GALOIS THEORY TOPIC V
POLYNOMIAL ARITHMETIC

PAUL L. BAILEY

Abstract. Algebra originated as the study of polynomial equations. What is required of for a polynomial to define a function on a set is that the set is closed under addition, subtraction, and multiplication. Thus we first create an abstract definition for such a set, called a *ring*, explicitly stating the properties that we require of addition, subtraction, and multiplication.

Next we proceed to study polynomials over a field, and fully develop the analogy between rings of such polynomials and the ring of integers. Specifically, we see that polynomials admit the division algorithm, the Euclidean algorithm, and unique factorization.

1. Rings and Fields

Definition 1. A *ring* consists of a set R together with two binary operations

$$+ : R \times R \to R \quad \text{and} \quad \cdot : R \times R \to R$$

satisfying

(R1) $a + b = b + a$ for every $a, b \in R$;
(R2) $(a + b) + c = a + (b + c)$ for every $a, b, c \in R$;
(R3) there exists $0 \in R$ such that $a + 0 = a$ for every $a \in R$;
(R4) for every $a \in R$ there exists $-a \in R$ such that $a + (-a) = 0$;
(R5) $ab = ba$ for every $a, b \in R$;
(R6) $(ab)c = a(bc)$ for every $a, b, c \in R$;
(R7) there exists $1 \in R$ such that $a \cdot 1 = a$ for every $a \in R$;
(R8) $a(b + c) = ab + bc$ for every $a, b, c \in R$.

This is commonly called a *commutative ring*, in expositions where a ring need not satisfy **(R1)**. However, all of the rings we will consider are commutative.

In a ring, we define subtraction by $a - b = a + (-b)$.

Definition 2. A *field* is a ring which satisfies the additional property

(R9) for every $a \in R \smallsetminus \{0\}$ there exists $a^{-1} \in R$ such that $aa^{-1} = 1$.

In a field, we define division by $\frac{a}{b} = ab^{-1}$, where b is nonzero.

We have already met the rings $\mathbb{Z}$, and the fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, as well as various subsets of $\mathbb{C}$ which are also fields, We have also considered the rings $\mathbb{Z}_n$, for $n \in \mathbb{N}$ with $n \geq 2$, and the field $\mathbb{Z}_p$, where p is a prime integer. The last type of ring we which to consider are rings of polynomials.

Date: October 16, 2006.

2. Basic Ring Properties

After this section, the only fields with which we will be dealing are $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}_p$, and the subfields of $\mathbb{C}$, and the primary rings of interest for us will be $\mathbb{Z}$, $\mathbb{Z}_n$, and $F[x]$ (the ring of polynomials over a field, which is the subject of this document). However, before we continue, it is informative and useful to point out some aspects of rings which are completely general; these properties follow directly from the axioms.

First we point out that the additive identity 0 is unique, since if e and f are additive identities, we have $e = e + f = f$. Moreover, additive inverses are unique, since if $a + b = 0$ and $a + c = 0$, then $b = b + (a + c) = (b + a) + c = (a + b) + c = c$. Similarly, the multiplicative identity 1 is unique, and multiplicative inverses are unique when they exist.

Proposition 1. (Cancellation Law of Addition)
Let R be a ring and let $a, b, c \in R$. If $a + c = b + c$, then $a = b$.

Proof. Add $-c$ be both sides on the right to obtain

$$\begin{aligned} a + c = b + c &\Rightarrow (a + c) + (-c) = (b + c) + (-c) \quad \text{because } + \text{ is a function} \\ &\Rightarrow a + (c + (-c)) + b + (c + (-c)) \quad \text{by } \textbf{(R8)} \\ &\Rightarrow a + 0 = b + 0 \quad \text{by } \textbf{(R4)} \\ &\Rightarrow a = b \quad \text{by } \textbf{(R3)}. \end{aligned}$$

This completes the proof. □

Proposition 2. (Multiplication by Zero)
Let R be a ring and let $a \in R$. Then $a \cdot 0 = 0$.

Proof. We have $a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$. Thus $0 + a \cdot 0 = a \cdot 0 + a \cdot 0$, so by cancellation, $0 = a \cdot 0$. □

Proposition 3. *Let R be a ring and let $a, b \in R$. Then $(-a)b = -(ab)$.*

Proof. Note what this is saying: if you take the multiplicative inverse of a and multiply it by b, you get the multiplicative inverse of the product ab.

Now since additive inverses are unique, it suffices to show that $(-a)b$ acts like an additive inverse of ab. This is true by the distributive property, since $ab + (-a)b = (a + (-a))b = 0 \cdot b = b \cdot 0 = 0$. □

3. Invertible and Entire Elements

Definition 3. Let R be a ring and let $a \in R$. We say that a is *invertible* if there exists $a^{-1} \in R$ such that $aa^{-1} = 1$. We call a^{-1} the *inverse* of a.

The invertible element of $\mathbb{Z}$ are ± 1, and the invertible elements of $\mathbb{Z}_n$ are those $a \in \mathbb{Z}_n$ such that $\gcd(a, n) = 1$. Every nonzero element of a field is invertible.

Definition 4. Let R be a ring and let $a \in R$. We say that a is *entire* if

$$ab = 0 \Rightarrow b = 0, \qquad \text{for all } b \in R.$$

If a is not entire, we say that a is a *zero-divisor*.

We say that R is *entire* if every nonzero element of R is entire.

Thus a is a zero divisor if there exists a nonzero element $b \in R$ such that $ab = 0$. Note that, in this case, b is also a zero-divisor.

Proposition 4. *Let R be a ring and let $a \in R$. If a is invertible, then a is entire.*

Proof. Suppose a is invertible, and that $ab = 0$. Multiply by a^{-1} to get $a^{-1}ab = a^{-1} \cdot 0$, so $b = 0$. □

Since every nonzero element of a field is invertible, every nonzero element of a field is entire, so a field is entire. Thus $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are entire rings. Also, the ring of integers $\mathbb{Z}$ is an entire ring. We have seen that $\mathbb{Z}_n$ is entire if and only if it is a field, which happens if and only if n is prime.

Definition 5. Let R be a ring and let $a \in R$. We say that a is *cancellable* if whenever $ab = ac$, then $b = c$.

Proposition 5. *Let R be a ring and let $a \in R$. Then a is entire if and only if a is cancellable.*

Proof. Suppose that a is entire, and that $ab = ac$. Then $a(b - c) = 0$, and since a is entire, we have $b - c = 0$, whence $b = c$. Thus a is cancellable.

On the other hand, suppose a is cancellable, and that $ab = 0$. Then $ab = a \cdot 0$, so by the cancellability of a, $b = 0$. Thus a is entire. □

4. Irreducible and Prime Elements

Definition 6. Let R be a ring and let $p \in R$ be an entire noninvertible element.

We say that p is *irreducible* if whenever $p = ab$, then either a is invertible or b is invertible.

We say that p is *prime* if whenever $p \mid ab$, then either $p \mid a$ or $p \mid b$.

Proposition 6. *Let R be a ring and let $p \in R$. If a is prime, then p is invertible.*

Proof. Suppose that p is prime, and suppose that $p = ab$. We wish to show that either a is invertible or b is invertible.

Since $p = ab$, we have $p \mid ab$, and since p is prime, either $p \mid a$ or $p \mid b$. Suppose that $p \mid a$; then $a = pc$ for some $c \in R$. Thus $p = pcb$, and since p is entire, it is cancellable, so $bc = 1$. Thus b is invertible.

Similarly, if $p \mid b$, then a is invertible. □

Notice that in the case $R = \mathbb{Z}$, our definition of irreducible is the same as Euclid's definition of prime. This is standard, and is allowable because, in the case of the integers, prime and irreducible are equivalent. This, however, is not the case in general. For example, consider the set $R = \mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$. This is clearly a subring of $\mathbb{C}$. Let $z = 2 + \sqrt{-5}$, and notice that $3^2 = 9 = a\overline{a}$; now 3 is irreducible and divides 9, but it does not divide either of the factors a or $\overline{a}$.

It is the equivalence of primeness and irreducibility which leads to unique factorization in the integers. This equivalence comes from the Euclidean algorithm, which in turn comes from the division algorithm.

5. Subrings

Definition 7. Let R be a ring and let $S \subset R$. We say that S is a *subring* of R if

(S0) $0, 1 \in S$;
(S1) $a, b \in S$ implies $a + b \in S$;
(S2) $a \in R$ implies $-a \in S$;
(S3) $a, b \in S$ implies $ab \in S$.

If S is a subring of R, it is also a *subfield* if additionally

(S4) $a \in S \smallsetminus \{0\}$ implies $a^{-1} \in S$.

Thus the fields we have previously considered, such as $\mathbb{Q}[\sqrt{2}]$, are actually subfields of the field $\mathbb{C}$. The ring $\mathbb{Z}[\sqrt{-5}]$ is a subring of the field $\mathbb{Q}[i]$.

6. Polynomials over a Ring

Definition 8. Let R be a field. A *polynomial over* R is a function $f : R \to R$ of the form

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n,$$

where $a_i \in F$ for $i = 0, \dots, n$. It is the *zero polynomial* if $n = 0$ and $a_n = 0$; otherwise assume that $a_n \neq 0$. The numbers a_i are called the *coefficients* of f.

We call n the *degree* of f, denoted $\deg(f)$. We call a_n the *leading coefficient* of f, denoted $\mathrm{LC}(f)$. We call a_0 the *constant coefficient* of f, denoted $\mathrm{CC}(f)$. We say that f is *monic* if $a_n = 1$.

The set of all polynomials over R is denoted $R[x]$.

We identify a constant polynomial of the form $f(x) = a_0$ with the number a_0; in this way, we view R as a subring of $R[x]$.

Definition 9. We give names to polynomials based on their degrees, as follows:

- A *constant* polynomial is a polynomial of degree 0.
- A *linear* polynomial is a polynomial of degree 1.
- A *quadratic* polynomial is a polynomial of degree 2.
- A *cubic* polynomial is a polynomial of degree 3.
- A *quartic* polynomial is a polynomial of degree 4.
- A *quintic* polynomial is a polynomial of degree 5.

Definition 10. Let R be a ring and let $f, g \in R[x]$. Define the *sum* and *product* of f and g by

$$(f+g)(x) = f(x) + g(x) \quad \text{and} \quad (fg)(x) = f(x) \cdot g(x).$$

This produces the operations of addition and multiplication of polynomials on the set $R[x]$.

Proposition 7. *Let R be a ring. Then $R[x]$ is a ring.*

Reason. Since addition and multiplication in the set $R[x]$ of polynomial functions are defined pointwise, the ring properties carry over directly from the ring properties of R. □

We will typically consider the ring of polynomials over a field F; the only case where this won't hold it the case of polynomials with integer coefficients, the set of which is the ring $\mathbb{Z}[x]$. Actually, since integers are rational numbers, we view $\mathbb{Z}[x]$ as a subring of $\mathbb{Q}[x]$.

If R is not entire, we can run into problems with standard properties of polynomials which we want, as indicated in the next proposition.

Proposition 8. *Let $f, g \in R[x]$. Then*

(a) $\deg(f+g) = \max\{\deg(f), \deg(g)\}$, *unless* $\mathrm{LC}(f) - \mathrm{LC}(g) = 0$.

(b) $\deg(fg) = \deg(f) + \deg(g)$, *unless* $\mathrm{LC}(f)\,\mathrm{LC}(g) = 0$.

We give an example where the condition of entireness of the leading coefficients is important. Let $R = \mathbb{Z}_6$, $f(x) = 2x^3 + x^2 + 5$, and $g(x) = 3x^2 + 5x + 3$. Then, computing in $\mathbb{Z}_6[x]$, we have

$$(fg)(x) = \overline{6}x^5 + \overline{10+3}x^4 + \overline{6+5}x^3 + \overline{3+15}x^2 + \overline{25}x + \overline{3} = x^4 + 5x^3 + x + 3.$$

7. The Division Algorithm

Henceforth, let F be a field and let $F[x]$ be the ring of polynomials over F. In this case, we can divide by the leading coefficients and obtain an inductive process that allows polynomial division; this produces a strong analogy between integer arithmetic and polynomial arithmetic, which we now develop.

Proposition 9. (Division Algorithm for Polynomials)
Let $f, g \in F[x]$, where f and g are nonzero. Then there exist polynomials $q, r \in F[x]$ such that

$$g = fq + r, \qquad \textit{where} \qquad \deg(r) < \deg(f).$$

We call q the quotient *and r the* remainder*.*

Proof. If $\deg(f) > \deg(g)$, let $q = 0$ and $r = g$. Otherwise, we have $\deg(f) \leq \deg(g)$.

We write the proof to mimic the well-known algorithm for division; we proceed by strong induction on the larger degree $\deg(g)$.

If $\deg(f) > \deg(g)$, let $q = 0$ and $r = g$. Otherwise, we have $\deg(f) \leq \deg(g)$.

Let $d = \deg(g) - \deg(f)$; then $d \geq 0$. Since f is nonzero, $\mathrm{LC}(f) \neq 0$; set $a = \frac{\mathrm{LC}(g)}{\mathrm{LC}(f)}$. Then $a \in F$, and the highest order term of $g(x)$ is f times ax^d.

Let $q_1 = g - fax^d$; then $q_1 \in F[x]$, and $\deg(q_1) < \deg(g)$. By induction on the degree, there exist polynomials q_2, r such that $q_1 = fq_2 + r$, with $\deg(r) < \deg(f)$. Thus $fq_2 + r = g - fax^d$. With $q = q_2 + ax^d$, we have $g = fq + r$. □

Proposition 10. (Remainder Theorem)
Let $g \in F[x]$ and let $a, r \in F$. Define $f \in F[x]$ by $f(x) = x - a$. and let $g = fq + r$ where $\deg(r) < \deg(f)$*. Then $r \in F$, and $f(a) = r$.*

Proof. Since $\deg(f) = 1$ and $\deg(r) < \deg(f)$, we must have $\deg(r) = 0$, so $r \in F$. Now $g(a) = f(a)q(a) + r = (a - a)q(a) + r = r$. □

Definition 11. Let $f, g \in F[x]$. We say that f *divides* g, and write $f \mid g$, if there exists $q \in F[x]$ such that $g = fq$.

The following are synonyms: f divides g, f is a factor of g, g is a multiple of f.

Proposition 11. (Factor Theorem)
Let $g \in F[x]$ and let $a \in F$. Define $f \in F[x]$ by $f(x) = x - a$. Then

$$f \mid g \quad \Leftrightarrow \quad g(a) = 0.$$

Proof. If $f \mid g$, then $g = fq$ for some $q \in F[x]$, and $g(a) = (a - a)q(a) = 0$.

On the other hand, if $g(a) = 0$, we divide g by f to obtain $g = fq + r$ with $\deg(r) < \deg(f)$. By the Remainder Theorem, $g(a) = r = 0$, so $g = fq$, and $f \mid g$. □

Definition 12. Let $f \in F[x]$ and $a \in F$. We say that a is a *root* of f if $f(a) = 0$.

Proposition 12. (Bound on Roots Corollary)
Let $g \in F[x]$. Then the number of roots of g cannot exceed $\deg(g)$*.*

Proof. Let $n = \deg(f)$, and let $a \in F$ is a root and set $f(x) = x - a$. Then $g = fq$ for some q, where $\deg(q) = n - 1$. By induction, q has at most $n - 1$ roots, and these together with a make at most n roots for f. □

8. The Euclidean Algorithm

Definition 13. Let $f, g \in F[x]$. A *greatest common divisor* of f and g is a polynomial $d \in F[x]$ satisfying

(a) $d \mid f$ and $d \mid g$;
(b) $e \mid f$ and $e \mid g$ implies $e \mid d$, for any $e \in F[x]$.

Proposition 13. *Let $f, g \in F[x]$. If $f \mid g$ and $g \mid f$, then $g = af$ for some $a \in F$.*

Proof. If $f \mid g$ and $g \mid f$, then $g = hf$ and $f = kg$ for some $h, k \in F[x]$. Then $g = hkg$, so $\deg(g) = \deg(hk) + \deg(g)$, so $\deg(hk) = 0$, and $hk \in F$. Set $a = hk$. □

Proposition 14. *Let $f, g, d, e \in F[x]$. If d and e are greatest common divisors of f and g, then $d = ae$ for some $a \in F$. Thus there is a unique monic greatest common divisor, which we denote by* $\gcd(f, g)$.

Proof. Suppose d and e are greatest common divisors of f and g. By **(b)** of the definition, $e \mid d$ and $d \mid e$. Thus $d = ae$ for some $a \in F$. Divide any greatest common divisor by its leading coefficient to obtain the unique monic greatest common divisor. □

Lemma 1. *Let $g, f, q, r \in F[x]$ with $g = fq + r$ and* $\deg(r) < \deg(f)$. *Then* $\gcd(g, f) = \gcd(f, r)$.

Proof. Let $d = \gcd(g, f)$; we wish to show that $d = \gcd(f, r)$.

Now $r = g - fq$, and d divides g and f; this means that $g = dq_1$ and $f = dq_2$ for some $q_1, q_2 \in F[x]$. Thus $r = d(q_1 - q_2)$, and $d \mid r$; **(a)** is satisfied.

Suppose that $e \mid f$ and $e \mid r$ for some $e \in F[x]$. Then $f = eq_3$ and $r = eq_4$ for some $q_3, q_4 \in F[x]$. Then $g = fq + r = eq_3q + eq_4 = e(q_3q + q_4)$. Thus $e \mid g$. Since $d = \gcd(g, f)$, $e \mid d$; **(b)** is satisfied. □

Proposition 15. (Euclidean Algorithm for Polynomials)
Let $f, g \in F[x]$. Then $d = \gcd(f, g)$ exists, and there exist polynomials $s, t \in F[x]$ such that

$$fs + gt = d.$$

Proof. Without loss of generality, we may assume that $\deg(g) \geq \deg(f)$; we proceed by induction on the smaller degree $\deg(f)$.

By the division algorithm, there exist $q, r \in F[x]$ such that $g = fq + r$, and $\deg(r) < \deg(f)$. By induction, there exist polynomials $s_1, t_1, d \in F[x]$ such that $d = \gcd(f, r)$ and $rs_1 + ft_1 = d$. Since $r = g - fq$, we have $(g - fq)s_1 + ft_1 = d$. Set $s = t_1 - q$ and $t = s_1$ to obtain $fs + gt = d$. Moreover, $d = \gcd(g, f)$ by the lemma. □

9. Irreducibility

Definition 14. Let $f \in F[x]$ be nonconstant. We say that f is *reducible over F* if there exist $g, h \in F[x]$, with $\deg(g) < \deg(f)$ and $\deg(h) < \deg(f)$, such that $f = gh$. Otherwise, we say that f is *irreducible*.

Note that if $f \in F[x]$ is nonzero and $a \in F$, we can always let $g = af$ and $h = \frac{1}{a}f$ to get $f = gh$. This is referred to as an improper factorization. We are not interested in these. Thus, f is irreducible if

$$f = gh \Rightarrow g \in F \text{ or } h \in F.$$

What we call irreducible here is the analog of what was called prime by Euclid in the case of the integers.

Proposition 16. (Euler's Argument for Polynomials)
Let $f, g, h \in F[x]$ be nonzero with f irreducible, and suppose that $f \mid gh$. Then $f \mid g$ or $f \mid h$.

Proof. Suppose that f does not divide g; we show that f divides h.

Since f is irreducible, the only factors of f are constants and constant multiples of f. Since f does not divide g, the only common factors are constants. Thus $\gcd(f, g) = 1$, and there exist polynomials $s, t \in F[x]$ such that

$$fs + gt = 1.$$

Multiplying this equation by h gives $fhs + ght = h$. Since f divides gh, we have $gh = fk$ for some $k \in F[x]$. Thus $fhs + fkt = h$, so $f(hs + kt) = h$, whence f divides h. □

Proposition 17. (Fundamental Theorem of Polynomial Arithmetic)
Let $f \in F[x]$. Then there exist irreducible polynomials $p_1, \dots, p_r \in F[x]$, unique up to order and multiplication by constants, such that $f = \prod_{i=1}^{r} p_i$.

Proof. If f is irreducible, set $r = 1$ and $p_1 = f$. Otherwise, f has a proper factorization $f = gh$ where $\deg(g) < \deg(f)$ and $\deg(h) < \deg(f)$. By induction on the degree, we declare that g and h are products of irreducible elements, and so then if f.

Uniqueness follows from Euclid's argument. Thus suppose that $f = p_1 \cdots p_r = q_1 \cdots q_s$ for some positive integers r, s and irreducible polynomials p_i, q_j. Then p_1 divides $q_1 \dots q_s$, so by repeated use of Euclid's argument, p_1 divides one of the q_j's; without loss of generality, we may assume that p_1 divides q_1. Since q_1 is irreducible and p_1 is nonconstant, we must have $q_1 = ap_1$ for some $a \in F$. Factoring out q_1 and continuing this process, we see that the each of the q_j's is a constant multiple of one of the p_i's, and that $r = s$. □

Department of Mathematics and CSci, Southern Arkansas University
E-mail address: plbailey@saumag.edu