SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF SCIENCE AND HUMANITIES
DEPARTMENT OF MATHEMATICS

UNIT – I –Theory of Divisibility and Congruencies– SMTA5203

Unit – I

Theory of Divisibility and Congruences

Division Algorithm

Theorem:(Division Algorithm)**.** *Given integers a and b, with b* > 0*, there exist unique integer q and r satisfying* $a = qb + r, 0 \leq r < b$*. The integers q and r are called, respectively, the quotient and remainder in the division of a by b.*

Proof. Consider the set

$$S = \{a - xb \mid x \text{ an integer}; a - xb \geq 0\}$$

Claim 1: S is non-empty

That is to prove that there exists a value of *x* that makes $a - xb$ nonnegative. It is given that the integer $b \geq 1$. Multiplying both sides by $|a|$, $|a|b \geq |a|$, and so

$$a - (-|a|) = a + |a|b \geq a + |a| \geq 0$$

Hence for the choice $x = -|a|$, $a - xb$ lies in *S*. By the application of the *Well-Ordering Principle*, it is guaranteed that the set *S* contains a smallest integer, say *r*. From the definition of *S*, it follows that there exists an integer *q* satisfying $r = a - qb$ $0 \leq r$.

Claim 2: $r < b$

Suppose the contradiction that $r \geq b$ and

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

The implication is that the integer $a - (q + 1)b$ has the proper form to belong to the set *S*. But $a - (q + 1)b = r - b < r$, leading to a contradiction of the choice of *r* as the smallest member of *S*. Hence, $r < b$.

Claim 3: Uniqueness

Next we turn to the task of showing the uniqueness of *q* and *r*. Suppose that *a* has two representations of the desired form, say,

$a = qb + r = q'b + r'$, where $0 \leq r < b$, $0 \leq r' < b$.

Then $r' - r = (q - q')$.

$$|r' - r| = b|q - q'|$$

Upon adding the two inequalities $-b < -r \leq 0$ and $0 \leq r' < b$, we obtain

$-b < r' - r < b$ or, in equivalent terms, $|r' - r| < b$. Thus, $b|q - q'| < b$, which yields

$$0 \leq |q - q'| < 1$$

Because $|q - q'|$ is a nonnegative integer, the only possibility is that $|q - q'| = 0$, whence $q = q'$; this, in turn, gives $r = r'$, ending the proof.

Corollary *If a and b are integers, with* $b \neq 0$*, then there exist unique integers q and r such that* $a = qb + r$*,*

$$0 \leq r < |b|$$

Proof. It is enough to consider the case in which *b* is negative. Then $|b| > 0$, and from the division algorithm there exits unique integers q' and *r* for which

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

As $b < 0$, $|b| = -b$, assuming $q = -q'$ it follows that $a = qb + r$, with $0 \leq r < |b|$.

The Greatest Common Divisor

Definition An integer *b* is said to be *divisible* by an integer $a \neq 0$, denoted by $a|b$, if there exists a integer *c* such that $b = ac$. Otherwise, we say that *b* is not divisible by *a* and is denoted by $a \nmid b$.

Theorem *For integers a,b and c, the following statements hold*

(a) $a|0,1|a, a|a.$

(b) $a|1$ if and only if $a = \pm 1$.

(c) If $a|b$ and $c|d$, then $ac|bd$.

(d) If a|b and b|c, then a|c.

(e) $a|b$ and $b|a$ if and only if $a = \pm b$.

(f) If $a|b$ and $b \neq 0$, then $|a| \leq |b|$.

(g) If $a|b$ and $a|c$, then $a|(bx + cy)$ for arbitrary integers x and y.

Proof:

(a) $0 = (0), 0 \in \mathbb{Z} \Rightarrow a|0.$

$a = 1(a), a \in \mathbb{Z} \Rightarrow 1|a.$

$a = (1), 1 \in \mathbb{Z} \Rightarrow a|a.$

(b) Case (i): Suppose $a|1 \Rightarrow 1 = a(k), k = \frac{1}{a} \in \mathbb{Z}$. Hence $a = \pm 1$.

Case (ii): Suppose $a = \pm 1$. $1 = 1(1)$ or $1 = -1(-1)$. Hence $a|1$.

(c) Let it be true that $a|b$ and $c|d$. It follows that $b = a(k_1)$, $k_1 \in \mathbb{Z}$ and $d = c(k_2)$, $k_2 \in \mathbb{Z}$. Hence $bd = ac(k_1k_2)$, $k_1k_2 \in \mathbb{Z}$. Hence by divisibility conditions, $ac|bd$.

(d) Let it be true that $a|b$ and $b|c$. It follows that $b = a(k_1)$, $k_1 \in \mathbb{Z}$ and $c = b(k_2)$, $k_2 \in \mathbb{Z}$. Hence $c = a(k_1k_2)$, $k_1k_2 \in \mathbb{Z}$. Hence by divisibility conditions, $a|c$.

(e) Let it be true that $a|b$ and $b|a$. It follows that $b = a(k_1)$, $k_1 \in \mathbb{Z}$ and $a = b(k_2)$, $k_2 \in \mathbb{Z}$. Hence $ab = ab(k_1k_2), k_1k_2 \in \mathbb{Z}. \Rightarrow 1 = k_1k_2 \Rightarrow k_1 = \frac{1}{k_2} \in \mathbb{Z} \Rightarrow k_2 = \pm 1$. It follows that $a = \pm b$.

(f) 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 $|b| = |a||c| \geq |a|$.

(g) Given that $|b$ and $a|c$. This ensures that $b = ar$ and $c = as$ for suitable integers *r* and *s*. For some integer values of *x* and *y*, $bx + cy = arx + asy = (rx + sy)$. As $rx + sy$ is an integer, by divisibility conditions, $a|(bx + cy)$.

Definition: 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:

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

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

Theorem: *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*:

$$S = \{au + bv \mid au + bv > 0; u, v\ integers\}$$ **Claim 1** *S* is not empty.

Since, if $a \neq 0$, then the integer $|a| = au + (0)$ lies in *S*, where we choose $u = 1$ or $u = -1$ according as *a* is positive or negative. By virtue of the Well-ordering Principle, *S* must contain a smallest element *d*. Thus, from the very definition of *S*, there exists integers *x* and *y* for which $d = ax + by$.

Claim 2 d is a common divisor of a and b.

By the Division Algorithm, there exists unique integers *q* and *r* such that $a = qd + r$, where $0 \leq r <$ d. Then *r* can be written in the form

$$r = a - qd = a - (ax + by) = a(1 - qx) + b(-qy)$$

If *r* were positive, then this representation would imply that $r < d$ is an element of *S*, contradicting the fact that *d* is the least integer in *S*. Therefore, $r = 0$, and so $a = qd$, or equivalently $d|a$. By similar reasoning, $d|b$, This assures that *d* a common divisor of *a* and *b*.

Claim 3: $d = \gcd(a, b)$.

Let *c* be an arbitrary positive common divisor of the integers *a* and *b*, then from the theorem it easily follows that $c|(ax + by) \Rightarrow c|d$. Hence, $c = |c| \leq |d| = d$, so that *d* is greater than every positive common divisor of *a* and *b*. Hence, $d = gc(a, b)$.

Corollary: *If a and b are given integers, not both zero, then the set* $T = \{ax + by|x, y \in \mathbb{Z}\}$ *is precisely the set of all multiples of* $d = gc(a, b)$.

Proof.

Given $d = \gcd(a, b)$. It follows that $d|a$ and $d|b$. By the above theorem $d|(ax + by)$ for all integers *x*, *y*. Thus, every member of *T* is a multiple of *d*.

Conversely, $d \in T$ may be written as $d = ax_0 + by_0$ for suitable x_0, and y_0, so that any multiple nd of *d* is of the form $nd = (ax0 + by0) = a(nx0) + b(ny0)$. Hence, *nd* is a linear combination of *a* and *b*, and, by definition, lies in *T*.

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

Theorem: *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 there exists 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* an *d*|*b*, it follows that *d*|(*ax* + *by*), or *d*|1. Hence, $d|1 \Rightarrow d = \pm 1$. By assumption d is positive. Hence, $d = \gcd(a, b) = 1$. i.e., a and b are relatively prime.

Corollary: *If* $gc(a, b) = d$, *then* $gcd\left(\frac{a}{d}, \frac{b}{d}\right) = 1$

Proof. Before starting with the proof proper, we should observe that although *a/d* and *b/d* have the appearance of fractions, in fact, they are integers because *d* is a divisor both of *a* and of *b*. Now, knowing that *gcd*(*a,b*) = *d*, it is possible to find integers *x* and *y* such that *d* = *ax*+*by*. Upon dividing each side of this equation by *d*, we obtain the expression

$1 = \frac{a}{d}) x + \frac{b}{d}) y$

Because *a/d* and *b/d* are integers, an appeal to the theorem is legitimate. The conclusion is that *a/d* and *b/d* are relatively prime. 2

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

Proof. If $a|c$ and $b|c$, then there exists integers *r* and *s* such that $c = ar = bs$. Given $gc(a, b) = 1$. It then follows that there exists integers x and y such that $1 = ax + by$. Multiplying, the last equation by *c*:

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

appropriate substitutions on the right-hand side:

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

follows that, $ab|c$.

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

Proof. As $\gcd(a, b) = 1$, it is true that 1 = *ax* + *by*, where *x* and *y* are integers. Multiplication of this equation by *c* leads to

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

Because $a|ac$ and $a|bc$, it follows that $a|(acx + bcy)$. From the above equation it follows that $a|c$.

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

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

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

Proof. Suppose that $d = gc(a, b)$. By definition of greatest common divisor it follows that $d|a$ and $d|b$, so that (a) holds. From the theorem, d is expressible as $d = ax + by$ for some integers x, y. We know that, if $c|a$ and $c|b$, then $c|(ax + by)$ hence it follows that $c|d$. In short, condition (b) holds.

Conversely, let d be any positive integer satisfying the stated conditions. Given any common divisor c of a and b, we have $c|d$ from hypothesis (b). The implication is that $d \geq c$, and consequently d is the greatest common divisor of a and b.

The Euclidean Algorithm

Let a and b be two integers whose greatest common divisors need to be computed. By the properties of GCD, $gc(|a|, |b|) = gcd(a, b)$. Assume that $a \geq b > 0$. The first step is to apply the Division Algorithm to a and b to get

$$a = q_1 b + r_1, 0 \leq r_1 < b$$

If it happens that $r_1 = 0$, then $b|a$ and $gc(a, b) = b$.

When $r_1 \neq 0$, again from the division algorithm there exists integers q_2 and r_2 satisfying

$$a = q_2 r_1 + r_2, 0 \leq r_2 < r_1$$

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

$$r_1 = q_3 r_2 + r_3, 0 \leq r_3 < r_2$$

This division process continues until some zero remainder appears, say, at the $(n + 1)$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:

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

We argue that r_n, the last nonzero remainder that appears in this manner, is equal to $gcd(a,b)$. Our proof is based on the lemma below.

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

Proof. If $d = gc(a, b)$, 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 $gc(b, r)$ that $d = gcd(b, r)$.

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

the equations appearing in the Euclidean Algorithm for a and b is multiplied by k, we obtain $ak = q_1(bk) + r_1k$ $0 < r_1k < bk$ bk $q_2(r_1k) + r_2k$ $0 < r_2k < r_1k$

=

.

.

.

$r_{n-2}k = q_n(r_{n-1}k) + r_nk$ $0 < r_nk < r_{n-1}k$ $r_{n-1}k = q_{n+1}(r_nk) + 0$

But this is clearly the Euclidean Algorithm applied to the integers ak and bk, so that their greatest common divisor is the r_nk; that is, $gcd(ka,kb) = r_nk = k.gcd(a,b)$.

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

Proof. It suffices to consider the case in which $k < 0$. Then $-k = |k| > 0$ and, by Theorem

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

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

$a|m$ and $b|m$.

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

Theorem 1.5.6. *For positive integers a and b,* $gc(a, b)lcm(a, b) = ab$

Proof. Let $d = gc(a, b)$. It follows that $a = dr$, $b = ds$ for integers r and s. If $m = ab/d$, then $m = as = rb$. (put $a = dr$ gives $m = rb$ and put $b = ds$ given $m = as$). This shows that m is a (positive) common multiple of a and b.

Now let c be any positive integer that is a common multiple of a and b; say, for definiteness, $c = au = bv$. As we know, there exist integers x and y satisfying $d = ax + by$. In consequence, $\frac{c}{m} = \frac{c}{\frac{ab}{d}} = \frac{cd}{ab} = \frac{c(ax+by)}{ab} = \frac{c}{b}x + \frac{c}{a}y = vx + uy$.

This equation states that $m|c$, allowing us to conclude that $m \leq c$. Thus, in accordance with Definition of lcm, $m = lcm(a,b)$; that is, $lcm(a, b) = \frac{ab}{d} = \frac{ab}{\gcd(a,b)}$. The theorem follows.

Corollary. *For any choice of positive integers a and b,* $lc(a, b) = ab$ *if and only if* $gcd(a, b) = 1$. **The Diophantine Equation** *ax+by =c*

Theorem. *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*

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

where t is an arbitrary integer.

Proof. To establish the second assertion of the theorem, let us suppose that a solution x_0, y_0 of the given equation is known. If x', y' is any other solution, then

$ax_0 + by_0 = c = ax' + by'$

which is equivalent to $a(x' -$

$x_0) = b(y_0 - y')$

By the corollary to Theorem 1.4.8, there exist relatively prime integers r and S such that $a = dr$, $b = ds$. Substituting these values into the last-written equation and canceling the common factor d, we find that $r(x' - x_0) = s(y_0 - y')$

The situation is now this: $r|s(y_0 - y')$, with $gcd(r,s) = 1$. Using Euclid's lemma, it must be the case that $r|(y_0 - y')$; or, in other words, $y_0 - y' = rt$ for some integer t. Substituting, we obtain

$x' - x_0 = st$

This leads us to the formulas

$x' = x_0 + st = x_0 + \frac{b}{d})\, t$

$y' = y_0 - rt = y_0 - \frac{a}{d})\, t$

It is easy to see that these values satisfy the Diophantine equation,regardless of the choice of the integer t; for

$$\begin{aligned} ax' + by' &= a\left[x_0 + (\frac{b}{d})t\right] + b\left[y_0 - (\frac{a}{d})t\right] \\ &= (ax_0 + by_0) + \left(\frac{ab}{d} - \frac{ab}{d}\right)t \\ &= c + 0.t \\ &= c \end{aligned}$$

Thus, there are an infinite number of solutions of the given equation, one for each value of

t.

Corollary *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.

The Fundamental Theorem of Arithmetic

Definition. *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.*

Theorem . *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 $gcd(p,a) = 1$. (In general, $gcd(p,a) = p$ or $gcd(p,a) = 1$ according as $p|a$ or $p \nmid a$.) Hence, citing

Euclid's lemma, we get $p|b$.

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

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. 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 $p|a_1a_2 \cdots a_n$. From Theorem, 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 $1 \le k \le n - 1$. In any event, p divides one of the integers $a_1,a_2,...,a_n$.

Corollary. *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 $1 \le k \le n$.*

Proof. By virtue of Corollary above, we know that $p|q_k$ for some k, with $1 \le k \le n$.

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 $p = q_k$. 2

Theorem(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 or it is composite; in the former case, there is nothing more 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:

$n = p_1p_2p_3n_3$ $1 < n_3 < n_2$ 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

$n = p_1 p_2 \cdots p_k$

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

$n = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s \qquad r \le s$ where the p_i and q_j are all primes,

written in increasing magnitude so that $p_1 \le p_2 \le \cdots p_r$ and $q_1 \le q_2 \le \cdots q_s$

Because $p_1 | q_1 q_2 \cdots q_s$, Corollary it follows that $p_1 = q_k$ for some k; but then $p_1 \ge q_1$. Similar reasoning gives $q_1 \ge p_1$, whence $p_1 = q_1$. We may cancel this common factor and obtain $p_2 p_3 \cdots p_r = q_2 q_3 \cdots q_s$

Now repeat the process to get $p_2 = q_2$ and, in turn, $p_3 p_4 \cdots p_r = q_3 q_4 \cdots q_s$

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. *Any positive integer $n > 1$ can be written uniquely in a canonical form* $n = p_1^{k_1} p_2^{k_2} \cdots p_3^{k_r}$ *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 (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 $b | a^2$. If $b > 1$, then the Fundamental Theorem of Arithmetic guarantees the existence of a prime p such that $p|b$. It follows that $p|a^2$ and, by Theorem, that $p|a$; hence, $gcd(a,b) \ge p$. We therefore arrive at a contradiction, unless $b = 1$. But if this happens, then $a^2 = 2$, which is impossible (we assume that the reader is willing to grant that no integer can be multiplied by itself to give 2). Our supposition that $\sqrt{2}$ is a rational number is untenable, and so $\sqrt{2}$ must be irrational. 2

Theorem (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 may put Theorem to work once again and conclude that 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 $p_1, p_2, \cdots, p_n$. 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.

Theorem. *If p_n is the nth prime number, then* $p_n \leq 2^{2^{n-1}}$.

Proof. Let us proceed by induction on *n*, the asserted inequality being clearly true when *n* = 1. As the hypothesis of the induction, we assume that *n* > 1 and that the result holds for all integers up to *n*. Then

$$p_{n+1} \leq p_1 p_2 \cdots p_n + 1$$
$$\leq 2.2^2 \cdots 2^{2^{n-1}} + 1$$
$$= 2^{1+2+2^2+\cdots 2^{n-1}} + 1$$

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

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

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

$$p_{n+1} \leq 2_{2^{n-1}} + 2_{2^{n-1}}$$
$$= 2.2_{2^{n-1}} = 2_{2^n}$$

completing the induction step, and the argument.

Corollary. *For n* ≥ 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}.

Basic properties of congruence

Definition Let *n* be a fixed positive integer. Two integers *a* and *b* are said to be *congruent modulo n*, symbolized by $a \equiv (mod\ n)$ if *n* divides the difference *a* − *b*; that is, provided that *a* − *b* = *kn* for some integer *k*.

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

Proof. First take $a \equiv (mod\ n)$, 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, $a = b + kn = (qn + r) + kn = (q + k) + r$ 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 $a - b = (q_1 n + r) - (q_2 n + r) = (q_1 - q_2)$ whence $n|a - b$. In the language of congruences, we have $a \equiv (mod\ n)$.

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

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

(b) *If* $a \equiv b (mod\ n)$*, then* $b \equiv a (mod\ n)$.

(c) *If* $a \equiv b (mod\ n)$ *and* $b \equiv c (mod\ n)$*, then* $a \equiv c (mod\ n)$.

(d) *If $a \equiv b(\text{mod } n)$ and $c \equiv d(\text{mod } n)$, then $a + c \equiv b + d(\text{mod } n)$ and $ac \equiv bd(\text{mod } n)$.*

(e) *If $a \equiv b(\text{mod } n)$, then $a + c \equiv b + c(\text{mod } n)$ and $ac \equiv be(\text{mod } n)$. (f) If $a \equiv b(\text{mod } n)$, then $ak \equiv bk(\text{mod } n)$ for any positive integer k.*

Proof. For any integer a, we have $a - a = 0 \cdot n$, so that $a \equiv a(\text{mod } n)$. Now if $a \equiv b(\text{mod } n)$, then $a - b = kn$ for some integer k. Hence, $b - a = -(kn) = (-k)n$ and because $-k$ is an integer, this yields property (b).

Property (c) is slightly less obvious: Suppose that $a \equiv b(\text{mod } n)$ and also $b \equiv c(\text{mod } n)$. Then there exist integers h and k satisfying $a - b = hn$ and $b - c = kn$. It follows that $a - c = (a - b) + (b - c) = hn + kn = (h + k)n$ which is $a \equiv c(\text{mod } n)$ in congruence notation.

In the same vein, if $a \equiv b(\text{mod } n)$ and $c \equiv d(\text{mod } n)$, then we are assured that $a - b = k_1 n$ and $c - d = k_2 n$ for some choice of k_1 and k_2. Adding these equations, we obtain

$$(a + c) - (b + d) = (a - b) + (c - d) = k_1 n + k_2 n = (k_1 + k_2)$$

or, as a congruence statement, $a + c \equiv b + d(\text{mod } n)$. As regards the second assertion of property (d), note that $ac = (b + k_1 n)(d + k_2 n) = bd + (bk_2 + dk_1 + k_1 k_2 n)n$

Because $bk_2 + dk_1 + k_1 k_2 n$ is an integer, this says that $ac - bd$ is divisible by n, whence $ac \equiv bd(\text{mod } n)$.

The proof of property (e) is covered by (d) and the fact that $c \equiv c(\text{mod } n)$. Finally, we obtain property (f) by making an induction argument. The statement certainly holds for $k = 1$, and we will assume it is true for some fixed k. From (d), we know that $a \equiv b(\text{mod } n)$ and $a^k \equiv b^k(\text{mod } n)$ together imply that $aa^k \equiv bb^k(\text{mod } n)$, or equivalently $a^{k+1} \equiv b^{k+1}(\text{mod } n)$. This is the form the statement should take for $k + 1$, and so the induction step is complete.

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

Proof. By hypothesis, we can write

$$c(a - b) = ca - cb = kn$$

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 the displayed equation 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 may be recast as $a \equiv b(\text{mod } s)$; in other words, $a \equiv b(\text{mod } n/d)$.

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

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

Proof. The conditions $p \nmid c$ and p a prime imply that $gcd(c,p) = 1$.

Binary and Decimal Representations of Integers

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

Proof. Because $a \equiv b(\text{mod } n)$, part(f) of Theorem 3.1.4 can be applied to give $a^k \equiv b^k(\text{mod } n)$ for $k = 0,1,\cdots,m$. Therefore, $c_k a^k \equiv c_k b^k(\text{mod } n)$ for all such k. Adding these $m + 1$ congruences, we conclude that

$$\sum_{k=0}^{m} c_k a^k = \sum_{k=0}^{m} c_k b^k (mod\ n)$$

or, in different notation, $P(a) \equiv P(b)(mod\ n)$. 2

Corollary *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. 2

Theorem *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,* $0 \le a_k < 10$*, and let* $S = a_0 + a_1 + \cdots + a_m$*. Then* $9|N$ *if and only if* $9|S$*.*

Proof. Consider $P(x) = \sum_{k=0}^{m} a_k x^k$, a polynomial with integral coefficients. The key observation is that $10 \equiv 1(mod\ 9)$, whence by Theorem 3.2.2, $P(10) \equiv P(l)(mod\ 9)$. But $P(10) = N$ and $P(1) = a_0 + a_1 + \cdots + a_m = S$, so that $N \equiv S(mod\ 9)$. It follows that $N \equiv 0(mod\ 9)$ if and only if $S \equiv 0(mod\ 9)$, which is what we wanted to prove.

Theorem *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,* $0 \le a_k < 10$*, and let* $T = a_0 - a_1 + a_2 - \cdots + (-l)^m a_m$*. Then* $11|N$ *if and only if* $11|T$*.*

Proof. As in the proof of Theorem 3.2.4, put $P(x) = \sum_{k=0}^{m} a_k x^k$. Because $10 \equiv -1(mod\ 11)$, we get $P(10) \equiv P(-1)(mod\ 11)$. But $P(10) = N$, whereas $P(-1) = a_0 - a_1 + a_2 - \cdots + (-l)^m a_m = T$, so that $N = T(mod\ 11)$. The implication is that either both N and T are divisible by 11 or neither is divisible by 11.

Linear Congruence and The Chinese Remainder Theorem

Theorem. *The linear congruence* $ax = b(mod\ n)$ *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.*

Proof. We already have observed that the given congruence is equivalent to the linear Diophantine equation $ax - ny = b$. From Theorem 1.6.1, it is known that the latter equation can be solved if and only if $d|b$; moreover, if it is solvable and x_0, y_0 is one specific solution, then any other solution has the form

$$x = x_0 + \frac{n}{d}t \qquad y = y_0 + \frac{a}{d}t$$

for some choice of t.

Among the various integers satisfying the first of these formulas, consider those that occur when t takes on the successive values $t = 0,1,2,\cdots,d-1$:

$$x_0, x_0 + \frac{n}{d}, x_0 + \frac{2n}{d}, \cdots, x_0 + \frac{(d-1)n}{d}$$

We claim that these integers are incongruent modulo n, and all other such integers x are congruent to some one of them. If it happened that

$$x_0 + \frac{n}{d}t_1 \equiv x_0 + \frac{n}{d}t_2 (mod\ n)$$

where $0 \le t_1 < t_2 \le d-1$, then we would have $\frac{n}{d}t_1 \equiv \frac{n}{d}t_2 (mod\ n)$

Now $gcd(n/d,n) = n/d$, and therefore by Theorem 2.1.7 the factor n/d could be canceled to arrive at the congruence $t_1 \equiv t_2(mod\ d)$ which is to say that $d|t_2 - t_1$. But this is impossible in view of the inequality $0 < t_2 - t_1 < d$.

It remains to argue that any other solution $x_0 + (n/d)t$ is congruent modulo n to one of the d integers listed above. The Division Algorithm permits us to write t as $t = qd + r$, where $0 \leq r \leq d - 1$. Hence

$$\begin{aligned} x_0 + \frac{n}{d}t &= x_0 + \frac{n}{d}(qd + r) \\ &= x_0 + nq + \frac{n}{d}r \\ &= x_0 + \frac{n}{d}r(mod\ n) \end{aligned}$$

with $x_0 + (n/d)r$ being one of our d selected solutions. This ends the proof. 2

Corollary *If* $gcd(a,n) = 1$*, then the linear congruence* $ax \equiv b(mod\ n)$ *has a unique solution modulo n.*

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

$a_1(mod\ n_1)\ x \equiv a_2(mod\ n_2)$

.

.

.

$x \equiv a_r(mod\ n_r)$ *has a simultaneous solution, which is unique modulo the integer*

$n_1 n_2 \cdots n_r$. **Proof.** We start by forming the product $n = n_1 n_2 \cdots n_r$. For each $k =$

$1,2,\cdots,r$, let

,

$$N_k = \frac{n}{n_k} = n_1 \cdots n_{k-1} n_{k+1} \cdots n_r$$

In words, N_k is the product of all the integers n_i with the factor n_k omitted. By hypothesis, the n_i are relatively prime in pairs, so that $gcd(N_k, n_k) = 1$. According to the theory of a single linear congruence, it is therefore possible to solve the congruence

$N_k x \equiv 1(mod\ n_k)$; call the unique solution x_k. Our aim is to prove that the integer $\bar{x}$

$= a_1 N_1 x_1 + a_2 N_2 x_2 + \cdots + a_r N_r x_r$ is a simultaneous solution of the given system.

First, observe that $N_i \equiv 0(mod\ n_k)$ for $i \neq k$, because $n_k | N_i$ in this case. The result is $\bar{x} =$

$a_1 N_1 x_1 + \cdots + a_r N_r x_r \equiv a_k N_k x_k (mod\ nk)$

But the integer x_k was chosen to satisfy the congruence $N_k x \equiv 1(mod\ n_k)$, which forces $\bar{x} \equiv$

$a_k \cdot 1 \equiv a_k(mod\ n_k)$

This shows that a solution to the given system of congruences exists.

As for the uniqueness assertion, suppose that x' is any other integer that satisfies these congruences. Then $\bar{x} \equiv a_k \equiv x'(mod\ n_k) \quad k = 1,2,\cdots,r$ and so $n_k | \bar{x} - x'$ for each value of k. Because $gcd(n_i, n_j) = 1$, Corollary 2 to Theorem

1.4.8 supplies us with the crucial point that $n_1 n_2 \cdots n_r | \bar{x} - x'$; hence $\bar{x} \equiv x'(mod\ n)$.

With this, the Chinese Remainder Theorem is proven.

Theorem *The system of linear congruences* $ax + by \equiv r(mod\ n)$; $cx + dy \equiv s(mod\ n)$ *has a unique solution modulo n whenever* $gcd(ad - bc, n) = 1$.

Proof. Let us multiply the first congruence of the system by d, the second congruence by b, and subtract the lower result from the upper. These calculations yield

$(ad - bc)x \equiv dr - bs(mod\ n)$ (3.1)

The assumption $gcd(ad - bc, n) = 1$ ensures that the congruence

$(ad - bc)z \equiv 1(mod\ n)$ possess a unique solution; denote the solution by t. When congruence (3.1) is multiplied by t, we obtain $x \equiv t(dr - bs)(mod\ n)$

A value for y is found by a similar elimination process. That is, multiply the first congruence of the system by c, the second one by a, and subtract to end up with

$(ad - bc)y \equiv as - cr(mod\ n)$ Multiplication of this congruence by t leads to $y \equiv t(as - cr)(mod\ n)$

A solution of the system is now established.

Part-A

1	Using division algorithm, prove that the cube of nay integer has one of the forms $9k$, $9k + 1$ or $9k + 8$
2	Prove that $3a^2 - 1$ is never a perfect square.
3	Determine all the solutions in positive integers of the Diophantine equation $18x + 5y = 48$.
4	Prove that $\sqrt{2}$ is irrational.
5	Find all the prime numbers that divide 50!
6	Prove that prime factorization of any positive integer $n > 1$ is unique.
7	State and prove any one property of congruent modulo n.
8	State and prove the divisibility condition for 5.
9	Find all solution to the linear congruence $3x - 7y \equiv 11 (\mathrm{mod} 13)$.
10	Show that if gcd(a, n) = 1, then the linear congruence $ax \equiv 1 (\mathrm{mod}\, n)$ has a unique solution modulo n.

Part-B

1	State and prove division algorithm.
2	Prove that if a and b are integers, with $b>0$, then there exists unique integers q an d r satisfying $a = qb + r$, where $2b \le r \le 3b$
3	Show that for non-zero integers a and b, there exits unique integer x and y such that gcd(a, b)=$ax+by$
4	State and prove Euclid's lemma.
5	Show that the linear Diophantine equation $ax + by = c$ has a solution if and only if $d \mid c$, where d = gcd(a, b).
6	State and prove fundamental theorem of arithmetic.
7	State and prove Euclid's theorem on number of primes.
8	Show that if $P(x) = \sum_{k=0}^{n} c_k x^k$ be a polynomial function of x with integral coefficients c_k, and $a \equiv b (\mathrm{mod}\, n)$, then $P(a) \equiv P(b) (\mathrm{mod}\, n)$.
9	State and prove Chinese remainder theorem.
10	Show that the system of linear congruence, $ax + by \equiv r (\mathrm{mod}\, n)$; $cx + dy \equiv s (\mathrm{mod}\, n)$ has a unique solution modulo n, whenever gcd($ad - bc$, n) = 1.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF SCIENCE AND HUMANITIES
DEPARTMENT OF MATHEMATICS

UNIT – II –Number Theoretic Functions– SMTA5203

2.1 Fermat's Little Theorem and Pseudo primes

Theorem: (Fermat's theorem)
Let p be a prime and suppose that $p \nmid a$. Then, $a^{p-1} \equiv 1(mod\ p)$.

Proof. Consider the positive integers:

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

The first $p-1$ multiples of a.

The $p-1$ multiples are mutually incongruent and not a multiple of p. Suppose that a randomly selected two multiples are congruent modulo p to each other,

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

then as $p \nmid a$, dividing by a we get:

$$r \equiv s(mod\ p)$$

which is impossible.

It then follows by Euclid's lemma that the set of multiples should be congruent modulo p to $1,2,3,\cdots,p-1$, taken in some order.

Multiplying all these congruencies together, we get:

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

Grouping the common factors together:

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

Since, $p \nmid (p-1)$, dividing both sides by $(p-1)!$ It follows that

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

The theorem follows.

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

Proof. Suppose $p|a$, then, $a^p \equiv 0 \equiv a(mod\ p)$. Hence the statement is proved.
If $p \nmid a$, then according to Fermat's theorem, we have

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

When this congruence is multiplied by a, the conclusion

$$a^p \equiv a(mod\ p)$$

The statement follows.

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

Proof. From the corollary prove above, letting $a = a^q$, we get

$$(a^q)^p \equiv a^q(mod\ p)$$

By our hypothesis

$$a^q \equiv a(mod\ p)$$

From these congruencies we obtain

$$a^{pq} \equiv a(mod\ p)$$

Hence the proof.

Definition
A composite integer n is called pseudoprime whenever, $n|2^n - 2$. In general, a composite integer n for which $a^n \equiv a\ (mod\ n)$ is called a pseudoprime to the base a.

The composite numbers n which are pseudoprimes to every base a; that is, $a^n \equiv a(mod\ n)$ for all integers a are called absolute pseudoprimes.

Theorem: If n is an odd pseudo prime, then $M_n = 2^n - 1$ is also an odd pseudo prime.

Proof. By the definition of pseudo prime, n is a composite number and hence there exists non trivial factors r and s such that $n = rs$, with $1 < r \leq s < n$. Hence, $2^r - 1|2^n - 1$, or equivalently $2^r - 1|M_n$. It is clear that M_n is a composite number as it has a non-trivial factor $2^r - 1$.
Since n is a pseudo prime, $2^n \equiv 2(mod\ n)$. Hence $2^n - 2 = kn$ for some integer k. It follows that
Hence, $2^{M_n-1} = 2^{2\,n-1-1} = 2^{2n-2} = 2^{kn}$.
It follows that:

$$2^{M_n-1} - 1 = 2^{kn} - 1 = (2^n - 1)(2^{n(k-1)} + 2^{n(k-2)} + \cdots + 2^n + 1)$$
$$\Rightarrow 2^{M_n-1} - 1 = M_n\big((2^{n(k-1)} + 2^{n(k-2)} + \cdots + 2^n + 1)\big)$$

and hence $2^{M_n} - 1 \equiv 0(mod\ M_n) \Rightarrow M_n|(2^{M_n} - 2)$.
This proves that M_n is a pseudoprime.

Definition
An integer is said to be square-free if it is not divisible by the square of any integer greater than 1.

Theorem: Let n be a composite square-free integer, say, $n = p_1p_2 \cdots p_r$, where the p_i are distinct primes. If $p_i - 1|n - 1$ for $i = 1,2,\cdots,r$, then n is an absolute pseudo prime.

Proof. Suppose that a is an integer satisfying, $gcd(a,n) = 1$, so that $gcd(a,p_i) = 1$ for each i. Then by Fermat's theorem $p_i|a^{p_i-1} - 1$.

Given that, $p_i - 1|n - 1 \Rightarrow n - 1 = k(p_i - 1)$.
Since $gcd(a,p_i) = 1$, for some prime p_i, $gcd(a^k,p_i) = 1$. We have $p_i|(a^k)^{p_i-1} - 1$ and therefore $p_i|a^{n-1} - a \Rightarrow p_i|a^n - a$ for all a and $i = 1,2,\cdots,r$.

Hence $p_1p_2 \ldots p_r|a^n - a \Rightarrow n|a^n - a$. This proves that n is an absolute pseudoprime.

2.2 Wilson's Theorem

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

Proof. The cases $p = 2$ and $p = 3$ as being evident, consider $p > 3$. Suppose that a is any one of the $p - 1$ positive integers

$$1,2,3,\cdots,p-1$$

and consider the linear congruence $ax \equiv 1(mod\ p)$. It is cleat that $gcd(a,p) = 1$ and hence the 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 $(a-1)\cdot(a+1) \equiv 0(mod\ p)$. 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 $\frac{p-3}{2}$ congruencies are multiplied together and the factors rearranged, we get

$$2\cdot 3\cdots(p-2) \equiv 1(mod\ p)$$

or rather

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

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

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

as was to be proved.

Example: A concrete example should help to clarify the proof of Wilson's theorem. Specifically, let us take $p = 13$. It is possible to divide the integers $2,3,\cdots,11$ into $(p-3)/2 = 5$ pairs, each product of which is congruent to 1 modulo 13. To write these congruences out explicitly:

$$\begin{aligned} 2 \cdot 7 &= 1(mod 13) \\ 3 \cdot 9 &= 1(mod 13) \\ 4 \cdot 10 &= 1(mod 13) \\ 5 \cdot 8 &= 1(mod\ 13) \\ 6 \cdot 11 &= 1(mod\ 13) \end{aligned}$$

Multiplying these congruences gives the result

$$11! = (2 \cdot 7)(3 \cdot 9)(4 \cdot 10)(5 \cdot 8)(6 \cdot 11) \equiv 1(mod\ 13)$$

and so

$$12! \equiv 12 \equiv -1(mod\ 13)$$

Thus, $(p-1)! \equiv -1(mod\ p)$, with $p = 13$.

Theorem 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. Let a be any solution of $x^2 + 1 \equiv 0(mod\ p)$, so that $a^2 \equiv -1(mod\ p)$. It follows that $p \nmid a$, the outcome of applying Fermat's theorem is

$$1 \equiv a^{p-1} \equiv (a^2)^{\frac{p-1}{2}} \equiv (-1)^{\frac{p-1}{2}} (mod\ p)$$

The possibility that $p = 4k + 3$ for some k does not arise. If it did, we would have

$$(-1)^{\frac{p-1}{2}} = (-1)^{2k+1} = -1$$

hence, $1 \equiv 1(mod\ p)$. The net result of this is that $p|2$, which is patently false.
Therefore, p must be of the form $4k + 1$ or equivalently $p \equiv 1(mod\ 4)$.

Conversly,

$$(p-1)! = 1 \cdot 2 \cdots \tfrac{p-1}{2} \cdot \tfrac{p+1}{2} \cdots (p-2)(p-1)$$

we have the congruences

$$\begin{aligned} p-1 &\equiv -1(mod\ p) \\ p-2 &\equiv -2(mod\ p) \\ &\vdots \\ \frac{p+1}{2} &\equiv -\frac{p-1}{2}(mod\ p) \end{aligned}$$

Rearranging the factors produces

$$\begin{aligned} (p-1)! &\equiv 1 \cdot (-1) \cdot 2 \cdot (-2) \cdots \frac{p-1}{2} \cdot \left(-\frac{p-1}{2}\right) (mod\ p) \\ &\equiv (-1)^{(p-l)/2} \left(1 \cdot 2 \cdots \frac{p-1}{2}\right)^2 (mod\ p) \end{aligned}$$

because there are $(p-1)/2$ minus signs involved. It is at this point that Wilson's theorem can be brought to bear; for, $(p-1)! \equiv -1(mod\ p)$, whence

$$-1 \equiv (-1)^{(p-1)/2} \left[\left(\tfrac{p-1}{2}\right)!\right]^2 (mod\ p)$$

If we assume that p is of the form $4k + 1$, then $(-1)^{(p-1)/2} = 1$, leaving us with the congruence

$$-1 \equiv \left[\left(\frac{p-1}{2}\right)!\right]^2 (mod\ p)$$

The conclusion is that the integer $[(p - 1)/2]!$ satisfies the quadratic congruence
$x^2 + 1 \equiv 0 (mod\ p)$.

2.3 The sum and number of divisors

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

Theorem. 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 $0 \le a_i \le k_i (i = 1,2,\cdots, r)$.

Proof. Note that the divisor $d = 1$ is obtained when $a_1 = a_2 = \cdots = a_r = 0$, and n itself occurs when $a_1 = k_1, a_2 = k_2, \cdots, a_r = k_r$. Suppose that d divides n non trivially; say, $n = dd'$, where $d > 1, d' > 1$. Express both d and d' as products of (not necessarily distinct) primes:

$$d = q_1 q_2 \cdots q_s; d' = t_1 t_2 \cdots t_u$$

with q_i, t_j prime. Then

$$p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} = q_1 \cdots q_s t_1 \cdots t_u$$

are two prime factorizations of the positive integer n. By the uniqueness of the prime factorization, each prime q_i must be one of the p_j. Collecting the equal primes into a single integral power, we get

$$d = q_1 q_2 \cdots q_s = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$$

where the possibility that $a_i = 0$ is allowed.

Conversely, every number $d = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$ turns out to be a divisor of n. For we can write

$$\begin{aligned} n &= p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} \\ &= \left(p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}\right)\left(p_1^{k_1 - a_1} p_2^{k_2 - a_2} \cdots p_r^{k_r - a_r}\right) \\ &= dd' \end{aligned}$$

With $d' = p_1^{k_1 - a_1} p_2^{k_2 - a_2} \cdots p_r^{k_r - a_r}$ and $k_i - a_i \ge 0$ for each i. Then $d' > 0$ and $d|n$.

Theorem *If $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ is the prime factorization of $n > 1$, then*
(a) $\tau(n) = (k_1 + 1)(k_2 + 1)\cdots(k_r + 1)$, and
(b) $\sigma(n) = \frac{p_1^{k_1+1}-1}{p_1-1} \frac{p_2^{k_2+1}-1}{p_2-1} \cdots \frac{p_r^{k_r+1}-1}{p_r-1}$.

Proof: The positive divisors of n are precisely those integers

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

where $0 \le a_i \le k_i$. There are $k_1 + 1$ choices for the exponent a_1; $k_2 + 1$ choices for $a_2, \cdots$; and $k_r + 1$ choices for a_r. Hence, there are

$$(k_1 + 1)(k_2 + 1)\cdots(k_r + 1)$$

possible divisors of n.

To evaluate $\sigma(n)$, consider the product

$$(1 + p_1 + p_1^2 + \cdots + p_1^{k_1})(1 + p_2 + p_2^2 + \cdots + p_2^{k_2}) \cdots (1 + p_r + p_r^2 + \cdots + p_r^{k_r})$$

Each positive divisor of n appears once and only once as a term in the expansion of this product, so that

$$\sigma(n) = (1 + p_1 + p_1^2 + \cdots + p_1^{k_1}) \cdots (1 + p_r + p_r^2 + \cdots + p_r^{k_r})$$

Applying the formula for the sum of a finite geometric series to the ith factor on the right-hand side, we get

$$1 + p_i + p_i^2 + \cdots + p_i^{k_i} = \frac{p_i^{k_i+1}-1}{p_i-1}$$

It follows that

$$\sigma(n) = \frac{p_1^{k_1+1}-1}{p_1-1} \frac{p_2^{k_2+1}-1}{p_2-1} \cdots \frac{p_r^{k_r+1}-1}{p_r-1}$$

2

Example. The number $180 = 2^2 \cdot 3^2 \cdot 5$ has

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

positive divisors. These are integers of the form

$$2^{a_1} \cdot 3^{a_2} \cdot 5^{a_3}$$

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

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

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

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

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

Theorem *The functions τ and σ are both multiplicative functions.*

Proof. Let *m* and *n* be relatively prime integers. Because the result is trivially true if either *m* or *n* is equal to 1, we may assume that $m > 1$ and $n > 1$. If

$$m = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r} \qquad \text{and} \quad n = q_1^{j_1}q_2^{j_2}\cdots q_s^{j_s}$$

are the prime factorizations of *m* and *n*, then because $gcd(m,n) = 1$, no p_i can occur among the q_j. It follows that the prime factorization of the product *mn* is given by

$$mn = p_1^{k_1}\cdots p_r^{k_r}q_1^{j_1}\cdots q_s^{j_s}$$

Appealing to Theorem 5.1.3, we obtain

$$\tau(mn) = [(k_i+1)\cdots(k_r+1)][(j_1+1)\cdot(j_s+1)]$$

$$= \tau(m)\tau(n)$$

In a similar fashion, Theorem 5.1.3 gives

$$\sigma(mn) = \left[\frac{p_1^{k_1+1}-1}{p_1-1}\cdots\frac{p_r^{k_r+1}-1}{p_r-1}\right]\left[\frac{q_1^{j_1+1}-1}{q_1-1}\cdots\frac{q_s^{j_s+1}-1}{q_s-1}\right]$$

$$= \tau(m)\sigma(n)$$

Thus, *τ* and *σ* are multiplicative functions.

Lemma. *If $gcd(m,n) = 1$, then the set of positive divisors of mn consists of all products d_1d_2, where $d_1|m$, $d_2|n$ and $gcd(d_1,d_2) = 1$; furthermore, these products are all distinct.*

Proof. It is harmless to assume that $m > 1$ and $n > 1, m = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$ and $n = q_1^{j_1}q_2^{j_2}\cdots q_s^{j_s}$ be their respective prime factorizations. In as much as the primes $p_1,\cdots,p_r,q_1,\cdots,q_s$ are all distinct, the prime factorization of *mn* is

$$mn = p_1^{k_1}\cdots p_r^{k_r}q_1^{j_1}\cdots q_s^{j_s}$$

Hence, any positive divisor *d* of *mn* will be uniquely representable in the form

$$d = p_1^{a_1}\cdots p_r^{a_r}q_1^{b_1}\cdots q_s^{b_s} \qquad 0 \le a_i \le k_i, 0 \le b_i \le j_i$$

This allows us to write *d* as $d = d_1d_2$, where $d_1 = p_1^{a_1}\cdots p_r^{a_r}$ divides *m* and $d_2 = q_1^{b_1}\cdots q_s^{b_s}$ divides *n*. Because no p_i is equal to any q_j. we surely must have $gcd(d_1,d_2) = 1$.

Theorem . *If f is a multiplicative function and F is defined by*

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

then F is also multiplicative.

Proof. Let *m* and *n* be relatively prime positive integers. Then

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

$$= \sum_{d_1|m;d_2|n} f(d_1 d_2)$$

n as a product of a divisor d_1 of m and a divisor d_2 of n, where $gcd(d_1,d_2) = 1$. By the definition of a multiplicative function,

$$f(d_1d_2) = f(d_1)f(d_2)$$

It follows that

$$= \sum_{d_1|m;d_2|n} f(d_1)f(d_2)$$

$$= \sum_{d_1|m} f(d_1) \sum_{d_2|n} f(d_2)$$

Corollary. *The functions τ and σ are multiplicative functions.*

Proof. We have mentioned that the constant function $f(n) = 1$ is multiplicative, as is the identity function $f(n) = n$. Because τ and σ may be represented in the form $\tau(n) = \sum_{d|n} 1$ and $\sigma(n) = \sum_{d|n} d$, which are constant functions and hence are multiplicative.

2.4 The Greatest Integer Function

Definition . 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 $x - 1 < [x] \leq x$.

Theorem. *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* $p^k > n$.

Proof. Among the first n positive integers, those divisible by p are $p,2p,\cdots,tp$, where t is the largest integer such that $tp \leq n$; in other words, t is the largest integer less than or equal to n/p (which is to say $t = [n/p]$). Thus, there are exactly $[n/p]$ multiples of p occurring in the product that defines $n!$, namely,

$$p, 2p, \cdots, \left[\frac{n}{p}\right] p$$

The exponent of p in the prime factorization of $n!$ is obtained by adding to the number of integers in Equation (5.3), the number of integers among $1,2,\cdots,n$ divisible by p^2, and then the number divisible by p^3, and so on. Reasoning as in the first paragraph, the integers between 1 and n that are divisible by p^2 are

$$p^2, 2p^2, \cdots, \left[\frac{n}{p^2}\right] p^2$$

which are $[n/p^2]$ in number. Of these, $[n/p^3]$ are again divisible by p:

$$p^3, 2p^3, \cdots, \left[\frac{n}{p^3}\right] p^3$$

After a finite number of repetitions of this process, we are led to conclude that the total number of times p divides $n!$ is

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

2

Example. We would like to find the number of zeros with which the decimal representation of 50! terminates. In determining the number of times 10 enters into the product 50!, it is enough to find the exponents of 2 and 5 in the prime factorization of 50!, and then to select the smaller figure.

By direct calculation we see that

$[50/2] + [50/2^2] + [50/2^3] + [50/2^4] + [50/2^5]$

6 + 3 + 1 = 47

Theorem 6.9 tells us that 2^{47} divides 50!, but 2^{48} does not. Similarly,

$$[50/5] + [50/5^2] = 10 + 2 = 12$$

and so the highest power of 5 dividing 50! is 12. This means that 50! ends with 12 zeros.

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

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

is also an integer.

Proof. The argument rests on the observation that if a and b are arbitrary real numbers, then $[a + b] \leq [a] + [b]$. In particular, for each prime factor p of $r!(n - r)!$,

$$\left[\frac{n}{p^k}\right] \geq \left[\frac{r}{p^k}\right] + \left[\frac{(n-r)}{p^k}\right] \qquad k = 1, 2, \cdots$$

Adding these inequalities, we obtain

$$\sum_{k\geq 1}\left[\frac{n}{p^k}\right] \geq \sum_{k\geq 1}\left[\frac{r}{p^k}\right] + \sum_{k\geq 1}\left[\frac{(n-r)}{p^k}\right]$$

The left-hand side of Equation gives the exponent of the highest power of the prime p that divides $n!$, whereas the right-hand side equals the highest power of this prime contained in $r!(n - r)!$. Hence, p appears in the numerator of $n!/r!(n - r)!$ at least as many times as it occurs in the denominator. Because this holds true for every prime divisor of the denominator, $r!(n - r)!$ must divide $n!$, making $n!/r!(n - r)!$ an integer.

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

Proof. The product of r consecutive positive integers, the largest of which is n, is

$$n(n - 1)(n - 2) \cdots (n - r + 1)$$

Now we have

$$n(n - 1)(n - 2) \cdots (n - r + 1) = \frac{n!}{r!\,(n-r)!}r!$$

Because $n!/r!(n - r)!$ is an integer by the theorem, it follows that $r!$ must divide the product $n(n - 1)\cdots(n - r + 1)$, as asserted.

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

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

Then, for any positive integer N,

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

Proof. We begin by noting that

$$\sum_{n=1}^{N} F(n) = \sum_{n=1}^{N}\sum_{d|n} f(d)$$

The strategy is to collect terms with equal values of $f(d)$ in this double sum. For a fixed positive integer $k \leq N$, the term $f(k)$ appears in P$d_{|n}f(d)$ if and only if k is a divisor of n. (Because each integer has itself as a divisor, the right-hand side of

Equation (5.7) includes $f(k)$, at least once.) Now, to calculate the number of sums $d|n$ $f(d)$ in which $f(k)$ occurs as a term, it is sufficient to find the number of integers among $1,2,\cdots,N$, which are divisible by k. There are exactly $[N/k]$ of them:

P

$$k, 2k, 3k, \cdots, \left[\frac{N}{k}\right]k$$

Thus, for each k such that $1 \le k \le N$, $f(k)$ is a term of the sum $\sum_{d|n} f(d)$ for $[N/k]$ different positive integers less than or equal to N. Knowing this, we may rewrite the double sum in Equation (5.7) as

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

and our task is complete.

Corollary *If N is a positive integer, then*

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

Proof. Noting that $\tau(n) = \sum_{d|n} 1$, we may writer for F and take f to be the constant function $f(n) = 1$ for all n.

Corollary. *If N is a positive integer, then*

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

Example *Consider the case $N = 6$. The definition of τ tells us that*

$$\sum_{n=1}^{6} \tau(n) = 14$$

By above Corollary,

$$\sum_{n=1}^{6}\left[\frac{6}{n}\right] = [6] + [3] + [2] + [3/2] + [6/5] + [1]$$

$$= 6 + 3 + 2 + 1 + 1 + 1$$

$$= 14$$

as it should. In the present case, we also have

$$\sum_{n=1}^{6} \sigma(n) = 33$$

and a simple calculation leads to

$$\sum_{n=1}^{6} n\left[\frac{6}{n}\right] = 1[6] + 2[3] + 3[2] + 4[3/2] + 5[6/5] + 6[1]$$

$$= 16 + 23 + 32 + 41 + 51 + 61$$

$$= 33$$

Definition

For $n \ge 1$, $\phi(n)$ denotes the number of positive integers not exceeding n and relatively prime to n. The function $\phi(n)$is usually called the *Euler phi-function* (*indicator* or *totient)*.

Note:

If n is a prime number, then every integer less than n is relatively prime to it; whence, $\phi(n) = n - 1$.

Theorem

If p is a prime and $k > 0$, then $\phi(p^k) = p^k - p^{k-1} = p^k\left(1 - \frac{1}{p}\right)$

Proof.

Since p is prime, $gcd(n, p^k) = 1$ if and only if $p \nmid n$. There are p^{k-1} integers between 1 and p^k that are divisible by *p,* namely, $p, 2p, 3p, \ldots (p^{k-1})p$. Thus, the set $\{1, 2, \ldots, p^k\}$ contains exactly $p^k - p^{k-1}$integers that are relatively prime to p^k, and so by the definition of the phi-function, $\phi(p^k) = p^k - p^{k-1}$.

Lemma.

Given integers *a, b,* c, $gcd(a, bc) = 1$ if and only if $gcd(a, b) = 1$ and $gcd(a, c) = 1$.

Proof.

Case (i)

Suppose that $gcd(a, bc) = 1$ and let $d = gcd(a, b)$. Then $d|a$ and $d|b$ hence it follows that $d|a$ and $d|bc$. This implies that $gcd(a, bc) = d$, which forces *d* = 1. Similarly it can be proved that $gcd(a, c) = 1$.

Case (ii)

Assume that $gcd(a, b) = 1 = gcd(a, c)$ and $gcd(a, bc) = d_1 > 1$. Then d_1 must have a prime divisor p. Because $d_1|bc$, it follows that $p|bc$; in consequence, $p|b$ or $p|c$. If $p|b$, then (by virtue of the fact that pI a) we have $gcd(a, b) \geq p$, a contradiction. In the same way, the condition $p|c$ leads to the equally false conclusion that $gcd(a, c) \geq p$. Thus, $d_1 = 1$ and the lemma is proven.

Theorem.

The Euler phi function is a multiplicative function. i.e., if m and n are two positive integers such that $\gcd(m, n) = 1$, then $\phi(mn) = \phi(m)\phi(n)$.

Proof.

We know that $\phi(1) = 1$, hence the result obviously holds if either *m* or *n* equals 1. Let us suppose that *m* > 1 and *n* > 1. Arranging the integers from 1 to *mn* in *m* columns of *n* integers each, as follows:

1	2	$\cdots r \cdots$	m
$m + 1$	$(m + 1)$	$\cdots (m + r) \cdots$	$2m$
$(2m + 2)$	$(2m + 2)$	$\cdots (2m + r) \cdots$	3m
.	.	.	.
.	.	.	.
$(n - 1)m + 1$	$(n - 1)m + 2$	$\cdots (n - 1)m + r \cdots$	mn

From the above array of mn elements we have identify numbers that are relatively prime to mn. From the previous lemma it is the same as the number of integers that are relatively prime to both *m* and *n.* We know that, $gcd(qm + r, m) = gcd(r, m)$, the numbers in the r^{th} column are relatively prime to *m* if and only if *r* itself

is relatively prime to *m.* Therefore, only ¢(m) columns contain integers relatively prime to m, and every entry in the column will be relatively prime to *m.* Now the entries in the r^{th} column (where it is assumed that gcd(r, m) = 1) are $r, m + r, 2m + r, \ldots, (n - 1)m + r$. The listed n integers are incongruent to modulo n. For if any two integers are congruent modulo n i.e. $km + r \equiv sm + r(mod\ n), 0 \leq k < s < n \Rightarrow km \equiv sm(mod\ n) \Rightarrow k \equiv s(mod\ n)$. Thus, the numbers in the *rth* column are congruent modulo *n* to 0, 1, 2, ... , *n*-1, in some order. But if $s \equiv t\ \ (mod\ n)$, then $gcd(s, n) = 1$ if and only if $gcd(t, n) = 1$. The implication is that the rth column contains as many integers that are relatively prime to *n* as does the set $\{0, 1, 2, \ldots, n - 1\}$, namely, ¢(n) integers. Therefore, the total number of entries in the array that are relatively prime to both *m* and *n* is ¢(m)¢(n). This completes the proof of the theorem.

Theorem.

If the integer n > 1 has the prime factorization $= p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots p_r^{k_r}$, then $\phi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_r}\right)$

Proof.

Let us prove this theorem by the method of induction, using induction on r, the number of distinct prime factors of n. When $r = 1$, the statement follows from the previous theorem. Since , it is true for $r = 1$, let us assume it is true for $r = i$. i.e., $\phi\left(p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots p_i^{k_i}\right) = p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots p_i^{k_i}\left(\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_i}\right)\right)$

For $r = i + 1$, $\phi\left(p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots p_i^{k_i} p_{i+1}^{k_i+1}\right) = \phi\left(p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots p_i^{k_i}\right)\phi\left(p_{i+1}^{k_i+1}\right)$

$$= p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots p_i^{k_i}\left(\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_i}\right)\right) p_{i+1}^{k_i+1}\left(1 - \frac{1}{p_{i+1}^{k_i+1}}\right)$$

Hence, whenever the statement is true for $n = i$, it is true for $n = i + 1$ by principle of mathematical induction the statement is true for all $n > 1$. This proves the theorem.

Theorem.

For n > 2, $\phi(n)$ is an even integer.

Proof.

If $n > 2$, is prime then $\phi(n) = n - 1$ is even. As every prime number greater than 2 is odd. If n is an even composite number with the prime factorisation $n = p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots p_r^{k_r}$ then $\phi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_r}\right)$ which is even as n is even. If n is odd, then the prime factorization of n involves only the odd prime factors. Let $= p_i^{k_i} m$. Since, Euler's phi function is multiplicative $\phi(n) = \phi\left(p_i^{k_i}\right)\phi(m) = p_i^{k_i - 1}(p_i - 1)\phi(m)$. As $p_i - 1$ is even $\phi(n)$ is even. This proves the theorem

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

$$aa_1, aa_2, \ldots, aa_{\phi(n)}$$

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

Proof. Observe that no two of the integers $aa_1, aa_2, \ldots, aa_{\phi(n)}$ are congruent modulo n. For if $aa_i \equiv aa_j \pmod{n}$, with $1 \leq i < j \leq \phi(n)$, then the cancellation law yields $a_i \equiv a_j \pmod{n}$, and thus $a_i = a_j$, a contradiction. Furthermore, because $\gcd(a_i, n) = 1$ for all i and $\gcd(a, n) = 1$, the lemma preceding Theorem 7.2 guarantees 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 $aa_i \equiv b \pmod{n}$. Because

$$\gcd(b, n) = \gcd(aa_i, n) = 1$$

b must be one of the integers $a_1, a_2, \ldots, a_{\phi(n)}$. All told, this proves that the numbers $aa_1, aa_2, \ldots, aa_{\phi(n)}$ and the numbers $a_1, a_2, \ldots, a_{\phi(n)}$ are identical (modulo n) in a certain order.

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

Proof. There is no harm in taking $n > 1$. Let $a_1, a_2, \ldots, a_{\phi(n)}$ be the positive integers less than n that are relatively prime to n. Because $\gcd(a, n) = 1$, it follows from the lemma that $aa_1, aa_2, \ldots, aa_{\phi(n)}$ are congruent, not necessarily in order of appearance, to $a_1, a_2, \ldots, a_{\phi(n)}$. Then

$$\begin{aligned} aa_1 &\equiv a_1' \pmod{n} \\ aa_2 &\equiv a_2' \pmod{n} \\ &\vdots \qquad\qquad \vdots \\ aa_{\phi(n)} &\equiv a_{\phi(n)}' \pmod{n} \end{aligned}$$

where $a_1', a_2', \ldots, a_{\phi(n)}'$ are the integers $a_1, a_2, \ldots, a_{\phi(n)}$ in some order. On taking the product of these $\phi(n)$ congruences, we get

$$\begin{aligned} (aa_1)(aa_2)\cdots(aa_{\phi(n)}) &\equiv a_1'a_2'\cdots a_{\phi(n)}' \pmod{n} \\ &\equiv a_1a_2\cdots a_{\phi(n)} \pmod{n} \end{aligned}$$

and so

$$a^{\phi(n)}(a_1a_2\cdots a_{\phi(n)}) \equiv a_1a_2\cdots a_{\phi(n)} \pmod{n}$$

Because $\gcd(a_i, n) = 1$ for each i, the lemma preceding Theorem 7.2 implies that $\gcd(a_1a_2\cdots a_{\phi(n)}, n) = 1$. Therefore, we may divide both sides of the foregoing congruence by the common factor $a_1a_2\cdots a_{\phi(n)}$, leaving us with

Hence the proof.

Part-A

1 Using Fermat's theorem, find the prime factorization of 12499.

2 Show that 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 the integers d of the form $d = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$ where $0 \le a_i \le k_i$.

3 Prove that τ and σ are both multiplicative functions.

4 Prove that for any positive integer n and r, the binomial coefficient ${}^nC_r$ is an integer.

5 Show that, if f and F are number-theoretic functions such that $F(n) = \sum_{d|n} f(d)$ then for any positive integer N, $\sum_{n=1}^{N} F(n) = \sum_{k=1}^{N} f(k)\left[\frac{n}{k}\right]$.

6 Show that, for the integer $n > 1$ having the prime factorization $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$, $\phi(n) = \left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_r}\right)$.

Part-B

1 State and prove Fermat's theorem

2 Show that if n is an odd pseudo prime, then $M_n = 2^n - 1$ is a larger one.

3 State and prove Wilson's theorem.

4 If $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ is the prime factorization of $n > 1$, then find $\tau(n)$, $\sigma(n)$.

5 Prove that the function ϕ is an multiplicative function.

6 Prove that if n is a positive integer and p a prime, then the exponent of the highest power of p that divides n! is $\sum_{k=1}^{\infty}\left[\frac{n}{p^k}\right]$.

7 State and prove Euler's theorem.

8 Prove that for n > 1, the sum of the positive integers less than n and relatively prime to n is $\frac{1}{2} n\phi(n)$.

SCHOOL OF SCIENCE AND HUMANITIES
DEPARTMENT OF MATHEMATICS

UNIT – III – Elementary Transformations– SMTA5203

Sub-Matrix

A matrix, which is obtained by deleting some rows or some columns or both of a matrix A, is called a sub-matrix of A. For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 0 & 11 \end{bmatrix}, \text{ then}$$

$\begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 10 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 3 \\ 6 & 7 \\ 10 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 3 \\ 6 & 7 \end{bmatrix}, \begin{bmatrix} 7 & 8 \\ 0 & 11 \end{bmatrix}$, etc. are all sub-matrices of the matrix A.

Particular Case. The matrix A is a sub-matrix of itself.

Minor of a Matrix

If any r rows and any r columns from an $m \times n$ matrix A are retained and the remaining $(m - r)$ rows and $(n - r)$ columns deleted, then the determinant of the remaining $r \times r$ sub-matrix of A is called a minor of A of order r. For example,

$$\text{Let } A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ a_{51} & a_{52} & a_{53} & a_{54} \end{bmatrix}_{5 \times 4}, \text{ then}$$

(i) The elements $a_{11}, a_{12}, a_{21}, a_{32}, a_{44}$, etc. are minors of A of order 1.

(ii) The determinants $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}, \begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix}, \begin{vmatrix} a_{13} & a_{14} \\ a_{53} & a_{54} \end{vmatrix}$, etc. are minors of A of order 2.

(iii) $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{41} & a_{42} & a_{43} \\ a_{51} & a_{52} & a_{53} \end{vmatrix}, \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{52} & a_{53} & a_{54} \end{vmatrix}$, etc. are minors of A of order 3.

(iv) $\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}, \begin{vmatrix} a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ a_{51} & a_{52} & a_{53} & a_{54} \end{vmatrix}$, etc. are minors of A of order 4.

Rank of a Matrix

A natural number r is called the rank of the matrix A if

(i) There exists at least one non-zero minor of order r.

(ii) Every minor of order $(r + 1)$, if any, vanishes. The rank of the matrix A is denoted by $\rho(A)$ or rank (A).

Theorem 1. The rank of a matrix is equal to the rank of the transposed matrix.

Theorem 2. $\rho(A^*) = \rho(A)$

Theorem 3. If A is a non-zero column matrix and B is a non-zero row matrix, then show that $\rho(AB) = 1$.

Example 1.

Find the rank of the matrix A, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 2 & 1 & 2 \end{bmatrix}$$

$|A| = 6 \neq 0$

$\therefore \quad \rho(A) = 3$

Example 2.

Find the rank of the matrix A, where

$$A = \begin{bmatrix} 8 & 0 & 0 & 1 \\ 1 & 0 & 8 & 1 \\ 0 & 0 & 1 & 8 \\ 0 & 8 & 1 & 8 \end{bmatrix}$$

Solution:

$$|A| = \begin{vmatrix} 8 & 0 & 0 & 1 \\ 1 & 0 & 8 & 1 \\ 0 & 0 & 1 & 8 \\ 0 & 8 & 1 & 8 \end{vmatrix}$$

$$= 8\begin{vmatrix} 0 & 8 & 1 \\ 0 & 1 & 8 \\ 8 & 1 & 8 \end{vmatrix} - 1\begin{vmatrix} 1 & 0 & 8 \\ 0 & 0 & 1 \\ 0 & 8 & 1 \end{vmatrix}$$

$= 8[8(64 - 1)] - [1(0 - 8)]$

$= 4032 + 8$

$= 4040 \neq 0$

$\therefore \quad \rho(A) = 4$

Normal Form

By a finite number of elementary transformations, every non-zero matrix A of order $m \times n$ and rank r (> 0) can be

reduced to one of the following forms:

$$\left[\begin{array}{c|c} I_r & O \\ \hline O & O \end{array}\right], \left[\begin{array}{c} I_r \\ \hline O \end{array}\right], [I_r \mid O], [I_r]$$

where I_r denotes identity matrix of order r. Each one of these four forms is called *Normal Form* or *Canonical Form* or *Orthogonal Form*.

Procedure for Reduction to Normal Form

Let $A = [a_{ij}]$ be any matrix of order $m \times n$. Then, we can get the normal form of the matrix A by subjecting it to a finite number of elementary transformations in the following manner:

(1) We first interchange a pair of rows (or columns), if necessary, to obtain a non-zero element (preferably 1) in the first row and first column of the matrix A.

(2) Divide the first row by this non-zero element, if it is not 1.

(3) We subtract appropriate multiples of the elements of the first row from other rows so as to obtain zeroes in the remainder of the first column.

(4) We subtract appropriate multiples of the elements of the first column from other columns so as to obtain zeroes in the remainder of the first row.

(5) We repeat the above four steps starting with the element in the second row and the second column.

(6) Continue this process down the leading diagonal until the end of the diagonal is reached or until all the remaining elements in the matrix are zero.

Theorem

If A is any $m \times n$ matrix of rank r, then there exist non-singular matrices R and C such that

$$RAC = \left[\begin{array}{c|c} I_r & O \\ \hline O & O \end{array}\right]$$

Proof. If A is a matrix of rank r, then it can be transformed into the form $\left[\begin{array}{c|c} I_r & O \\ \hline O & O \end{array}\right]$ by means of elementary transformations. Since elementary row (or column) operations are equivalent to pre (or post) multiplication of the corresponding elementary matrices, therefore, we have the following result:

$$R_p \ldots\ldots R_2 R_1 A C_1 C_2 \ldots\ldots C_q = \left[\begin{array}{c|c} I_r & O \\ \hline O & O \end{array}\right]$$

where $R_1 R_2 \ldots\ldots R_p$; $C_1 C_2 \ldots\ldots C_q$ are elementary matrices corresponding to the row (or column) elementary transformations.

Since the elementary matrices are non-singular, therefore,

$$R_1 R_2 \ldots\ldots R_p = R$$

and $C_1 C_2 \ldots\ldots C_q = C$

will be non-singular matrices.

$$RAC = \left[\begin{array}{c|c} I_r & O \\ \hline O & O \end{array}\right]$$

The matrix $\left[\begin{array}{c|c} I_r & O \\ \hline O & O \end{array}\right]$ is of order $m \times n$. This matrix is called normal matrix and is denoted by N_r. Thus,

$$N_r = RAC$$

which is of the form

$$A = PBQ$$

$$\Rightarrow B = P^{-1}AQ^{-1}$$

Rank of a Matrix Product

Theorem 1. The rank of the product of two matrices cannot exceed the rank of either matrix, i.e.

$\rho(AB) \leq \rho(A)$ and $\rho(AB) \leq \rho(B)$

Proof. Let r_1, r_2, r be the ranks of the matrices A, B, AB respectively. We have to show that

$r \leq r_1$ and $r \leq r_2$

Lemma. If A be an $m \times n$ matrix of rank r, then there exists a non-zero matrix P such that

$$PA = \begin{bmatrix} G \\ O \end{bmatrix}$$

where G is an $r \times n$ matrix of rank r and O is a zero matrix of order $(m - r) \times n$.

Now, there exists a non-singular matrix P and a matrix G of rank r_1 and r_2 such that

$$PA = \begin{bmatrix} G \\ O \end{bmatrix}$$

The matrix P, being a product of elementary matrices, is non-singular.

We have

$$r = \rho(AB) = \rho(PAB)$$

Also,

$$PAB = \begin{bmatrix} G \\ O \end{bmatrix} B$$

has at the most r_1 non-zero rows which arise on multiplying the r_1 non-zero rows of G with columns of B so that

$$\rho(PAB) \leq r_1$$

$$\Rightarrow r \leq r_1$$

$$\Rightarrow \rho(AB) \leq \rho(A)$$

Again,

$$\rho(AB) = \rho(AB)'$$

$$= \rho(B'A') \leq \rho(B') = \rho(B) \qquad \text{| as proved above}$$

$$\therefore \rho(AB) \leq \rho(B)$$

$$\Rightarrow r \leq r_2$$

Theorem 2. The rank of a matrix does not alter by pre-multiplication or post-multiplication with any non-singular matrix.

Proof. Let A be a matrix of order $m \times n$. Let P be a singular matrix of order $n \times n$. Then, the product AP exists and is a $m \times n$ matrix. We have to prove that

$$\text{Rank } (AP) = \text{Rank } (A)$$

Let $B = AP$

Then,

$$BP^{-1} = APP^{-1} = AI = A$$

P^{-1} exists since P is non-singular.

Now, $B = AP$

$$\Rightarrow \text{Rank } (B) = \text{Rank } (AP) \leq \text{Rank } (A)$$

$$\Rightarrow \text{Rank } (B) \leq \text{Rank } (A)$$

Also,

$$A = BP^{-1}$$

$$\Rightarrow \text{Rank } (A) = \text{Rank } (BP^{-1}) \leq \text{Rank } B$$

| By Theorem 1 above

$$\Rightarrow \text{Rank } (A) \leq \text{Rank } (B)$$

$$\text{Rank } (A) = \text{Rank } (B)$$

$$\Rightarrow \text{Rank } (A) = \text{Rank } (AP)$$

Theorem 3. Prove that Rank (AA') = Rank (A)

Proof. Let $B = AA'$. Then,

Rank (B) = Rank (AA')

$\Rightarrow$ Rank $(B) \leq$ Rank (A)

$\Rightarrow A = P^{-1}NQ^{-1}$

$\Rightarrow AB = P^{-1}NQ^{-1}B$

$\Rightarrow O = P^{-1}NQ^{-1}B$ | $\because AB = O$ (given)

$\Rightarrow PO = PP^{-1}NQ^{-1}B$

$\Rightarrow O = NQ^{-1}B$

A is of order $m \times p$, Q is of order $p \times p$ and $Q^{-1}B$ is of order $p \times n$.

$NQ^{-1}B = O$ implies that the first r rows of $Q^{-1}B$ must be zeroes while the remaining $(p - r)$ rows may be arbitrary. Thus, the rank of $Q^{-1}B$ and hence the rank of B cannot exceed $p - r$.

Hence, the theorem.

Theorem

If A is of order n and rank $(n - 1)$, then prove that adj A is of rank 1.

Proof. $\because$ A is of rank $(n - 1)$.

$\therefore$ These exists at least one non-zero cofactor and $|A| = 0$.

Now,

A (adj A) $= |A|\, I = O$ | $\because |A| = O$

$\therefore$ Rank of adj $A = n - (n - 1)$ | By Th. 5.12.

$= 1$ | $\because \rho(A) = n - 1$

Theorem

Show that the equivalence of matrices is an equivalence relation.

Proof. Let A and B be any two matices of order $m \times n$ each. If there exist non-singular matrices P and Q such that $A = PBQ$, then we say that A is equivalent to B and denote it by $A \sim B$.

We see that

1. Reflexivity. For any matrix A of order $m \times n$, there exist two identity matrices I_m and I_n such that

$$A = I_m A I_n$$

$$\Rightarrow A = PAQ$$

where $P = I_m$, $Q = I_n$

So every matrix is equivalent to itself. Hence, the relation of equivalence is reflexive.

2. Symmetry. For any two $m \times n$ matrices A and B, $A \sim B$ $\Rightarrow A = PBQ$ for some non-singular matricces P and Q.

$$\Rightarrow P^{-1}AQ^{-1} = B$$

$$\Rightarrow B \sim A$$

Hence, the relation of equivalence is commutative.

3. Transitivity. For any three matrices of the same order $m \times n$,

$$A \sim B,\ B \sim C \Rightarrow A = PBQ \text{ and}$$

$$B = P_1 C Q_1$$

where P, Q, P_1 and Q_1 are non-singular.

$$\Rightarrow A = P(P_1 C Q_1)Q$$

$$\Rightarrow A = (PP_1)\ C(Q_1 Q)$$

$$\Rightarrow A \sim C \mid \because PP_1 \text{ and } Q_1 Q \text{ are non-singular}$$

Hence, the relation of equivalence is transitive. Since the relation of equivalence is reflexive, symmetric and transitive, therefore, it is an equivalence relation.

Hence, the theorem.

Example 1. Find the rank of the following matrix using elementary transformations:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 2 \\ 2 & 6 & 5 \end{bmatrix}$$

Solution: We have

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 2 \\ 2 & 6 & 5 \end{bmatrix}$$

Operating $R_{21}(-1)$, $R_{31}(-2)$

$$A \sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & -1 \\ 0 & 2 & -1 \end{bmatrix}$$

Operating $R_{32}(-1)$

$$A \sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

The single minor of order 3 is zero.

A minor of order 2 is

$$\begin{vmatrix} 1 & 2 \\ 0 & 2 \end{vmatrix} = 2 \neq 0$$

$\therefore \quad \rho(A) = 2$

Example 2. Reduce the matrix

$$A=\begin{bmatrix} 8 & 1 & 3 & 6 \\ 0 & 3 & 2 & 2 \\ -8 & -1 & -3 & 4 \end{bmatrix}$$

to normal form and find its rank.

Solution: We have

$$A=\begin{bmatrix} 8 & 1 & 3 & 6 \\ 0 & 3 & 2 & 2 \\ -8 & -1 & -3 & 4 \end{bmatrix}$$

Operating $R_{31}(1)$

$$A \sim \begin{bmatrix} 8 & 1 & 3 & 6 \\ 0 & 3 & 2 & 2 \\ 0 & 0 & 0 & 10 \end{bmatrix}$$

$\Rightarrow A \sim [I_3 \; O_{3\times 1}]$
which is the normal form.
Hence, $\rho(A) = 3$

Example 3.

If $A=\begin{bmatrix} 3 & -3 & 4 \\ 2 & -3 & 4 \\ 0 & -1 & 1 \end{bmatrix}$, determine two non-singular

matrices P and Q such that $PAQ = I$. Hence, find A^{-1}.

Solution: Let us write

$$A = I_3 A I_3$$

$$\Rightarrow \begin{bmatrix} 3 & -3 & 4 \\ 2 & -3 & 4 \\ 0 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Operating $R_{12}(-1)$

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & -3 & 4 \\ 0 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Operating $R_{21}(-2)$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 4 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Operating $R_{23}(-4)$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 3 & -4 \\ 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Operating $C_{23}(1)$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 3 & -4 \\ 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\Rightarrow I = PAQ$$

$$P = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 3 & -4 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$PAQ = I$$

Pre-multiplying by P^{-1},

$$P^{-1}PAQ = P^{-1}I$$

$$\Rightarrow IAQ = P^{-1}I \quad | \because P^{-1}P = I$$

$$\Rightarrow AQ = P^{-1}$$

Post-multiplying by P,

$$AQP = P^{-1}P$$

$$\Rightarrow AQP = I$$

Pre-multiplying by A^{-1},

$$A^{-1}AQP = A^{-1}I$$

$$\Rightarrow IQP = A^{-1}I$$

$$\Rightarrow QP = A^{-1}$$

$$\therefore \quad A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ -2 & 3 & -4 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 3 & -4 \\ -2 & 3 & -3 \end{bmatrix}$$

Operating $R_{32}(-1)$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & \frac{1}{3} & -\frac{5}{3} \\ \frac{1}{2} & -\frac{1}{3} & -\frac{1}{6} \end{bmatrix} A \begin{bmatrix} 1 & \frac{4}{7} & \frac{9}{119} & \frac{9}{217} \\ 0 & \frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} \\ 0 & 0 & -\frac{1}{17} & 0 \\ 0 & 0 & 0 & \frac{1}{31} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} I_{2\times 2} & O_{2\times 2} \\ O_{1\times 2} & O_{1\times 2} \end{bmatrix} = PAQ$$

where,

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & \frac{1}{3} & -\frac{5}{6} \\ \frac{1}{2} & -\frac{1}{3} & -\frac{1}{6} \end{bmatrix} \text{ and } Q = \begin{bmatrix} 1 & \frac{4}{7} & \frac{9}{119} & \frac{9}{217} \\ 0 & \frac{1}{7} & -\frac{1}{7} & -\frac{1}{7} \\ 0 & 0 & -\frac{1}{17} & 0 \\ 0 & 0 & 0 & \frac{1}{31} \end{bmatrix}$$

$\therefore \quad \rho(A) = 2$

Part-A

1 Show that equivalent system of linear equations has exactly the same solutions.

2 Show that the inverse of an elementary row operation exists and is an elementary row operation of the same type.

3 If $A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}$ find all the solutions of $AX = 0$ by row-reducing A.

4 If A is a square matrix of order n, then A is row-equivalent to I_n if and only if the system of equations $AX = 0$ has only the trivial solution.

5 Find a row-reduced echelon matrix which is row-equivalent to $A = \begin{bmatrix} 1 & -i \\ 2 & 2 \\ i & 1+i \end{bmatrix}$.

6 Show that the following statements are true for any square matrix A and B over F:
i) If A is invertible then so is A^{-1} and $(A^{-1})^{-1} = A$.
ii) If both A and B are invertible, so is AB and $(AB)^{-1} = B^{-1}A^{-1}$.

7 Prove that an elementary matrix is invertible.

8 A square matrix with left and right inverse is invertible.

Part-B

1 Prove that every m x n matrix over the field F is row-equivalent to a row-reduced matrix.

2 Show that every m x n matrix over the field F is row-equivalent to a row-reduced echelon matrix.

3 Prove that for any A and B, n x n matrices over the field F. the B is row-equivalent to A if and only if $B = PA$, where P is a product of m x m elementary matrices.

4 Show that if A is a m x n matrix and m < n, then the homogenous system of linear equations AX = 0 has a non-trivial solution.

5 Prove that if A and B are n x n matrices over the field F, then the following statements are true:
i) A is invertible, so is A^{-1} and $(A^{-1})^{-1} = A$
ii) If A and B are invertible then so is AB and $(AB)^{-1}=B^{-1}A^{-1}$.

6 Prove that the following statements are equivalent for any square matrix:
i) A is invertible
ii) A is row-equivalent to the n x n identity matrix.
iii) A is a product of elementary matrices.

7 Show that for an n x n matrix A, the following are equivalent:
i) A is invertible
ii) The homogeneous system $AX = 0$ has only the trivial solution $X = 0$.
iii) The system of equations $AX = Y$ has a solution X for each n x 1 matrix Y.

8 Let $A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ -1 & 0 & 3 & 5 \\ 1 & -2 & 1 & 1 \end{bmatrix}$. Find a row-reduced echelon matrix R which is row-equivalent to A and an invertible 3 x 3 matrix P such that $R = PA$.

SCHOOL OF MATHEMATICS
DEPARTMENT OF MATHEMATICS

UNIT – IV – System of Linear Equations – SMTA5203

Linear Equation

An equation of first degree in n unknowns $x_1, x_2, x_3, \ldots, x_n$ is called a linear equation.

Thus,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n = b_1$$

is a linear equation in n unknowns $x_1, x_2, x_3, \ldots, x_n$ with coefficients $a_{11}, a_{12}, a_{13}, \ldots, a_{1n}$ and b_1 as constants.

If $b_1 = 0$, then Eq. (6.1) takes the form

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n = 0$$

and is called a homogeneous linear equation in n unknowns $x_1, x_2, x_3, \ldots, x_n$.

System of Linear Equations

Consider a system of m linear equations in n unknowns ($m > n$, $m = n$ or $m < n$) given below:

$$\left.\begin{array}{r} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots + a_{2n}x_n = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \ldots + a_{mn}x_n = b_m \end{array}\right\} \quad \ldots(A)$$

In matrix notation, these equations can be put in the form

$$AX = B$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{bmatrix}_{m \times n}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n \times 1} \quad \text{and} \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_m \end{bmatrix}_{m \times 1}$$

The matrix A is called the coefficient matrix. The matrix X is called the column matrix of n unknowns $x_1, x_2, x_3, ..., x_n$. The matrix B is called the column matrix of m constants $b_1, b_2, ..., b_m$.

The matrix $C = [A : B]$ obtained by placing the constant column matrix B to the right of the matrix A is called augmented matrix. Thus, the matrix

$$C = [A : B] = \left[\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & ..... & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & ..... & a_{2n} & b_2 \\ ..... & ..... & .... & ..... & ..... & ... \\ ..... & ..... & ..... & ..... & ..... & ... \\ a_{m1} & a_{m2} & a_{m3} & ..... & a_{mn} & b_m \end{array}\right]$$

is called the augmented matrix.

Any set of values of x_1, x_2, x_3, x_n which simultaneously satisfy the system of equation (A) is called the solution of the system (A). If the system has one or more solutions, it is called consistent. If it has no solution, it is called inconsistent. A consistent system has either one solution or infinitely many solutions.

Non-singular or Regular System of Linear Equations

If we take $m = n$ in the system of equations (A), then we have

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & ..... & a_{1n} \\ a_{21} & a_{22} & a_{23} & ..... & a_{2n} \\ ..... & ..... & ..... & ..... & ..... \\ ..... & ..... & ..... & ..... & ..... \\ a_{n1} & a_{n2} & a_{n3} & ..... & a_{nn} \end{bmatrix}$$

$$\therefore \quad |A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & ..... & a_{1n} \\ a_{21} & a_{22} & a_{23} & ..... & a_{2n} \\ ..... & ..... & ..... & ..... & ..... \\ ..... & ..... & ..... & ..... & ..... \\ a_{n1} & a_{n2} & a_{n3} & ..... & a_{nn} \end{vmatrix}$$

$|A|$ is called the determinant of coefficients.

If $|A| \neq 0$, then the system of equations (when $m = n$) is called regular. This system has a unique solution given by

$\frac{x_1}{|A_1|} = \frac{x_2}{|A_2|} = \frac{x_3}{|A_3|} = ... = \frac{x_n}{|A_n|} = \frac{1}{|A|}$ where $|A_i|$ represents the determinant obtained by replacing the i^{th} column of $|A|$ by the column of b's.

This is known as Cramer's Rule for solving a system of n linear equation in n unknowns.

Note. If $|A| = 0$, then Cramer's Rule fails.

When $|A| \neq 0$, then A^{-1} exists. Hence, pre-multiplying the matrix equation

$$AX = B$$

by A^{-1}, we obtain,

$$A^{-1}(AX) = A^{-1}B$$

$$\Rightarrow (A^{-1}A)X = A^{-1}B$$

$$\Rightarrow IX = A^{-1}B$$

$$\Rightarrow X = A^{-1}B$$

This gives the solution of the system of n equations in n unknowns when the system is non-singular (or regular).

Singular System of Linear Equations

If $|A| = 0$, the system of n equations in n unknowns is called singular. This case will be dealt with later on.

System of Linear Equations, in General

Consider the system of linear equations (A) as given in Section 6.2.

It is simple to see that the rank of the augmented matrix C cannot be less than the rank of the coefficient matrix A because every sub-matrix of A is also a sub-matrix of C.

Let $\rho(A) = r$. Then, by a suitable sequence of elementary row operations, the matrix A can be reduced to an equivalent matrix in which each of the first r elements of the leading diagonal is 1 and every element below this diagonal and/or below the r^{th} row is zero. The matrix, so reduced, is said to be in *Echelon Form.*

If the same sequence of elementary operations is performed on the system of equations (A), then these will be transformed in the following form:

$$\left.\begin{aligned} y_1 + \alpha_{12}y_2 + \alpha_{13}y_3 + \ldots\ldots + \alpha_{1n}y_n &= \beta_1 \\ y_2 + \alpha_{23}y_3 + \ldots\ldots + \alpha_{2n}y_n &= \beta_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ y_r + \ldots\ldots + \alpha_{rn}y_n &= \beta_r \\ 0 &= \beta_{r+1} \\ \ldots\ldots &= \ldots\ldots \\ \ldots\ldots &= \ldots\ldots \\ 0 &= \beta_m \end{aligned}\right\} \quad \ldots (B)$$

in which $y_1, y_2, \ldots\ldots, y_n$ is some permutation of $x_1, x_2, \ldots\ldots, x_n$.

Also, the coefficient matrix and the augmented matrix of the system of equations (B) will be equivalent to A and C respectively.

Now the following cases arise:

Case I. When $r = 1$, then the system of equations (B) becomes,

$$\left.\begin{aligned} y_1 + \alpha_{12}y_2 + \alpha_{13}y_3 + \ldots\ldots + \alpha_{1n}y_n &= \beta_1 \\ 0 &= \beta_1 \\ 0 &= \beta_3 \\ &\vdots \\ &\vdots \\ 0 &= \beta_m \end{aligned}\right\} \quad \ldots (C)$$

In this case, $\rho(C) = 2$ or 1, since C has one column more than A.

When $\rho(C) = 2$, $\beta_2, \beta_3, \ldots\ldots, \beta_m$ cannot all be zero. Hence, the equations are inconsistent and there will be no solution.

When $\rho(C) = 2$, $\beta_2, \beta_3, \ldots\ldots, \beta_m$ will be all zero and the system of equations (B) will be equivalent to a single equation from which y_1 will be expressible in terms of $y_2, y_3, \ldots\ldots, y_n$ which can have arbitrary values.

Case II. When $r = 2$, then the system of equations (B) becomes

$$\left.\begin{aligned} y_1 + \alpha_{12}y_2 + \alpha_{13}y_3 + \ldots\ldots + \alpha_{1n}y_n &= \beta_1 \\ y_2 + \alpha_{23}y_3 + \ldots\ldots + \alpha_{2n}y_n &= \beta_2 \\ 0 &= \beta_3 \\ 0 &= \beta_4 \\ &\vdots \\ 0 &= \beta_m \end{aligned}\right\} \ldots (D)$$

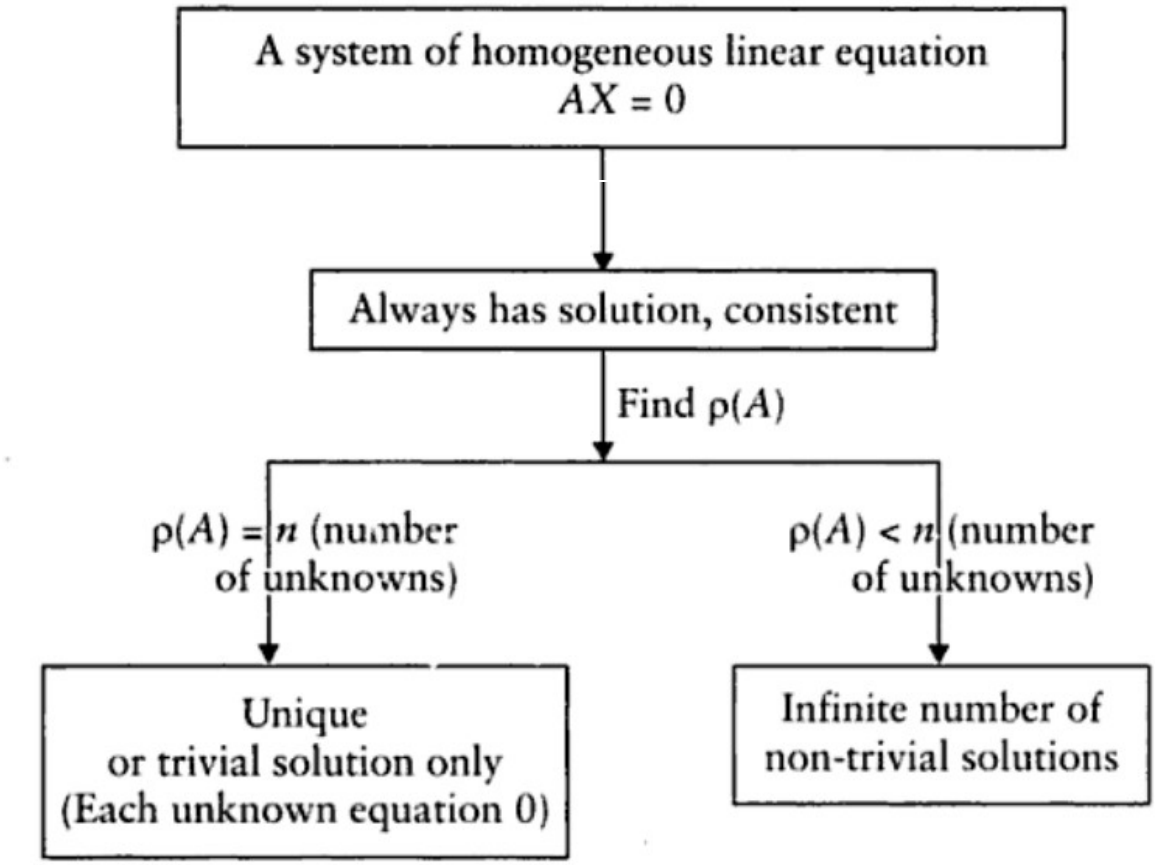
In this case, $\rho(C) = 3$ or 2.

When $\rho(C) = 3$, $\beta_3, \beta_4, \ldots\ldots, \beta_m$ cannot all be zero. Hence, the equations are inconsistent and there will be no solution.

But when $\rho(C) = 2$, $\beta_3, \beta_4, \ldots\ldots, \beta_m$ will be all zero and the equations will be equivalent to two independent equations from which y_1 and y_2 will be expressible in terms of $y_3, y_4, \ldots\ldots, y_n$ which can have arbitrary values.

Similarly, when $r = 3$, then $\rho(C)$ must also be 3 in order that the equations may be consistent and in that case y_1, y_2, y_3 will be expressible in terms of $y_4, y_5, \ldots, y_n$ which are arbitrary.

In general, the necessary and sufficient conditions that the equations (B) may be consistent is that $\rho(C) = \rho(A)$, i.e. if the coefficient matrix A and the augmented matrix C have the same rank and if each rank = r, the equations will be equivalent to r equations from which r unknowns can be



Check the consistency of the following system of homogeneous equations

$$x_1 - 2x_2 + x_3 - x_4 + 1 = 0$$

$$3x_1 - 2x_3 + 3x_4 + 4 = 0$$

$$5x_1 - 4x_2 + x_4 + 3 = 0$$

$$\text{Coefficient matrix } A = \begin{bmatrix} 1 & -2 & 1 & -1 \\ 3 & 0 & -2 & 3 \\ 5 & -4 & 0 & 1 \end{bmatrix}$$

$$\text{Augmented matrix } [A:B] = \left[\begin{array}{cccc:c} 1 & -2 & 1 & -1 & -1 \\ 3 & 0 & -2 & 3 & -4 \\ 5 & -4 & 0 & 1 & -3 \end{array}\right]$$

Operating $R_{21}(-3)$, $R_{31}(-5)$

$$\sim \left[\begin{array}{cccc:c} 1 & -2 & 1 & -1 & -1 \\ 0 & 6 & -5 & 6 & -1 \\ 0 & 6 & -5 & 6 & 2 \end{array}\right]$$

Operating $R_{32}(-1)$

$$\sim \left[\begin{array}{cccc:c} 1 & -2 & 1 & -1 & -1 \\ 0 & 6 & -5 & 6 & -1 \\ 0 & 0 & 0 & 0 & 3 \end{array}\right]$$

which is Echelon Form.

Clearly, $\rho(A) = 2$

$\rho(B) = 3$

$\therefore \quad \rho(A) \neq \rho(B)$

Hence, the given system of equations is inconsistent.

Solve the following equations using matrix method:

$$x_1 + 3x_2 + 2x_3 = 0$$
$$2x_1 - x_2 + 3x_3 = 0$$
$$3x_1 - 5x_2 + 4x_3 = 0$$
$$x_1 + 17x_2 + 4x_3 = 0$$

Solution:

$$\text{Coefficient matrix } A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & -1 & 3 \\ 3 & -5 & 4 \\ 1 & 17 & 4 \end{bmatrix}$$

Operating $R_{21}(-2)$, $R_{31}(-3)$, $R_{41}(-1)$

$$\sim \begin{bmatrix} 1 & 3 & 2 \\ 0 & -7 & -1 \\ 0 & -14 & -2 \\ 0 & 14 & 2 \end{bmatrix}$$

Operating $R_{32}(-2)$, $R_{42}(2)$

$$\sim \begin{bmatrix} 1 & 3 & 2 \\ 0 & -7 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which is Echelon Form.

$\therefore$ $\rho(A) = 2 <$ no. of unknowns

Hence, the given system has infinite number of non-trivial solutions given by

$$x_1 + 3x_2 + 2x_3 = 0 \quad \ldots (6.5)$$

$$-7x_2 - x_3 = 0 \quad \ldots (6.6)$$

Let $x_2 = k$. Then from Eq. (6.6),

$$-7k - x_3 = 0$$

$\Rightarrow x_3 = -7k$

$\therefore$ From Eq. (6.5),

$$x_1 + 3k - 14k = 0$$

$\Rightarrow x_1 = 11k$

Hence, the required solutions are

$$x_1 = 11k$$

$$x_2 = k$$

$$x_3 = -7k$$

where k is arbitrary. Different values of k give different solutions thus making the number of solutions infinite.

Matrix Polynomial

An expression of the form $F(\lambda) = A_0 + A_1\lambda + A_2\lambda^2 + \ldots + A_{m-1}\lambda^{m-1} + A_m\lambda^m$ is called a matrix polynomial of degree m if

(i) $A_0, A_1, A_2, \ldots, A_{m-1}, A_m$ all are square matrices of the same order n (say) and

(ii) $A_m \neq O$.

Such a matrix polynomial is called n-rowed and the symbol λ is called *intermediate*. A_m is called the leading coefficient.

Note: Every square matrix can be expressed as a polynomial of degree zero because if A is a square matrix, then we can write

$$A = \lambda^0 A$$

Characteristic Roots and Vectors

1. Characteristic Matrix

The matrix $A - \lambda I$ is known as the characteristic matrix of A.

2. Characteristic Polynomial

The determinant of the matrix $A - \lambda I$, i.e. $|A - \lambda I|$ is known as the characteristic polynomial of A and is denoted by $\varphi(\lambda)$.

3. Characteristic Equation

The equation $\varphi(\lambda) = 0$, i.e. $|A - \lambda I| = 0$ is known as the characteristic equation (or secular equation) of A.

4. Characteristic Roots

The roots of the characteristic equation of A are called characteristic roots of A. These are also called as latent roots or invariant roots or proper roots or eigen values. The set of characteristic roots of A is called the spectrum of A.

5. Characteristic Vectors

Let $\lambda = \lambda_1$ be any characteristic root of A. Then, we have

$$(A - \lambda_1 I)X = O$$

The non-zero vector X which satisfies the above equation is called characteristic vector of A corresponding to the characteristic root $\lambda = \lambda_1$.

6. Characteristic Space

The collection of all X such that $AX = \lambda X$ is called the characteristic space associated with λ.

Lemma If A is a square matrix of order n, then the adjoint of the characteristic matrix $A - \lambda I$ can be expressed as a matrix polynomial in λ of degree $n - 1$.

Cayley-Hamilton Theorem

Statement. Every square matrix satisfies its own characteristic equation.

OR

If $|A - \lambda I| = (-1)^n [\lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \ldots + a_n]$ be the characteristic polynomial of an $n \times n$ matrix $A = [a_{ij}]$, then the matrix equation

$$X^n + a_1X^{n-1} + \ldots + a_nI = O$$

is satisfied by $X = A$, i.e.

$$A^n + a_1A^{n-1} + \ldots + a_nI = O$$

Verify Cayley-Hamilton theorem for the matrix $A = \begin{bmatrix} 1 & 3 & 7 \\ 4 & 2 & 3 \\ 1 & 2 & 1 \end{bmatrix}$

The characteristic equation of A is

$$\begin{vmatrix} 1-\lambda & 3 & 7 \\ 4 & 2-\lambda & 3 \\ 1 & 2 & 1-\lambda \end{vmatrix} = 0$$

i.e. $(1-\lambda)(\lambda^2-3\lambda-4)-3(4-4\lambda-3)+7(8-2+\lambda)=0$

i.e. $\lambda^3-4\lambda^2-20\lambda-35=0$

Cayley-Hamilton theorem states that

$$A^3-4A^2-20A-35I=0$$

which is to be verified.

Now, $$A^2=\begin{bmatrix}1&3&7\\4&2&3\\1&2&1\end{bmatrix}\begin{bmatrix}1&3&7\\4&2&3\\1&2&1\end{bmatrix}=\begin{bmatrix}20&23&23\\15&22&37\\10&9&14\end{bmatrix}$$

$$A^3=A\cdot A^2=\begin{bmatrix}1&3&7\\4&2&3\\1&2&1\end{bmatrix}\begin{bmatrix}20&23&23\\15&22&37\\10&9&14\end{bmatrix}=\begin{bmatrix}135&152&232\\140&163&208\\60&76&111\end{bmatrix}$$

Substituting these values in (1), we get,

$$\text{L.S.}=\begin{bmatrix}135&152&232\\140&163&208\\60&76&111\end{bmatrix}-\begin{bmatrix}80&92&92\\60&88&148\\40&36&56\end{bmatrix}-\begin{bmatrix}20&60&140\\80&40&60\\20&40&20\end{bmatrix}-\begin{bmatrix}35&0&0\\0&35&0\\0&0&35\end{bmatrix}$$

$$=\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix}$$

$$=\text{R.S.}$$

Thus Cayley-Hamilton theorem is verified. Premultiplying (1) by A^{-1},

$$A^2-4A-20I-35A^{-1}=0$$

$\therefore$ $$A^{-1}=\frac{1}{35}\left(A^2-4A-20I\right)$$

$$=\frac{1}{35}\left(\begin{bmatrix}20&23&23\\15&22&37\\10&9&14\end{bmatrix}-\begin{bmatrix}4&12&28\\16&8&12\\4&8&4\end{bmatrix}-\begin{bmatrix}20&0&0\\0&20&0\\0&0&20\end{bmatrix}\right)$$

$$=\frac{1}{35}\begin{bmatrix}-4&11&-5\\-1&-6&25\\6&1&-10\end{bmatrix}$$

Part-A

1. If A is m x n matrix with entries in the filed F, then show that row –rank(A) = column-rank(A).
2. A linear transformation $T : P_2(t) \to P_2(t)$ is defined as $T(a_0 + a_1 t + a_2 t^2) = (-1 + 2t - 2t^2)a_0 + (-2 + 3t - 2t^2)a_1$ find the eigenvalues and eigenvectors of T.
3. If λ is a characteristic root of A, then show that λ^k is a characteristic root of A^k.
4. Let T be a finite-dimensional vector space V and let α be a scalar, the following statements are equivalent: (i) α is a characteristic value of T (ii) the operator (T – α I) is invertible (iii) det(T – αI)=0

Part-B

1. If V and W are vector spaces over the filed F and T is a linear transformation form V into W, show that rank(T) + nullity(T) = dim V.
2. State and prove Sylvester's law of nullity.
3. Show that two similar matrices have the same characteristic polynomial and hence the same characteristic roots.
4. Show that the eigenvectors associated with distinct eigenvalues of an n-square matrix A are linearly independent.
5. Show that an n^{th} order matrix A with distinct eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_n$ is similar to a diagonal matrix D with these eigenvalues as diagonal elements.

SCHOOL OF SCIENCE AND HUMANITIES
DEPARTMENT OF MATHEMATICS

UNIT – V – Complex Matrices – SMTA5203

Characteristic Roots and Characteristic Vectors of a Square Matrix

The important notions of the characteristic roots and the characteristic vectors of a square matrix over the field of complex numbers will now be introduced.

Def. *Any non-zero vector, X is said to be a characteristic vector of a matrix A, if there exists a number λ such that*

$$AX = \lambda X$$

Also then λ is said to be a characteristic root of the matrix A corresponding to the characteristic vector X and vice-versa.

Characteristic roots (vectors) are also often called *Proper*, *Latent* or *Eigen values* (vectors).

Ex. If A, B are two matrices such that

$$AB = \lambda B$$

show that each of non-zero column of B is a characteristic vector of A corresponding to the characteristic root λ.

Note. It will be found useful to remember that

(*i*) to a characteristic vector of a matrix cannot correspond two different characteristic roots, but

(*ii*) to a characteristic root of a matrix can, and will correspond different characteristic vectors.

Thus, if $AX = \lambda_1 X;\ AX = \lambda_2 X;\ \lambda_2 \neq \lambda_1$

then, $\lambda_1 X = \lambda_2 X \Rightarrow (\lambda_1 - \lambda_2) X = O$

But $X \neq O$ and $(\lambda_1 - \lambda_2) \neq 0$

and therefore $(\lambda_1 - \lambda_2) X \neq 0$

Thus we have a contradiction and as such we see the truth of the statement (*i*).

But if $AX = \lambda X$

then also $A(kX) = \lambda(kX)$

so that kX is also a characteristic vector of A corresponding to the same characteristic root λ. Thus we have the truth of the statement (*ii*).

Determination of Characteristic Roots and Vectors

If, λ, be a characteristic root and, X, a corresponding characteristic vector of a matrix A, then we have

$$AX = \lambda X = \lambda IX$$

$$\Rightarrow \quad (A - \lambda I) X = O$$

Since $X \neq O$, we deduce that the matrix $(A - \lambda I)$ is singular so that its determinant

$$|A - \lambda I| = 0$$

Every characteristic root λ of a matrix A is a root of its characteristic equation

$$|A - xI| = 0.$$

Thus, every root of characteristic equation is a characteristic root of the matrix.

Characteristic Subspace of a Matrix

Let λ be a characteristic root of an n-rowed square matrix A. Consider the matrix equation

$$(A - \lambda I) X = O \qquad \ldots(1)$$

every non-zero solution of which is a characteristic vector of the matrix A corresponding to the characteristic root λ.

If r, be the rank of the matrix $(A - \lambda I)$, then the equation (1) possesses a linearly independent system of $(n - r)$ solutions. Every non-zero linear combination of these solutions, being also a solution of (1), is a characteristic vector corresponding to λ.

The set of all these linear combinations including the zero vector is a subspace of $V_n(C)$ called the *characteristic space* of the matrix A corresponding to the characteristic root λ. Thus *the characteristic space of a matrix A corresponding to a characteristic root λ is just the null space of the matrix* $(A - \lambda I)$.

SOME FUNDAMENTAL THEOREMS

Theorem 1. *Corresponding to a characteristic vector X of a square matrix A, there exists one and only one characteristic root whereas corresponding to a characteristic root there exists more than one characteristic vectors.*

Proof. Let us assume that there exist two distinct characteristic roots λ_1 and λ_2 corresponding to a given characteristic vector X of a square matrix A. Then, we have

$$AX = \lambda_1 X, \; AX = \lambda_2 X$$

On subtracting, we get

$$(\lambda_1 - \lambda_2) X = 0$$

as

$$\lambda_1 - \lambda_2 = 0, \text{ hence } X = 0$$

This is a contradiction that X is a non-zero vector. Hence corresponding to a characteristic vector X there is only one characteristic root of the square matrix A.

Again, if λ be the characteristic root of A, then corresponding characteristic vector X will be given by

$$AX = \lambda X$$

Let k be any non-zero scalar, then

$$k\,(AX) = k\,(\lambda X)$$

i.e.,

$$A\,(kX) = \lambda\,(kX).$$

Thus, kX is also a characteristic vector of A corresponding to the same characteristic root λ.

Theorem 2. *The product of the characteristic roots of a square matrix of order n is equal to the determinant of the matrix.* (*Jiwaji, 1999; Bilaspur, 2000; Garhwal, 2000*)

Proof. Let $A = [a_{ij}]$ be a given square matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the characteristic roots of A. If $\phi(\lambda)$ is the characteristic function, then

$$= (-1)^n [\lambda^n + p_1\lambda^{n-1} + p_2\lambda^{n-2} + \ldots + p_n]$$

$$= (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n)$$

On putting $\lambda = 0$, we have

$$\phi(0) = |A| = \lambda_1 \cdot \lambda_2 \cdot \lambda_3 \ldots \lambda_n = (-1)^n p_n$$

Theorem 3. *For a square matrix A, λ is a characteristic root, if and only if there exists a non-zero vector X such that $AX = \lambda X$.*

or

The equation $AX = \lambda X$ has a non-trivial solution X if λ is a latent root of A.

or

The scalar λ is a characteristic root of the matrix A if and only if the matrix $(A - \lambda I)$ is singular.

Proof. Let λ be a characteristic root of the square matrix A. Then by definition λ must satisfy the characteristic equation of A,

i.e.,

$$|A - \lambda I| = 0$$

This implies that the matrix $A - \lambda I$ must be singular. Hence, if $(A - \lambda I)$ is singular, then λ is a characteristic root of a matrix. Conversely, if $|A - \lambda I| = 0$ then for some non-zero vector X, we have

$$(A - \lambda I) X = 0$$

or

$$AX = \lambda X$$

which shows that λ is a characteristic root of the square matrix A.

Find the Characteristic roots and vectors for each of the following matrices

(i) $\begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix}$ **(ii)** $\begin{bmatrix} 6 & -2 & 2 \\ -2 & 3 & -1 \\ 2 & -1 & 3 \end{bmatrix}$ **(iii)** $\begin{bmatrix} 3 & 10 & 5 \\ -2 & -3 & -4 \\ 3 & 5 & 7 \end{bmatrix}$

(iv) $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$

Sol. (*i*) The characteristic equation is

$$0 = |A - xI| = \begin{bmatrix} 8-x & -6 & 2 \\ -6 & 7-x & -4 \\ 2 & -4 & 3-x \end{bmatrix} = -x^3 + 18x^2 - 45x$$

so that 0, 3, 15 are the three characteristic roots of the matrix.

If x, y, z be the components of a characteristic vector corresponding to the characteristic root, 0, we have

$$O = (A - 0I)X = \begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$\Rightarrow \quad 8x - 6y + 2z = 0, \; -6x + 7y - 4z = 0, \; 2x - 4y + 3z = 0$

These equations determine a single linearly independent solution which we may take as

$$[1 \quad 2 \quad 2]'$$

so that every non-zero multiple of this column vector is a characteristic vector corresponding to the characteristic root 0.

It may similarly be shown by considering the equations

$$(A - 3I)X = O, \; (A - 15I)X = O$$

that the characteristic vectors corresponding to the characteristic roots 3 and 15 are arbitrary non-zero multiples of the vectors

$$\begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

The subspaces of V_3 spanned by these three vectors separately are the three characteristic spaces.

(*ii*) The characteristic equation is

$$0 = |A - xI| = \begin{bmatrix} 6-x & 2 & 2 \\ -2 & 3-x & -1 \\ 2 & -1 & 3-x \end{bmatrix} = -x^3 + 12x^2 - 36x + 32$$

so that 2, 2, 8 are the characteristic roots, only two roots being distinct.

Considering $(A - 8I)X = O$, we may show that we obtain only one linearly independent solution

$$\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

so that every non-zero multiple of the same is a characteristic vector for the characteristic root 8.

For the characteristic root 2, we have

$$O = (A - 2I)X = \begin{bmatrix} 4 & -2 & 2 \\ -2 & 1 & -1 \\ 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$\Rightarrow \quad 4x - 2y + 2z = 0, \; -2x + y - z = 0, \; 2x - y + z = 0$

which are equivalent to a single equation.

Thus we obtain two linearly independent solutions which we may take as $\begin{bmatrix}-1\\0\\2\end{bmatrix}, \begin{bmatrix}1\\2\\0\end{bmatrix}$

The subspace of V_3 spanned by these two vectors is the characteristic space for the root 2.

(*iii*) The characteristic equation of the matrix is

$$-x^3 + 7x^2 - 16x + 12 = 0$$

so that the characteristic roots are 2, 2, 3.

Corresponding to the characteristic root, 3, we find only one linearly independent characteristic vector which may be taken as $\begin{bmatrix}1\\1\\-2\end{bmatrix}$

For the repeated root, 2, we have

$$O = (A - 2I)\, K$$

which gives

$$x + 10y + 5z = 0, -2x - 5y - 4z = 0, 3x + 5y + 5z = 0$$

These equations determine a single linearly independent solution which we take as $\begin{bmatrix}5\\2\\-5\end{bmatrix}$

(*iv*) The characteristic equation is

$$(2 - x)^3 = 0$$

so that, 2, is the only distinct characteristic root.

It may be seen that $(A - 2I)\, X = O$ determines only one linearly independent solution which we may take as $\begin{bmatrix}1\\0\\0\end{bmatrix}$.

2. *If $a + b + c = 0$, find the characteristic roots of the matrix* $A = \begin{bmatrix}a & c & b\\ c & b & a\\ b & a & c\end{bmatrix}$.

(Garhwal, 1996, 2001)

Sol. We have the characteristic equation of A

$$|A - \lambda I| = 0$$

or
$$\begin{vmatrix} a-\lambda & c & b \\ c & b-\lambda & a \\ b & a & c-\lambda \end{vmatrix} = \begin{vmatrix} a+b+c-\lambda & c & b \\ a+b+c-\lambda & b-\lambda & a \\ a+b+c-\lambda & a & c-\lambda \end{vmatrix}$$

On replacing C_1 by $C_1 + C_2 + C_3$,

$$= \begin{vmatrix} -\lambda & c & b \\ -\lambda & b-\lambda & a \\ -\lambda & a & c-\lambda \end{vmatrix} \qquad [\because\ a+b+c = 0]$$

$$= \begin{vmatrix} -\lambda & c & b \\ 0 & b-\lambda-c & c-b \\ 0 & a-c & c-\lambda-b \end{vmatrix}$$

On operating $R_2 - R_1$ and $R_3 - R_1$

$$= \lambda[(a^2+b^2+c^2-ab-bc-ca)-\lambda^2]$$

But $a + b + c = 0$, *i.e.*, $(a+b+c)^2 = 0$

or $a^2 + b^2 + c^2 + 2ab + 2bc + 2ca = 0$

or $-(ab+bc+ca) = \frac{1}{2}(a^2+b^2+c^2)$

$\therefore$ Characteristic equation becomes

$$\lambda\left[a^2+b^2+c^2+\frac{1}{2}(a^2+b^2+c^2)-\lambda^2\right] = 0$$

or $$\lambda\left[\frac{3}{2}(a^2+b^2+c^2)-\lambda^2\right] = 0$$

which gives $\lambda = 0$ or $\lambda = \pm\left[\frac{3}{2}(a^2+b^2+c^2)\right]^{1/2}$

3. *Find the latent roots and latent vectors of the matrix* $A = \begin{bmatrix} a & h & g \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$.

(*Avadh, 2003*; *Kanpur, 2001*)

Sol. The characteristic equation of the matrix A is given by

$$|A - \lambda I| = 0$$

or $$\begin{vmatrix} a-\lambda & h & g \\ 0 & b-\lambda & 0 \\ 0 & 0 & c-\lambda \end{vmatrix} = 0$$

or $$(a-\lambda)(b-\lambda)(c-\lambda) = 0$$

i.e., $$\lambda = a, b, c.$$

Hence the latent root of the matrix corresponding to $\lambda = a$ will be given by

$$\begin{bmatrix} a-a & h & g \\ 0 & b-a & 0 \\ 0 & 0 & c-a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

This gives
$$hx_2 + gx_3 = 0$$
$$(b-a)\,x_2 = 0$$
$$(c-a)\,x_3 = 0$$

On solving these equations, we get

$$x_2 = 0,\ x_3 = 0,\ x_1 = k_1 \text{ (say)},\ k_1 \neq 0.$$

Hence the latent vector corresponding to $\lambda_1 = a$ will be

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} k_1 \\ 0 \\ 0 \end{bmatrix}.$$

Similarly, latent vector corresponding to $\lambda_2 = b$ is given by

$$\begin{bmatrix} a-b & h & g \\ 0 & b-b & 0 \\ 0 & 0 & c-b \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This gives
$$(a-b)\,x_1 + hx_2 + gx_3 = 0$$
and
$$(c-b)\,x_3 = 0$$

On solving these equations, we get

$$x_3 = 0 \text{ and } \frac{x_1}{h} = \frac{x_2}{(b-a)} = k_2 \text{ (say)}$$

Hence, the latent vector corresponding to the root $\lambda = b$ is given by

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} hk_2 \\ (b-a)\,k_2 \\ 0 \end{bmatrix}$$

Now, if we take $\lambda = c$, then latent vector is given by

$$\begin{bmatrix} a-c & h & g \\ 0 & b-c & 0 \\ 0 & 0 & c-c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This implies that

$$(a-c)\,x_1 + hx_2 + gx_3 = 0$$
$$(b-c)\,x_2 = 0$$

On solving these equations, we get

$$x_2 = 0 \text{ and } \frac{x_1}{g} = \frac{x_3}{(c-a)} = k_3 \text{ (say)}$$

Hence latent vector corresponding to the latent root $\lambda_3 = c$ is given by

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} gk_3 \\ 0 \\ (c-a)\,k_3 \end{bmatrix}.$$

4. *Show that a characteristic vector, X, corresponding to the characteristic root, λ, of a matrix A is also a characteristic vector of every matrix f (A); f (x) being any scalar polynomial, and the corresponding root for f (A) is f (λ). In general, show that if*

$$g(x) = f_1(x)/f_2(x);\ |f_2(A)| \neq 0$$

then g (λ) is a characteristic root of

$$g(A) = f_1(A)\ \{f_2(A)\}^{-1}.$$

Sol. If $AX = \lambda X$

then $A^2X = A(AX) = A(\lambda X) = \lambda(AX) = \lambda\lambda X = \lambda^2 X$

Repeating this process k times, we obtain

$$A^k X = \lambda^k X$$

$$\Rightarrow \quad f(A)X = (a_0 I + a_1 A + a_2 A^2 + \ldots\ldots + a_m A^m)X = a_0 X + a_1 \lambda X + \ldots\ldots + a_m \lambda^m X$$

$$= (a_0 + a_1\lambda + \ldots\ldots + a_m\lambda^m)X = f(\lambda)X$$

so that X is a characteristic vector of the matrix $f(A)$ and $f(\lambda)$ is the corresponding characteristic root.

Since $|f_2(A)| \neq 0$, the matrix $f_2(A)$ is non-singular and as such no characteristic root of $f_2(A)$ is zero. In particular,

$$f_2(\lambda) \neq 0$$

for $f_2(\lambda)$ is a characteristic root of $f_2(A)$. Now

$$f_1(A)X = f_1(\lambda)X \qquad \ldots(i)$$

$$f_2(A)X = f_2(\lambda)X \qquad \ldots(ii)$$

From *(ii)*, $\{f_2(\lambda)\}^{-1}X = \{f(A)\}^{-1}X$

$$g(A)X = f_1(A)[\{f_2(A)\}^{-1}X] \qquad \ldots(iii)$$

$$= f_1(A)[\{f_2(\lambda)\}^{-1}X] = \{f_2(\lambda)\}^{-1} f_1(A)X$$

$$= \{f_2(\lambda)\}^{-1} f_1(\lambda)X = g(\lambda)X$$

Thus X is also a characteristic vector of $g(A)$ with corresponding root $g(\lambda)$.

5. *Show that the two matrices A, $P^{-1}AP$ have the same characteristic roots.*

(M.D.U. Rohtak, 2000)

Sol. We write

$$P^{-1}AP = B$$

$$\therefore \quad B - xI = P^{-1}AP - xI = P^{-1}AP - P^{-1}xIP = P^{-1}(A - xI)P$$

$$\Rightarrow \quad |B - xI| = |P^{-1}||A - xI||P| = |A - xI||P^{-1}||P|$$

$$= |A - xI||P^{-1}P| = |A - xI||I| = |A - xI|$$

Thus the two matrices A and B have the same characteristic determinants and hence the same characteristic equations and the same characteristic roots.

The same thing may also be seen in another way.

Now $AX = \lambda X$

$$\Rightarrow \quad P^{-1}AX = \lambda P^{-1}X$$

$$\Rightarrow \quad (P^{-1}AP)(P^{-1}X) = \lambda(P^{-1}X)$$

so that λ is also a characteristic root of $P^{-1}AP$ and $P^{-1}X$ is a corresponding characteristic vector.

6. *If A and B are two square matrices, then the matrices AB and BA have the same characteristic roots.*

$$AB = B^{-1}(BA)B \text{ or } AB = A(BA)A^{-1}$$

so that by the preceding result AB, BA have the same characteristic roots.

We now give a proof which holds in the general case.

If, r, be the rank of A, then there exist two non-singular matrices P and Q such that,

$$PAQ = \text{Diag. } [I_r, O]$$

We have $PABP^{-1} = (PAQ)(Q^{-1}BP^{-1})$

Let $$Q^{-1}BP^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

where C_{11} is $r \times r$.

$$\therefore \quad PABP^{-1} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ O & O \end{bmatrix}$$

Again $$Q^{-1}BAQ = (Q^{-1}BP^{-1})(PAQ)$$

$$= \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = \begin{bmatrix} C_{11} & O \\ C_{21} & O \end{bmatrix}$$

Thus the characteristic roots of AB and BA are the same as those of C_{11} along with $(n - r)$ roots each equal to 0.

7. *Show that the characteristic roots of A^{Θ} are the conjugates of the characteristic roots of A.*

Sol. We have $|A^{\Theta} - \bar{\lambda}I| - |(A - \lambda I)^{\Theta}| = |\overline{A - \lambda I}|$

$\therefore \quad |A^{\Theta} - \bar{\lambda}I| = 0$ iff $|\overline{A - \lambda I}| = 0$

$\Rightarrow \quad |A^{\Theta} - \bar{\lambda}I| = 0$ iff $|A - \lambda I| = 0$

or $\bar{\lambda}$ is an eigen value of A^{Θ} if, and only if, λ is an eigen value of A.

8. *Show that the characteristic roots of a triangular matrix are just the diagonal elements of the matrix.* (*Jabalpur, 2001*)

Sol. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix}$$

$$\therefore \quad |A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ 0 & a_{22} - \lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) \dots (a_{nn} - \lambda)$$

Hence the characteristic roots of A are $a_{11}, a_{22}, \dots, a_{nn}$ which are just the diagonal elements of A.

9. *If A is non-singular, prove that the eigen values of A^{-1} are the reciprocals of the eigen values of A.* (*Jabalpur, 1998; Sagar, 1999; Vikram, 1998, 2000; Indore, 1999; Rewa, 2000*)

Sol. Let A be an eigen value of A and X be a corresponding eigen vector. Then

$$AX = \lambda X \Rightarrow X = A^{-1}(\lambda X) = \lambda (A^{-1}\lambda)$$

$$\Rightarrow \frac{1}{\lambda} X = A^{-1}X \qquad [\because A \text{ is non-singular} \Rightarrow \lambda \neq 0]$$

$$\Rightarrow A^{-1}X = \frac{1}{\lambda} X$$

$$\Rightarrow \frac{1}{\lambda} \text{ is an eigen value of } A^{-1} \text{ and X is a corresponding eigen vector.}$$

Conversely, suppose that k is an eigen value of A^{-1}. Since A is non-singular $\Rightarrow A^{-1}$ is non-singular and $(A^{-1})^{-1} = A$, therefore it follows from the first part that $1/k$ is an eigen value of A. Thus each eigen value of A^{-1} is equal to the reciprocal of some eigen value of A.

10. *If α is a characteristic root of a non-singular matrix A, then prove that $\frac{|A|}{\alpha}$ is a characteristic root of adj. A.* *(Jiwaji, 2000; Rewa, 1999; Vikram, 1999, 2001)*

Sol. Since α is a characteristic root of a non-singular matrix, therefore $\alpha \neq 0$. Also α is a characteristic root of A implies that there exists a non-zero vector X such that

$$AX = \alpha X$$

$$\Rightarrow \quad (\text{adj. A})(AX) = (\text{adj. A})(\alpha X)$$

$$\Rightarrow \quad [(\text{adj. A})\,A]\,X = \alpha\,(\text{adj. A})\,X$$

$$\Rightarrow \quad |A|IX = \alpha\,(\text{adj. A})\,X \qquad [\because (\text{adj. A})\,A = |A|I]$$

$$\Rightarrow \quad |A|X = \alpha\,(\text{adj. A})\,X$$

$$\Rightarrow \quad \frac{|A|}{\alpha} X = (\text{adj. A})\,X$$

$$\Rightarrow \quad (\text{adj. A})\,X = \frac{|A|}{\alpha} X$$

Since X is a non-zero vector, therefore $\frac{|A|}{\alpha}$ is a characteristic root of the matrix adj. A.

11. *Show that if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are n eigen values of a square matrix A of order n then the eigen values of the matrix A^2 be $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$.* *(Jabalpur, 1999; Rewa, 1994)*

Sol. We know that if λ be the eigen value of a square matrix A, then there exists a non-zero vector X such that

$$AX = \lambda X$$

$$\Rightarrow \quad A(AX) = A(\lambda X)$$

$$\Rightarrow \quad A^2X = \lambda(AX) = \lambda(\lambda X) = \lambda^2 X$$

i.e.,

$$A^2X = \lambda^2 X$$

$\Rightarrow$ Eigen value of A^2 is λ^2.

Similarly if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigen values of A then $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$ are eigen values of A^2.

12. *Show that the characteristic roots of an idempotent matrix are either zero or unity.*

Sol. Since A is an idempotent matrix, hence

$$A^2 = A$$

Let X be a latent vector of the matrix A corresponding to the latent root λ so that,

$$AX = \lambda X \qquad \ldots(i)$$

or

$$(A - \lambda I)\,X = 0$$

such that $X \neq 0$

On pre-multiplying (i) by A, we get

$$A(AX) = A(\lambda X) = \lambda(AX)$$

i.e., $(AA)X = \lambda(\lambda X)$

or $AX = \lambda^2 X$

or $\lambda X = \lambda^2 X$

or $(\lambda^2 - \lambda)X = 0$

or $\lambda^2 - \lambda = 0$

or $\lambda(\lambda - 1) = 0$

or $\lambda = 0, \lambda = 1.$

NATURE OF THE CHARACTERISTIC ROOTS OF SOME SPECIAL TYPES OF MATRICES

Theorem 1. The characteristic roots of a Hermitian matrices are all real.

Let, λ, be a characteristic root of a Hermitian matrix A so that there exists a vector $X \neq O$, such that

$$AX = \lambda X$$

Pre-multiplying with X^{Θ}, we obtain

$$X^{\Theta}AX = X^{\Theta}\lambda X = \lambda X^{\Theta} = AX^{\Theta}IX$$

Being the values of Hermitian forms, $X^{\Theta}AX$ and $X^{\Theta}IX$ are both real. (§ 9.1, p. ???). Also $X^{\Theta}X \neq 0$, for $X \neq O$. Thus

$$\lambda = X^{\Theta}AX/X^{\Theta}IX$$

is real.

Cor. I. *The characteristic roots of a real symmetric matrix are all real, for every such matrix is Hermitian.* (*Ravishankar, 1997*)

An independent proof can, of course, be given exactly along the lines of the proof above.

Cor. II. *A characteristic root of a skew-Hermitian matrix is either zero or a pure imaginary number.* (*M.D.U. Rohtak, 1998, 2000*)

If A be a skew-Hermitian matrix and

$$AX = \lambda X$$

then $$(iA)X = (i\lambda)X$$

But iA is Hermitian and, as such, $i\lambda$, a characteristic root of iA, is real. Thus either $\lambda = 0$ or is a pure imaginary number.

Cor. III. *A characteristic root of real skew matrix is either zero or a pure imaginary number, for every such matrix is skew-Hermitian.*

Also the imaginary characteristic roots occur in conjugate pairs, for the coefficients of the characteristic equation of a real matrix are all real.

Theorem 2. *The modulus of such characteristic root of a unitary matrix is unity.* (*Bilaspur, 1998; Jabalpur, 1996, 98; Vikram, 1999; Sagar, 2001; M.D.U. Rohtak, 1996, 98, 2000*)

If A is a unitary matrix and

$$AX = \lambda X$$

hen on taking conjugate transpose of each side, we have

$$X^{\Theta}A^{\Theta} = \bar{\lambda} X^{\Theta}$$

These give $X^{\Theta}A^{\Theta}AX = \bar{\lambda}\lambda X^{\Theta}X$

As A is unitary,

e., $A^{\Theta}A = I$

ve obtain

$$X^{\Theta}A = \lambda\bar{\lambda} X^{\Theta}X$$

$\Rightarrow$ $(1 - \lambda\bar{\lambda}) X^{\Theta}X = 0$

Now $X \neq O \Rightarrow X^{\Theta}X \neq 0$

Hence $1 - \lambda\bar{\lambda} = 0 \Rightarrow \lambda\bar{\lambda} = 1$

So that the modulus of λ is unity.

Cor. The modulus of each characteristic root of an orthogonal matrix is unity, for every such matrix is unitary.

11.3.1. Algebraic and Geometric Multiplicity of a Characteristic Root

If, λ, be a *t*-ple root of the characteristic equation

$$| A - xI | = 0$$

then, *t*, is called the *Algebraic multiplicity* of λ and the dimension, *s*, of the characteristic space of A corresponding to λ, *i.e.*, the number of linearly independent solutions of

$$(A - \lambda I) X = O$$

is called the *Geometric multiplicity* of A.

Ex. 1. The characteristic roots of a diagonal matrix are the same as its diagonal elements.

Ex. 2. Zero and unity are the characteristic roots of algebraic multiplicity n of O_n and I_n respectively.

Ex. 3. Point out the algebraic and geometric multiplicities of each characteristic roots of each of the matrices in Q. 1, and observe that for every characteristic root :

7 Eigenvalues and Eigenvectors

7.1 Introduction

The simplest of matrices are the diagonal ones. Thus a linear map will be also easy to handle if its associated matrix is a diagonal matrix. Then again we have seen that the matrix associated depends upon the choice of the bases to some extent. This naturally leads us to the problem of investigating the existence and construction of a suitable basis with respect to which the matrix associated to a given linear transformation is diagonal.

Definition 7.1 *A $n \times n$ matrix A is called* diagonalizable *if there exists an invertible $n \times n$ matrix M such that $M^{-1}AM$ is a diagonal matrix. A linear map $f : V \longrightarrow V$ is called diagonalizable if the matrix associated to f with respect to some basis is diagonal.*

Remark 7.1
(i) Clearly, f is diagonalizable iff the matrix associated to f with respect to some basis (any basis) is diagonalizable.
(ii) Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis. The matrix M_f of a linear transformation f w.r.t. this basis is diagonal iff $f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i,\ 1 \leq i \leq n$ for some scalars λ_i. Naturally a subquestion here is: does there exist such a basis for a given linear transformation?

Definition 7.2 *Given a linear map $f : V \longrightarrow V$ we say $\mathbf{v} \in V$ is an* eigenvector *for f if $\mathbf{v} \neq 0$ and $f(\mathbf{v}) = \lambda \mathbf{v}$ for some $\lambda \in \mathbb{K}$. In that case λ is called as* eigenvalue *of f. For a square matrix A we say λ is an eigenvalue if there exists a non zero column vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda \mathbf{v}$. Of course $\mathbf{v}$ is then called the eigenvector of A corresponding to λ.*

Remark 7.2
(i) It is easy to see that eigenvalues and eigenvectors of a linear transformation are same as those of the associated matrix.
(ii) Even if a linear map is not diagonalizable, the existence of eigenvectors and eigenvalues itself throws some light on the nature of the linear map. Thus the study of eigenvalues becomes extremely important. They arise naturally in the study of differential equations. Here we shall use them to address the problem of diagonalization and then see some geometric applications of diagonalization itself.

7.2 Characteristic Polynomial

Proposition 7.1
(1) Eigenvalues of a square matrix A are solutions of the equation

$$\chi_A(\lambda) = \det(A - \lambda I) = 0.$$

(2)The null space of $A - \lambda I$ is equal to the eigenspace

$$E_A(\lambda) := \{\mathbf{v} \ : \ A\mathbf{v} = \lambda \mathbf{v}\} = \mathcal{N}(A - \lambda I).$$

Proof: (1) If $\mathbf{v}$ is an eigenvector of A then $\mathbf{v} \neq 0$ and $A\mathbf{v} = \lambda \mathbf{v}$ for some scalar λ. Hence $(A - \lambda I)\mathbf{v} = 0$. Thus the nullity of $A - \lambda I$ is positive. Hence $\text{rank}(A - \lambda I)$ is less than n. Hence $\det(A - \lambda I) = 0$.
(2) $E_A(\lambda) = \{\mathbf{v} \in V \ : \ A\mathbf{v} = \lambda \mathbf{v}\} = \{\mathbf{v} \in V \ : \ (A - \lambda I)\mathbf{v} = 0\} = \mathcal{N}(A - \lambda I)$. ♠

Definition 7.3 *For any square matrix A, the polynomial $\chi_A(\lambda) = \det(A - \lambda I)$ in λ is called the* **characteristic polynomial** *of A.*

Example 7.1

(1) $A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$. *To find the eigenvalues of A, we solve the equation*

$$\det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 2 \\ 0 & 3-\lambda \end{bmatrix} = (1-\lambda)(3-\lambda) = 0.$$

Hence the eigenvalues of A are 1 and 3. Let us calculate the eigenspaces $E(1)$ and $E(3)$. By definition

$$E(1) = \{\mathbf{v} \mid (A - I)\mathbf{v} = 0\} \text{ and } E(3) = \{\mathbf{v} \mid (A - 3I)\mathbf{v} = 0\}.$$

$A - I = \begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix}$. *Hence* $(x, y)^t \in E(1)$ *iff* $\begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2y \\ 2y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. *Hence* $E(1) = L\{(1, 0)\}$.

$A - 3I = \begin{bmatrix} 1-3 & 2 \\ 0 & 3-3 \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix}$. *Suppose* $\begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Then $\begin{bmatrix} -2x + 2y \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. *This is possible iff $x = y$. Thus* $E(3) = L(\{(1, 1)\})$.

(2) Let $A = \begin{bmatrix} 3 & 0 & 0 \\ -2 & 4 & 2 \\ -2 & 1 & 5 \end{bmatrix}$. *Then* $\det(A - \lambda I) = (3-\lambda)^2(6-\lambda)$.

Hence eigenvalues of A are 3 and 6. The eigenvalue $\lambda = 3$ is a double root of the characteristic polynomial of A. We say that $\lambda = 3$ has **algebraic multiplicity** *2. Let us find the eigenspaces $E(3)$ and $E(6)$.*

$\underline{\lambda = 3}$: $A - 3I = \begin{bmatrix} 0 & 0 & 0 \\ -2 & 1 & 2 \\ -2 & 1 & 2 \end{bmatrix}$. *Hence* $rank(A - 3I) = 1$. *Thus* $nullity(A - 3I) = 2$. *By solving the system* $(A - 3I)\mathbf{v} = 0$, *we find that*

$$\mathcal{N}(A - 3I) = E_A(3) = L(\{(1, 0, 1), (1, 2, 0)\}).$$

The dimension of $E_A(\lambda)$ is called the **geometric multiplicity** *of λ. Hence geometric multiplicity of $\lambda = 3$ is 2.*

$\underline{\lambda = 6}$: $A - 6I = \begin{bmatrix} -3 & 0 & 0 \\ -2 & -2 & 2 \\ -2 & 1 & -1 \end{bmatrix}$. *Hence* $rank(A - 6I) = 2$. *Thus* $\dim E_A(6) = 1$. *(It can be shown that $\{(0, 1, 1)\}$ is a basis of $E_A(6)$.) Thus both the algebraic and geometric multiplicities of the eigenvalue 6 are equal to 1.*

(3) $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. *Then* $\det(A - \lambda I) = (1-\lambda)^2$. *Thus $\lambda = 1$ has algebraic multiplicity*

o

$A - I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. *Hence nullity* $(A - I) = 1$ *and* $E_A(1) = L\{e_1\}$. *In this case the geometric multiplicity is less than the algebraic multiplicity of the eigenvalue* 1.

Remark 7.3
(i) Observe that $\chi_A(\lambda) = \chi_{M^{-1}AM}(\lambda)$. *Thus the characteristic polynomial is an invariant of similarity. Thus the characteristic polynomial of any linear map* $f : V \longrightarrow V$ *is also defined (where* V *is finite dimensional) by choosing some basis for* V, *and then taking the characteristic polynomial of the associated matrix* $\mathcal{M}(f)$ *of* f. *This definition does not depend upon the choice of the basis.*
(ii) If we expand $\det(A - \lambda I)$ *we see that there is a term*

$$(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda).$$

This is the only term which contributes to λ^n *and* λ^{n-1}. *It follows that the degree of the characteristic polynomial is exactly equal to* n, *the size of the matrix; moreover, the coefficient of the top degree term is equal to* $(-1)^n$. *Thus in general, it has* n *complex roots, some of which may be repeated, some of them real, and so on. All these patterns are going to influence the geometry of the linear map.*
(iii) If A *is a real matrix then of course* $\chi_A(\lambda)$ *is a real polynomial. That however, does not allow us to conclude that it has real roots. So while discussing eigenvalues we should consider even a real matrix as a complex matrix and keep in mind the associated linear map* $\mathbb{C}^n \longrightarrow \mathbb{C}^n$. *The problem of existence of real eigenvalues and real eigenvectors will be discussed soon.*
(iv) Next, the above observation also shows that the coefficient of λ^{n-1} *is equal to*

$$(-1)^{n-1}(a_{11} + \cdots + a_{nn}) = (-1)^{n-1} tr\, A.$$

Lemma 7.1 Suppose A is a real matrix with a real eigenvalue λ. Then there exists a real column vector $\mathbf{v} \neq 0$ such that $A\mathbf{v} = \lambda\mathbf{v}$.

Proof: Start with $A\mathbf{w} = \lambda\mathbf{w}$ where $\mathbf{w}$ is a non zero column vector with complex entries. Write $\mathbf{w} = \mathbf{v} + \imath\mathbf{v}'$ where both $\mathbf{v}, \mathbf{v}'$ are real vectors. We then have

$$A\mathbf{v} + \imath A\mathbf{v}' = \lambda(\mathbf{v} + \imath\mathbf{v}')$$

Compare the real and imaginary parts. Since $\mathbf{w} \neq 0$, at least one of the two $\mathbf{v}, \mathbf{v}'$ must be a non zero vector and we are done. ♠

Proposition 7.2 *Let* A *be an* $n \times n$ *matrix with eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$. *Then*
(i) $tr\,(A) = \lambda_1 + \lambda_2 + \ldots + \lambda_n$.
(ii) $\det A = \lambda_1\lambda_2 \ldots \lambda_n$.

Proof: The characteristic polynomial of A is

$$\det(A - \lambda I) = \det \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix}$$

7.4 Eigenvalues of Special Matrices

In this section we discuss eigenvalues of special matrices. We will work in the n-dimensional complex vector space $\mathbb{C}^n$. If $\mathbf{u} = (u_1, u_2, \ldots, u_n)^t$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)^t \in \mathbb{C}^n$, we have defined their inner product in $\mathbb{C}^n$ by

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^*\mathbf{v} = \overline{u_1}v_1 + \overline{u_2}v_2 + \cdots + \overline{u_n}v_n.$$

The length of $\mathbf{u}$ is given by $\|\mathbf{u}\| = \sqrt{|u_1|^2 + \cdots + |u_n|^2}$.

Definition 7.6 *Let A be a square matrix with complex entries. A is called*
(i) **Hermitian** *if $A = A^*$;*
(ii) **Skew Hermitian** *if $A = -A^*$.*

Lemma 7.2 *A is Hermitian iff for all column vectors $\mathbf{v}, \mathbf{w}$ we have*

$$(A\mathbf{v})^*\mathbf{w} = \mathbf{v}^*A\mathbf{w}; \quad \textit{i.e.,} \ (\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, A\mathbf{w}\rangle) \tag{52}$$

Proof: If A is Hermitian then $(A\mathbf{v})^*\mathbf{w} = \mathbf{v}^*A^*\mathbf{w} = \mathbf{v}^*A\mathbf{w}$. To see the converse, take $\mathbf{v}, \mathbf{w}$ to be standard basic column vectors. ♠

Remark 7.5
(i) If A is real then $A = A^*$ means $A = A^t$. Hence real symmetric matrices are Hermitian. Likewise a real skew Hermitian matrix is skew symmetric.
(ii) A is Hermitian iff $\imath A$ is skew Hermitian.

Proposition 7.7 *Let A be an $n \times n$ Hermitian matrix. Then :*

1. *For any $\mathbf{u} \in \mathbb{C}^n$, $\mathbf{u}^*A\mathbf{u}$ is a real number.*
2. *All eigenvalues of A are real.*
3. *Eigenvectors of a Hermitian matrix corresponding to distinct eigenvalues are mutually orthogonal.*

Proof: (1) Since $\mathbf{u}^*A\mathbf{u}$ is a complex number, to prove it is real, we prove that $(\mathbf{u}^*A\mathbf{u})^* = \mathbf{u}^*A\mathbf{u}$. But $(\mathbf{u}^*A\mathbf{u})^* = \mathbf{u}^*A^*(\mathbf{u}^*)^* = \mathbf{u}^*A\mathbf{u}$. Hence $\mathbf{u}^*A\mathbf{u}$ is real for all $\mathbf{u} \in \mathbb{C}^n$.

(2) Suppose λ is an eigenvalue of A and $\mathbf{u}$ is an eigenvector for λ. Then

$$\mathbf{u}^*A\mathbf{u} = \mathbf{u}^*(\lambda\mathbf{u}) = \lambda(\mathbf{u}^*\mathbf{u}) = \lambda\|\mathbf{u}\|^2.$$

Since $\mathbf{u}^*A\mathbf{u}$ is real and $\|\mathbf{u}\|$ is a nonzero real number, it follows that λ is real.

(3) Let λ and μ be two distinct eigenvalues of A and $\mathbf{u}$ and $\mathbf{v}$ be corresponding eigenvectors. Then $A\mathbf{u} = \lambda\mathbf{u}$ and $A\mathbf{v} = \mu\mathbf{v}$. Hence

$$\lambda\mathbf{u}^*\mathbf{v} = (\lambda\mathbf{u})^*\mathbf{v} = (A\mathbf{u})^*\mathbf{v} = \mathbf{u}^*(A\mathbf{v}) = \mathbf{u}^*\mu\mathbf{v} = \mu(\mathbf{u}^*\mathbf{v}).$$

Hence $(\lambda - \mu)\mathbf{u}^*\mathbf{v} = 0$. Since $\lambda \neq \mu$, $\mathbf{u}^*\mathbf{v} = 0$. ♠

Corollary 7.1 *Let A be an $n \times n$ skew Hermitian matrix. Then :*

1. *For any $\mathbf{u} \in \mathbb{C}^n, \mathbf{u}^* A\mathbf{u}$ is either zero or a purely imaginary number.*
2. *Each eigenvalue of A is either zero or a purely imaginary number.*
3. *Eigenvectors of A corresponding to distinct eigenvalues are mutually orthogonal.*

Proof: All this follow straight way from the corresponding statement about Hermitian matrix, once we note that A is skew Hermitian implies $\imath A$ is Hermitian and the fact that a complex number c is real iff $\imath c$ is either zero or purely imaginary.

Definition 7.7 *Let A be a square matrix over $\mathbb{C}$. A is called*
(i) **unitary** *if $A^*A = I$;*
(ii) **orthogonal** *if A is real and unitary.*

Thus a real matrix A is orthogonal iff $A^T = A^{-1}$. Also observe that A is unitary iff A^T is unitary iff $\overline{A}$ is unitary.

Example 7.2 *The matrices*

$$U = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad \textit{and} \quad V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$$

are orthogonal and unitary respectively.

Proposition 7.8 *Let A be a square matrix. Then the following conditions are equivalent.*
(i) U is unitary.
(ii) The rows of U form an orthonormal set of vectors.
(iii) The columns of U form an orthonormal set of vectors.
(iv) U preserves the inner product, i.e., for all vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, we have $\langle U\mathbf{x}, U\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$.

Proof: Write the matrix U column-wise :

$$U = [\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_n] \quad \text{so that} \quad U^* = \begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \vdots \\ \mathbf{u}_n^* \end{bmatrix}.$$

Hence

$$\begin{aligned} U^*U &= \begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \vdots \\ \mathbf{u}_n^* \end{bmatrix} [\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_n] \\ &= \begin{bmatrix} \mathbf{u}_1^*\mathbf{u}_1 & \mathbf{u}_1^*\mathbf{u}_2 & \cdots & \mathbf{u}_1^*\mathbf{u}_n \\ \mathbf{u}_2^*\mathbf{u}_1 & \mathbf{u}_2^*\mathbf{u}_2 & \cdots & \mathbf{u}_2^*\mathbf{u}_n \\ \vdots & & \cdots & \\ \mathbf{u}_n^*\mathbf{u}_1 & \mathbf{u}_n^*\mathbf{u}_2 & \cdots & \mathbf{u}_n^*\mathbf{u}_n \end{bmatrix}. \end{aligned}$$

Thus $U^*U = I$ iff $\mathbf{u}_i^*\mathbf{u}_j = 0$ for $i \neq j$ and $\mathbf{u}_i^*\mathbf{u}_i = 1$ for $i = 1, 2, \ldots, n$ iff the column vectors of U form an orthonormal set. This proves $(i) \iff (ii)$. Since $U^*U = I$ implies $UU^* = I$, the proof of $(i) \iff (iii)$ follows.
To prove $(i) \iff (iv)$ let U be unitary. Then $U^*U = Id$ and hence $\langle U\mathbf{x}, U\mathbf{y}\rangle = \langle \mathbf{x}, U^*U\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$. Conversely, iff U preserves inner product take $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$ to get

$$\mathbf{e}_i^*(U^*U)\mathbf{e}_j = \mathbf{e}_i^*\mathbf{e}_j = \delta_{ij}$$

where δ_{ij} are Kronecker symbols ($\delta_{ij} = 1$ if $i = j$; $= 0$ otherwise.) This means the $(i,j)^{th}$ entry of U^*U is δ_{ij}. Hence $U^*U = I_n$. ♠

Remark 7.6 *Observe that the above theorem is valid for an orthogonal matrix also by merely applying it for a real matrix.*

Corollary 7.2 *Let U be a unitary matrix. Then :*
(1) For all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, $\langle U\mathbf{x}, U\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$. Hence $\|U\mathbf{x}\| = \|\mathbf{x}\|$.
(2) If λ is an eigenvalue of U then $|\lambda| = 1$.
(3) Eigenvectors corresponding to different eigenvalues are orthogonal.

Proof: (1) We have, $\|U\mathbf{x}\|^2 = \langle U\mathbf{x}, U\mathbf{x}\rangle = \langle \mathbf{x}, \mathbf{x}\rangle = \|\mathbf{x}\|^2$.
(2) If λ is an eigenvalue of U with eigenvector $\mathbf{x}$ then $U\mathbf{x} = \lambda\mathbf{x}$. Hence $\|\mathbf{x}\| = \|U\mathbf{x}\| = |\lambda|\ \|\mathbf{x}\|$. Hence $|\lambda| = 1$.
(3) Let $U\mathbf{x} = \lambda\mathbf{x}$ and $U\mathbf{y} = \mu\mathbf{y}$ where $\mathbf{x}, \mathbf{y}$ are eigenvectors with distinct eigenvalues λ and μ respectively. Then

$$\langle \mathbf{x}, \mathbf{y}\rangle = \langle U\mathbf{x}, U\mathbf{y}\rangle = \langle \lambda\mathbf{x}, \mu\mathbf{y}\rangle = \overline{\lambda}\mu\langle \mathbf{x}, \mathbf{y}\rangle.$$

Hence $\overline{\lambda}\mu = 1$ or $\langle \mathbf{x}, \mathbf{y}\rangle = 0$. Since $\overline{\lambda}\lambda = 1$, we cannot have $\overline{\lambda}\mu = 1$. Hence $\langle \mathbf{x}, \mathbf{y}\rangle = 0$, i.e., $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. ♠

Example 7.3 $U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ *is an orthogonal matrix. The characteristic polynomial of U is :*

$$D(\lambda) = \det(U - \lambda I) = \det, \begin{bmatrix} \cos\theta - \lambda & -\sin\theta \\ \sin\theta & \cos\theta - \lambda \end{bmatrix} = \lambda^2 - 2\lambda\cos\theta + 1.$$

Roots of $D(\lambda) = 0$ are :

$$\lambda = \frac{2\cos\theta \pm \sqrt{4cos^2\theta - 4}}{2} = \cos\theta \pm \imath\sin\theta = e^{\pm\imath\theta}.$$

Hence $|\lambda| = 1$. Check that eigenvectors are :
for $\underline{\lambda = e^{\imath\theta}}$: $x = \begin{bmatrix} 1 \\ -\imath \end{bmatrix}$ *and for* $\underline{\lambda = e^{-\imath\theta}}$: $y = \begin{bmatrix} 1 \\ \imath \end{bmatrix}$.
Thus $\mathbf{x}^*\mathbf{y} = [1\ \imath]\begin{bmatrix} 1 \\ \imath \end{bmatrix} = 1 + \imath^2 = 0$. *Hence* $\mathbf{x} \perp \mathbf{y}$. *Normalize the eigenvectors* $\mathbf{x}$ *and* $\mathbf{y}$. *Therefore if we take,*

$$C = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -\imath & \imath \end{bmatrix}$$

then $C^{-1}UC = D(e^{\imath\theta}, e^{-\imath\theta})$.

Part-A

1	Show that the eigenvalues of a Hermitian matrix are real and of skew-Hermitian matrix are purely imaginary.
2	Show that if α is a eigenvalue of a unitary matrix U, then (i) α^{-1} is an eigenvalue of U and (ii) Every eigenvalue of U has unit modulus.
3	Show that eigenvectors associated with distinct eigenvectors of a Hermitian matrix are orthogonal.
4	Show that a normal matric is unitarily similar to a diagonal matrix.

Part-B

1	Show that every Hermitian matrix H is unitarily similar to a diagonal matrix whose diagonal elements are the eigenvalues of H.
2	Show that if α is an eigenvalue of multiplicity *m* of a Hermitian matrix H, then the number of linearly independent eigenvectors associated with α is *m*.
3	Show that if α is an eigenvalue of Hermitian matrix H of multiplicity *m*, then there exists *m* orthogonal vectors associated with α.
4	Show that a Hermitian matrix H of order *n* possesses an orthogonal set of *n* eigenvectors.
5	Prove that if A and B are two Hermitian matrices of the same order *n* with A having positive eigenvalues, then there exists an *n* x *n* non-singular matrix P such that $P^*AP = I$, $P^*BP=diag(c_1, c_2,\ldots c_n)$ where c_i's are real.
6	State and prove Cayley-Hamilton theorem.