Historical Note

The question of unique factorization in an integral domain other than the integers was first raised in public in connection with the attempted proof by Gabriel Lamé (1795–1870) of Fermat's Last Theorem, the conjecture that $x^n + y^n = z^n$ has no nontrivial integral solutions for $n > 2$. It is not hard to show that the conjecture is true if it can be proved for all odd primes p. At a meeting of the Paris Academy on March 1, 1847, Lamé announced that he had proved the theorem and presented a sketch of the proof. Lamé's idea was first to factor $x^p + y^p$ over the complex numbers as

$$x^p + y^p = (x + y)(x + \alpha y)(x + \alpha^2 y) \cdots (x + \alpha^{p-1} y)$$

where α is a primitive pth root of unity. He next proposed to show that if the factors in this expression are relatively prime and if $x^p + y^p = z^p$, then each of the p factors must be a pth power. He could then demonstrate that this Fermat equation would be true for a triple x', y', z', each number smaller than the corresponding number in the original triple. This would lead to an infinite descending sequence of positive integers, an impossibility that would prove the theorem.

After Lamé finished his announcement, however, Joseph Liouville (1809–1882) cast serious doubts on the purported proof, noting that the conclusion that each of the relatively prime factors was a pth power because their product was a pth power depended on the result that any integer can be uniquely factored into a product of primes. It was by no means clear that "integers" of the form $x + \alpha^k y$ had this unique factorization property. Although Lamé attempted to overcome Liouville's objections, the matter was settled on May 24, when Liouville produced a letter from Ernst Kummer noting that in 1844 he had already proved that unique factorization failed in the domain $\mathbb{Z}[\alpha]$, where α is a 23rd root of unity.

It was not until 1994 that Fermat's Last Theorem was proved, and by techniques of algebraic geometry unknown to Lamé and Kummer. In the late 1950s, Yutaka Taniyama and Goro Shimura noticed a curious relationship between two seemingly disparate fields of mathematics, elliptic curves and modular forms. A few years after Taniyama's tragic death at age 31, Shimura clarified this idea and eventually formulated what became known as the Taniyama–Shimura Conjecture. In 1984, Gerhard Frey asserted and in 1986 Ken Ribet proved that the Taniyama–Shimura Conjecture would imply the truth of Fermat's Last Theorem. But it was finally Andrew Wiles of Princeton University who, after secretly working on this problem for seven years, gave a series of lectures at Cambridge University in June 1993 in which he announced a proof of enough of the Taniyama–Shimura Conjecture to derive Fermat's Last Theorem. Unfortunately, a gap in the proof was soon discovered, and Wiles went back to work. It took him more than a year, but with the assistance of his student Richard Taylor, he finally was able to fill the gap. The result was published in the *Annals of Mathematics* in May 1995, and this 350-year-old problem was now solved.

The fact that $F[x]$ is a UFD, where F is a field (by Theorem 28.21), illustrates both theorems. For by Theorem 31.24, $F[x]$ is a PID. Also, since F has no nonzero elements that are not units, F satisfies our definition for a UFD. Thus Theorem 34.30 would give another proof that $F[x]$ is a UFD, except for the fact that we shall actually use Theorem 28.21 in proving Theorem 34.30. In the following section we shall study properties of a certain special class of UFDs, the *Euclidean domains*.

Let us proceed to prove the two theorems.

Every PID Is a UFD

The steps leading up to Theorem 28.21 and its proof indicate the way for our proof of Theorem 34.18. Much of the material will be repetitive. We inefficiently handled the special case of $F[x]$ separately in Theorem 28.21, since it was easy and was the only case we needed for our field theory in general.

To prove that an integral domain D is a UFD, it is necessary to show that both Conditions 1 and 2 of the definition of a UFD are satisfied. For our special case of $F[x]$ in Theorem 28.21, Condition 1 was very easy and resulted from an argument that in a factorization of a polynomial of degree > 0 into a product of two nonconstant polynomials, the degree of each factor was less than the degree of the original polynomial. Thus we couldn't keep on factoring indefinitely without running into unit factors, that is, polynomials of degree 0. For the general case of a PID, it is harder to show that this is so. We now turn to this problem. We shall need the definition of the union of an arbitrary collection of sets. The definition must include the possibility that the collection of sets is infinite.

34.8 Definition If $\{A_i \mid i \in I\}$ is a collection of sets, then the **union** $\cup_{i\in I} A_i$ **of the sets** A_i is the set of all x such that $x \in A_i$ for at least one $i \in I$. ■

34.9 Lemma Let R be a commutative ring and let $N_1 \subseteq N_2 \subseteq \cdots$ be an ascending chain of ideals N_i in R. Then $N = \cup_i N_i$ is an ideal of R.

Proof Let $a, b \in N$. Then there are ideals N_i and N_j in the chain, with $a \in N_i$ and $b \in N_j$. Now either $N_i \subseteq N_j$ or $N_j \subseteq N_i$; let us assume that $N_i \subseteq N_j$, so both a and b are in N_j. This implies that $a \pm b$ and ab are in N_j, so $a \pm b$ and ab are in N. Taking $a = 0$, we see that $b \in N$ implies $-b \in N$, and $0 \in N$ since $0 \in N_i$. Thus N is a subring of D. For $a \in N$ and $d \in D$, we must have $a \in N_i$ for some N_i. Then since N_i is an ideal, $da = ad$ is in N_i. Therefore, $da \in \cup_i N_i$, that is, $da \in N$. Hence N is an ideal. ◆

34.10 Lemma **(Ascending Chain Condition for a PID)** Let D be a PID. If $N_1 \subseteq N_2 \subseteq \cdots$ is an ascending chain of ideals N_i, then there exists a positive integer r such that $N_r = N_s$ for all $s \geq r$. Equivalently, every strictly ascending chain of ideals (all inclusions proper) in a PID is of finite length. We express this by saying that the **ascending chain condition** (ACC) holds for ideals in a PID.

Proof By Lemma 34.9, we know that $N = \cup_i N_i$ is an ideal of D. Now as an ideal in D, which is a PID, $N = \langle c \rangle$ for some $c \in D$. Since $N = \cup_i N_i$, we must have $c \in N_r$, for some $r \in \mathbb{Z}^+$. For $s \geq r$, we have

$$\langle c \rangle \subseteq N_r \subseteq N_s \subseteq N = \langle c \rangle.$$

Thus $N_r = N_s$ for $s \geq r$.

The equivalence with the ACC is immediate. ◆

34.11 Definition A commutative ring with unity R that satisfies the ascending chain condition is a **Noetherian ring**. That is, a commutative ring with unity R is Noetherian if for every chain of ideals $N_1 \subseteq N_2 \subseteq N_3 \subseteq \ldots$ in R, there is an integer r such that if $s \geq r$, then $N_r = N_s$. ■

Lemma 34.10 states that every PID is a Noetherian ring. In Section 37 we will see that if R is a Noetherian ring, then $R[x]$ is also a Noetherian ring.

In what follows, it will be useful to remember that for elements a and b of a domain D,

$\langle a \rangle \subseteq \langle b \rangle$ if and only if b divides a, and

$\langle a \rangle = \langle b \rangle$ if and only if a and b are associates.

For the first property, note that $\langle a \rangle \subseteq \langle b \rangle$ if and only if $a \in \langle b \rangle$, which is true if and only if $a = bd$ for some $d \in D$, so that b divides a. Using this first property, we see that $\langle a \rangle = \langle b \rangle$ if and only if $a = bc$ and $b = ad$ for some $c, d \in D$. But then $a = adc$ and by canceling, we obtain $1 = dc$. Thus d and c are units, so a and b are associates.

We can now prove Condition 1 of the definition of a UFD for an integral domain that is a PID.

34.12 Theorem Let D be a PID. Every element that is neither 0 nor a unit in D is a product of irreducibles.

Proof Let $a \in D$, where a is neither 0 nor a unit. We first show that a has at least one irreducible factor. If a is an irreducible, we are done. If a is not an irreducible, then $a = a_1 b_1$, where *neither* a_1 nor b_1 is a unit. Now

$$\langle a \rangle \subset \langle a_1 \rangle,$$

for $\langle a \rangle \subseteq \langle a_1 \rangle$ follows from $a = a_1 b_1$, and if $\langle a \rangle = \langle a_1 \rangle$, then a and a_1 would be associates and b_1 would be a unit, contrary to construction. Continuing this procedure then, starting now with a_1, we arrive at a strictly ascending chain of ideals

$$\langle a \rangle \subset \langle a_1 \rangle \subset \langle a_2 \rangle \subset \cdots .$$

By the ACC in Lemma 34.10, this chain terminates with some $\langle a_r \rangle$, and a_r must then be irreducible. Thus a has an irreducible factor a_r.

By what we have just proved, for an element a that is neither 0 nor a unit in D, either a is irreducible or $a = p_1 c_1$ for p_1 an irreducible and c_1 not a unit. By an argument similar to the one just made, in the latter case we can conclude that $\langle a \rangle \subset \langle c_1 \rangle$. If c_1 is not irreducible, then $c_1 = p_2 c_2$ for an irreducible p_2 with c_2 not a unit. Continuing, we get a strictly ascending chain of ideals

$$\langle a \rangle \subset \langle c_1 \rangle \subset \langle c_2 \rangle \subset \cdots .$$

This chain must terminate, by the ACC in Lemma 34.10, with some $c_r = q_r$ that is an irreducible. Then $a = p_1 p_2 \cdots p_r q_r$. ◆

This completes our demonstration of Condition 1 of the definition of a UFD. Let us turn to Condition 2. Our arguments here are parallel to those leading to Theorem 28.21. The results we encounter along the way are of some interest in themselves.

34.13 Lemma **(Generalization of Theorem 31.25)** An ideal $\langle p \rangle$ in a PID is maximal if and only if p is an irreducible.

Proof Let $\langle p \rangle$ be a maximal ideal of D, a PID. Suppose that $p = ab$ in D. Then $\langle p \rangle \subseteq \langle a \rangle$. Suppose that $\langle a \rangle = \langle p \rangle$. Then a and p would be associates, so b must be a unit. If $\langle a \rangle \neq \langle p \rangle$, then we must have $\langle a \rangle = \langle 1 \rangle = D$, since $\langle p \rangle$ is maximal. But then a and 1 are associates, so a is a unit. Thus, if $p = ab$, either a or b must be a unit. Hence p is an irreducible of D.

Conversely, suppose that p is an irreducible in D. Then if $\langle p \rangle \subseteq \langle a \rangle$, we must have $p = ab$. Now if a is a unit, then $\langle a \rangle = \langle 1 \rangle = D$. If a is not a unit, then b must be a unit, so there exists $u \in D$ such that $bu = 1$. Then $pu = abu = a$, so $\langle a \rangle \subseteq \langle p \rangle$, and we have $\langle a \rangle = \langle p \rangle$. Thus $\langle p \rangle \subseteq \langle a \rangle$ implies that either $\langle a \rangle = D$ or $\langle a \rangle = \langle p \rangle$, and $\langle p \rangle \neq D$ or p would be a unit. Hence $\langle p \rangle$ is a maximal ideal. ◆

34.14 Lemma **(Generalization of Theorem 31.27)** In a PID, if an irreducible p divides ab, then either $p \mid a$ or $p \mid b$.

Proof Let D be a PID and suppose that for an irreducible p in D we have $p \mid ab$. Then $(ab) \in \langle p \rangle$. Since every maximal ideal in D is a prime ideal by Corollary 31.16, $(ab) \in \langle p \rangle$ implies that either $a \in \langle p \rangle$ or $b \in \langle p \rangle$, giving either $p \mid a$ or $p \mid b$. ◆

34.15 Corollary If p is an irreducible in a PID and p divides the product $a_1 a_2 \cdots a_n$ for $a_i \in D$, then $p \mid a_i$ for at least one i.

Proof Proof of this corollary is immediate from Lemma 34.14 if we use mathematical induction. ◆

34.16 Definition A nonzero nonunit element p of an integral domain D is a **prime** if, for all $a, b \in D$, $p \mid ab$ implies either $p \mid a$ or $p \mid b$. ■

Lemma 34.14 focused our attention on the defining property of a prime. In Exercises 25 and 26, we ask you to show that a prime in an integral domain is always an irreducible and that in a UFD an irreducible is also a prime. Thus the concepts of prime and irreducible coincide in a UFD. Example 34.17 will exhibit an integral domain containing some irreducibles that are not primes, so the concepts do not coincide in every domain.

34.17 Example Let F be a field and let D be the subdomain $F[x^3, xy, y^3]$ of $F[x, y]$. Then x^3, xy, and y^3 are irreducibles in D, but

$$(x^3)(y^3) = (xy)(xy)(xy).$$

Since xy divides x^3y^3 but not x^3 or y^3, we see that xy is not a prime. Similar arguments show that neither x^3 nor y^3 is a prime. ▲

The defining property of a prime is precisely what is needed to establish uniqueness of factorization, Condition 2 in the definition of a UFD. We now complete the proof of Theorem 34.18 by demonstrating the uniqueness of factorization in a PID.

34.18 Theorem **(Generalization of Theorem 28.21)** Every PID is a UFD.

Proof Theorem 34.12 shows that if D is a PID, then each $a \in D$, where a is neither 0 nor a unit, has a factorization

$$a = p_1p_2 \cdots p_r$$

into irreducibles. It remains for us to show uniqueness. Let

$$a = q_1q_2 \cdots q_s$$

be another such factorization into irreducibles. Then we have $p_1 \mid (q_1q_2 \cdots q_s)$, which implies that $p_1 \mid q_j$ for some j by Corollary 34.15. By changing the order of the q_j if necessary, we can assume that $j = 1$ so $p_1 \mid q_1$. Then $q_1 = p_1u_1$, and since q_1 is an irreducible, u_1 is a unit, so p_1 and q_1 are associates. We have then

$$p_1p_2 \cdots p_r = p_1u_1q_2 \cdots q_s,$$

so by the cancellation law in D,

$$p_2 \cdots p_r = u_1q_2 \cdots q_s.$$

Continuing this process, starting with p_2 and so on, we finally arrive at

$$1 = u_1u_2 \cdots u_rq_{r+1} \cdots q_s.$$

Since the q_j are irreducibles, we must have $r = s$. ◆

Example 34.32 at the end of this section will show that the converse to Theorem 34.18 is false. That is, a UFD need not be a PID.

Many algebra texts start by proving the following corollary of Theorem 34.18. We have assumed that you were familiar with this corollary and used it freely in our other work.

34.19 Corollary **(Fundamental Theorem of Arithmetic)** The integral domain $\mathbb{Z}$ is a UFD.