Study Material on CC-14

Department of Mathematics, P. R. Thakur Govt. College
MTMACOR14T: (Unit 1) (Semester - 6)

Syllabus:

Unit 1 : Polynomial rings over commutative rings, division algorithm and consequences, principal ideal domains, factorization of polynomials, reducibility tests, irreducibility tests, Eisenstein criterion, and unique factorization in $\mathbb{Z}[x]$. Divisibility in integral domains, irreducible, primes, unique factorization domains, Euclidean domains.

Unit 2 : Dual spaces, dual basis, double dual, transpose of a linear transformation and its matrix in the dual basis, annihilators. Eigen spaces of a linear operator, diagonalizability, invariant subspaces and Cayley-Hamilton theorem, the minimal polynomial for a linear operator, canonical forms.

Unit 3 : Inner product spaces and norms, Gram-Schmidt orthogonalisation process, orthogonal complements, Bessel's inequality, the adjoint of a linear operator, Least Squares Approximation, minimal solutions to systems of linear equations, Normal and self-adjoint operators, Orthogonal projections and Spectral theorem.

1 Polynomial Rings

1.1 Definition and Examples

DEFINITION. 1.1 Let R be a ring. A formal expression

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_0, a_1, \ldots, a_n \in R$, is called a *polynomial* of the variable x over the ring R. The elements $a_0, a_1, \ldots, a_n$ are called the *coefficients* of the polynomial $f(x)$. If $a_n \neq 0$ then a_n is called the *leading coefficient* and n is called the *degree* of $f(x)$, written as $\deg f(x) = n$. If R is a ring with unity 1 and $a_n = 1$ then the polynomial $f(x)$ is called a *monic polynomial*.

Two polynomials $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ and $g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$ are said to be equal if $m = n$ and $a_k = b_k$ for all $k = 0, 1, 2, \ldots n$.

For $a \in R$, $f(x) = a$ is called a *constant polynomial* and the degree of a constant polynomial is defined to be zero. The set of all the polynomials over a ring R of a variable x is denoted by $R[x]$.

EXAMPLE. 1.2 1. $f(x) = 2x^3 + 0x^2 + 3x + 5$ is a polynomial over $\mathbb{Z}$

2. $g(x) = \bar{3}x^2 + \bar{1}x + \bar{4}$ is a polynomial over $\mathbb{Z}_6$.

For convenience we shall write $1x^k = x^k$ and omit the term $0x^m$ in the formal expression of a polynomial in $R[x]$. Thus in the above example $f(x)$ is written as $f(x) = 2x^3+3x+5$ and $g(x)$ is written as $g(x) = \bar{3}x^2 + x + \bar{4}$.

DEFINITION. 1.3 Let R be a ring, addition and multiplication in $R[x]$ can be defined as follows: for $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ and $g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$ in $R[x]$ define

$$\begin{aligned} f(x) + g(x) &= (a_k + b_k)x^k + (a_{k-1} + b_{k-1}x^{k-1} + \cdots + (a_1 + b_1)x + (a_0 + b_0) \\ f(x)g(x) &= c_{m+n}x^{m+n} + c_{m+n-1}x^{m+n-1} + \cdots + c_1 x + c_0, \end{aligned}$$

where $k = \max\{m, n\}, a_i = 0$ for $i > n$ and $b_i = 0$ for $i > m$. Also for $0 \leq j \leq m + n$, $c_j = a_0 b_j + a_1 b_{j-1} + a_2 b_{j-2} \cdots + a_{j-1} b_1 + a_j b_0$.

EXAMPLE. 1.4 In $\mathbb{Z}_6[x]$ let $f(x) = \bar{3}x^3 + \bar{2}x^2 + \bar{2}x + \bar{5}$ and $g(x) = \bar{3}x^2 + \bar{4}x + \bar{2}$. Then,

$$\begin{aligned} f(x) + g(x) &= (\bar{3} + 0)x^3 + (\bar{2} + \bar{3})x^2 + (\bar{2} + \bar{4})x + (\bar{5} + \bar{2}) \\ &= \bar{3}x^3 + \bar{5}x^2 + \bar{1}, \\ f(x).g(x) &= (\bar{3}x^3 + \bar{2}x^2 + \bar{2}x + \bar{5})(\bar{3}x^2 + \bar{4}x + \bar{2}) \\ &= \bar{3}x^5 + \bar{2}x^3 + \bar{3}x + \bar{4}. \end{aligned}$$

Henceforth we shall write k instead of $\bar{k}$ to denote a member of $\mathbb{Z}_n$.

THEOREM. 1.5 *If R is a ring then $R[x]$ is a ring.*

PROOF. It is a routine verification and hence the proof is omitted. ■

It also immediately follows that the ring $R[x]$ is commutative if and only if R is commutative and if R contains the unity element 1 then the constant polynomial 1 is the unity element of $R[x]$.

For a ring R, $R[x]$ may contain divisors of zero, for example in $\mathbb{Z}_8[x]$, $f(x) = 2x^2 + 4x + 6$ and $g(x) = 4x + 4$ both are non-zero polynomials but $f(x)g(x) = (2x^2 + 4x + 6)(4x + 4) = 8x^3 + 24x^2 + 40x + 24 = 0$.

THEOREM. 1.6 *If D is an integral domain then $D[x]$ is also an integral domain.*

PROOF. Since D is a commutative ring with unity so is $D[x]$. To show that $D[x]$ contains no divisors of zero, take $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ and $g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$, two non-zero polynomials with $a_n \neq 0, b_m \neq 0$. Then the leading term of $f(x)g(x) = a_n b_m x^{n+m}$. D being an integral domain we must have $a_n b_m \neq 0$ and hence $f(x)g(x) \neq 0$.

Hence $D[x]$ is an integral domain. ■

It immediately follows that for an integral domain D and for $f(x), g(x) \in D[x]$, $\deg(f(x)g(x)) = \deg f(x) + \deg g(x)$ and $\deg(f(x) + g(x)) \leq \max\{\deg f(x), \deg g(x)\}$.

1.2 Division Algorithm

DEFINITION. 1.7 Let D be an integral domain, $f(x), g(x) \in D[x]$. $g(x)$ is called a *factor of* $f(x)$ if there exists $h(x) \in D[x]$ such that $f(x) = g(x)h(x)$. In such a case we also say that $g(x)$ *divides* $f(x)$ and write as $g(x) \mid f(x)$.

DEFINITION. 1.8 Let R be a ring, $f(x) \in R[x]$. For $a \in R$, $f(a)$ denotes the element of R obtained by substituting x by a in the expression of $f(x)$. The element a is called a *zero* of $f(x)$ is $f(a) = 0$.

If F is a field then $a \in F$ is called a *zero of multiplicity* k ($k \in \mathbb{N}$) of the polynomial $f(x) \in F[x]$ if $(x - a)^k$ is a factor of $f(x)$ but $(x - a)^{k+1}$ is not a factor of $f(x)$.

THEOREM. 1.9 *Let F be a field and $f(x), g(x) \in F[x]$ with $g(x) \neq 0$. Then there exist unique polynomials $q(x), r(x)$ in $F[x]$ such that $f(x) = g(x)q(x) + r(x)$, where either $r(x) = 0$ or $\deg r(x) < g(x)$.*

PROOF. The result will be proven by induction method. If $f(x) = 0$ or $\deg f(x) < \deg g(x)$ we take $q(x) = 0$ and $r(x) = f(x)$ so that $f(x) = g(x)q(x) + r(x)$. Hence the result is proved for this case.

Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, $a_n \neq 0$, $g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$, $b_m \neq 0$ and $n > m$. Note that the result is proved for $n = 0$. Assume that the result is proved for all integers less than n.

Define $f_1(x) = f(x) - a_n b_m^{-1} x^{n-m} g(x)$. Then $\deg f_1(x) < n$ and hence by induction hypothesis there exist $q_1(x), r(x) \in F[x]$ such that $f_1(x) = g(x)q_1(x) + r(x)$ where either $r(x) = 0$ or $\deg r(x) < m$. Then

$$\begin{aligned} f(x) &= f_1(x) + a_n b_m^{-1} x^{n-m} g(x) \\ &= g(x)q_1(x) + r(x) + a_n b_m^{-1} x^{n-m} g(x) \\ &= (q_1(x) + a_n b_m^{-1} x^{n-m})g(x) + r(x) \;=\; q(x)g(x) + r(x), \end{aligned}$$

where $q(x) = q_1(x) + a_n b_m^{-1} x^{n-m}$ and either $r(x) = 0$ or $\deg r(x) < \deg g(x) = m$. So the result is true for n. Hence by induction is true for every integer n.

To prove the uniqueness let us assume that $f(x) = q(x)g(x) + r(x) = q_1(x)g(x) + r_1(x)$, where either $r(x) = 0$ or $\deg r(x) < \deg g(x)$ and either $r_1(x) = 0$ or $\deg r_1(x) < \deg g(x)$. Subtracting we have $(q(x) - q_1(x))g(x) + r(x) - r_1(x) = 0$, where either $r(x) - r_1(x) = 0$ or $\deg(r(x) - r_1(x)) < \deg g(x)$. Since $r(x) - r_1(x) = g(x)(q_1(x) - q(x))$, $\deg(r(x) - r_1)$ must be equal to that of $g(x)(q_1(x) - q(x))$ which is impossible since $\deg g(x) > \deg(r(x) - r_1))$. Hence $r(x) - r_1(x) = 0$, i.e., $r(x) = r_1(x)$ and $q(x) - q_1(x) = 0$, i.e., $q(x) = q_1(x)$. ■

DEFINITION. 1.10 The polynomial $q(x), r(x)$ obtained by division algorithm are respectively called the *quotient* and *remainder* when the polynomial $f(x)$ is divided by $g(x)$.

EXAMPLE. 1.11 Find the quotient and remainder when $f(x) = x^3 + 2x^2 + 5x + 3$ is divided by $g(x) = 2x^2 + x + 1$, $f(x), g(x)$ are in $\mathbb{Z}_7[x]$.

$$\begin{array}{r|l|l} 2x^2 + x + 1 & x^3 + 2x^2 + 5x + 3 \;\;| & 4x + 6 \\ & x^3 + 4x^2 + 4x \;\;\;\;\;\;| & \\ \hline & \quad 5x^2 + x + 3 & \\ & \quad 5x^2 + 6x + 6 & \\ \hline & \qquad\qquad 2x + 4 & \end{array}$$

Hence the quotient is $q(x) = 4x + 6$ and the remainder is $r(x) = 2x + 1$.

COROLLARY. 1.12 *If F is a field then for any $f \in F[x]$ and for any $a \in F$, $f(a)$ is the remainder when $f(x)$ is divided by $x - a$.*

PROOF. Taking $g(x) = x - a$, by Division algorithm $f(x) = (x - a)q(x) + r(x)$, where either $r(x) = 0$ or $\deg r(x) < \deg(x - a) = 1$. Thus $r(x)$ is a constant polynomial, say $r(x) = c$. Hence $f(x) = (x - a)q(x) + c$ which gives $f(a) = r$. ■

COROLLARY. 1.13 *If F is a field and $f(x) \in F[x]$ then $a \in F$ is a zero of $f(x)$ if and only if $(x - a)$ is a factor of $f(x)$.*

PROOF. immediately follows. ■

THEOREM. 1.14 *If F is a field then a polynomial of degree n in $F[x]$ can have at most n zeros counting multiplicity.*

PROOF. The proof is done with help of mathematical induction. If $f(x) \in F[x]$ is of degree 0 then it has no root and hence the result is true for $n = 0$.

Assume that the result is true for every integer $m \leq n$ and $f(x)$ is a polynomial over F with degree n. If $f(x)$ has no zero then it is done, otherwise let a be a zero of $f(x)$ of multiplicity k. Then $f(x) = (x-a)^k q(x)$ and $q(a) \neq 0$. Now $\deg f(x) = \deg(x-a)^k + \deg q(x)$, i.e., $n = k + \deg q(x)$ which shows that $n \geq k$. Also $\deg q(x) = n - k$. If $q(x)$ has no zero then the prove is done, otherwise if b is a zero of $q(x)$ then $f(b) = (b-a)^k q(b) = 0$ which shows that b is also a zero of $f(x)$. Also multiplicity of b in $f(x)$ is same as that in $q(x)$.

By induction hypothesis $q(x)$ can have at most $n-k$ zeros and hence $f(x)$ can have at most $k + (n-k) = n$ zeros. This completes the proof. ■

EXAMPLE. 1.15

1. It can be observed that when F is not a field the result may not be true. take the polynomial $f(x) = x^2 + 3x + 2$ in $\mathbb{Z}_6[x]$. Then $f(1) = 0, f(2) = 0, f(4) = 0$ and $f(5) = 0$. Thus the polynomial of the second degree has four zeros.

2. List all the polynomials of degree 2 in $\mathbb{Z}_2[x]$. Which of these are equal as functions from $\mathbb{Z}_2$ to $\mathbb{Z}_2$?

   $\mathbb{Z}_2 = \{0, 1\}$. The polynomials of degree 2 are $f_1(x) = x^2$ and $f_2(x) = x^2 + 1$, $f_3(x) = x^2 + x$ and $f_4(x) = x^2 + x + 1$. Here $f_1(x)$ and $f_2(x)$ are functions from $\mathbb{Z}_2$ onto $\mathbb{Z}_2$, $f_3(0) = f_3(1) = 0$ and $f_4(0) = f_4(1) = 1$.

3. Let R be a commutative ring. Show that $R[x]$ has a subring isomorphic to R.

   The mapping $a \mapsto \underline{a}$, where $\underline{a}$ denotes the constant polynomial a, is the required isomorphism.

4. If $\phi : R \to S$ is a ring homomorphism. For $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in R[x]$ define $\bar{\phi} : R[x] \to S[x]$ by

   $$\bar{\phi}(f(x)) = \phi(a_n)x^n + \phi(a_{n-1})x^{n-1} + \cdots + \phi(a_1)x + \phi(a_0)$$

   for all $f(x) \in R[x]$, is a ring homomorphism. Moreover if ϕ is an isomorphism then so is $\bar{\phi}$.

5. Show that the polynomial $f(x) = 2x + 1$ in $\mathbb{Z}_4$ is a unit.

   $(2x + 1)(2x + 1) = 4x^2 + 4x + 1 = 1$. Hence $(2x + 1)^{-1} = 2x + 1$, i.e., $2x_1$ is a unit in $\mathbb{Z}_4$.

6. Let F be a field. Show that there exist a, b in F with the property that $x^2 + x + 1$ divides $x^{43} + ax + b$.

   By division algorithm there exist $q(x), r(x)$ in $F[x]$ such that $x^{43} = (x^2+x+1)q(x)+r(x)$ where either $r(x) = 0$ or $\deg r(x) < 2 = \deg(x^2 + x + 1)$. Let $r(x) = cx + d$. Then $x^{43} = (x^2 + x + 1)q(x) + cx + d$, i.e. $x^{43} - cx - d = (x^2 + x + 1)q(x)$, which shows that $x^{43} - cx - d$ is divisible by $x^2 + x + 1$. Hence taking $a = -c, b = -d$, $x^2 + x + 1$ divides $x^{43} + ax + b$.

1.3 Principal Ideal Domain

Recall that a subring I of a ring R is called a *left ideal* if for all $x \in R$ for all $a \in I$, $xa \in I$. I is called a right ideal of R if for all $x \in R$ for all $a \in I$, $ax \in I$. An ideal I is both a left ideal and a right ideal. In a commutative ring every left ideal is a right ideal and vice versa.

In a ring R with unity 1 and $A \subset R$ the set

$$AR = \{a_1x_1 + \cdots + a_nx_n : a_1, a_2, \ldots a_n \in A, x_1, x_2 \ldots, x_n \in R, n \in \mathbb{N}\}$$

is the smallest (with respect to set inclusion) right ideal containing A, called the *right ideal generated by* A. The left ideal RA generated by A is defined analogously. Ideal generated by A, denoted by $\langle A\rangle$, is the set

$$\langle A\rangle = \left\{\sum_{i=1}^{n} x_i a_i y_i : a_1, \ldots a_n \in A, x_1, \ldots, x_n, y_1, \ldots, y_n \in R, n \in \mathbb{N}\right\}$$

If A is a finite set then $\langle A\rangle$ is called a *finitely generated ideal*. If $a = \{a\}$, a singleton set then the ideal $\langle A\rangle$ is called a *principal ideal*, in this case $\langle A\rangle$ is written as $\langle a\rangle$.

Definition. 1.16 A commutative ring with unity is called a *Principal Ideal Ring* if every ideal in that ring is a principal ideal. An integral domain D is called a *Principal Ideal Domain* (PID) if every ideal in D is a principal ideal, i.e., of the form $\langle a\rangle$ for some $a \in D$.

Theorem. 1.17 *For a field F, $F[x]$ is a PID.*

PROOF. Let I be an ideal in $F[x]$. If $I = \{0\}$ then $I = \langle 0\rangle$. Otherwise there is non-zero elements in I. Let us choose $g(x) \in I$ with $g(x)$ is of minimum degree in I. Then $\langle g(x)\rangle \subset I$. Choose $f(x) \in I$. Then by division algorithm there exist $q(x), r(x)$ such that $f(x) = g(x)q(x) + r(x)$, where either $r(x) = 0$ or $\deg r(x) < \deg g(x)$. This shows that $r(x) = f(x) - g(x)q(x) \in I$. Since degree of $g(x)$ is minimum in I and $\deg r(x) < \deg g(x)$ we must have $r(x) = 0$ and hence $f(x) = g(x)q(x) \in \langle g(x)\rangle$, i.e., $I \subset \langle g(x)\rangle$. Hence $I = \langle g(x)\rangle$. ■

COROLLARY. 1.18 *Let F be a field and I be an ideal in $F[x]$. Then $I = \langle g(x)\rangle$ if and only if $g(x)$ is a non-zero polynomial of minimum degree in I.*

PROOF. Assume that $I = \langle g(x)\rangle$ and choose $f(x) \in I$. Then $f(x) = g(x)q(x)$ for some $q(x) \in F[x]$ and hence $\deg f(x) = \deg g(x) + \deg q(x) \geq \deg g(x)$. Thus degree of $g(x)$ is minimum in I. Conversely, if $g(x) \in I$ with minimum degree then by the above theorem $I = \langle g(x)\rangle$. Hence the result. ■

COROLLARY. 1.19 *If F is a field and I is an ideal then there is a unique monic polynomial (a polynomial whose coefficient of highest degree term is 1) g such that $I = \langle g(x)\rangle$.*

PROOF. Taking any $g \in I$ with minimal degree, $I = \langle g(x)\rangle$. We can make g monic by multiplying it by the inverse of its coefficient of highest degree term. If there is another monic polynomial $h \in I$ such that $I = \langle h(x)\rangle$, then there exist polynomials $q_1, q_2 \in F[x]$ such that $h = gq_1$ and $g = hq_2$. Hence $g = hq_2 = gq_1q_2$ which implies that $\deg g = \deg g + \deg q_1 + \deg q_2$. Thus $\deg q_1 = \deg q_2 = 0$, i.e., q_1, q_2 are constant polynomials. Since g, h are monic polynomials, we must have $q_1 = q_2 = 1$, i.e., $g = h$. ■

THEOREM. 1.20 *Let R be a commutative ring with unity. If $R[x]$ is a PID then R is a field.*

PROOF. Assume that $R[x]$ is a PID. It is sufficient to show that every non-zero element of R is a unit. Let $a \in R$, $a \neq 0$. Consider the ideal I generated by a and x, i.e., $I = \langle\{a, x\}\rangle$. multiple $R[x]$ being a PID there exists $f(x) \in R[x]$ such that $I = \langle f(x)\rangle$. Since $a \in I, x \in I$ there exist $p(x), q(x) \in R[x]$ such that $a = f(x)p(x)$ and $x = f(x)q(x)$. Hence $f(x)$ must be a constant, say $f(x) = r$.

Also $x = f(x)q(x) = rq(x)$. Thus $\deg q(x) = 1$, say, $q(x) = cx + d$ and hence $x = r(cx + d)$ which shows that $rc = 1$ and $rd = 0$. Thus r is an unit.

Since $r \in I$ it follows that $I = R[x]$. Since $1 \in I$, there exist $u(x), v(x) \in R[x]$ such that $1 = au(x) + xv(x)$ which shows that the constant term of $au(x)$ must be 1, i.e., if the constant term of $u(x)$ is b then $ab = 1$. Thus a is a unit.

Since $a \neq 0$ has been chosen in R arbitrarily, it follows that every non-zero element of R is a unit. Hence R is a field. ■

EXAMPLE. 1.21 $\mathbb{R}[x]/\langle x^2+1\rangle \simeq \mathbb{C}$.

Let us define a mapping $\psi : \mathbb{R}[x] \to \mathbb{C}$ by $\psi(f(x)) = f(i)$ for all $f(x) \in \mathbb{R}[x]$. It can be easily be verified that ψ is a ring homomorphism. Also for $z = a + ib \in \mathbb{C}$, taking $f(x) = bx + a$ we have $\psi(f(x)) = bi + a = z$. Thus ψ is onto.

$f(x) \in \ker\psi$ if and only if $\psi(f(x)) = 0$ if and only if $f(i) = 0$ if and only if $x^2 + 1$ is a factor of $f(x)$. Hence $\ker\psi = \langle x^2+1\rangle$. By First isomorphism theorem, $\mathbb{R}[x]/\ker\psi \simeq \mathbb{C}$, i.e., $\mathbb{R}[x]/\langle x^2+1\rangle \simeq \mathbb{C}$.

EXAMPLE. 1.22 Let F be a field, and suppose that $f(x), g(x) \in F[x]$. If there is no polynomial of positive degree in $F[x]$ that divides both $f(x)$ and $g(x)$ [that is, $f(x)$ and $g(x)$ are *relatively prime*], prove that there exist polynomials $h(x)$ and $k(x)$ in $F[x]$ such that $f(x)h(x) + g(x)k(x) = 1$.

Consider the ideal $I = \langle\{f(x), g(x)\}\rangle$. F being a field, $F[x]$ is a PID, so there exists $u(x) \in F[x]$ such that $I = \langle u(x)\rangle$. Since $f(x), g(x) \in I$, we have $f(x) = v(x)u(x)$ and $g(x) = w(x)u(x)$ for some $v(x), w(x) \in F[x]$. This shows that $u(x)$ is a common divisor of $f(x)$ and $g(x)$. By condition $u(x)$ must be a constant, say $u(x) = c$. Also $c \neq 0$, otherwise $I = \langle 0\rangle$. Thus $1 = cc^{-1} \in I$ and hence $I = F[x]$.

Since $1 \in F[x] = \langle\{f(x), g(x)\}\rangle$ there exist polynomials $h(x)$ and $k(x)$ in $F[x]$ such that $1 = f(x)h(x) + g(x)k(x)$.

EXAMPLE. 1.23 Let $f(x)$ be a non-constant element in $\mathbb{Z}[x]$. Prove that $\langle f(x)\rangle$ is not a maximal ideal in $\mathbb{Z}[x]$.

Since $f(x)$ is a non-constant polynomial, we can find $a \in \mathbb{Z}$ such that $f(a) \neq \pm 1$ and $f(a) \neq 0$. If $I = \langle\{f(a), f(x)\}\rangle$ then $f(a) \in I$, also $f(a) \notin \langle f(x)\rangle$ since elements of $\langle f(x)\rangle$ have degree at least 1, whereas $f(a)$ has degree 0. Thus the ideal I properly contains the ideal $\langle f(x)\rangle$.

I is a proper ideal, otherwise if $I = \mathbb{Z}[x]$ then $1 \in I$, so that there exist $h(x), k(x) \in \mathbb{Z}[x]$ such that $1 = h(x)f(a) + k(x)f(x)$ which gives $1 = h(a)f(a) + k(a)f(a) = (h(a) + k(a))f(a)$, i.e., $f(a)$ divides 1 — a contradiction. Hence $I \neq \mathbb{Z}[x]$.

So, $\langle f(x)\rangle \subsetneqq I \subsetneqq \mathbb{Z}[x]$ shows that $\langle f(x)\rangle$ is not a maximal ideal.

1.4 Factorisation of Polynomials

DEFINITION. 1.24 Let D be an integral domain. A polynomial $f(x) \in D[x]$ is said to be an *irreducible polynomial* over D if:

(i) $f(x)$ is neither a zero polynomial nor a unit of $D[x]$,

(ii) if $f(x)$ is expressed as $f(x) = g(x)h(x)$, where $g(x), h(x) \in D[x]$, then either $g(x)$ or $h(x)$ is a unit in $D[x]$

If a non-zero non-unit polynomial is not irreducible then it is called a *reducible* polynomial.

For a field F the definition becomes as follows: a polynomial $f(x) \in F[x]$ is called an *irreducible polynomial* over F if $f(x)$ can not be expressed as a product of two polynomials of degree lower than that of $f(x)$.

EXAMPLE. 1.25

1. The polynomial $2x^2+2$ is reducible in $\mathbb{Q}[x]$ as $2x^2+2 = 2(x^2+1)$ and 2 is a unit in $\mathbb{Q}[x]$ but is irreducible in $\mathbb{Z}[x]$ since neither 2 nor $2x+1$ is a unit in $\mathbb{Z}[x]$.
2. The polynomial $2x^2 - 3$ is reducible in $\mathbb{R}[x]$ since $x^2 - 3 = (x+\sqrt{3})(x-\sqrt{3})$, a product of two polynomials of lower degree, but is irreducible in $\mathbb{Q}[x]$.
3. The polynomial $x^2 + 1$ is irreducible in $\mathbb{R}[x]$ but reducible in $\mathbb{Q}[x]$ as $x^2 + 1 = (x+i)(x-i)$.
4. Every polynomial in $\mathbb{C}[x]$ with degree more than 1 is reducible.

THEOREM. 1.26 *If F is a field then a polynomial $f(x)$ in $F[x]$ of degree 2 or 3 is reducible over F if and only if $f(x)$ has a zero in F.*

PROOF. Assume that $f(x) \in F[x]$ has a zero a in F. Then $x - a$ is a factor of $f(x)$, i.e., $f(x) = (x-a)h(x)$ for some polynomial $h(x)$ and hence $f(x)$ is reducible.

Conversely, assume that $f(x)$ is reducible over F. Then there exists $g(x), h(x) \in F[x]$ such that $f(x) = g(x)h(x)$. Since $\deg f(x) = 3 = \deg g(x) + \deg h(x)$, one of $g(x), h(x)$ must have degree 1, say $g(x) = ax + b$. Then $x = a^{-1}b$ is a zero of $g(x)$ and hence a zero of $f(x)$. ■

EXAMPLE. 1.27 Consider the polynomial $f(x) = x^2 + x + 1$. In $\mathbb{Z}_7[x]$, $x = 2$ is a zero of $f(x)$ and hence $f(x)$ is reducible over $\mathbb{Z}_7$, $f(x) = (x-2)(x-4)$. But in $\mathbb{Z}_5$, $f(0) = 1, f(1) = 3, f(2) = 2, f(3) = 3, f(4) = 1$, i.e., none of $0, 1, 2, 3, 4$ is a zero of $f(x)$ and hence $f(x)$ is irreducible in $\mathbb{Z}_5[x]$.

EXAMPLE. 1.28 A polynomial in $F[x]$ of degree larger than 3 may be reducible even without having a zero in F. For example, $f(x) = x^4 + 5x^2 + 6 = (x^2 + 2)(x^2 + 3)$ is reducible in $\mathbb{R}[x]$ but has no zero in $\mathbb{R}$.

DEFINITION. 1.29 Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 \in \mathbb{Z}[x]$. The gcd of $a_0, a_1, \ldots, a_n$ is called the *content* of $f(x)$. The polynomial $f(x)$ is called a *primitive polynomial* if the content of $f(x)$ is 1.

EXAMPLE. 1.30 The polynomial $f(x) = 2x^3 + 4x + 6$ is not primitive as its content is 2, whereas the polynomial $g(x) = 3x^2 + 5x + 2$ is a primitive polynomial as its content is $\gcd(3, 5, 2) = 1$.

THEOREM. 1.31 *The product of two primitive polynomials is primitive.*

PROOF. Assume that $f(x), f(x)$ are primitive polynomials. If possible, suppose that $h(x) = f(x)g(x)$ is not a primitive polynomial, let p be a prime factor of the content of $h(x)$.

Let $\bar{f}(x), \bar{g}(x)$ denote the polynomials in $\mathbb{Z}_p[x]$ obtained by reducing the coefficients of $f(x), g(x)$ respectively modulo p. Hence $\bar{h}(x) = \bar{f}(x)\bar{g}(x)$. Since p is a divisor of the content of $h(x)$, p divides every coefficient of $h(x)$ and hence $h(x)$ is the zero polynomial of $\mathbb{Z}_p[x]$.

$\mathbb{Z}_p[x]$ being an integral domain, and $\bar{f}(x)\bar{g}(x) = 0$, we have either $\bar{f}(x) = 0$ or $\bar{g}(x) = 0$ in $\mathbb{Z}_p[x]$, i.e., either content of $f(x)$ is a multiple of p or content of $g(x)$ is a multiple of p, i.e., either $f(x)$ or $g(x)$ is not a primitive polynomial — a contradiction. Hence $h(x)$ is a primitive polynomial. ■

THEOREM. 1.32 *Let F be a field and $f(x) \in F[x]$, $a \in F$, $a \neq 0$. Then*

1. *If $af(x)$ is irreducible over F then $f(x)$ is irreducible over F.*
2. *If $f(ax)$ is irreducible over F then $f(x)$ is irreducible over F.*
3. *If $f(x + a)$ is irreducible over F then $f(x)$ is irreducible over F.*

PROOF. 1. If $f(x)$ is reducible over F then there exist $g(x), h(x) \in F[x]$ with $\deg g(x) \geq 1, \deg h(x) \geq 1$ such that $f(x) = g(x) \cdot h(x)$. Hence $af(x) = ag(x) \cdot h(x)$ showing that $af(x)$ is reducible.

2. Assume that $f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ is reducible over F. Then there exist $g(x), h(x) \in F[x]$ such that $f(x) = g(x) \cdot h(x)$. Let $g(x) = b_mx^m + \cdots + b_1x + b_0$ and $h(x) = c_kx^k + \cdots + c_1x + c_0$ where $m + k = n$, $m \geq 1, k \geq 1$. Here for $0 \leq r \leq n$, $a_r = b_rc_0 + b_{r-1}c_1 + \cdots + b_0c_r$. So the coefficient of x^r in $g(ax) \cdot h(ax)$ is $b_ra^rc_0 + b_{r-1}a^{r-1}c_1r + \cdots + b_0c_ra^r = a^r(b_rc_0 + b_{r-1}c_1 + \cdots + b_0c_r) = a^ra_r$ which is the coefficient of coefficient of x^r in $f(ar)$. Hence $f(ar) = g(ar) \cdot h(ar)$ showing that $f(ar)$ is reducible — contradiction.

3. Assume that $f(x)$ is irreducible. If $f(x + a)$ is reducible then there exist $g(x), h(x) \in F[x]$ such that $f(x + a) = g(x) \cdot h(x)$. Put $x + a = z$, then $f(z) = g(z - a) \cdot h(z - a)$, i.e., $f(x) = g(x - a) \cdot h(x - a)$ showing that $f(x)$ is reducible — contradiction. ■

THEOREM. 1.33 *If a polynomial $f(x) \in \mathbb{Z}[x]$ is reducible over $\mathbb{Q}$ then it is reducible over $\mathbb{Z}$.*

PROOF. Assume that $f(x) = g(x)h(x)$ where $g(x), h(x) \in \mathbb{Q}[x]$. Without any loss of generality we may assume that $f(x)$ is a primitive polynomial, otherwise we may divide both $f(x)$ and $g(x)$ by the content of $f(x)$. Let a be the lcm of the denominators of the coefficients of $g(x)$ and b be the lcm of the denominators of the coefficients of $h(x)$. Then $ag(x)$ and $bh(x)$ both belong to $\mathbb{Z}[x]$. Let c be the content of $ag(x)$ and d be the content of $bh(c)$ so that $ag(x) = cg_1(x)$ and $bh(x) = dh_1(x)$ where $g_1(x), h_1(x)$ are primitive polynomials.

Thus $abf(x) = ag(x) \cdot bh(x) = cg_1(x) \cdot dh_1(x) = cdg_1(x)h_1(x)$. Since the product of two primitive polynomials is a primitive polynomial, we have $g_1(x)h_1(x)$ is a primitive polynomial and also since $f(x)$ is a primitive polynomial the content of $abf(x)$ is ab. Thus $ab = cd$, which shows that $f(x) = g_1(x)h_1(x)$ where $g_1(x), h_1(x) \in \mathbb{Z}[x]$. Hence $f(x)$ is reducible over $\mathbb{Z}$. ■

It is also observed in the above theorem that $\deg g(x) = \deg g_1(x)$ and $\deg h(x) = \deg h_1(x)$. Since $\deg g(x) < \deg f(x)$ and $\deg h(x) < \deg f(x)$ in $\mathbb{Q}[x]$ we have $\deg g_1(x) < \deg f(x)$ and $\deg h_1(x) < \deg f(x)$ in $\mathbb{Z}[x]$.

The next theorem is known as mod p test of irreducibility.

THEOREM. 1.34 *Let $f(x) \in \mathbb{Z}[x]$ with $\deg f(x) \geq 1$ and p be a prime. Let $\bar{f}(x)$ denote the polynomial in $\mathbb{Z}_p[x]$ obtained from $f(x)$ by reducing all of its coefficients modulo p. If $\bar{f}(x)$ is irreducible over $\mathbb{Z}_p$ and $\deg f(x) = \deg \bar{f}(x)$ then $f(x)$ is irreducible over $\mathbb{Q}$.*

PROOF. Assume that $\bar{f}(x)$ is irreducible over $\mathbb{Z}_p$ and $\deg f(x) = \deg \bar{f}(x)$. If possible, let $f(x)$ be reducible over $\mathbb{Q}$. Then $f(x)$ is reducible over $\mathbb{Z}$ where $f(x) = g(x)h(x)$;

$g(x), h(x) \in \mathbb{Z}[x]$ with $\deg g(x) < \deg f(x)$ and $\deg h(x) < \deg f(x)$. Let $\bar{g}(x), \bar{h}(x)$ denote the polynomials obtained from $g(x)$ and $h(x)$ respectively by reducing their coefficients modulo p. Then $\bar{f}(x) = \bar{g}(x)\bar{h}(x)$. Also $\deg \bar{g}(x) \leq \deg g(x) < \deg f(x) = \deg \bar{f}(x)$ and $\deg \bar{h}(x) \leq \deg h(x) < \deg f(x) = \deg \bar{f}(x)$. This shows that $\bar{f}(x)$ is reducible over $\mathbb{Z}_p$ — a contradiction. Hence $f(x)$ must be irreducible over $\mathbb{Q}$. ■

EXAMPLE. 1.35 Let $f(x) = 21x^3 - 3x^2 + 2x + 9$.

In $\mathbb{Z}_2$, $\bar{f}(x) = x^3 - x^2 + 1$. Then $\bar{f}(0) = 1, \bar{f}(1) = 1$ and hence $\bar{f}(x)$ has no zero in $\mathbb{Z}_2$ which implies that $\bar{f}$ is irreducible in $\mathbb{Z}_2[x]$. Also $\deg \bar{f}(x) = \deg f(x) = 3$. Hence $f(x)$ is irreducible over $\mathbb{Q}$.

EXAMPLE. 1.36 1. Test irreducibility of the polynomial $f(x) = x^4 + x + 1$.

In $\mathbb{Z}_2[x]$, $\bar{f}(x) = x^4 + x + 1$. Note that $\bar{f}(0) = 1, \bar{f}(1) = 1$, so $\bar{f}(x)$ has no zero in $\mathbb{Z}_2$. So it has no linear factor in $\mathbb{Z}_2[x]$.

The possible quadratic polynomials in $\mathbb{Z}_2[x]$ are $x^2, x^2 + x, x^2 + 1$ and $x^2 + x + 1$. The first three have zeros in $\mathbb{Z}_2$ and hence can not be a factor of $\bar{f}(x)$.

$$\begin{array}{r|l|l} x^2+x+1 & x^4+x+1 & x^2-x \\ & x^4+x^3+x^2 & \\ \hline & -x^3-x^2+x+1 & \\ & -x^3-x^2-x & \\ \hline & 1 & \end{array}$$

For $x^2 + x + 1$ a long division shows that it is not a factor of $\bar{f}(x)$.

Hence $\bar{f}(x)$ is irreducible over $\mathbb{Z}_2$ and hence $f(x)$ is irreducible over $\mathbb{Q}$.

2. Test the irreducibility of the polynomial $f(x) = x^5 + 5x^2 + 1$.

In $\mathbb{Z}_2[x]$, $\bar{f}(x) = x^5 + x^2 + 1$. Here $f(0) = f(1) = 1$, so $\bar{f}(x)$ has no zero in $\mathbb{Z}_2$ and hence $\bar{f}(x)$ has no linear factor in $\mathbb{Z}_2[x]$. The only quadratic irreducible polynomial in $\mathbb{Z}_2[x]$ is $x^2 + x + 1$.

$$\begin{array}{r|l|l} x^2+x+1 & x^5+x^2+1 & x^3-x^2 \\ & x^5+x^4+x^3 & \\ \hline & -x^4-x^3+x^2+1 & \\ & -x^4-x^3-x^2 & \\ \hline & 1 & \end{array}$$

Thus a long division shows that $x^2 + x + 1$ is not a factor of $\bar{f}(x)$. So $\bar{f}(x)$ has no quadratic factor.

Since $\deg \bar{f}(x) = 5$ and $\bar{f}(x)$ has no linear or quadratic factor it can not have a cubic factor of a biquadratic factor. Hence $\bar{f}(x)$ is irreducible over $\mathbb{Z}_2$ and hence $f(x)$ is irreducible over $\mathbb{Q}$.

3. Test the irreducibility of $f(x) = \frac{3}{7}x^4 - \frac{2}{7}x^2 + \frac{9}{35}x + \frac{3}{5}$.

   The lcm of the denominators of the coefficients of $f(x)$ is 35. So $h(x) = 35f(x) = 15x^4 - 10x^2 + 9x + 21 \in \mathbb{Z}[x]$. Then $f(x)$ is irreducible over $\mathbb{Q}$ if $h(x)$ is irreducible over $\mathbb{Z}$.

   Now, on $\mathbb{Z}_2[x]$, $\bar{h}(x) = x^4 + x + 1$. Since $\bar{h}(0) = \bar{h}(1) = 1$, $\bar{h}(x)$ has no zero in $\mathbb{Z}_2$ and hence $\bar{h}(x)$ has no linear factor in $\mathbb{Z}_2[x]$. The only irreducible quadratic in $\mathbb{Z}_2[x]$ is $x^2 + x + 1$ and a long division shows that $x^2 + x + 1$ is not a factor of $\bar{h}(x)$. Thus $\bar{h}(x)$ is irreducible over $\mathbb{Z}_2$. Hence $f(x)$ is irreducible over $\mathbb{Q}$.

THEOREM. 1.37 (EISENSTEIN'S CRITERION) *Let*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x].$$

If there exists a prime p such that $p \nmid a_n$, $p \mid a_{n-1}, \ldots, p \mid a_0$ and $p^2 \nmid a_0$ then $f(x)$ is irreducible over $\mathbb{Q}$.

PROOF. If possible, assume that $f(x)$ is reducible over $\mathbb{Q}$. Then $f(x)$ is reducible over $\mathbb{Z}$ and $f(x) = g(x)h(x)$ where $g(x), h(x) \in \mathbb{Z}[x]$ with $1 \leq \deg g(x) < n$ and $1 \leq \deg h(x) < n$.

Let $g(x) = b_k x^k + b_{k-1}x^{k-1} + \cdots + b_1 x + b_0$ and $h(x) = c_l x^l + c_{l-1}x^{l-1} + \cdots + c_1 x + c_0$, where $k + l = n$. Then $a_n = b_k c_l$ and $a_0 = b_0 c_0$. Since $p \mid a_0$ we have $p \mid b_0 c_0$, p being a prime either $p \mid b_0$ or $p \mid c_0$. Also $p^2 \nmid a_0$ implies that p does not divide both of b_0, c_0. Assume that $p \mid b_0$ and $p \nmid c_0$.

Since $p \nmid a_n = b_k c_l$, $p \nmid b_k$. Let r be the least integer such that $p \nmid b_r$, then $1 \leq r \leq k$. $a_r = b_r c_0 + b_{r-1}c_1 + \cdots + b_1 c_{r-1} + b_0 c_r$. Since p divides $b_{r-1}, b_{r-2}, \ldots, b_1, b_0$ we have $p \mid b_{r-1}c_1 + \cdots + b_1 c_{r-1} + b_0 c_r$. Also $p \mid a_r$ and hence $p \mid a_r - (b_{r-1}c_1 + \cdots + b_1 c_{r-1} + b_0 c_r)$ which implies that $p \mid b_r c_0$. This is a contradiction since p is a prime and $p \nmid b_r$ and $p \nmid c_0$. Hence $f(x)$ must be irreducible over $\mathbb{Q}$. ■

COROLLARY. 1.38 *For a prime number p, The Cyclotomic polynomial $\Phi_p(x) = x^{p-1} + \cdots + x + 1$ is irreducible over $\mathbb{Q}$.*

PROOF. Note that $\Phi_p(x) = \dfrac{x^p - 1}{x - 1}$. Then

$$\begin{aligned}
\Phi_p(x+1) &= \frac{(x+1)^p - 1}{(x+1) - 1} \\
&= \frac{1}{x}\left[\left(x^p + \binom{p}{1}x^{p-1} + \cdots + \binom{p}{p-1}x + 1\right) - 1\right] \\
&= x^{p-1} + \binom{p}{1}x^{p-2} + \cdots + \binom{p}{p-2}x + \binom{p}{1}
\end{aligned}$$

Here every coefficient except the leading coefficient is divisible by p and the constant term is p not divisible by p^2. Hence by Eisenstein's criterion $\Phi_p(x+1)$ is irreducible over $\mathbb{Q}$ and hence $\Phi_p(x)$ is irreducible over $\mathbb{Q}$. ■

EXAMPLE. 1.39 Test the irreducibility of the following polynomials:

1. $x^5 + 9x^4 + 12x^2 + 6$ over $\mathbb{Q}$.

   Here all the coefficients other than the leading coefficient is divisible by 3 and the constant term is not divisible by 3^2. Hence by Eisenstein's criterion the polynomial is irreducible.

2. $\frac{5}{2}x^5 + \frac{9}{2}x^4 + 15x^3 + \frac{3}{7}x^2 + 6x + \frac{3}{14}$ over $\mathbb{Q}$.

   $f(x) = \frac{5}{2}x^5 + \frac{9}{2}x^4 + 15x^3 + \frac{3}{7}x^2 + 6x + \frac{3}{14}$. To clear the fractions multiplying by 14 we have $14f(x) = 35x^5 + 63x^4 + 210x^3 + 6x^2 + 90x + 3$. All the coefficients other than the leading coefficient is divisible by 3 and the constant term is not divisible by $3^2 = 9$. Hence by Eisenstein's criterion $14f(x)$ is irreducible over $\mathbb{Q}$ and hence $f(x)$ is irreducible over $\mathbb{Q}$.

EXAMPLE. 1.40 1. Show that the polynomial x^4+1 is irreducible over $\mathbb{Q}$ but reducible over $\mathbb{R}$.

Let $f(x) = x^4 + 1$. Then $f(x+1) = (x+1)^4 + 1 = x^4 + 4x^3 + 6x^2 + 4x + 2$. Each of the coefficients other than the leading one is divisible by 2. Also the constant term is not divisible by 2^2. hence by Eisenstein's criterion $f(x+1)$ is irreducible in $\mathbb{Q}$ and hence $f(x)$ is irreducible over $\mathbb{Q}$.

To fine the zeros of x^4+1 in $\mathbb{C}$, we note that $x^4 = -1 = \cos(2k+1)\pi + i\sin(2k+1)\pi$. Hence $x = (\cos(2k+1)\pi + i\sin(2k+1)\pi)^{\frac{1}{4}} = \cos\frac{(2k+1)\pi}{4} + i\sin\frac{(2k+1)\pi}{4}$, $k = 0, 1, 2, 3$. When $k = 0$, let $\alpha = \cos\frac{\pi}{4} + i\sin\frac{\pi}{4}$, then the other roots correspond to $k = 1, 2, 3$,

i.e., $\cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4} = \alpha^3$, $\cos\frac{5\pi}{4} + i\sin\frac{5\pi}{4} = \alpha^5$ and $\cos\frac{7\pi}{4} + i\sin\frac{7\pi}{4} = \alpha^7$. Thus

$$\begin{aligned} x^4+1 &= (x-\alpha)(x-\alpha^3)(x-\alpha^5)(x-\alpha^7) \\ &= (x-\alpha)(x-\alpha^7)(x-\alpha^3)(x-\alpha^5) \\ &= (x^2-(\alpha+\alpha^7)+\alpha^8)(x^2-(\alpha^3+\alpha^5)+\alpha^8). \end{aligned}$$

Now, $\alpha^8 = (\alpha^4)^2 = (-1)^2 = 1$. Also $\alpha + \alpha^7 = (\frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}}) + (\frac{1}{\sqrt{2}} - i\frac{1}{\sqrt{2}}) = \sqrt{2}$ and $\alpha^3 + \alpha^5 = (-\frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}}) + (-\frac{1}{\sqrt{2}} - i\frac{1}{\sqrt{2}}) = -\sqrt{2}$

Hence $x^4+1 = (x^2-\sqrt{2}x+1)(x^2+\sqrt{2}x+1)$ which shows that x^4+1 is reducible over $\mathbb{R}$.

2. Show that the polynomial $f(x) = x^3+6$ is reducible in $\mathbb{Z}_7$. Write it as a product of irreducible polynomials.

   In $\mathbb{Z}_7$, $f(0) = 6, f(1) = 7 = 0, f(2) = 14 = 0, f(3) = 33 = 5, f(4) = 70 = 0, f(5) = 131 = 5$ and $f(6) = 216 = 6$. Thus $f(x)$ has zeros $1, 2$ and 4 in $\mathbb{Z}_7$ and hence $f(x)$ is reducible in $\mathbb{Z}_7$.

   The factors of $f(x)$ are $(x-1), (x-2)$ and $(x-4)$ which are equivalent to $(x+6), (x+5)$ and $(x+3)$ in $\mathbb{Z}_7$. Hence $x^3+6 = (x+3)(x+5)(x+6)$ in $\mathbb{Z}_7$.

It can be recalled that for a field F, the integral domain $F[x]$ is a PID. We study some more results on $F[x]$ where F is a field.

THEOREM. 1.41 *If F is a field then a polynomial $p(x) \in F[x]$ is irreducible if and only if $\langle p(x)\rangle$ is a maximal ideal.*

PROOF. Assume that $\langle p(x)\rangle$ is a maximal ideal. If possible, suppose that $p(x)$ is a reducible polynomial, say $p(x) = f(x)g(x)$ where $1 \le \deg f(x) < \deg p(x)$ and $1 \le \deg g(x) < \deg p(x)$. Then for $h(x) \in F[x]$,

$$\begin{aligned} h(x) \in \langle p(x)\rangle &\Rightarrow h(x) = p(x)q(x) \text{ for some } q(x) \in F[x] \\ &\Rightarrow h(x) = f(x)g(x)q(x) = f(x)q_1(x) \text{ where } q_1(x) = g(x)q(x) \\ &\Rightarrow h(x) \in \langle f(x)\rangle. \end{aligned}$$

Thus $\langle p(x)\rangle \subseteq \langle f(x)\rangle$. Since $\langle p(x)\rangle$ is a maximal ideal, either $\langle f(x)\rangle = F[x]$ or $\langle p(x)\rangle = \langle f(x)\rangle$. In the first case $\deg f(x) = 0$. In the second case $\deg p(x) = \deg f(x)$ and hence $\deg g(x) = 0$. Both lead to contradiction. Hence $p(x)$ is an irreducible polynomial over F.

Conversely, assume that $p(x)$ is irreducible polynomial. Let I be an ideal in $F[x]$ such that $\langle p(x)\rangle \subset I \subset F[x]$. Since $F[x]$ is a PID there exists $g(x) \in F[x]$ such that $I = \langle g(x)\rangle$. Since $p(x) \in I$ there exists $q(x) \in F[x]$ such that $p(x) = g(x)q(x)$. $p(x)$ being irreducible either $\deg g(x) = 0$ or $\deg q(x) = 0$ In the first case $g(x)$ is a constant polynomial and hence $I = F[x]$. In the second case $\deg p(x) = \deg g(x)$ and hence $I = \langle g(x)\rangle = \langle p(x)\rangle$. Both the cases indicate that $\langle p(x)\rangle$ is a maximal ideal in $F[x]$. ■

COROLLARY. 1.42 *If F is a field and $p(x)$ is an irreducible polynomial over F then $F[x]/\langle p(x)\rangle$ is a field.*

PROOF. Since $F[x]$ is a commutative ring with unity and $\langle p(x)\rangle$ is a maximal ideal the result follows immediately. ■

COROLLARY. 1.43 *Let F be a field and let $p(x), f(x), g(x) \in F[x]$. If $p(x)$ is irreducible over F and $p(x) \mid f(x)g(x)$, then either $p(x) \mid f(x)$ or $p(x) \mid h(x)$.*

PROOF. Since $p(x)$ is an irreducible polynomial $\langle p(x)\rangle$ is a maximal ideal. Also $F[x]$ being a commutative ring with unity, $\langle p(x)\rangle$ is a prime ideal. Now, $p(x) \mid f(x)g(x)$ implies that $f(x)g(x) \in \langle p(x)\rangle$. Since $\langle p(x)\rangle$ is a prime ideal, either $f(x) \in \langle p(x)\rangle$ or $g(x) \in \langle p(x)\rangle$. Hence either $p(x) \mid f(x)$ or $p(x) \mid g(x)$. ■

EXAMPLE. 1.44 Let F be a field and let $f(x)$ be a polynomial in $F[x]$. If $f(x)$ is reducible over F prove that $\langle f(x)\rangle$ is not a prime ideal in $F[x]$.

There exist $g(x), h(x) \in F[x]$ such that $f(x) = g(x)h(x)$ where $1 \leq \deg g(x) < \deg f(x)$ and $1 \leq \deg h(x) < \deg f(x)$. Then $f(x) \in \langle f(x)\rangle$, i.e., $g(x)h(x) \in \langle f(x)\rangle$ but none of $g(x), h(x)$ belong to $\langle f(x)\rangle$. Hence $\langle f(x)\rangle$ is not a prime ideal.

1.5 Unique Factorization Domain

Recall the Fundamental Theorem of Arithmetic, which states that every integer greater than 1 can be expressed as a product of prime numbers in an unique way up to the order of appearance of the primes. This property of integers can be generalised to an arbitrary integral domain.

DEFINITION. 1.45 let D be an integral domain. Two elements $a, b \in D$ are said to be *associates* of each other if $a \mid b$ and $b \mid a$.

THEOREM. 1.46 *Let D be an integral domain, $a, b \in D$. Then the followings are equivalent:*

1. *a and b are associates to each other.*
2. *There exists an unit $u \in D$ such that $a = ub$.*
3. *$\langle a \rangle = \langle b \rangle$.*

PROOF. $(1 \Rightarrow 2)$: Assume that a and b are associates of each other. Then there exist $c, d \in D$ such that $a = cb$ and $b = da$ and hence $a = cda$. Since D is an integral domain, by cancellation $cd = 1$ and hence c, d are units. Taking $u = c$ the result follows.

$(2 \Rightarrow 3)$: Assume the condition holds. Then $a = ub$ implies that $a \in \langle b \rangle$ which implies that $\langle a \rangle \subset \langle b \rangle$. Also since u is a unit, $b = u^{-1}a$ which implies that $\langle b \rangle \subset \langle a \rangle$. Hence $\langle a \rangle = \langle b \rangle$.

$(3 \Rightarrow 1)$: Assume $\langle a \rangle = \langle b \rangle$. Then $a \in \langle b \rangle \Rightarrow a = cb$ for some $c \in D$, hence $b \mid a$. Similarly, $\langle a \rangle = \langle b \rangle \Rightarrow b \in \langle a \rangle \Rightarrow b = da$ for some $d \in D$, hence $a \mid b$. Thus a, b are associates of each other. ■

DEFINITION. 1.47 A non-zero non-unit element a in a ring R is called a *prime* element if for $b, c \in R$, $a \mid bc$ implies that either $a \mid b$ or $a \mid c$.

Recall that in a commutative ring R an ideal I is called a *prime ideal* if for all $a, b \in R$, $ab \in I$ implies that either $a \in I$ or $b \in I$.

THEOREM. 1.48 *In a commutative ring R an element $a \in R$ is prime if and only if $\langle a \rangle$ is a prime ideal.*

PROOF. Assume that a is a prime. Let $x, y \in R$ such that $xy \in \langle a \rangle$. Then $xy = ca$ for some $c \in R$. Since $a \mid ca$ it follows that $a \mid xy$. a being prime we have either $a \mid x$ or $a \mid y$ which shows that either $x \in \langle a \rangle$ or $y \in \langle a \rangle$. Thus $\langle a \rangle$ is a prime ideal.

Conversely, assume that $\langle a \rangle$ is a non-trivial prime ideal in R. Since a is a non-unit, $\langle a \rangle \neq R$. Let $x, y \in R$ such that $a \mid xy$. Then there exists $c \in R$ such that $xy = ca$ which shows that $xy \in \langle a \rangle$. Since $\langle a \rangle$ is a prime ideal, either $x \in \langle a \rangle$ or $y \in \langle a \rangle$ which implies that either $a \mid x$ or $a \mid y$. Hence a is prime. ■

DEFINITION. 1.49 A non-zero non-unit element a in a ring R is called an *irreducible element* if $a = bc$ then either b or c is a unit in R, i.e., a can not be written as a product of two non-units. An element that is not irreducible is called a *reducible* element.

Thus the only divisors of an irreducible element are its associates or units.

THEOREM. 1.50 *In a ring with unity a prime element is always irreducible.*

PROOF. Let p be a prime in a ring R. Suppose that $p = ab$ then $p \mid ab$ and hence either $p \mid a$ or $p \mid b$. If $p \mid a$ then $a = pc$ for some $c \in R$, thus $p = ab = pcb$ showing that $cb = 1$ which shows that $c = b^{-1}$, ie.e, b is a unit. Similarly if $p \mid b$ then we can show that a is a unit. Thus p is an irreducible. ■

Converse of the result is not in general true. However for a PID the converse is also true.

EXAMPLE. 1.51 Let $R = Z[i\sqrt{3}] = \{a+ib\sqrt{3} : a, b \in \mathbb{Z}\}$. We shall show that $\alpha = 1+i\sqrt{3}$ is not a prime but irreducible.

Since $(1 + i\sqrt{3})(1 - i\sqrt{3}) = 4$, it follows that $1 + i\sqrt{3} \mid 4$, but $4 = 2 \cdot 2$ and $1 + i\sqrt{3} \nmid 2$ which shows that $1 + i\sqrt{3}$ is not a prime.

Let $1 + i\sqrt{3}$ be factorized as $1 + i\sqrt{3} = (a + ib\sqrt{3})(c + id\sqrt{3})$. Then taking square of modulus, $4 = (a^2 + 3b^2)(c^2 + 3d^2)$. Then possible value of $a^2 + 3b^2$ can be 1 or 4 because for any integers a, b, $a^2 + 3b^2 = 2$ is not possible. If $a^2 + 3b^2 = 1$ then $a = \pm 1, b = 0$ and hence $a+ib\sqrt{3}$ is a unit. If $a^2+3b^2 = 4$ then $c^2+3d^2 = 1$ which is only possible if $c = \pm 1$ and $d = 0$, i.e., $c + id\sqrt{3}$ is a unit. Hence $1 + i\sqrt{3}$ is irreducible.

THEOREM. 1.52 *If D is a PID then an irreducible element is prime.*

PROOF. Let $a \in D$ be an irreducible element. Then a is a non-unit and hence $\langle a \rangle \neq D$. By Zorn's lemma the ideal $\langle a \rangle$ is contained in a maximal ideal M. Since D is a PID there is $p \in D$ such that $M = \langle p \rangle$. Thus $\langle a \rangle \subset \langle p \rangle$ which implies that $a \in \langle p \rangle$. So there exists $c \in D$ such that $a = pc$. Since a is irreducible, c must be a unit, i.e., a and p are associates. Thus $\langle a \rangle = \langle p \rangle = M$.

Thus $\langle a \rangle$ is a maximal ideal and hence a prime ideal. So a is prime. ■

REMARK. 1.53 In view of the above theorem and the preceding example it can be concluded that $\mathbb{Z}[\sqrt{-3}]$ is not a PID.

DEFINITION. 1.54 An integral domain D is called a *Factorization Domain* (FD) if every nonzero element $x \in D$ can be written uniquely as a product of a unit and some irreducible elements of D, i.e., $x = up_1p_2 \dots p_n$, $n \geq 0$, where u is a unite of D and $p_1, p_2, \dots, p_n$ are irreducible elements of D.

DEFINITION. 1.55 An integral domain D is called an *Unique Factorization Domain* (UFD) if

1. D is a Factorization Domain and

2. The factorization of each non zero element is unique in the sense that if $x \neq 0$ and $x = up_1p_2 \dots p_n = vq_1, q_2 \dots q_m$, where u, v are units, then $m = n$ and for every $1 \leq i \leq n$, p_i is an associate of some $q_j, 1 \leq j \leq n$.

EXAMPLE. 1.56 1. $\mathbb{Z}$ is an UFD as stated in the Fundamental Theorem of Arithmetic.

2. The set of all Gaussian Integers $Z[i] = \{a + ib : a, b \in \mathbb{Z}\}$ where $i = \sqrt{-1}$, is an example of UFD.

3. Let $D = \mathbb{Z}[\sqrt{-5}] = \{a + ib\sqrt{5} : a, b \in \mathbb{Z}\}$. Then D is an integral domain. Here $6 \in D$ $(a = 6, b = 0)$ can be factorized as $6 = (3 + 0\sqrt{-5})(2 + 0\sqrt{-5})$ and $6 = (1 + i\sqrt{5})(1 - i\sqrt{5})$, i.e., two different ways. It can be verified that 2, 3, $1 + i\sqrt{5}$ and $1 - i\sqrt{5}$ are all irreducibles. Hence D is not an UFD.

LEMMA. 1.57 *Let D be a PID. Then any strictly increasing chain of ideals $I_1 \subset I_2 \subset I_3 \subset \cdots$ in D is finite.*

PROOF. Let $I = \cup I_k$. Then since the ideals I_k form a chain, the union I is an ideal in D. Since D is a PID, there exists $a \in D$ such that $I = \langle a \rangle$. Now $a \in I = \cup I_k$, there exists I_n in the chain such that $a \in I_n$. Since every $I_k \subset I = I_n$ we have I_n is the last member of the chain. Hence the chain is finite. ■

THEOREM. 1.58 *Every PID is an UFD.*

PROOF. Let D be a PID and a_0 be a non-zero element of D. First we shall show that a_0 can be expressed as a product of a unit and some irreducible elements of D.

If a_0 is irreducible then it is done. Otherwise let $a_0 = ua_1b_1$ where u is a unit, $a_1 \neq 0$ and both a_1 and b_1 are non-units. If a_1 is not irreducible then we can write $a_1 = a_2b_2$ where $a_2 \neq 0$ and both of a_2, b_2 are non-units. Proceeding this way we get two sequences of non-units $a_1, a_2, \dots$ and $b_1, b_2, \dots$, where each a_i is non-zero, such that for all $i = 1, 2, \dots$, $a_i = a_{i+1}b_{i+1}$. Hence we get of chain of strictly increasing ideals $\langle a_0 \rangle \subset \langle a_1 \rangle \subset \langle a_2 \rangle \subset \cdots$. By the above lemma (Lemma 1.57) this chain must be finite. So the chain ends at $\langle a_k \rangle$ for some $k \in \mathbb{N}$. Hence the element a_k must be irreducible. Thus, the element a_0 has at least one irreducible factor.

Let $a_0 = up_1c_1$, where u is a unit, p_1 is irreducible and c_1 is a non-unit. If c_1 is not irreducible we can write it as $c_1 = p_2c_2$ where p_2 is irreducible and c_2 is a non-unit. Proceeding this way we get a strictly increasing chain $\langle a_0 \rangle \subset \langle p_1 \rangle \subset \langle p_2 \rangle \subset \cdots$ of ideals

in D. By the above lemma this chain is finite and hence there is m such that c_m is irreducible. Thus $a_0 = up_1p_2 \dots p_mc_m$ where u is a unit, $p_1, p_2, \dots, p_m, c_m$ are irreducibles. Hence D is a Factorisation domain.

To see the uniqueness, let a_0 be expressed as $a_0 = up_1p_2 \dots p_r = vq_1q_2 \dots q_s$, where u, v are units, $p_1, p_2, \dots, p_r$ and $q_1, q_1, \dots, q_s$ are irreducibles. If $r = 1$ then a_0 becomes irreducible and hence $s = 1$. Thus $a_0 = up_1 = v_q1$, i.e., p_1, q_1 are associates.

To use induction on r assume that the expression is unique up to associates for the products of fewer than r irreducible factors. Let

$$a_0 = up_1p_2 \dots p_r = vq_1q_2 \dots q_s.$$

Since D is a PID each p_i is a prime, in particular p_1 is a prime. Now, $up_1p_2 \dots p_r = vq_1q_2 \dots q_s$ implies that p_1 divides $q_1q_1 \dots q_s$ and primeness of p_1 says $p_1 \mid q_i$ for at least one q_i, say $p_1 \mid q_1$. So there exists a unit u_1 such that $q_1 = u_1p_1$, hence $uu_1p_1p_2 \dots p_r = vu_1q_1q_2 \dots q_s$, i.e., $uq_1p_2 \dots p_r = vu_1q_1q_2 \dots q_s$. By cancellation $up_2 \dots p_r = vu_1q_2 \dots q_s$. The left hand side is a product of $r-1$ irreducible factors, hence by induction hypothesis $s = r$ and each q_i is an associate of some p_i, $2 \leq i \leq r$. This completes the induction and hence D is an UFD. ■

COROLLARY. 1.59 *If F is a field then $F[x]$ is an UFD.*

PROOF. This follows immediately as $F[x]$ is a PID. ■

EXAMPLE. 1.60

1. Let D be an integral domain. Show that the relation $\sim$ on D defined by $a \sim b$ *if and only if a and b are associates* is an equivalence relation.

2. If D is an integral domain and $a, b \in D$, $b \neq 0$, then show that $\langle ab \rangle = \langle b \rangle$ if and only if a is a unit.

   Since $ab \in \langle b \rangle$ it immediately follows that $\langle ab \rangle \subset \langle a \rangle$. Assume that a is a unit. Then $b = a^{-1}(ab) \in \langle ab \rangle$ and hence $\langle b \rangle \subset \langle ab \rangle$. Thus $\langle b \rangle = \langle ab \rangle$.

   Conversely, if $\langle b \rangle = \langle ab \rangle$ then there exists $r \in D$ such that $b = rab$, by cancellation $ra = 1$, i.e., $a^{-1} = r$, hence a is a unit.

3. In an integral domain D show that the product of an irreducible and a unit is an irreducible.

   Let a be an irreducible and u be a unit in D. If $ua = cd$, where c, d are non-units, then $a = (u^{-1}c)d$, both $u^{-1}c$ and d are non-units — contradicting that a is irreducible.

1.6 Euclidean Domain

DEFINITION. 1.61 An integral domain D is called an *Euclidean Domain* if there exists a map $d : D^* \to \mathbb{Z}_+$ satisfying conditions:

1. For $a, b \in D^*, d(a) \leq d(ab)$
2. For $a, b \in D$, $a \neq 0$ there exist $q, r \in D$ such that $b = aq + r$ where either $r = 0$ or $d(r) < d(a)$.

The property 2 is called the *division algorithm* and the map d is called the *algorithm map* or *measure*. (Here D^* denotes the set of all non-zero elements of D). The element b is called the dividend, a is called the divisor, q is called the quotient and r is called the remainder.

EXAMPLE. 1.62
1. $\mathbb{Z}$ is the simplest example of Euclidean Domain. Here for $n \in \mathbb{Z}$, $d(n) = |n|$, the modulus of n.
2. Another example of Euclidean Domain is $F[x]$, where F is a field. For $f(x) \in F[x]$, $d(f(x)) = \deg f(x)$.
3. The Gaussian integers $\mathbb{Z}[i] = \{m + in : m, n \in \mathbb{Z}\}$ is another example of Euclidean domain. Here for $a = m + in$, $m \neq 0 \neq n$, $d(a) = a\bar{a} = m^2 + n^2$. To see the division algorithm take $a, b \in \mathbb{Z}[i]$ with $a \neq 0$. Let $z = ba^{-1}$. Then $z \in \mathbb{Q}[i]$, let $z = \frac{p}{q} + \frac{s}{t}i$. Choose the integers m, n such that $m \leq \frac{p}{q} < m+1$ and $n \leq \frac{s}{t} < n+1$. Consider the square with the vertices $m + in, (m + 1) + in, m + i(n + 1)$ and $(m + 1) + i(n + 1)$. Take q is the number among the four such that $|z - q| < 1$. Let $r = b - qa$ so that $b = qa + r$. Now

$$\frac{b(r)}{b(a)} = \frac{r\bar{r}}{a\bar{a}} = \frac{|r|^2}{|a|^2} = \left|\frac{r}{a}\right|^2 = \left|\frac{b - qa}{a}\right|^2 = \left|\frac{b}{a} - q\right|^2 = |z - q|^2 < 1.$$

This shows that $b(r) < b(a)$.

THEOREM. 1.63 *In an Euclidean domain D the set of all the units is the set*

$$\{a \in D^* : d(a) = d(1)\}.$$

PROOF. The set $d(D^*)$ is a subset of $\mathbb{Z}_+$ and hence have a smallest element. Let $d(e)$ be the smallest element where $e \in D^*$. By the property of d we have $d(1) \leq d(1.e) = d(e)$,

also $d(e)$ is the smallest of all the values taken by d, i.e., $d(e) \leq d(x)$ for all $x \in D^*$, in particular $d(e) \leq d(1)$. Hence $d(e) = d(1)$.

If x is a unit in D then $d(x) \leq d(xx^{-1}) = d(1)$. Since $d(1)$ is the minimum value we have $d(x) = d(1)$. Conversely, if $x \in D$ such that $d(x) = d(1)$ then by division algorithm there is $q, r \in D$ such that $1 = qx + r$, where either $r = 0$ or $d(r) < d(x)$. But $d(x) = d(1)$ being the minimum value $d(r) < d(x)$ is not possible. Hence $r = 0$. Thus $1 = qx$ showing that x is a unit and $x^{-1} = q$. ■

THEOREM. 1.64 *Every Euclidean domain is a principal ideal domain.*

PROOF. Let D be an Euclidean domain, I be an ideal in D. Then $d(I) = \{d(a) : a \in I\}$ is a subset of $\mathbb{Z}_+$ and hence has a smallest member. Let us choose $a \in I$ such that $d(a)$ is the minimum value in $d(I)$. To show that $I = \langle a \rangle$ take any $b \in I$. By division algorithm there exist $q, r \in D$ such that $b = qa + r$ where either $r = 0$ or $d(r) < d(a)$. Since $r = b - qa \in I$ and $d(a)$ is the minimum value in $d(I)$ it follows that $d(r) < d(a)$ is not possible. Hence $r = 0$. Thus $b = qa$ which shows that $I = \langle a \rangle$. So D is PID. ■

COROLLARY. 1.65 *Every Euclidean domain is an UFD.*

PROOF. This follows immediately as every Euclidean domain ia a PID and PID is an UFD. (Theorem 2.24) ■

EXAMPLE. 1.66 1. Show that in $\mathbb{Z}[i]$, for all $x, y \in D^*$, $d(xy) = d(x)d(y)$.

Let $x = a + ib$ and $y = c + id$, $a, b, c, d \in \mathbb{Z}$. Then $d(x) = a^2 + b^2$ and $d(y) = c^2 + d^2$. Thus $d(x)d(y) = (a^2 + b^2)(c^2 + d^2) = a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2$.

On the other hand $xy = (a + ib)(c + id) = (ac - bd) + i(ad + bc)$ and $d(xy) = (ac - bd)^2 + (ad + bc)^2 = a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2$. Hence the result.

2. In an Euclidean domain D if a and b are associates then prove that $d(a) = d(b)$.

   If a, b are associates then there exists a unit u such that $a = bu$. So $d(a) = d(bu) \leq d(b)d(u) = d(b)$, since $d(u) = 1$, u being a unit. Similarly, $b = au^{-1}$ shows that $d(b) \leq d(a)$. Thus $d(a) = d(b)$.

3. Find q and r in $\mathbb{Z}[i]$ when $3 - 4i$ is divided by $2 + 5i$.

   To find q and r such that $3 - 4i = (2 + 5i)q + r$. Let us consider

$$z = \frac{3 - 4i}{2 + 5i} = \frac{(3 - 4i)(2 - 5i)}{(2 + 5i)(2 - 5i)} = \frac{-14 - 23i}{29} = -\frac{14}{29} - \frac{23}{29}i.$$

Now, $-1 < -\frac{14}{29} < 0$ and $-1 < -\frac{23}{29} < 0$. Here $-\frac{14}{29}$ is nearer to 0 and $-\frac{23}{29}$ is nearer to -1, hence we can take $q = 0 - i = -i$. Consequently $r = b - aq = (3-4i) - (2+5i)(-i) = -2-2i$.

Thus $3 - 4i = (2+5i)(-i) + (-2-2i)$.

4. Show that $1 - i$ is irreducible in $\mathbb{Z}[i]$.

   Let $1 - i = xy$ where $x, y \in \mathbb{Z}[i]$ are non-units. Then $d(1-i) = 2 = d(x)d(y)$. Then either $d(x) = 1$ or $d(x) = 2$. $d(x) = 1$ shows that x is a unit and $d(x) = 2$ implies that $d(y) = 1$, i.e., y is a unit — both lead to contradiction.

5. In $\mathbb{Z}[\sqrt{-5}]$, show that 21 does not factor uniquely as a product of irreducibles.

   Note that $21 = 3 \cdot 7 = (3 + 0\sqrt{-5})(7 + \sqrt{-5})$. To show that 3 is irreducible, let $3 = xy$ where $x, y \in \mathbb{Z}[\sqrt{-5}]$ are non-units. Then $d(3) = d(xy) = d(x)d(y)$, or, $9 = d(x)d(y)$. Since x, y are non-units, $d(x) \neq 1 \neq d(y)$ and hence $d(x) = d(y) = 3$. If $x = a + b\sqrt{-5}$ then $d(x) = a^2 + 5b^2$. Since there is no integers a, b such that $a^2 + 5b^2 = 3$ $d(x) = 3$ is not possible. Hence 3 is irreducible. Similarly 7 is irreducible.

   Also $21 = (1+2\sqrt{-5})(1-2\sqrt{-5})$. If $1+2\sqrt{-5} = xy$ where x, y are irreducibles then $d(1+2\sqrt{-5}) = d(xy) = d(x)d(y)$, or $21 = d(x)d(y)$. Since x, y are non-units, either $d(x) = 3, d(y) = 7$ or vice versa. But both $d(x) = 3$ or $d(x) = 7$ are impossible as there are no integers a, b such that $a^2 + 5b^2 = 3$ or $a^2 + 5b^2 = 7$.

   Thus $21 = 3 \cdot 7 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5})$ are two factorizations into irreducibles.

6. In $\mathbb{Z}[i]$ show that 2 and 5 are not irreducibles.

   $2 = (1+i)(1-i)$, both the factors are non-units. Also $2 = -i(1+i)^2$, here $1 + i$ is non-unit.

   $5 = (1+2i)(1-2i)$, both the factors are non-units. (The only units in $\mathbb{Z}[i]$ are ± 1 and $\pm i$.)

7. Prove that $\mathbb{Z}[\sqrt{-3}]$ is not a principal ideal domain.

   Referred to Remark 1.53.

8. In $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$, show that every element of the form $(3 + 2\sqrt{2})^n$ is a unit, where n is a positive integer.

   Note that $(3+2\sqrt{2})(3-2\sqrt{2}) = 1$. Hence $(3+2\sqrt{2})$ is a unit. Also $(3+2\sqrt{2})^n(3-2\sqrt{2})^n = [(3+2\sqrt{2})(3-2\sqrt{2})]^n = 1$. Hence the result follows.

9. Let D be a PID and p an irreducible element of D. Prove that $D/\langle p\rangle$ is a field.

   p being an irreducible element in the PID D it follows that $\langle p\rangle$ is a maximal ideal. Hence the result.

10. Let $n > 1$ be an integer which is not divisible by the square of a prime. Show that the only units of $\mathbb{Z}[\sqrt{-n}]$ are ± 1.

    If $x = a + ib\sqrt{n}$ is a unit then $d(x) = 1$, i.e., $a^2 + nb^2 = 1$. Since $n > 1$ this is only possible when $b = 0$ and $a = \pm 1$.

11. In $Z[\sqrt{-7}]$, show that $d(6 + 2\sqrt{-7}) = d(1 + 3\sqrt{-7})$ but $6 + 2\sqrt{-7}$ and $1 + 3\sqrt{-7}$ are not associates.

    $d(6 + 2\sqrt{-7}) = 36 + 4 \times 7 = 64$ and $d(1 + 3\sqrt{-7}) = 1 + 9 \times 7 = 64$. Since $7 > 1$ and 7 is not divisible by square of a prime the only units in $Z[\sqrt{-7}]$ are 1 and -1. Hence the elements can not be associates.

2 Dual Spaces

2.1 Review of the previous study

In what follows we recall some definitions and elementary results, without proof or with a sketchy proof, from the part Linear Algebra already studied in a previous course.

DEFINITION. 2.1 A set V is called a *vector space* over a field F if a binary operation $+$, called *vector addition*, on V and an operation $F \times V \to V$, called *scalar multiplication*, are defined which satisfy the following axioms:

1. $(V, +)$ is a commutative group, the identity element is called the *null vector* and is usually denoted by θ.
2. For every scalar $x \in F$ for every vector $v \in V$ the scalar product $xv \in V$ satisfies the following properties:
   (a) For $x, y \in F, v \in V$, $(x + y)v = xv + yv$,
   (b) For $x, y \in F, v \in V$, $(xy)v = x(yv)$,
   (c) For $x \in F, v, w \in V$, $x(v + w) = xv + xw$,
   (d) For all $v \in V$, $1v = v$, where 1 is the unity element of F.

The following results immediately follow from the definition:

THEOREM. 2.2 *Let V be a vector space over a field F. Then*

1. *$x\theta = \theta$ for every scalar x.*
2. *$0v = \theta$ for every vector v.*
3. *If $xv = \theta$, where $x \in F, v \in V$, then either $x = 0$ or $v = \theta$.*
4. *For $x \in F, v \in V$, $(-x)v = x(-v) = -xv$.*
5. *For $x \in F, v, w \in V$, $x(v - w) = xv - xw$.*
6. *For $x, y \in F, v \in V$, $(x - y)v = xv - yv$.*

DEFINITION. 2.3 Let V be a vector space over a field F. A vector v is called a *linear combination* of the vectors $v_1, v_2, \ldots, v_n \in V$ if there exist scalars $x_1, x_2, \ldots, x_n \in F$ such that $v = x_1v_1 + x_2v_2 + \cdots + x_nv_n = \sum_{k=1}^{n} x_kv_k$.

DEFINITION. 2.4 Let V be a vector space over a field F. A subset $W \subset V$ is called a *vector subspace* or *linear subspace* of V if W is itself a vector space over the same field F.

THEOREM. 2.5 *If V is a vector space over a field F then $W \subset V$ is a vector subspace of V if and only if (i) $\theta \in W$ and (ii) for every $x, y \in F$, for every $v, w \in W$, $xv + yw \in W$.*

DEFINITION. 2.6 Let V be a vector space over a field F and S be a non-empty subset of V. Then the set of all the possible linear combinations of the members of S is called the *linear span* of S and is denoted by $L(S)$. Thus,

$$L(S) = \{x_1v_1 + x_2v_2 + \cdots + x_nv_n : x_1, x_2, \ldots, x_n \in F, v_1, v_2, \ldots, v_n \in V, n \in \mathbb{N}\}$$

THEOREM. 2.7 *For a set $S \subset V$, where V is a vector space over a field F, $L(S)$ is a vector space. It is the smallest, with respect to set inclusion, vector subspace of V containing S.*

DEFINITION. 2.8 Let V be a vector space over a field F. A finite set of vectors $v_1, v_2, \ldots, v_n \in V$ is called *Linearly independent* if for any set of scalars $x_1, x_2, \ldots x_n \in F$, $x_1v_1 + x_2v_2 + \cdots + x_nv_n = \theta$ implies that $x_1 = x_2 = \cdots x_n = 0$. An infinite set of vectors is called linearly independent if every finite subset of it is linearly independent. A set which is not linearly independent is called *Linearly dependent*.

PROPOSITION. 2.9 *If a set of vectors contains the null vector θ then the set is linearly dependent.*

DEFINITION. 2.10 A set $B \subset V$ of a vector space over a field F is called a basis for V if (i) B spans V, i.e., $L(B) = V$ and (ii) B is linearly independent. If B contains only finite number of elements then V is called a *finite dimensional vector space* otherwise V is called *infinite dimensional*.

PROPOSITION. 2.11 *In a finite dimensional vector space the number of element of any basis is the same.*

DEFINITION. 2.12 The number of elements in a basis of a finite dimensional vector space V is called the dimension of V and is denoted by $\dim V$.

EXAMPLE. 2.13 1. Consider the field $\mathbb{R}$ and the set $\mathbb{R}^3 = \{(x, y, z) : x, y, z \in \mathbb{R}\}$. Then $\mathbb{R}^3$ is a vector space over $\mathbb{R}$. The set $B = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ is a basis

of $\mathbb{R}^3$, called the standard basis of $\mathbb{R}^3$. Another basis of $\mathbb{R}^3$ is $\{(1,1,0),(1,0,1),(0,1,1)\}$, note that both the basis have three elements and dimension of $\mathbb{R}^3$ is 3.

A subspace of $\mathbb{R}^3$ is $W = \{(x,y,0) : x,y \in \mathbb{R}\}$. A basis for W is $\{(1,0,0),(0,1,0)\}$ hence $\dim W = 2$. The set $V = \{(x,y,z) : x + 2y - 4z = 0\}$ is another example of a subspace of $\mathbb{R}^3$. Here it can be verified that $\{(0,2,1),(4,0,1)$ form a basis for V, this is also a subspace of dimension 2. Take $U = \{(t,2t,4t) : t \in \mathbb{R}\}$. Here $\{(4,-2,1)\}$ is a basis for U and hence U is a subspace of $\mathbb{R}^3$ with dimension 1.

2. Consider $P(x) = \mathbb{R}[x]$ the set of all the polynomials with coefficients from $\mathbb{R}$. Then $P(x)$ is a vector space over $\mathbb{R}$. A basis for $P(x)$ is $\{1, x, x^2, x^3, \ldots\}$, hence $P(x)$ is an infinite dimensional vector space. For a given positive integer n let $P_n(x)$ denote the set of all polynomials over $\mathbb{R}$ with degree less than or equal to n. Then $P_n(x)$ is a subspace of $P(x)$. A basis of $P_n(x)$ is $\{1, x, x^2, \ldots, x^n\}$ which have $n+1$ elements and hence $\dim P_n(x) = n + 1$.

3. It can be noted that if F is a field then F itself is a vector space over itself, any non-zero element of it form a basis for it. Obviously the dimension of this vector space is 1.

DEFINITION. 2.14 Let V, W be two vector spaces over the same field F. A function $T : V \to W$ is called a *linear transformation* if it satisfies the following axioms:

1. For all $v, w \in V$, $T(v + w) = T(v) + T(w)$.
2. For all $c \in F$, for all $v \in V$, $T(cv) = cT(v)$.

If T is one-one and onto then it is called an *isomorphism.*

It can be noted that in this definition in the condition 1 the + sign in left hand side is the addition in V whereas in right hand side the + sign denotes the addition in W. In the same manner we can say that in condition 2, since $v \in V$, cv is the scalar multiplication in V whereas in the right hand side since $T(v) \in W$, $cT(v)$ is the scalar multiplication in W.

THEOREM. 2.15 *A linear transformation $T : V \to W$ is uniquely determined by the effects of T on any basis of V.*

The above theorem states that if B is a basis of V and if $T(u)$ is knows for all $u \in B$ then T can be determined for any $v \in V$. For $v \in V$ there are basis vectors $u_1, u_2, \ldots, u_n \in B$ and scalars $x_1, x_2, \ldots, x_n \in F$ such that $v = x_1u_1 + x_2u_2 + \cdots + x_nu_n$, thus by linearity $T(v) = x_1T(u_1) + x_2T(u_2) + \cdots + x_nT(u_n)$.

THEOREM. 2.16 *If a vector space V over a field F is finite dimensional and* $\dim V = n$ *then V is isomorphic to F^n.*

PROOF. Let $B = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis for V. For $v \in V$ there are scalars $a_1, a_2, \ldots, a_n \in F$ such that $v = a_1v_1 + a_2v_2 + \cdots + a_nv_n$. Then $(a_1, a_2, \ldots, a_n) \in F^n$ and define $T(v) = (a_1, a_2, \ldots, a_n)$ for all $v \in V$. This T is the required isomorphism from V onto F^n can be proved easily. ■

THEOREM. 2.17 *Let V, W be vector spaces over the same field F, $\{v_1, v_2, \ldots, v_n\}$ be a basis for V. Then for any set of n vectors $\{w_1, w_2, \ldots, w_n\}$ in W there exists a unique linear transformation T such that $T(v_i) = w_i$ for $i = 1, 2, \ldots, n$.*

PROOF. For an arbitrary $v \in V$, since $\{v_1, v_2, \ldots, v_n\}$ is a basis for V there exist scalars $c_1, c_2, \ldots, c_n \in F$ such that $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$. Define $T(v) = c_1w_1 + c_2w_2 + \cdots + c_nw_n$. Then T is a linear transformation with the required property. Uniqueness of T can also be proved easily. ■

THEOREM. 2.18 *Let $T : V \to W$ be a linear transformation where V, W are vector spaces over the same field F. Then $T(V) = \{T(v) : v \in V\}$ is a subspace of W and $\{v \in V : T(v) = \theta_W\}$ is a subspace of V, where θ_W is the null vector in W..*

DEFINITION. 2.19 For a linear transformation $T : V \to W$ the subspace $\{v \in V : T(v) = \theta_W\}$ of V is called the *null space* of the linear transformation T and the subspace $T(V) = \{T(v) : v \in V\}$ is called the *range space* of T. If V is finite dimensional then the dimension of the null space is called the *nullity* of T, written as nullity(T) and the dimension of the range space is called the *rank* of T, written as rank(T).

THEOREM. 2.20 *If V, W are vector spaces with V is finite dimensional then for any linear transformation $T : V \to W$,* $\text{rank}(T) + \text{nullity}(T) = \dim V$.

If V, W are finite dimensional vector spaces over a field F then for given ordered bases $B = \{v_1, v_2, \ldots, v_n\}$ of V and $C = \{w_1, w_2, \ldots, w_m\}$ of W any linear transformation $T : V \to W$ can be represented by a matrix of order $n \times m$ over F. The process is given below.

For $v \in V$ there are scalars $c_1, c_2, \ldots, c_n \in F$ such that $v = \sum_{j=1}^{n} c_jv_j$. The ordered n-tuple $(c_1, c_2, \ldots, c_n)$ is called the coordinates of v with respect to the basis B. Similarly for every vector $w \in W$ there are scalars $d_1, d_2, \ldots d_m \in F$ such that $w = \sum_{i=1}^{m} d_iw_i$. The ordered m-tuple $(d_1, d_2, \ldots, d_m)$ is the coordinates of w with respect to the ordered basis

C. Now for each $v_j \in B$ let $T(v_j) = a_{1j}w_1 + a_{2j}w_2 + \cdots + a_{mj}w_m = \sum_{i=1}^{m} a_{ij}w_i$. Thus for $v = \sum_{j=1}^{n} c_j v_j \in V$,

$$\begin{aligned} T(v) &= T(c_1v_1 + c_2v_2 + \cdots + c_nv_n) = c_1T(v_1) + c_2T(v_2) + \cdots + c_nT(v_n) \\ &= \sum_{j=1}^{n} c_jT(v_j) = \sum_{j=1}^{n} c_j(a_{1j}w_1 + a_{2j}w_2 + \cdots + a_{mj}w_m) \\ &= \sum_{j=1}^{n} c_j \sum_{i=1}^{m} a_{ij}w_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} c_j a_{ij} \right) w_i. \end{aligned}$$

Hence for $1 \leq i \leq m$ the i-th coordinate of $T(v)$ with respect to the basis C is $\sum_{j=1}^{n} a_{ij}c_j$.

If we set $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ and denote the vector v as the column matrix $v = \begin{bmatrix} c_1 & c_2 & \ldots & c_n \end{bmatrix}^T$ of its coordinates with respect to the basis B, then

$$\begin{aligned} Av &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} a_{11}c_1 + a_{12}c_2 + \cdots + a_{1n}c_n \\ a_{21}c_1 + a_{22}c_2 + \cdots + a_{2n}c_n \\ \vdots \\ a_{m1}c_1 + a_{m2}c_2 + \cdots + a_{mn}c_n \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^{n} a_{1j}c_j \\ \sum_{j=1}^{n} a_{2j}c_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}c_j \end{bmatrix} = \begin{bmatrix} d_1 & d_2 & \cdots & d_m \end{bmatrix}^T, \end{aligned}$$

where for $1 \leq i \leq m$, $d_i = \sum_{j=1}^{n} a_{ij}c_j$ is the i-th coordinate of $T(v)$ in W with respect to the basis C. Hence $T(v)$ is represented by the $m \times n$ matrix A over F.

It may be observed that the matrix A is the array of the column matrices $A_1, A_2, \ldots, A_n$, where for $1 \leq j \leq n$, A_j is the coordinates of $T(v_j)$ with respect to C.

EXAMPLE. 2.21 Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ is defined by $T(x, y) = (x + y, 2x - y, 3x + 5y)$ for all $(x, y) \in \mathbb{R}^2$. Show that T is a linear transformation and find its matrix representation with respect to the ordered bases $B = \{(1, 1), (0, -1)\}$ and $C = \{(1, 1, 1), (1, 0, 1), (0, 0, 1)\}$ of $\mathbb{R}^2$ and $\mathbb{R}^3$ respectively.

That verification of linearity of T is easy and left. Let $v_1 = (1, 1), v_2 = (0, -1)$ and $w_1 = (1, 1, 1), w_2 = (1, 0, 1), w_3 = (0, 0, 1)$ be the respective basis vectors.

Now, $T(v_1) = (2, 1, 8)$. We express $T(v_1)$ as a linear combination of w_1, w_2, w_3. Then $(2, 1, 8) = a_1(1, 1, 1) + a_2(1, 0, 1) + a_3(0, 0, 1) = (a_1 + a_2, a_1, a_1 + a_2 + a_3)$ which gives

$a_1 + a_2 = 2$, $a_1 = 1$ and $a_1 + a_2 + a_3 = 8$. Thus $a_1 = 1, a_2 = 1, a_3 = 6$. Hence the A_1 column matrix is $\begin{bmatrix}1 & 1 & 6\end{bmatrix}^T$.

$T(v_2) = (-1, 1, -5)$, we express it as a linear combination of w_1, w_2, w_3. Then $(-1, 1, -5) = b_1(1, 1, 1) + b_2(1, 0, 1) + b_3(0, 0, 1) = (b_1 + b_2, b_1, b_1 + b_3)$ which gives $b_1 + b_2 = -1$, $b_1 = 1$ and $b_1 + b_3 = -5$. Thus $b_1 = 1, b_2 = -2, b_3 = -4$. Hence the A_2 column matrix is $\begin{bmatrix}1 & -2 & -4\end{bmatrix}^T$.

Thus the matrix representation of T is $A = \begin{bmatrix}1 & 1\\ 1 & -2\\ 6 & -4\end{bmatrix}$.

2.2 Vector space of Linear Transformations

DEFINITION. 2.22 Let V and W be vector spaces over a field F. The set of all linear transformations from V to W is denoted by $\mathcal{L}(V, W)$. For $S, T \in \mathcal{L}(V, W)$, define $(S + T)(v) = S(v) + T(v)$ and for $c \in F$ define $(cT)(v) = cT(v)$ for all $v \in V$.

PROPOSITION. 2.23 *For $S, T \in \mathcal{L}(V, W)$ and $c \in F$, $S+T \in \mathcal{L}(V, W)$ and $cT \in \mathcal{L}(V, W)$. Hence $\mathcal{L}(V, W)$ is a vector space over the field F.*

PROOF. For $v, w \in V$ and $c \in F$

$$\begin{aligned}(S + T)(cv + w) &= S(cv + w) + T(cv + w)\\ &= cS(v) + S(w) + cT(v) + T(w)\\ &= c(S(v) + T(v)) + S(w) + T(w)\\ &= c(S + T)(v) + (S + T)(w).\end{aligned}$$

Hence $S + T$ is a linear transformation. Similar for cT. ■

THEOREM. 2.24 *If V, W are vector spaces of dimension n and m respectively then the dimension of $\mathcal{L}(V, W)$ is mn.*

PROOF. Let $B = \{v_1, v_2, \ldots, v_n\}$ be a basis for V and $C = \{w_1, w_2, \ldots, w_m\}$ be a basis for W. For each pair (i, j), $1 \le i \le n, 1 \le j \le m$ define $T_{ij} : V \to W$ as follows: For $v \in V$ there are scalars $c_1, c_2, \ldots, c_n \in F$ such that $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$, define $T_{ij}(v) = c_iw_j$. In particular, $T_{ij}(v_k) = 0$ if $i \neq k$ and $T_{ij}(v_k) = w_j$ if $i = k$. Thus, $T_{ij}(v_k) = \delta_{ik}w_j$, where δ_{ij} is the Kronecker delta.

First to show that T_{ij} is linear. Let $v, w \in V$ so that $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$ and

$w = d_1v_1 + d_2v_2 + \cdots + d_nv_n$, where $c_1, c_2 \ldots c_n, d_1, d_2, \ldots, d_n \in F$. Also take $c \in F$. Then

$$\begin{aligned} T_{ij}(cv + w) &= T_{ij}(c(c_1v_1 + c_2v_2 + \cdots + c_nv_n) + d_1v_1 + d_2v_2 + \cdots + d_nv_n) \\ &= T_{ij}((cc_1 + d_1)v_1 + (cc_2 + d_2)v_2 + \cdots + (cc_n + d_n)v_n) \\ &= (cc_i + d_i)w_j \;=\; cc_iw_j + d_iw_j \\ &= cT_{ij}(v) + T_{ij}(w). \end{aligned}$$

Hence T_{ij} is linear.

To show that $\{T_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$ form a basis for $\mathcal{L}(V, W)$ let us take $T \in \mathcal{L}(V, W)$ arbitrarily. For any $v_i \in B$, $1 \leq i \leq n$, since $T(v_i) \in W$ there are scalars $a_{i1}, a_{i2}, \ldots, a_{im} \in F$ such that for any $1 \leq i \leq n$

$$T(v_i) \;=\; a_{i1}w_1 + a_{i2}w_2 + \cdots + a_{im}w_m \;=\; \sum_{j=1}^{m} a_{ij}w_j$$

Consider the sum $T_0 = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}T_{ij}$. Then $T_0 \in \mathcal{L}(V, W)$. For any $1 \leq k \leq n$,

$$T_0(v_k) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}T_{ij}(v_k) \;=\; \sum_{j=1}^{m} a_{kj}w_j \quad (\text{since } T_{ij}(v_k) = \delta_{ik}w_j).$$

Hence $T(v_i) = T_0(v_i)$ for all $v_i \in B$. Since B is a basis for V we have $T = T_0$. Hence T can be expressed as a linear combination of members of $\{T_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$, i.e. $\mathcal{L}(V, W)$ is linear span of $\{T_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$.

To show that $\{T_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$ is linearly independent let us take scalars $c_{ij} \in F$, $1 \leq i \leq n, 1 \leq j \leq m$ such that $\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}T_{ij} = 0$. Applying this linear transformation to v_k, $1 \leq k \leq n$,

$$0 \;=\; \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}T_{ij}(v_k) \;=\; \sum_{j=1}^{m} c_{kj}w_j.$$

Since $w_1, w_2, \ldots, w_m$ are linearity independent we must have $c_{k1} = c_{k2} = \cdots = c_{km} = 0$. Since this is true for every $k = 1, 2, \ldots, n$ we have $c_{ij} = 0$ for all $i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$. Hence $\{T_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$ is linearly independent.

Thus the set $\{T_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$ containing mn elements form a basis for $\mathcal{L}(V, W)$ and hence the dimension of $\mathcal{L}(V, W)$ is mn. ■

2.3 Dual Space

Let V be a vector space over a field F. It is known that F is itself a vector space of dimension 1 over itself. A linear transformation from V to F is called a *linear functional* or a *linear form*. If V is of dimension n then from the last theorem of previous subsection it follows that $\mathcal{L}(V, F)$ is a vector space of dimension n. This vector space is called the Dual space of V and is denoted by V^*. We define it formally as:

DEFINITION. 2.25 If V is a vector space over a field F then the space of all the linear functionals on V is called the *dual space* of V and is denoted by V^*. Thus $V^* = \mathcal{L}(V, F)$.

THEOREM. 2.26 *If V is a finite dimensional vector space then for every basis $B = \{v_1, v_2, \ldots, v_n\}$ of V there is a unique dual basis $\{f_1, f_2, \ldots, f_n\}$ satisfying $f_i(v_j) = \delta_{ij}$ for $1 \leq i, j \leq n$.*

PROOF. For each $i = 1, 2, \ldots, n$ define $f_i : V \to F$ by

$$\begin{aligned} f_i(v_j) &= 1, \quad \text{if } i = j \\ &= 0, \quad \text{if } i \neq j \end{aligned}$$

and extend f_i over V linearly, i.e., if $v = \sum_{j=1}^{n} c_j v_j$ then $f_i(v) = \sum_{j=1}^{n} c_j f_i(v_j) = c_i$.

To show that $\hat{B} = \{f_1, f_2, \ldots, f_n\}$ spans V^* take $f \in V^*$ arbitrarily. Let $f(v_i) = b_i$ for each $i = 1, 2, \ldots, n$. Then $b_1 f_1 + b_2 f_2 + \cdots + b_n f_n \in V^*$ and for each $i = 1, 2, \ldots, n$,

$$\begin{aligned} (b_1 f_1 + b_2 f_2 + \cdots + b_n f_n)(v_i) &= b_1 f_1(v_i) + b_2 f_2(v_i) + \cdots + b_n f_n(v_i) \\ &= b_i f_i(v_i) \;=\; b_i = f(v_i). \end{aligned}$$

Thus f agrees with $b_1 f_1 + b_2 f_2 + \cdots + b_n f_n$ in each member of B and hence $f = b_1 f_1 + b_2 f_2 + \cdots + b_n f_n$. Hence $\hat{B}$ spans V^*.

To show that $\hat{B}$ is linearly independent let us take $a_1, a_2, \ldots, a_n \in F$ such that $a_1 f_1 + a_2 f_2 + \cdots + a_n f_n = 0$, the zero function. Then for any $v_i \in B$,

$$\begin{aligned} 0 &= (a_1 f_1 + a_2 f_2 + \cdots + a_n f_n)(v_i) \;=\; a_1 f_1(v_i) + a_2 f_2(v_i) + \cdots + a_n f_n(v_i) \\ &= a_i f_i(v_i) = a_i. \end{aligned}$$

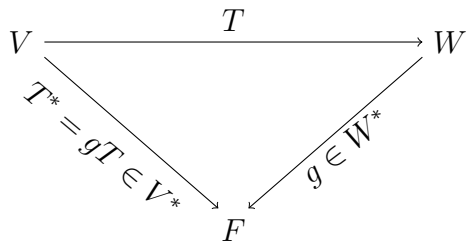
Thus $a_i = 0$ for all $i = 1, 2, \ldots, n$. Thus, $\hat{B}$ is a linearly independent set and hence $\hat{B}$ is a basis for V^*. ■

It can be noted that the functionals f_i are nothing but T_{ij} as defined in the proof of Theorem 2.24 where W is replaced by F.

It is convenient to write $\hat{v}_i$ for f_i as defined above. Thus if $B = \{v_1, v_2, \ldots, v_n\}$ is a basis for V then $\hat{B} = \{\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n\}$ is a basis for V^*, called the *dual basis* of B, where for each $i = 1, 2, \ldots, n$, $\hat{v}_i$ is defined by $\hat{v}_i(v_j) = \delta_{ij}$ on B and extended linearly over whole of V.

DEFINITION. 2.27 Let $T : V \to W$ be a linear transformation. Then $T^* : W^* \to V^*$, defined by $T^*(g) = gT$ for all $g \in W^*$, is called the *dual transformation* or *transpose* of T. Here gT denotes the composition of maps T and g.

The following figure describes the composition



The next theorem is stated without proof.

THEOREM. 2.28 *Let V be a finite dimensional vector space and $T : V \to W$ be a linear transformation. If $\mathcal{M}(T)$ is the matrix representation of T then the matrix representation of $T^* : W^* \to V^*$ is the transpose of T. Thus $\mathcal{M}(T^*) = \mathcal{M}(T)^t$.*

PROOF. Let us consider the following basis sets: $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ a basis of V and $\mathcal{D} = \{\beta_1, \beta_2, \ldots, \beta_m\}$ a basis of W. $\hat{\mathcal{B}} = \{\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_n\}$ and $\hat{\mathcal{D}} = \{\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_m\}$ are the dual basis of $\mathcal{B}$ and $\mathcal{D}$ respectively.

Let $\mathcal{M}(T) = A = (a_{ij})_{m\times n}$ and $\mathcal{M}(T^*) = B = (b_{ij})_{n\times m}$ with respect to the above basis sets. Then for $v = \sum_{i=1}^{n} c_i\alpha_i \in V$, the i-th coordinate of $T(v)$ in W is $\sum_{j=1}^{n} a_{ij}c_j$, $1 \le i \le m$. In particular, if $v = \alpha_j$, then the i-th coordinate of $T(\alpha_j)$ in W is a_{ij} and hence $T(\alpha_j) = \sum_{i=1}^{m} a_{ij}\beta_i$. Similar for T^*, i.e.,

$$\begin{aligned} T(\alpha_j) &= \sum_{i=1}^{m} a_{ij}\beta_i \text{ for } 1 \le j \le n \\ T^*(\hat{\beta}_j) &= \sum_{i=1}^{n} b_{ij}\hat{\alpha}_i \text{ for } 1 \le j \le m. \end{aligned}$$

Also from the definition of T^*, for $1 \leq i \leq n$ and $1 \leq j \leq m$,

$$T^*(\hat{\beta}_j)(\alpha_i) \;=\; \hat{\beta}_j(T(\alpha_i)) \;=\; \hat{\beta}_j\left(\sum_{k=1}^{m} a_{ki}\beta_k\right) \;=\; \sum_{k=1}^{m} a_{ki}\hat{\beta}_j(\beta_k) \;=\; a_{ji}.$$

For any $f \in V^*$, $f = \sum_{i=1}^{n} f(\alpha_i)\hat{\alpha}_i$, hence in particular $T^*(\hat{\beta}_j) = \sum_{i=1}^{n} T^*(\hat{\beta}_j)(\alpha_i)\hat{\alpha}_i = \sum_{i=1}^{n} a_{ji}\hat{\alpha}_i$. But we have $T^*(\hat{\beta}_j) = \sum_{i=1}^{n} b_{ij}\hat{\alpha}_i$. Since $\{\hat{\beta}_1, \hat{\beta}_1, \ldots, \hat{\beta}_m\}$ form a basis of W^* and a linear map is uniquely determined by its effect on basis vectors, we have $a_{ji} = b_{ij}$ for $1 \leq i \leq n, 1 \leq j \leq m$. Hence $B = A^t$. ■

DEFINITION. 2.29 Let V be a vector space and $S \subset V$. The annihilator of S, denoted by S^0, is a subset of V^* defined by $S^0 = \{f \in V^* : f(s) = 0 \; \forall s \in S\}$.

THEOREM. 2.30 *For a subspace S of V, S^0 is a subspace of V^*.*

PROOF. If $f, g \in S^0$ then $f(s) = g(s) = 0$ for all $s \in S$. So $(f+g)(s) = f(s) + g(s) = 0 + 0 = 0$ for all $s \in S$, hence $f + g \in S^0$. Also for $c \in F$, $(cf)(s) = cf(s) = c0 = 0$ for all $s \in S$, i.e., $cf \in S^0$. Hence S^0 is a subspace of V^*. ■

THEOREM. 2.31 *If $S \subset V$ then $(L(S))^0 = S^0$.*

PROOF. Since $S \subset L(S)$ if $f \in (L(S))^0$ then $f(s) = 0$ for all $s \in L(S)$ and hence in particular $f(s) = 0$, or all $s \in S$ i.e., $f \in S^0$. Thus $(L(S))^0 \subset S^0$.

Conversely, let $v \in L(S)$. Then there exist $s_1, s_2, \ldots, s_n \in S$ and scalars $c_1, c_2, \ldots, c_n$ such that $v = c_1 s_1 + c_2 s_2 + \cdots + c_n s_n$. Hence for $f \in S^0$ we have $f(v) = f(c_1 s_1 + c_2 s_2 + \cdots + c_n s_n) = c_1 f(s_1) + c_2 f(s_2) + \cdots + c_n f(s_n) = 0$, thus $f \in (L(S))^0$, hence $S^0 \subset (L(S))^0$. Combining these two we have $S^0 = (L(S))^0$. ■

THEOREM. 2.32 *Let V be a finite dimensional vector space over a field F. Then for a subspace S of V,* $\dim S + \dim S^0 = \dim V$.

PROOF. Let $\dim V = n$ and $\dim S = k < n$. Let $v_1, v_2, \ldots, v_k$ be a basis for S. We choose $v_{k+1}, v_{k+2}, \ldots, v_n$ so that $B = \{v_1, v_2, \ldots v_n\}$ is a basis for V. Let $\hat{B} = \{\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_n\}$ be the dual basis of B. Since for $i > k$ and $j \leq k$, $\hat{v}_i(v_j) = 0$ it follows that $\hat{v}_i \in S^0$ for all $i = k+1, k+2, \ldots, n$. Thus for each $v \in S$ since v is linear combination of $v_1, v_2, \ldots, v_k$ it follows that $\hat{v}_i(v) = 0$ for all $i > k$. Hence $\{\hat{v}_{k+1}, \hat{v}_{k+2}, \ldots, \hat{v}_n\} \subset S^0$. Also $\{\hat{v}_{k+1}, \hat{v}_{k+2}, \ldots, \hat{v}_n\}$ are linearly independent.

Let $f \in V^*$. Then there are scalars $c_1, c_2, \ldots, c_n$ such that $f = \sum_{i=1}^{n} c_i \hat{v}_i$. For each $j = 1, 2, \ldots, n$ we have $f(v_j) = \sum_{i=1}^{n} c_i \hat{v}_i(v_j) = c_j$ since $\hat{v}_i(v_j) = \delta_{ij}$. Hence $f = \sum_{i=1}^{n} f(v_i)\hat{v}_i$.

In particular, if $f \in S^0$ then $f(v_i) = 0$ for all $i = 1, 2, \ldots, k$, thus $f = \sum_{i=k+1}^{n} f(v_i)\hat{v}_i$. This shows that S^0 is the linear span of $\{\hat{v}_{k+1}, \hat{v}_{k+2}, \ldots, \hat{v}_n\}$. Thus $\{\hat{v}_{k+1}, \hat{v}_{k+2}, \ldots, \hat{v}_n\}$ is a basis of S^0 and hence its dimension is $n - k$.

Hence $\dim S + \dim S^0 = k + (n - k) = n = \dim V$. ■

COROLLARY. 2.33 *If a vector space V is of dimension n any subspace S of V with dimension k is an intersection of $n - k$ hyperspaces of V.*

PROOF. Let $\{v_1, v_2, \ldots, v_k\}$ be a basis of S. Extend this to to a basis of V, say $\{v_1, v_2, \ldots,$-$v_k, \ldots, v_n\}$. If $v \in S$ then v can be expressed as $v = c_1v_1 + \cdots + c_kv_k$, hence for $i > k$, $\hat{v}_i(v) = \hat{v}_i(c_1v_1 + \cdots + c_kv_k) = c_1\hat{v}_i(v_1) + \cdots + c_k\hat{v}_i(v_k) = 0$. On the other hand if for $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$, and for $i > k$, $\hat{v}_i(v) = 0$ then $\hat{v}_i(c_1v_1 + c_2v_2 + \cdots + c_nv_n) = c_1\hat{v}_i(v_1) + c_2\hat{v}_i(v_2) + \cdots + c_n\hat{v}_i(v_n) = 0$ we must have $c_i = 0$. Thus $v \in S$. Hence $v \in S$ if and only if $\hat{v}_i(v) = 0$ for all $i = k + 1, k + 2, \ldots, n$.

If for $k + 1 \leq i \leq n$, H_i denotes the null space of $\hat{v}_i$ then $\dim H_i = n - 1$ and hence a hyperspace of V. Thus $S = \bigcap_{i=k+1}^{n} H_i$ is an intersection of $n - k$ hyperspaces. ■

COROLLARY. 2.34 *For two subspaces S_1, S_2 of a finite dimensional vector space V, $S_1^0 = S_2^0$ if and only if $S_1 = S_2$.*

PROOF. If $S_1 = S_2$ then it immediately follows that $S_1^0 = S_2^0$. For the converse part, assume that $S_1 \neq S_2$. Then either $S_1 - S_2 \neq \emptyset$ or $S_2 - S_1 \neq \emptyset$. In the former case choose $v \in S_1 - S_2$, then there exists $f \in S_2^0$, $f(v) \neq 0$ but for all $f \in S_1^0$, $f(v) = 0$. Thus $S_1^0 \neq S_2^0$. Similarly if $S_2 - S_1 \neq \emptyset$ then $S_2^0 \neq S_1^0$. ■

EXAMPLE. 2.35 1. In $\mathbb{R}^3$ let $\alpha = (1, 0, 1), \beta = (0, 1, -2)$ and $\gamma = (-1, -1, 0)$, f is a linear functional on $\mathbb{R}^3$ such that $f(\alpha) = 1, f(\beta) = -1$ and $f(\gamma) = 3$. If $v = (a, b, c)$ find $f(v)$.

First we express v as a linear combination of α, β, γ, i.e., $v = x\alpha + y\beta + z\gamma$ where $x, y, z \in \mathbb{R}$, in matrix form, $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & -2 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

To solve this we reduced the augmented matrix:

$$\begin{pmatrix} 1 & 0 & -1 & | & a \\ 0 & 1 & -1 & | & b \\ 1 & -2 & 0 & | & c \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -1 & | & a \\ 0 & 1 & -1 & | & b \\ 0 & -2 & 1 & | & c-a \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & 0 & -1 & | & a \\ 0 & 1 & -1 & | & b \\ 0 & 0 & -1 & | & c-a+2b \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -1 & | & a \\ 0 & 1 & -1 & | & b \\ 0 & 0 & 1 & | & a-2b-c \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & 0 & -1 & | & a \\ 0 & 1 & -1 & | & b \\ 0 & 0 & 1 & | & a-2b-c \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & | & 2a-2b-c \\ 0 & 1 & 0 & | & a-b-c \\ 0 & 0 & 1 & | & a-2b-c \end{pmatrix}.$$

Thus $x = 2a - 2b - c, y = a - b - c, z = a - 2b - c$. Therefore,

$$\begin{aligned} f(v) &= f(x\alpha + y\beta + z\gamma) = xf(\alpha) + yf(\beta) + zf(\gamma) \\ &= (2a - 2b - c) \cdot 1 + (a - b - c) \cdot -1 + (a - 2b - c) \cdot 3 \\ &= 4a - 7b - 3c. \end{aligned}$$

EXAMPLE. 2.36 1. Find the dual of the basis $\mathcal{B} = \{\alpha, \beta\}$ of $\mathbb{R}^2$, where $\alpha = (2, 1)$ and $\beta = (3, 1)$.

The dual basis is $\hat{\mathcal{B}} = \{\hat{\alpha}, \hat{\beta}\}$, where the linear functionals $\hat{\alpha}, \hat{\beta}$ are such that $\hat{\alpha}(\alpha) = 1, \hat{\alpha}(\beta) = 0$ and $\hat{\beta}(\alpha) = 0, \hat{\beta}(\beta) = 1$.

Any element of $\mathbb{R}^2$ is written as $(x, y) = a\alpha + b\beta = a(2, 1) + b(3, 1) = (2a + 3b, a + b)$. Thus $x = 2a + 3b, y = a + b$, solving for a, b, $a = 3y - x$ and $b = x - 2y$. Thus $(x, y) = (3y - x)\alpha + (x - 2y)\beta$.

$\hat{\alpha}(x, y) = \hat{\alpha}((3y - x)\alpha + (x - 2y)\beta) = (3y - x)\hat{\alpha}(\alpha) + (x - 2y)\hat{\alpha}(\beta) = 3y - x$ for all $(x, y) \in \mathbb{R}^2$. Similarly $\hat{\beta}(x, y) = x - 2y$ for all $(x, y) \in \mathbb{R}^2$.

We can do it more conveniently with the help of matrix. We write the linear combination as $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$. The augmented matrix is $\begin{pmatrix} 2 & 3 & | & x \\ 1 & 1 & | & y \end{pmatrix}$. Applying row reduction,

$$\begin{pmatrix} 2 & 3 & | & x \\ 1 & 1 & | & y \end{pmatrix} \to \begin{pmatrix} 2 & 3 & | & x \\ 0 & 1 & | & x-2y \end{pmatrix} \to \begin{pmatrix} 2 & 0 & | & -2x+6y \\ 0 & 1 & | & x-2y \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & 0 & | & -x+3y \\ 0 & 1 & | & x-2y \end{pmatrix}.$$

Thus, $a = 3y - x, b = x - 2y$. So, $\hat{\alpha}(x, y) = \hat{\alpha}((3y - x)\alpha + (x - 2y)\beta) = (3y - x)\hat{\alpha}(\alpha) + (x - 2y)\hat{\alpha}(\beta) = 3y - x$ and $\hat{\beta}(x, y) = x - 2y$ for all $(x, y) \in \mathbb{R}^2$.

2. Let $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ be a basis for $\mathbb{R}^3$ where $\alpha_1 = (1, 0, -1), \alpha_2 = (1, 1, 1)$ and $\alpha_3 = (2, 2, 0)$. Find the dual basis $\hat{\mathcal{B}}$.

   If $\hat{\mathcal{B}} = \{\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3\}$ then $\hat{\alpha}_i(\alpha_j) = \delta_{ij}$. Thus $\hat{\alpha}_1(\alpha_1) = 1, \hat{\alpha}_1(\alpha_2) = 0, \hat{\alpha}_1(\alpha_3) = 1$ and so on.

   Any $(x, y, z) \in \mathbb{R}^3$ can be written as a linear combination of $\alpha_1, \alpha_2, \alpha_3$ as $(x, y, z) = a_1\alpha_1 + a_2\alpha_2 + a_3\alpha_3$. In matrix notation, $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$. To solve for a_1, a_2, a_3 we apply row reduction on the augmented matrix

$$\begin{pmatrix} 1 & 1 & 2 & | & x \\ 0 & 1 & 2 & | & y \\ -1 & 1 & 0 & | & z \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 2 & | & x \\ 0 & 1 & 2 & | & y \\ 0 & 2 & 2 & | & x+z \end{pmatrix}$$

$$\to \quad \begin{pmatrix} 1 & 1 & 2 & | & x \\ 0 & 1 & 2 & | & y \\ 0 & 0 & -2 & | & x+z-2y \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & | & x-y \\ 0 & 1 & 2 & | & y \\ 0 & 0 & -2 & | & x+z-2y \end{pmatrix}$$

$$\to \quad \begin{pmatrix} 1 & 0 & 0 & | & x-y \\ 0 & 1 & 0 & | & x-y+z \\ 0 & 0 & -2 & | & x+z-2y \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & | & x-y \\ 0 & 1 & 0 & | & x-y+z \\ 0 & 0 & 1 & | & -\frac{x}{2}+y-\frac{z}{2} \end{pmatrix}.$$

   Thus $a_1 = x - y, a_2 = x - y + z$ and $a_3 = -\frac{x}{2} + y - \frac{z}{2}$ which gives $(x, y, z) = (x - y)\alpha_1 + (x - y + z)\alpha_2 + (-\frac{x}{2} + y - \frac{z}{2})\alpha_3$. Thus we get

$$\begin{aligned} \hat{\alpha}_1(x, y, z) &= \hat{\alpha}_1((x - y)\alpha_1 + (x - y + z)\alpha_2 + (-\frac{x}{2} + y - \frac{z}{2})\alpha_3) \\ &= (x - y)\hat{\alpha}_1(\alpha_1) + (x - y + z)\hat{\alpha}_1(\alpha_2) + (-\frac{x}{2} + y - \frac{z}{2})\hat{\alpha}_1(\alpha_3) \\ &= x - y \end{aligned}$$

$$\begin{aligned} \hat{\alpha}_2(x, y, z) &= \hat{\alpha}_2((x - y)\alpha_1 + (x - y + z)\alpha_2 + (-\frac{x}{2} + y - \frac{z}{2})\alpha_3) \\ &= (x - y)\hat{\alpha}_2(\alpha_1) + (x - y + z)\hat{\alpha}_2(\alpha_2) + (-\frac{x}{2} + y - \frac{z}{2})\hat{\alpha}_2(\alpha_3) \\ &= x - y + z \\ \hat{\alpha}_3(x, y, z) &= \hat{\alpha}_3((x - y)\alpha_1 + (x - y + z)\alpha_2 + (-\frac{x}{2} + y - \frac{z}{2})\alpha_3) \\ &= (x - y)\hat{\alpha}_3(\alpha_1) + (x - y + z)\hat{\alpha}_3(\alpha_2) + (-\frac{x}{2} + y - \frac{z}{2})\hat{\alpha}_3(\alpha_3) \\ &= -\frac{x}{2} + y - \frac{z}{2}. \end{aligned}$$

3. Find the dual basis of $\mathcal{B} = \{(1, -2, 3), (1, -1, 1), (2, -4, 7)\}$ in $\mathbb{R}^3$.

   Let $\alpha_1 = (1, -2, 3), \alpha_2 = (1, -1, 1), \alpha_3 = (2, -4, 7)$, then $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$. Any vector $v = (x, y, z) \in \mathbb{R}^3$ can be written as $v = c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3$, where $c_1, c_2, c_3 \in$

$\mathbb{R}$. Then $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ -2 & -1 & -4 \\ 3 & 1 & 7 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$. To solve for c_1, c_2, c_3 we reduce the augmented matrix by row operation:

$$\begin{pmatrix} 1 & 1 & 2 & | & x \\ -2 & -1 & -4 & | & y \\ 3 & 1 & 7 & | & z \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 2 & | & x \\ -2 & -1 & -4 & | & y \\ 0 & -1 & 1 & | & z+y-x \end{pmatrix}$$

$$\to \quad \begin{pmatrix} 1 & 1 & 2 & | & x \\ 0 & 1 & 0 & | & y+2x \\ 0 & -1 & 1 & | & z+y-x \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 3 & | & x+(z+y-x) \\ 0 & 1 & 0 & | & y+2x \\ 0 & -1 & 1 & | & -x+y+z \end{pmatrix}$$

$$\to \quad \begin{pmatrix} 1 & 0 & 3 & | & y+z \\ 0 & 1 & 0 & | & 2x+y \\ 0 & 0 & 1 & | & x+2y+z \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & | & -3x-5y-2z \\ 0 & 1 & 0 & | & 2x+y \\ 0 & 0 & 1 & | & x+2y+z \end{pmatrix}.$$

Thus, $c_1 = -3x - 5y - 2z, c_2 = 2x + y, c_3 = x + 2y + z$. Hence the dual basis $\hat{\mathcal{B}} = \{\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3\}$ is given by,

$$\begin{aligned} \hat{\alpha}_1(v) &= c_1\hat{\alpha}_1(\alpha_1) + c_2\hat{\alpha}_1(\alpha_2) + c_3\hat{\alpha}_1(\alpha_3) = c_1 = -3x - 5y - 2z \\ \hat{\alpha}_2(v) &= c_1\hat{\alpha}_2(\alpha_1) + c_2\hat{\alpha}_2(\alpha_2) + c_3\hat{\alpha}_2(\alpha_3) = c_2 = 2x + y \\ \hat{\alpha}_3(v) &= c_1\hat{\alpha}_3(\alpha_1) + c_2\hat{\alpha}_3(\alpha_2) + c_3\hat{\alpha}_3(\alpha_3) = c_3 = x + 2y + z. \end{aligned}$$

It is a good practice to verify whether there is any mistake in calculation.

$$\begin{aligned} \hat{\alpha}_1(\alpha_1) &= \hat{\alpha}_1(1,-2,3) = -3.1 - 5(-2) - 2.3 = 1 \\ \hat{\alpha}_1(\alpha_2) &= \hat{\alpha}_1(1,-1,1) = -3.1 - 5(-1) - 2.1 = 0 \\ \hat{\alpha}_1(\alpha_3) &= \hat{\alpha}_1(2,-4,7) = -3.2 - 5(-4) - 2.7 = 0 \\ \hat{\alpha}_2(\alpha_1) &= \hat{\alpha}_2(1,-2,3) = 2.1 + (-2) = 0 \\ \hat{\alpha}_2(\alpha_2) &= \hat{\alpha}_2(1,-1,1) = 2.1 + (-1) = 1 \\ \hat{\alpha}_2(\alpha_3) &= \hat{\alpha}_2(2,-4,7) = 2.2 + (-4) = 0 \\ \hat{\alpha}_3(\alpha_1) &= \hat{\alpha}_3(1,-2,3) = 1 + 2(-2) + 3 = 0 \\ \hat{\alpha}_3(\alpha_2) &= \hat{\alpha}_3(1,-1,1) = 1 + 2(-1) + 1 = 0 \\ \hat{\alpha}_3(\alpha_3) &= \hat{\alpha}_3(2,-4,7) = 2 + 2(-4) + 7 = 1. \end{aligned}$$

EXAMPLE. 2.37 If S is the subspace of $\mathbb{R}^4$ spanned by $\alpha_1 = (1, 2, -3, 4)$ and $\alpha_2 = (0, 1, 4, -1)$ find the annihilator S^0.

A linear functional on $\mathbb{R}^4$ is of the form $\phi(x, y, z, w) = ax+by+cz+dw$ for all $(x, y, z, w) \in \mathbb{R}^4$, where $a, b, c, d \in \mathbb{R}$.

Let $\phi \in S^0$. Then $\phi(x,y,z,w) = 0$ for all $(x,y,z,w) \in S$. If $\phi(x,y,z,w) = ax+by+cz+dw$ then since α_1, α_2 form a basis for S, $\phi(\alpha_1) = \phi(\alpha_2) = 0$, i.e.,

$$\begin{aligned}\phi(1,2,-3,4) &= a+2b-3c+4d = 0\\ \phi(0,1,4,-1) &= b+4c-d = 0.\end{aligned}$$

We have two equations of four unknowns a, b, c, d. The solutions contain two parameters. Let us take particular values $c = 1, d = 0$, then we have $a + 2b - 3 = 0$ and $b + 4 = 0$ which give $a = 11, b = -4$. This gives a member ϕ_1 of S^0, $\phi_1(x,y,z,w) = 11x - 4y + z$. Again taking $c = 0, d = 1$ we have $a + 2b + 4 = 0$ and $b - 1 = 0$ which give $b = 1, a = -6$. Thus another member of S^0 is $\phi_2(x,y,z,w) = -6x + y + z$. Since dimension of S^0 is 2 and ϕ_1, ϕ_2 are linearly independent, $\{\phi_1, \phi_2\}$ forms a basis for S^0.

EXAMPLE. 2.38 Let V be the vector space of all polynomial functions p from $\mathbb{R}$ to $\mathbb{R}$ have degree 2 or less : $p(x) = c_0 + c_1x + c_2x^2$. Define three linear functionals on V by

$$f_1(p) = \int_0^1 p(x)\ dx, \quad f_2(p) = \int_0^2 p(x)\ dx, \quad f_3(p) = \int_0^{-1} p(x)\ dx,$$

Show that $\{f_1, f_2, f_3\}$ is a basis for V^* by exhibiting the basis for V of which it is the dual.

Let p_1, p_2, p_3 be the members of V of which f_1, f_2, f_3 are dual. If $p(x) = c_0+c_1x+c_2x^2 \in V$ then for $a \in \mathbb{R}$, $\displaystyle\int_0^a p(x)\ dx = \int_0^a (c_0 + c_1x + c_2x^2)\ dx = c_0a + c_1\frac{a^2}{2} + c_2\frac{a^3}{3}$. Thus,

$$\begin{aligned}f_1(p) &= \int_0^1 p(x)\ dx = c_0 + \frac{1}{2}c_1 + \frac{1}{3}c_2\\ f_2(p) &= \int_0^2 p(x)\ dx = 2c_0 + \frac{2^2}{2}c_1 + \frac{2^3}{3}c_2 \ = \ 2c_0 + 2c_1 + \frac{8}{3}c_2\\ f_3(p) &= \int_0^{-1} p(x)\ dx = -c_0 + \frac{1}{2}c_1 - \frac{1}{3}c_2\end{aligned}$$

When $p = p_1$ we have $f_1(p_1) = 1, f_2(p_1) = 0, f_3(p_1) = 0$, i.e.,

$$c_0 + \frac{1}{2}c_1 + \frac{1}{3}c_2 = 1, \quad 2c_0 + 2c_1 + \frac{8}{3}c_2 = 0, \quad -c_0 + \frac{1}{2}c_1 - \frac{1}{3}c_2 = 0.$$

To get p_1 we need to solve for c_0, c_1, c_2. In matrix notation the system becomes,

$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 2 & 2 & \frac{8}{3} \\ -1 & \frac{1}{2} & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Similarly, for p_2 and p_3 respectively we need to solve the following systems, $\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 2 & 2 & \frac{8}{3} \\ -1 & \frac{1}{2} & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 2 & 2 & \frac{8}{3} \\ -1 & \frac{1}{2} & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

For the three systems the augmented matrix is written together and reduced as

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & | & 1 & 0 & 0 \\ 2 & 2 & \frac{8}{3} & | & 0 & 1 & 0 \\ -1 & \frac{1}{2} & -\frac{1}{3} & | & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & | & 1 & 0 & 0 \\ 2 & 2 & \frac{8}{3} & | & 0 & 1 & 0 \\ 0 & 1 & 0 & | & 1 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & | & 1 & 0 & 0 \\ 0 & 1 & 2 & | & -2 & 1 & 0 \\ 0 & 1 & 0 & | & 1 & 0 & 1 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & | & 1 & 0 & 0 \\ 0 & 1 & 2 & | & -2 & 1 & 0 \\ 0 & 0 & -2 & | & 3 & -1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & | & 1 & 0 & 0 \\ 0 & 1 & 2 & | & -2 & 1 & 0 \\ 0 & 0 & 1 & | & -\frac{3}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & -\frac{2}{3} & | & 2 & -\frac{1}{2} & 0 \\ 0 & 1 & 2 & | & -2 & 1 & 0 \\ 0 & 0 & 1 & | & -\frac{3}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & | & 1 & -\frac{1}{6} & -\frac{1}{3} \\ 0 & 1 & 0 & | & 1 & 0 & 1 \\ 0 & 0 & 1 & | & -\frac{3}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

Thus for p_1, $(c_0, c_1, c_2) = (1, 1, -\frac{3}{2})$, i.e., $p_1(x) = 1 + x - \frac{3}{2}x^2$. Similarly, for p_2, $(c_0, c_1, c_2) = (-\frac{1}{6}, 0, \frac{1}{2})$, i.e., $p_2(x) = -\frac{1}{6} + \frac{1}{2}x^2$ and for p_3, $(c_0, c_1, c_2) = (-\frac{1}{3}, 1, -\frac{1}{2})$, i.e., $p_3(x) = -\frac{1}{3} + x - \frac{1}{2}x^2$. We verify the calculations as follows:

$$\begin{aligned}
f_1(p_1) &= \int_0^1 p_1(x)\,dx = \int_0^1 (1 + x - \frac{3}{2}x^2)\,dx = \left[x + \frac{x^2}{2} - \frac{3}{2}\frac{x^3}{3}\right]_0^1 = \frac{3}{2} - \frac{1}{2} = 1 \\
f_2(p_1) &= \int_0^2 p_1(x)\,dx = \int_0^2 (1 + x - \frac{3}{2}x^2)\,dx = \left[x + \frac{x^2}{2} - \frac{x^3}{2}\right]_0^2 = 0 \\
f_3(p_1) &= \int_0^{-1} p_1(x)\,dx = \int_0^{-1} 1(1 + x - \frac{3}{2}x^2)\,dx = \left[x + \frac{x^2}{2} - \frac{x^3}{2}\right]_0^{-1} = 0 \\
f_1(p_2) &= \int_0^1 p_2(x)\,dx = \int_0^1 (-\frac{1}{6} + \frac{1}{2}x^2)\,dx = \left[-\frac{1}{6}x + \frac{1}{2}\frac{x^3}{3}\right]_0^1 = 0 \\
f_2(p_2) &= \int_0^2 p_2(x)\,dx = \int_0^2 (-\frac{1}{6} + \frac{1}{2}x^2)\,dx = \left[-\frac{1}{6}x + \frac{x^3}{6}\right]_0^2 = 1 \\
f_3(p_2) &= \int_0^{-1} p_2(x)\,dx = \int_0^{-1} (-\frac{1}{6} + \frac{1}{2}x^2)\,dx = \left[-\frac{1}{6}x + \frac{x^3}{6}\right]_0^{-1} = 0 \\
f_1(p_3) &= \int_0^1 p_3(x)\,dx = \int_0^1 (-\frac{1}{3} + x - \frac{1}{2}x^2)\,dx = \left[-\frac{1}{3}x + \frac{x^2}{2} - \frac{x^3}{6}\right]_0^1 = 0 \\
f_2(p_3) &= \int_0^2 p_3(x)\,dx = \int_0^2 (-\frac{1}{3} + x - \frac{1}{2}x^2)\,dx = \left[-\frac{1}{3}x + \frac{x^2}{2} - \frac{x^3}{6}\right]_0^2 = 0 \\
f_3(p_3) &= \int_0^{-1} p_3(x)\,dx = \int_0^{-1} (-\frac{1}{3} + x - \frac{1}{2}x^2)\,dx = \left[-\frac{1}{3}x + \frac{x^2}{2} - \frac{x^3}{6}\right]_0^{-1} = 1.
\end{aligned}$$

Hence the dual of $\{p_1, p_2, p_3\}$ is $\{f_1, f_2, f_3\}$. Since $\{p_1, p_2, p_3\}$ is a basis for V, $\{f_1, f_2, f_3\}$ is a basis for V^*.

2.4 The Double Dual

If V is finite dimensional vector space over F then since V and V^* are both are of same dimension over F, they are both isomorphic to F^n. Thus V is isomorphic to V^*. However defining an isomorphism from V to V^* depends on the particular choice of the basis of V.

DEFINITION. 2.39 For a vector space V the dual of V^* is called the *double dual* of V and is denoted by $(V^*)^*$ or simply by V^{**}.

THEOREM. 2.40 *Let V be a finite dimensional vector space over a field F. The map $e_V : V \to V^{**}$, defined by $e_V(v)(f) = f(v)$ for all $v \in V$ for all $f \in V^*$, is an isomorphism of V onto V^{**}.*

[Note that the members of V^* are linear functionals of V, i.e. linear maps from V to F, and the members of V^{**} are linear functionals of V^*, i.e., linear maps from V^* to F. Thus in the statement of the theorem $e_V(v) \in V^{**}$, i.e., $e_V(v)$ is a linear map from V^* to F. According to the statement $e_V(v)(f)$ is the value of $e_V(v)$ at the point f, which, by definition is $f(v)$.]

PROOF. For $v_1, v_2 \in V, c \in F$ and $f \in V^*$, $e_V(cv_1+v_2)(f) = f(cv_1+v_2) = cf(v_1)+f(v_2) = ce_V(v_1)(f) + e_V(v_2)(f)$. Hence $e_V(cv_1 + v_2) = ce_V(v_1) + e_V(v_2)$. Thus e_V is a linear transformation.

Since both of V and V^{**} are of same dimension it is sufficient to show that e_V is injective. Let $v \in$ null e_V. Then $e_V(v) = 0$, where 0 denotes the zero element of V^{**} which is the the zero mapping on V^*. Thus $e_V(v)(f) = 0$ for all $f \in V^*$, i.e., $f(v) = 0$ for all $f \in V^*$. Hence $v = \theta$ which implies that e_V is injective. Thus e_V is an isomorphism. ■

Since e_V is bijection the following is immediate.

COROLLARY. 2.41 *If V is a finite dimensional vector space then for each $L \in V^{**}$ there exists unique $v \in V$ such that $L = e_V(v)$.*

COROLLARY. 2.42 *If V is a finite dimensional vector space then every basis of V^* is the dual of some basis of V.*

PROOF. Let $\mathcal{B}^* = \{f_1, f_2, \ldots, f_n\}$ be a basis of V^*. Then it has a dual basis $\mathcal{B}^{**} = \{\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_n\}$ in V^{**}. By the above corollary for each $\hat{f}_i$ in V^{**} there is unique α_i in V such that $\hat{f}_i = e_V(\alpha_i)$, $1 \leq i \leq n$. Since $\mathcal{B}^{**}$ is a basis of V^{**} and e_V is an isomorphism, $\mathcal{B} = \{\alpha_1\alpha_2, \ldots, \alpha_n\}$ is a basis of V.

By definition of e_V, $e_V(\alpha_i)(f_j) = f_j(\alpha_i)$ for $1 \leq j \leq n$. On the other hand $e_V(\alpha_i)(f_j) = \hat{f}_i(f_j) = \delta_{ij}$. Hence $f_j(\alpha_i) = \delta_{ij}$, $1 \leq i, j \leq n$. This shows that $f_i = \hat{\alpha}_i$ for all $i = 1, 2, \ldots n$. Thus the basis $\mathcal{B}^*$ is the dual of the basis $\mathcal{B}$. ■

The isomorphism from V to V^{**} defined in the above theorem is independent of any choice of particular basis. We can identify V with V^{**} under this isomorphism $v \mapsto e_V(v)$.

THEOREM. 2.43 *Let V be a finite dimensional vector space. For a subset $S \subset V$, $(S^0)^0$ is the subspace spanned by S.*

PROOF. It is known that for $S \subset V$, S^0 is a subspace of V^* and hence $(S^0)^0$ is a subspace of V^{**}. Also $(L(S))^0 = S^0$. So it is sufficient to show that for a subspace S of V, $(S^0)^0 = S$.

Note that $\dim S + \dim S^0 = \dim V$ and $\dim S^0 + \dim(S^0)^0 = \dim V^*$. Since $\dim V = \dim V^*$ we have $\dim S = \dim(S^0)^0$. Since S is a subspace of $(S^0)^0$ we have $S = (S^0)^0$. ■

EXAMPLE. 2.44 Let $V = \mathbb{R}^n$, $S = \{(x_1, x_2, \ldots, x_n) : x_1 + x_2 + \cdots + x_n = 0\}$. (i) Show that S is a subspace of V (ii) Find S^0 and (iii) S^*.

(i) Follows from the fact for $v, w \in S$ and for $c \in \mathbb{R}$, $v + w \in S$ and $cv \in S$.

(ii) If $f \in V^*$, then the linear map $f : \mathbb{R}^n \to \mathbb{R}$ can be represented by an $1 \times n$ matrix $A = (c_1\ c_2\ \cdots\ c_n)$ and for $v = (x_1, x_2, \ldots x_n) \in \mathbb{R}^n$, $f(v) = Av = A(x_1\ x_2\ \cdots\ x_n)^t = c_1x_2 + c_2x_2 + \cdots + c_nx_n$.

$f \in S^0$ if and only if $f(s) = 0$ for all $s \in S$. Since for $s = (x_1, x_2, \ldots, x_n) \in S$, $x_1 + x_2 + \cdots x_n = 0$, $c_1x_2 + c_2x_2 + \cdots + c_nx_n = 0$ if and only if $c_1 = c_2 = \cdots = c_n$. Hence f must be of the form $f(v) = cv$ for some scalar c. Thus

$$S^0 = \{f : \mathbb{R}^n \to \mathbb{R} : \exists c \in \mathbb{R}, f(v) = cv\ \forall v \in \mathbb{R}^n\}.$$

(iii) Define $\phi : S \to S^*$ as follows: for $v = (c_1, c_2, \ldots, c_n) \in S$, $\phi(v)(x) = c_1x_1 + c_2x_2 + \cdots + c_nx_n$ for all $(x_1, x_2, \ldots, x_n) \in S$. Then it immediately follows that ϕ is linear and bijective. Hence S^* is precisely the collection of all those f, $f(x_1, x_2, \ldots, x_n) = c_1x_2 + c_2x_2 + \cdots + c_nx_n$ such that $c_1 + c_2 + \cdots + c_n = 0$.

2.5 Eigenvalues and diagonalization

DEFINITION. 2.45 Let V be a vector space over a field F and $T : V \to V$ be a linear operator. A scalar $c \in F$ is called a *characteristic root* or *eigenvalue* of the operator T if there exists a non-null vector $\alpha \in V$ such that $T(\alpha) = c\alpha$. The vector α is called the *characteristic vector* or *eigenvector* associated with the characteristic root c.

THEOREM. 2.46 *If c is a characteristic root of a linear operator T on a vector space V then the set of all characteristic vectors form a subspace of V.*

PROOF. Let I denote the identity transformation, then cI is a linear operator on V and hence $T - cI$ is a linear operator on V. A vector $\alpha \in V$ is a characteristic vector associated with c iff $T(\alpha) = c\alpha$ iff $T(\alpha) = cI(\alpha)$ iff $(T - cI)(\alpha) = 0$. Thus the set of all the characteristic vectors is the null space of the linear operator $T - cI$ and hence a subspace of V. ■

DEFINITION. 2.47 The subspace of all the characteristic vectors associated with a characteristic root c is called the *characteristic spaces associated with c.*

For a finite dimensional vector space V the characteristic vectors can be found with the help of associated determinant.

THEOREM. 2.48 *Let V be a finite dimensional vector space and $T : V \to V$ be a linear operator. Then the followings are equivalent:*

1. *c is a characteristic root of T.*
2. $\det(T - cI) = 0$.

PROOF. (1) $\Rightarrow$ (2): We do not distinguish the linear operator T with its matrix representation. If α is a characteristic vector associated with c then $T\alpha = c\alpha$. if I denotes the unit matrix, then $T\alpha - cI\alpha = 0$, i.e, $(T - cI)\alpha = 0$. This shows that the system of homogeneous linear equations $(T - cI)x = 0$, where $x = (x_1\ x_2\ \ldots\ x_n)^t$, has a non-null solution α. Hence the matrix $T - cI$ must be singular, i.e., $\det(T - cI) = 0$.

(2) $\Rightarrow$ (1): If $\det(T-cI) = 0$ then the system of homogeneous linear equations $(T-cI)x = 0$, where $x = (x_1\ x_2\ \ldots\ x_n)^t$, has a non-null solution, say $x = \alpha$. So $(T - cI)\alpha = 0$ which gives $T\alpha - c\alpha$. Hence c is a characteristic root of T and α is a characteristic vector associated with c. ■

DEFINITION. 2.49 If T is a linear operator on a vector space V of dimension n then $\det(T - cI)$ is a polynomial in c of degree n. This polynomial is called the *characteristic polynomial* of the linear operator T. The equation $\det(T - cI) = 0$ is called the *characteristic equation* of T.

DEFINITION. 2.50 If A is an $n \times n$ matrix then the polynomial $\det(A - \lambda I)$ in λ is called the *characteristic polynomial* and the equation $\det(A - \lambda I) = 0$ is called the *characteristic*

equation of the matrix A. A root of the characteristic equation is called a *characteristic root* or *eigenvalue* of the matrix A and any vector x satisfying the relation $Ax = \lambda x$ is called a *characteristic vector* or *eigen vector* associated with the characteristic root λ.

Recall that two matrices A, B are called *similar matrices* if there exists a matrix P such that $B = P^{-1}AP$.

THEOREM. 2.51 *Similar matrices have the same characteristic equation and hence the same characteristic roots.*

PROOF. Let the matrix A and B be similar. Then there exists a non-singular matrix P such that $B = P^{-1}AP$. Then

$$\begin{aligned}\det(B-\lambda I) &= \det(P^{-1}AP-\lambda I) \;=\; \det(P^{-1}AP-P^{-1}\lambda IP)\\ &= \det(P^{-1}(A-\lambda I)P) \;=\; \det(P^{-1})\det(A-\lambda I)\det P\\ &= \det(A-\lambda I).\end{aligned}$$

Hence the result. ■

DEFINITION. 2.52 A linear operator T on a finite dimensional vector space V is called *diagonalizable* if V has a basis each vector of which is a characteristic vector of T.

Let T be a diagonalizable linear operator on a finite dimensional vector space V. Then there is an ordered basis $\{v_1, v_2, \ldots, v_n\}$ of V, each v_i being a characteristic vector. Let c_i be the characteristic root such that $T(v_i) = c_i v_i$, $1 \le i \le n$. Let us denote the matrix representation of T by T itself and v_i^t be the transpose of the vector v_i. Then $Tv_i^t = c_i v_i^t$, $1 \le i \le n$. Let $P = (v_1^t\ v_2^t \cdots\ v_n^t)$, the columns of the matrix P are the transposes of v_i. Then $TP = T(v_1^t\ v_2^t \cdots\ v_n^t) = (c_1v_1^t\ c_2v_2^t \cdots\ c_nv_n^t) = (v_1^t\ v_2^t \cdots\ v_n^t)\cdot\mathrm{diag}(c_1, c_2, \ldots, c_n) = P\cdot\mathrm{diag}(c_1, c_2, \ldots, c_n)$, where $\mathrm{diag}(c_1, c_2, \ldots, c_n)$ represents the diagonal matrix whose entries are $c_1, c_2, \ldots, c_n$. Hence $P^{-1}TP = \mathrm{diag}(c_1, c_2, \ldots, c_n)$, i.e., the matrix representation of a diagonalizable linear operator is similar to a diagonal matrix.

EXAMPLE. 2.53 Show that the matrix $A = \begin{pmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{pmatrix}$ over $\mathbb{R}$ is diagonalizable. Find a matrix P such that $P^{-1}AP$ is a diagonal matrix.

The characteristic polynomial is give by

$$\begin{vmatrix} -9-\lambda & 4 & 4 \\ -8 & 3-\lambda & 4 \\ -16 & 8 & 7-\lambda \end{vmatrix} = \begin{vmatrix} -1-\lambda & 1+\lambda & 0 \\ -8 & 3-\lambda & 4 \\ -16 & 8 & 7-\lambda \end{vmatrix}$$

$$= (1+\lambda)\begin{vmatrix} -1 & 1 & 0 \\ -8 & 3-\lambda & 4 \\ -16 & 8 & 7-\lambda \end{vmatrix} = (1+\lambda)\begin{vmatrix} -1 & 0 & 0 \\ -8 & -5-\lambda & 4 \\ -16 & -8 & 7-\lambda \end{vmatrix}$$

$$= (1+\lambda)\begin{vmatrix} 1 & 0 & 0 \\ 8 & 5+\lambda & 4 \\ 16 & 8 & 7-\lambda \end{vmatrix} = (1+\lambda)((5+\lambda)(7-\lambda)-32)$$

$$= (1+\lambda)(3+2\lambda-\lambda^2) = (1+\lambda)(1+\lambda)(3-\lambda).$$

Hence the characteristic roots are $\lambda_1 = \lambda_2 = -1, \lambda_3 = 3$.

For λ_1 and λ_2 the matrix $A - \lambda_1 I = \begin{pmatrix} -8 & 4 & 4 \\ -8 & 4 & 4 \\ -16 & 8 & 8 \end{pmatrix}$ has rank 1 and hence nullity 2, i.e., its null space has 2 linearly independent vectors. If $x = (x_1 \;\; x_2 \;\; x_3)^t$ is a characteristic vector associated with $\lambda_1 = -1$ then $\begin{pmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_1 \\ -x_2 \\ -x_3 \end{pmatrix}$ which gives a single equation $-2x_1 + x_2 + x_3 = 0$. Taking $x_3 = 0$ we get a solution $(1, 2, 0)$ and taking $x_2 = 0$ we get a solution $(1, 0, 2)$ which are linearly independent.

For $\lambda_3 = 3$, $A - \lambda_3 I = \begin{pmatrix} -12 & 4 & 4 \\ -8 & 0 & 4 \\ -16 & 8 & 4 \end{pmatrix} \sim \begin{pmatrix} -3 & 1 & 1 \\ -2 & 0 & 1 \\ -4 & 2 & 1 \end{pmatrix} \sim \begin{pmatrix} -3 & 1 & 1 \\ -2 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 0 & 0 & 0 \\ -2 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$ has rank 2 and hence nullity 1, i.e., its null space has 1 vector. If $x = (x_1 \;\; x_2 \;\; x_3)^t$ is a characteristic vector then $\begin{pmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3x_1 \\ 3x_2 \\ 3x_3 \end{pmatrix}$ which gives the homogeneous equations $-2x_1 + x_3 = 0$, and $-x_1 + x_2 = 0$. This gives a solution $x_1 = 1, x_2 = 1, x_3 = 2$.

The characteristic vectors $v_1 = (1, 2, 0), v_2 = (1, 0, 2)$ and $v_3 = (1, 1, 2)$ are linearly independent and hence form a basis for $\mathbb{R}^3$. Thus the matrix is diagonalizable.

If we take the matrix $P = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 0 & 2 & 2 \end{pmatrix}$, whose columns are transposes of v_1, v_2, v_3 then $AP = (\lambda_1 v_1 \;\; \lambda_2 v_2 \;\; \lambda_3 v_3) = (-v_1 \;\; -v_2 \;\; 3v_3) = P\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. Hence, $P^{-1}AP = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ which is the diagonalization of A.

EXAMPLE. 2.54 Verify whether the matrix $A = \begin{pmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{pmatrix}$ over $\mathbb{R}$ is diagonalizable.

The characteristic equation is $\begin{vmatrix} 6-\lambda & -3 & -2 \\ 4 & -1-\lambda & -2 \\ 10 & -5 & -3-\lambda \end{vmatrix} = 0$. Using row and column operations on the determinant,

$$\begin{vmatrix} 6-\lambda & -3 & -2 \\ 4 & -1-\lambda & -2 \\ 10 & -5 & -3-\lambda \end{vmatrix} = \begin{vmatrix} 2-\lambda & -2+\lambda & 0 \\ 4 & -1-\lambda & -2 \\ 10 & -5 & -3-\lambda \end{vmatrix}$$

$$= (2-\lambda)\begin{vmatrix} 1 & -1 & 0 \\ 4 & -1-\lambda & -2 \\ 10 & -5 & -3-\lambda \end{vmatrix} = (2-\lambda)\begin{vmatrix} 1 & 0 & 0 \\ 4 & 3-\lambda & -2 \\ 10 & 5 & -3-\lambda \end{vmatrix}$$

$$= -(2-\lambda)\begin{vmatrix} 1 & 0 & 0 \\ 4 & 3-\lambda & 2 \\ 10 & 5 & 3+\lambda \end{vmatrix} = -(2-\lambda)((3-\lambda)(3+\lambda)-10)$$

$$= (2-\lambda)(1+\lambda^2).$$

Hence the characteristic equation is $(2-\lambda)(1+\lambda^2) = 0$ which has only one real root $\lambda = 2$. So the characteristic vector can not span V. Hence the matrix A is not diagonalizable over the field $\mathbb{R}$.

The filed concerned is important. It can be noted that the characteristic equation has three roots over the field of complex numbers, $\lambda_1 = 2, \lambda_2 = i, \lambda_3 = -i$. Hence it can be shown that the matrix is diagonalizable over the field $\mathbb{C}$.

THEOREM. 2.55 *If A is a non-singular matrix over a field F and c is a characteristic root of A the c^{-1} is a characteristic root of A^{-1} with same characteristic vector.*

PROOF. We have $c \neq 0$. Then $|A - cI| = 0 \Rightarrow |AA^{-1} - cIA^{-1}| = 0 \Rightarrow |I - cA^{-1}| = 0 \Rightarrow |\frac{1}{c}I - A^{-1}| = 0$. Hence $\frac{1}{c}$ is a characteristic root of A^{-1}. ■

THEOREM. 2.56 *The characteristic roots of a matrix A are the same as those of A^t.*

PROOF. If c is a characteristic root of the matrix A then $|A - cI| = 0$. This implies that $|A - cI|^t = 0 \Rightarrow |A^t - cI^t| = 0 \Rightarrow |A^t - cI| = 0$. Hence c is a characteristic root of A^t. ■

THEOREM. 2.57 *The characteristic roots of a real symmetric matrix are all real.*

PROOF. Let λ be a characteristic root of a real symmetric matrix A and x be a characteristic vector associated with λ. Then $Ax = \lambda x$, taking complex conjugate on both sides,

since A is real,

$$A\bar{x} = \bar{\lambda}\bar{x} \;\Rightarrow\; (A\bar{x})^t = \bar{\lambda}\bar{x}^t \;\Rightarrow\; \bar{x}^t A = \bar{\lambda}\bar{x}^t$$

Hence,

$$\bar{\lambda}\bar{x}^t x \;=\; \bar{x}^t Ax = \bar{x}^t \lambda x = \lambda \bar{x}^t x.$$

If $x = (x_1, x_2, \ldots, x_n)^t \in \mathbb{C}^n$ then $\bar{x}^t x = \bar{x}_1 x_1 + \bar{x}_2 x_2 + \cdots + \bar{x}_n x_n = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 > 0$. Hence $\bar{\lambda} = \lambda$, i.e., λ is real. ■

THEOREM. 2.58 *The characteristic roots of a real skew-symmetric matrix are purely imaginary.*

PROOF. Let λ be a characteristic root of a real skew-symmetric matrix A and x be a characteristic vector associated with λ. Then $Ax = \lambda x$, taking complex conjugate on both sides, since A is real,

$$A\bar{x} = \bar{\lambda}\bar{x} \;\Rightarrow\; (A\bar{x})^t = \bar{\lambda}\bar{x}^t \;\Rightarrow\; \bar{x}^t(-A) = \bar{\lambda}\bar{x}^t$$

Hence,

$$\bar{\lambda}\bar{x}^t x \;=\; \bar{x}^t(-Ax) = -\bar{x}^t \lambda x = -\lambda \bar{x}^t x.$$

If $x = (x_1, x_2, \ldots, x_n)^t$ then $\bar{x}^t x = \bar{x}_1 x_1 + \bar{x}_2 x_2 + \cdots + \bar{x}_n x_n > 0$. Hence $\bar{\lambda} = -\lambda$, i.e., λ is purely imaginary. ■

THEOREM. 2.59 *The characteristic roots of a real orthogonal matrix have modulus 1.*

PROOF. If A is an orthogonal matrix then $AA^t = A^t A = I$. If λ is a characteristic root and v is an associated characteristic vector then $Av = \lambda v$. Then $(Av)^t = \lambda v^t$. Hence

$$\begin{aligned}(Av)^t(Av) = \lambda^2 v^t v \;&\Rightarrow\; v^t A^t A v = \lambda^2 v^t v \;\Rightarrow\; v^t I v = \lambda^2 v^t v \;\Rightarrow\; v^t v = \lambda^2 v^t v \\ &\Rightarrow\; v^t v(1 - \lambda^2) = 0.\end{aligned}$$

Since $v^t v = \|v\| \neq 0$ we must have $1 - \lambda^2 = 0$, i.e., $\lambda = \pm 1$. ■

Recall that for arbitrary matrices $A = (a_{ij})_{n\times n}$ and $B = (b_{ij})_{n\times n}$, $AB = C = (c_{ij})_{n\times n}$, where $c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$. Hence $Tr(AB) = \sum_{i=1}^n c_{ii} = \sum_{i=1}^n \sum_{k=1}^n a_{ik}b_{ki}$. Similarly, $Tr(BA) = \sum_{i=1}^n \sum_{k=1}^n b_{ik}a_{ki} = Tr(AB)$. Thus $Tr(AB) = Tr(BA)$. We use the fact in the next result.

THEOREM. 2.60 *The sum of characteristic roots of a matrix is the trace of the matrix.*

PROOF. Let the matrix A over a field F have characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_n$. These roots all may not belong to the field F but belong to some algebraic extension F_1 of F. Then there exists a non-singular matrix P over F_1 such that $B = P^{-1}AP$ is the diagonal matrix $diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Since the characteristic roots of a a diagonal matrix are the diagonal elements and similar matrices have same characteristic roots, it follows that $Tr(A) = Tr((AP)P^{-1}) = Tr(P^{-1}AP) = Tr(B) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. ■

2.6 Minimal Polynomial

For a field F the set $F[x]$ of all the polynomials with coefficients from F is a commutative ring with unity. Moreover $F[x]$ is a PID and every ideal in $F[x]$ is generated by a unique monic polynomial.

If V is a vector space over a field F and $T \in L(V, V)$ is a linear operator on V then for any $n \in \mathbb{N}$, $T^n \in L(V, V)$ and for $c \in F$, $cT \in L(V, V)$. Hence for $p \in F[x]$, $p(T) \in L(V, V)$.

LEMMA. 2.61 *Let V be a vector space over a field F and T be a linear operator on V. If c is a characteristic root of T and v is a characteristic vector associated with c then for any polynomial p over F $p(T)(v) = p(c)v$.*

PROOF. We have $T(v) = cv$. Assume $T^k(v) = c^k v$ for some $k \geq 1$ then $T^{k+1}(v) = TT^k(v) = T(c^k v) = c^k T(v) = c^k cv = c^{k+1}v$. Hence by induction $T^m(v) = c^m v$ for all $m \in \mathbb{N}$.

If $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ be a polynomial over F then

$$\begin{aligned} p(T)(v) &= a_0 v + a_1 T(v) + a_2 T^2(v) + \cdots + a_n T^n(v) \\ &= a_0 v + a_1 cv + a_2 c^2 v + \cdots + a_n c^n v \\ &= (a_0 + a_1 c + a_2 c^2 + \cdots + a_n c^n)v \;=\; p(c)v. \end{aligned}$$

DEFINITION. 2.62 Let T be a linear operator on a vector space V over a field F. A polynomial $p \in F[x]$ is called an *annihilating polynomial* of T if $p(T) = 0$.

LEMMA. 2.63 *The set of all the annihilating polynomials of a linear operator T on a vector space V over a field F form an ideal in $F[x]$.*

PROOF. Let p, q be two annihilating polynomials of the linear operator T. Then $p(T) = q(T) = 0$. Hence $(p - q)(T) = p(T) - q(T) = 0$ showing that $p - q$ is an annihilating

polynomial. Also for any polynomial $f \in F[x]$, $(pf)(T) = p(T)f(T) = 0f(T) = 0$, hence pf is an annihilating polynomial. ■

It is known that for a field F the ring $F[x]$ is a PID, hence for a linear operator T on a vector space V over F the ideal of annihilating polynomials of T is generated by a unique monic polynomial g.

DEFINITION. 2.64 Let T be a linear operator on a vector space V over a field F. The unique monic polynomial g is called the *minimal polynomial* of T if the ideal of all the annihilating polynomials of T is generated by g.

DEFINITION. 2.65 For a square matrix A over a field F a monic polynomial $g(x)$ is called the *minimal polynomial* of A if the ideal of all the annihilating polynomials of A is generated by g

THEOREM. 2.66 *If V is finite dimensional vector space then for any linear operator T on V the ideal of annihilating polynomial is non-zero.*

PROOF. Let $\dim V = n$, then $\dim L(V,V) = n^2$. If $T \in L(V,V)$ is a linear operator, consider the $n^2 + 1$ linear operators $I, T, T^2, \ldots, T^{n^2}$ on V. Since $\dim L(V,V) = n^2$ these $n^2 + 1$ vectors must be linearly dependent. So there are scalars, not all zero, $c_0, c_1, c_2, \ldots, c_{n^2}$ such that $c_0 I + c_1 T + c_2 T^2 + \cdots + c_{n^2} T^{n^2} = 0$. Taking $p(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n^2} x^{n^2}$ then $p \in F[x]$ and $p(T) = 0$. Thus p is a non-zero annihilating polynomial of T. ■

THEOREM. 2.67 *Let A be an $n \times n$ matrix over a field F. Then the minimal polynomial and the characteristic polynomial of A have the same roots except for multiplicities.*

The same result is true for a linear operator on an n dimensional vector space V also.

PROOF. Let p be the minimal polynomial of A. Then $p(A) = 0$. It is sufficient to show that a scalar c is a root of $p(x) = 0$ if and only if c is a characteristic root of A.

Assume that c is a root of $p(x) = 0$. Then $p(x) = (x - c)q(x)$ for some polynomial $q(x)$. So, $0 = p(A) = (A - cI)q(A)$. Since p is minimal polynomial and $\deg q < \deg p$ it follows that $q(A) \neq 0$. We choose a vector w such that $q(A)w = v \neq 0$. Then $0 = p(A)w = (A - cI)q(A)w = (A - cI)v$ which shows that c is a characteristic root and v is an associated characteristic vector.

Conversely, assume that c is a characteristic root of the matrix A. If v is a characteristic vector associated with c then by Lemma 2.61, $p(A)v = p(c)v$. Since $p(A) = 0$ we have $p(c) = 0$, i.e., c is a root of the minimal polynomial. ■

REMARK. 2.68 It follows from the above that for V is a vector space over a field F

1. The minimal polynomial divides the characteristic polynomial and hence the degree of minimal polynomial can not exceed the dimension of the vector space.
2. if A is a diagonalizable matrix with the distinct characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_k$ then $p(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$ is the minimal polynomial of A.
3. If A is not diagonalizable then the roots of the characteristic polynomial belong to some extended field F_1. The product of linear factors for distinct characteristic roots in the extended field is the minimal polynomial. However, it can be proved that the coefficients of the minimal polynomial belongs to the original field F.

EXAMPLE. 2.69 Find the minimal polynomial of the matrix $A = \begin{pmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 2 \end{pmatrix}$.

The characteristic polynomial of A is $\begin{vmatrix} 3-x & -1 & 0 \\ 0 & 2-x & 0 \\ 1 & -1 & 2-x \end{vmatrix} = (3-x)(2-x)^2$. Hence the minimal polynomial will be either $(3-x)(2-x)$ or $(3-x)(2-x)^2$.

Now, (3I-A)(2I-A) =

$$\left[\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} - \begin{pmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 2 \end{pmatrix}\right] \times \left[\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} - \begin{pmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 2 \end{pmatrix}\right]$$

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

Hence the minimal polynomial is $p(x) = (3-x)(2-x)$.

EXAMPLE. 2.70 Find the characteristic polynomial and the minimal polynomial of the matrix $A = \begin{pmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{pmatrix}$

The characteristic polynomial is given by

$$\begin{vmatrix} 0-x & 0 & c \\ 1 & 0-x & b \\ 0 & 1 & a-x \end{vmatrix} = \begin{vmatrix} -x & 0 & c \\ 1 & 0 & b+ax-x^2 \\ 0 & 1 & a \end{vmatrix} = (-1)(x^3 - ax^2 - bx - c).$$

The minimal polynomial $p(x)$ has degree ≤ 3. Let us take a polynomial of degree 2, $f(x) = x^2 + px + q$. Then

$$\begin{aligned} f(A) &= \begin{vmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{vmatrix}^2 + p \begin{vmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{vmatrix} + qI \\ &= \begin{vmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{vmatrix} \cdot \begin{vmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{vmatrix} + \begin{vmatrix} 0 & 0 & c \\ p & 0 & b \\ 0 & 1 & a \end{vmatrix} + \begin{vmatrix} q & 0 & 0 \\ 0 & q & 0 \\ 0 & 0 & q \end{vmatrix} \\ &= \begin{vmatrix} 0 & c & ac \\ 0 & b & c+ab \\ 1 & a & b+a^2 \end{vmatrix} + \begin{vmatrix} 0 & 0 & c \\ p & 0 & b \\ 0 & 1 & a \end{vmatrix} + \begin{vmatrix} q & 0 & 0 \\ 0 & q & 0 \\ 0 & 0 & q \end{vmatrix} = \begin{vmatrix} q & c & ac+c \\ p & b+q & c+ab+b \\ 1 & a+1 & b+a^2+a+q \end{vmatrix} \neq 0. \end{aligned}$$

Thus $f(x)$ can not be the minimal polynomial. So $\deg p(x) = 3$. Since the characteristic polynomial divides $p(x)$ we have $p(x) = x^3 - ax^2 - bx - c$.

EXAMPLE. 2.71 Let n be a positive integer, and let V be the space of polynomials over $\mathbb{R}$ which have degree at most n. Let D be the differentiation operator on V. What is the minimal polynomial for D?

A basis for V is $\{1, x.x^2, \ldots, x^n\}$. D operates on the basis as follows: $D1 = 0, Dx = 1, Dx^2 = 2x, \ldots, Dx^n = nx^{n-1}$. We observe that for any $f \in V$, $D^{n+1}f = 0$ and for $k \leq n$, $D^k x^n = n(n-1)(n-2)\cdots(n-k+1)x^{n-k} \neq 0$. Hence the minimal polynomial of D is $p(x) = x^{n+1}$.

3 Inner Product Spaces

DEFINITION. 3.1 A vector space V over the field of complex numbers $\mathbb{C}$ is called an *inner product space* if for each $x, y \in V$ there exists a unique complex number, denoted by $\langle x, y\rangle$, which satisfy the following properties:

1. $\langle x, y\rangle = \langle y, x\rangle^*$ for all $x, y \in V$, where for $z \in \mathbb{C}$, z^* denotes the complex conjugate of z,
2. $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ for all $x, y \in V$ and for all $\lambda \in \mathbb{C}$,
3. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ for all $x, y, z \in V$,
4. $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$ for all $x \in V$.

If V is a vector space over the field of real numbers $\mathbb{R}$, then condition 1 stated above becomes

1.* $\langle x, y\rangle = \langle y, x\rangle$ for all $x, y \in V$.

In such case V is called a real inner product space.

It immediately follows from definition that if V is an inner product space, $x_1, x_2, \ldots, x_n \in V$, $y_1, y_2, \ldots, y_n \in V$, $\lambda_1, \lambda_2, \ldots, \lambda_n, \mu_1, \mu_2, \ldots \mu_n \in \mathbb{C}$ then

$$\left\langle \sum_{i=1}^{n} \lambda_i x_i, \sum_{i=1}^{n} \mu_i y_i \right\rangle = \sum_{i=1}^{n} \lambda_i \mu_i^* \langle x_i, y_i\rangle.$$

EXAMPLE. 3.2

1. Let n be a positive integer. For all $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ define $\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i$. Then $\langle\ \rangle$ defines an inner product on $\mathbb{R}^n$. This inner product is nothing but the usual dot product in $\mathbb{R}^n$.
2. Consider the unitary space $\mathbb{C}^n, n \in \mathbb{N}$. For $x = (\xi_1, \xi_2, \ldots, \xi_n)$, $y = (\eta_1, \eta_2, \ldots, \eta_n) \in \mathbb{C}^n$, define $\langle x, y\rangle = \sum_{i=1}^{n} \xi_i \bar{\eta}_i$, then $\mathbb{C}^n$ is an inner product space.
3. Consider the ℓ^p space, where $p \geq 1$ is a real number, whose members are all the sequences of numbers (real or complex) $\{x_n\}$ such that $\sum |x_n|^p$ is convergent. ℓ^p is a normed linear space. It can be verified that ℓ^2 is an inner product space where the inner product is defined by $\langle x, y\rangle = \sum_{i=1}^{\infty} x_i \bar{y_i}$, $x = \{x_i\}, y = \{y_i\}$.
4. For all $f, g \in L_E^2$, the set of all square summable functions, define $\langle f, g\rangle = \int_E fg\, d\mu$. Then $\langle\ \rangle$ is an inner product in L_E^2.

PROPOSITION. 3.3 (Schwarz inequality) *Let V be a real inner product space and $x, y \in V$. Then*

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle.$$

PROOF. Put $A = \langle x, x\rangle, B = \langle x, y\rangle$ and $C = \langle y, y\rangle$. Then for all $\lambda \in \mathbb{R}$,

$$\begin{aligned} A\lambda^2 + 2B\lambda + C &= \lambda^2 A + \lambda B + \lambda B + C \\ &= \lambda^2\langle x, x\rangle + \lambda\langle x, y\rangle + \lambda\langle x, y\rangle + \langle y, y\rangle \\ &= \langle \lambda x, \lambda x\rangle + \langle \lambda x, y\rangle + \langle \lambda x, y\rangle + \langle y, y\rangle \\ &= \langle \lambda x, \lambda x\rangle + \langle y, \lambda x\rangle + \langle \lambda x, y\rangle + \langle y, y\rangle \\ &= \langle \lambda x + y, \lambda x\rangle + \langle \lambda x + y, y\rangle \\ &= \langle \lambda x, \lambda x + y\rangle + \langle y, \lambda x + y\rangle \\ &= \langle \lambda x + y, \lambda x + y\rangle \\ &\geq 0. \end{aligned}$$

Thus $B^2 \leq AC$, i.e., $\langle x, y\rangle^2 \leq \langle x, x\rangle\langle y, y\rangle$. Taking positive square root, we have $|\langle x, y\rangle| \leq \sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}$.

THEOREM. 3.4 *Let V be a real inner product space. Define for each $x \in V$, $\|x\| = \sqrt{\langle x, x\rangle}$. Then $\|, \|$ is a norm on V.*

PROOF. For $x \in V$, $\|x\| = \sqrt{\langle x, x\rangle} \geq 0$. Also $\|x\| = 0$ iff $\|x\|^2 = 0$ iff $\langle x, x\rangle = 0$ iff $x = 0$.

For $x \in V, \lambda \in \mathbb{R}$, $\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\lambda^2\langle x, x\rangle} = |\lambda|\|x\|$.

Finally, for $x, y \in V$, $\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + 2|\langle x, y\rangle| + \langle y, y\rangle \leq \|x\|^2 + 2\|x\|.\|y\| + \|y\|^2$ (by Schwarz inequality). Hence $\|x + y\|^2 \leq (\|x\| + \|y\|)^2$ which implies that $\|x + y\| \leq \|x\| + \|y\|$.

Thus $\|, \|$ is a norm.

REMARK. 3.5 An inner product on a real inner product space induces a norm on the space. So we can talk of continuity and convergence with respect to the metric induced by this norm.

DEFINITION. 3.6 A complete real inner product space is called a *Hilbert space.*

THEOREM. 3.7 *The norm of an inner product space V satisfies the parallelogram equality:*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2),\ x, y \in V.$$

PROOF. For $x, y \in V$,

$$\begin{aligned}\|x+y\|^2 + \|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle + \langle x,x\rangle - \langle x,y\rangle - \langle y,x\rangle + \langle y,y\rangle \\ &= 2(\langle x,x\rangle + \langle y,y\rangle) \;=\; 2(\|x\|^2 + \|y\|^2).\end{aligned}$$

The following example says that a normed linear space in general may not satisfy the above equality.

EXAMPLE. 3.8 Consider the ℓ^p space, $p \neq 2$. Then ℓ^p is a normed linear space. Take $x = \{1, 1, 0, 0, 0, \ldots\}$, $y = \{1, -1, 0, 0, 0, \ldots\}$. Then $\|x\| = \|y\| = 2^{\frac{1}{p}}$ and $\|x+y\| = \|x-y\| = 2$. Hence it does not satisfy the parallelogram equality.

It follows from the above that ℓ^p is not an inner product space for $p \neq 2$.

In an inner product space V, if the norm is known then the inner product can be obtained from the following formula: for $x, y \in V$, where V a real inner product space,

$$\langle x, y\rangle \;=\; \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2),$$

For a complex inner product space V the formula is the polariation identity, given by,

$$\begin{aligned}Re\langle x, y\rangle &= \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2), \\ Im\langle x, y\rangle &= \frac{1}{4}(\|x+iy\|^2 - \|x-iy\|^2),\end{aligned}$$

DEFINITION. 3.9 Let V be an inner product space, $x, y \in V$. Then x and y are said to be *orthogonal* if $\langle x, y\rangle = 0$ and is written as $x \perp y$. Two sets $A, B \subset V$ are called orthogonal if for all $x \in A, y \in B$, $\langle x, y\rangle = 0$ and is written as $A \perp B$.

A set $A \subset V$ is called an orthogonal set if for $x, y \in A$, $x \perp y$. In particular, a basis B of V is called an *orthogonal basis* if the set B is orthogonal. If Moreover, $\|x\| = 1$ for all $x \in B$, then the basis B is called *orthonormal*.

EXAMPLE. 3.10

1. In $\mathbb{R}^n$, for $i \neq j$, $\alpha_i \perp \alpha_j$, where $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is the standard basis of $\mathbb{R}^n$. This is an orthonormal basis.
2. In $L^2_{[-\pi,\pi]}$, the functions $\sin mx, \sin nx, m \neq n, m, n \in \mathbb{N}$, are orthogonal.

3. In $L^2_{[-1,1]}$, $P_m(x) \perp P_n(X)$ whenever $m \neq n$, where $P_n(x)$ is the Legendre polynomial of degree n.

THEOREM. 3.11 (Pythagorean Theorem) *In an inner product space if $x \perp y$ then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.*

For a real inner product space the converse is also true. However converse is not true for a complex inner product space.

THEOREM. 3.12 *In an inner product space any orthogonal set is linearly independent.*

PROOF. Assume that $B \subset V$ is an orthogonal set. Let $x_1, x_2, \ldots, x_n \in B$ and $c_1, c_2, \ldots c_n$ be scalars such that $c_1x_1 + c_2x_2 + \cdots + c_nx_n = \theta$. Then for $1 \leq k \leq n$, $0 = \langle \theta, x_k \rangle = \langle \sum_{i=1}^n c_i x_i, x_k \rangle = \sum_{i=1}^n c_i \langle x_i, x_k \rangle = c_k \langle x_k, x_k \rangle = c_k \|x_k\|^2$. Since x_k is non-zero vector, $c_k = 0$. Hence B is linearly independent. ■

THEOREM. 3.13 *If V is a finite dimensional inner product space with a basis $B = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ then it has an orthogonal basis $B' = \{\beta_1, \beta_2, \ldots, \beta_n\}$ such that for each $1 \leq k \leq n$, $L(\{\alpha_1, \alpha_2, \ldots, \alpha_k\}) = L(\{\beta_1, \beta_2, \ldots, \beta_k\})$.*