17

Finite Fields and Combinatorial Designs

It is time now to recall the ring structure of Chapter 14 as we examine rings of polynomials and their role in the construction of finite fields. We know that for every prime p, $(\mathbf{Z}_p, +, \cdot)$ is a finite field, but here we shall find other finite fields. Just as the order of a finite Boolean algebra is restricted to powers of 2, for finite fields the possible orders are p^n, where p is a prime and $n \in \mathbf{Z}^+$. Applications of these finite fields will include a discussion of such combinatorial designs as Latin squares. Finally, we shall investigate the structure of a finite geometry and discover how these geometries and combinatorial designs are interrelated.

17.1 Polynomial Rings

We recall that a ring $(R, +, \cdot)$ consists of a nonempty set R, where $(R, +)$ is an abelian group, $(R, \cdot)$ is closed under the associative operation $\cdot$, and the two operations are related by the distributive laws: $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$, for all $a, b, c \in R$. (We write ab for $a \cdot b$.)

In order to introduce the formal concept of a polynomial with coefficients in R we let x denote an indeterminate — that is, a formal symbol that is not an element of the ring R. We then use this symbol x to define the following.

Definition 17.1

Given a ring $(R, +, \cdot)$, an expression of the form $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x^1 + a_0 x^0$, where $a_i \in R$ for all $0 \leq i \leq n$, is called a *polynomial in the indeterminate x with coefficients from R*.

If a_n is not the zero element of R, then a_n is called *the leading coefficient of* $f(x)$ and we say that $f(x)$ has *degree n*. Hence the degree of a polynomial is the highest power of x that occurs in a summand of the polynomial. The term $a_0 x^0$ is called the *constant*, or *constant term*, of $f(x)$.

If $g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x^1 + b_0 x^0$ is also a polynomial in x over R, then $f(x) = g(x)$ if $m = n$ and $a_i = b_i$ for all $0 \leq i \leq n$.

Finally, we use the notation $R[x]$ to represent the set of all polynomials in the indeterminate x with coefficients from R.

EXAMPLE 17.1

a) Over the ring $R = (\mathbf{Z}_6, +, \cdot)$, the expression $5x^2 + 3x^1 - 2x^0$ is a polynomial of degree 2, with leading coefficient 5 and constant term $-2x^0$. As before, here we are using a to denote $[a]$ in $\mathbf{Z}_6$. This polynomial may also be written as $5x^2 + 3x^1 + 4x^0$ since $[4] = [-2]$ in $\mathbf{Z}_6$.

b) If z is the zero element of ring R, then the zero polynomial $zx^0 = z$ is also the zero element of $R[x]$ and is said to have *no degree* and no leading coefficient. A polynomial over R that is the zero element or is of degree 0 is called a *constant polynomial*. For example, the polynomial $5x^0$ over $\mathbf{Z}_7$ has degree 0 and leading coefficient 5 and is a constant polynomial.

For a ring of coefficients $(R, +, \cdot)$, let

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x^1 + a_0 x^0$$
$$g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x^1 + b_0 x^0,$$

where $a_i, b_j \in R$ for all $0 \le i \le n$, $0 \le j \le m$. We introduce (closed binary) operations of addition and multiplication for these polynomials in order to obtain a new ring.

Assume that $n \ge m$. We define

$$f(x) + g(x) = \sum_{i=0}^{n} (a_i + b_i)x^i, \tag{1}$$

where $b_i = z$ for $i > m$, and

$$f(x)g(x) = (a_n b_m)x^{n+m} + (a_n b_{m-1} + a_{n-1}b_m)x^{n+m-1} + \cdots + (a_1 b_0 + a_0 b_1)x^1 + (a_0 b_0)x^0. \tag{2}$$

In the definition of $f(x) + g(x)$, the coefficient $(a_i + b_i)$, for each $0 \le i \le n$, is obtained from the addition of elements in R. For $f(x)g(x)$, the coefficient of x^t is $\sum_{k=0}^{t} a_{t-k}b_k$, where all additions and multiplications occur within R, and $0 \le t \le n + m$. Here is one such example to demonstrate the types of calculations that are involved.

Let $f(x) = 4x^3 + 2x^2 + 3x^1 + 1x^0$ and $g(x) = 3x^2 + x^1 + 2x^0$ be polynomials from $\mathbf{Z}_5[x]$. Here

$$a_3 = 4, \qquad a_2 = 2, \qquad a_1 = 3, \qquad a_0 = 1,$$

and

$$b_2 = 3, \qquad b_1 = 1, \qquad b_0 = 2.$$

For all $n \ge 4$ we find that $a_n = 0$. When $m \ge 3$ we have $b_m = 0$. Using the definitions in Eqs. (1) and (2), where the addition and multiplication of the coefficients are now performed modulo 5, we obtain

$$\begin{aligned} f(x) + g(x) &= (4+0)x^3 + (2+3)x^2 + (3+1)x^1 + (1+2)x^0 \\ &= 4x^3 + 0x^2 + 4x^1 + 3x^0 = 4x^3 + 4x^1 + 3x^0 \end{aligned}$$

and

$$\begin{aligned} f(x)g(x) = &\left(\sum_{k=0}^{5} a_{5-k}b_k\right)x^5 + \left(\sum_{k=0}^{4} a_{4-k}b_k\right)x^4 + \left(\sum_{k=0}^{3} a_{3-k}b_k\right)x^3 \\ &+ \left(\sum_{k=0}^{2} a_{2-k}b_k\right)x^2 + \left(\sum_{k=0}^{1} a_{1-k}b_k\right)x^1 + \left(\sum_{k=0}^{0} a_{0-k}b_k\right)x^0 \end{aligned}$$

$$
\begin{aligned}
&= (0 \cdot 2 + 0 \cdot 1 + 4 \cdot 3 + 2 \cdot 0 + 3 \cdot 0 + 1 \cdot 0)x^5 \\
&\quad + (0 \cdot 2 + 4 \cdot 1 + 2 \cdot 3 + 3 \cdot 0 + 1 \cdot 0)x^4 \\
&\quad + (4 \cdot 2 + 2 \cdot 1 + 3 \cdot 3 + 1 \cdot 0)x^3 \\
&\quad + (2 \cdot 2 + 3 \cdot 1 + 1 \cdot 3)x^2 + (3 \cdot 2 + 1 \cdot 1)x^1 + (1 \cdot 2)x^0 \\
&= 2x^5 + 0x^4 + 4x^3 + 0x^2 + 2x^1 + 2x^0 = 2x^5 + 4x^3 + 2x^1 + 2x^0.
\end{aligned}
$$

The closed binary operations defined in Eqs. (1) and (2) were designed to give us the following result.

THEOREM 17.1 If R is a ring, then under the operations of addition and multiplication given in Eqs. (1) and (2), $(R[x], +, \cdot)$ is a ring, called the *polynomial ring*, or *ring of polynomials, over* R.

Proof: The ring properties for $R[x]$ hinge upon those of R. Consequently, we shall prove the associative law of multiplication here, as an example, and shall then leave the proofs of the other properties to the reader. Let $h(x) = \sum_{k=0}^{p} c_k x^k$, with $f(x)$, $g(x)$ as defined earlier. A typical summand in $(f(x)g(x))h(x)$ has the form Ax^t, where $0 \leq t \leq (m+n)+p$ and A is the sum of all products of the form $(a_i b_j)c_k$, with $0 \leq i \leq n$, $0 \leq j \leq m$, $0 \leq k \leq p$, and $i + j + k = t$. In $f(x)(g(x)h(x))$ the coefficient of x^t is the sum of all products of the form $a_i(b_j c_k)$, again with $0 \leq i \leq n$, $0 \leq j \leq m$, $0 \leq k \leq p$, and $i + j + k = t$. Since R is associative under multiplication, $(a_i b_j)c_k = a_i(b_j c_k)$ for each of these terms, and so the coefficient of x^t in $(f(x)g(x))h(x)$ is the same as it is in $f(x)(g(x)h(x))$. Hence $(f(x)g(x))h(x) = f(x)(g(x)h(x))$.

COROLLARY 17.1 Let $R[x]$ be a polynomial ring.

a) If R is commutative, then $R[x]$ is commutative.

b) If R is a ring with unity, then $R[x]$ is a ring with unity.

c) $R[x]$ is an integral domain if and only if R is an integral domain.

Proof: The proof of this corollary is left for the reader.

From this point on, we shall write x instead of x^1. If R has unity u, we define $x^0 = u$, and for all $r \in R$ we write rx^0 as r.

EXAMPLE 17.2 Let $f(x), g(x) \in \mathbf{Z}_8[x]$ with $f(x) = 4x^2 + 1$ and $g(x) = 2x + 3$. Then $f(x)$ has degree 2 and $g(x)$ has degree 1. From our past experiences with polynomials, we expect the degree of $f(x)g(x)$ to be 3, the sum of the degrees of $f(x)$ and $g(x)$. Here, however, $f(x)g(x) = (4x^2 + 1)(2x + 3) = 8x^3 + 12x^2 + 2x + 3 = 4x^2 + 2x + 3$ because $[8] = [0]$ in $\mathbf{Z}_8$. So degree $f(x)g(x) = 2 < 3 =$ degree $f(x) +$ degree $g(x)$.

The cause of the phenomenon in Example 17.2 is the existence of proper divisors of zero in the ring $\mathbf{Z}_8$. This observation leads us to the following theorem.

THEOREM 17.2 Let $(R, +, \cdot)$ be a commutative ring with unity u. Then R is an integral domain if and only if for all $f(x), g(x) \in R[x]$, if neither $f(x)$ nor $g(x)$ is the zero polynomial, then

$$\text{degree } f(x)g(x) = \text{degree } f(x) + \text{degree } g(x).$$

Proof: Let $f(x) = \sum_{i=0}^{n} a_i x^i$, $g(x) = \sum_{j=0}^{m} b_j x^j$, with $a_n \neq z$, $b_m \neq z$. If R is an integral domain, then $a_n b_m \neq z$, so degree $f(x)g(x) = n + m =$ degree $f(x) +$ degree $g(x)$. Conversely, if R is not an integral domain, let $a, b \in R$ with $a \neq z$, $b \neq z$, but $ab = z$. The polynomials $f(x) = ax + u$, $g(x) = bx + u$ each have degree 1, but $f(x)g(x) = (a + b)x + u$ and degree $f(x)g(x) \leq 1 < 2 =$ degree $f(x) +$ degree $g(x)$.

Before we can proceed we need to recall an idea that was introduced in Section 14.2—in Exercise 21. If R is a ring with unity u and $r \in R$, we define $r^0 = u$, $r^1 = r$, and $r^{n+1} = r^n r$ for all $n \in \mathbf{Z}^+$. [From these definitions one can show, for example, that for all $m, n \in \mathbf{Z}^+$, $(r^m)(r^n) = r^{m+n}$ and $(r^m)^n = r^{mn}$.] So now we continue as follows.

Let R be a ring with unity u and let $f(x) = a_n x^n + \cdots + a_1 x + a_0 \in R[x]$. If $r \in R$, then $f(r) = a_n r^n + \cdots + a_1 r + a_0 \in R$. We are especially interested in those values of r for which $f(r) = z$, and this interest leads us to the following concept.

Definition 17.2 Let R be a ring with unity u and let $f(x) \in R[x]$, with degree $f(x) \geq 1$. If $r \in R$ and $f(r) = z$, then r is called a *root* of the polynomial $f(x)$.

EXAMPLE 17.3

a) If $f(x) = x^2 - 2 \in \mathbf{R}[x]$, then $f(x)$ has $\sqrt{2}$ and $-\sqrt{2}$ as roots because $(\sqrt{2})^2 - 2 = 0 = (-\sqrt{2})^2 - 2$. In addition, we can write $f(x) = (x - \sqrt{2})(x + \sqrt{2})$, with $x - \sqrt{2}, x + \sqrt{2} \in \mathbf{R}[x]$. However, if we regard $f(x)$ as an element of $\mathbf{Q}[x]$, then $f(x)$ has no roots because $\sqrt{2}$ and $-\sqrt{2}$ are irrational numbers. Consequently, the existence of roots for a polynomial is dependent on the underlying ring of coefficients.

b) For $f(x) = x^2 + 3x + 2 \in \mathbf{Z}_6[x]$, we find that

$$f(0) = (0)^2 + 3(0) + 2 = 2 \qquad f(3) = (3)^2 + 3(3) + 2 = 20 = 2$$
$$f(1) = (1)^2 + 3(1) + 2 = 6 = 0 \qquad f(4) = (4)^2 + 3(4) + 2 = 30 = 0$$
$$f(2) = (2)^2 + 3(2) + 2 = 12 = 0 \qquad f(5) = (5)^2 + 3(5) + 2 = 42 = 0$$

Consequently, $f(x)$ has four roots: 1, 2, 4, and 5. This is more than we expected. In our prior experiences, a polynomial of degree 2 had at most two roots.

In this chapter we shall be primarily concerned with polynomial rings $F[x]$, where F is a field (and $F[x]$ is an integral domain). Consequently, we shall not dwell any further on situations where degree $f(x)g(x) <$ degree $f(x) +$ degree $g(x)$. In addition, unless it is stated otherwise, we shall denote the zero element of a field by 0 and use 1 to denote its unity.

As a result of Example 17.3(b), we shall now develop the concepts needed to find out when a polynomial of degree n has at most n roots.

Definition 17.3 Let F be a field. For $f(x), g(x) \in F[x]$, where $f(x)$ is not the zero polynomial, we call $f(x)$ a *divisor* (or *factor*) of $g(x)$ if there exists $h(x) \in F[x]$ with $f(x)h(x) = g(x)$. In this situation we also say that $f(x)$ *divides* $g(x)$ and that $g(x)$ is a *multiple* of $f(x)$.

This leads to the *division algorithm* for polynomials. Before proving the general result, however, we shall examine two particular examples.

EXAMPLE 17.4

Early in algebra we were taught how to perform the long division of polynomials with real coefficients. Given two polynomials $f(x)$, $g(x)$ with degree $f(x) \leq$ degree $g(x)$, we organized our work in the form

$$\begin{array}{r} q_1(x) + q_2(x) + \cdots + q_t(x)\ (= q(x)) \\ \hline f(x)\overline{)\,g(x)} \\ \underline{f(x)q_1(x)} \\ g(x) - f(x)q_1(x) \\ \underline{\cdots\cdots\cdots} \\ r(x) \end{array}$$

where we continued to divide until we found either

$$r(x) = 0 \qquad \text{or} \qquad \text{degree } r(x) < \text{degree } f(x).$$

It then followed that $g(x) = q(x)f(x) + r(x)$.

For example, if $f(x) = x - 3$ and $g(x) = 7x^3 - 2x^2 + 5x - 2$, then $f(x)$, $g(x) \in \mathbf{Q}[x]$ (or $\mathbf{R}[x]$, or $\mathbf{C}[x]$), and we find

$$\begin{array}{r} 7x^2 + 19x + 62 \quad (= q(x)) \\ \hline x - 3\overline{)\,7x^3 - 2x^2 + 5x - 2} \\ \underline{7x^3 - 21x^2} \\ 19x^2 + 5x - 2 \\ \underline{19x^2 - 57x} \\ 62x - 2 \\ \underline{62x - 186} \\ 184\ (= r(x)) \end{array}$$

Checking these results, we have

$$q(x)f(x) + r(x) = (7x^2 + 19x + 62)(x - 3) + 184 = 7x^3 - 2x^2 + 5x - 2 = g(x).$$

EXAMPLE 17.5

The technique illustrated in Example 17.4 also applies when the coefficients of our polynomials are taken from a *finite field*.

If $f(x) = 3x^2 + 4x + 2$ and $g(x) = 6x^4 + 4x^3 + 5x^2 + 3x + 1$ are polynomials in $\mathbf{Z}_7[x]$, then the process of long division provides the following calculations:

$$\begin{array}{r} 2x^2 + x + 6 \quad (= q(x)) \\ \hline 3x^2 + 4x + 2\overline{)\,6x^4 + 4x^3 + 5x^2 + 3x + 1} \\ \underline{6x^4 + x^3 + 4x^2} \\ 3x^3 + x^2 + 3x + 1 \\ \underline{3x^3 + 4x^2 + 2x} \\ 4x^2 + x + 1 \\ \underline{4x^2 + 3x + 5} \\ 5x + 3 \quad (= r(x)) \end{array}$$

Performing all arithmetic in $\mathbf{Z}_7$, we find (as in Example 17.4) that

$$\begin{aligned} q(x)f(x) + r(x) &= (2x^2 + x + 6)(3x^2 + 4x + 2) + (5x + 3) \\ &= 6x^4 + 4x^3 + 5x^2 + 3x + 1 = g(x) \end{aligned}$$

We turn now to the general situation.

THEOREM 17.3

Division Algorithm. Let $f(x), g(x) \in F[x]$ with $f(x)$ not the zero polynomial. There exist unique polynomials $q(x), r(x) \in F[x]$ such that $g(x) = q(x)f(x) + r(x)$, where $r(x) = 0$ or degree $r(x) <$ degree $f(x)$.

Proof: Let $S = \{g(x) - t(x)f(x) | t(x) \in F[x]\}$.

If $0 \in S$, then $0 = g(x) - t(x)f(x)$ for some $t(x) \in F[x]$. Then with $q(x) = t(x)$ and $r(x) = 0$, we have $g(x) = q(x)f(x) + r(x)$.

If $0 \notin S$, consider the degrees of the elements of S, and let $r(x) = g(x) - q(x)f(x)$ be an element in S of minimum degree. Since $r(x) \neq 0$, the result follows if degree $r(x)$ $<$ degree $f(x)$. If not, let

$$\begin{aligned} r(x) &= a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0, \qquad a_n \neq 0, \\ f(x) &= b_m x^m + b_{m-1}x^{m-1} + \cdots + b_2 x^2 + b_1 x + b_0, \qquad b_m \neq 0, \end{aligned}$$

with $n \geq m$. Define

$$\begin{aligned} h(x) = r(x) - [a_n b_m^{-1} x^{n-m}]f(x) &= (a_n - a_n b_m^{-1} b_m)x^n + (a_{n-1} - a_n b_m^{-1} b_{m-1})x^{n-1} \\ &+ \cdots + (a_{n-m} - a_n b_m^{-1} b_0)x^{n-m} + a_{n-m-1}x^{n-m-1} + \cdots + a_1 x + a_0. \end{aligned}$$

Then $h(x)$ has degree less than n, the degree of $r(x)$. More important, $h(x) = [g(x) - q(x)f(x)] - [a_n b_m^{-1} x^{n-m}]f(x) = g(x) - [q(x) + a_n b_m^{-1} x^{n-m}]f(x)$, so $h(x) \in S$ and this contradicts the choice of $r(x)$ as having minimum degree. Consequently, degree $r(x) <$ degree $f(x)$ and we have the existence part of the theorem.

For uniqueness, let $g(x) = q_1(x)f(x) + r_1(x) = q_2(x)f(x) + r_2(x)$ where $r_1(x) = 0$ or degree $r_1(x) <$ degree $f(x)$, and $r_2(x) = 0$ or degree $r_2(x) <$ degree $f(x)$. Then $[q_2(x) - q_1(x)]f(x) = r_1(x) - r_2(x)$, and if $q_2(x) - q_1(x) \neq 0$, then degree $([q_2(x) - q_1(x)]f(x)) \geq$ degree $f(x)$, whereas $r_1(x) - r_2(x) = 0$ or degree $[r_1(x) - r_2(x)] \leq \max\{\text{degree } r_1(x), \text{ degree } r_2(x)\} <$ degree $f(x)$. Consequently, $q_1(x) = q_2(x)$, and $r_1(x) = r_2(x)$.

The division algorithm provides the following results on roots and factors.

THEOREM 17.4

The Remainder Theorem. For $f(x) \in F[x]$ and $a \in F$, the remainder in the division of $f(x)$ by $x - a$ is $f(a)$.

Proof: From the division algorithm, $f(x) = q(x)(x - a) + r(x)$, with $r(x) = 0$ or degree $r(x) <$ degree $(x - a) = 1$. Hence $r(x) = r$ is an element of F. Substituting a for x, we find $f(a) = q(a)(a - a) + r = 0 + r = r$.

THEOREM 17.5

The Factor Theorem. If $f(x) \in F[x]$ and $a \in F$, then $x - a$ is a factor of $f(x)$ if and only if a is a root of $f(x)$.

Proof: If $x - a$ is a factor of $f(x)$, then $f(x) = q(x)(x - a)$. With $f(a) = q(a)(a - a) = 0$, it follows that a is a root of $f(x)$. Conversely, suppose that a is a root of $f(x)$. By the

division algorithm, $f(x) = q(x)(x - a) + r$, where $r \in F$. Since $f(a) = 0$ we have $r = 0$, so $f(x) = q(x)(x - a)$, and $x - a$ is a factor of $f(x)$.

EXAMPLE 17.6

a) Let $f(x) = x^7 - 6x^5 + 4x^4 - x^2 + 3x - 7 \in \mathbf{Q}[x]$. From the remainder theorem it follows that when $f(x)$ is divided by $x - 2$, the remainder is

$$f(2) = 2^7 - 6(2^5) + 4(2^4) - 2^2 + 3(2) - 7 = -5.$$

If we were to divide $f(x)$ by $x + 1$, then the remainder would be $f(-1) = -2$.

b) If $g(x) = x^5 + 3x^4 + x^3 + x^2 + 2x + 2 \in \mathbf{Z}_5[x]$ is divided by $x - 1$, then the remainder here is $g(1) = 1 + 3 + 1 + 1 + 2 + 2 = 0$ (in $\mathbf{Z}_5$). Consequently, $x - 1$ divides $g(x)$, and by the factor theorem,

$$g(x) = q(x)(x - 1) \qquad \text{(where degree } q(x) = 4\text{)}.$$

Using the results of Theorems 17.4 and 17.5, we now establish the last major idea for this section.

THEOREM 17.6

If $f(x) \in F[x]$ has degree $n \geq 1$, then $f(x)$ has at most n roots in F.

Proof: The proof is by mathematical induction on the degree of $f(x)$. If $f(x)$ has degree 1, then $f(x) = ax + b$, for $a, b \in F$, $a \neq 0$. With $f(-a^{-1}b) = 0$, $f(x)$ has at least one root in F. If c_1 and c_2 are both roots, then $f(c_1) = ac_1 + b = 0 = ac_2 + b = f(c_2)$. By cancellation in a ring, $ac_1 + b = ac_2 + b \Rightarrow ac_1 = ac_2$. Since F is a field and $a \neq 0$, we have $ac_1 = ac_2 \Rightarrow c_1 = c_2$, so $f(x)$ has only one root in F.

Now assume the result of the theorem is true for all polynomials of degree k (≥ 1) in $F[x]$. Consider a polynomial $f(x)$ of degree $k + 1$. If $f(x)$ has no roots in F, the theorem follows. Otherwise, let $r \in F$ with $f(r) = 0$. By the factor theorem, $f(x) = (x - r)g(x)$ where $g(x)$ has degree k. Consequently, by the induction hypothesis, $g(x)$ has at most k roots in F, and $f(x)$, in turn, has at most $k + 1$ roots in F.

EXAMPLE 17.7

a) Let $f(x) = x^2 - 6x + 9 \in \mathbf{R}[x]$. Then $f(x)$ has at most two roots in $\mathbf{R}$—namely, the roots 3, 3. So here we say that 3 is a root of *multiplicity* 2. In addition $f(x) = (x - 3)(x - 3)$, a factorization into two first-degree, or *linear*, factors.

b) For $g(x) = x^2 + 4 \in \mathbf{R}[x]$, $g(x)$ has no real roots, but Theorem 17.6 is not contradicted. (Why?) In $\mathbf{C}[x]$, $g(x)$ has the roots $2i$, $-2i$ and can be factored as $g(x) = (x - 2i)(x + 2i)$.

c) If $h(x) = x^2 + 2x + 6 \in \mathbf{Z}_7[x]$, then $h(2) = 0$, $h(3) = 0$ and these are the only roots. Also, $h(x) = (x - 2)(x - 3) = x^2 - 5x + 6 = x^2 + 2x + 6$, because $[-5] = [2]$ in $\mathbf{Z}_7$.

d) As we saw in Example 17.3(b), the polynomial $x^2 + 3x + 2$ has four roots. This is not a contradiction to Theorem 17.6 because $\mathbf{Z}_6$ is not a field. Also, $x^2 + 3x + 2 = (x + 1)(x + 2) = (x + 4)(x + 5)$, two distinct factorizations.

We close with one final remark, without proof, on the idea of factorization in $F[x]$. If $f(x) \in F[x]$ has degree n, and $r_1, r_2, \ldots, r_n$ are the roots of $f(x)$ in F (where it is

possible for a root to be repeated—that is, $r_i = r_j$ for some $1 \le i < j \le n$), then $f(x) = a_n(x - r_1)(x - r_2) \cdots (x - r_n)$, where a_n is the leading coefficient of $f(x)$. This representation of $f(x)$ is unique up to the order of the first-degree factors.

EXERCISES 17.1

1. Let $f(x), g(x) \in \mathbf{Z}_7[x]$ where $f(x) = 2x^4 + 2x^3 + 3x^2 + x + 4$ and $g(x) = 3x^3 + 5x^2 + 6x + 1$. Determine $f(x) + g(x)$, $f(x) - g(x)$, and $f(x)g(x)$.

2. Determine all of the polynomials of degree 2 in $\mathbf{Z}_2[x]$.

3. How many polynomials are there of degree 2 in $\mathbf{Z}_{11}[x]$? How many have degree 3? degree 4? degree n, for $n \in \mathbf{N}$?

4. a) Find two nonzero polynomials $f(x), g(x)$ in $\mathbf{Z}_{12}[x]$ where $f(x)g(x) = 0$.

b) Find polynomials $h(x), k(x) \in \mathbf{Z}_{12}[x]$ such that degree $h(x) = 5$, degree $k(x) = 2$, and degree $h(x)k(x) = 3$.

5. Complete the proofs of Theorem 17.1 and Corollary 17.1.

6. For each of the following pairs $f(x)$, $g(x)$, find $q(x)$, $r(x)$ so that $g(x) = q(x)f(x) + r(x)$, where $r(x) = 0$ or degree $r(x) <$ degree $f(x)$.

a) $f(x), g(x) \in \mathbf{Q}[x]$, $f(x) = x^4 - 5x^3 + 7x$, $g(x) = x^5 - 2x^2 + 5x - 3$

b) $f(x), g(x) \in \mathbf{Z}_2[x]$, $f(x) = x^2 + 1, g(x) = x^4 + x^3 + x^2 + x + 1$

c) $f(x), g(x) \in \mathbf{Z}_5[x]$, $f(x) = x^2 + 3x + 1, g(x) = x^4 + 2x^3 + x + 4$

7. a) If $f(x) = x^4 - 16$, find its roots and factorization in $\mathbf{Q}[x]$.

b) Answer part (a) for $f(x) \in \mathbf{R}[x]$.

c) Answer part (a) for $f(x) \in \mathbf{C}[x]$.

d) Answer parts (a), (b), and (c) for $f(x) = x^4 - 25$.

8. a) Find all roots of $f(x) = x^2 + 4x$ if $f(x) \in \mathbf{Z}_{12}[x]$.

b) Find four distinct linear polynomials $g(x), h(x), s(x), t(x) \in \mathbf{Z}_{12}[x]$ so that $f(x) = g(x)h(x) = s(x)t(x)$.

c) Do the results in part (b) contradict the statements made in the paragraph following Example 17.7?

9. In each of the following, find the remainder when $f(x)$ is divided by $g(x)$.

a) $f(x), g(x) \in \mathbf{Q}[x]$, $f(x) = x^8 + 7x^5 - 4x^4 + 3x^3 + 5x^2 - 4, g(x) = x - 3$

b) $f(x), g(x) \in \mathbf{Z}_2[x]$, $f(x) = x^{100} + x^{90} + x^{80} + x^{50} + 1, g(x) = x - 1$

c) $f(x), g(x) \in \mathbf{Z}_{11}[x]$, $f(x) = 3x^5 - 8x^4 + x^3 - x^2 + 4x - 7, g(x) = x + 9$

10. For each of the following polynomials $f(x) \in \mathbf{Z}_7[x]$, determine all of the roots in $\mathbf{Z}_7$ and then write $f(x)$ as a product of first-degree polynomials.

a) $f(x) = x^3 + 5x^2 + 2x + 6$

b) $f(x) = x^7 - x$

11. How many units are there in the ring $\mathbf{Z}_5[x]$? How many in $\mathbf{Z}_7[x]$? How many in $\mathbf{Z}_p[x]$, p a prime?

12. Given a field F, let $f(x) \in F[x]$ where $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0$. Prove that $x - 1$ is a factor of $f(x)$ if and only if

$$a_n + a_{n-1} + \cdots + a_2 + a_1 + a_0 = 0.$$

13. Let R, S be rings, and let $g: R \to S$ be a ring homomorphism. Prove that the function $G: R[x] \to S[x]$ defined by

$$G\left(\sum_{i=0}^{n} r_i x^i\right) = \sum_{i=0}^{n} g(r_i)x^i$$

is a ring homomorphism.

14. If R is an integral domain, prove that if $f(x)$ is a unit in $R[x]$, then $f(x)$ is a constant and is a unit in R.

15. Verify that $f(x) = 2x + 1$ is a unit in $\mathbf{Z}_4[x]$. Does this contradict the result of Exercise 14?

16. For $n \in \mathbf{Z}^+$, $n \ge 2$, let $f(x) \in \mathbf{Z}_n[x]$. Prove that if $a, b \in \mathbf{Z}$ and $a \equiv b \pmod{n}$, then $f(a) \equiv f(b) \pmod{n}$.

17. If F is a field, let $S \subseteq F[x]$ where $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0 \in S$ if and only if $a_n + a_{n-1} + \cdots + a_2 + a_1 + a_0 = 0$. Prove that S is an ideal of $F[x]$.

18. Let $(R, +, \cdot)$ be a ring. If I is an ideal of R, prove that $I[x]$, the set of all polynomials in the indeterminate x with coefficients in I, is an ideal in $R[x]$.

17.2 Irreducible Polynomials: Finite Fields

We now wish to construct finite fields other than those of the type $(\mathbf{Z}_p, +, \cdot)$, where p is a prime. The construction will use the following special polynomials.

Definition 17.4 Let $f(x) \in F[x]$, with F a field and degree $f(x) \geq 2$. We call $f(x)$ *reducible* (over F) if there exist $g(x), h(x) \in F[x]$, where $f(x) = g(x)h(x)$ and each of $g(x), h(x)$ has degree ≥ 1. If $f(x)$ is not reducible it is called *irreducible*, or *prime*.

Theorem 17.7 contains some useful observations about irreducible polynomials.

THEOREM 17.7 For polynomials in $F[x]$,

a) every nonzero polynomial of degree ≤ 1 is irreducible.

b) if $f(x) \in F[x]$ with degree $f(x) = 2$ or 3, then $f(x)$ is reducible if and only if $f(x)$ has a root in the field F.

Proof: The proof is left for the reader.

EXAMPLE 17.8

a) The polynomial $x^2 + 1$ is irreducible in $\mathbf{Q}[x]$ and $\mathbf{R}[x]$, but in $\mathbf{C}[x]$ we find $x^2 + 1 = (x + i)(x - i)$.

b) Let $f(x) = x^4 + 2x^2 + 1 \in \mathbf{R}[x]$. Although $f(x)$ has no real roots, it is reducible because $(x^2 + 1)^2 = x^4 + 2x^2 + 1$. Hence part (b) of Theorem 17.7 is not applicable for polynomials of degree > 3.

c) In $\mathbf{Z}_2[x]$, $f(x) = x^3 + x^2 + x + 1$ is reducible because $f(1) = 0$. But $g(x) = x^3 + x + 1$ is irreducible because $g(0) = g(1) = 1$.

d) Let $h(x) = x^4 + x^3 + x^2 + x + 1 \in \mathbf{Z}_2[x]$. Is $h(x)$ reducible in $\mathbf{Z}_2[x]$? Since $h(0) = h(1) = 1$, $h(x)$ has no first-degree factors, but perhaps we can find $a, b, c, d \in \mathbf{Z}_2$ such that $(x^2 + ax + b)(x^2 + cx + d) = x^4 + x^3 + x^2 + x + 1$.

By expanding $(x^2 + ax + b)(x^2 + cx + d)$ and comparing coefficients of like powers of x, we find $a + c = 1$, $ac + b + d = 1$, $ad + bc = 1$, and $bd = 1$. With $bd = 1$, it follows that $b = 1$ and $d = 1$, so $ac + b + d = 1 \Rightarrow ac = 1 \Rightarrow a = c = 1 \Rightarrow a + c = 0$. This contradicts $a + c = 1$. Consequently, $h(x)$ is irreducible in $\mathbf{Z}_2[x]$.

All of the polynomials in Example 17.8 share a common property, which we shall now define.

Definition 17.5 A polynomial $f(x) \in F[x]$ is called *monic* if its leading coefficient is 1, the unity of F.

Some of our next results (up to and including the discussion in Example 17.11) awaken memories of Chapters 4 and 14.

Definition 17.6 If $f(x), g(x) \in F[x]$, then $h(x) \in F[x]$ is *a greatest common divisor* of $f(x)$ and $g(x)$

a) if $h(x)$ divides each of $f(x)$ and $g(x)$, and

b) if $k(x) \in F[x]$ and $k(x)$ divides both $f(x), g(x)$, then $k(x)$ divides $h(x)$.

We now state the following results on the existence and uniqueness of what we shall call *the* greatest common divisor, which we shall abbreviate as gcd. Furthermore, there is a method for finding this gcd that is called the Euclidean algorithm for polynomials. A proof for the first result is outlined in the Section Exercises.

THEOREM 17.8

Let $f(x), g(x) \in F[x]$, with at least one of $f(x), g(x)$ not the zero polynomial. Then each polynomial of minimum degree that can be written as a linear combination of $f(x)$ and $g(x)$ — that is, in the form $s(x)f(x) + t(x)g(x)$, for $s(x), t(x) \in F[x]$ — will be a greatest common divisor of $f(x), g(x)$. If we require a gcd to be monic, then it will be unique.

THEOREM 17.9

Euclidean Algorithm for Polynomials. Let $f(x), g(x) \in F[x]$ with degree $f(x) \leq$ degree $g(x)$ and $f(x) \neq 0$. Applying the division algorithm, we write

$$\begin{aligned}
g(x) &= q(x)f(x) + r(x), && \text{degree } r(x) < \text{degree } f(x)\\
f(x) &= q_1(x)r(x) + r_1(x), && \text{degree } r_1(x) < \text{degree } r(x)\\
r(x) &= q_2(x)r_1(x) + r_2(x), && \text{degree } r_2(x) < \text{degree } r_1(x)\\
&\vdots && \vdots\\
r_{k-2}(x) &= q_k(x)r_{k-1}(x) + r_k(x), && \text{degree } r_k(x) < \text{degree } r_{k-1}(x)\\
r_{k-1}(x) &= q_{k+1}(x)r_k(x) + r_{k+1}(x), && r_{k+1}(x) = 0.
\end{aligned}$$

Then $r_k(x)$, the last nonzero remainder, is a greatest common divisor of $f(x), g(x)$, and is a constant multiple of the monic greatest common divisor of $f(x), g(x)$. [Multiplying $r_k(x)$ by the inverse of its leading coefficient allows us to obtain the unique monic polynomial we call *the* greatest common divisor.]

Definition 17.7

If $f(x), g(x) \in F[x]$ and their gcd is 1, then $f(x)$ and $g(x)$ are called *relatively prime.*

The last results we need to construct our new finite fields provide the analog of a construction we developed in Section 14.3.

THEOREM 17.10

Let $s(x) \in F(x)$, $s(x) \neq 0$. Define relation $\mathcal{R}$ on $F[x]$ by $f(x)\,\mathcal{R}\,g(x)$ if $f(x) - g(x) = t(x)s(x)$, for some $t(x) \in F[x]$ — that is, $s(x)$ divides $f(x) - g(x)$. Then $\mathcal{R}$ is an equivalence relation on $F[x]$.

Proof: The verification of the reflexive, symmetric, and transitive properties of $\mathcal{R}$ is left for the reader.

When the situation in Theorem 17.10 occurs, we say that $f(x)$ is *congruent* to $g(x)$ *modulo* $s(x)$ and write $f(x) \equiv g(x) \pmod{s(x)}$. The relation $\mathcal{R}$ is referred to as *congruence modulo* $s(x)$.

Let us examine the equivalence classes for one such relation.

EXAMPLE 17.9

Let $s(x) = x^2 + x + 1 \in \mathbf{Z}_2[x]$. Then

a) $[0] = [x^2 + x + 1] = \{0, x^2 + x + 1, x^3 + x^2 + x, (x + 1)(x^2 + x + 1), \ldots\}$
$= \{t(x)(x^2 + x + 1) | t(x) \in \mathbf{Z}_2[x]\}$

b) $[1] = \{1, x^2 + x, x(x^2 + x + 1) + 1, (x + 1)(x^2 + x + 1) + 1, \ldots\}$
$= \{t(x)(x^2 + x + 1) + 1 | t(x) \in \mathbf{Z}_2[x]\}$

c) $[x] = \{x, x^2 + 1, x(x^2 + x + 1) + x, (x + 1)(x^2 + x + 1) + x, \ldots\}$
$= \{t(x)(x^2 + x + 1) + x | t(x) \in \mathbf{Z}_2[x]\}$

d) $[x + 1] = \{x + 1, x^2, x(x^2 + x + 1) + (x + 1), (x + 1)(x^2 + x + 1)$
$+ (x + 1), \ldots\} = \{t(x)(x^2 + x + 1) + (x + 1) | t(x) \in \mathbf{Z}_2[x]\}$

Are these all of the equivalence classes? If $f(x) \in \mathbf{Z}_2[x]$, then by the division algorithm $f(x) = q(x)s(x) + r(x)$, where $r(x) = 0$ or degree $r(x) <$ degree $s(x)$. Since $f(x) - r(x) = q(x)s(x)$, it follows that $f(x) \equiv r(x) \pmod{s(x)}$, so $f(x) \in [r(x)]$. Consequently, to determine all the equivalence classes, we consider the possibilities for $r(x)$. Here $r(x) = 0$ or degree $r(x) < 2$, so $r(x) = ax + b$, where $a, b \in \mathbf{Z}_2$. With only two choices for each of a, b, there are four possible choices for $r(x)$: 0, 1, x, $x + 1$.

We now place a ring structure on the equivalence classes of Example 17.9. Recalling how this was accomplished in Chapter 14 for $\mathbf{Z}_n$, we define addition by $[f(x)] + [g(x)] = [f(x) + g(x)]$. Since degree $(f(x) + g(x)) \leq \max\{\text{degree } f(x), \text{degree } g(x)\}$, we can find the equivalence class for $[f(x) + g(x)]$ without too much trouble. Here, for example, $[x] + [x + 1] = [x + (x + 1)] = [2x + 1] = [1]$ because $2 = 0$ in $\mathbf{Z}_2$.

In defining the multiplication of these equivalence classes, we run into a little more difficulty. For instance, what is $[x][x]$ in Example 17.9? If, in general, we define $[f(x)][g(x)] = [f(x)g(x)]$, it is possible that degree $f(x)g(x) \geq$ degree $s(x)$, so we may not readily find $[f(x)g(x)]$ in the list of equivalence classes. However, if degree $f(x)g(x) \geq$ degree $s(x)$, then using the division algorithm, we can write $f(x)g(x) = q(x)s(x) + r(x)$, where $r(x) = 0$ or degree $r(x) <$ degree $s(x)$. With $f(x)g(x) = q(x)s(x) + r(x)$, it follows that $f(x)g(x) \equiv r(x) \pmod{s(x)}$, and we define $[f(x)g(x)] = [r(x)]$, where $[r(x)]$ *does* occur in the list of equivalence classes.

From these observations we construct Tables 17.1 and 17.2 for the addition and multiplication, respectively, of $\{[0], [1], [x], [x + 1]\}$. (In these tables we write a for $[a]$.)

Table 17.1

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

Table 17.2

$\cdot$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	$x+1$	1
$x+1$	0	$x+1$	1	x

From the multiplication table (Table 17.2), we find that these equivalence classes form not only a ring but also a field, where $[1]^{-1} = [1]$, $[x]^{-1} = [x + 1]$, and $[x + 1]^{-1} = [x]$. This field of order 4 is denoted by $\mathbf{Z}_2[x]/(x^2 + x + 1)$, and we observe that it contains (an isomorphic copy of) the *subfield* $\mathbf{Z}_2$. [In general, a subring $(R, +, \cdot)$ of a field $(F, +, \cdot)$ is called a subfield when $(R, +, \cdot)$ is a field.] In addition, for the nonzero elements of this field we find that $[x]^1 = [x]$, $[x]^2 = [x + 1]$, $[x]^3 = [1]$, so we have a cyclic group of order 3. But the nonzero elements of any field form a group under multiplication, and any group of order 3 is cyclic, so why bother with this observation? In general, the nonzero elements of *any* finite field form a cyclic group under multiplication. (A proof for this can be found in Chapter 12 of reference [10].)

The preceding construction is summarized in the following theorem. An outline of the proof is given in the Section Exercises.

THEOREM 17.11 Let $s(x)$ be a nonzero polynomial in $F[x]$.

a) The equivalence classes of $F[x]$ for the relation of congruence modulo $s(x)$ form a commutative ring with unity under the closed binary operations

$$[f(x)] + [g(x)] = [f(x) + g(x)], \qquad [f(x)][g(x)] = [f(x)g(x)] = [r(x)],$$

where $r(x)$ is the remainder obtained upon dividing $f(x)g(x)$ by $s(x)$. This ring is denoted by $F[x]/(s(x))$.

b) If $s(x)$ is irreducible in $F[x]$, then $F[x]/(s(x))$ is a field.

c) If $|F| = q$ and degree $s(x) = n$, then $F[x]/(s(x))$ contains q^n elements.

Before we continue we wish to emphasize that for $s(x)$ irreducible in $F[x]$ the elements in the field $F[x]/(s(x))$ are *not* simply polynomials (in x). But how can this be, considering the presence of the symbol x in each of the elements $[x]$ and $[x + 1]$ in the field $\mathbf{Z}_2[x]/(x^2 + x + 1)$ of Example 17.9? In order to make our point more apparent we consider an infinite example that is somewhat familiar to us.

EXAMPLE 17.10 Here we let $F = (\mathbf{R}, +, \cdot)$, the field of real numbers, and we consider the irreducible polynomial $s(x) = x^2 + 1$ in $\mathbf{R}[x]$. From part (b) of Theorem 17.11 we learn that $\mathbf{R}[x]/(s(x)) = \mathbf{R}[x]/(x^2 + 1)$ is a field.

For all $f(x) \in \mathbf{R}[x]$ it follows by the division algorithm that

$$f(x) = q(x)(x^2 + 1) + r(x), \quad \text{where } r(x) = 0 \text{ or } 0 \leq \deg r(x) \leq 1.$$

Therefore,

$$\mathbf{R}[x]/(x^2 + 1) = \{[a + bx] | a, b \in \mathbf{R}\},$$

where it can be shown that $[a + bx] = [a] + [bx] = [a] + [b][x]$.

Among the (infinitely many) elements of $\mathbf{R}[x]/(x^2 + 1)$ are the following:

1) $[1] = \{1 + t(x)(x^2 + 1) | t(x) \in \mathbf{R}[x]\}$, where we find the elements $x^2 + 2$ and $3x^3 + 3x + 1$ (from $\mathbf{R}[x]$);

2) $[r] = \{r + t(x)(x^2 + 1) | t(x) \in \mathbf{R}[x]\}$, where r is any (fixed) real number;

3) $[-1] = \{-1 + t(x)(x^2 + 1) | t(x) \in \mathbf{R}[x]\}$, where we find the polynomial $-1 + (1)(x^2 + 1) = x^2$—so, $[x][x] = [x^2] = [-1]$; and

4) $[\sqrt{2}x - 3] = \{(\sqrt{2}x - 3) + t(x)(x^2 + 1) | t(x) \in \mathbf{R}[x]\}$.

Now let us consider the field $(\mathbf{C}, +, \cdot)$ of complex numbers and the correspondence

$$h: \mathbf{R}[x]/(x^2 + 1) \to \mathbf{C},$$

where $h([a + bx]) = a + bi$.

For all $[a + bx], [c + dx] \in \mathbf{R}[x]/(x^2 + 1)$, we have $[a + bx] = [c + dx] \iff (a + bx) - (c + dx) = t(x)(x^2 + 1)$, for some $t(x) \in \mathbf{R}[x] \iff (a - c) + (b - d)x = t(x)(x^2 + 1)$. If $t(x)$ is not the zero polynomial, then we have $(a - c) + (b - d)x$, a polynomial of degree less than 2, equal to $t(x)(x^2 + 1)$, a polynomial of degree at least 2. Consequently, $t(x) = 0$, so $a + bx = c + dx$ and $a = c$, $b = d$. This guarantees that the

correspondence given by h is actually a function. In fact, h is an isomorphism of fields. (See Exercise 24 in the exercises at the end of this section.) To establish that h preserves the operation of multiplication, for example, we observe that

$$\begin{aligned} h([a+bx][c+dx]) &= h([ac+adx+bcx+bdx^2]) \\ &= h([ac+(ad+bc)x]+[bd][x^2]) \\ &= h([ac+(ad+bc)x]+[bd][-1]) \\ &= h([ac-bd)+(ad+bc)x]) \\ &= (ac-bd)+(ad+bc)i = (a+bi)(c+di) \\ &= h([a+bx])h([c+dx]). \end{aligned}$$

Since $\mathbf{R}[x]/(x^2+1)$ is isomorphic to $\mathbf{C}$, the correspondence $h([x]) = i$ makes us think of $[x]$ as a *number* in $\mathbf{R}[x]/(x^2+1)$ and not as a polynomial in x (in $\mathbf{R}[x]$). The number $[x]$ represents an equivalence class of polynomials in $\mathbf{R}[x]$, and this number $[x]$ behaves like the complex number i in the field $(\mathbf{C}, +, \cdot)$. We should also note that for each real number r, $h([r]) = r$, and $\{[r] | r \in \mathbf{R}\}$ is a subfield of $\mathbf{R}[x]/(x^2+1)$, which is isomorphic to the subfield $\mathbf{R}$ of $\mathbf{C}$.

Finally, if we identify the field $\mathbf{R}[x]/(x^2+1)$ with the field $(\mathbf{C}, +, \cdot)$, we can summarize what has happened above as follows: We started with the irreducible polynomial $s(x) = x^2+1$ in $\mathbf{R}[x]$, which had no root in the field $(\mathbf{R}, +, \cdot)$. We then enlarged $(\mathbf{R}, +, \cdot)$ to $(\mathbf{C}, +, \cdot)$ and in $\mathbf{C}$ we found the root i (and the root $-i$) for $s(x)$, which can now be factored as $(x+i)(x-i)$ in $\mathbf{C}[x]$.

Since our major concern in the chapter is with finite fields, we now examine another example of a finite field that arises by virtue of Theorem 17.11.

EXAMPLE 17.11

In $\mathbf{Z}_3[x]$ the polynomial $s(x) = x^2+x+2$ is irreducible because $s(0) = 2$, $s(1) = 1$, and $s(2) = 2$. Consequently, $\mathbf{Z}_3[x]/(s(x))$ is a field containing all equivalence classes of the form $[ax+b]$, where $a, b \in \mathbf{Z}_3$. These arise from the possible remainders when a polynomial $f(x) \in \mathbf{Z}_3[x]$ is divided by $s(x)$. The nine equivalence classes are $[0]$, $[1]$, $[2]$, $[x]$, $[x+1]$, $[x+2]$, $[2x]$, $[2x+1]$, and $[2x+2]$.

Instead of constructing a complete multiplication table, we examine four sample multiplications and then make two observations.

a) $[2x][x] = [2x^2] = [2x^2+0] = [2x^2+(x^2+x+2)] = [3x^2+x+2] = [x+2]$ because $3 = 0$ in $\mathbf{Z}_3$.

b) $[x+1][x+2] = [x^2+3x+2] = [x^2+2] = [x^2+2+2(x^2+x+2)] = [2x]$.

c) $[2x+2]^2 = [4x^2+8x+4] = [x^2+2x+1] = [(-x-2)+(2x+1)]$ since $x^2 \equiv (-x-2) \pmod{s(x)}$. Consequently, $[2x+2]^2 = [x-1] = [x+2]$.

d) Often we write the equivalence classes without brackets and concentrate on the coefficients of the powers of x. For example, 11 is written for $[x+1]$ and 21 represents $[2x+1]$. Consequently, $(21)\cdot(12) = [2x+1][x+2] = [2x^2+5x+2] = [2x^2+2x+2] = [2(-x-2)+2x+2] = [-4+2] = [-2] = [1]$, so $(21)^{-1} = (12)$.

e) We also observe that

$$\begin{array}{llll} [x]^1 = [x] & [x]^3 = [2x+2] & [x]^5 = [2x] & [x]^7 = [x+1] \\ [x]^2 = [2x+1] & [x]^4 = [2] & [x]^6 = [x+2] & [x]^8 = [1] \end{array}$$

Therefore the nonzero elements of $\mathbf{Z}_3[x]/(s(x))$ form a cyclic group under multiplication.

f) Finally, when we consider the equivalence classes [0], [1], [2], we realize that they provide us with a subfield of $\mathbf{Z}_3[x]/(s(x))$—a subfield we identify with the field $(\mathbf{Z}_3, +, \cdot)$.

In Example 17.9 (and in the discussion that follows it) and in Example 17.11, we constructed finite fields of orders 4 $(= 2^2)$ and 9 $(= 3^2)$, respectively. Now we shall close this section as we investigate other possibilities for the order of a finite field. To accomplish this we need the following idea.

Definition 17.8

Let $(R, +, \cdot)$ be a ring. If there is a least positive integer n such that $nr = z$ (the zero of R) for all $r \in R$, then we say that R has *characteristic* n and write $\text{char}(R) = n$. When no such integer exists, R is said to have *characteristic* 0.

EXAMPLE 17.12

a) The ring $(\mathbf{Z}_3, +, \cdot)$ has characteristic 3; $(\mathbf{Z}_4, +, \cdot)$ has characteristic 4; in general, $(\mathbf{Z}_n, +, \cdot)$ has characteristic n.

b) The rings $(\mathbf{Z}, +, \cdot)$ and $(\mathbf{Q}, +, \cdot)$ both have characteristic 0.

c) A ring can be infinite and still have positive characteristic. For example, $\mathbf{Z}_3[x]$ is an infinite ring but it has characteristic 3.

d) The ring in Example 17.9 has characteristic 2. In Example 17.11 the characteristic of the ring is 3. Unlike the examples in part (a), the order of a finite ring can be different from its characteristic.

Examples 17.9 and 17.11, however, are more than just rings. They are fields with prime characteristic. Could this property be true for all finite fields?

THEOREM 17.12

Let $(F, +, \cdot)$ be a field. If $\text{char}(F) > 0$, then $\text{char}(F)$ must be prime.

Proof: In this proof we write the unity of F as u so that it is distinct from the positive integer 1. Let $\text{char}(F) = n > 0$. If n is not prime, we write $n = mk$, where $m, k \in \mathbf{Z}^+$ and $1 < m < n$, $1 < k < n$. By the definition of characteristic, $nu = z$, the zero of F. Hence $(mk)u = z$. But

$$(mk)(u) = \underbrace{(u + u + \cdots + u)}_{mk \text{ summands}} = \underbrace{(u + u + \cdots + u)}_{m \text{ summands}} \underbrace{(u + u + \cdots + u)}_{k \text{ summands}} = (mu)(ku).$$

With F a field, $(mu)(ku) = z \Rightarrow (mu) = z$ or $(ku) = z$. Assume without loss of generality that $ku = z$. Then for each $r \in F$, $kr = k(ur) = (ku)r = zr = z$, contradicting the choice of n as the characteristic of F. Consequently, $\text{char}(F)$ is prime.

(The proof of Theorem 17.12 actually requires that F only be an integral domain.)

If F is a finite field and $m = |F|$, then $ma = z$ for all $a \in F$ because $(F, +)$ is an additive group of order m. (See Exercise 8 of Section 16.3.) Consequently, F has positive characteristic and by Theorem 17.12 this characteristic is prime. This leads us to the following theorem.

THEOREM 17.13 A finite field F has order p^t, where p is a prime and $t \in \mathbf{Z}^+$.

Proof: Since F is a finite field, let $\text{char}(F) = p$, a prime, and let u denote the unity and z the zero element. Then $S_0 = \{u, 2u, 3u, \ldots, pu = z\}$ is a set of p distinct elements in F. If not, $mu = nu$ for $1 \leq m < n \leq p$ and $(n - m)u = z$, with $0 < n - m < p$. So for all $x \in F$ we now find that $(n - m)x = (n - m)(ux) = [(n - m)u]x = zx = z$, and this contradicts $\text{char}(F) = p$. If $F = S_0$, then $|F| = p^1$ and the result follows. If not, let $a \in F - S_0$. Then $S_1 = \{ma + nu | 0 < m, n \leq p\}$ is a subset of F with $|S_1| \leq p^2$. If $|S_1| < p^2$, then $m_1 a + n_1 u = m_2 a + n_2 u$, with $0 < m_1, m_2, n_1, n_2 \leq p$ and at least one of $m_1 - m_2, n_2 - n_1 \neq 0$. Should $m_1 - m_2 = 0$, then $(m_1 - m_2)a = z = (n_2 - n_1)u$, with $0 < |n_2 - n_1| < p$. Consequently, for all $x \in F$, $|n_2 - n_1|x = |n_2 - n_1|(ux) = (|n_2 - n_1|u)x = zx = z$ with $0 < |n_2 - n_1| < p = \text{char}(F)$, another contradiction. If $n_1 - n_2 = 0$, then $(m_1 - m_2)a = z$ with $0 < |m_1 - m_2| < p$. Since F is a field and $a \neq z$ we know that $a^{-1} \in F$, so $|m_1 - m_2|u = |m_1 - m_2|aa^{-1} = za^{-1} = z$ with $0 < |m_1 - m_2| < p$—yet another contradiction. Hence neither $m_1 - m_2$ nor $n_1 - n_2$ is 0. Therefore, $(m_1 - m_2)a = (n_2 - n_1)u \neq z$. Choose $k \in \mathbf{Z}^+$ such that $0 < k < p$ and $k(m_1 - m_2) \equiv 1 \pmod{p}$. Then $a = k(m_1 - m_2)a = k(n_2 - n_1)u$, and $a \in S_0$, one more contradiction. Hence $|S_1| = p^2$, and if $F = S_1$ the theorem is proved. If not, continue this process with an element $b \in F - S_1$. Then $S_2 = \{\ell b + ma + nu | 0 < \ell, m, n \leq p\}$ will have order p^3. (Prove this.) Since F is finite, we reach a point where $F = S_{t-1}$ for some $t \in \mathbf{Z}^+$, and $|F| = |S_{t-1}| = p^t$.

As a result of this theorem there can be no finite fields with orders such as 6, 10, 12, 14, 15, ... In addition, for each prime p and each $t \in \mathbf{Z}^+$, there is really only one field of order p^t. Any two finite fields of the same order are isomorphic. These fields were discovered by the French mathematician Evariste Galois (1811–1832) in his work on the nonexistence of formulas for solving general polynomial equations of degree ≥ 5 over $\mathbf{Q}$. As a result, a finite field of order p^t is denoted by $GF(p^t)$, where the letters GF stand for *Galois field.*

EXERCISES 17.2

1. Determine whether or not each of the following polynomials is irreducible over the given fields. If it is reducible, provide a factorization into irreducible factors.

a) $x^2 + 3x - 1$ over $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$

b) $x^4 - 2$ over $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$

c) $x^2 + x + 1$ over $\mathbf{Z}_3$, $\mathbf{Z}_5$, $\mathbf{Z}_7$

d) $x^4 + x^3 + 1$ over $\mathbf{Z}_2$

e) $x^3 + 3x^2 - x + 1$ over $\mathbf{Z}_5$

2. Give an example of a polynomial $f(x) \in \mathbf{R}[x]$ where $f(x)$ has degree 6, is reducible, but has no real roots.

3. Determine all polynomials $f(x) \in \mathbf{Z}_2[x]$ such that $1 \leq$ degree $f(x) \leq 3$ and $f(x)$ is irreducible (over $\mathbf{Z}_2$).

4. Let $f(x) = (2x^2 + 1)(5x^3 - 5x + 3)(4x - 3) \in \mathbf{Z}_7[x]$. Write $f(x)$ as the product of a unit and three monic polynomials.

5. How many monic polynomials in $\mathbf{Z}_7[x]$ have degree 5?

6. Prove Theorem 17.7.

7. An outline for a proof of Theorem 17.8 follows.

a) Let $S = \{s(x)f(x) + t(x)g(x) | s(x), t(x) \in F[x]\}$. Select an element $m(x)$ of minimum degree in S. (Recall that the zero polynomial has no degree, so it is not selected.) Can we guarantee that $m(x)$ is monic?

b) Show that if $h(x) \in F[x]$ and $h(x)$ divides both $f(x)$ and $g(x)$, then $h(x)$ divides $m(x)$.

c) Show that $m(x)$ divides $f(x)$. If not, use the division algorithm and write $f(x) = q(x)m(x) + r(x)$, where $r(x) \neq 0$ and degree $r(x) <$ degree $m(x)$. Then show that $r(x) \in S$ and obtain a contradiction.

d) Repeat the argument in part (c) to show that $m(x)$ divides $g(x)$.

8. Prove Theorems 17.9 and 17.10.

9. Use the Euclidean algorithm for polynomials to find the gcd of each pair of polynomials, over the designated field F. Then write the gcd as $s(x)f(x) + t(x)g(x)$, where $s(x), t(x) \in F[x]$.

a) $f(x) = x^2 + x - 2$, $g(x) = x^5 - x^4 + x^3 + x^2 - x - 1$ in $\mathbf{Q}[x]$

b) $f(x) = x^4 + x^3 + 1$, $g(x) = x^2 + x + 1$ in $\mathbf{Z}_2[x]$

c) $f(x) = x^4 + 2x^2 + 2x + 2$, $g(x) = 2x^3 + 2x^2 + x + 1$ in $\mathbf{Z}_3[x]$

10. If F is any field, let $f(x), g(x) \in F[x]$. If $f(x)$, $g(x)$ are relatively prime, prove that there is no element $a \in F$ with $f(a) = 0$ and $g(a) = 0$.

11. Let $f(x), g(x) \in \mathbf{R}[x]$ with $f(x) = x^3 + 2x^2 + ax - b$, $g(x) = x^3 + x^2 - bx + a$. Determine values for a, b so that the gcd of $f(x)$, $g(x)$ is a polynomial of degree 2.

12. For Example 17.9, determine which equivalence class contains each of the following:

a) $x^4 + x^3 + x + 1$

b) $x^3 + x^2 + 1$

c) $x^4 + x^3 + x^2 + 1$

13. An outline for the proof of Theorem 17.11 follows.

a) Prove that the operations defined in part (a) of Theorem 17.11 are well-defined by showing that if $f(x) \equiv f_1(x) \pmod{s(x)}$ and $g(x) \equiv g_1(x) \pmod{s(x)}$, then $f(x) + g(x) \equiv f_1(x) + g_1(x) \pmod{s(x)}$ and $f(x)g(x) \equiv f_1(x)g_1(x) \pmod{s(x)}$.

b) Verify the ring properties for the equivalence classes in $F[x]/(s(x))$.

c) Let $f(x) \in F[x]$, with $f(x) \neq 0$ and degree $f(x) <$ degree $s(x)$. If $s(x)$ is irreducible in $F[x]$, why does it follow that 1 is the gcd of $f(x)$ and $s(x)$?

d) Use part (c) to prove that if $s(x)$ is irreducible in $F[x]$, then $F[x]/(s(x))$ is a field.

e) If $|F| = q$ and degree $s(x) = n$, determine the order of $F[x]/(s(x))$.

14. a) Show that $s(x) = x^2 + 1$ is reducible in $\mathbf{Z}_2[x]$.

b) Find the equivalence classes for the ring $\mathbf{Z}_2[x]/(s(x))$.

c) Is $\mathbf{Z}_2[x]/(s(x))$ an integral domain?

15. For the field in Example 17.11, find each of the following:

a) $[x + 2][2x + 2] + [x + 1]$

b) $[2x + 1]^2[x + 2]$

c) $(22)^{-1} = [2x + 2]^{-1}$

16. Let $s(x) = x^4 + x^3 + 1 \in \mathbf{Z}_2[x]$.

a) Prove that $s(x)$ is irreducible.

b) What is the order of the field $\mathbf{Z}_2[x]/(s(x))$?

c) Find $[x^2 + x + 1]^{-1}$ in $\mathbf{Z}_2[x]/(s(x))$. (*Hint*: Find a, b, c, $d \in \mathbf{Z}_2$ so that $[x^2 + x + 1][ax^3 + bx^2 + cx + d] = [1]$.)

d) Determine $[x^3 + x + 1][x^2 + 1]$ in $\mathbf{Z}_2[x]/(s(x))$.

17. For p a prime, let $s(x)$ be irreducible of degree n in $\mathbf{Z}_p[x]$.

a) How many elements are there in the field $\mathbf{Z}_p[x]/(s(x))$?

b) How many elements in $\mathbf{Z}_p[x]/(s(x))$ generate the multiplicative group of nonzero elements of this field?

18. Give the characteristic for each of the following rings:

a) $\mathbf{Z}_{11}$ **b)** $\mathbf{Z}_{11}[x]$ **c)** $\mathbf{Q}[x]$

d) $\mathbf{Z}[\sqrt{5}] = \{a + b\sqrt{5} | a, b \in \mathbf{Z}\}$, under the binary operations of ordinary addition and multiplication of real numbers.

19. In each of the following rings, the operations are componentwise addition and multiplication, as in Exercise 18 of Section 14.2. Determine the characteristic in each case.

a) $\mathbf{Z}_2 \times \mathbf{Z}_3$ **b)** $\mathbf{Z}_3 \times \mathbf{Z}_4$ **c)** $\mathbf{Z}_4 \times \mathbf{Z}_6$

d) $\mathbf{Z}_m \times \mathbf{Z}_n$, for $m, n \in \mathbf{Z}^+$, $m, n \geq 2$

e) $\mathbf{Z}_3 \times \mathbf{Z}$

20. For Theorem 17.13, prove that $|S_2| = p^3$.

21. Find the orders n for all fields $GF(n)$, where $100 \leq n \leq 150$.

22. Construct a finite field of 25 elements.

23. Construct a finite field of 27 elements.

24. a) Prove that the function h in Example 17.10 is one-to-one and onto and preserves the operation of addition.

b) Let $(F, +, \cdot)$ and $(K, \oplus, \odot)$ be two fields. If $g\colon F \to K$ is a *ring* isomorphism and a is a nonzero element of F (that is, a is a unit in F), prove that $g(a^{-1}) = [g(a)]^{-1}$. (Consequently, this function g establishes an isomorphism of fields. In particular, the function h of Example 17.10 is such a function.)

25. a) Let $\mathbf{Q}[\sqrt{2}] = \{a + b\sqrt{2} | a, b \in \mathbf{Q}\}$. Prove that $(\mathbf{Q}[\sqrt{2}], +, \cdot)$ is a subring of the field $(\mathbf{R}, +, \cdot)$. (Here the binary operations in $\mathbf{R}$ and $\mathbf{Q}[\sqrt{2}]$ are those of ordinary addition and multiplication of real numbers.)

b) Prove that $\mathbf{Q}[\sqrt{2}]$ is a field and that $\mathbf{Q}[x]/(x^2 - 2)$ is isomorphic to $\mathbf{Q}[\sqrt{2}]$.

26. Let p be a prime. (a) How many monic quadratic (degree 2) polynomials $x^2 + bx + c$ in $\mathbf{Z}_p[x]$ can we factor into linear factors in $\mathbf{Z}_p[x]$? (For example, if $p = 5$, then the polynomial $x^2 + 2x + 2$ in $\mathbf{Z}_5[x]$ would be one of the quadratic polynomials for which we should account, under these conditions.) (b) How many quadratic polynomials $ax^2 + bx + c$ in $\mathbf{Z}_p[x]$ can we factor into linear factors in $\mathbf{Z}_p[x]$? (c) How many monic quadratic polynomials $x^2 + bx + c$ in $\mathbf{Z}_p[x]$ are irreducible over $\mathbf{Z}_p$? (d) How many quadratic polynomials $ax^2 + bx + c$ in $\mathbf{Z}_p[x]$ are irreducible over $\mathbf{Z}_p$?

17.3 Latin Squares

Our first application for this chapter deals with the structure called a Latin square. Such configurations arise in the study of combinatorial designs and play a role in statistics — in the design of experiments. We introduce the structure in the following example.

EXAMPLE 17.13

A petroleum corporation is interested in testing four types of gasoline additives to determine their effects on mileage. To do so, a research team designs an experiment wherein four different automobiles, denoted A, B, C, and D, are run on a fixed track in a laboratory. Each run uses the same prescribed amount of fuel with one of the additives present. To see how each additive affects each type of auto, the team follows the schedule in Table 17.3, where the additives are numbered 1, 2, 3, and 4. This schedule provides a way to test each additive thoroughly in each type of auto. If one additive produces the best results in all four types, the experiment will reveal its superior capability.

The same corporation is also interested in testing four other additives developed for cleaning engines. A similar schedule for these tests is shown in Table 17.4, where these engine-cleaning additives are also denoted as 1, 2, 3, and 4.

Table 17.3

	Day			
Auto	Mon	Tues	Wed	Thurs
A	1	2	3	4
B	2	1	4	3
C	3	4	1	2
D	4	3	2	1

Table 17.4

	Day			
Auto	Mon	Tues	Wed	Thurs
A	1	2	3	4
B	3	4	1	2
C	4	3	2	1
D	2	1	4	3

Furthermore, the research team is interested in the combined effect of both types of additives. It requires 16 days to test the 16 possible pairs of additives (one for improved mileage, the other for cleaning engines) in every automobile. If the results are needed in four days, the research team must design the schedules so that every pair is tested once by some auto. There are 16 ordered pairs in $\{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$, so this can be done in the allotted time if the schedules in Tables 17.3 and 17.4 are superimposed to obtain the schedule in Table 17.5. Here, for example, the entry (4, 3) indicates that on Tuesday, auto C is used to test the combined effect of the fourth additive for improved mileage and the third additive for maintaining a clean engine.

Table 17.5

	Day			
Auto	Mon	Tues	Wed	Thurs
A	(1, 1)	(2, 2)	(3, 3)	(4, 4)
B	(2, 3)	(1, 4)	(4, 1)	(3, 2)
C	(3, 4)	(4, 3)	(1, 2)	(2, 1)
D	(4, 2)	(3, 1)	(2, 4)	(1, 3)

What has happened here leads us to the following concepts.

Definition 17.9

An $n \times n$ *Latin square* is a square array of symbols, usually $1, 2, 3, \ldots, n$, where each symbol appears exactly once in each row and each column of the array.

EXAMPLE 17.14

a) Tables 17.3 and 17.4 are examples of 4×4 Latin squares.

b) For all $n \geq 2$, we can obtain an $n \times n$ Latin square from the table of the group $(\mathbf{Z}_n, +)$ if we replace the occurrences of 0 by the value of n.

From the two Latin squares in Example 17.13 we were able to produce all of the ordered pairs in $S \times S$, for $S = \{1, 2, 3, 4\}$. We now question whether or not we can do this for $n \times n$ Latin squares in general.

Definition 17.10

Let $L_1 = (a_{ij})$, $L_2 = (b_{ij})$ be two $n \times n$ Latin squares, where $1 \leq i, j \leq n$ and each a_{ij}, $b_{ij} \in \{1, 2, 3, \ldots, n\}$. If the n^2 ordered pairs (a_{ij}, b_{ij}), $1 \leq i, j \leq n$, are distinct, then L_1, L_2 are called a *pair of orthogonal Latin squares*.

EXAMPLE 17.15

a) There is no pair of 2×2 orthogonal Latin squares because the only possibilities are

$$L_1: \begin{matrix} 1 & 2 \\ 2 & 1 \end{matrix} \quad \text{and} \quad L_2: \begin{matrix} 2 & 1 \\ 1 & 2 \end{matrix}$$

b) In the 3×3 case we find the orthogonal pair

$$L_1: \begin{matrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{matrix} \quad \text{and} \quad L_2: \begin{matrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{matrix}$$

Table 17.6

1	2	3	4
4	3	2	1
2	1	4	3
3	4	1	2

c) The two 4×4 Latin squares in Example 17.13 form an orthogonal pair. The 4×4 Latin square shown in Table 17.6 is orthogonal to each of the Latin squares in that example.

We could continue listing some larger Latin squares, but we've seen enough of them at this point to ask the following questions:

1) Is there any $n > 2$ for which there is no pair of orthogonal $n \times n$ Latin squares? If so, what is the smallest such n?

2) For $n > 1$, what can we say about the number of $n \times n$ Latin squares that can be constructed so that each pair of them is orthogonal?

3) Is there a method to assist us in constructing a pair of orthogonal $n \times n$ Latin squares for certain values of $n > 2$?

Before we can examine these questions, we need to standardize some of our results.

Definition 17.11

If L is an $n \times n$ Latin square, then L is said to be in *standard form* if its first row is $1 \quad 2 \quad 3 \quad \cdots \quad n$.

Except for the Latin square L_2 in Example 17.15(a), all the Latin squares we've seen in this section are in standard form. If a Latin square is not in standard form, it can be put in that form by interchanging some of the symbols.

EXAMPLE 17.16

The 5×5 Latin square shown in (a) is not in standard form. If, however, we replace each occurrence of 4 with 1, each occurrence of 5 with 4, and each occurrence of 1 with 5, then the result is the (standard) 5×5 Latin square shown in (b).

4	2	3	5	1
1	3	5	4	2
3	4	2	1	5
2	5	1	3	4
5	1	4	2	3

(a)

1	2	3	4	5
5	3	4	1	2
3	1	2	5	4
2	4	5	3	1
4	5	1	2	3

(b)

It is often convenient to deal with Latin squares in standard form. But will this affect our results on orthogonal pairs in any way?

THEOREM 17.14

Let L_1, L_2 be an orthogonal pair of $n \times n$ Latin squares. If L_1, L_2 are standardized as L_1^*, L_2^*, then L_1^*, L_2^* are orthogonal.

Proof: The proof of this result is left for the reader.

These ideas are needed for the main results of this section.

THEOREM 17.15

In $n \in \mathbf{Z}^+$, $n > 2$, then the largest possible number of $n \times n$ Latin squares that are orthogonal in pairs is $n - 1$.

Proof: Let $L_1, L_2, \ldots, L_k$ be k distinct $n \times n$ Latin squares that are in standard form and orthogonal in pairs. We write $a_{ij}^{(m)}$ to denote the entry in the ith row and jth column of L_m, where $1 \leq i, j \leq n$, $1 \leq m \leq k$. Since these Latin squares are in standard form, we have $a_{11}^{(m)} = 1$, $a_{12}^{(m)} = 2, \ldots,$ and $a_{1n}^{(m)} = n$ for all $1 \leq m \leq k$. Now consider $a_{21}^{(m)}$, for all $1 \leq m \leq k$. These entries in the second row and first column are below $a_{11}^{(m)} = 1$. Thus $a_{21}^{(m)} \neq 1$, for all $1 \leq m \leq k$, or the configuration is not a Latin square. Further, if there exists $1 \leq \ell < m \leq k$ with $a_{21}^{(\ell)} = a_{21}^{(m)}$, then the pair L_ℓ, L_m cannot be an orthogonal pair. (Why not?) Consequently, there are at best $n - 1$ choices for the a_{21} entries in any of our $n \times n$ Latin squares, and the result follows from this observation.

This theorem places an upper bound on the number of $n \times n$ Latin squares that are orthogonal in pairs. We shall find that for certain values of n, this upper bound can be attained. In addition, our next theorem provides a method for constructing these Latin squares, though initially not in standard form. The construction uses the structure of a finite field. Before proving this theorem for the general situation, however, we shall examine one special case.

EXAMPLE 17.17

Let $F = \{f_i | 1 \leq i \leq 5\} = \mathbf{Z}_5$ with $f_1 = 1$, $f_2 = 2$, $f_3 = 3$, $f_4 = 4$, and $f_5 = 5$, the zero of $\mathbf{Z}_5$.

For $1 \leq k \leq 4$, let L_k be the 5×5 array $(a_{ij}^{(k)})$, where $1 \leq i, j \leq 5$ and

$$a_{ij}^{(k)} = f_k f_i + f_j.$$

When $k = 1$, we construct $L_1 = (a_{ij}^{(1)})$ as follows. Here $a_{ij}^{(1)} = f_1 f_i + f_j = f_i + f_j$, for $1 \leq i, j \leq 5$. With $i = 1$, the first row of L_1 is calculated as follows:

$$a_{11}^{(1)} = f_1 + f_1 = 2 \qquad a_{12}^{(1)} = f_1 + f_2 = 3 \qquad a_{13}^{(1)} = f_1 + f_3 = 4$$
$$a_{14}^{(1)} = f_1 + f_4 = 5 \qquad a_{15}^{(1)} = f_1 + f_5 = 1$$

The entries in the second row of L_1 are computed when $i = 2$. Here we find

$$a_{21}^{(1)} = f_2 + f_1 = 3 \qquad a_{22}^{(1)} = f_2 + f_2 = 4 \qquad a_{23}^{(1)} = f_2 + f_3 = 5$$
$$a_{24}^{(1)} = f_2 + f_4 = 1 \qquad a_{25}^{(1)} = f_2 + f_5 = 2$$

Continuing these calculations, we obtain the Latin square L_1 as

$$\begin{array}{ccccc} 2 & 3 & 4 & 5 & 1 \\ 3 & 4 & 5 & 1 & 2 \\ 4 & 5 & 1 & 2 & 3 \\ 5 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

For $k = 2$, the entries of L_2 are given by the formula $a_{ij}^{(2)} = f_2 f_i + f_j = 2f_i + f_j$. To obtain the first row of L_2, we set i equal to 1 and compute

$$a_{11}^{(2)} = 2f_1 + f_1 = 3 \qquad a_{12}^{(2)} = 2f_1 + f_2 = 4 \qquad a_{13}^{(2)} = 2f_1 + f_3 = 5$$
$$a_{14}^{(2)} = 2f_1 + f_4 = 1 \qquad a_{15}^{(2)} = 2f_1 + f_5 = 2$$

When i is set equal to 2, the entries in the second row of L_2 are calculated as follows:

$$a_{21}^{(2)} = 2f_2 + f_1 = 5 \qquad a_{22}^{(2)} = 2f_2 + f_2 = 1 \qquad a_{23}^{(2)} = 2f_2 + f_3 = 2$$
$$a_{24}^{(2)} = 2f_2 + f_4 = 3 \qquad a_{25}^{(2)} = 2f_2 + f_5 = 4$$

Similar calculations for $i = 3$, 4, and 5 result in the Latin square L_2 given by

$$\begin{array}{ccccc} 3 & 4 & 5 & 1 & 2 \\ 5 & 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 & 1 \\ 4 & 5 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

It is straightforward to check that the two Latin squares L_1 and L_2 are orthogonal. In Exercise 5 (at the end of this section) the reader will be asked to calculate L_3 and L_4. Our next result will verify that the four arrays L_1, L_2, L_3, and L_4 are Latin squares and that they are orthogonal in pairs.

THEOREM 17.16 Let $n \in \mathbf{Z}^+$, $n > 2$. If p is a prime and $n = p^t$, for $t \in \mathbf{Z}^+$, then there are $n - 1$ Latin squares that are $n \times n$ and orthogonal in pairs.

Proof: Let $F = GF(p^t)$, the Galois field of order $p^t = n$. Consider $F = \{f_1, f_2, \ldots, f_n\}$, where f_1 is the unity and f_n is the zero element.

We construct $n-1$ Latin squares as follows.

For each $1 \le k \le n-1$, let L_k be the $n \times n$ array $(a_{ij}^{(k)})$, $1 \le i, j \le n$, where $a_{ij}^{(k)} = f_k f_i + f_j$.

First we show that each L_k is a Latin square. If not, there are two identical elements of F in the same row or column of L_k. Suppose that a repetition occurs in a column — that is, $a_{rj}^{(k)} = a_{sj}^{(k)}$ for $1 \le r, s \le n$. Then $a_{rj}^{(k)} = f_k f_r + f_j = f_k f_s + f_j = a_{sj}^{(k)}$. This implies that $f_k f_r = f_k f_s$, by the cancellation for addition in F. Since $k \ne n$, it follows that $f_k \ne f_n$, the zero of F. Consequently, f_k is invertible, so $f_r = f_s$ and $r = s$. A similar argument shows that there are no repetitions in any row of L_k.

At this point we have $n-1$ Latin squares, $L_1, L_2, \ldots, L_{n-1}$. Now we shall prove that they are orthogonal in pairs. If not, let $1 \le k < m \le n-1$ with

$$a_{ij}^{(k)} = a_{rs}^{(k)}, \qquad a_{ij}^{(m)} = a_{rs}^{(m)}, \qquad 1 \le i, j, r, s \le n, \qquad \text{and} \qquad (i, j) \ne (r, s).$$

(Then the same ordered pair occurs twice when we superimpose L_k and L_m.) But

$$a_{ij}^{(k)} = a_{rs}^{(k)} \iff f_k f_i + f_j = f_k f_r + f_s, \qquad \text{and}$$
$$a_{ij}^{(m)} = a_{rs}^{(m)} \iff f_m f_i + f_j = f_m f_r + f_s.$$

Subtracting these equations, we find that $(f_k - f_m) f_i = (f_k - f_m) f_r$. With $k \ne m$, $(f_k - f_m)$ is not the zero of F, so it is invertible and we have $f_i = f_r$. Putting this back into either of the prior equations, we find that $f_j = f_s$. Consequently, $i = r$ and $j = s$. Therefore for $k \ne m$, the Latin squares L_k and L_m form an orthogonal pair.

The first value of n that is not a power of a prime is 6. The existence of a pair of 6×6 orthogonal Latin squares was first investigated by Leonhard Euler (1707–1783) when he sought a solution to the "problem of the 36 officers." This problem deals with six different regiments wherein six officers, each with a different rank, are selected from each regiment. (There are only six possible ranks.) The objective is to arrange the 36 officers in a 6×6 array so that in each row or column of the array, every rank and every regiment is represented exactly once. Hence each officer in the square array corresponds to an ordered pair (i, j) where $1 \le i, j \le 6$, with i for his regiment and j for his rank. In 1782 Euler conjectured that the problem could not be solved — that there is no pair of 6×6 orthogonal Latin squares. He went further and conjectured that for all $n \in \mathbf{Z}^+$, if $n \equiv 2 \pmod 4$, then there is no pair of $n \times n$ orthogonal Latin squares. In 1900 G. Tarry verified Euler's conjecture for $n = 6$ by a systematic enumeration of all possible 6×6 Latin squares. However, it was not until 1960, through the combined efforts of R. C. Bose, S. S. Shrikhande, and E. T. Parker, that the remainder of Euler's conjecture was proved false. They showed that if $n \in \mathbf{Z}^+$ with $n \equiv 2 \pmod 4$ and $n > 6$, then there exists a pair of $n \times n$ orthogonal Latin squares.

For more on this result and Latin squares in general, the reader should consult the chapter references.

EXERCISES 17.3

1. **a)** Rewrite the following 4×4 Latin square in standard form.

1	3	4	2
3	1	2	4
2	4	3	1
4	2	1	3

b) Find a 4×4 Latin square in standard form that is orthogonal to the result in part (a).

c) Apply the reverse of the process in part (a) to the result in part (b). Show that your answer is orthogonal to the given 4×4 Latin square.

2. Prove Theorem 17.14.

3. Complete the proof of the first part of Theorem 17.16.

4. The three 4×4 Latin squares in Tables 17.3, 17.4, and 17.6 are orthogonal in pairs. Can you find another 4×4 Latin square that is orthogonal to each of these three?

5. Complete the calculations in Example 17.17 in order to obtain the two 5×5 Latin squares L_3 and L_4. Rewrite each Latin square L_i, for $1 \leq i \leq 4$, in standard form.

6. Find three 7×7 Latin squares that are orthogonal in pairs. Rewrite these results in standard form.

7. Extend the experiment in Example 17.13 so that the research team needs three 4×4 Latin squares that are orthogonal in pairs.

8. A Latin square L is called *self-orthogonal* if L and its transpose L^{tr} form an orthogonal pair.

a) Show that there is no 3×3 self-orthogonal Latin square.

b) Give an example of a 4×4 Latin square that is self-orthogonal.

c) If $L = (a_{ij})$ is an $n \times n$ self-orthogonal Latin square, prove that the elements a_{ii}, for $1 \leq i \leq n$, must all be distinct.

17.4 Finite Geometries and Affine Planes

In the Euclidean geometry of the real plane, we find that (a) two distinct points determine a unique line and (b) if ℓ is a line in the plane, and P a point not on ℓ, then there is a unique line ℓ' that contains P and is parallel to ℓ. During the eighteenth and nineteenth centuries, non-Euclidean geometries were developed when alternatives to condition (b) were investigated. Yet all of these geometries contained infinitely many points and lines. The notion of a finite geometry did not appear until the end of the nineteenth century in the work of Gino Fano (*Giornale di Matematiche*, 1892).

How can we construct such a geometry? To do so, we return to the more familiar Euclidean geometry. In order to describe points and lines in this plane algebraically, we introduced a set of coordinate axes and identified each point P by an ordered pair (c, d) of real numbers. This description set up a one-to-one correspondence between the points in the plane and the set $\mathbf{R} \times \mathbf{R}$. By using the idea of slope, we could uniquely represent each line in this plane by either (1) $x = a$, where the slope is infinite, or (2) $y = mx + b$, where m is the slope; a, m, and b are real numbers. We also found that two distinct lines are parallel if and only if they have the same slope. When their slopes are distinct, the lines intersect in a unique point.

Instead of using real numbers a, b, c, d, m for the point (c, d) and the lines $x = a$, $y = mx + b$, we now turn to a comparable *finite* structure, the finite field. Our objective is to construct what is called a (finite) affine plane.

Definition 17.12

Let $\mathcal{P}$ be a finite set of points, and let $\mathcal{L}$ be a set of subsets of $\mathcal{P}$, called lines. A (*finite*) *affine plane* on the sets $\mathcal{P}$ and $\mathcal{L}$ is a finite structure satisfying the following conditions.

A1) Two distinct points of $\mathcal{P}$ are (simultaneously) in only one element of $\mathcal{L}$; that is, they are on only one line.

A2) For each $\ell \in \mathcal{L}$, and each $P \in \mathcal{P}$ with $P \notin \ell$, there exists a unique element $\ell' \in \mathcal{L}$ where $P \in \ell'$ and ℓ, ℓ' have no point in common.

A3) There are four points in $\mathcal{P}$, no three of which are collinear (that is, no three of these four points are in any one of the subsets $\ell \in \mathcal{L}$).

Figure 17.1

The reason for condition (A3) is to avoid uninteresting situations like the one shown in Fig. 17.1. If only conditions (A1) and (A2) were considered, then this system would be an affine plane.

We return now to our construction. Let $F = GF(n)$, where $n = p^t$ for some prime p and $t \in \mathbf{Z}^+$. In constructing our affine plane, denoted by $AP(F)$, we let $\mathcal{P} = \{(c, d) | c, d \in F\}$. Thus we have n^2 points.

How many lines should we have for the set $\mathcal{L}$?

The lines fall into two categories. For a line of infinite slope the equation is $x = a$, where $a \in F$. Thus we have n such "vertical lines." The other lines are given algebraically by $y = mx + b$, where $m, b \in F$. With n choices for each of m and b, it follows that there are n^2 lines that are not "vertical." Hence $|\mathcal{L}| = n^2 + n$.

Before we verify that $AP(F)$, with $\mathcal{P}$ and $\mathcal{L}$ as constructed, is an affine plane, we make two other observations.

First, for each line $\ell \in \mathcal{L}$, if ℓ is given by $x = a$, then there are n choices for y on $\ell = \{(a, y) | y \in F\}$. Thus ℓ contains exactly n points. If ℓ is given by $y = mx + b$, for $m, b \in F$, then for each choice of x we have y uniquely determined, and again ℓ consists of n points.

Now consider any point $(c, d) \in \mathcal{P}$. This point is on the line $x = c$. Furthermore, on each line $y = mx + b$ of finite slope m, $d - mc$ uniquely determines b. With n choices for m, we see that the point (c, d) is on the n lines of the form $y = mx + (d - mc)$. Overall, (c, d) is on $n + 1$ lines.

Thus far in our construction of $AP(F)$ we have a set $\mathcal{P}$ of points and a set $\mathcal{L}$ of lines where (a) $|\mathcal{P}| = n^2$; (b) $|\mathcal{L}| = n^2 + n$; (c) each $\ell \in \mathcal{L}$ contains n points; and (d) each point in $\mathcal{P}$ is on exactly $n + 1$ lines. We shall now prove that $AP(F)$ satisfies the three conditions to be an affine plane.

A1) Let $(c, d), (e, f) \in \mathcal{P}$. Using the two-point formula for the equation of a line, we have

$$(e - c)(y - d) = (f - d)(x - c) \tag{1}$$

as a line on which we find both (c, d) and (e, f). Each of these points is on $n + 1$ lines. Could there be a second line containing both of them?

The point (c, d) is on the line $x = c$. If (e, f) is also on this line, then $e = c$, but $f \neq d$ because the points are distinct. With $e = c$, Eq. (1) reduces to $0 = (f - d)(x - c)$, or $x = c$ because $f - d \neq 0$, and so we do not have a second line.

With $c \neq e$, if (c, d), (e, f) are on a second line of the form $y = mx + b$, then $d = mc + b$, $f = me + b$, and $(f - d) = m(e - c)$. Our coefficients are taken from a field and $e \neq c$, so $m = (f - d)(e - c)^{-1}$ and $b = d - mc = d - (f - d) \cdot (e - c)^{-1}c$. Consequently, this second line containing (c, d) and (e, f) is

$$y = (f - d)(e - c)^{-1}x + [d - (f - d)(e - c)^{-1}c]$$

or, because multiplication in F is commutative, $(e - c)(y - d) = (f - d)(x - c)$, which is Eq. (1). Thus two points from $\mathcal{P}$ are on only one line, and condition (A1) is satisfied.

A2) To verify this condition, consider the point P and the line ℓ as shown in Fig. 17.2. Since there are n points on any line, let $P_1, P_2, \ldots, P_n$ be the points of ℓ. (These are the only points on ℓ, although the figure might suggest others.) The point P is not on ℓ, so P and P_i determine a unique line ℓ_i, for each $1 \leq i \leq n$. We showed earlier that each point is on $n + 1$ lines, so now there is one additional line ℓ' with P on ℓ' and with ℓ' not intersecting ℓ.

A3) The last condition uses the field F. Since $|F| \geq 2$, there is the unity 1 and the zero element 0 in F. Considering the points $(0, 0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$, if line ℓ

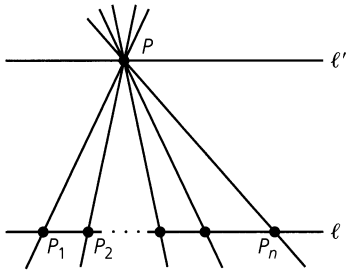


Figure 17.2

contains any three of these points, then two of the points have the form (c, c), (c, d). Consequently the equation for ℓ is given by $x = c$, which is not satisfied by either (d, c) or (d, d). Hence no three of these points are collinear.

We have now shown the following.

THEOREM 17.17 If F is a finite field, then the system based on the set $\mathcal{P}$ of points and the set $\mathcal{L}$ of lines, as described above, is an affine plane denoted by $AP(F)$.

Some particular examples will indicate a connection between these finite geometries, or affine planes, and the Latin squares of the previous section.

EXAMPLE 17.18

For $F = (\mathbf{Z}_2, +, \cdot)$, we have $n = |F| = 2$. The affine plane in Fig. 17.3 has $n^2 = 4$ points and $n^2 + n = 6$ lines. For example, the line $\ell_4 = \{(1, 0), (1, 1)\}$, and ℓ_4 contains no other points that the figure might suggest. Furthermore, ℓ_5 and ℓ_6 are parallel lines in this finite geometry because they do not intersect.

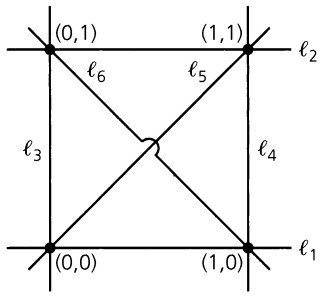


Figure 17.3

EXAMPLE 17.19

Let $F = GF(2^2)$ — the field of Example 17.9. Recall the notation of Example 17.11(d) and write $F = \{00, 01, 10, 11\}$, with addition and multiplication given by Table 17.7. We use this field to construct a finite geometry with $n^2 = 16$ points and $n^2 + n = 20$ lines. The 20 lines can be partitioned into five *parallel classes* of four lines each.

Class 1: Here we have the lines of infinite slope. These four "vertical" lines are given by the equations $x = 00$, $x = 01$, $x = 10$, and $x = 11$.

Class 2: For the "horizontal" class, or class of slope 0, we have the four lines $y = 00$, $y = 01$, $y = 10$, and $y = 11$.

Table 17.7

+	00	01	10	11
00	00	01	10	11
01	01	00	11	10
10	10	11	00	01
11	11	10	01	00

·	00	01	10	11
00	00	00	00	00
01	00	01	10	11
10	00	10	11	01
11	00	11	01	10

Class 3: The lines with slope 01 are those whose equations are $y = 01x + 00$, $y = 01x + 01$, $y = 01x + 10$, and $y = 01x + 11$.

Class 4: This class consists of the lines with equations $y = 10x + 00$, $y = 10x + 01$, $y = 10x + 10$, and $y = 10x + 11$.

Class 5: The last class contains the four lines given by $y = 11x + 00$, $y = 11x + 01$, $y = 11x + 10$, and $y = 11x + 11$.

Since each line in $AP(F)$ contains four points and each parallel class contains four lines, we shall see now how three of these parallel classes partition the 16 points of $AP(F)$.

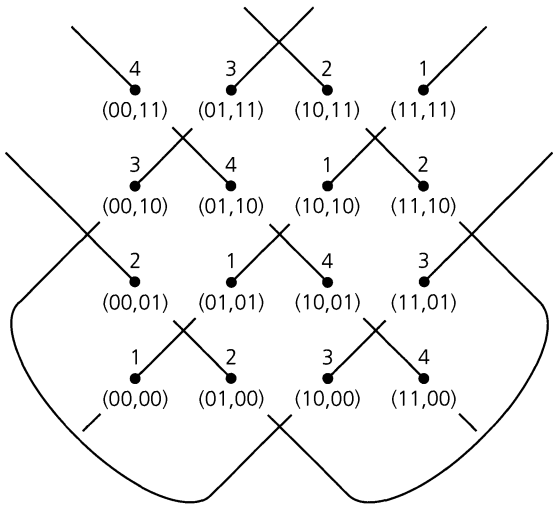


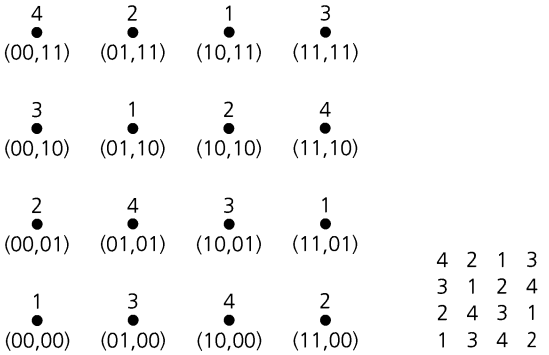
Figure 17.4

For the class with $m = 01$, there are four lines: (1) $y = 01x + 00$; (2) $y = 01x + 01$; (3) $y = 01x + 10$; and (4) $y = 01x + 11$. Above each point in $AP(F)$ we write the number corresponding to the line it is on. (See Fig. 17.4.) This configuration can be given by the following Latin square:

$$\begin{array}{cccc} 4 & 3 & 2 & 1 \\ 3 & 4 & 1 & 2 \\ 2 & 1 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{array}$$

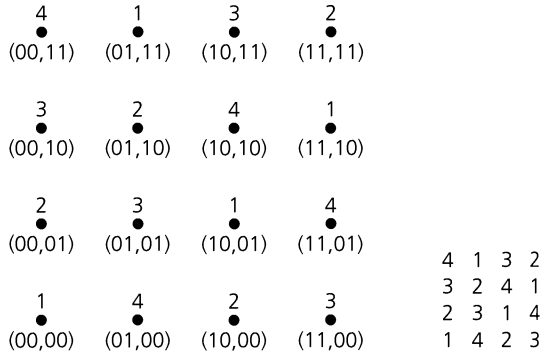
If we repeat this process for classes 4 and 5, we get the partitions shown in Figs. 17.5 and 17.6, respectively. In each class the lines are listed, for the given slope, in the same order as for Fig. 17.4. Within each figure is the corresponding Latin square.

These figures give us three 4×4 Latin squares that are orthogonal in pairs.



4 2 1 3
3 1 2 4
2 4 3 1
1 3 4 2

Figure 17.5



4 1 3 2
3 2 4 1
2 3 1 4
1 4 2 3

Figure 17.6

The results of this example are no accident, as demonstrated by the following theorem.

THEOREM 17.18 Let $F = GF(n)$, where $n \geq 3$ and $n = p^t$, p a prime, $t \in \mathbf{Z}^+$. The Latin squares that arise from $AP(F)$ for the $n - 1$ parallel classes, where the slope is neither 0 nor infinite, are orthogonal in pairs.

Proof: A proof of this result is outlined in the Section Exercises.

EXERCISES 17.4

1. Complete the following table dealing with affine planes.

Field	Number of Points	Number of Lines	Number of Points on a Line	Number of Lines on a Point
	25			
$GF(3^2)$				
		56		
				17
			31	

2. How many parallel classes do each of the affine planes in Exercise 1 determine? How many lines are in each class?

3. Construct the affine plane $AP(\mathbf{Z}_3)$. Determine its parallel classes and the corresponding Latin squares for the classes of finite nonzero slope.

4. Repeat Exercise 3 with $\mathbf{Z}_5$ taking the place of $\mathbf{Z}_3$.

5. Determine each of the following lines.

a) The line in $AP(\mathbf{Z}_7)$ that is parallel to $y = 4x + 2$ and contains (3, 6).

b) The line in $AP(\mathbf{Z}_{11})$ that is parallel to $2x + 3y + 4 = 0$ and contains (10, 7).

c) The line in $AP(F)$, where $F = GF(2^2)$, that is parallel to $10y = 11x + 01$ and contains (11, 01). (See Table 17.7.)

6. Suppose we try to construct an affine plane $AP(\mathbf{Z}_6)$ as we did in this section.

a) Determine which of the conditions (A1), (A2), and (A3) fail in this situation.

b) Find how many lines contain a given point P and how many points are on a given line ℓ, for this "geometry."

7. The following provides an outline for a proof of Theorem 17.18.

a) Consider a parallel class of lines given by $y = mx + b$, where $m \in F$, $m \neq 0$. Show that each line in this class inter-

sects each "vertical" line and each "horizontal" line in exactly one point of $AP(F)$. Thus the configuration obtained by labeling the points of $AP(F)$, as in Figs. 17.4, 17.5, and 17.6, is a Latin square.

b) To show that the Latin squares corresponding to two different classes, other than the classes of slope 0 or infinite slope, are orthogonal, assume that an ordered pair (i, j) appears more than once when one square is superimposed upon the other. How does this lead to a contradiction?

17.5 Block Designs and Projective Planes

In this final section, we examine a type of combinatorial design and see how it is related to the structure of a finite geometry. The following example will illustrate this design.

EXAMPLE 17.20

Dick (d) and his wife Mary (m) go to New York City with their five children — Richard (r), Peter (p), Christopher (c), Brian (b), and Julie (j). While staying in the city they receive three passes each day, for a week, to visit the Empire State Building. Can we make up a schedule for this family so that everyone gets to visit this attraction the same number of times?

The following schedule is one possibility.

1) b, c, d **2)** b, j, r **3)** b, m, p **4)** c, j, m
5) c, p, r **6)** d, j, p **7)** d, m, r

Here the result was obtained by trial and error. For a problem of this size such a technique is feasible. However, in general, a more effective strategy is needed. Furthermore, in asking for a certain schedule, we may be asking for something that doesn't exist. In this problem, for example, each pair of family members is together on only one visit. If the family had received four passes each day, we would not be able to construct a schedule that maintained this property.

The situation in this example generalizes as follows.

Definition 17.13

Let V be a set with v elements. A collection $\{B_1, B_2, \ldots, B_b\}$ of subsets of V is called a *balanced incomplete block design*, or (v, b, r, k, λ)-*design*, if the following conditions are satisfied:

a) For each $1 \leq i \leq b$, the subset B_i contains k elements, where k is a fixed constant and $k < v$.

b) Each element $x \in V$ is in r $(\leq b)$ of the subsets B_i, $1 \leq i \leq b$.

c) Every pair x, y of elements of V appears together in λ $(\leq b)$ of the subsets B_i, $1 \leq i \leq b$.

The elements of V are often called *varieties* because of the early applications in the design of experiments that dealt with tests on fertilizers and plants. The b subsets $B_1, B_2, \ldots, B_b$ of V are called *blocks*, where each block contains k varieties. The number r is referred to as the *replication number* of the design. Finally, λ is termed the *covalency* for the design. This parameter makes the design balanced in the following sense. For general block designs we have a number λ_{xy} for each pair $x, y \in V$; if λ_{xy} is the same for all pairs of elements from

V, then λ represents this common measure and the design is called *balanced*. In this text we only deal with balanced designs.

EXAMPLE 17.21

a) The schedule in Example 17.20 is an example of a (7, 7, 3, 3, 1)-design.

b) For $V = \{1, 2, 3, 4, 5, 6\}$, the ten blocks

1 2 4	1 3 4	1 5 6	2 3 6	3 4 6
1 2 6	1 3 5	2 3 5	2 4 5	4 5 6

constitute a (6, 10, 5, 3, 2)-design.

c) If F is a finite field, with $|F| = n$, then the affine plane $AP(F)$ yields an $(n^2, n^2 + n, n + 1, n, 1)$-design. Here the varieties are the n^2 points in $AP(F)$; the $n^2 + n$ lines are the blocks of the design.

At this point there are five parameters determining our design. We now examine how these parameters are related.

THEOREM 17.19

For a (v, b, r, k, λ)-design, (1) $vr = bk$ and (2) $\lambda(v - 1) = r(k - 1)$.

Proof:

1) With b blocks in the design and k elements per block, listing all the elements of the blocks, we get bk symbols. This collection of symbols consists of the elements of V with each element appearing r times, for a total of vr symbols. Hence $vr = bk$.

2) For this property we introduce the *pairwise incidence matrix* A for the design. With $|V| = v$, let $t = \binom{v}{2}$, the number of pairs of elements in V. We construct the $t \times b$ matrix $A = (a_{ij})$ by defining $a_{ij} = 1$ if the ith pair of elements from V is in the jth block of the design; if not, $a_{ij} = 0$.

$$\begin{array}{c} \\ x_1x_2 \\ x_1x_3 \\ \vdots \\ x_1x_v \\ x_2x_3 \\ \vdots \\ x_{v-1}x_v \end{array} \begin{array}{c} \begin{array}{cccc} B_1 & B_2 & \cdots & B_b \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1b} \\ a_{21} & a_{22} & \cdots & a_{2b} \\ \vdots & \vdots & \vdots & \vdots \\ a_{v-1\,1} & a_{v-1\,2} & \cdots & a_{v-1\,b} \\ a_{v\,1} & a_{v\,2} & \cdots & a_{v\,b} \\ \vdots & \vdots & \vdots & \vdots \\ a_{t\,1} & a_{t\,2} & \cdots & a_{t\,b} \end{bmatrix} \end{array}$$

We now count the number of 1's in matrix A in two ways.

a) Consider the rows. Since each pair x_i, x_j, for $1 \le i < j \le v$, appears in λ blocks, it follows that each row contains λ 1's. With t rows in the matrix, the number of 1's is then $\lambda t = \lambda v(v - 1)/2$.

b) Now consider the columns. As each block contains k elements, this determines $\binom{k}{2} = k(k - 1)/2$ pairs, and this is the number of 1's in each column of matrix A. With b columns, the total number of 1's is $bk(k - 1)/2$.

Then, $\lambda v(v - 1)/2 = bk(k - 1)/2 = vr(k - 1)/2$, so $\lambda(v - 1) = r(k - 1)$.

As we mentioned earlier, when n is a power of a prime, an $(n^2, n^2+n, n+1, n, 1)$-design can be obtained from the affine plane $AP(F)$, where $F = GF(n)$. Here the points are the varieties and the lines are the blocks. We shall now introduce a construction that enlarges $AP(F)$ to what is called a finite projective plane. From this projective plane we can construct an $(n^2+n+1, n^2+n+1, n+1, n+1, 1)$-design. First let us see how these two kinds of planes compare.

Definition 17.14

If $\mathcal{P}'$ is a finite set of points and $\mathcal{L}'$ a set of lines, each of which is a nonempty subset of $\mathcal{P}'$, then the (finite) plane based on $\mathcal{P}'$ and $\mathcal{L}'$ is called a *projective plane* if the following conditions are satisfied.

P1) Two distinct points of $\mathcal{P}'$ are on only one line.

P2) Any two lines from $\mathcal{L}'$ intersect in a unique point.

P3) There are four points in $\mathcal{P}'$, no three of which are collinear.

The difference between the affine and projective planes lies in the condition dealing with the existence of parallel lines. Here the parallel lines of the affine plane based on $\mathcal{P}$ and $\mathcal{L}$ will intersect when the given system is enlarged to the projective plane based on $\mathcal{P}'$ and $\mathcal{L}'$.

The construction proceeds as follows.

EXAMPLE 17.22

Start with an affine plane $AP(F)$ where $F = GF(n)$. For each point $(x, y) \in \mathcal{P}$, rewrite the point as $(x, y, 1)$. We then think of the points as ordered triples (x, y, z) where $z = 1$. Rewrite the equations of the lines $x = c$ and $y = mx + b$ in $AP(F)$ as $x = cz$ and $y = mx + bz$, where $z = 1$. We still have our original affine plane $AP(F)$, but with a change of notation.

Add the set of points $\{(1, 0, 0)\} \cup \{(x, 1, 0) | x \in F\}$ to $\mathcal{P}$ to get the set $\mathcal{P}'$. Then $|\mathcal{P}'| = n^2 + n + 1$. Let ℓ_∞ be the subset of $\mathcal{P}'$ consisting of these new points. This new line can be given by the equation $z = 0$, with the stipulation that we never have $x = y = z = 0$. Hence $(0, 0, 0) \notin \mathcal{P}'$.

Now let us examine these ideas for the affine plane $AP(\mathbf{Z}_2)$. Here $\mathcal{P} = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$, so

$$\mathcal{P}' = \{(0, 0, 1), (1, 0, 1), (0, 1, 1), (1, 1, 1)\} \cup \{(1, 0, 0), (0, 1, 0), (1, 1, 0)\}.$$

The six lines in $\mathcal{L}$ were originally

$x = 0$: $\{(0, 0), (0, 1)\}$ $\quad$ $y = 0$: $\{(0, 0), (1, 0)\}$ $\quad$ $y = x$: $\{(0, 0), (1, 1)\}$

$x = 1$: $\{(1, 0), (1, 1)\}$ $\quad$ $y = 1$: $\{(0, 1), (1, 1)\}$ $\quad$ $y = x + 1$: $\{(0, 1), (1, 0)\}$

We rewrite these as

$$x = 0 \qquad y = 0 \qquad y = x \qquad x = z \qquad y = z \qquad y = x + z$$

and add a new line ℓ_∞ defined by $z = 0$. These constitute the set $\mathcal{L}'$ of lines for our projective plane. And now at this point we consider z as a *variable*. Consequently, the line $x = z$ consists of the points $(0, 1, 0)$, $(1, 0, 1)$, and $(1, 1, 1)$. In fact, each line of $\mathcal{L}$ that contained

two points will now contain three points when considered in $\mathscr{L}'$. The set $\mathscr{L}'$ consists of the following seven lines.

$x = 0$: {(0, 0, 1), (0, 1, 0), (0, 1, 1)}
$y = 0$: {(0, 0, 1), (1, 0, 0), (1, 0, 1)}
$x = z$: {(0, 1, 0), (1, 0, 1), (1, 1, 1)}
$z = 0$ (ℓ_∞): {(1, 0, 0), (0, 1, 0), (1, 1, 0)}
$y = z$: {(1, 0, 0), (0, 1, 1), (1, 1, 1)}
$y = x$: {(0, 0, 1), (1, 1, 0), (1, 1, 1)}
$y = x + z$: {(0, 1, 1), (1, 1, 0), (1, 0, 1)}

In the original affine plane the lines $x = 0$ and $x = 1$ were parallel because no point in this plane satisfied both equations simultaneously. Here in this new system $x = 0$ and $x = z$ intersect in the point (0, 1, 0), so they are no longer parallel in the sense of $AP(\mathbf{Z}_2)$. Likewise, $y = x$ and $y = x + 1$ were parallel in $AP(\mathbf{Z}_2)$, whereas here the lines $y = x$ and $y = x + z$ intersect at (1, 1, 0). We depict this projective plane based on $\mathscr{P}'$ and $\mathscr{L}'$ as shown in Fig. 17.7. Here the "circle" through (1, 0, 1), (1, 1, 0), and (0, 1, 1) is the line $y = x + z$. Note that every line intersects ℓ_∞, which is often called the *line at infinity*. This line consists of the three *points at infinity*. We define two lines to be parallel in the projective plane when they intersect in a point at infinity (or on ℓ_∞).

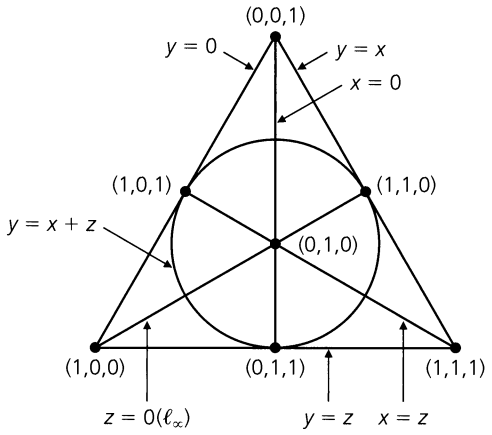


Figure 17.7

This projective plane provides us with a (7, 7, 3, 3, 1)-design like the one we developed by trial and error in Example 17.20.

We generalize the results of Example 17.22 as follows: Let n be a power of a prime. The affine plane $AP(F)$, for $F = GF(n)$, provides an example of an $(n^2, n^2 + n, n + 1, n, 1)$-design. In $AP(F)$ the $n^2 + n$ lines fall into $n + 1$ parallel classes. For each parallel class we add a point at infinity to $AP(F)$. The point (0, 1, 0) is added for the class of lines $x = cz$, $c \in F$; the point (1, 0, 0) for the class of lines $y = bz$, $b \in F$. When $m \in F$ and $m \neq 0$, then we add the point $(m^{-1}, 1, 0)$ for the class of lines $y = mx + bz$, $b \in F$. The line at infinity, ℓ_∞, is then defined as the set of $n + 1$ points at infinity. In this way we obtain the projective plane over $GF(n)$, which has $n^2 + n + 1$ points and $n^2 + n + 1$ lines. Here each point is on $n + 1$ lines, and each line contains $n + 1$ points. Furthermore, any two points in this plane are on only one line. Consequently, we have an example of an $(n^2 + n + 1, n^2 + n + 1, n + 1, n + 1, 1)$-design.

EXERCISES 17.5

1. Let $V = \{1, 2, \ldots, 9\}$. Determine the values of v, b, r, k, and λ for the design given by the following blocks.

1 2 6 1 4 7 2 3 4 2 7 9 3 7 8 4 6 8

1 3 5 1 8 9 2 5 8 3 6 9 4 5 9 5 6 7

2. Find an example of a $(4, 4, 3, 3, \lambda)$-design.

3. Find an example of a $(7, 7, 4, 4, \lambda)$-design.

4. Complete the following table so that the parameters v, b, r, k, λ in any row may be possible for a balanced incomplete block design.

v	b	r	k	λ
4			3	2
9	12		3	
10		9		2
13		4	4	
	30	10		3

5. Is it possible to have a (v, b, r, k, λ)-design where (a) $b = 28$, $r = 4$, $k = 3$? (b) $v = 17$, $r = 8$, $k = 5$?

6. Given a (v, b, r, k, λ)-design with $b = v$, prove that if v is even, then λ is even.

7. A (v, b, r, k, λ)-design is called a *triple system* if $k = 3$. When $k = 3$ and $\lambda = 1$, we call the design a *Steiner triple system*.

a) Prove that in every triple system, $\lambda(v - 1)$ is even and $\lambda v(v - 1)$ is divisible by 6.

b) Prove that in every Steiner triple system, v is congruent to 1 or 3 modulo 6.

8. Verify that the following blocks constitute a Steiner triple system on nine varieties.

1 2 8 1 4 7 2 3 4 2 7 9 3 8 9 4 6 8

1 3 5 1 6 9 2 5 6 3 6 7 4 5 9 5 7 8

9. In a Steiner triple system with $b = 12$, find the values of v and r.

10. In each of the following, $\mathcal{P}'$ is a set of points and $\mathcal{L}'$ a set of lines, each of which is a nonempty subset of $\mathcal{P}'$. Which of the conditions (P1), (P2), and (P3) of Definition 17.14 hold for the given $\mathcal{P}'$ and $\mathcal{L}'$?

a) $\mathcal{P}' = \{a, b, c\}$
$\mathcal{L}' = \{\{a, b\}, \{a, c\}, \{b, c\}\}$

b) $\mathcal{P}' = \{(x, y, z) | x, y, z \in \mathbf{R}\} = \mathbf{R}^3$
$\mathcal{L}'$ is the set of all lines in $\mathbf{R}^3$.

c) $\mathcal{P}'$ is the set of all lines in $\mathbf{R}^3$ that pass through $(0, 0, 0)$.
$\mathcal{L}'$ is the set of all planes in $\mathbf{R}^3$ that pass through $(0, 0, 0)$.

11. Bowling teams of five students each are formed from a class of 15 college freshmen. Each of the students bowls on the same number of teams; each pair of students bowls together on two teams. (a) How many teams are there in all? (b) On how many different teams does each student bowl?

12. Mrs. Mackey gave her computer science class a list of 28 problems and directed each student to write algorithms for the solutions of exactly seven of these problems. If each student did as instructed and if for each pair of problems there was exactly one pair of students who wrote algorithms to solve them, how many students did Mrs. Mackey have in her class?

13. Consider a (v, b, r, k, λ)-design on the set V of varieties, where $|V| = v \geq 2$. If $x, y \in V$, how many blocks in the design contain either x or y?

14. In a programming class Professor Madge has a total of n students, and she wants to assign teams of m students to each of p computer projects. If each student must be assigned to the same number of projects, (a) in how many projects will each individual student be involved? (b) in how many projects will each pair of students be involved?

15. **a)** If a projective plane has six lines through every point, how many points does this projective plane have in all?

b) If there are 57 points in a projective plane, how many points lie on each line of the plane?

16. In constructing the projective plane from $AP(\mathbf{Z}_2)$ in Example 17.22, why didn't we want to include the point $(0, 0, 0)$ in the set $\mathcal{P}'$?

17. Determine the values of v, b, r, k, and λ for the balanced incomplete block design associated with the projective plane that arises from $AP(F)$ for the following choices of F: (a) $\mathbf{Z}_5$ (b) $\mathbf{Z}_7$ (c) $GF(8)$.

18. **a)** List the points and lines in $AP(\mathbf{Z}_3)$. How many parallel classes are there for this finite geometry? What are the parameters for the associated balanced incomplete block design?

b) List the points and lines for the projective plane that arises from $AP(\mathbf{Z}_3)$. Determine the points on ℓ_∞, and use them to determine the "parallel" classes for this geometry. What are the parameters for the associated balanced incomplete block design?

17.6 Summary and Historical Review

The structure of a field was first developed in Chapter 14. In this chapter we examined polynomial rings and their role in the structure of finite fields, directing our attention to applications in finite geometries and combinatorial designs.

In Chapter 15 we saw that the order of a finite Boolean algebra could only be a power of 2. Now we find that for a finite field the order can only be a power of a prime and that for each prime p and each $n \in \mathbf{Z}^+$, there is only one field, up to isomorphism, of order p^n. This field is denoted by $GF(p^n)$, in honor of the French mathematician Evariste Galois (1811–1832).

Evariste Galois (1811–1832)

The finite fields $(\mathbf{Z}_p, +, \cdot)$, for p a prime, were obtained in Chapter 14 by means of the equivalence relation, congruence modulo p, defined on $\mathbf{Z}$. Using these finite fields, we developed here the integral domains $\mathbf{Z}_p[x]$. Then, with $s(x)$ an irreducible polynomial of degree n in $\mathbf{Z}_p[x]$, a similar equivalence relation — namely, congruence modulo $s(x)$ — gave us a set of p^n equivalence classes, denoted $\mathbf{Z}_p[x]/(s(x))$. These p^n equivalence classes became the elements of the field $GF(p^n)$. (Although we did not prove every possible result in general, it can be shown that over the finite field $\mathbf{Z}_p$, there is an irreducible polynomial of degree n for each $n \in \mathbf{Z}^+$.)

The theory of finite fields was developed by Galois in his work addressing the problem of the solutions of polynomial equations. As we mentioned in the summary of Chapter 16, the study of polynomial equations was an area of research that challenged many mathematicians from the sixteenth to the nineteenth centuries. In the nineteenth century, Niels Henrik Abel (1802–1829) first showed that the solution of the general quintic could not be given by radicals. Galois showed that for any polynomial of degree n over a field F, there is a corresponding group G that is isomorphic to a subgroup of S_n, the group of permutations of $\{1, 2, 3, \ldots, n\}$. The essence of Galois's work is that such a polynomial equation can be solved by (addition, subtraction, multiplication, division, and) radicals if its corresponding group is *solvable*. Now what makes a finite group solvable? We say that a finite group G is solvable if it has a chain of subgroups $G = K_1 \supset K_2 \supset K_3 \supset \cdots \supset K_t = \{e\}$, where for all

$2 \leq i \leq t$, K_i is a normal subgroup of K_{i-1} (that is, $xyx^{-1} \in K_i$ for all $y \in K_i$ and for all $x \in K_{i-1}$), and the quotient group K_{i-1}/K_i is abelian. One finds that all subgroups of S_i, for $1 \leq i \leq 4$, are solvable, but for $n \geq 5$ there are subgroups of S_n that are not solvable.

Though it seems that Galois theory is concerned predominantly with groups, there is a great deal more on the theory of fields that we have not mentioned. As a consequence of Galois's work, the areas of field theory and finite group theory became topics of great mathematical interest.

For more on *Galois theory*, the reader will find Chapter 6 of the text by V. H. Larney [8] and Chapter 12 in the book by N. H. McCoy and T. R. Berger [10] good places to start. Chapter 5 of I. N. Herstein [6] has more on the topic, while a detailed presentation can be found in the text by S. Roman [11] and the classic work by O. Zariski and P. Samuel [17]. Appendix E in the text by V. H. Larney [8] includes an interesting short account of the life of Galois; more on his life can be found in the somewhat fictional account by L. Infeld [7]. The article by T. Rothman [12] provides a more contemporary discussion of the inaccuracies and myths surrounding the life, and especially the death, of Galois. The biographical notes on pages 287–291 of the text by J. Stillwell [14] relate more on the life and work of this great genius.

The Latin squares, combinatorial designs, and finite geometries of the later sections of the chapter showed us how the finite field structure entered into problems of design. Dating back to the time of Leonhard Euler (1707–1783) and the problem of the "36 officers," the study of orthogonal Latin squares has been developed considerably since 1900, and especially since 1960 with the work of R. C. Bose, S. S. Shrikhande, and E. T. Parker. Chapter 7 of the monograph by H. J. Ryser [13] provides the details of their accomplishments. The text by C. L. Liu [9] includes ideas from coding theory in its discussion of Latin squares.

The study of finite geometries can be traced back to the work of Gino Fano, who, in 1892, considered a finite three-dimensional geometry consisting of 15 points, 35 lines, and 15 planes. However, it was not until 1906 that these geometries gained any notice, when O. Veblen and W. Bussey began their study of finite projective geometries. For more on this topic, the reader should find the texts by A. A. Albert and R. Sandler [1] and H. L. Dorwart [4] very interesting. The text by P. Dombowski [3] provides an extensive coverage for those seeking something more advanced.

Finally, the notion of designs was first studied by statisticians in the area called the design of experiments. Through the research of R. A. Fisher and his followers, this area has come to play an important role in the modern theory of statistical analysis. In our development, we examined conditions under which a (v, b, r, k, λ)-design could exist and how such designs were related to affine planes and finite projective planes. The text by M. Hall, Jr. [5] provides more on this topic, as does the work by A. P. Street and W. D. Wallis [15]. Chapter XIII of reference [15] includes material relating to designs and coding theory. A rather thorough coverage of the topic of designs is given in the work by W. D. Wallis [16], and the text edited by J. H. Dinitz and D. R. Stinson [2] provides the reader with a collection of more work in this area.

REFERENCES

1. Albert, A. Adrian, and Sandler, R. *An Introduction to Finite Projective Planes*. New York: Holt, 1968.
2. Dinitz, Jeffrey H., and Stinson, Douglas R., eds. *Contemporary Design Theory*. New York: Wiley, 1992.
3. Dombowski, Peter. *Finite Geometries*. New York: Springer-Verlag, 1968.

4. Dorwart, Harold L. *The Geometry of Incidence*. Englewood Cliffs, N.J.: Prentice-Hall, 1966.
5. Hall, Marshall, Jr. *Combinatorial Theory*. Waltham, Mass.: Blaisdell, 1967.
6. Herstein, Israel Nathan. *Topics in Algebra*, 2nd ed. Lexington, Mass.: Xerox College Publishing, 1975.
7. Infeld, Leopold. *Whom the Gods Love*. New York: McGraw-Hill, 1948.
8. Larney, Violet H. *Abstract Algebra: A First Course*. Boston: Prindle, Weber & Schmidt, 1975.
9. Liu, C. L. *Topics in Combinatorial Mathematics*. Mathematical Association of America, 1972.
10. McCoy, Neal H., and Berger, Thomas R. *Algebra: Groups, Rings, and Other Topics*. Boston: Allyn and Bacon, 1977.
11. Roman, Steven. *Field Theory*. New York: Springer-Verlag, 1995.
12. Rothman, Tony. "Genius and Biographers: The Fictionalization of Evariste Galois." *The American Mathematical Monthly* 89, no. 2 (1982): pp. 84–106.
13. Ryser, Herbert J. *Combinatorial Mathematics*. Carus Mathematical Monographs, Number 14, Mathematical Association of America, 1963.
14. Stillwell, John. *Mathematics and Its History*. New York: Springer-Verlag, 1989.
15. Street, Anne Penfold, and Wallis, W. D. *Combinatorial Theory: An Introduction*. Winnipeg, Canada: The Charles Babbage Research Center, 1977.
16. Wallis, W. D. *Combinatorial Designs*. New York: Marcel Dekker, Inc., 1988.
17. Zariski, Oscar, and Samuel, Pierre. *Commutative Algebra*, Vol. I. New York: Van Nostrand, 1958.

SUPPLEMENTARY EXERCISES

1. Determine n if over $GF(n)$ there are 6561 monic polynomials of degree 5 with no constant term.

2. **a)** Let $f(x) = a_n x^n + \cdots + a_1 x + a_0 \in \mathbf{Z}[x]$. If $r/s \in \mathbf{Q}$, with $\gcd(r, s) = 1$ and $f(r/s) = 0$, prove that $s | a_n$ and $r | a_0$.

b) Find the rational roots, if any exist, of the following polynomials over $\mathbf{Q}$. Factor $f(x)$ in $\mathbf{Q}[x]$.

i) $f(x) = 2x^3 + 3x^2 - 2x - 3$

ii) $f(x) = x^4 + x^3 - x^2 - 2x - 2$

c) Show that the polynomial $f(x) = x^{100} - x^{50} + x^{20} + x^3 + 1$ has no rational root.

3. **a)** For how many integers n, where $1 \le n \le 1000$, can we factor $f(x) = x^2 + x - n$ into the product of two first degree factors in $\mathbf{Z}[x]$?

b) Answer part (a) for $f(x) = x^2 + 2x - n$.

c) Answer part (a) for $f(x) = x^2 + 5x - n$.

d) Let $g(x) = x^2 + kx - n \in \mathbf{Z}[x]$, for $1 \le n \le 1000$. Find the smallest positive integer k so that $g(x)$ cannot be factored into two first degree factors in $\mathbf{Z}[x]$ for all $1 \le n \le 1000$.

4. Verify that the polynomial $f(x) = x^4 + x^3 + x + 1$ is reducible over every field F (finite or infinite).

5. If p is a prime, prove that in $\mathbf{Z}_p[x]$,

$$x^p - x = \prod_{a \in \mathbf{Z}_p} (x - a).$$

6. For any field F, let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in F[x]$. If $r_1, r_2, \ldots, r_n$ are the roots of $f(x)$, and $r_i \in F$ for all $1 \le i \le n$, prove that

a) $-a_{n-1} = r_1 + r_2 + \cdots + r_n$.

b) $(-1)^n a_0 = r_1 r_2 \cdots r_n$.

7. Four of the seven blocks in a (7, 7, 3, 3, 1)-design are {1, 3, 7}, {1, 5, 6}, {2, 6, 7}, and {3, 4, 6}. Determine the other three blocks.

8. Find the values of b and r for a Steiner triple system where $v = 63$.

9. **a)** If a projective plane has 73 points, how many points lie on each line?

b) If each line in a projective plane passes through 10 points, how many lines are there in the projective plane?

10. A projective plane is coordinatized with the elements of a field F. If this plane contains 91 lines, what are $|F|$ and $\text{char}(F)$?

11. Let $V = \{x_1, x_2, \ldots, x_v\}$ be the set of varieties and $\{B_1, B_2, \ldots, B_b\}$ the collection of blocks for a (v, b, r, k, λ)-design. We define the *incidence matrix* A for the design by

$$A = (a_{ij})_{v \times b}, \qquad \text{where } a_{ij} = \begin{cases} 1, & \text{if } x_i \in B_j \\ 0, & \text{otherwise.} \end{cases}$$

a) How many 1's are there in each row and column of A?

b) Let $J_{m \times n}$ be the $m \times n$ matrix where every entry is 1. For $J_{n \times n}$ we write J_n. Prove that for the incidence matrix A, $A \cdot J_b = r \cdot J_{v \times b}$ and $J_v \cdot A = k \cdot J_{v \times b}$.

c) Show that

$$A \cdot A^{\text{tr}} = \begin{bmatrix} r & \lambda & \lambda & \cdots & \lambda \\ \lambda & r & \lambda & \cdots & \lambda \\ \lambda & \lambda & r & \cdots & \lambda \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \lambda & \lambda & \lambda & \cdots & r \end{bmatrix}$$
$$= (r - \lambda)I_v + \lambda J_v,$$

where I_v is the $v \times v$ (multiplicative) identity.

d) Prove that

$$\det(A \cdot A^{\text{tr}}) = (r - \lambda)^{v-1}[r + (v - 1)\lambda] = (r - \lambda)^{v-1}rk.$$

12. Given a (v, b, r, k, λ)-design based on the v varieties of V, replace each of the blocks B_i, for $1 \leq i \leq b$, by its complement $\overline{B}_i = V - B_i$. Then the collection $\{\overline{B}_1, \overline{B}_2, \ldots, \overline{B}_b\}$ provides the blocks for a (v, b, r', k', λ')-design, also based on the set V.

a) Find this corresponding complementary (v, b, r', k', λ')-design for the design given in Exercise 1 of Section 17.5.

b) In general, how are the parameters r', k', λ' of the complementary design related to the parameters v, b, r, k, λ of the original design?