On Factorization in Domains

R. C. Daileda

October 11, 2017

Recall that the Fundamental Theorem of Arithmetic (FTA) guarantees that every $n \in \mathbb{N}$, $n \geq 2$ has a unique factorization (up to the order of the factors) into prime numbers. If we allow our factorizations to have signs, this statement extends to all of $\mathbb{Z} \setminus (\mathbb{Z}^{\times} \cup \{0\})$. A natural question to ask is how this result might generalize to rings other than \mathbb{Z} . This immediately leads us to two subquestions:

Q1. How do we "correctly" generalize the notion of prime number?

Q2. Where do prime factorizations "come from?" What makes them unique?

To answer **Q1** we turn to the proof of the FTA wherein we find that there are two properties required of a prime number *p*:

i. p = ab implies $a = \pm 1$ or $b = \pm 1$;

ii. p|ab implies p|a or p|b.

In fact, these two properties are equivalent by Bézout's Lemma. But because we don't expect Bézout's Lemma to be true in an arbitrary ring, there's no reason to expect these properties to be equivalent in general. We therefore begin by defining two separate terms.

Definition 1. Let D be a domain, $a \in D \setminus (D^{\times} \cup \{0\})$. Then we say a is *irreducible* if a = bc with $b, c \in D$ implies $b \in D^{\times}$ or $c \in D^{\times}$. We say a is *prime* if a|bc with $b, c \in D$ implies a|b or a|c.

Lemma 1. Let D be a domain, $a \in D \setminus (D^{\times} \cup \{0\})$. Then:

- a is prime if and only if (a) is a prime ideal;
- a is irreducible if and only if (a) is maximal among principal ideals.

Proof. The proof of is a straightforward application of the definitions and is left as an exercise.

Lemma 2 (Prime Implies Irreducible). Let D be a domain and $a \in D \setminus (D^{\times} \cup \{0\})$. Then if a is prime, it is irreducible.

Proof. Suppose $a \in D \setminus (D^{\times} \cup \{0\})$ is prime and a = bc for some $b, c \in D$. Clearly a|bc so, without loss of generality, a|b. Write b = ad for some $d \in D$. Then

$$a = bc = adc \Rightarrow 1 = dc \Rightarrow c \in D^{\times}.$$

Hence a is irreducible.

We have just seen that prime implies irreducible, but, unlike the case with \mathbb{Z} , the converse of this statement is false in general.

Example 1. We will show that $1 + \sqrt{-3} \in \mathbb{Z}[\sqrt{-3}]$ is irreducible but not prime. Our main tool will be the *norm*,

$$N(a+b\sqrt{-3}) = (a+b\sqrt{-3})(a-b\sqrt{-3}) = a^2 + 3b^2 \in \mathbb{N}_0,$$

which is readily seen to be multiplicative. Suppose that $1 + \sqrt{-3} = xy$ with $x, y \in \mathbb{Z}[\sqrt{-3}]$. Then

$$4 = N(1 + \sqrt{-3}) = N(xy) = N(x)N(y)$$

Since N(x)|4 but N(x) = 2 is impossible, we must have N(x) = 1 or N(y) = 1. But the definition of the norm shows that this means x or y is a unit in $\mathbb{Z}[\sqrt{-3}]$. Hence $1 + \sqrt{-3}$ is irreducible.

But it is not prime. Again we appeal to the norm to see that

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

which shows that $(1 + \sqrt{-3})|2 \cdot 2$. However, $(1 + \sqrt{-3}) \nmid 2$, since if it did we'd have

$$2 = (a + b\sqrt{-3})(1 + \sqrt{-3}) = (a - 3b) + (a + b)\sqrt{-3} \implies a - 3b = 2, a + b = 0,$$

and the system on the right-hand side does not have an integral solution. Therefore $1 + \sqrt{-3}$ is *not* prime. \blacklozenge

So when does the converse statement, that irreducible implies prime, hold? It depends on the ring. Here are two examples.

Theorem 1. Let D be a PID and $a \in D \setminus (D^{\times} \cup \{0\})$. Then a is irreducible if and only if (a) is maximal if and only if a is prime.

Proof. According to Lemma 1, a is irreducible if and only if (a) is maximal among principal ideals. But in D every ideal is principal, so the second statement simply means that (a) is maximal. But maximal ideals are prime, so that by Lemma 1 again, maximality of (a) implies that a is prime. This proves that a is prime when it is irreducible. We already know that the converse holds in any domain, so we're finished.

Theorem 2. Let D be a Bézout domain, i.e. a domain in which the sum of principal ideals is principal. Then for $a \in D \setminus (D^{\times} \cup \{0\})$, a is prime if and only if a is irreducible.

Proof. Exercise. Mimic the proof of this fact in \mathbb{Z} (which is a Bézout domain!).

Corollary 1 (Existence of Roots of Polynomials). Let F be a field and $f \in F[x]$. If f is irreducible, then F[x]/(f) is a field containing (an isomorphic image of) F and a root of f.

Proof. F[x] is a PID so the first statement follows from Theorem 1. We the composition of homomorphisms

$$F \hookrightarrow F[x] \twoheadrightarrow F[x]/(f)$$

whose kernel is $F \cap (f) = \{0\}$, since deg $f \ge 1$. So the map $a \mapsto a + (f)$ embeds F into K = F[x]/(f). Let $\overline{\cdot}$ denote the associate embedding of F[x] into K[x]. Then it is not hard to see that

$$\overline{f}(x+(f)) = f(x) + (f) = (f),$$

which is zero in K. So x + (f) is the desired root of f.

Example 2. The polynomial $x^2 + 1 \in \mathbb{F}_3[x]$ has no roots so is irreducible. Let $p(x) \in \mathbb{F}_3[x]$ and use the division algorithm to write $p(x) = q(x)(x^2 + 1) + a + bx$ over \mathbb{F}_3 . Then $p(x) + (x^2 + 1) = a + bx + (x^2 + 1)$. Moreover, because deg $a + bx < \deg x^2 + 1$, it is not difficult to conclude that each coset $a + bx + (x^2 + 1)$ is distinct. Denote it by a + bi and notice that $i^2 = x^2 + (x^2 + 1) = -1 + (x^2 + 1) = -1 + 0i$. We conclude that

$$K = \mathbb{F}_3[x]/(x^2 + 1) = \{a + bi \,|\, i^2 = -1\},\$$

a field extension of \mathbb{F}_3 with nine elements, containing the root *i* of $x^2 + 1$.

We now turn to the questions posed in $\mathbf{Q2}$. The answer to the first question is tied up with what are called *chain conditions* on ideals.

Definition 2. In a ring R an ascending chain of ideals is a sequence of ideals $\{I_k\}$ in R satisfying

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq I_4 \subseteq \cdots, \tag{1}$$

We say that such a chain *stabilizes* if there is an $n \in \mathbb{N}$ so that $I_n = I_k$ for all $k \ge n$. Finally, we say that R satisfies the *ascending chain condition* (ACC) if every ascending chain of ideals in R stabilizes. A ring satisfying the ACC is called *Noetherian*.

Before examining how the ACC is related to factorizations we provide an alternate formulation.

Theorem 3. Let R be a commutative ring with unity. Then R is Noetherian if and only if every ideal in R has the form

 $Ra_1 + Ra_2 + \dots + Ra_n = (a_1, a_2, \dots, a_n)$

for some $n \in \mathbb{N}$ and $a_i \in R$, *i.e.*, every ideal is finitely generated.

Proof. (\Leftarrow) Suppose every ideal in R is finitely generated and consider a chain (1). Let

$$I = \bigcup_{k=1}^{\infty} I_k.$$

I is an ideal in R and consequently $I = (a_1, a_2, \dots, a_m)$ for some $a_i \in R$. For each *i*, there is an $n_i \in \mathbb{N}$ so that $a_i \in I_{n_i}$. Let $n = \max_i \{n_i\}$. Then $a_i \in I_{n_i} \subseteq I_n$ for all *i* and hence

$$I \subseteq I_n \subseteq I_{n+1} \subseteq I_{n+2} \subseteq \cdots I,$$

so that $I = I_n = I_k$ for all $k \ge n$. Hence (1) must stabilize and R is Noetherian.

 (\Rightarrow) We prove the contrapositive. Suppose that there exists an ideal I in R that is not finitely generated. Construct a chain of ideals in R as follows. Begin by choosing a nonzero $a_1 \in I$ and let $I_1 = (a_1)$. In general, given $I_k = (a_1, a_2, \ldots, a_k)$, choose $a_{k+1} \in I \setminus I_{k+1}$ and let $I_{k+1} = (a_1, a_2, \ldots, a_k, a_{k+1})$. We can continue this construction indefinitely because I is not finitely generated, producing a chain

$$I_1 \subset I_2 \subset I_3 \subset \cdots$$

of proper containments which cannot stabilize. This shows that R is not Noetherian and completes the proof.

Corollary 2. Every PID is Noetherian.

Remark 1. We will primarily be concerned with chains of principal ideals. In that context it is worth noting that if D is a domain and $a, b \in D \setminus \{0\}$ then:

- $(a) \subseteq (b)$ if and only if b|a (as my advisor used to say, "To contain is to divide.");
- (a) = (b) if and only if a = bc for some $c \in D^{\times}$ (a and b are associates);
- $(a) \subset (b)$ if and only if a = bc for some $c \in D \setminus (D^{\times} \cup \{0\})$.

Lemma 3. Let D be a domain. If $a \in D \setminus (D^{\times} \cup \{0\})$ is not equal to a product of irreducibles, then there exists $b \in D \setminus (D^{\times} \cup \{0\})$ so that $(a) \subset (b)$ and b is not equal to a product of irreducibles.

▼

Proof. Since a is not a product of irreducibles, it is not itself irreducible. Therefore a = bc for some $b, c \in D \setminus (D^{\times} \cup \{0\})$. It cannot be the case that both b and c are products of irreducibles, lest a be, so without loss of generality b is not a product of irreducibles. Since a = bc and c is not a unit, we have $(a) \subset (b)$. This is what we needed to show.

Theorem 4. Let D be a domain satisfying the ACC on principal ideals. Then every element of $D \setminus (D^{\times} \cup \{0\})$ can be written as a product of irreducible elements in D.

Proof. According to Lemma 3, any given chain

$$(a_1) \subset (a_2) \subset (a_3) \subset \cdots \subset (a_n)$$

of principal ideals generated by elements in $D \setminus (D^{\times} \cup \{0\})$ that are not products of irreducibles can be properly extended to the right by an ideal of the same type. Therefore, if there is an element in $D \setminus (D^{\times} \cup \{0\})$ that is not a product of irreducibles (to initiate the chain), then we can create an ascending chain of principal ideals that does not stabilize, so that D does not satisfy the ACC on principal ideals. This proves the contrapositive of the statement we sought to establish, so we are finished.

Corollary 3. If D is Noetherian, then every element of $D \setminus (D^{\times} \cup \{0\})$ has a factorization into irreducible elements of D.

Corollary 4. If D is a PID, then every element of $D \setminus (D^{\times} \cup \{0\})$ has a factorization into irreducible elements of D.

As we've just seen, Theorem 4 is a useful tool to establish the existence of irreducible factorizations in a given domain or class of domains. So we now turn to the second question posed in **Q2**, that of the uniqueness of factorizations. The first thing we notice is that in passing from N to Z, both a prime number and its negative are irreducible/prime. This suddenly means that the uniqueness of prime factorizations guaranteed by the FTA no longer holds, e.g. $6 = 2 \cdot 3 = (-2)(-3)$ exhibits two distinct prime factorizations of 6. However, the two are only off by the signs of the primes (units in Z). We can get around this irritation and restore uniqueness of prime factorizations in Z if we identify a prime with its negative (its associate). We will take this tack with our general domains as well. Two factorizations will be considered equivalent if they both are made up of the same factors, up to order and association (multiplication of a factor by a unit).

Remark 2. But even with this "correction" we find immediately that irreducible factorizations need not be unique. Consider the example of 4 in $D = \mathbb{Z}[\sqrt{-3}]$. We have seen that $4 = 2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3}) = \alpha \overline{\alpha}$, that α is irreducible in D, and α does not divide 2. In particular, 2 and α are not associates in D. It is not hard to show that the same statements are true of $\overline{\alpha}$. We will therefore have two distinct irreducible factorizations of 4 in D if we can show that 2 is irreducible in D. Once again we use the norm. If 2 = xy in D then

$$4 = N(2) = N(x)N(y),$$

and we have already seen that this equation implies x or y is a unit in D. This shows 2 is irreducible and concludes our argument.

Again taking the FTA as our guide, we find that in its proof the uniqueness of prime factorizations is established by using property **ii** of primes, not property **i**. This property we generalized to the actual definition of a prime element in a domain. This suggests that if we want *unique* factorizations, our factorizations need to be into primes, which are of course irreducible. We will see that this is exactly the correct line of reasoning. We begin with a fundamental lemma.

Lemma 4. Let D be a domain and let $p, q \in D \setminus (D^{\times} \cup \{0\})$ be prime elements. If p|q then p and q are associates.

Proof. If p|q then $(q) \subseteq (p)$. Since all primes are irreducible, Lemma 1 tells us that (q) is maximal among principal ideals, and hence (p) = (q) or (p) = D. Since $p \notin D^{\times}$, $(p) \neq D$. So we must have (p) = (q) and the statement of the lemma follows from our earlier remark.

Theorem 5 (Uniqueness of Prime Factorizations). Let D be a domain and let $p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_\ell \in D \setminus (D^{\times} \cup \{0\})$ be prime. If

$$\prod_{i=1}^{k} p_i = \prod_{j=1}^{\ell} q_j$$

then $k = \ell$ and, possibly after reordering, p_i is associate to q_i .

Proof. We induct on $k \in \mathbb{N}$. When k = 1 we have $p_1 = q_1 q_2 \cdots q_\ell$. As p_1 is irreducible, if $\ell \geq 2$ then either $q_1 \in D^{\times}$ or $q_2 \cdots q_\ell \in D^{\times}$, implying $q_i \in D^{\times}$ for $i \geq 2$. Both situations are contradictory, so we must have $\ell = 1$ and the equation $p_1 = q_1$, which establishes what we need for the case k = 1.

Now suppose we have established the theorem for some $k\geq 1$ and that

$$\prod_{i=1}^{k+1} p_i = \prod_{j=1}^{\ell} q_j,$$

all p_1 and q_j primes in D. Since p_1 is prime, by reordering $p_1|q_1$. Then by Lemma 4 $uq_1 = p_1$ for some $u \in D^{\times}$. Cancellation then yields

$$u \prod_{i=2}^{k+1} p_i = \prod_{j=2}^{\ell} q_j,$$

The product on the right cannot be empty, otherwise the primes appearing on the left (of which there is at least one) would all be units. Replacing p_2 with its associate up_2 , we find that we may apply the inductive hypothesis to conclude that $\ell = k+1$, and, possibly after reordering, p_i is associate to q_i for all $2 \le i \le k+1$. Since we have already shown that p_1 and q_1 are associates, this concludes the inductive step and therefore the proof.

So factorizations into primes are always unique, in the sense of Theorem 5. Therefore if factorizations into irreducibles are to be unique, we should demand that they in fact are prime factorizations. We therefore make the following (non-standard, but still useful) definition.

Definition 3. An *FTA domain* is a domain *D* is which every $a \in D \setminus (D^{\times} \cup \{0\})$ can be written as a product of irreducibles and every irreducible element is prime

Remark 3. The condition that being irreducible is equivalent to being prime is independent from the existence of factorizations into irreducibles: a domain may have one without having the other.

- $\mathbb{Z}[\sqrt{-3}]$ is Noetherian, therefore has factorizations into irreducibles. But we have already seen that irreducibles in this ring need not be prime.
- In the monoid-ring $F[x; \mathbb{Q}_0^+]$, the "polynomials" in nonnegative rational powers of x with coefficients in F, irreducible elements are prime, but the element x does not have a factorization into irreducibles. See [1]

An immediate consequence of this definition and Theorem 5 is that if D is an FTA domain and $a \in D \setminus (D^{\times} \cup \{0\})$, then a has a factorization into irreducibles that is unique (in the sense of that theorem). Another way to obtain this result is, perhaps obviously, to simply assert it.

Definition 4. Let *D* be a domain. Suppose every element of $a \in D \setminus (D^{\times} \cup \{0\})$ has a unique factorization into irreducibles, i.e. there exist irreducibles $r_1, r_2, \ldots, r_k \in D$ so that $a = r_1 r_2 \cdots r_k$, and if $s_1, s_2, \ldots, s_\ell \in D$ are irreducible and satisfy $a = s_1 s_2 \cdots s_\ell$, then $k = \ell$ and, possibly after reordering, r_i is associate to s_i for all *i*. Then *D* is called a *unique factorization domain* (UFD).

Corollary 5. Every FTA domain is a UFD.

Theorem 6. If D is a Noetherian domain in which every irreducible is prime, then D is an FTA domain, hence a UFD.

Proof. By Corollary 3, every element of D possesses a factorization into irreducibles. Thus D is an FTA domain. By Corollary 5, D is a UFD.

Corollary 6 (PID implies UFD). If D is a PID, then D is a UFD.

▼

Proof. By Corollary 2, D is Noetherian. By Theorem 1 irreducibles and primes coincide in D. Now apply the preceding result.

Example 3. If F is a field, then F[x] is a PID and hence a UFD.

(Non)Example. Although we mentioned Bézout domains in connection with the potential equality of prime and irreducible, they need not possess an irreducible factorization for every element, and so are not necessarily UFDs. A somewhat advanced example of this is the ring of entire functions on the complex plane. Incidentally, this proves that this ring is non-Noetherian.

Theorem 7 (Irreducible Equals Prime in a UFD). Let D be a UFD. If $a \in D \setminus (D^{\times} \cup \{0\})$ is irreducible, then a is prime.

Proof. Suppose $a \in D \setminus (D^{\times} \cup \{0\})$ is irreducible and a|bc in $D \setminus (D^{\times} \cup \{0\})$. Write ad = bc and express each of b, c and d as a product of irreducibles. Without loss of generality, by uniqueness of factorizations, an associate of a must occur in the factorization of b. Hence a|b and a is prime.

Corollary 7 (FTA iff UFD). Every UFD is an FTA domain.

Moral. The ACC on principal ideals yields irreducible factorizations, while prime factorizations are those that are unique. Therefore one way to prove a domain is a UFD is to establish that it has the ACC on principal ideals (or is more generally Noetherian) and that irreducibles are prime.

If D is a UFD, then it is possible to define the notion of content in D[x], prove the analogue of Gauss' Lemma and establish the following result, by essentially the same means.

Theorem 8. Let D be a UFD, $f \in D[x]$ of degree at least 1, and F denote the quotient field of D. If f reduces over F, it reduces over D, and the factors have the same degree.

Although we won't provide the details, a this and a little more work allows one to prove the following result.

Corollary 8. If D is a UFD, then so is D[x].

We have seen that PIDs are UFDs, and the two examples we have of this so far are \mathbb{Z} and F[x]. The techniques used to prove that these are PIDs were essentially the same: both used the division algorithm in a specific context. This can be abstracted as follows.

Definition 5. Let *D* be a domain. We call *D* a *Euclidean domain* (ED) if there exists a function $d : D \setminus \{0\} \to \mathbb{N}_0$ so that:

a. $d(a) \leq d(ab)$ for all $a, b \in D \setminus \{0\}$;

b. if $a, b \in D$ and $b \neq 0$, then there exist $q, r \in D$ so that a = bq + r, and r = 0 or d(r) < d(b).

Example 4. \mathbb{Z} in an ED with d(n) = |n|. F[x] is an ED with $d(f) = \deg(f)$.

Example 5. The Gaussian integers $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z} \text{ and } i^2 = -1\}$ is an ED with

$$d(a + bi) = N(a + bi) = (a + bi)(a - bi) = a^{2} + b^{2}.$$

To prove this, we need a few properties of the complex $modulus |z|, z \in \mathbb{C}$.

Given a complex number z = x + iy we define its *conjugate* to be $\overline{z} = x - iy$. Clearly $\overline{\overline{z}} = z$ and it is not hard to show that $\overline{zw} = \overline{zw}$. We define the *modulus* of $z = x + iy \in \mathbb{C}$ to be

$$|z| = \sqrt{z\overline{z}} = \sqrt{x^2 + y^2}.$$

It satisfies |zw| = |z||w| and (although we won't need it) the triangle inequality $|z+w| \le |z|+|w|$.

The key observation to be made is that if we identify the complex numbers z = x + iy and w = u + iv with the points $(x, y), (u, v) \in \text{Re}^2$, then

$$|z-w| = |(x-u) + i(y-v)| = \sqrt{(x-u)^2 + (y-v)^2} =$$
distance from (x,y) to (u,v) .

We now verify that d has the properties required to make $\mathbb{Z}[i]$ an ED. Since $d(z) = |z|^2$, for any nonzero $x, y \in \mathbb{Z}[i]$ we have

$$d(xy) = d(x)d(y) \ge d(x)$$

since $d(x) \in \mathbb{N}$. As for the division algorithm, let $a, b \in \mathbb{Z}[i]$ with $b \neq 0$. Consider $a/b \in \mathbb{C}$. If $a/b \in \mathbb{Z}[i]$, then a = bq with $q \in \mathbb{Z}[i]$ and we're done. Otherwise, notice that the points of $\mathbb{Z}[i]$ geometrically make up a square lattice in the complex plane, each square having side length 1. Since a/b will lie within at least one of these squares, it will be within $\sqrt{2}/2$ of at least one of the corners $q \in \mathbb{Z}[i]$. That is

$$0 < \left|\frac{a}{b} - q\right| \le \frac{\sqrt{2}}{2} \quad \Rightarrow \quad 0 < \left|\frac{a - bq}{b}\right| \le \frac{\sqrt{2}}{2} \quad \Rightarrow \quad 0 < |a - bq| \le \frac{\sqrt{2}}{2}|b| \quad \Rightarrow \quad 0 < d(a - bq) \le \frac{d(b)}{2} < d(b),$$

where to obtain the final inequality we have squared the terms of the previous one. If we now set r = a - bq, we are finished.

Because they all have a division algorithm, as one might expect, every Euclidean domain is principal.

Theorem 9. (ED implies PID) Let D be a ED. Then D is a PID.

Proof. Let I be a nonzero ideal of D. The Well Ordering Principal then implies that there exists a nonzero $a \in I$ so that $d(a) \leq d(b)$ for all $b \in I$. Let $b \in I$. Write b = aq + r with r = 0 or d(r) < d(a). Since $r = b - aq \in I$, we cannot have d(r) < d(a) by our choice of a, and so r = 0. That is, $b = aq \in (a)$. Since b was an arbitrary element of I, we conclude that $I \subseteq (a) \subseteq I$, and hence that I = (a). As I was an arbitrary ideal in D, we conclude that D is a PID.

Remark 4. We have just finished proving the chain of implications

$$ED \Rightarrow PID \Rightarrow UFD.$$

It is worth noting that none of these implications is reversible. By Corollary 8, $\mathbb{Z}[x]$ and F[x, y] provide examples of UFDs that are not PIDs. An example of a PID that is not an ED is

$$\mathbb{Z}\left[\frac{1+\sqrt{-19}}{2}\right] = \left\{ \left. \frac{a+b\sqrt{-19}}{2} \right| \ a, b \in \mathbb{Z} \text{ and } a \equiv b \pmod{2} \right\}.$$

▼

For details, see [2].

References

- [1] Daileda, R. C., A Non-UFD Integral Domain in Which Irreducibles are Prime, preprint, available at http://ramanujan.math.trinity.edu/rdaileda/teach/m4363s07/non_ufd.pdf.
- [2] Wilson, J. C., A Principal Ideal Ring That is Not a Euclidean Ring, Math. Mag. 46 (1) (1973), 34–38.