NUMBER THEORY AND LINEAR ALGEBRA: MAT6B12

Shyama M.P.
Assistant Professor,
Department of Mathematics
Malabar Christian College, Calicut

January 4, 2017

Contents

1 DIVISIBILITY THEORY IN THE INTEGERS 3

2 PRIMES AND THEIR DISTRIBUTION 17

3 THE THEORY OF CONGRUENCES 21

4 FERMAT'S THEOREM 27

5 NUMBER-THEORETIC FUNCTIONS 31

6 EULER'S GENERALIZATION OF FERMAT'S THEOREM 35

7 VECTOR SPACE 39

8 LINEAR MAPPINGS 49

MODULE I

Chapter 1

DIVISIBILITY THEORY IN THE INTEGERS

Well- Ordering Principle
Every non empty set S of nonnegative integers contains a least element. That is, there exists some integer a in S such that $a \leq b$ for all b in S.

THE DIVISION ALGORITHM
Division Algorithm, the result is familiar to most of us roughly, it asserts that an integer a can be "divided" by a positive integer b in such a way that the remainder is smaller than b. The exact statement of this fact is Theorem 1.0.1:

Theorem 1.0.1. *Given integers a and b, with $b > 0$, there exist unique integers q and r satisfying*

$$a = qb + r \qquad 0 \leq r < b$$

The integers q and r are called, respectively, the quotient and remainder in the division of a by b.

Proof. Let a and b be integers with $b > 0$ and consider the set

$$S = \{a - xb : xisaninteger; a - xb \geq 0\}.$$

Claim: The set S is nonempty

It suffices to find a value x which making $a - xb$ nonnegative. Since $b \geq 1$, we have $|a|b \geq |a|$ and so, $a - (-|a|)b = a + |a|b \geq a + |a| \geq 0$. For the choice $x = -|a|$, then $a - xb$ lies in S. Therefore S is nonempty, hence the claim.

Therefore by Well-Ordering Principle, S contains a small integer, say r. By the definition of S there exists an integer q satisfying

$$r = a - qb \qquad 0 \leq r.$$

Claim: $r < b$

Suppose $r \geq b$. Then we have

$$a - (q+1)b = (a - qb) - b = r - b \geq 0.$$

This implies that, $a-(q+1)b \in S$. But $a-(q+1)b = r-b < r$, since $b > 0$, leading to a contradiction of the choice of r as the smallest member of S. Hence, $r < b$, hence the claim.
Next we have to show that the uniqueness of q and r. Suppose that a as two representations of the desired form, say,

$$a = qb + r = q'b + r',$$

where $0 \leq r < b$ and $0 \leq r' < b$. Then $(r' - r) = b(q - q')$. Taking modulus on both sides,

$$|(r' - r)| = |b(q - q')| = |b||(q - q')| = b|(q - q')|.$$

But we have $-b < -r \leq 0$ and $0 \leq r' < b$, upon adding these inequalities we obtain $-b < r' - r < b$. This implies $b|(q - q')| < b$, which yields $0 \leq |q - q'| < 1$. Because $|q - q'|$ is a nonnegative integer, the only possibility is that $|q - q'| = 0$, hence, $q = q'$. This implies $|r' - r| = 0$, that is, $r = r'$. Hence the proof. □

Corollary 1.0.1. *If a and b are integers, with $b \neq 0$, then there exists integers q and r such that*

$$a = qb + r \qquad 0 \leq r < |b|.$$

Proof. It is enough to consider the case in which b is negative. Then $|b| > 0$, and Theorem 1. produces unique integers q' and r for which

$$a = q'|b| + r \qquad 0 \leq r < |b|.$$

Noting that $|b| = -b$, we may take $q = -q'$ to arrive at $a = qb + r$, with $0 \leq r < |b|$. □

Application of the Division Algorithm

1. Square of any integer is either of the form $4k$ or $4k+1$. That is, the square of integer leaves the remainder 0 or 1 upon division by 4.
   Solution: Let a be any integer. If a is even, we can let $a = 2n, n$ is an integer, then $a^2 = (2n)^2 = 4n^2 = 4k$. If a is odd, we can let $a = 2n+1, n$ is an integer, then $a^2 = (2n+1)^2 = 4n^2+4n+1 = 4(n^2+n)+1 = 4k+1$.

2. The square of any odd integer is of the form $8k+1$.
   Solution: Let a be an integer and let $b = 4$, then by division algorithm a is representable as one of the four forms: $4q, 4q+1, 4q+2, 4q+3$. In this representation, only those integers of the forms $4q+1$ and $4q+3$ are odd. If $a = 4q+1$, then

   $$a^2 = (4q+1)^2 = 16q^2 + 8q + 1 = 8(2q^2+q) + 1 = 8k+1.$$

   If $a = 4q+3$, then

   $$a^2 = (4q+3)^2 = 16q^2+24q+9 = 16q^2+24q+8+1 = 8(2q^2+3q+1)+1 = 8k+1.$$

3. For all integer $a \geq 1$, $\frac{a(a^2+2)}{3}$ is an integer.
   Solution: Let $a \geq 1$ be an integer. According to division algorithm, a is of the form $3q, 3q+1$ or $3q+2$. If $a = 3q$, then

   $$\frac{3q((3q)^2+2)}{3} = 9q^3 + 2q,$$

   which is clearly an integer. Similarly we can prove other two cases also.

THE GREATEST COMMON DIVISOR

Definition 1.0.1. *An integer b is said to be divisible by an integer $a \neq 0$, in symbols $a|b$, if there exists some integer c such that $b = ac$. We write $a \nmid b$ to indicate that b is not divisible by a.*

Thus, for example, -22 is divisible by 11, because $-22 = 11(-2)$. However, 22 is not divisible by 3; for there is no integer c that makes the statement $22 = 3c$ true.
There is other language for expressing the divisibility relation $a|b$. We could say that a is a divisor of b, that a is a factor of b, or that b is a multiple of a. Notice that in Definition 1 there is a restriction on the divisor a: Whenever the notation $a|b$ is employed, it is understood that a is different from zero.
If a is a divisor of b, then b is also divisible by $-a$ (indeed, $b = ac$ implies that $b = (-a)(-c)$), so that the divisors of an integer always occur in pairs.

To find all the divisors of a given integer, it is sufficient to obtain the positive divisors and then adjoin to them the corresponding negative integers. For this reason, we shall usually limit ourselves to a consideration of positive divisors. It will be helpful to list some immediate consequences of Definition 1.0.1.

Theorem 1.0.2. *For integers a, b, c, the following hold:*

1. $a|0$, $1|a$, $a|a$.
2. $a|1$ *if and only if* $a = \pm 1$.
3. *If* $a|b$ *and* $c|d$, *then* $ac|bd$.
4. *If* $a|b$ *and* $b|c$, *then* $a|c$.
5. $a|b$ *and* $b|a$ *if and only if* $a = \pm b$.
6. *If* $a|b$ *and* $b \neq 0$, *then* $|a| \leq |b|$.
7. *If* $a|b$ *and* $a|c$, *then* $a|(bx + cy)$ *for arbitrary integers* x *and* y.

Proof. 1. Since $0 = a.0$, $a|0$. Since $a = 1.a$, $1|a$. Since $a = a.1$, $a|a$.

2. We have $a|1$ if and only if $1 = a.c$ for some c, this is if and only if $a = \pm 1$.

3. Clear from definition.

4. Clear from definition.

5. Clear from definition.

6. If $a|b$, then there exists an integer c such that $b = ac$; also, $b \neq 0$ implies that $c \neq 0$. Upon taking absolute values, we get $|b| = |ac| = |a||c|$. Because $c \neq 0$, it follows that $|c| \geq 1$, whence $|b| = |a||c| \geq |a|$.

7. The relations $a|b$ and $a|c$ ensure that $b = ar$ and $c = as$ for suitable integers r and s. But then whatever the choice of x and y, $bx + cy = arx + asy = a(rx + sy)$. Because $rx + sy$ is an integer, this says that $a|(bx + cy)$, as desired.

□

Definition 1.0.2. *Let a and b be given integers, with at least one of them different from zero. The greatest common divisor of a and b, denoted by $gcd(a, b)$, is the positive integer d satisfying the following:*

(i) $d|a$ *and* $d|b$.

(ii) *If* $c|a$ *and* $c|b$, *then* $c \leq d$.

Example: The positive divisors of -12 are $1, 2, 3, 4, 6, 12$, whereas those of 30 are $1, 2, 3, 5, 6, 10, 15, 30$; hence, the positive common divisors of -12 and 30 are $1, 2, 3, 6$. Because 6 is the largest of these integers, it follows that $gcd(-12, 30) = 6$. In the same way, we can show that $gcd(-5, 5) = 5$, $gcd(8, 17) = 1$, $gcd(-8, -36) = 4$.

Theorem 1.0.3. *Given integers* a *and* b, *not both of which are zero, there exist integers* x *and* y *such that*

$$gcd(a, b) = ax + by.$$

Proof. Consider the set S of all positive linear combinations of a and b :

$$S = \{au + bv : au + bv > 0; u, v \; integers\}.$$

Since, if $a \neq 0$ then $|a| = au + b.0 \in S$, where $u = 1, if\, a > 0; u = -1, if\, a < 0$, S is nonempty. Therefore by the Well-Ordering Principle, S must contain a smallest element, say d. Thus, from the very definition of S, there exist integers x and y for which $d = ax + by$.
Claim: $d = gcd(a, b)$

By using the Division Algorithm, we can obtain integers q and r such that $a = qd + r$, where $0 \leq r < d$. Then r can be written in the form:

$$\begin{aligned} r &= a - qd \\ &= a - q(ax + by) \\ &= a(1 - qx) + b(-qy). \end{aligned}$$

If r were positive, then this representation would imply that r is a member of S, contradicting the fact that d is the least integer in S (recall that $r < d$). Therefore, $r = 0$, and so $a = qd$, or equivalently $d|a$. By similar reasoning, $d|b$, this implies d is a common divisor of a and b.
Now if c is an arbitrary positive common divisor of the integers a and b, then part (7) of Theorem 2 allows us to conclude that $c|(ax + by)$; that is, $c|d$. By part (6) of the same theorem, $c = |c| \leq |d| = d$, so that d is greater than every positive common divisor of a and b. Hence $d = gcd(a, b)$. Hence the claim. Therefore $gcd(a, b) = ax + by$. □

Corollary 1.0.2. *If* a *and* b *are given integers, not both zero, then the set*

$$T = ax + by : x, y \; are \; integers$$

is precisely the set of all multiples of $d = gcd(a, b)$.

Proof. Because $d|a$ and $d|b$, we know that $d|(ax + by)$ for all integers x, y. Thus, every member of T is a multiple of d. Conversely, d may be written as $d = ax_0 + by_0$ for suitable integers x_0 and y_0 , so that any multiple nd of d is of the form

$$nd = n(ax_0 + by_0) = a(nx_0) + b(ny_0).$$

Hence, nd is a linear combination of a and b, and, by definition, lies in T. □

Definition 1.0.3. *Two integers a and b, not both of which are zero, are said to be relatively prime whenever $gcd(a, b) = 1$.*

Theorem 1.0.4. *Let a and b be integers, not both zero. Then a and b are relatively prime if and only if there exist integers x and y such that $1 = ax + by$.*

Proof. If a and b are relatively prime so that $gcd(a, b) = 1$, then Theorem 3 guarantees the existence of integers x and y satisfying $1 = ax+by$. Conversely, suppose that $1 = ax + by$ for some choice of x and y, and that $d = gcd(a, b)$. Because $d|a$ and $d|b$, Theorem 2 yields $d|(ax+by)$, or $d|1$. This implies $d = \pm 1$. But d is a positive integer, $d = 1$. That is a and b are relatively prime. □

Corollary 1.0.3. *If $gcd(a, b) = d$, then $gcd(a/d, b/d) = 1$.*

Proof. Since $d|a$ and $d|b$, a/d and b/d are integers. We have, if $gcd(a, b) = d$, then there exists x and y such that $d = ax + by$. Upon dividing each side of this equation by d, we obtain the expression

$$1 = (a/d)x + (b/d)y.$$

Because a/d and b/d are integers, a/d and b/d are relatively prime. Therefore $gcd(a/d, b/d) = 1$. □

Corollary 1.0.4. *If $a|c$ and $b|c$, with $gcd(a, b) = 1$, then $ab|c$.*

Proof. Since $a|c$ and $b|c$, we can find integers r and s such that $c = ar = bs$. Given that $gcd(a, b) = 1$, so there exists integers x and y such that $1 = ax+by$. Multiplying the last equation by c, we get,

$$c = c1 = c(ax + by) = acx + bcy.$$

If the appropriate substitutions are now made on the right-hand side, then

$$c = a(bs)x + b(ar)y = ab(sx + ry).$$

This implies, $ab|c$. □

Theorem 1.0.5. *(Euclid's lemma.) If $a|bc$, with $gcd(a,b) = 1$, then $a|c$.*

Proof. Since $gcd(a,b) = 1$, we have $1 = ax + by$ for some integers x and y. Multiplication of this equation by c produces

$$c = 1c = (ax + by)c = acx + bcy.$$

Since $a|bc$ and $a|ac$, we have $a|acx + bcy$. This implies $a|c$. □

Note: If a and b are not relatively prime, then the conclusion of Euclid's lemma may fail to hold. For example: $6|9.4$ but $6 \nmid 9$ and $6 \nmid 4$.

Theorem 1.0.6. *Let a, b be integers, not both zero. For a positive integer d, $d = gcd(a,b)$ if and only if*

(i) *$d|a$ and $d|b$.*

(ii) *Whenever $c|a$ and $c|b$, then $c|d$.*

Proof. Suppose that $d = gcd(a,b)$. Certainly, $d|a$ and $d|b$, so that (i) holds. By Theorem 3, d is expressible as $d = ax + by$ for some integers x, y. Thus, if $c|a$ and $c|b$, then $c|(ax + by)$, or rather $c|d$. This implies, condition (ii) holds. Conversely, let d be any positive integer satisfying the stated conditions (i) and (ii). Given any common divisor c of a and b, we have $c|d$ from hypothesis (ii). This implies that $d \geq c$, and consequently d is the greatest common divisor of a and b. □

THE EUCLIDEAN ALGORITHM

Lemma 1.0.1. *If $a = qb + r$, then $gcd(a,b) = gcd(b,r)$.*

Proof. If $d = gcd(a,b)$, then the relations $d|a$ and $d|b$ together imply that $d|(a - qb)$, or $d|r$. Thus, d is a common divisor of both b and r. On the other hand, if c is an arbitrary common divisor of b and r, then $c|(qb + r)$, whence $c|a$. This makes c a common divisor of a and b, so that $c \leq d$. It now follows from the definition of $gcd(b,r)$ that $d = gcd(b,r)$. □

The Euclidean algorithm

The Euclidean Algorithm may be described as follows: Let a and b be two integers whose greatest common divisor is desired. Because $gcd(|a|,|b|) = gcd(a,b)$,, with out loss of generality we may assume $a \geq b > 0$. The first step is to apply the Division Algorithm to a and b to get

$$a = q_1 b + r_1 \qquad 0 \leq r_1 < b.$$

If it happens that $r_1 = 0$, then $b|a$ and $gcd(a, b) = b$. When $r_1 \neq 0$, divide b by r_1 to produce integers q_2 and r_2 satisfying

$$b = q_2 r_1 + r_2 \qquad 0 \leq r_2 < r_1.$$

If $r_2 = 0$, then we stop; otherwise, proceed as before to obtain

$$r_1 = q_3 r_2 + r_3 \qquad 0 \leq r_3 < r_2.$$

This division process continues until some zero remainder appears, say, at the $(n+l)^{th}$ stage where r_{n-1} is divided by r_n (a zero remainder occurs sooner or later because the decreasing sequence $b > r_1 > r_2 > \cdots \geq 0$ cannot contain more than b integers). The result is the following system of equations:

$$\begin{aligned}
a &= q_1 b + r_1 \qquad 0 \leq r_1 < b \\
b &= q_2 r_1 + r_2 \qquad 0 \leq r_2 < r_1 \\
r_1 &= q_3 r_2 + r_3 \qquad 0 \leq r_3 < r_2 \\
&\vdots \\
r_{n-2} &= q_n r_{n-1} + r_n \qquad 0 \leq r_n < r_{n-1} \\
r_{n-1} &= q_{n+1} r_n + 0.
\end{aligned}$$

By Lemma 1.0.1,

$$gcd(a, b) = gcd(b, r_1) = \ = gcd(r_{n-1}, r_n) = gcd(r_n, 0) = r_n.$$

Note: Start with the next-to-last equation arising from the Euclidean Algorithm, we can determine x and y such that $gcd(a, b) = ax + by$.
Example: Let us see how the Euclidean Algorithm works in a concrete case by calculating, say, $gcd(12378, 3054)$. The appropriate applications of the Division Algorithm produce the equations

$$\begin{aligned}
12378 &= 4.3054 + 162 \\
3054 &= 18.162 + 138 \\
162 &= 1.138 + 24 \\
138 &= 5.24 + 18 \\
24 &= 1.18 + 6 \\
18 &= 3.6 + 0
\end{aligned}$$

This tells us that the last nonzero remainder appearing in these equations, namely, the integer 6, is the greatest common divisor of 12378 and 3054:

$$6 = gcd(12378, 3054).$$

To represent 6 as a linear combination of the integers 12378 and 3054, we start with the next-to-last of the displayed equations and successively eliminate the remainders 18, 24, 138, and 162:

$$\begin{aligned} 6 &= 24 - 18 \\ &= 24 - (138 - 5.24) \\ &= 6.24 - 138 \\ &= 6(162 - 138) - 138 \\ &= 6.162 - 7.138 \\ &= 6.162 - 7(3054 - 18.162) \\ &= 132.162 - 7.3054 \\ &= 132(12378 - 4.3054) - 7.3054 \\ &= 132.12378 + (-535)3054 \end{aligned}$$

Thus, we have

$$6 = gcd(12378, 3054) = 12378x + 3054y,$$

where $x = 132$ and $y = -535$. Note that this is not the only way to express the integer 6 as a linear combination of 12378 and 3054; among other possibilities, we could add and subtract 3054.12378 to get

$$\begin{aligned} 6 &= (132 + 3054)12378 + (-535 - 12378)3054 \\ &= 3186.12378 + (-12913)3054. \end{aligned}$$

Theorem 1.0.7. *If $k > 0$, then $gcd(ka, kb) = k\ gcd(a, b)$.*

Proof. If each of the equations appearing in the Euclidean Algorithm for a and b, multiplied by k, we obtain

$$\begin{aligned} ak &= q_1(bk) + r_1k \qquad 0 \leq r_1k < bk \\ bk &= q_2(r_1k) + r_2k \qquad 0 \leq r_2k < r_1k \\ &\vdots \\ r_{n-2}k &= q_n(r_{n-1}k) + r_nk \qquad 0 \leq r_nk < r_{n-1}k \\ r_{n-1}k &= q_{n+1}(r_nk) + 0. \end{aligned}$$

But this is clearly the Euclidean Algorithm applied to the integers ak and bk, so that their greatest common divisor is the last nonzero remainder $r_n k$; that is,

$$gcd(ka, kb) = r_n k = k\ gcd(a, b),$$

Hence the theorem. □

Corollary 1.0.5. *For any integer $k \neq 0$, $gcd(ka, kb) = |k|\ gcd(a, b)$.*

Proof. We already have, if $k > 0$, then $gcd(ka, kb) = k\ gcd(a, b)$. Therefore it suffices to consider the case in which $k < 0$. Then $-k = |k| > 0$ and, by Theorem 1.0.7,

$$\begin{aligned} gcd(ak, bk) &= gcd(-ak, -bk) \\ &= gcd(a|k|, b|k|) \\ &= |k|\ gcd(a, b). \end{aligned}$$

Hence the result. □

Definition 1.0.4. *The least common multiple of two nonzero integers a and b, denoted by $lcm(a, b)$, is the positive integer m satisfying the following:*

(i) *$a|m$ and $b|m$.*

(ii) *If $a|c$ and $b|c$, with $c > 0$, then $m \leq c$.*

As an example, the positive common multiples of the integers -12 and 30 are 60, 120, 180, ... hence, $lcm(-12, 30) = 60$.

Theorem 1.0.8. *For positive integers a and b*

$$gcd(a, b)\ lcm(a, b) = ab.$$

Proof. Let $d = gcd(a, b)$ and let $m = ab/d$, then $m > 0$.
Claim: $m = lcm(a, b)$

Since d is the common divisor of a and b we have $a = dr$, $b = ds$ for integers r and s. Then $m = as = rb$. This implies, m a (positive) common multiple of a and b.
Now let c be any positive integer that is a common multiple of a and b, then $c = au = bv$ for some integers u and v. As we know, there exist integers x and y satisfying d =ax+ by. In consequence,

$$\frac{c}{m} = \frac{cd}{ab} = \frac{c(ax + by)}{ab} = (\frac{c}{b})x + \frac{c}{a})y = vx + uy.$$

This equation states that $m|c$, this implies, $m \leq c$. By the definition of least common multiple, we have $m = lcm(a, b)$. Hence the claim. Therefore $gcd(a, b)\ lcm(a, b) = ab$. □

Corollary 1.0.6. *For any choice of positive integers a and b, $lcm(a,b) = ab$ if and only if $gcd(a,b) = 1$.*

Definition 1.0.5. *If $a, b, c,$ are three integers, not all zero, $gcd(a,b,c)$ is defined to be the positive integer d having the following properties:*

(i) *d is a divisor of each of a, b, c.*

(ii) *If e divides the integers $a, b, c,$ then $e \leq d$.*

For example $gcd(39, 42, 54) = 3$ and $gcd(49, 210, 350) = 7$.

THE DIOPHANTINE EQUATION $ax + by = c$

The simplest type of Diophantine equation that we shall consider is the linear Diophantine equation in two unknowns:

$$ax + by = c,$$

where a, b, c are given integers and a, b are not both zero. A solution of this equation is a pair of integers $x_0,\ y_0$ that, when substituted into the equation, satisfy it; that is, we ask that $ax_0 + by_0 = c$.

Theorem 1.0.9. *The linear Diophantine equation $ax+by = c$ has a solution if and only if $d|c$, where $d = gcd(a,b)$. If $x_0,\ y_0$ is any particular solution of this equation, then all other solutions are given by*

$$x = x_0 + (\frac{b}{d})t \qquad y = y_0 - (\frac{a}{d})t,$$

where t is an arbitrary integer.

Example: Consider the linear Diophantine equation

$$172x + 20y = 1000$$

Applying the Euclidean's Algorithm to the evaluation of $gcd(172, 20)$, we find that

$$\begin{aligned} 172 &= 8.20 + 12 \\ 20 &= 1.12 + 8 \\ 12 &= 1.8 + 4 \\ 8 &= 24, \end{aligned}$$

whence $gcd(172, 20) = 4$. Because $4|1000$, a solution to this equation exists. To obtain the integer 4 as a linear combination of 172 and 20, we work backward through the previous calculations, as follows:

$$\begin{aligned} 4 &= 12 - 8 \\ &= 12 - (20 - 12) \\ &= 212 - 20 \\ &= 2(172 - 8.20) - 20 \\ &= 2.172 + (-17)20 \end{aligned}$$

Upon multiplying this relation by 250, we arrive at

$$\begin{aligned} 1000 &= 250.4 \\ &= 250(2.172 + (-17)20) \\ &= 500.172 + (-4250)20, \end{aligned}$$

so that $x = 500$ and $y = -4250$ provide one solution to the Diophantine equation in question. All other solutions are expressed by

$$x = 500 + (20/4)t = 500 + 5t$$

$$y = -4250 - (172/4)t = -4250 - 43t,$$

for some integer t.
If we want to find positive solution,if any happen to exist. For this, t must be chosen to satisfy simultaneously the inequalities

$$5t + 500 > 0 \qquad -43t - 4250 > 0$$

or

$$-98\frac{36}{43} > t > -100.$$

Because t must be an integer, we are forced to conclude that $t = -99$. Thus, our Diophantine equation has a unique positive solution $x = 5$, $y = 7$ corresponding to the value $t = -99$.

Corollary 1.0.7. *If $gcd(a, b) = 1$ and if x_0 , y_0 is a particular solution of the linear Diophantine equation $ax + by = c$, then all solutions are given by*

$$x = x_0 + bt \qquad y = Y_0 - at,$$

for integral values of t.

Problem 1: A customer bought a dozen pieces of fruit, apples and oranges, for $1.32. If an apple costs 3 rupees more than an orange and more apples than oranges were purchased, how many pieces of each kind were bought?
Solution: Let x be the number of apples and y be the number of oranges purchased. Also let z be the cost (in rupees) of an orange. Then the conditions of the problem lead to:

$$(z+3)x + zy = 132,$$

or equivalently

$$3x + (x+y)z = 132.$$

Because $x + y = 12$, the previous equation may be replaced by

$$3x + 12z = 132$$

this implies,

$$x + 4z = 44.$$

Now the problem is to find integers x and z satisfying the Diophantine equation $x + 4z = 44$. We have $gcd(1, 4) = 1$ is a divisor of 44, there is a solution to this equation. Upon multiplying the relation 1 = 1 (-3) + 4 1 by 44 to get

$$44 = 1(-132) + 4.44 \tag{1.1}$$

it follows that $x_0 = -132$, $z_0 = 44$, serves as one solution. All other solutions of Eq. (1.1) are of the form

$$x = -132 + 4t \qquad z = 44 - t,$$

where t is an integer.
Not all of the choices fort furnish solutions to the original problem. Only values of t that ensure $12 \geq x > 6$ should be considered. This requires obtaining those values of t such that

$$12 \geq -132 + 4t > 6.$$

Now,$12 \geq -132 + 4t$ implies that $t \leq 36$, whereas $-132 + 4t > 6$ gives $t > 34\frac{1}{2}$. The only integral values of t to satisfy both inequalities are $t = 35$ and $t = 36$. Thus, there are two possible purchases: a dozen apples costing 11 rupees apiece (the case where $t = 36$), or 8 apples at 12 rupees each and 4 oranges at 9 rupees each (the case where $t = 35$).
Problem 2: If a cock is worth 5 coins, a hen 3 coins, and three chicks together 1 coin, how many cocks, hens, and chicks, totaling 100, can be bought for 100 coins?

Solution: Let x be the number of cocks, y be the number of hens, z be the number of chicks. Then the conditions of the problem lead to

$$5x + 3y + \frac{1}{3}z = 100 \qquad x + y + z + 100.$$

Eliminating one of the unknowns, we are left with a linear Diophantine equation in the two other unknowns. Specifically, because the quantity $z = 100 - x - y$, we have $5x + 3y + \frac{1}{3}(100 - x - y) = 100$, or

$$7x + 4y = 100.$$

This equation has the general solution $x = 4t$, $y = 25-7t$, so that $z = 75+3t$, where t is an arbitrary integer. Some solutions are:

$$\begin{array}{cccccccc} x & = & 4 & y & = & 18 & z & = & 78 \\ x & = & 8 & y & = & 11 & z & = & 81 \\ x & = & 12 & y & = & 4 & z & = & 84 \end{array}$$

To obtain all solutions in the positive integers, t must be chosen to satisfy simultaneously the inequalities

$$4t > 0 \quad 25 - 7t > 0 \quad 75 + 3t > 0.$$

The last two of these are equivalent to the requirement $-25 < t < 3\frac{4}{7}$. Because t must have a positive value, we conclude that $t = 1, 2, 3$, leading to precisely the values given above.

Chapter 2

PRIMES AND THEIR DISTRIBUTION

THE FUNDAMENTAL THEOREM OF ARITHMETIC

Definition 2.0.6. *An integer $p > 1$ is called a prime number, or simply a prime, if its only positive divisors are 1 and p. An integer greater than 1 that is not a prime is termed composite.*

Among the first ten positive integers, 2, 3, 5, 7 are primes and 4, 6, 8, 9, 10 are composite numbers. Note that the integer 2 is the only even prime, and according to our definition the integer 1 plays a special role, being neither prime nor composite.

Theorem 2.0.10. *If p is a prime and $p|ab$, then $p|a$ or $p|b$.*

Proof. If $p|a$, then we need go no further, so let us assume that $p \nmid a$. Because the only positive divisors of p are 1 and p itself, this implies that $gcd(p, a) = 1$. Hence, by Euclid's lemma, we get $p|b$. □

Corollary 2.0.8. *If p is a prime and $p|a_1a_2 \cdots a_n$, then $p|a_k$ for some k, where $1 \leq k \leq n$.*

Proof. We proceed by induction on n, the number of factors. When $n = 1$, the stated conclusion obviously holds; whereas when $n = 2$, the result is the content of Theorem 10. Suppose, as the induction hypothesis, that $n > 2$ and that whenever p divides a product of less than n factors, it divides at least one of the factors. Now let $p|a_1a_2 \cdots a_n$. From Theorem 10, either $p|a_n$ or $p|a_1a_2 \cdots a_{n-1}$ If $p|a_n$, then we are through. As regards the case where $p|a_1a_2 \cdots a_{n-1}$, the induction hypothesis ensures that $p|a_k$ for some choice of k, with $1 \leq k \leq n - 1$. In any event, p divides one of the integers $a_1, a_2, \cdots, a_n$. □

Corollary 2.0.9. *If p, q_1, $q_2, \cdots$, q_n are all primes and $p|q_1q_2\cdots q_n$, then $p = q_k$ for some k, where $1 \leq k \leq n$.*

Proof. By Corollary 2.0.8, we know that $p|q_k$ for some k, with $1 \leq k \leq n$. Being a prime, q_k is not divisible by any positive integer other than 1 or q_k itself. Because $p > 1$, we are forced to conclude that $p = q_k$. □

Theorem 2.0.11. *(Fundamental Theorem of Arithmetic.) Every positive integer $n > 1$ can be expressed as a product of primes; this representation is unique, apart from the order in which the factors occur.*

Proof. Either n is a prime, there is nothing to prove. If n is composite, then there exists an integer d satisfying $d|n$ and $1 < d < n$. Among all such integers d, choose p_1 to be the smallest (this is possible by the Well-Ordering Principle). Then P_1 must be a prime number. Otherwise it too would have a divisor q with $1 < q < p_1$; but then $q|p_1$ and $p_1|n$ imply that $q|n$, which contradicts the choice of p_1 as the smallest positive divisor, not equal to 1, of n. We therefore may write $n = p_1n_1$, where p_1 is prime and $1 < n_1 < n$.If n_1 happens to be a prime, then we have our representation. In the contrary case, the argument is repeated to produce a second prime number p_2 such that $n_1 = p_2n_2$; that is,

$$n = p_1P_2n_2 \qquad 1 < n_2 < n_1.$$

If n_2 is a prime,then it is not necessary to go further. Otherwise, write $n_2 = p_3n_3$, with p_3 a prime:

$$n = p_1P_2p_3n_3 \qquad 1 < n_3 < n_2.$$

The decreasing sequence $n > n_1 > n_2 > \cdots > 1$ cannot continue indefinitely, so that after a finite number of steps n_{k-1} is a prime, call it, p_k. This leads to the prime factorization

$$n = p_1p_2\cdots p_k.$$

To establish the second part of the proof-the uniqueness of the prime factorization, let us suppose that the integer n can be represented as a product of primes in two ways, say,

$$n = p_1p_2\cdots p_r = q_1q_2\cdots q_s \qquad r \leq s,$$

where the p_i and q_j are all primes, written in increasing magnitude so that

$$p_1 \leq p_2 \leq \cdots \leq p_r \qquad q_1 \leq q_2 \leq \cdots \leq q_s.$$

Because $p_1|q_1q_2\cdots q_s$, Corollary 9 tells us that $p_1 = q_k$ for some k; but then $p_1 \geq q_1$. Similar reasoning gives $q_1 \geq p_1$, whence $p_1 = q_1$. We may cancel this common factor and obtain

$$p_2p_3\cdots p_r = q_2q_3\cdots q_s.$$

Now repeat the process to get $p_2 = q_2$ and, in turn,

$$p_3p_4\cdots p_r = q_3q_4\cdots q_s.$$

Continue in this fashion. If the inequality $r < s$ were to hold, we would eventually arrive at

$$1 = q_{r+1}q_{r+2}\cdots q_s,$$

which is absurd, because each $q_j > 1$. Hence, $r = s$ and

$$p_1 = q_1,\ p_2 = q_2,\ \cdots, p_r = q_r,$$

making the two factorizations of n identical. The proof is now complete. □

Corollary 2.0.10. *Any positive integer $n > 1$ can be written uniquely in a canonical form*

$$n = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r},$$

where, for $i = 1, 2, \cdots, r$, each k_i is a positive integer and each p_i is a prime, with $p_1 < p_2 < \cdots < p_r$.

Theorem 2.0.12. *(Pythagoras.) The number $\sqrt{2}$ is irrational.*

Proof. Suppose, to the contrary, that $\sqrt{2}$ is a rational number, say, $\sqrt{2} = a/b$, where a and b are both integers with $gcd(a, b) = 1$. Squaring, we get $a^2 = 2b^2$, so that $b|a^2$. Claim: $b = 1$ If $b > 1$, then the Fundamental Theorem of Arithmetic guarantees the existence of a prime p such that $p|b$. It follows that $p|a^2$. This implies that $p|a$; hence, $gcd(a, b) \geq p$. We therefore arrive at a contradiction, unless $b = 1$, hence the claim.
But if this happens, then $a^2 = 2$, which is impossible. Our supposition that $\sqrt{2}$ is a rational number is not true, and so $\sqrt{2}$ must be irrational. □

THE SIEVE OF ERATOSTHENES

Given a particular integer, how can we determine whether it is prime or composite and, in the latter case, how can we actually find a nontrivial divisor? The most obvious approach consists of successively dividing the integer in question by each of the numbers preceding it; if none of them (except 1) serves as a divisor, then the integer must be prime. Although this method

is very simple to describe, it cannot be regarded as useful in practice. For even if one is undaunted by large calculations, the amount of time and work involved may be prohibitive.

Theorem 2.0.13. *(Euclid.) There is an infinite number of primes.*

Proof. Euclid's proof is by contradiction. Let $p_1 = 2,\ p_2 = 3,\ p_3 = 5,\ p_4 = 7, \cdots$ be the primes in ascending order, and suppose that there is a last prime, called p_n Now consider the positive integer $P = p_1 p_2 \cdots p_{n+1}$. Because $P > 1$, we have P is divisible by some prime p. But $p_1,\ p_2, \cdots, p_n$ are the only prime numbers, so that p must be equal to one of $p_1, p_2, ..., p_n$. Combining the divisibility relation $p|p_1, p_2, \cdots, p_n$ with $p|P$, we arrive at $p|P - p_1, p_2, \cdots, p_n$ or, equivalently, $p|1$. The only positive divisor of the integer 1 is 1 itself and, because $p > 1$, a contradiction arises. Thus, no finite list of primes is complete, whence the number of primes is infinite. □

Definition 2.0.7. *For a prime p, define $p^{\#}$ = the product of all primes that are less than or equal to p. Numbers of the form $p^{\#} + 1$ called Euclidean numbers.*

Note: Not all Euclidean numbers are primes. For example, $13^{\#}+1 = 59.509$, $17^{\#} + 1 = 19.97.277$, $19^{\#} + 1 = 347.27953$.

Theorem 2.0.14. *If p_n is the nth prime number, then $p_n \leq 2^{2^{n-l}}$.*

Proof. Proof is by induction on n.When $n = 1$, $p_n = p_1 = 2$ and $2^{2^{n-1}} = 2^{2^{1-1}} = 2^{2^0} = 2^1 = 2$, the result is true.Suppose that $n > 1$ and that the result holds for all integers up to n. Then

$$\begin{array}{rcl} p_{n+1} & \leq & p_1 p_2 \cdots p_n + 1 \\ & \leq & 22^2 \cdots 2^{n-1} + 1 \quad = \quad 2^{1+2+2^2+\cdots+2^{n-1}} + 1. \end{array}$$

Recalling the identity $1 + 2 + 2^2 + \ + 2^{n-1} = 2^n - 1$, we obtain

$$p_{n+1} \leq 2^{2^n - 1} + 1.$$

However, $1 \leq 2^{2^n - 1}$ for all n; whence

$$\begin{array}{rcl} p_{n+1} & \leq & 2^{2^n-1} + 2^{2^n-1} \\ & = & 2.2^{2^n-1} \quad = \quad 2^{2^n} \end{array}$$

completing the induction step, and the argument. □

Corollary 2.0.11. *For $n \geq 1$, there are at least $n + 1$ primes less than 2^{2^n}.*

Proof. From the theorem, we know that $p_1, p_2, \cdots, p_{n+1}$ are all less than 2^{2^n}. □

Chapter 3

THE THEORY OF CONGRUENCES

Definition 3.0.8. *Let n be a fixed positive integer. Two integers a and b are said to be congruent modulo n, symbolized by*

$$a \equiv b(mod n)$$

if n divides the difference $a - b$; that is, provided that $a - b = kn$ for some integer k.

Theorem 3.0.15. *For arbitrary integers a and b, $a \equiv b(mod n)$ if and only if a and b leave the same nonnegative remainder when divided by n.*

Proof. Suppose $a \equiv b(mod n)$, so that $a = b + kn$ for some integer k. Upon division by n, b leaves a certain remainder r; that is, $b = qn + r$, where $0 \leq r < n$. Therefore,

$$a = b + kn = (qn + r) + kn = (q + k)n + r$$

which indicates that a has the same remainder as b.
On the other hand, suppose we can write $a = q_1 n + r$ and $b = q_2 n + r$, with the same remainder r $(0 \leq r < n)$. Then

$$a - b = (q_1 n + r) - (q_2 n + r) = (q_1 - q_2)n,$$

whence $n|a - b$. That is, $a \equiv b(mod n)$. □

Theorem 3.0.16. *Let $n > 1$ be fixed and a, b, c, d be arbitrary integers. Then the following properties hold:*

1. $a \equiv a(mod n)$.

2. *If* $a \equiv b(modn)$, *then* $b \equiv a(modn)$.

3. *If* $a \equiv b(modn)$ *and* $b \equiv c(modn)$, *then* $a \equiv c(modn)$.

4. *If* $a \equiv b(modn)$ *and* $c \equiv d(modn)$, *then* $a + c \equiv b + d(modn)$ *and* $ac \equiv bd(modn)$.

5. *If* $a \equiv b(modn)$, *then* $a + c \equiv b + c(modn)$ *and* $ac \equiv bc(modn)$.

6. *If* $a \equiv b(modn)$, *then* $a^k \equiv b^k(modn)$ *for any positive integer* k.

Problem 1: Show that $41|2^{20} - 1$.
Solution: We have

$$2^5 \equiv -9(mod\ 41).$$

Therefore

$$(2^5)^4 \equiv (-9)^4(mod\ 41).$$

This implies that

$$2^{20} \equiv (-9)^4(mod\ 41).$$

But we have $(-9)^4 = 81.81$ and $81 \equiv -1(mod\ 41)$. Therefore

$$2^{20} \equiv (-1)(-1)(mod\ 41).$$

This implies $41|2^{20} - 1$.
Problem 2: Find the remainder obtained upon dividing the sum

$$1! + 2! + 3! + 4! + \cdots + 99! + 100!$$

by 12.
Solution: We have $4! \equiv 24 \equiv 0(mod\ 12)$; thus, for $k \geq 4$,

$$k! \equiv 4!.5.6 \cdots k \equiv 0.5..6 \cdots k \equiv 0(mod\ 12).$$

Therefore

$$1! + 2! + 3! + 4! + \cdots + 100! \equiv 1! + 2! + 3! + 0 + \cdots + 0 \equiv 9(mod\ 12).$$

The remainder 9.

Theorem 3.0.17. *If* $ca \equiv cb(mod\ n)$, *then* $a \equiv b(mod\ n/d)$, *where* $d = gcd(c, n)$.

Proof. By hypothesis, we can write

$$c(a-b) = ca - cb = kn, \tag{3.1}$$

for some integer k. Knowing that $gcd(c,n) = d$, there exist relatively prime integers r and s satisfying $c = dr$, $n = ds$. When these values are substituted in Eq. 3.1 and the common factor d canceled, the net result is

$$r(a-b) = ks.$$

Hence, $s|r(a-b)$ and $gcd(r,s) = 1$. Euclid's lemma yields $s|(a-b)$, which implies $a \equiv b(mod\ s)$; in other words, $a \equiv b(mod\ n/d)$. □

Corollary 3.0.12. *If $ca \equiv cb(mod\ n)$ and $gcd(c,n) = 1$, then $a \equiv b(mod\ n)$.*

Corollary 3.0.13. *If $ca \equiv cb(mod\ p)$ and $p \nmid c$, where p is a prime number, then $a \equiv b(mod\ p)$.*

Proof. The conditions $p \nmid c$ and p a prime imply that $gcd(c,p) = 1$. Then by Corollary 12, $a \equiv b(mod\ p)$. □

BINARY AND DECIMAL REPRESENTATIONS OF INTEGERS.

One of the more interesting applications of congruence theory involves finding special criteria under which a given integer is divisible by another integer. At their heart, these divisibility tests depend on the notational system used to assign "names" to integers and, more particularly, to the fact that 10 is taken as the base for our number system. Let us, therefore, start by showing that, given an integer $b > 1$, any positive integer N can be written uniquely in terms of powers of b as

$$N = a_m b^m + a_{m-l} b^{m-l} + \cdots + a_2 b^2 + a_1 b + a_0$$

where the coefficients a_k can take on the b different values $0, 1, 2, \cdots, b-1$. For the Division Algorithm yields integers q_1 and a_0 satisfying

$$N = q_1 b + a_0 \qquad 0 \leq a_0 < b.$$

If $q_1 \geq b$, we can divide once more, obtaining

$$q_1 = q_2 b + a_1 \qquad 0 \geq a_1 < b.$$

Now substitute for q_1 in the earlier equation to get

$$N = (q_2 b + a_1)b + a_0 = q_2 b^2 + a_1 b + a_0.$$

Because $N > q_1 > q_2 > \cdots \geq 0$ is a strictly decreasing sequence of integers, this process must eventually terminate, say, at the $(m-1)$th stage, where

$$q_{m-1} = q_m b + a_{m-1} \qquad 0 \leq a_{m-1} < b$$

and $0 \leq q_m < b$. Setting $a_m = q_m$, we reach the representation

$$N = a_m b^m + a_{m-l} b_{m-l} + \cdots + a_1 b + a_0.$$

which was our aim. To show uniqueness, refer text.

Theorem 3.0.18. *Let $P(x) = \sum_{k=0}^{m} c_k x^k$ be a polynomial function of x with integral coefficients c_k. If $a \equiv b(mod\ n)$, then $P(a) \equiv P(b)(mod\ n)$.*

Proof. Please refer text. □

Theorem 3.0.19. *If a is a solution of $P(x) \equiv 0(mod\ n)$ and $a \equiv b(mod\ n)$, then b also is a solution.*

Proof. From the last theorem, it is known that $P(a) \equiv P(b)(mod\ n)$. Hence, if a is a solution of $P(x) \equiv 0(mod\ n)$, then $P(b) \equiv P(a) \equiv 0(mod\ n)$, making b a solution. □

Theorem 3.0.20. *Let $N = a_m 10^m + a_{m-1}, 10^{m-1} + \cdots + a_1 10 + a_0$ be the decimal expansion of the positive integer N, $0 \leq a_k < 10$, and let $S = a_0 + a_1 + \cdots + a_m$. Then $9|N$ if and only if $9|S$.*

Proof. Please refer text. □

Theorem 3.0.21. *Let $N = a_m 10^m + a_{m-1} 10^{m-1} + \cdots + a_1 10 + a_0$ be the decimal expansion of the positive integer N, $0 \leq a_k < 10$, and let $T = a_0 - a_1 + a_2 - \cdots + (-1)^m a_m$. Then $11|N$ if and only if $11|T$.*

Proof. Please refer text. □

LINEAR CONGRUENCES AND THE CHINESE REMAINDER THEOREM.

Theorem 3.0.22. *The linear congruence $ax \equiv b(mod\ n)$ has a solution if and only if $d|b$, where $d = gcd(a, n)$. If $d|b$, then it has d mutually incongruent solutions modulo n.*

Theorem 3.0.23. *Chinese Remainder Theorem: Let $n_1, n_2, \cdots, n$, be positive integers such that $gcd(n_i, n_j) = 1$ for $i \neq j$. Then the system of linear congruences*

$$\begin{aligned} x &\equiv a_1 (mod\, n_1) \\ x &\equiv a_2 (mod\, n_2) \\ &\vdots \\ x &\equiv a_r (mod\, n_r) \end{aligned}$$

has a simultaneous solution, which is unique modulo the integer $n_1, n_2, \cdots, n_r$.

Problem 1: Solve

$$\begin{aligned} x &\equiv 2 (mod\, 3) \\ x &\equiv 3 (mod\, 5) \\ x &\equiv 2 (mod\, 7). \end{aligned}$$

Solution: We have $n = 3\ .\ 5\ .7 = 105$ and

$$N_1 = \frac{n}{3} = 35 \qquad N_2 = \frac{n}{5} = 21 \qquad N_3 = \frac{n}{7} = 15$$

Now the linear congruences

$$35x \equiv 1 (mod\, 3) \qquad 21x \equiv 1 (mod\, 5) \qquad 15x \equiv 1 (mod\, 7)$$

are satisfied by $x_1 = 2$, $x_2 = 1$, $x_3 = 1$, respectively. Thus, a solution of the system is given by

$$x = 2352 + 3211 + 2151 = 233$$

Modulo 105, we get the unique solution $x = 233 = 23 (mod\, 105)$.

MODULE II

Chapter 4

FERMAT'S THEOREM

FERMAT'S LITTLE THEOREM AND PSEUDOPRIMES

The most significant of Fermat's correspondents in number theory was Bernhard Frenicle de Bessy (1605-1675), an official at the French mint who was renowned for his gift of manipulating large numbers. (Frenicle 's facility in numerical calculation is revealed by the following incident: On hearing that Fermat had proposed the problem of finding cubes that when increased by their proper divisors become squares, as is the case with $7^3+(1+7+72) = 20^2$, he immediately gave four different solutions, and supplied six more the next day.) Though in no way Fermat's equal as a mathematician, Frenicle alone among his contemporaries could challenge Fermat in number theory and Frenicle's challenges had the distinction of coaxing out of Fermat some of his carefully guarded secrets. One of the most striking is the theorem that states: If p is a prime and a is any integer not divisible by p, then p divides $a^{p-1} - 1$. Fermat communicated the result in a letter to Frenicle dated October 18, 1640, along with the comment, "I would send you the demonstration, if I did not fear its being too long." This theorem has since become known as "Fermat's Little Theorem," or just "Fermat's Theorem," to distinguish it from Fermat's "Great" or "Last Theorem," which is the subject of Chapter 12. Almost 100 years were to elapse before Euler published the first proof of the little theorem in 1736. Leibniz, however, seems not to have received his share of recognition, for he left an identical argument in an unpublished manuscript sometime before 1683.
We now proceed to a proof of Fermat's theorem.

Theorem 4.0.24. *Fermat's theorem. Let p be a prime and suppose that $p|a$. Then*

$$a^{P-1} \equiv 1(mod\ p).$$

Proof. We begin by considering the first $p-1$ positive multiples of a; that is, the integers

$$a, 2a, 3a, \cdots, (p-1)a$$

None of these numbers is congruent modulo p to any other, nor is any congruent to zero. Indeed, if it happened that

$$ra \equiv sa(mod\ p) \qquad 1 \leq r < s \leq p-1$$

then a could be canceled to give $r \equiv s(mod\ p)$, which is impossible. Therefore, the previous set of integers must be congruent modulo p to $1, 2, 3, \cdots, p-1$, taken in some order. Multiplying all these congruences together, we find that

$$a.2a.3a \cdots (p-1)a \equiv 1.2.3 \cdots (p-1)(mod\ p)$$

whence

$$a^{p-1}(p-1)! \equiv (p-1)!(mod\ p).$$

Once $(p-1)!$ is canceled from both sides of the preceding congruence (this is possible because since $p|(p-1)!$), our line of reasoning culminates in the statement that $a^{p-l} = 1(mod\ p)$, which is Fermat's theorem. □

This result can be stated in a slightly more general way in which the requirement that $p \nmid a$ is dropped.

Corollary 4.0.14. *If p is a prime, then $a^p = a(mod\ p)$ for any integer a.*

Proof. When $p|a$, the statement obviously holds; for, in this setting, $a^p \equiv 0 \equiv a(mod\ p)$. If $p|a$, then according to Fermat's theorem, we have $a^p - l \equiv 1(mod\ p)$. When this congruence is multiplied by a, the conclusion $a^p \equiv a(mod\ p)$ follows. □

Lemma 4.0.2. *If p and q are distinct primes with $a^p \equiv a(mod\ q)$ and $a^q \equiv a(mod\ p)$, then $a^{pq} = a(mod\ pq)$.*

Proof. The last corollary tells us that $(a^q)^p = a^q(mod\ p)$, whereas $a^q = a(mod\ p)$ holds by hypothesis. Combining these congruences, we obtain $a^{pq} = a(mod\ p)$. or, in different terms, $p|a^{pq} - a$. In similar manner, $q|a^{pq} - a$. Therefore $pq|a^{pq} - a$, that is, $a^{pq} = a(mod\ pq)$. □

Definition 4.0.9. *A composite integer n is called pseudoprime whenever $n|2n-2$. It can be shown that there are infinitely many pseudoprimes, the smallest four being 341, 561, 645, and 1105.*

Theorem 4.0.25. *If n is an odd pseudoprime, then $M_n = 2^n - 1$ is a larger one.*

Proof. Please refer text. □

WILSON'S THEOREM

We now turn to another milestone in the development of number theory. In his Meditationes Algebraicae of 1770, the English mathematician Edward Waring (1734-1798) announced several new theorems. Foremost among these is an interesting property of primes reported to him by one of his former students, a certain John Wilson. The property is the following: If p is a prime number, then p divides $(p-1)!+1$. Wilson appears to have guessed this on the basis of numerical computations; at any rate, neither he nor Waring knew how to prove it. Confessing his inability to supply a demonstration, Waring added, "Theorems of this kind will be very hard to prove, because of the absence of a notation to express prime numbers." (Reading the passage, Gauss uttered his telling comment on "notationes versus notiones," implying that in questions of this nature it was the notion that really mattered, not the notation.) Despite Waring's pessimistic forecast, soon afterward Lagrange (1771) gave a proof of what in literature is called "Wilson's theorem" and observed that the converse also holds. Perhaps it would be more just to name the theorem after Leibniz, for there is evidence that he was aware of the result almost a century earlier, but published nothing on the subject.
Now we give a proof of Wilson's theorem.

Theorem 4.0.26. *Wilson. If p is a prime, then $(p-1)! \equiv -1(mod\ p)$.*

Proof. When $p=2$ and $p=3$ are trivial, let us take $p>3$. Suppose that a is any one of the $p-1$ positive integers $1,2,3,\cdots,p-1$ and consider the linear congruence $ax \equiv 1(mod\ p)$. Then $gcd(a,p)=1$. Therefore this congruence admits a unique solution modulo p; hence, there is a unique integer a', with $1 \leq a' \leq p-1$, satisfying $aa' \equiv 1(mod\ p)$. Because p is prime, $a=a'$ if and only if $a=1$ or $a=p-1$. Indeed, the congruence $a^2 \equiv 1(mod\ p)$ is equivalent to $(a-1).(a+1) = 0(mod\ p)$. Therefore, either $a-1 \equiv 0(mod\ p)$, in which case $a=1$, or $a+1 \equiv 0(mod\ p)$, in which case $a=p-1$.
If we omit the numbers 1 and $p-1$, the effect is to group the remaining integers $2,3,\cdots,p-2$ into pairs a,a', where $a \neq a'$, such that their product $aa' \equiv 1(mod\ p)$. When these $(p-3)/2$ congruences are multiplied together and the factors rearranged, we get

$$2.3\cdots(p-2) \equiv 1(mod\ p)$$

or rather

$$(p-2)! \equiv 1(mod\ p)$$

Now multiply by $p-1$ to obtain the congruence

$$(p-1)! \equiv p-1 \equiv -1(mod\ p)$$

this completes the proof. □

Theorem 4.0.27. *The quadratic congruence* $x^2+1 \equiv 0(mod\ p)$, *where* p *is an odd prime, has a solution if and only if* $p \equiv 1(mod 4)$.

Proof. Please refer text. □

Chapter 5

NUMBER-THEORETIC FUNCTIONS

THE SUM AND NUMBER OF DIVISORS

Certain functions are found to be of special importance in connection with the study of the divisors of an integer. Any function whose domain of definition is the set of positive integers is said to be a number-theoretic (or arithmetic)function. Although the value of a number-theoretic function is not required to be a positive integer or, for that matter, even an integer, most of the number-theoretic functions that we shall encounter are integer-valued. Among the easiest to handle, and the most natural, are the functions τ and σ.

Definition 5.0.10. *Given a positive integer n, let $\tau(n)$ denote the number of positive divisors of n and $\sigma(n)$ denote the sum of these divisors.*

For an example of these notions, consider $n = 12$. Because 12 has the positive divisors 1, 2, 3, 4, 6, 12, we find that $\tau(12) = 6$ and $\sigma(12) = 1 + 2 + 3 + 4 + 6 + 12 = 28$ For the first few integers,

$$\tau(l) = 1\tau(2) = 2\tau(3) = 2\tau(4) = 3\tau(5) = 2\tau(6) = 4, \cdots$$

and

$$\sigma(1) = 1, \sigma(2) = 3, \sigma(3) = 4, \sigma(4) = 7, \sigma(5) = 6, \sigma(6) = 12, \cdots$$

It is not difficult to see that $\tau(n) = 2$ if and only if n is a prime number; also, $\sigma(n) = n + 1$ if and only if n is a prime.

Theorem 5.0.28. *If $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ is the prime factorization of $n > 1$, then the positive divisors of n are precisely those integers d of the form*

$$d = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$$

where $0 \leq a_i \leq k_i \quad (i = 1, 2, ..., r)$.

Proof. Please refer text. □

Theorem 5.0.29. *If $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ is the prime factorization of $n > 1$, then*

1. *$\tau(n) = (k_1 + 1)(k_2 + 1) \cdots (k_r + 1)$, and*
2. *$\sigma(n) = \frac{p_1^{k_1+1}-1}{p_1-1} \frac{p_2^{k_2+1}-1}{p_2-1} \frac{p_r^{k_r+1}-1}{p_r-1}$*

Proof. Please refer text. □

Problem: Find the number of positive divisors and their sum of 180.
Solution: The number $180 = 2^2 \,.\, 3^2 \,.\, 5$ has

$$\tau(180) = (2 + 1)(2 + 1)(1 + 1) = 18$$

positive divisors. These are integers of the form

$$2^{a_1} \,.\, 3^{a_2} \,.\, 5^{a_3}$$

where $a_1 = 0, 1, 2$; $a_2 = 0, 1, 2$; and $a_3 = 0, 1$. Specifically, we obtain $1, 2, 3, 4, 5, 6, 9, 10, 12, 15, 18, 20, 30, 36, 45, 60, 90, 180$.
The sum of these integers is

$$\sigma(180) = \frac{2^3 - 1}{2 - 1} \frac{3^3 - 1}{3 - 1} \frac{5^2 - 1}{5 - 1} = \frac{7}{1} \frac{26}{2} \frac{24}{4} = 7 \,.\, 13 \,.\, 6 = 546$$

Definition 5.0.11. *A number-theoretic function f is said to be multiplicative if*

$$f(mn) = f(m)f(n)$$

whenever $gcd(m, n) = 1$.

Example: The functions τ and σ are both multiplicative functions.

THE GREATEST INTEGER FUNCTION

The greatest integer or "bracket" function [ ] is especially suitable for treating divisibility problems. Although not strictly a number-theoretic function, its study has a natural place in this chapter.

Definition 5.0.12. *For an arbitrary real number x, we denote by $[x]$ the largest integer less than or equal to x; that is, $[x]$ is the unique integer satisfying $x - 1 < [x] \leq x$.*

$$[-3/2] = -2 \qquad [\sqrt{2}] = 1 \qquad [1/3] = 0 \qquad [\pi] = 3[-\pi] = -4$$

The important observation to be made here is that the equality $[x] = x$ holds if and only if x is an integer. Also from the definition we have any real number x can be written as

$$x = [x] + \theta$$

for a suitable choice of θ with $0 \leq \theta < 1$.

Results:

1. If n is a positive integer and p a prime, then the exponent of the highest power of p that divides $n!$ is

$$\sum_{k=1}^{\infty} \left[\frac{n}{p^k} \right]$$

   where the series is finite, because $[n/p^k] = 0$ for $p^k > n$.

2. If n and r are positive integers with $1 \leq r < n$, then the binomial coefficient

$$\left(\begin{array}{c} n \\ r \end{array} \right) = \frac{n!}{r!(n-r)!}$$

   is also an integer.

3. For a positive integer r, the product of any r consecutive positive integers is divisible by $r!$.

4. Let f and F be number-theoretic functions such that

$$F(n) = \sum_{d|n} f(d).$$

   Then, for any positive integer N,

$$\sum_{n=1}^{N} F(n) = \sum_{k=1}^{N} f(k) \left[\frac{N}{k} \right].$$

5. If N is a positive integer, then

$$\sum_{n=1}^{N} \tau(n) = \sum_{n=1}^{N} \left[\frac{N}{n} \right].$$

6. If N is a positive integer, then

$$\sum_{n=1}^{N} \sigma(n) = \sum_{n=1}^{N} n \left[\frac{N}{n} \right].$$

Chapter 6

EULER'S GENERALIZATION OF FERMAT'S THEOREM

EULER'S PHI-FUNCTION

This chapter deals with that part of the theory arising out of the result known as Euler's Generalization of Fermat's Theorem. In a nutshell, Euler extended Fermat's theorem, which concerns congruences with prime moduli, to arbitrary moduli. While doing so, he introduced an important number-theoretic function:

Definition 6.0.13. *For $n \geq 1$, let $\phi(n)$ denote the number of positive integers not exceeding n that are relatively prime to n.*

As an illustration of the definition, we find that $\phi(30) = 8$; for, among the positive integers that do not exceed 30, there are eight that are relatively prime to 30; specifically,

$$1, 7, 11, 13, 17, 19, 23, 29.$$

Similarly, for the first few positive integers, $\phi(l) = 1, \phi(2) = 1, \phi(3) = 2, \phi(4) = 2, \phi(5) = 4, \phi(6) = 2, \phi(7) = 6, \cdots$
Notice that $\phi(1) = 1$, because $gcd(1, 1) = 1$. In the event $n > 1$, then $gcd(n, n) = n \neq 1$, so that $\phi(n)$ can be characterized as the number of integers less than n and relatively prime to it. The function $\phi(n)$ is usually called the Euler phi-function (sometimes, the indicator or totient) after its originator; the functional notation $\phi(n)$, however, is credited to Gauss.
If n is a prime number, then every integer less than n is relatively prime to it; whence, $\phi(n) = n - 1$. On the other hand, if $n > 1$ is composite, then n has a divisor d such that $1 < d < n$. It follows that there are at least two

integers among $1, 2, 3, \cdots, n$ that are not relatively prime to n, namely, d and n itself. As a result, $\phi(n) \leq n - 2$. This proves that for $n > 1$,

$$\phi(n) = n - 1 \quad \textit{if and only if } n \textit{ is prime.}$$

Results:

1. If p is a prime and $k > 0$, then
$$\phi(p^k) = p^k - p^{k-1} = p^k \left(1 - \frac{1}{p}\right).$$
2. Given integers a, b, c, $gcd(a, bc) = 1$ if and only if $gcd(a, b) = 1$ and $gcd(a, c) = 1$.
3. The function ϕ is a multiplicative function.
4. If the integer $n > 1$ has the prime factorization $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$, then
$$\begin{aligned}\phi(n) &= (p_1^{k_1} - p_1^{k_1-1})(p_2^{k_2} - p_2^{k_2-1}) \cdots (p_r^{k_r} - p_r^{k_r-1}) \\ &= n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_r}\right)\end{aligned}$$
5. For $n > 2$, $\phi(n)$ is an even integer.

Problem 1: Find $\phi(16)$.
Solution: $\phi(16) = \phi(2^4) = 2^4 - 2^3 = 16 - 8 = 8$.
Problem 2: Find $\phi(360)$.
Solution: The prime-power decomposition of 360 is $2^3 \cdot 3^2 \cdot 5$, therefore

$$\begin{aligned}\phi(360) &= 360(1 - \frac{1}{2})(1 - \frac{1}{3})(1 - \frac{1}{5}) \\ &= 360 \cdot \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{4}{5} = 96.\end{aligned}$$

EULER'S THEOREM

Lemma 6.0.3. *Let $n > 1$ and $gcd(a, n) = 1$. If $a_1, a_2, \cdots, a_{\phi(n)}$ are the positive integers less than n and relatively prime to n, then*

$$aa_1,\ aa_2,\ \cdots a_{\phi(n)}$$

are congruent modulo n to $a_1, a_2, \cdots, a_{\phi(n)}$ in some order.

Proof. Observe that no two of the integers $a_1, a_2, \cdots, a_{\phi(n)}$ are congruent modulo n. For if $aa_i \equiv aa_j (mod\ n)$, with $1 \leq i < j \leq \phi(n)$, then the cancellation law yields $a_i \equiv a_j (mod\ n)$, and thus $a_i = a_j$, a contradiction. Furthermore, because $gcd(a_i, n) = 1$ for all i and $gcd(a, n) = 1$, this implies that each of the aa_i is relatively prime to n.
Fixing on a particular aa_i, there exists a unique integer b, where $0 \leq b < n$, for which $aa_i \equiv b(mod\ n)$. Because $gcd(b, n) = gcd(aa_i, n) = 1$, b must be one of the integers $a_1, a_2, \cdots, a_{\phi(n)}$. All told, this proves that the numbers $aa_1,\ aa_2,\ \cdots a_{\phi(n)}$ and the numbers $a_1, a_2, \cdots, a_{\phi(n)}$ are identical modulo n in a certain order. □

Theorem 6.0.30. *Euler. If $n \geq 1$ and $gcd(a, n) = 1$, then $a^{\phi(n)} \equiv 1\ (mod\ n)$.*

Corollary 6.0.15. *Fermat. If p is a prime and $p \nmid a$, then $a^{p-1} \equiv 1(mod\ p)$.*

Problem: Find the last two digits in the decimal representation of 3^{256}.
Solution: This is equivalent to obtaining the smallest nonnegative integer to which 3^{256} is congruent modulo 100. Because $gcd(3, 100) = 1$ and

$$\phi(100) = \phi(2^2 \,.\, 5^2) = 100\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{5}\right) = 40.$$

Euler's theorem yields

$$3^{\phi(100)} \equiv 1(mod\ 100)$$

$$3^40 \equiv 1(mod\ 100)$$

By the Division Algorithm, $256 = 6 \,.\, 40 + 16$; whence

$$3^{256} \equiv 3^{6\,.\,40+16} \equiv (3^{40})^6\ 3^{16} \equiv 3^{16}(mod\ 100)$$

and our problem reduces to one of evaluating 3^{16}, modulo 100. The method of successive squaring yields the congruences

$$3^2 \equiv 9(mod\ 100) \qquad 3^4 \equiv 81(mod\ 100)$$

$$3^8 \equiv 61(mod\ 100) \qquad 3^{16} \equiv 21(mod\ 100).$$

Hence the last two digits of 3^{256} is 21.

SOME PROPERTIES OF THE PHI-FUNCTION

The next theorem points out a curious feature of the phi-function; namely, that the sum of the values of $\phi(d)$, as d ranges over the positive divisors of n, is equal to n itself. This was first noticed by Gauss.

Theorem 6.0.31. *For each positive integer $n \geq 1$,*

$$n = \sum_{d|n} \phi(d)$$

the sum being extended over all positive divisors of n.

Theorem 6.0.32. *For $n > 1$, the sum of the positive integers less than n and relatively prime to n is $\frac{1}{2}n\phi(n)$.*

MODULE III

Chapter 7

VECTOR SPACE

Definition 7.0.14. *By a vector space we shall mean a set V on which there are defined two operations, one called 'addition' and the other called 'multiplication by scalars', such that the following properties hold:*

(V_1) $x + y = y + x$ *for all* $x, y \in V$;

(V_2) $(x + y) + z = x + (y + z)$ *for all* $x, y, z \in V$;

(V_3) *there exists an element* $0 \in V$ *such that* $x + 0 = x$ *for every* $x \in V$;

(V_4) *for every* $x \in V$ *there exists* $-x \in V$ *such that* $x + (-x) = 0$;

(V_5) $\lambda(x + y) = \lambda x + \lambda y$ *for all* $x, y \in V$ *and all scalars* λ;

(V_6) $(\lambda + \mu)x = \lambda x + \mu x$ *for all* $x \in V$ *and all scalars* λ, μ;

(V_7) $(\lambda\mu)x = \lambda(\mu x)$ *for all* $x \in V$ *and all scalars* λ, μ;

(V_8) $1x = x$*for all* $x \in V$.

When the scalars are all real numbers we shall often talk of a real vector space; and when the scalars are all complex numbers we shall talk of a complex vector space.

* It should be noted that in the definition of a vector space the scalars need not be restricted to be real or complex numbers. They can in fact belong to any 'field' F (which may be regarded informally as a number system in which every nonzero element has a multiplicative inverse). Although in what follows we shall find it convenient to say that 'V is a vector space over a field F' to indicate that the scalars come from a field F, we shall in fact normally assume (that is, unless explicitly mentioned otherwise) that F is either the field $\mathbb{R}$ of real numbers or the field C$\mathbb{C}$ of complex numbers .

** Axioms (V_1) to (V_4) above can be summarized by saying that the algebraic structure $(V;+)$ is an abelian group. If we denote by F the field of scalars (usually $\mathbb{R}$ or $\mathbb{C}$) then multiplication by scalars can be considered as an action by F on V, described by $(\lambda, x) \to \lambda x$, which relates the operations in F (addition and multiplication) to that of V (addition) in the way described by the axioms (V_5) to (V_8).

Example 7.0.1. *$Mat_{m\times n}\mathbb{R}$, the set of all $m \times n$ matrices with real entries is a real vector space under the usual operations of addition of matrices and multiplication by scalars.*

Example 7.0.2. *The set $\mathbb{R}^n$ of n-tuples $(x_1, x_2, ..., x_n)$ of real numbers is a real vector space under the following component-wise definitions of addition and mutiplication by scalars;*

$$(x_1, x_2, ..., x_n) + (y_1, y_2, ..., y_n) = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n),$$

$$\lambda(x_1, x_2, ..., x_n) = (\lambda x_1, \lambda x_2, ..., \lambda x_n).$$

Geometrically, $\mathbb{R}^2$ represents the cartesian plane, whereas $\mathbb{R}^3$ represents three dimensional space.
Similarly, the set $\mathbb{C}^n$ of n-tuples of complex numbers can be made into both a real vector space (with the scalars real numbers) or a complex vector space (with the scalars complex numbers).

Example 7.0.3. *Let Map $(\mathbb{R}, \mathbb{R})$ be the set of all mappings $f : \mathbb{R} \to \mathbb{R}$. For two such mappings f, g define $f + g : \mathbb{R} \to \mathbb{R}$ to be the mapping given by the prescription $(f + g)(x) = f(x) + g(x)$, and for every scalar $\lambda \in \mathbb{R}$ define $\lambda f : \mathbb{R} \to \mathbb{R}$ to be the mapping given by the prescription $(\lambda f)(x) = \lambda f(x)$. These operations make Map $(\mathbb{R}, \mathbb{R})$ into a real vector space.*

Theorem 7.0.33. *If V is a vector space over a field F then*

1. $(\forall \lambda \in F)\ \lambda 0_V = 0_V$
2. $(\forall x \in V)\ 0_F x = 0_V$
3. *if $\lambda x = 0_V$ then either $\lambda = 0_F$ or $x = 0_V$*
4. $(\forall x \in V)\ (\forall \lambda \in F)\ (-\lambda)x = -(\lambda x) = \lambda(-x)$.

Proof. By (V_3) and (V_5) we have

$$\begin{aligned} \lambda 0_V &= \lambda(0_V + 0_V) \\ &= \lambda 0_V + \lambda 0_V \end{aligned}$$

Now add $-(\lambda 0_V)$ to each side, we get

$$\lambda 0_V + -(\lambda 0_V) = \lambda 0_V + \lambda 0_V + -(\lambda 0_V).$$

This implies that $\lambda 0_V = 0_V$.

2. By (V_6) we have

$$\begin{aligned} \lambda 0_F x &= \lambda(0_F + 0_F)x \\ &= \lambda 0_F + \lambda 0_F \end{aligned}$$

Now add $-(\lambda 0_F x)$ to each side, we get

$$\lambda 0_F + -(\lambda 0_F) = \lambda 0_F + \lambda 0_F + -(\lambda 0_F).$$

This implies that $0_F x = 0_V$.

3. Suppose that $\lambda x = 0_V$ and that $\lambda \neq 0_F$. Then λ has a multiplicative inverse λ^{-1}, and so, by (V_7) and 1., $x = 1_F x = (\lambda\lambda^{-1})x = \lambda^{-1}(\lambda x) = \lambda^{-1}0_V = 0_V$.

4. By 2. and (V_6) we have

$$0_V = [\lambda + (-\lambda)]x = \lambda x + (-\lambda)x.$$

Now add -$(\lambda)x$ to each side. Also, by 1. and (V_5) we have

$$0_V = \lambda[x + (-x)] = \lambda x + \lambda(-x).$$

Now add -$(\lambda)x$ to each side. We have $(-\lambda)x = -(\lambda x) = \lambda(-x)$. □

Definition 7.0.15. *Let V be a vector space over a field F. By a subspace of V we shall mean a non-empty subset W of V that is closed under the operations of V, in the sense that*

(1) if $x, y \in W$ then $x + y \in W$;

(2) if $x \in W$ and $\lambda \in F$ then $\lambda x \in W$.

Example 7.0.4. *Every vector space V is (trivially) a subspace of itself. V itself is therefore the biggest subspace of V.*

Example 7.0.5. *The singleton subset $\{0_V\}$ is a subspace of V. This is then the smallest subspace of V since, as observed above, we have that $0_V \in W$ for every subspace W of V.*

Example 7.0.6. *The real vector space $\mathbb{R}$ is a subspace of the complex vector space $\mathbb{C}$.*

Example 7.0.7. *In the real vector space* $\mathbb{R}^2$ *the set* $X = \{(x, 0) : x \in \mathbb{R}\}$ *is a subspace. This subspace is simply the 'x-axis' in the cartesian plane* $\mathbb{R}$. *Similarly,the 'y-axis'* $Y = \{(0, y) : y \in \mathbb{R}\}$ *is a subspace of* $\mathbb{R}^2$.

Theorem 7.0.34. *The intersection of any set of subspaces of a vector space V is a subspace of V.*

Proof. Let C be a set of subspaces of V and let T be their intersection. Then $T \neq \phi$ since every subspace of V (and therefore every subspace in C) contains 0_V, whence so also does T. Suppose now that $x, y \in T$. Since x and y belong to every subspace W in the set C, so does x + y and hence $x + y \in T$. Also, if $x \in T$ then x belongs to every subspace W in the set C, whence so does λx and so $\lambda x \in T$. Thus we see that T is a subspace of V. □

Remark: The union of a set of subspaces of a vector space V need not be a subspace of V.

Example 7.0.8. *In* $\mathbb{R}^2$ *the x-axis X and the y-axis Y are subspaces, but* $X \bigcup Y$ *is not. For example, we have* $(1, 0) \in X$ *and* $(0, 1) \in Y$, *but*

$$(1, 0) + (0, 1) = (1, 1) \notin X \bigcup Y$$

so the subset $X \bigcup Y$ *is not closed under addition and therefore cannot be a subspace.*

Suppose now that we are given a subset S of a vector space V (with no restrictions, so that S may be empty if we wish). The collection C of all the subs paces of V that contain S is not empty, for clearly V itself belongs to C. By Theorem 7.0.34, the intersection of all the subspaces in C is also a subspace of V, and clearly this intersection also contains S. This intersection is therefore the smallest subspace of V that contains S (and is, of course, S itself whenever S is a subspace). We shall denote this subspace by $\langle S \rangle$.

Definition 7.0.16. *Let V be a vector space over a field F and let S be a non-empty subset of V. Then we say that* $v \in V$ *is a linear combination of elements of S if there exist* $x_1, x_2, ..., x_n \in S$ *and* $\lambda_1, \lambda_2, ..., \lambda_n \in F$ *such that*

$$v = \lambda_1 x_1 + \lambda_2 x_2 + ... + \lambda_n x_n = \sum_{i=1}^{n} \lambda_i x_i.$$

It is clear that $v = \sum_{i=1}^{n} \lambda_i x_i$ *and* $w = \sum_{i=1}^{n} \mu_i y_i$ *are linear combinations of elements of S then so is* $v + w$; *moreover, so is* λv *for every* $\lambda \in F$. *Thus the set of linear combinations of elements of S is a subspace of V. We call this the subspace spanned by S and denote it by Span S.*

Theorem 7.0.35. $\langle S\rangle = SpanS$.

Proof. For every $x \in S$ we have $x = 1_F x \in SpanS$ and therefore we see that $S \subseteq SpanS$. Since, by definition, $\langle S\rangle$ is the smallest subspace that contains S, and since Span S is a subspace, we see that $\langle S\rangle \subseteq SpanS$.
For the reverse inclusion, let $x_1, x_2, ..., x_n \in S$ and $\lambda_1, \lambda_2, ..., \lambda_n \in F$. If W is any subspace of V that contains S we clearly have $x_1, x_2, ..., x_n \in W$ and $\lambda_1 x_1 + \lambda_2 x_2 + ... + \lambda_n x_n \in W$. Consequently we see that $SpanS \subseteq W$. Taking W in particular to be $\langle S\rangle$, we obtain the result. □

Example 7.0.9. *If the n-tuple*

$$e_i = (0, ..., 0, 1, 0, ..., 0)$$

has the 1 in the i-th position then for every $(x_1, x_2, ..., x_n) \in \mathbb{R}^n$ *we have* $(x_1, x_2, ..., x_n) = x_1 e_1 + x_2 e_2 + ... x_n e_n$. *Consequently,* $\{e_1, e_2, ..., e_n\}$ *spans* $\mathbb{R}^n$.

Definition 7.0.17. *Let S be a non-empty subset of a vector space V over a field F. Then S is said to be linearly independent if the only way of expressing* 0_V *as a linear combination of elements of S is the trivial way (in which all scalars are* 0_F*)' Equivalently, S is linearly independent if, for any given* $x_1, x_2, ..., x_n \in S$, *we have* $\lambda_1 x_1 + \lambda_2 x_2 + ... + \lambda_n x_n = 0_V$. *This implies that* $\lambda_= \lambda_2 = ... = \lambda_n = 0_F$.
A subset that is not linearly independent is said to be linearly dependent.

Example 7.0.10. *If the n-tuple*

$$e_i = (0, ..., 0, 1, 0, ..., 0)$$

has the 1 in the i-th position then $\{e_1, e_2, ..., e_n\}$ *is a linearly independent subset of the vector space* $\mathbb{R}^n$.

Example 7.0.11. *Every singleton subset* $\{x\}$ *of a vector space V with* $x \neq 0$ *is linearly independent.*

Theorem 7.0.36. *No linearly independent subset of a vector space V can contain* 0_V.

Theorem 7.0.37. *Let V be a vector space over a field F. If S is a subset of V that contains at least two elements then the following statements are equivalent:*

(1) S is linearly dependent;

(2) at least one element of S can be expressed as a linear combination of the other elements of S.

Definition 7.0.18. *A basis of a vector space V is a linearly independent subset of V that spans V.*

Example 7.0.12. *If the n-tuple*

$$e_i = (0, ..., 0, 1, 0, ..., 0)$$

has the 1 in the i-th position then $\{e_1, e_2, ..., e_n\}$ is a basis for the vector space $\mathbb{R}^n$. This basis are called the natural (or canonical) basis.

A fundamental characterisation of bases is the following.

Theorem 7.0.38. *A non-empty subset S of a vector space V is a basis of V if and only if every element of V can be expressed in a unique way as a linear combination of elements of S.*

Proof. Suppose first that S is a basis of V. Then V = Span S and so, by Theorem 7.0.35, every $x \in V$ is a linear combination of elements of S. Now since S is linearly independent, only one such linear combination is possible for each $x \in V$; for if $\sum_{i=1}^{n} \lambda_i x_i = \sum_{i=1}^{n} \mu_i x_i$ where $x_i \in S$ then $\sum_{i=1}^{n} (\lambda_i - \mu_i) x_i = 0_V$ whence each $\lambda_i - \mu_i = 0_V$ and therefore $\lambda_i = \mu_i$ for each i.
Conversely, suppose that every element of V can be expressed in a unique way as a linear combination of elements of S. Then, by Theorem 7.0.35, Span S is the whole of V. Moreover, by the hypothesis, 0_V can be expressed in only one way as a linear combination of elements of S. This can only be the linear combination in which all the scalars are 0_F, It follows, therefore, that S is also linearly independent. Hence S is a basis of V. □

Theorem 7.0.39. *Let V be a vector space that is spanned by the finite set $G = \{v_1, ..., v_n\}$. If $I = \{w_1, ..., w_n\}$ is a linearly independent subset of V then necessarily $m \leq n$.*

Proof. Consider $w_i \in I$. Since G is a spanning set of V, there exist scalars $\lambda_1, ..., \lambda_n$ such that

$$w_1 = \lambda_1 v_1 + ... + \lambda_n v_n$$

and at least one of theλ_i is non-zero (otherwise every $\lambda_i = 0$ whence $w_1 = 0_V$ and this contradicts Theorem 7.0.36). By a suitable change of indices if necessary, we may assume without loss that $\lambda_1 \neq 0$. We then have

$$v_1 = \lambda_1^{-1} w_1 - \lambda_1^{-1} \lambda_2 v_2 - ... - \lambda_1^{-1} \lambda_n v_n,$$

which shows that

$$V = SpanG = Span\{v_1, v_2, ..., v_n\} \subseteq Span\{w_l, v_2, v_3, ..., v_n\}.$$

It follows that

$$V = Span\{w_l, v_2, v_3, ..., v_n\}.$$

Now w_2 can be written as a linear combination of $w_l, v_2, v_3, ..., v_n$ in which at least one of the coefficients of the v_j is non-zero (otherwise w_2 is a linear combination of w_1, a contradiction). Repeating the above argument we therefore obtain

$$V = Span\{w_l, w_2, v_3, ..., v_n\}.$$

Continuing in this way, we see that if $p = \min\{m, n\}$ then

$$V = Span\{w_l, ..., w_p, v_{p+1}, ..., v_n\}.$$

Now we see that $m > n$ is impossible; for in this case $p = n$ and we would have $V = Span\{w_l, w_2, ..., w_n\}$ whence the elements $w_{n+1}, ..., w_m$ would be linear combinations of $w_l, w_2, ..., w_n$ and this would contradict the fact that I is independent. Thus we conclude that $m \leq n$. □

Corollary 7.0.16. *If V has a finite basis B then every basis of V is finite and has the same number of elements as B.*

Proof. Suppose that B^* were an infinite basis of V. Since, clearly, every subset of a linearly independent set is also linearly independent, every subset of B* is linearly independent Now B^*, being infinite, contains finite subsets that have more elements than B. There would therefore exist a finite independent subset having more elements than B. Since this contradicts Theorem 7.0.39, we conclude that all bases of V must be finite. Suppose now that the basis B has n elements and let B^* be a basis with n^* elements. By Theorem 7.0.39, we have $n^* \leq n$. But, inverting the roles of B and B^*, we deduce also that $n \leq n^*$. Thus $n^* = n$ and so all bases have the same number of elements. □

Corollary 7.0.17. *If V has a finite basis then all linearly independent subsets of V are finite.*

Proof. If V has a finite basis of n elements and if there existed an infinite independent subset then this would contain an independent subset of n + 1 elements, and by the above this is not possible. □

Definition 7.0.19. *By a finite-dimensional vector space we shall mean a vector space V that has a finite basis. The number of elements in any basis of V is called the dimension of V and will be denoted by dim V.*

Example 7.0.13. *The vector space $\mathbb{R}^n$ has dimension n.*

Example 7.0.14. *The vector space $Mat_{m\times n}\mathbb{R}$ has dimension mn.*

Example 7.0.15. *The set V of complex matrices of the form forms a real vector space of dimension 6. In fact, V is a subspace of the real vector space $Mat_{2\times 2}\mathbb{C}$. Moreover, the matrix*

$$\begin{bmatrix} \alpha & \beta \\ \gamma & -\alpha \end{bmatrix} = \begin{bmatrix} a+ib & c+id \\ e+if & -a-ib \end{bmatrix}$$

can be written as

$$a\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} + c\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & i \\ 0 & 0 \end{bmatrix} + e\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + f\begin{bmatrix} 0 & 0 \\ i & 0 \end{bmatrix}$$

and as the six matrices involved in this belong to V and are clearly linearly independent over $\mathbb{R}$, they form a basis of the subspace that they span, which is V.

We shall now establish some important facts concerning bases.

Theorem 7.0.40. *Let V be a finite-dimensional vector space. If G is a finite spanning set of V and if I is a linearly independent subset of V such that $I \subseteq G$ then there is a basis B of V such that $I \subseteq B \subseteq G$.*

Proof. Observe first that if I also spans V then I is a basis of V and there is nothing to prove. Suppose then that $V \neq SpanI$. Then we must have $l \subset G$ (for otherwise I = G and is a spanning set of V). We note first that there exists $g_l \in G\backslash I$ such that $g_l \notin SpanI$; for otherwise every element of $G\backslash I$ belongs to Span I whence $V = SpanG \subseteq SpanI$ and we have the contradiction $V = SpanI$. We then observe that $I \bigcup \{g_1\}$ is linearly independent; otherwise we have the contradiction $g_1 \in SpanI$. Now if $I \bigcup \{g_1\}$ spans V then it is a basis, in which case no more proof is required since we can take $B = I \bigcup \{g_1\}$. If $I \bigcup \{g_1\}$ does not span V then we can repeat the above argument to produce an element $g_2 \in G$ $(I \bigcup \{g_1\})$ with $I \bigcup \{g_1, g_2\}$linearly independent. Proceeding in this way we see, since G is finite by hypothesis, that for some m the set $B = I \bigcup \{g_1, g_2, ..., g_m\}$ is a basis of V with $I \subset B \subseteq G$. □

Corollary 7.0.18. *Every linearly independent subset I of a finite-dimensional vector space V can be extended to form a basis.*

Proof. By Corollary 7.0.17, I is finite. Take $G = I \bigcup B$ where B is any basis of V. Then by the above there is a basis B^* with $I \subseteq B^* \subseteq I \bigcup B$. □

Corollary 7.0.19. *If V is of dimension n then every linearly independent set consisting of n elements is a basis of V.*

Corollary 7.0.20. *If S is a subset of V then the following statements are equivalent:*

(1) S is a basis;

(2) S is a maximal independent subset (in the sense that if I is an independent subset with $S \subseteq I$ then $S = I$);

(3) S is a minimal spanning set (in the sense that if G spans V and $G \subseteq S$ then G= S).

Proof. Let S be a subset of V

(1)⇒(2) If I is independent with $S \subseteq I$ then by Corollary 7.0.18 there is a basis B such that $I \subseteq B$. Since S is a basis, and since all bases have the same number of elements, we deduce that S = B = I.

(2)⇒(1) By CorolIary 7.0.18 there is a basis B with $S \subseteq B$. But B is also independent so, by (2), we have S = B and therefore S is a basis.

(1)⇒(3) If G spans V then (recalling that ϕ is independent) there is a basis B with $\phi \subseteq B \subseteq G$. If $G \subseteq S$ then $B \subseteq S$ and both are bases. Again since bases have the same number of elements, we deduce that B = G = S.

(3)⇒(1) There is a basis B with $\phi \subseteq B \subseteq S$. But B also spans V so, by (3), we have B = S and so S is a basis.

This completes the proof. □

Corollary 7.0.21. *If V is of dimension n then every subset containing more than n elements is linearly dependent. No subset containing fewer than n elements can span V.*

Theorem 7.0.41. *Let V be a finite-dimensional vector space. If W is a subspace of V then W is also of finite dimension, and dim W ≤ dim V. Moreover, we have dim W= dim V ⇔ W= V.*

Proof. Suppose that V is of dimension n. If I is a linearly independent subset of W then, by Theorem 7.0.39, I has at most n elements. A maximal such subset B is then, by Corollary 7.0.20, a basis of W. Hence W is also of finite dimension, and dim W≤ dim V. Finally, if dim W = dim V = n then B is a linearly independent subset of V having n elements whence, by Corollary 7.0.19, B is a basis of V. Hence W = Span B = V. □

MODULE IV

Chapter 8

LINEAR MAPPINGS

Definition 8.0.20. *If V and W are vector spaces over the same field F then by a linear mapping (or linear transformation) from V to W we shall mean a mapping $f : V \to W$ such that*

1. $(\forall x, y \in V)\ f(x + y) = f(x) + f(y)$;
2. $(\forall x \in V)\ (\forall \lambda \in F)\ f(\lambda x) = \lambda f(x)$.

If $f : V \to W$ is linear then V is sometimes called the departure space and W the arrival space off.

Example 8.0.16. *The mapping $f : \mathbb{R}^2 \to \mathbb{R}^3$ given by*

$$f(a, b) = (a + b, a - b, b)$$

is linear.

Proof. For all (a, b) and (a', b') in $\mathbb{R}^2$ we have

$$\begin{aligned} f((a, b) + (a', b')) &= f(a + a', b + b') \\ &= (a + a' + b + b', a + a' - b - b', b + b') \\ &= (a + b, a - b, b) + (a' + b', a' - b', b') \\ &= f(a, b) + f(a', b') \end{aligned}$$

and for all $(a, b) \in \mathbb{R}^2$ and all $\lambda \in \mathbb{R}$

$$\begin{aligned} f(\lambda(a, b)) &= f(\lambda a, \lambda b) \\ &= (\lambda a + \lambda b, \lambda a - \lambda b, \lambda b) \\ &= \lambda(a + b, a - b, b) \\ &= \lambda f(a, b). \end{aligned}$$

This implies that f is linear. □

The following result contains two important properties of linear mappings.

Theorem 8.0.42. *If the mapping $f : V \to W$ is linear then*

1. $f(0_V) = 0_W$;
2. $(\forall x \in V)\ f(-x) = -f(x)$.

Proof. 1. $f(0_V) = f(0_F 0_V) = 0_F f(0_V) = 0_W$.

2. Using 1. we have $\forall x \in V$,

$$f(x) + f(-x) = f(x + (-x)) = f(0_V) = 0_W,$$

adding $-f(x)$ to each side we get the result. □

Definition 8.0.21. *If $f : V \to W$ is linear then for every subset X of V we define $f^{\to}(X)$ to be the subset of W given by*

$$f^{\to}(X) = \{f(x) : x \in X\}$$

and for every subset Y of W we define $f^{\leftarrow}(Y)$ to be the subset of V given by

$$f^{\leftarrow}(Y) = x \in V : f(x) \in Y.$$

We often call $f^{\to}(X)$ the direct image of X under f, and $f^{\leftarrow}(Y)$ the inverse image of Y under f.

Theorem 8.0.43. *Let $f : V \to W$ be linear. If X is a subspace of V then $f^{\to}(X)$ is a subspace of W; and if Y is a subspace of W then $f^{\leftarrow}(Y)$ is a subspace of V.*

Proof. Observe first that if X is a subspace of V then we have $0_V \in X$ and therefore $0_W = f(0_V) \in f^{\to}(X)$. Thus $f^{\to}(X) \neq \phi$. If now $y_1, y_2 \in f^{\to}(X)$ then $y_l = f(x_l)$ and $y_2 = f(x_2)$ for some $x_l, x_2 \in X$. Consequently, since X is a subspace of V,

$$y_l + y_2 = f(x_l) + f(x_2) = f(x_1 + x_2) \in f^{\to}(X)$$

and, for every scalar λ,

$$\lambda y_1 = \lambda f(x_1) = f(\lambda x_1) \in f^{\to}(X).$$

Thus $f^{\to}(X)$ is a subspace of W. Suppose now that Y is a subspace of W. Observe that $f(0_V) = 0_W \in Y$ gives $0_V \in f^{\leftarrow}(Y)$, and therefore $f^{\leftarrow}(Y) \neq \phi$. If now $x_1, x_2 \in f^{\leftarrow}(Y)$ and therefore

$$f(x_1 + x_2) = f(x_1) + f(x_2) \in Y$$

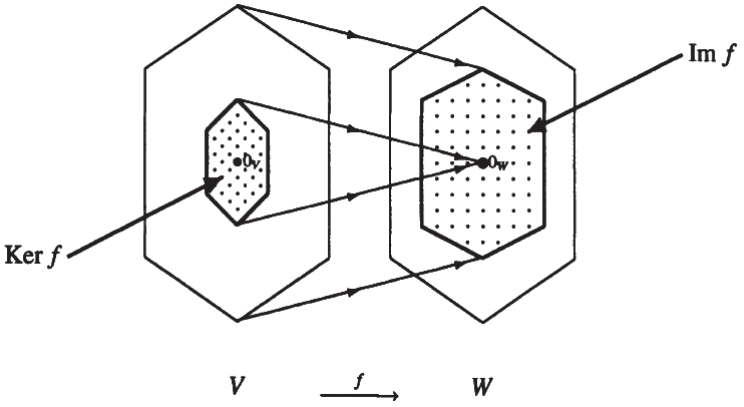
whence $x_1 + x_2 \in f^{\leftarrow}(Y)$ and for every scalar λ,

$$f(\lambda x_1) = \lambda f(x_1) \in Y$$

whence $\lambda x_1 \in f^{\leftarrow}(Y)$. Thus $f^{\leftarrow}(Y)$ is a subspace of V. □

Definition 8.0.22. *Let $f : V \to W$ be linear. The biggest possible direct image $f^{\rightarrow}(V)$ is called the image (or range) of f and is denoted by Im f. The smallest possible inverse image $f^{\leftarrow}(\{0_V\})$ is called the kernel (or null-space) of f and is denoted by Ker f.*

Pictorially, these sets can be depicted as follows:



Example 8.0.17. *If A is a given real $n \times n$ matrix, consider the linear mapping $f_A : Mat_{n\times 1}\mathbb{R} \to Mat_{n\times 1}\mathbb{R}$ by $f_A(x) = Ax$. The image of f_A consists of all $n \times 1$ column matrices $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ for which there exists $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ such that $Ax = y$; that is, the set of all y such that there exist $x_1, ..., x_n$ with $y = x_1a_1 + ... + x_na_n$. In other words, Im f_A is the subspace of $Mat_{n\times 1}\mathbb{R}$ that is spanned by the columns of A. As for the kernel of f_A this is the subspace of $Mat_{n\times 1}\mathbb{R}$ consisting of the column matrices x such that $Ax = 0$; that is, the solution space of the system $Ax = 0$.*

Example 8.0.18. *Consider the mapping $f : \mathbb{R}^3 \to \mathbb{R}^4$ given by*

$$f(a, b, c, d) = (a + b, b - c, a + d).$$

Since

$$(a + b, b - c, a + d) = a(1, 0, 1) + b(1, 1, 0) + c(0, -1, 0) + d(0, 0, 1)$$

we see that Im $f = Span\{(1,0,1),(1,1,0),(0,-1,0),(0,0,1)\}$. To find a basis for 1m f, proceed as follows. Observe that Im f is the subspace spanned by the rows of the matrix

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The Hermite form of A is

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since the rows of this matrix span the same subspace, and since they are linearly independent, we deduce that a basis for Im f is $\{(1,0,1),(1,1,0),(0,-1,0)\}$. Thus Im $f = \mathbb{R}^3$.

Definition 8.0.23. *A linear mapping $f : V \to W$ is said to be surjective if Im $f = W$ (in other words, if every element of W is the image under f of some element of V); and injective if $f(x) \neq f(y)$ whenever $x \neq y$ (in other words, if f carries distinct elements to distinct elements). We say that f is bijective if it is both injective and surjective.*

Example 8.0.19. *The i-th projection $pr_i : \mathbb{R}^n \to \mathbb{R}$ is surjective but not injective.*

Example 8.0.20. *The linear mapping $f : \mathbb{R}^2 \to \mathbb{R}^3$ given by*

$$f(x,y) = (y,0,x)$$

is injective but not surjective.

Example 8.0.21. *The differentiation map $D : \mathbb{R}_n[X] \to \mathbb{R}_n[X]$ is neither injective nor surjective.*

Theorem 8.0.44. *If $f : V \to W$ is linear then the following statements are equivalent:*

(1) f is injective;

(2) Ker $f = \{0\}$.

Proof. Let $f : V \to W$ be linear.

(1) $\Rightarrow$ (2) : Suppose that f is injective. Then f is such that $x \neq y \Rightarrow f(x) \neq f(y)$ or, equivalently, $f(x) = f(y) \Rightarrow x = y$. Suppose now that $x \in Ker\, f$. Then we have $f(x) = 0_W = f(0_V)$ whence we see that $x = 0_V$ and consequently Ker $f = \{0\}$.

(2) $\Rightarrow$ (1) : Suppose that Ker $f = \{0\}$ and let $f(x) = f(y)$.

Then $f(x - y) = f[x + (-y)] = f(x) + f(-y) = f(x) - f(y) = 0_W$ so that $x - y \in Ker\, f = 0_V$ and hence $x = y$, that is f is injective. □

Example 8.0.22. *The linear mapping $f : \mathbb{R}^3 \to \mathbb{R}^3$ given by*

$$f(x, y, z) = (x + z, x + y + 2z, 2x + y + 3z)$$

is neither surjective nor injective. In fact, we have that $(a, b, c) \in Im\, f$ if and only if the system of equations

$$\begin{array}{ll} x + z & = a \\ x + y + 2z & = b \\ 2x + y + 3z & = c \end{array}$$

is consistent. The augmented matrix of the system is

$$\begin{bmatrix} 1 & 0 & 1 & a \\ 1 & 1 & 2 & b \\ 2 & 1 & 3 & c \end{bmatrix}$$

and this has Hermite form

$$\begin{bmatrix} 1 & 0 & 1 & a \\ 0 & 1 & 1 & b - a \\ 0 & 0 & 0 & c - b - a \end{bmatrix}$$

We deduce from this that $(a, b, c) \in Im\, f$ if and only if $c = a + b$, whence f is not surjective.
Now $(x, y, z) \in Ker\, f$ if and only if

$$\begin{array}{ll} x + z & = 0 \\ x + y + 2z & = 0 \\ 2x + y + 3z & = 0 \end{array}$$

which is the associated homogeneous system of equations. By Theorem 8.0.44, for Ker f to be the zero subspace we require this system to have a unique solution (namely the trivial solution $(0, 0, 0)$). But, from the above Hermite form, the coefficient matrix has rank 2 and so, non-trivial solutions exist. Hence f is not injective.

In the case of finite-dimensional vector spaces there is an important connection between the dimensions of the subspaces Im f and Ker f.

Theorem 8.0.45. *[Dimension Theorem] Let V and W be vector spaces of finite dimension over a field F. If $f : V \to W$ is linear then*

$$dim\ V = dim\ Im\ f + dim\ Ker\ f.$$

Proof. Let $\{w_1, w_2, ..., w_m\}$ be a basis of Im f, and let $\{v_1, v_2, ..., v_n\}$ be a basis of Ker f. Since each $w_i \in$ Im f, we can choose $v_1^*, v_2^*, ..., v_m^* \in V$ such that $f(v_i^*) = w_i$ for $i = 1, ..., m$. We shall show that

$$\{v_1^*, v_2^*, ..., v_m^*, v_1, v_2, ..., v_n\}$$

is a basis of V. whence the result follows. Suppose that x E V, Since $f(x) \in Im\ f$ there exist A $\lambda_1, ..., \lambda_m \in F$ such that

$$f(x) = \sum_{i=1}^{m} \lambda_i w_i = \sum_{i=1}^{m} \lambda_i f(v_i^*) = \sum_{i=1}^{m} f(\lambda_i v_i^*) = f(\sum_{i=1}^{m} \lambda_i v_i^*).$$

It follows that

$$x - \sum_{i=1}^{m} \lambda_i v_i^* \in Ker\ f$$

and so there exist $\mu_1, \mu_2, ..., \mu_n \in F$ such that

$$x - \sum_{i=1}^{m} \lambda_i v_i^* = \sum_{j=1}^{n} \mu_j v_j.$$

Thus every $x \in V$ is a linear combination of $v_1^*, v_2^*, ..., v_m^*, v_1, v_2, ..., v_n$ and so

$$V = Span\ \{v_1^*, v_2^*, ..., v_m^*, v_1, v_2, ..., v_n\}.$$

Suppose now that

$$\sum_{i=1}^{m} \lambda_i v_i^* + \sum_{j=1}^{n} \mu_j v_j = 0. \tag{8.1}$$

Then we have

$$\sum_{i=1}^{m} \lambda_i v_i^* = -\sum_{j=1}^{n} \mu_j v_j \in Ker\ f$$

and consequently

$$\sum_{i=1}^{m} \lambda_i w_i = f(\sum_{i=1}^{m} \lambda_i v_i^*) = 0.$$

whence $\lambda_1 = ... = \lambda_m = 0$ since $\{w_1, w_2, ..., w_m\}$ is a basis of $Im\ f$. It now follows from 8.1 that $\sum_{j=1}^{n} \mu_j v_j = 0$ whence $\mu_1 = ... = \mu_n = 0$ since $\{v_1, v_2, ..., v_n\}$ is a basis of Ker f. Thus we see that the spanning set $\{v_1^*, v_2^*, ..., v_m^*, v_1, v_2, ..., v_n\}$ is also linearly independent and is therefore a basis of V. □

Definition 8.0.24. *If f is a linear mapping then dim Im f is called the rank of f; and dim Ker f is called the nullity of f.*

Remark: The dimension theorem above can be stated in the form:

$$\text{rank} + \text{nullity} = \text{dimension of departure space.}$$

Example 8.0.23. *Consider $pr_1 : \mathbb{R}^3 \to \mathbb{R}$ given by $pr_1(x, y, z) = x$. We have $Im\ pr_1 = \mathbb{R}$ which is of dimension 1 since $\{1\}$ is a basis of the real vector space $\mathbb{R}$; so pr_1 is of rank 1. Also, Ker pr_1 is the y, z-plane which is of dimension 2. Thus pr_1 is of nullity 2.*

Theorem 8.0.46. *Let V and W be vector spaces each of dimension n over a field F. If $f : V \to W$ is linear then the following statements are equivalent:*
(1) f is injective;
(2) f is surjective;
(3) f is bijective;
(4) f carries bases to bases, in the sense that if $\{v_1, v_2, ..., v_n\}$ is a basis of V then $\{f(v_1), f(v_2), ..., f(v_n)\}$ is a basis of W.

Proof. $(l) \Rightarrow (3)$: Suppose that f is injective. Then Ker $f = \{0\}$ and so dim Ker $f = 0$. By Theorem 8.0.45, it follows that

$$dim\ Im\ f = n = dim\ V = dim\ W.$$

It now follows by Theorem 7.0.41 that Im $f = W$ and so f is also surjective, and hence is bijective.
$(2) \Rightarrow (3)$: Suppose that f is surjective. Then Im $f = W$ and so, by Theorem 8.0.45,

$$dim\ Im\ f = dim\ W = n = dim\ V = dim\ Im\ f + dim\ Ker\ f$$

whence dim Ker $f = 0$. Thus Ker $f = \{0\}$ and so, by Theorem 8.0.44,f is also injective, and hence is bijective.
$(3) \Rightarrow (1)$: and $(3) \Rightarrow$ are clear.
$(1) \Rightarrow (4)$: Suppose that f is injective. If $\{v_1, v_2, ..., v_n\}$ is a basis of V then the elements $f(v_1), f(v_2), ..., f(v_n)$ are distinct. If now $\sum_{i=1}^{n} \lambda_i f(v_i) = 0$ then $f(\sum_{i=1}^{n} \lambda_i v_i) = 0$ and so, since Ker $f = \{0\}$, we have $\sum_{i=1}^{n} \lambda_i v_i = 0$ and hence

$\lambda_1 = ... = \lambda_n = 0$. Thus $\{f(v_1), f(v_2), ..., f(v_n)\}$ is linearly independent. That it is now a basis follows from Corollary 7.0.19.
(4) $\Rightarrow$ (2) : Since every linear combination of $f(v_1), f(v_2), ..., f(v_n)$ belongs to Im f, it is clear from (4) that Im $f = W$ and so f is surjective. □

Definition 8.0.25. *A bijective linear mapping is called a linear isomorphism, or simply an isomorphism. We say that vector spaces V, Ware isomorphic, and write $V \simeq W$, if there is an isomorphism $f : V \to W$.*

Example 8.0.24. *Let $A = (x, y, 0); x, y \in \mathbb{R}$ be the x, y-plane in $\mathbb{R}^3$, and let $B = (x, 0, z); x, z \in \mathbb{R}$ be the x, z-plane. Consider the mapping $f : A \to B$ given by $f(x, y, 0) = (x, 0, y)$. Clearly, f is linear and bijective. Thus f is an isomorphism and so $A \simeq B$.*

Theorem 8.0.47. *Let V be a vector space 01 dimension $n \geq 1$ over a field F. Then V is isomorphic to the vector space F^n.*

Proof. Let $\{v_1, v_2, ..., v_n\}$ be a basis of V. Consider the mapping $f : V \to F^n$. given by the prescription

$$f(\sum_{i=1}^{n} \lambda_i v_i) = (\lambda_1, ..., \lambda_n).$$

Since for every $x \in V$ there are unique scalars $\lambda_1, ..., \lambda_n$ such that $x = \sum_{i=1}^{n} \lambda_i v_i$ it is clear that f is a bijection. It is clear that f is linear. Hence f is an isomorphism. □

Corollary 8.0.22. *If V and W are vector spaces of the same dimension n over F then V and W are isomorphic.*

Proof. There are isomorphisms $f_V : V \to F^n$ and I$f : W \to F^n$. Since the inverse of an isomorphism is clearly also an isomorphism, so then is the composite mapping $f_W^{-1} \circ f_V : V \to W$. □

Theorem 8.0.48. *Let V and W be vector spaces over a field F. If $\{v_1, v_2, ..., v_n\}$ is a basis of V and $\{w_1, w_2, ..., w_n\}$ are elements of W (not necessarily distinct) then there is a unique linear mapping $f : V \to W$ such that $(i = l, ..., n)$ $f(v_i) = w_i$.*

Proof. Since every element of V can be expressed uniquely in the form $\sum_{i=1}^{n} \lambda_i v_i$, we can define a mapping $f : V \to W$ by the prescription $f(\sum_{i=1}^{n} \lambda_i v_i) = \sum_{i=1}^{n} \lambda_i w_i$, that is, taking x as a linear combination of the basis elements,

define $f(x)$ to be the same linear combination of the elements $w_1, w_2, ..., w_n$. It is readily verified that f is linear. Moreover, for each i, we have

$$f(v_i) = f(\sum_{j=1}^{n} \delta_{ij} v_j) = \sum_{j=1}^{n} \delta_{ij} v_j = w_i.$$

As for the uniqueness, suppose that $g : V \to W$ is also linear and such that $g(v_i) = w_i$ for each i. Given $x \in V$, say $x = \sum_{i=1}^{n} \lambda_i v_i$, we have

$$f(v_i) = f(\sum_{j=1}^{n} \delta_{ij} v_j) = \sum_{j=1}^{n} \delta_{ij} v_j = w_i$$

whence $g = f$. □

Corollary 8.0.23. *A linear mapping is completely and uniquely determined by its action on a basis.*

Proof. If $f : V \to W$ is linear and $B = \{v_1, v_2, ..., v_n\}$ is a basis of V let $w_i = f(v_i)$ for each i. Then by the above f is the only linear mapping that sends v_i to w_i. Moreover, knowing the action of f on the basis B, we can compute $f(x)$ for every x; for $x = \sum_{i=1}^{n} \lambda_i v_i$ gives $f(x) = \sum_{i=1}^{n} \lambda_i f(v_i)$. □

Corollary 8.0.24. *Two linear mappings $f, g : V \to W$ are equal if and only if they agree on any basis of V.*

Proof. If $f(v_i) = g(v_i)$ for every basis element v_i then by the above uniqueness we have that $f = g$. □

Example 8.0.25. *Consider the basis $\{(1,1,0), (1,0,1), (0,1,1)\}$ of $\mathbb{R}^3$. If $f : \mathbb{R}^3 \to \mathbb{R}^2$ is linear and such that*

$$f(1,1,0) = (1,2), \quad f(1,0,1) = (0,0), \quad f(0,1,1) = (2,1),$$

then we can determine f completely. In fact, we have

$$(1,0,0) = \frac{1}{2}(1,1,0) + \frac{1}{2}(1,0,1) - \frac{1}{2}(0,1,1)$$

and therefore

$$\begin{aligned} f(1,0,0) &= \tfrac{1}{2} f(1,1,0) + \tfrac{1}{2} f(1,0,1) - \tfrac{1}{2} f(0,1,1) \\ &= \tfrac{1}{2}(1,2) + \tfrac{1}{2}(0,0) - \tfrac{1}{2}(2,1) \\ &= (-\tfrac{1}{2}, \tfrac{1}{2}). \end{aligned}$$

Likewise,

$$\begin{array}{ll} (0,1,0) & = \frac{1}{2}(1,1,0) - \frac{1}{2}(1,0,1) + \frac{1}{2}(0,1,1) \\ (0,0,1) = -\frac{1}{2}(1,1,0) + \frac{1}{2}(1,0,1) + \frac{1}{2}(0,1,1) & \end{array}$$

give

$$\begin{array}{ll} f(0,1,0) & = \frac{1}{2}(1,2) + \frac{1}{2}(2,1) = (\frac{3}{2},\frac{3}{2}) \\ f(0,0,1) & = -\frac{1}{2}(1,2) + \frac{1}{2}(2,1) = (\frac{1}{2},-\frac{1}{2}). \end{array} .$$

Consequently, f is given by

$$\begin{array}{rl} f(x,y,z) & = f[x(1,0,0) + y(1,0,0) + z(0,0,1)] \\ & = xf(1,0,0) + yf(1,0,0) + zf(0,0,1) \\ & = x(-\frac{1}{2},\frac{1}{2}) + y(\frac{3}{2},\frac{3}{2}) + (\frac{1}{2},-\frac{1}{2}) \\ & = (\frac{1}{2}(-x+3y+z), \frac{1}{2}(x+3y-z)). \end{array}$$

Remark 1: Note that, alternatively, we could first have expressed (x,y,z) as a linear combination of the given basis elements by solving an appropriate system of equations, then using the given data.
Remark 2: Note that Theorem 8.0.46 is not true for vector spaces of infinite dimension.

Example 8.0.26. *Let $V = Seq_f\mathbb{R}$ be the infinite-dimensional vector space of finite sequences of real numbers. Since every element of V is a (finite) linear combination of basis elements, we can define a linear mapping $f : V \to V$ by specifying $f(e_i)$ for the basis elements $e_1, e_2, e_3, ...$ and extending to all of V by linearity. Consider then the definition*

$$f(e_i) = \begin{cases} 0 & \text{if } i \text{ is odd;} \\ e_{\frac{1}{2}i} & \text{if } i \text{ is even.} \end{cases}$$

Since $f(e_1) = 0 = f(e_3)$ we see that f is not injective. But, given any basis element e_n we have $e_n = f(e2n) \in Im\ f$, so the subspace spanned by these elements (namely, the whole of V) is contained in Im f. Hence $Im\ f = V$ and so f is surjective. If we define $g : V \to V$ by specifying $g(e_i) = e_{2i}$ for every i then we obtain an injective linear mapping that is not surjective.