

**WHY FACTORIZATION THEORY:
THE BASICS**

opening address at the

MAA-PREP FACTORIZATION WORKSHOP

Monday, May 21, 2007

by

Scott Chapman

Trinity University

Department of Mathematics

One Trinity Place

San Antonio, Texas 78212-7200

Goals of this talk:

- (1) Introduce Factorization theory in the manner in which I first learned it - via algebraic rings of integers and Dedekind domains.
- (2) Give the basic notation and definitions that the remaining speakers this week will build upon.
- (3) Start considering an interesting “cast of characters” which will again be expanded as the week progresses.
- (4) From this “cast” derive some elementary examples that have interesting behavior with respect to their factorization properties.
- (5) Emphasize why this material is interesting and important.

Consider the integral domain

$$D = \mathbb{Z}[\sqrt{-5}].$$

In D , 6 factors as

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

Given that 1 and -1 are the only two units of D , one can argue (using the norm function) that

- (1) 2, 3, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are irreducible in D , and
- (2) 2 (respectively 3) is not an associate of either $1 + \sqrt{-5}$ or $1 - \sqrt{-5}$.

Hence, $(*)$ is usually the first example encountered by students of an integral domain where the Fundamental Theorem of Arithmetic fails.

A simpler example can be constructed as follows. Let K be any field and let $D = K[X^2, X^3]$. With even less work than above, we have that both X^2 and X^3 are irreducible in D and

$$(**) \quad X^2 \cdot X^2 \cdot X^2 = X^3 \cdot X^3.$$

Here, the number of irreducible factors on each side of $(**)$ is unequal.

The roots of the theory of non-unique factorizations lies in algebraic number theory and (*) allows us to get to the heart of the matter. I will require some basic structure from Commutative Algebra.

Let D be an integral domain, and let K be its field of fractions. A *fractional ideal* of D is a nonzero finitely generated D -submodule of K . A fractional ideal I is contained in D if and only if it is an *integral* ideal of D .

If \mathcal{O}_K represents the ring of integers in an algebraic extension K of \mathbb{Q} , then \mathcal{O}_K is a *Dedekind domain*. We give a formal definition.

Definition 1. An integral domain D is a Dedekind domain if and only if

- (1) every ideal in D is finitely generated,
- (2) every nonzero prime ideal is a maximal ideal, and
- (3) D is integrally closed in its fraction field.

A Dedekind domain D has several important properties:

- (1) Every ideal I can be factored uniquely as a product of prime ideals from D .
- (2) If I is an ideal from D , then there exists a fractional ideal I^{-1} of D such that $I \cdot I^{-1} = D$.

Given a Dedekind domain D , let

$$\mathcal{F}(D) = \text{fractional ideals of } D$$

and

$$\mathcal{P}(D) = \text{the principal fractional ideals of } D.$$

Then

$$\mathcal{C}(D) = \mathcal{F}(D)/\mathcal{P}(D)$$

forms a group known as the *ideal class group of D* . In general, this group may be infinite, but a classical result from algebraic number theory indicates that if $D = \mathcal{O}_K$, then the class group is finite. In this instance, $|\mathcal{C}(\mathcal{O}_K)|$ is known as the *class number of \mathcal{O}_K* . Another important classical result concerning $\mathcal{C}(\mathcal{O}_K)$ is that each equivalence class (or *ideal class*) of the class group contains an integral prime ideal. This definition of the class number leads to a basic result.

Theorem 2. \mathcal{O}_K is a unique factorization domain if and only if the class number of \mathcal{O}_K is 1.

Hence, the class number was interpreted in a classical sense as the distance a ring of integers was from having unique factorization of elements into irreducible elements.

The ring of integers used in (*) has an interesting property. If an integer α in $\mathbb{Z}[\sqrt{-5}]$ has non-unique factorizations into products of irreducibles, then the number of irreducibles in each factorization of α remains the same. This is reflected in the class number by this classical result of Carlitz (*Proc. Amer. Math. Soc.* **11**(1960), 391–392).

Theorem 3 (Carlitz). *The algebraic number ring \mathcal{O}_K has class number ≤ 2 if and only if for every nonzero integer $\alpha \in \mathcal{O}_K$ the number of primes π_j in every factorization*

$$\alpha = \pi_1 \pi_2 \cdots \pi_k$$

only depends on α .

One could prove Carlitz's Theorem, but let's consider a deeper structure which tells more of the story.

$$(\alpha) = P_1 \cdots P_k \rightarrow [(\alpha)] = [P_1 \cdots P_k] = [P_1] \cdots [P_k] = [(1)].$$

Let's demonstrate this by returning to (*).

$$\langle 2 \rangle = \langle 2, 1 + \sqrt{-5} \rangle^2$$

$$\langle 3 \rangle = \langle 3, 1 - \sqrt{-5} \rangle \langle 3, 1 + \sqrt{-5} \rangle$$

$$\langle 1 + \sqrt{-5} \rangle = \langle 2, 1 + \sqrt{-5} \rangle \langle 3, 1 + \sqrt{-5} \rangle$$

$$\langle 1 - \sqrt{-5} \rangle = \langle 2, 1 + \sqrt{-5} \rangle \langle 3, 1 - \sqrt{-5} \rangle$$

And so

$$\langle 6 \rangle = \langle 2 \rangle \langle 3 \rangle =$$

$$\langle 2, 1 + \sqrt{-5} \rangle^2 \langle 3, 1 - \sqrt{-5} \rangle \langle 3, 1 + \sqrt{-5} \rangle =$$

$$\langle 2, 1 + \sqrt{-5} \rangle \langle 3, 1 + \sqrt{-5} \rangle \langle 2, 1 + \sqrt{-5} \rangle \langle 3, 1 - \sqrt{-5} \rangle =$$

$$\langle 1 + \sqrt{-5} \rangle \langle 1 - \sqrt{-5} \rangle$$

$$\langle 6 \rangle = \langle 2 \rangle \langle 3 \rangle =$$

$$\langle 2, 1 + \sqrt{-5} \rangle^2 \langle 3, 1 - \sqrt{-5} \rangle \langle 3, 1 + \sqrt{-5} \rangle =$$

$$\langle 2, 1 + \sqrt{-5} \rangle \langle 3, 1 + \sqrt{-5} \rangle \langle 2, 1 + \sqrt{-5} \rangle \langle 3, 1 - \sqrt{-5} \rangle =$$

$$\langle 1 + \sqrt{-5} \rangle \langle 1 - \sqrt{-5} \rangle$$

This brings us to our first of several “characters” for the week.

Let G be an additively written abelian torsion group. We denote by \mathbb{Z}_n a cyclic group of order $n \in \mathbb{N}$.

Let $\mathcal{F}(G)$ denote the, multiplicatively written, free abelian monoid over G . An element $S \in \mathcal{F}(G)$ is called a sequence over G_0 . By definition

$$S = \prod_{g \in G_0} g^{v_g}$$

with $v_g \in \mathbb{N}_0$ where almost all v_g s equal 0, and $S = \prod_{i=1}^l g_i$ for some $l \in \mathbb{N}_0$ and $g_i \in G_0$, which are uniquely determined up to permutation. One calls

$$\sigma(S) = \sum_{i=1}^n g_i \in G$$

the sum of S .

The monoid

$$\mathcal{B}(G) = \{S \in \mathcal{F}(G) : \sigma(S) = 0\} \subset \mathcal{F}(G)$$

is called the *block monoid* over G . Its elements are called zero-sum sequences. We let $\mathcal{A}(\mathcal{B}(G))$ represent the irreducible elements or *atoms* of $\mathcal{B}(G)$. The elements of $\mathcal{A}(G)$ are the minimal zero-sum sequences over G_0 .

Example 4. Let $G = \mathbb{Z}/3\mathbb{Z} \sim \mathbb{Z}_3 = \{[0], [1], [2]\}$. Then

$$\mathcal{B}(\mathbb{Z}_3) = \{[0]^{x_1}[1]^{x_2}[2]^{x_3} \mid x_i \in \mathbb{N}_0 \text{ and } x_2 + 2x_3 \equiv 0 \pmod{3}\}.$$

The irreducible elements (or minimal zero-sequences) are

$$[0], [1]^3, [2]^3, \text{ and } [1][2].$$

We can further observe the following non-unique factorization in $\mathcal{B}(\mathbb{Z}_3)$.

$$[1]^3[2]^3 = ([1][2])^3.$$

NOTE: Factorizations problems on an integral domain D are merely factorization problems on the multiplicative monoid of D which we denote D^* .

Hence, we will state the definitions and notation of factorization theory using the language of monoids. The translation to domains, as mentioned above, is relatively simple.

Throughout, we assume that M is a commutative cancellative monoid. Unless otherwise noted, we write the operation of M multiplicatively and hence represent its identity element by 1_M . We use the standard notation of divisibility theory; if x and y are in M and there exists c in M with $cx = y$, then $x \mid y$. Denote by

$$M^\times = \{u \in M \mid uv = 1_M \text{ for some } v \in M\}$$

the set of units of M . The *irreducibles* (or *atoms*) of M are denoted $\mathcal{A}(M)$, where

$$\mathcal{A}(M) = \{x \in M \setminus M^\times \mid x = rs \text{ with}$$

$$r, s \in M \text{ implies } r \in M^\times \text{ or } s \in M^\times\}.$$

The monoid M is *atomic* if every element of $M \setminus M^\times = M^\bullet$ possesses a factorization into elements of $\mathcal{A}(M)$. We leave for later discussion the study of monoids which are not atomic. Hence, unless otherwise noted, we assume throughout the remainder of these notes that M is atomic.

Two elements x and y in $\mathcal{A}(M)$ are called *associates* if there exists a unit $u \in M^\times$ such that $x = uy$. If x and y are associates, then we write $x \simeq y$.

Given an element $x \in M$, suppose that

$$(\dagger) \quad x = p_1 \cdots p_t = q_1 \cdots q_k$$

where each p_i and q_j is in $\mathcal{A}(M)$. M is *factorial* if for every nonunit $x \in M$ and factorization of the form (\dagger) , then $t = k$ and there exists a permutation σ of $\{1, \dots, t\}$ with $p_i \simeq q_{\sigma(i)}$ for all i .

You are familiar with a particular type of irreducible element. Let M be a commutative cancellative monoid and $x \in M^\times$. We call x *prime* if whenever $x \mid yz$, for y and z in M , then either $x \mid y$ or $x \mid z$. With respect to studying elements in M which do not factor uniquely as a product of irreducibles, the prime elements essentially play no role, as the following Lemma indicates.

Lemma 5. *Let M be a commutative cancellative monoid $x \in M$ be a prime element.*

- (i) *x is irreducible in M .*
- (ii) *Suppose $y \in M^\times$ and $x \mid y$. If $y = x_1 \cdots x_t$ is a factorization of y into irreducibles of M , then $x \simeq x_i$ for some i .*

Hence, if a prime x appears in a irreducible factorization of an element $y \in M$, then some associate of x appears in *every* irreducible factorization of y in M . Lemma 5 yields a nice characterization of commutative cancellative monoids which are factorial.

Corollary 6. *Let M be an atomic commutative cancellative monoid. M is factorial if and only if every irreducible element of M is prime.*

Consider factorizations of the form (\dagger) which may not be unique. If $x \in M^\bullet$, then *the set of lengths of x* is

$$\mathcal{L}(x) = \{k \in \mathbb{N} \mid x = a_1 a_2 \cdots a_k \text{ where } a_i \in \mathcal{A}(M)\}.$$

If $|\mathcal{L}(x)| = 1$ for every $x \in M^\bullet$, then M is called a half-factorial monoid. By Carlitz's Theorem, any ring of integers \mathcal{O}_K with class number less than or equal to 2 is half-factorial.

Set

$$\ell(x) = \min \mathcal{L}(x) \text{ and } L(x) = \max \mathcal{L}(x).$$

We define the total set of lengths of M to be

$$\mathfrak{L}(M) = \{\mathcal{L}(x) \mid x \in M \setminus M^\times\}.$$

If $\mathcal{L}(x) = \{n_1, \dots, n_t\}$ with the n_i 's listed in increasing order, then set

$$\Delta(x) = \{n_i - n_{i-1} \mid 2 \leq i \leq t\}$$

and

$$\Delta(M) = \bigcup_{1 \neq x \in M} \Delta(x).$$

If $\Delta(M) \neq \emptyset$, then,

$$\min \Delta(M) = \gcd \Delta(M).$$

The *elasticity* of an element $x \in M$, denoted $\rho(x)$, is given by

$$\rho(x) = \max(\mathcal{L}(x)) / \min(\mathcal{L}(x)).$$

The *elasticity of M* is then defined as

$$\rho(M) = \sup\{\rho(x) \mid x \in M \setminus M^\times\}.$$

We say that M has *accepted elasticity* if there exists $x \in M$ such that $\rho(x) = \rho(M)$. We say that M is *fully elastic* if for all $q \in \mathbb{Q} \cap [1, \rho(M)]$ (or $[1, \infty)$ if the elasticity is infinite) there exists a nonunit $x \in H$ such that $\rho(x) = q$.

WHY IS $\mathcal{B}(G)$ SO IMPORTANT?

Theorem 7. *Let \mathcal{O}_K be a ring of algebraic integers. The monoid homomorphism*

$$\varphi : \mathcal{O}_K - \{0\} \rightarrow \mathcal{B}(\mathcal{C}(\mathcal{O}_K))$$

defined by

$$\varphi(\alpha) = [P_1] \cdots [P_k]$$

where

$$(\alpha) = P_1 \cdots P_k$$

preserves lengths of factorizations. In other words

$$\mathcal{L}(\alpha) = \mathcal{L}(\varphi(\alpha))$$

for all nonunits α in \mathcal{O}_K .

It follows directly that

$$\mathfrak{L}(\mathcal{O}_K) = \mathfrak{L}(\mathcal{B}(\mathcal{C}(\mathcal{O}_K))).$$

Theorem 8. *Let G be a finite abelian group. Then*

$$\rho(\mathcal{B}(G)) = D(G)/2$$

and the elasticity is accepted.

Proof. Let B be a nonunit of $\mathcal{B}(G)$ and suppose that $B = A_1 \dots A_n = C_1 \dots C_m$ are two different factorizations of B into irreducible elements of $\mathcal{B}(G)$. We show that $n/m \leq D(G)/2$. Now,

$$B = g_1 \cdots g_k,$$

where g_1, \dots, g_k are not necessarily distinct elements of G . If A is an irreducible of $\mathcal{B}(G)$ which divides B , then $A = g_{i_1} \cdots g_{i_w}$ where i_1, \dots, i_w is a subsequence of $1, \dots, k$. Since A is irreducible we have that $w \leq D(G)$. Thus

$$k/D(G) \leq l(B) \text{ and } L(B) \leq k/2.$$

It follows that,

$$n/m \leq L(B)/l(B) \leq D(G)/2.$$

If $B_1 = g_1 \cdots g_{D(G)}$ is a minimal zero-sequence of G of length $D(G)$, then so too is $B_2 = (g_1)^{-1} \cdots (g_{D(G)})^{-1}$. Hence, in $\mathcal{B}(G)$ we have

$$B_1 B_2 = E_1 \cdots E_{D(G)}$$

where $E_i = (g_i)(g_i)^{-1}$. Thus $\rho(B_1 B_2) = D(G)/2$. □

MORE OF THE CAST OF CHARACTERS

Definition 9. Let a and b be in \mathbb{N} . If 1) $a \leq b$ and 2) $a^2 \equiv a \pmod{b}$, then the set

$$M(a, b) = \{1, a, a + b, a + 2b, a + 3b, \dots\}$$

is a multiplicative submonoid of \mathbb{N} known as an *arithmetical congruence monoid (or ACM)*. Notice with the exception of the element 1 (which is the identity element) that all elements x of $M(a, b)$ satisfy $x \equiv a \pmod{b}$.

ACMs have been the focus of study in three recent papers. Among other things that have been shown are the following.

- (1) Necessary and sufficient conditions for $\rho(M(a, b)) < \infty$.
- (2) A formula to compute the elasticity when $\rho(M(a, b)) < \infty$.
- (3) If $\gcd(a, b) = p^\alpha$, then $\Delta(M(a, b))$ is completely determined.
- (4) There are ACMs with accepted elasticity and some without accepted elasticity.

□

Example 10 (The Hilbert Monoid). Set $a = 1$ and $b = 4$ above. Then

$$\begin{aligned} M(1, 4) &= \{1, 5, 9, 13, 17, \dots\} \\ &= \{x \mid x \in \mathbb{N} \text{ and } x \equiv 1 \pmod{4}\}. \end{aligned}$$

Factorization in M into irreducibles is not unique for

$$441 = 9 \cdot 49 = 21 \cdot 21$$

where 9, 49 and 21 are distinct elements of $\mathcal{A}(M)$. Using elementary Number Theory, one can show that if x is in $\mathcal{A}(M)$, then either $x = p$ where p is a prime number with $p \equiv 1 \pmod{4}$ or $x = q_1 q_2$ where q_1 and q_2 are prime numbers with $q_1 \equiv 3 \pmod{4}$ and $q_2 \equiv 3 \pmod{4}$. Hence, if y is in M and

$$y = p_1 \cdots p_s q_1 \cdots q_t$$

where each p_i a prime number with $p_i \equiv 1 \pmod{4}$ and q_j are prime numbers with $q_j \equiv 3 \pmod{4}$, then any irreducible factorization of y in M has length $s + \frac{t}{2}$. Thus M is half-factorial. □

Example 11 (Meyerson's Monoid). Set $a = 4$ and $b = 6$.

Here

$$M(4, 6) = \{1, 4, 10, 16, 22, 28, \dots\}.$$

By a formula you will likely see on Tuesday, $\rho(M(4, 6)) = 2$, but the elasticity is *not* accepted. I will not prove this, but the argument centers around the following observation. The atoms of the $M(4, 6)$ fall into two types: A) atoms of the form $2r$ where r is an odd number congruent to 2 (mod 3) and B) atoms of the form $4s$ where s is a product of odd primes all of which are congruent to 1 (mod 3). \square

An Interesting Observation: Consider

$$\mathbb{Z}[\sqrt{-5}]^* = \mathbb{Z}[\sqrt{-5}] - \{0\}$$

as a monoid under multiplication. As earlier, set $x \simeq y$ in $\mathbb{Z}[\sqrt{-5}]^*$ if and only if x and y are associates. The relation \simeq is a congruence on $\mathbb{Z}[\sqrt{-5}]^*$, and the quotient monoid $\mathbb{Z}[\sqrt{-5}]^*/\simeq$ has a unique unit. Using the theory of Krull monoids, it follows that

$$\mathbb{Z}[\sqrt{-5}]^*/\simeq \cong M(1, 4).$$

Definition 12. Let S be an additive submonoid of $\mathbb{N} \cup \{0\}$. S is called a *numerical monoid*. If $\{n_1, \dots, n_t\}$ is a set of elements of S such that every $x \in S$ can be written in the form

$$x = x_1 n_1 + \dots + x_t n_t$$

then $\{n_1, \dots, n_t\}$ is called a *generating set of S* . This is commonly denoted by

$$S = \langle n_1, \dots, n_t \rangle.$$

It follows from Elementary Number Theory that every numerical monoid S possesses a unique minimal set of generators. If $\gcd\{s \mid s \in S\} = 1$, then S is called *primitive*. It again follows easily from Number Theory that every numerical monoid S is isomorphic to a primitive numerical monoid.

The factorization properties of numerical monoids have been studied in depth and four principle results are obtained.

- (1) [Holden-Moore] If $S = \langle n_1, \dots, n_t \rangle$ where $n_1 < \dots < n_t$ and the generators $\{n_1, \dots, n_t\}$ form a minimal set, then $\rho(S) = n_t/n_1$ and the elasticity is accepted.
- (2) [Holden-Moore] If S is as in (1) and $t \geq 2$, then S is not fully elastic.

- (3) Let n, k and t be a positive integers with $n > 1$. If $S = \langle n, n + k, n + 2k, \dots, n + tk \rangle$, then $\Delta(S) = \{k\}$.
- (4) Numerical monoids are not determined by their sets of lengths (i.e., there exist $S \neq S'$ with $\mathfrak{L}(S) = \mathfrak{L}(S')$).

□

Example 13. We illustrate this with an example. Let

$$S = \{0, 2, 3, 4, 5, 6, 7, \dots\} = \langle 2, 3 \rangle.$$

Notice that 2 and 3 are the only irreducible elements of S . Hence, an irreducible factorization of $n \in S$ is of the form

$$n = x_1 \cdot 2 + n_2 \cdot 3.$$

Factorizations in S are far from unique as

$$\begin{aligned} 17 &= 7 \cdot 2 + 1 \cdot 3 \\ &= 4 \cdot 2 + 3 \cdot 3 \\ &= 1 \cdot 1 + 5 \cdot 3. \end{aligned}$$

Thus, $\mathfrak{L}(17) = \{6, 7, 8\}$, $\ell(17) = 6$, $L(17) = 8$, $\Delta(17) = \{1\}$ and $\rho(17) = 8/6 = 4/3$. Notice that the longest factorization of an element $n \in S$ contains the most possible copies of 2 and the shortest the most possible copies of 3. Using this fact and some Number Theory, all the invariants discussed to this point

can be worked out. For instance, for all $m \in S$ we have that

$$\ell(m) = \lceil \frac{m}{3} \rceil \text{ and } L(m) = \lfloor \frac{m}{2} \rfloor$$

and for all $m \geq 4$ in S that

$$\mathcal{L}(m) = \{\lceil \frac{m}{3} \rceil, \lceil \frac{m}{3} \rceil + 1, \dots, \lfloor \frac{m}{2} \rfloor - 1, \lfloor \frac{m}{2} \rfloor\}.$$

If $x \leq y$ are both integers, then set $\{z \in \mathbb{Z} \mid x \leq z \leq y\} = [x, y]$. We then have

$$\mathfrak{L}(M) = \{\{1\}, [\lceil \frac{4}{3} \rceil, \lfloor \frac{4}{2} \rfloor], [\lceil \frac{5}{3} \rceil, \lfloor \frac{5}{2} \rfloor], \dots\}.$$

and further that

$$\rho(m) = \frac{\lfloor \frac{m}{2} \rfloor}{\lceil \frac{m}{3} \rceil}, \Delta(m) = \{1\}, \text{ and } \Delta(S) = \{1\}.$$

Moreover, it is easy to verify that $\rho(m) \leq 3/2$ for all m and $\rho(m) = 3/2$ if $m \equiv 0 \pmod{6}$. Thus $\rho(S) = 3/2$ and the elasticity is accepted. \square

SO: WHY FACTORIZATION THEORY?

- (1) Problems involving non-unique factorizations originated with classical algebraic number theory and expand upon an area (unique factorization) which fueled much of the development of early 20th century algebra.
- (2) Solutions of such problems often involve mathematical techniques from many different branches of mathematics, such as commutative algebra, classical number theory, combinatorics, graph theory, abelian group theory, discrete geometry and additive number theory.
- (3) Problems involving factorizations can vary widely with respect to the amount of sophistication required for their solutions. Some are doctoral level in nature and require many years of background to solve. Others can easily be introduced and are accessible to even advanced undergraduates.
- (4) With many undergraduate programs shifting to a project based “capstone” requirement, problems which have reasonable levels of access and still contain challenging mathematics are now badly needed.