|s$) if there exists a $t \in R$ such that $s = tr$. An element $u \in R$ is a *unit* if it has a multiplicative inverse, i.e. if there exists an element $v \in R$ such that $uv = 1$. The (multiplicative) group of units is denoted $R^*$. If $r, s \in R$, then $r$ and $s$ are *associates* if there exists a unit $u \in R^*$ such that $r = us$. In this case, $s = u^{-1}r$, and indeed the relation that $r$ and $s$ are associates is an equivalence relation. We say that $r \in R$ is *irreducible* if $r \neq 0$, $r$ is not a unit, and, for all $s \in R$, if $s$ divides $r$ then either $s$ is a unit or $s$ is an associate of $r$. In other words, if $r = st$ for some $t \in R$, then one of $s$ or $t$ is a unit (and hence the other is an associate of $r$). If $r \in R$ with $r \neq 0$ and $r$ is not a unit, then $r$ is *reducible* if it is not irreducible.

**Examples:** 1) $R = \mathbb{Z}$. The units $\mathbb{Z}^* = \pm 1$. Two integers $n$ and $m$ are associates $\iff m = \pm n$.

2) $R = F[x]$, $F$ a field. The units in $F[x]$ are: $(F[x])^* = F^*$, the set of constant nonzero polynomials. Hence, if $F$ is infinite, there are an infinite number of units. Two polynomials $f(x)$ and $g(x)$ are associates $\iff$ there exists a $c \in F^*$ with $g(x) = cf(x)$.

3) $R = \mathbb{Z}[i]$, the *Gaussian integers*. The units $(\mathbb{Z}[i])^* = \{\pm 1, \pm i\}$. Two elements $\alpha, \beta \in \mathbb{Z}[i]$ are associates $\iff \alpha = \pm\beta$ or $\alpha = \pm i\beta$.

4) $R = \mathbb{Z}[\sqrt{2}]$. As we have seen on the homework, $1 + \sqrt{2}$ is a unit of infinite order. In fact, $(\mathbb{Z}[\sqrt{2}])^* \cong \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$.

4) $R = \mathbb{Z}[\sqrt{-2}]$. As we have seen on the midterm, $(\mathbb{Z}[\sqrt{-2}])^* = \pm 1$.

**Definition:** $R$ is a *unique factorization domain* (UFD) if

(i) for every $r \in R$ not 0 or a unit, there exist irreducibles $p_1, \dots, p_n \in R$ such that $r = p_1 \cdots p_n$, and

(ii) if $p_i, 1 \le i \le n$ and $q_j, 1 \le j \le m$ are irreducibles such that $p_1 \cdots p_n = q_1 \cdots q_m$, then $n = m$ and, after reordering, $p_i$ and $q_i$ are associates.

Note that two separate issues are involved: (i) the **existence** of some factorization of $r$ into irreducibles and (ii) the **uniqueness** of a factorization. As we shall see, these two questions are in general unrelated.

Given an element $r$ in a UFD, not 0 or a unit, it is often more natural to factor $r$ by grouping together all of the associated irreducibles (after making some choices). Hence, such an $r$ can always be written as

$$r = up_1^{a_1} \cdots p_n^{a_n},$$

where $u$ is a unit, the $p_i$ are irreducibles, $a_i > 0$, and, for $i \neq j$, $p_i$ and $p_j$ are not associates, and such a product is essentially unique in the following sense: if also

$$r = vq_1^{b_1} \cdots q_m^{b_m},$$

where $v$ is a unit, the $q_j$ are irreducibles, $b_j > 0$, and, for $k \neq \ell$, $q_k$ and $q_\ell$ are not associates, then $n = m$ and, after reordering, $p_i$ and $q_i$ are associates and $a_i = b_i$.

**Definition:** $R$ is a *principal ideal domain* (PID) if every ideal $I$ of $R$ is principal, i.e. for every ideal $I$ of $R$, there exists $r \in R$ such that $I = (r)$.

**Examples:** The rings $\mathbb{Z}$ and $F[x]$, where $F$ is a field, are PID's.

We shall prove later: A principal ideal domain is a unique factorization domain. However, there are many examples of UFD's which are not PID's. For example, if $n \geq 2$, then the polynomial ring $F[x_1, \ldots, x_n]$ is a UFD but not a PID. Likewise, $\mathbb{Z}[x]$ is a UFD but not a PID, as is $\mathbb{Z}[x_1, \ldots, x_n]$ for all $n \geq 1$.

**Definition:** Let $R$ be an integral domain. Let $r, s \in R$, not both 0. A *greatest common divisor* (gcd) of $r$ and $s$ is an element $d \in R$ such that $d|r$, $d|s$, and if $e \in R$ and $e|r$, $e|s$, then $e|d$. If a gcd of $r$ and $s$ exists, it is unique up to a unit (i.e. any two gcd's of $r$ and $s$ are associates). The elements $r$ and $s$ are *relatively prime* if $\gcd(r, s) = 1$; equivalently, if $d \in R$ and $d|r$, $d|s$, then $d$ is a unit.

**Proposition:** if $R$ is a UFD, then the gcd of two elements $r, s \in R$, not both 0, exists.

**Proof.** If say $r = 0$, then the gcd of $r$ and $s$ exists and is $s$. If $r$ is a unit, then the gcd of $r$ and $s$ exists and is a unit. So we may clearly assume that

$r$ is neither 0 nor a unit, and likewise that $s$ is neither 0 nor a unit. Then we can factor both $r$ and $s$ as in the comments after the definition of a UFD. In fact, it is clear that we can write

$$r = up_1^{a_1} \cdots p_k^{a_k}, \qquad s = vp_1^{b_1} \cdots p_k^{b_k}$$

where $u$ and $v$ are units, the $p_i$ are irreducibles, $a_i, b_i \geq 0$, and, for $i \neq j$, $p_i$ and $p_j$ are not associates. (Here, we set $a_i = 0$ if $p_i$ is not a factor of $r$, and similarly for $b_i$.) Then set

$$t = p_1^{c_1} \cdots p_k^{c_k},$$

where $c_i = \min\{a_i, b_i\}$. We claim that $t$ is a gcd of $r$ and $s$. Clearly $t|r$ and $t|s$. If now $w|r$ and $w|s$ and $q$ is an irreducible factor of $w$, then $q = p_i$ for some $i$, and if $d_i$ is the largest integer such that $p_i^{d_i}|w$, then since $p_i^{d_i}|r$ and $p_i^{d_i}|s$, $d_i \leq a_i$ and $d_i \leq b_i$. Hence $d_i \leq c_i$. It then follows by taking the factorization of $w$ into powers of the $p_i$ tines a unit that $w|t$. Hence $t$ is a gcd of $r$ and $s$. $\square$

**Lemma:** If $R$ is a UFD and $p, r, s \in R$ are such that $p$ is an irreducible and $p|rs$, then either $p|r$ or $p|s$. More generally, if $t$ and $r$ are relatively prime and $t|rs$ then $t|s$.

**Proof.** To see the first statement, write $rs = pt$ and factor $r, s, t$ into irreducibles. Then $p$ must be an associate of some irreducible factor of either $r$ or $s$, hence $p$ divides either $r$ or $s$. The second statement can be proved along similar but slightly more complicated lines. $\square$

As a consequence, we have:

**Proposition:** Let $R$ be a UFD and let $r \in R$, where $r \neq 0$. Then $(r)$ is prime ideal $\iff$ $r$ is irreducible.

**Proof.** $\implies$ : If $(r)$ is a prime ideal, then $r$ is not a unit, and $r \neq 0$ by assumption. If $r = st$, then one of $s, t \in (r)$, say $s \in (r)$, hence $s = ru$. Then $r = rut$ so that $ut = 1$ and $t$ is a unit. Hence $r$ is irreducible. (Note: this part did not use the fact that $R$ was a UFD, and holds in every integral domain.)

$\impliedby$ : If $r$ is irreducible, then it is not a unit and hence $(r) \neq R$. Suppose that $st \in (r)$. Then $r|st$. By the remark above, either $r|s$ or $r|t$, i.e. either $s \in (r)$ or $t \in (r)$. Hence $(r)$ is prime. $\square$

Note: in case $R$ is not a UFD, there will in general exist irreducibles $r$ such that $(r)$ is not a prime ideal.

**Theorem:** Let $R$ be a PID, and let $r, s \in R$, not both 0. Then a gcd $d$ of $r$ and $s$ exists. Moreover, $d$ is a linear combination of $r$ and $s$: there exist $a, b \in R$ such that $d = ar + bs$.

Note: for a general UFD, the gcd of two elements $r$ and $s$ will not in general be a linear combination of $r$ and $s$. For example, in $F[x, y]$, the elements $x$ and $y$ are relatively prime, hence their gcd is 1, but 1 is not a linear combination of $x$ and $y$, since if $f(x, y) = xp(x, y) + yq(x, y)$ is any linear combination of $x$ and $y$, then $f(0, 0) = 0$.

**Proof.** This argument is very similar to the corresponding argument for $F[x]$, or for $\mathbb{Z}$. Given $r, s \in R$, not both 0, consider the ideal

$$(r, s) = \{ar + bs : a, b \in R\} = (r) + (s).$$

Then $(r, s)$ is easily checked to be an ideal, hence there exists a $d \in R$ with $(r, s) = (d)$. By construction $d = ar + bs$ for some $r, s \in R$. Since $r = 1 \cdot r + 0 \cdot s \in (r, s) = (d)$, this says that $d|r$. Similarly $d|s$. Finally, if $e|r$ and $e|s$, then $e|(ar + bs) = d$. $\square$

**Corollary (of Theorem):** If $R$ is a PID, $r, s \in R$ are relatively prime and $r|st$, then $r|t$.

**Proof.** Write $1 = ar+bs$ for some $a, b \in R$. Then $t = tar+tbs = r(at)+b(st)$. By assumption $r|st$ and clearly $r|r(at)$. Hence $r|t$. $\square$

**Corollary:** If $R$ is a PID, and $r \in R$ is an irreducible, then for all $s, t \in R$, if $r|st$, then either $r|s$ or $r|t$.

**Proof.** Since $r$ is an irreducible, it is easy to see that a gcd of $r$ and $s$ is either a unit or an associate of $r$, i.e. if $r$ does not divide $s$, then $r$ and $s$ are relatively prime. Suppose then that $r$ does not divide $s$. Then by the previous corollary $r|t$. Hence either $r|s$ or $r|t$. $\square$

The following proves the uniqueness half of the assertion that a PID is a UFD:

**Corollary:** If $R$ is a PID, then uniqueness of factorization holds in $R$: if $p_i, 1 \leq i \leq n$ and $q_j, 1 \leq j \leq m$ are irreducibles such that $p_1 \cdots p_n = q_1 \cdots q_m$, then $n = m$ and, after reordering, $p_i$ and $q_j$ are associates.

**Proof.** This is proved in exactly the same way as the argument for $F[x]$ (or $\mathbb{Z}$). $\square$

**Theorem:** A PID is a UFD.

**Proof.** We have already seen that, if an irreducible factorization exists, it is unique. Thus the remaining point is to show that, if $R$ is a PID, then every element $r \in R$, not 0 or a unit, admits **some** factorization into a product of irreducibles. The proof will be in several steps.

**Lemma:** Let $R$ be an integral domain with the property that, if

$$(a_1) \subseteq (a_2) \subseteq \cdots \subseteq (a_n) \subseteq (a_{n+1}) \subseteq \cdots$$

is an increasing sequence of principal ideals, then the sequence is eventually constant, i.e. there exists an $N$ such that, for all $n \geq N$, $(a_n) = (a_{n+1}) = \cdots$. Then every nonzero $r \in R$ which is not a unit factors into a product of irreducibles.

We can paraphrase the hypothesis of the lemma by saying that $R$ *satisfies the ascending chain condition* (a.c.c) on principal ideals.

**Proof of the lemma.** Suppose by contradiction that $r \in R$ is an element, not zero or a unit, which does not factor into a product of irreducibles. In particular, $r$ itself is not irreducible, so that $r = r_1 s_1$ where neither $r_1$ nor $s_1$ is a unit. Thus $(r)$ is properly contained in $(r_1)$ and in $(s_1)$. Clearly, we can assume that at least one of $r_1$, $s_1$, say $r_1$, does not factor into irreducibles (if both so factor, so does the product). By applying the above to $r_1$, we see that $(r_1)$ is strictly contained in a principal ideal $(r_2)$, where $r_2$ does not factor into a product of irreducibles. Continuing in this way, we can produce a strictly increasing infinite chain of principal ideals $(r_1) \subset (r_2) \subset \cdots$, i.e. each $(r_{i+1})$ properly contains the previous ideal $(r_i)$, contradicting the hypothesis on $R$. $\square$

To complete the proof of the theorem that a PID is a UFD, it suffices to show that a PID $R$ satisfies the hypotheses of the above lemma. First suppose that $(r_1) \subseteq (r_2) \subseteq \cdots$ is an increasing sequence of ideals of $R$. It is easy to check that $I = \bigcup_i (r_i)$ is again an ideal. More generally, we have the following:

**Claim:** Let $R$ be a ring and let $I_1 \subseteq I_2 \subseteq \cdots$ be an increasing sequence of ideals of $R$. If $I = \bigcup_n I_n$, then $I$ is an ideal of $R$.

**Proof.** To see that $I$ is an additive subgroup, we show for example that it is closed under addition. Given $a, b \in I$, there exists a $j$ such that $a \in I_j$ and there exists a $k$ such that $b \in I_k$. Setting $\ell = \max\{j, k\}$, we have $a \in I_j \subseteq I_\ell$ and $b \in I_k \subseteq I_\ell$. Hence $a, b \in I_\ell$, and since $I_\ell$ is an ideal, $a + b \in I_\ell \subseteq I$. Thus $I$ is closed under addition. Similarly, if $a \in I$, then $-a \in I$ and $ta \in I$ for all $t \in R$. Thus $I$ is an ideal. $\square$

Returning to the proof of the theorem, given the increasing sequence of ideals $(r_1) \subseteq (r_2) \subseteq \cdots$, the claim implies that $I = \bigcup_i (r_i)$ is again an ideal of $R$. Since $R$ is a PID, $I = (r)$ for some $r \in R$. Necessarily $r \in (r_N)$ for some $N$. But then $(r) \subseteq (r_N) \subseteq (r_{N+1}) \cdots \subseteq \bigcup_i (r_i) = (r)$. Thus all inclusions are equalities, and $(r_n) = (r_N)$ for all $n \geq N$, i.e. the sequence is eventually constant. Hence $R$ satisfies the hypotheses of the previous lemma, so that every $r \in R$, not 0 or a unit, factors into a product of irreducibles. $\square$

The ascending chain condition and the arguments we have just given are so fundamental that we generalize them as follows:

**Proposition:** For a ring $R$, the following two conditions are equivalent:

(i) Every ideal $I$ of $R$ is *finitely generated*: if $I$ is an ideal of $R$, then $I = (r_1, \ldots, r_n)$ for some $r_i \in R$.

(ii) Every increasing sequence of ideals is eventually constant, in other words if
$$I_1 \subseteq I_2 \subseteq \cdots \subseteq I_n \subseteq I_{n+1} \subseteq \cdots,$$
where the $I_n$ are ideals of $R$, then there exists an $N \in \mathbb{N}$ such that for all $k \geq N$, $I_k = I_N$.

If the ring $R$ satisfies either of the equivalent conditions above, then $R$ is called a *Noetherian* ring.

**Proof.** (i) $\implies$ (ii): given an increasing sequence of ideals $I_1 \subseteq I_2 \subseteq \cdots$, let $I = \bigcup_n I_n$. Then by the claim above, $I$ is an ideal, and hence $I = (r_1, \ldots, r_n)$ for some $r_i \in R$. Thus $r_i \in I_{n_i}$ for some $n_i$. If $N = \max_i n_i$, then $r_i \in I_N$ for every $i$. Hence, for all $k \geq N$, $I = (r_1, \ldots, r_n) \subseteq I_N \subseteq I_k \subseteq I$. It follows that $I_k = I_N = I$ for all $k \geq N$.

(ii) $\implies$ (i): Let $I$ be an ideal of $R$ and choose an arbitrary $r_1 \in I$ (for example, $r_1$ could be 0). Set $I_1 = (r_1)$. If $I = I_1$, stop. Otherwise there exists an $r_2 \in I - I_1$. Set $I_2 = (r_1, r_2)$, and note that $I_2$ strictly contains $I_1$. If $I = I_2$, stop, otherwise there exists an $r_3 \in I - I_2$. Inductively suppose that we have found $I_k = (r_1, \dots, r_k)$ with $I_k \subseteq I$. If $I = I_k$ we are done, otherwise there exists $r_{k+1} \in I - I_k$ and we set $I_{k+1} = (r_1, \dots, r_{k+1})$. So if $I$ is not finitely generated, we have constructed a strictly increasing sequence $I_1 \subset I_2 \subset \cdots$, contradicting the assumption on $R$. Thus $I$ is finitely generated. □

Clearly, the arguments we have already discussed imply the following:

**Theorem:** Suppose that $R$ is a Noetherian integral domain. Then every element $r \in R$, not 0 or a unit, factors into a product of irreducibles. Moreover, the following are equivalent:

(i) $R$ is a UFD.

(ii) For every nonzero $r \in R$, the element $r$ is irreducible if and only if $(r)$ is a prime ideal. □

# 2 Euclidean domains

We turn now to finding new examples of PID's.

**Definition:** Let $R$ be an integral domain. A *Euclidean norm* on $R$ is a function $N\colon R - \{0\} \to \mathbb{Z}$ satisfying:

1. For all $r \in R - \{0\}$, $N(r) \geq 0$.
2. For all $a, b \in R$ with $a \neq 0$, there exist $q, r \in R$ with $b = aq + r$ and either $r = 0$ or $N(r) < N(a)$.

An integral domain $R$ such that there exists a Euclidean norm on $R$ is called a *Euclidean domain*.

**Definition:** The Euclidean norm $N$ is *submultiplicative* if in addition $N$ satisfies: For all $a, b \in R - \{0\}$, $N(a) \leq N(ab)$. It is *multiplicative* if $N$ satisfies: For all $a, b \in R - \{0\}$, $N(ab) = N(a)N(b)$. If $N$ is multiplicative and $N(a) > 0$ for all $a \in R - \{0\}$, then $N$ is submultiplicative. (In fact, the condition that $N(a) > 0$ for all $a \in R - \{0\}$ is automatically satisfied.)

**Examples:** $R = \mathbb{Z}$, $N(a) = |a|$; $R = F[x]$, $F$ a field, and $N(f(x)) = \deg f(x)$, defined for $f(x) \neq 0$. Here (1) is clear and (2) is the statement of long division in $\mathbb{Z}$ or in $F[x]$. In fact, it is easy to see that $N$ is submultiplicative in both cases.

**Remark:** In the definition of a Euclidean norm, we do **not** require that the $q, r \in R$ are unique. In fact, this even fails in $\mathbb{Z}$ if we allow $q$ and $r$ to be negative. For example, with $a = 3$, $b = 11$, we can write $11 = 3 \cdot 3 + 2 = 3 \cdot 4 + (-1)$.

**Proposition:** If $R$ is a Euclidean domain, then $R$ is a PID.

**Proof.** This argument should be very familiar. Let $I$ be an ideal of $R$. If $I = \{0\}$, then $I = (0)$ is principal. Otherwise, consider the nonempty set $A$ of nonnegative integers $\{N(r) : r \in I - \{0\}\}$. By the well-ordering principle, there exists an $a \in I - \{0\}$ such that $N(a)$ is a smallest element of $A$. We claim that $I = (a)$. Clearly $(a) \subseteq I$ since $a \in I$. Conversely, if $b \in I$, then there exist $q, r \in R$ such that $b = aq + r$ with either $r = 0$ or $N(r) < N(a)$. As $b, aq \in I$, $r = b - aq \in I$. Hence $N(r) < N(a)$ is impossible by the choice of $a$, so that $r = 0$ and $b = aq \in (a)$. Thus $(a) \subseteq I$ and hence $(a) = I$. $\square$

**Lemma:** Let $R$ be an integral domain and let $N$ be a submultiplicative Euclidean norm on $R$. For all $b \in R - \{0\}$, exactly one of the following holds:

1. $b$ is not a unit, $N(b) > N(1)$, and $N(a) < N(ab)$ for all $a \in R - \{0\}$.
2. $b$ is a unit, $N(b) = N(1)$, and $N(a) = N(ab)$ for all $a \in R - \{0\}$.

**Proof.** Since we always have $N(a) \leq N(ab)$, it suffices to show that $N(a) = N(ab) \iff b$ is a unit. First, if $b$ is a unit, then $N(a) \leq N(ab)$ and $N(ab) \leq N(abb^{-1}) = N(a)$, so that $N(a) = N(ab)$. It is then an easy exercise to see that $N(b) = N(1)$. Conversely, suppose that $N(a) = N(ab)$. Applying long division of $ab$ into $a$, we se that $a = (ab)q + r$, with either $r = 0$ or $N(r) < N(ab) = N(a)$. We claim that $r$ must be 0, since otherwise $r = a - abq = a(1 - bq)$ with $1 - bq \neq 0$, and hence

$$N(a) \leq N(a(1 - bq)) = N(r) < N(a),$$

a contradiction. Thus $r = 0$, so that $a = abq$ and thus $bq = 1$, i.e. $b$ is a unit. $\square$

**Corollary:** Let $R$ be an integral domain and let $N$ be a submultiplicative Euclidean norm on $R$. If $r \in R - \{0\}$ and $r = ab$ with neither $a$ nor $b$ a unit, then $N(a) < N(r)$ and $N(b) < N(r)$. $\square$

**Proposition:** If $R$ is a Euclidean domain with a submultiplicative Euclidean norm and $r \in R$ is not 0 or a unit, then $r$ is a product of irreducibles.

**Proof.** Given $r$, not 0 or a unit, if $r$ is irreducible we are done. Otherwise, $r = r_1 r_2$, with neither $r_1$ nor $r_2$ a unit. Hence $N(r_i) < N(r)$, $i = 1, 2$. If $r_i$ is irreducible for $i = 1, 2$, we are done. Otherwise at least one of $r_1$, $r_2$ factors into factors: say $r_1 = ab$, with $N(a) < N(r_1) < N(r)$ and $N(b) < N(r_1) < N(r)$. Clearly this process cannot continue indefinitely.

A more formal way to give this argument is as follows: if there exists an $r \in R$, not 0 or a unit, which is **not** a product of irreducibles, then there exists an $r$ such that $N(r)$ is minimal among all such, i.e. if $s \in R$ is not 0, a unit, or a product of irreducibles, then $N(r) \leq N(s)$, by the well-ordering principle. But such an $r$ cannot be irreducible (since a single irreducible is by convention a product of one irreducible). So $r = r_1 r_2$, with neither $r_1$ nor $r_2$ a unit, and so $N(r_i) < N(r)$, $i = 1, 2$. But at least one of $r_1$ and $r_2$ is not a product of irreducibles, since if both $r_1$ and $r_2$ were a product of irreducibles, then $r_1 r_2 = r$ would also be a product of irreducibles. Say $r_1$ is not a product of irreducibles. Then by the choice of $r$, $N(r) \leq N(r_1)$. This contradicts $N(r_1) < N(r)$. Hence no such $r$ can exist. $\square$

**Corollary:** If $R$ is a Euclidean domain with a submultiplicative Euclidean norm, then $R$ is a UFD. $\square$

Of course, the corollary follows from the more general fact that a PID is a UFD. But we were able to give a more direct proof using the proposition above.

**The Euclidean algorithm in a Euclidean domain:** Let $R$ be a Euclidean domain with Euclidean norm $N$. Begin with $a, b \in R$, with $b \neq 0$. Write $a = bq_1 + r_1$, with $q_1, r_1 \in R$, and either $r_1 = 0$ or $N(r_1) < N(b)$. Note that $r_1 = a + b(-q_1)$ is a linear combination of $a$ and $b$. If $r_1 = 0$, stop, otherwise repeat this process with $b$ and $r_1$ instead of $a$ and $b$, so that $b = r_1 q_2 + r_2$, with $r_2 = 0$ or $N(r_2) < N(b)$ If $r_2 = 0$, stop, otherwise repeat again. to find $r_1, \dots, r_k$ with $N(r_1) > N(r_2) > N(r_3) > \cdots > N(r_k) \geq 0$, with $r_{k-1} = r_k q_{k+1} + r_{k+1}$. Since the integers $N(r_i)$ decrease, and they are all nonnegative, eventually this procedure must stop with an $r_n$ such that $r_{n+1} = 0$, and hence $r_{n-1} = r_n q_{n+1}$. The procedure looks as follows:

$$\begin{aligned} a &= bq_1 + r_1 \\ b &= r_1 q_2 + r_2 \\ r_1 &= r_2 q_3 + r_3 \\ &\vdots \\ r_{n-2} &= r_{n-1} q_n + r_n \\ r_{n-1} &= r_n q_{n+1}. \end{aligned}$$

Then $r_n$ is a gcd of $a, b$ and tracing back through the steps shows how to write it as a linear combination of $a$ and $b$.

# 3 Factorization in the Gaussian integers

We now consider factorization in the *Gaussian integers*

$$\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}.$$

Consider the function $N\colon \mathbb{Z}[i] \to \mathbb{Z}$ defined by $N(\alpha) = \alpha\bar{\alpha}$, where if $\alpha = a + bi$, then $\bar{\alpha} = a - bi$ (i.e. $N(a + bi) = a^2 + b^2$). Note that, given $n \in \mathbb{Z}$, $n = N(\alpha)$ for some $\alpha \in \mathbb{Z}[i] \iff n$ is a sum of two integer squares.

**Lemma:** The function $N$ satisfies:

(a) $N(\alpha) \geq 0$ for all $\alpha \in \mathbb{Z}[i]$.

(b) For all $\alpha, \beta \in \mathbb{Z}[i]$, $N(\alpha\beta) = N(\alpha)N(\beta)$ ($N$ is multiplicative). Hence, if $n_1$ and $n_2$ are two integers which are each a sum of two integer squares, then $n_1 n_2$ is a sum of two integer squares.

(c) There is a natural extension of $N$ to a function $\mathbb{Q}(i) \to \mathbb{Q}$, satisfying (a) and (b) (and which we continue to denote by $N$).

(d) $N(\alpha) = 1 \iff \alpha$ is a unit.

**Proof.** (a) Clear. (b) $N(\alpha\beta) = (\alpha\beta)(\overline{\alpha\beta}) = (\alpha\beta)(\bar{\alpha}\bar{\beta}) = \alpha\bar{\alpha}\beta\bar{\beta} = N(\alpha)N(\beta)$. (c) Clear. (d) We can see this directly ($N(\alpha) = 1 \iff \alpha = \pm 1$ or

$\alpha = \pm i$) or as follows: if $N(\alpha) = 1$, then $\alpha\bar{\alpha} = 1$ and hence $\alpha$ is a unit with $\alpha^{-1} = \bar{\alpha}$. Conversely, if $\alpha$ is a unit, then $\alpha\beta = 1$ for some $\beta \in \mathbb{Z}[i]$, hence $N(\alpha\beta) = 1 = N(\alpha)N(\beta)$. Thus $N(\alpha)$ is a positive integer dividing 1, so $N(\alpha) = 1$. $\square$

**Proposition:** In the integral domain $\mathbb{Z}[i]$, the function $N(\alpha) = \alpha\bar{\alpha}$ is a (submultiplicative) Euclidean norm.

**Proof.** Given $\alpha, \beta \in \mathbb{Z}[i]$ with $\alpha \neq 0$, we must show that we can find $\xi, \rho \in \mathbb{Z}[i]$ with $\beta = \alpha\xi + \rho$ and $\rho = 0$ or $N(\rho) < N(\alpha)$. Consider the quotient $\beta/\alpha \in \mathbb{Q}[i]$. Write $\beta/\alpha = r + si$ with $r, s \in \mathbb{Q}$. Then there exist integers $n, m \in \mathbb{Z}$ with $|r - n| \leq \frac{1}{2}$ and $|s - m| \leq \frac{1}{2}$. Set $\xi = n + mi$ and $\gamma = \beta/\alpha - \xi$. Then $\beta = \alpha\xi + \alpha\gamma = \alpha\xi + \rho$, say, where $\rho = \alpha\gamma$. Since $\rho = \beta - \alpha\xi$, $\rho \in \mathbb{Z}[i]$. Moreover,

$$N(\gamma) = N(\beta/\alpha - \xi) = (r-n)^2 + (s-m)^2 \leq \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} < 1.$$

Then $\beta = \alpha\xi + \rho$ with either $\rho = 0$ or

$$N(\rho) = N(\alpha\gamma) = N(\alpha)N(\gamma) < N(\alpha).$$

Hence $N$ is a Euclidean norm and it is submultiplicative since it is multiplicative and $N(\alpha) \geq 1$ for all $\alpha \neq 0$. $\square$

**Corollary:** $\mathbb{Z}[i]$ is a PID and a UFD. $\square$

**Lemma:**

(i) If $N(\alpha) = p$, where $p$ is a prime number, then $\alpha$ is irreducible.

(ii) If $p$ is a prime number, then $p$ is not irreducible in $\mathbb{Z}[i] \iff p = N(\alpha)$ for some $\alpha \in \mathbb{Z}[i] \iff p$ is a sum of two integer squares. In this case, if $\alpha$ divides $p$ and $\alpha$ is not a unit or an associate of $p$, then $p = N(\alpha)$.

**Proof.** (i) If $\alpha = \beta\gamma$, then $p = N(\alpha) = N(\beta\gamma) = N(\beta)N(\gamma)$, and so one of $N(\beta)$, $N(\gamma)$ is 1. Hence either $\beta$ or $\gamma$ is a unit, so that $\alpha$ is irreducible.

(ii) If $p$ is not irreducible, then $p = \alpha\beta$ where neither $\alpha$ nor $\beta$ is a unit, hence $N(\alpha)$ and $N(\beta)$ are both greater than 1. Then $p^2 = N(p) =$

$N(\alpha)N(\beta)$, so that $N(\alpha) = N(\beta) = p$. Conversely, if $p = N(\alpha)$, then $p = \alpha\bar{\alpha}$ with $N(\alpha) = N(\bar{\alpha}) = p$, so that neither $\alpha$ nor $\bar{\alpha}$ is a unit. Hence $p$ is not irreducible in $\mathbb{Z}[i]$. $\square$

**Lemma:** If $\pi$ is an irreducible element of $\mathbb{Z}[i]$, then there exists a prime number $p$ such that $\pi$ divides $p$ in $\mathbb{Z}[i]$. If the prime number $p$ is also irreducible in $\mathbb{Z}[i]$, then $\pi$ and $p$ are associates, so that $\pi = \pm p$ or $\pm ip$. If the prime number $p$ is not irreducible in $\mathbb{Z}[i]$, then $p = N(\pi)$ and every irreducible factor of $p$ is other an associate of $\pi$ or an associate of $\bar{\pi}$.

**Proof.** Consider $N(\pi) \in \mathbb{Z}$. Since $\pi$ is not a unit, $N(\pi) > 1$, and hence $N(\pi)$ is a product of prime numbers $p_1 \cdots p_r$ (not necessarily distinct). Since $\mathbb{Z}[i]$ is a UFD and $\pi$ is an irreducible dividing the product $p_1 \cdots p_r$, there must exist an $i$ such that $\pi$ divides $p_i$, and we take $p = p_i$. If $p$ is also irreducible, then $\pi$ and $p$ are associates, and hence $\pi = \pm p$ or $\pm ip$. If $p$ is not irreducible, then we have seen that $p = \alpha\bar{\alpha}$ for every $\alpha \in \mathbb{Z}[i]$ which is a nontrivial factor of $p$, hence $\pi$ divides $p = \alpha\bar{\alpha}$. Moreover both $\alpha$ and $\bar{\alpha}$ are irreducible since both have norm $p$. It follows that $\pi$ divides either $\alpha$ or $\bar{\alpha}$, say $\pi$ divides $\alpha$, and hence that $\pi$ is an associate of $\alpha$ since $\alpha$ is irreducible. Since units have norm 1, it follows that $N(\pi) = N(\alpha) = p$. $\square$

Note that 2 is not irreducible in $\mathbb{Z}[i]$, and in fact $2 = N(1+i)$. The irreducible factors of 2 are $\pm 1 \pm i$, and they are all associates: up to a unit, 2 is a square since $2 = (-i)(1+i)^2$. For other primes $p$ of the form $N(\alpha) = \alpha\bar{\alpha}$, this does not happen: if $\alpha = a + bi$, the associates of $\alpha$ are $\pm(a+bi)$ and $\pm i(a+bi) = \pm(-b+ai)$. Hence $\bar{\alpha} = a - bi$ is an associate of $\alpha \iff a = b$. If moreover $\alpha$ is irreducible, then since $a|(a+ai)$, $a = \pm 1$ and $p = 2$.

We may now describe the irreducibles in $\mathbb{Z}[i]$ as follows:

**Theorem:** The irreducible elements in $\mathbb{Z}[i]$ are:

1. $1 + i$ and its associates $\pm 1 \pm i$;
2. Ordinary prime numbers $p \in \mathbb{Z} \subseteq \mathbb{Z}[i]$ congruent to 3 mod 4 and their associates $\pm p, \pm ip$;
3. Gaussian integers $\alpha = a + bi$ such that $N(\alpha) = a^2 + b^2 = p$, where $p$ is a prime number congruent to 1 mod 4. Moreover, for every prime number $p$ congruent to 1 mod 4, there exists an $\alpha = a + bi$ such that $N(\alpha) = a^2 + b^2 = p$.

**Proof.** Let $\pi$ be an irreducible in $\mathbb{Z}[i]$. We have seen that either $\pi$ is an associate of a prime $p$ which is irreducible in $\mathbb{Z}[i]$, or $N(\pi) = p$ is a prime number and that the irreducible factors of $p$ are exactly the associates of $\pi$ or $\bar{\pi}$. Moreover, 2 is not irreducible and the only irreducibles dividing 2 are $1+i$ and its associates. If $p$ is an odd prime, $p$ is not irreducible in $\mathbb{Z}[i] \iff p = a^2 + b^2$, where $a, b \in \mathbb{Z}$. Since $p$ is odd, $a$ and $b$ cannot be both odd or both even, so one of them, say $a$, is odd and the other, say $b$, is even. Then $a^2 \equiv 1 \bmod 4$ and $b^2 \equiv 0 \bmod 4$, so that $p = a^2 + b^2 \equiv 1 \bmod 4$. In other words, if $p$ is an odd prime which is not irreducible in $\mathbb{Z}[i]$, then $p \equiv 1 \bmod 4$. Hence, if $p$ is an odd prime with $p \equiv 3 \bmod 4$, then $p$ is irreducible in $\mathbb{Z}[i]$ and its irreducible factors are its associates $\pm p, \pm ip$.

Thus we will be done if we show that every odd prime number congruent to 1 mod 4 is not irreducible in $\mathbb{Z}[i]$, for then the remaining irreducibles of $\mathbb{Z}[i]$ will be the nontrivial factors of $p$ for such primes $p$, which are necessarily irreducible and of norm $p$. To see this statement, we use the following:

**Lemma:** If $p \equiv 1 \bmod 4$, then there exists a $k \in \mathbb{Z}$ such that $k^2 \equiv -1 \bmod p$.

**Proof.** The assumption $p \equiv 1 \bmod 4$ is exactly the statement that $4 | p - 1$. Now we know that $(\mathbb{Z}/p\mathbb{Z})^*$ is a cyclic group of order $p-1$. By known results on cyclic groups, there exists an element $k$ of $(\mathbb{Z}/p\mathbb{Z})^*$ of order 4. In other words, $k^4 = 1$ in $(\mathbb{Z}/p\mathbb{Z})^*$ but $k^2 \neq 1$ in $(\mathbb{Z}/p\mathbb{Z})^*$. Since $k^2$ is then a root of the polynomial $x^2 - 1 = (x + 1)(x - 1)$ in the field $\mathbb{Z}/p\mathbb{Z}$, we must have $k^2 = \pm 1$, and since by assumption $k^2 \neq 1$, $k^2 = -1$. This says that there is an integer $k$ such that $k^2 \equiv -1 \bmod p$. $\square$

To complete the proof of the theorem, if $p \equiv 1 \bmod 4$, then we shall show that $p$ is not irreducible in $\mathbb{Z}[i]$. Let $k \in \mathbb{Z}$ be such that $k^2 \equiv -1 \bmod p$, so that $p$ divides $k^2 + 1$. In $\mathbb{Z}[i]$, we can factor $k^2 + 1 = (k + i)(k - i)$. If $p$ were an irreducible, then since $p$ divides $k^2 + 1 = (k + i)(k - i)$, $p$ would divide one of the factors $k \pm i$. But

$$\frac{k \pm i}{p} = \frac{k}{p} \pm \frac{1}{p} i.$$

Since $\pm 1/p$ is not an integer, the quotient $(k \pm i)/p$ does not lie in $\mathbb{Z}[i]$. Hence $p$ does not divide either factor $k \pm i$ of $k^2 + 1$, and so cannot be an irreducible. $\square$

**Corollary:** Let $n \in \mathbb{N}$, $n > 1$, and write $n = p_1^{a_1} \cdots p_r^{a_r}$, where the $p_i$ are distinct prime numbers and $a_i \in \mathbb{N}$. Then $n$ is a sum of two integer squares if and only, for every prime factor $p_i$ of $n$ such that $p_i \equiv 3 \bmod 4$, $a_i$ is even.

**Proof.** $\Longleftarrow$ : If $n$ is as described, then every prime factor $p_i$ of $n$ which is either 2 or $\equiv 1 \bmod p$ is a sum of two squares, hence so is $p_i^{a_i}$ for an arbitrary positive power $a_i$. If $p_i \equiv 3 \bmod 4$, then, if $a_i$ is even, $p_i^{a_i}$ is also a square since it is an even power. Thus $n = p_1^{a_1} \cdots p_r^{a_r}$ is a sum of two squares since it is a product of factors, each of which is a sum of two squares.

$\Longrightarrow$ : Suppose that $n$ is a sum of two squares. Then $n = N(\alpha)$ for some $\alpha \in \mathbb{Z}[i]$, not 0 or a unit. Factor $\alpha$ into a product of irreducibles: $\alpha = u\pi_1^{b_1} \cdots \pi_s^{b_s}$, where $u$ is a unit, the $b_i$ are positive integers, and $\pi_i$ is not an associate . If $\pi_i$ is not an associate of a prime $p_i \equiv 3 \bmod 4$, then $N(\pi_i)$ is either 2 or a prime $\equiv 1 \bmod 4$. If $\pi_i$ is an associate of a prime $p_i \equiv 3 \bmod 4$, then $N(\pi_i) = p_i^2$ and thus $N(\pi_i^{b_i}) = p_i^{2b_i}$. Hence

$$n = N(\alpha) = (N(\pi_1))^{b_1} \cdots (N(\pi_s))^{b_s}$$

is a product of prime powers with the property that all of the primes $\equiv 3 \bmod 4$ occur to even powers. It follows that the prime factorization of $n$ is as claimed. $\square$

# 4 Examples where unique factorization fails

One can try to extend the above arguments to more general classes of rings. One very kind of ring to consider is $\mathbb{Z}[\sqrt{-d}]$, where $d \in \mathbb{N}$. We usually assume that $d$ has no squared prime factors, in other words that either $d = 1$ or $d = p_1 \cdots p_k$ is a product of distinct primes, since $\sqrt{-a^2e} = a\sqrt{-e}$. Note that $\mathbb{Z}[\sqrt{-d}]$ is a subring of the field $\mathbb{Q}(\sqrt{-d})$, which is called an *imaginary quadratic field*. Similarly, we could look at $\mathbb{Z}[\sqrt{d}]$, where $d \in \mathbb{N}$ and $d$ has no squared prime factors. In this case $\mathbb{Z}[\sqrt{d}]$ is a subring of the field $\mathbb{Q}(\sqrt{d})$, which is called a *real quadratic field*.

There is a natural multiplicative function $N \colon \mathbb{Z}[\sqrt{-d}] \to \mathbb{Z}$ defined by, if $\alpha = a + b\sqrt{-d} \in \mathbb{Z}[\sqrt{-d}]$,

$$N(\alpha) = \alpha\bar{\alpha} = a^2 + db^2.$$

Just as in the case $d = 1$, $N$ is multiplicative, i.e. $N(\alpha\beta) = N(\alpha)N(\beta)$, and $N$ extends to a function from $\mathbb{Q}(\sqrt{-d})$ to $\mathbb{Q}$ which is a homomorphism of multiplicative groups from $\mathbb{Q}(\sqrt{-d})^*$ to $\mathbb{Q}^*$. Adapting the arguments in the preceding section for $\mathbb{Z}[i]$, it is not hard to show:

**Proposition:** In the integral domain $\mathbb{Z}[\sqrt{-2}]$, the function $N(\alpha) = \alpha\bar{\alpha}$ is a (submultiplicative) Euclidean norm.

However, this fails for every $d > 2$.

**Example:** The integral domain $\mathbb{Z}[\sqrt{-3}]$ is not a UFD. In fact, in $\mathbb{Z}[\sqrt{-3}]$,

$$4 = 2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3}).$$

We will show that 2 and $1 \pm \sqrt{-3}$ are all irreducible, and that 2 is not an associate of $1 \pm \sqrt{-3}$. First, arguing as for $\mathbb{Z}[i]$, it is easy to check that $\alpha \in \mathbb{Z}[\sqrt{-3}]$ is a unit $\iff N(\alpha) = 1$. Now suppose that 2 factors in $\mathbb{Z}[\sqrt{-3}]$: say $2 = \alpha\beta$. Then $N(\alpha)N(\beta) = N(2) = 4$. If neither $\alpha$ nor $\beta$ is a unit, then $N(\alpha) > 1$ and $N(\beta) > 1$, hence $N(\alpha) = N(\beta) = 2$. But if say $\alpha = a + b\sqrt{-3}$ with $a, b \in \mathbb{Z}$, then $a^2 + 3b^2 = 2$, hence $b = 0$ and $a^2 = 2$, which is impossible. Thus 2 is irreducible, and since $N(1 \pm \sqrt{-3}) = 4$ as well, a similar argument shows that $1 \pm \sqrt{-3}$ is irreducible. Finally, 2 and $1 + \sqrt{-3}$ are not associates, since if they were, then 2 would divide $1 + \sqrt{-3}$ in $\mathbb{Z}[\sqrt{-3}]$. But $(1 + \sqrt{-3})/2 = 1/2 + (1/2)\sqrt{-3} \notin \mathbb{Z}[\sqrt{-3}]$. Likewise, 2 and $1 - \sqrt{-3}$ are not associates in $\mathbb{Z}[\sqrt{-3}]$. Hence $\mathbb{Z}[\sqrt{-3}]$ is not a UFD.

This example is slightly misleading, because $\mathbb{Z}[\sqrt{-3}]$ is a subring of a somewhat more natural ring which is in fact a UFD: Let $\omega = e^{2\pi i/3} = -\frac{1}{2} + \frac{1}{2}\sqrt{-3}$ be a cube root of unity. Note that $\omega$ is a root of the monic polynomial $x^2 + x + 1$, since $\omega$ is a root of $x^3 - 1$ and $x^3 - 1 = (x - 1)(x^2 + x + 1)$. Note that, since $\omega^3 = 1$, $\omega^2 = \omega^{-1} = \bar{\omega}$. Hence $\sqrt{-3} = \omega - \omega^2 \in \mathbb{Z}[\omega]$, so that $\mathbb{Z}[\sqrt{-3}]$ is a subring of $\mathbb{Z}[\omega]$. More generally, we say that an $\alpha \in \mathbb{C}$ is an *algebraic integer* if $\alpha$ is a root of a monic polynomial with integer coefficients, i.e. $f(\alpha) = 0$, where $f(x) \in \mathbb{Z}[x]$ is monic. (It is easy to see that every algebraic **number** is a root of a polynomial $f(x) \in \mathbb{Z}[x]$, but $f(x)$ is not usually monic.) Them if $E \leq \mathbb{C}$ is an algebraic extension of $\mathbb{Q}$, one can show that the set of algebraic integers in $E$ is a subring of $E$ whose quotient field is $E$, and this ring plays the role of the subring $\mathbb{Z}$ of $\mathbb{Q}$. For $E = \mathbb{Q}(i)$, for example, the subring of algebraic integers is just $\mathbb{Z}[i]$, but for $E = \mathbb{Q}(\sqrt{-3})$, the subring of algebraic integers is $\mathbb{Z}[\omega]$. In this particular example, $\mathbb{Z}[\omega]$ is in fact a PID and hence a UFD.

However, this situation does not persist for long. For example, $\mathbb{Z}[\sqrt{-5}]$ turns out to be the full subring of algebraic integers in $\mathbb{Q}(\sqrt{-5})$, but it is easy to check that

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

gives a factorization of 6 into a product of irreducibles in two essentially different ways. Hence $\mathbb{Z}[\sqrt{-5}]$ is not a UFD, and hence it is not a PID.

More generally, a famous theorem due to Heegner-Stark says that there is a finite (and relatively short) list of imaginary quadratic fields whose rings of integers are UFD's.

Much of the above discussion carries over to real quadratic fields. For example, for $\mathbb{Z}[\sqrt{2}]$, we have a multiplicative function $N\colon \mathbb{Z}[\sqrt{2}] \to \mathbb{Z}$ defined by

$$N(a + b\sqrt{2}) = |a^2 - 2b^2|.$$

One can check that, at least in this case, $N$ is a Euclidean norm. For general real quadratic fields, one can define an analogous multiplicative function $N$, which will usually not however be a Euclidean norm. It is unknown if there are finitely or infinitely many real quadratic fields whose rings of integers are UFD's.