Mappings, Inner Product Spaces & Orthogonalization

Vineet Sahula

sahula@ieee.org

Vector Spaces & Sub-Spaces

ROW SPACE OF A MATRIX

Let A be an arbitrary $m \times n$ matrix over a field K:

$$A = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

The rows of A,

$$R_1 = (a_{11}, a_{21}, \dots, a_{1n}), \quad \dots, \quad R_m = (a_{m1}, a_{m2}, \dots, a_{mn})$$

viewed as vectors in K^n, span a subspace of K^n called the *row space* of A. That is,

$$\text{row space of } A = L(R_1, R_2, \dots, R_m)$$

Analogously, the columns of A, viewed as vectors in K^m, span a subspace of K^m called the *column space* of A.

Vector Spaces & Sub-Spaces…

SUMS AND DIRECT SUMS

Let U and W be subspaces of a vector space V. The sum of U and W, written $U + W$, consists of all sums $u + w$ where $u \in U$ and $w \in W$:

$$U + W = \{u + w : u \in U, w \in W\}$$

Note that $0 = 0 + 0 \in U + W$, since $0 \in U, 0 \in W$. Furthermore, suppose $u + w$ and $u' + w'$ belong to $U + W$, with $u, u' \in U$ and $w, w' \in W$. Then

$$(u + w) + (u' + w') = (u + u') + (w + w') \in U + W$$

and, for any scalar k, $$k(u + w) = ku + kw \in U + W$$

Thus we have proven the following theorem.

Theorem 4.8: The sum $U + W$ of the subspaces U and W of V is also a subspace of V.

Vector Spaces & Sub-Spaces…

Definition: The vector space V is said to be the *direct sum* of its subspaces U and W, denoted by

$$V = U \oplus W$$

if every vector $v \in V$ can be written in one and only one way as $v = u + w$ where $u \in U$ and $w \in W$.

The following theorem applies.

Theorem 4.9: The vector space V is the direct sum of its subspaces U and W if and only if: (i) $V = U + W$, and (ii) $U \cap W = \{0\}$.

Basis & Dimension

LINEAR DEPENDENCE

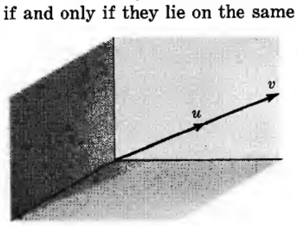
Definition: Let V be a vector space over a field K. The vectors $v_1, \ldots, v_m \in V$ are said to be *linearly dependent over K*, or simply *dependent*, if there exist scalars $a_1, \ldots, a_m \in K$, not all of them 0, such that

$$a_1v_1 + a_2v_2 + \cdots + a_mv_m = 0 \qquad (*)$$

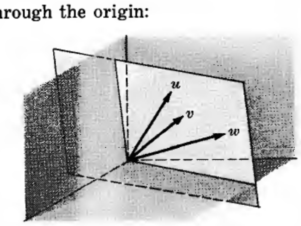
Otherwise, the vectors are said to be *linearly independent over K*, or simply *independent.*

Basis & Dimension…

Remark 6. In the real space $\mathbf{R}^3$, dependence of vectors can be described geometrically as follows: any two vectors u and v are dependent if and only if they lie on the same line through the origin; and any three vectors u, v and w are dependent if and only if they lie on the same plane through the origin:



u and v are dependent.



u, v and w are dependent.

Basis & Dimension…

BASIS AND DIMENSION

We begin with a definition.

Definition: A vector space V is said to be of *finite dimension* n or to be *n-dimensional*, written $\dim V = n$, if there exists linearly independent vectors $e_1, e_2, \ldots, e_n$ which span V. The sequence $\{e_1, e_2, \ldots, e_n\}$ is then called a *basis* of V.

The above definition of dimension is well defined in view of the following theorem.

Theorem 5.3: Let V be a finite dimensional vector space. Then every basis of V has the same number of elements.

The vector space $\{0\}$ is defined to have dimension 0. (In a certain sense this agrees with the above definition since, by definition, $\emptyset$ is independent and generates $\{0\}$.) When a vector space is not of finite dimension, it is said to be of *infinite dimension*.

Basis & Dimension…

Theorem 5.5: Suppose S generates V and $\{v_1, \ldots, v_m\}$ is a maximal independent subset of S. Then $\{v_1, \ldots, v_m\}$ is a basis of V.

The main relationship between the dimension of a vector space and its independent subsets is contained in the next theorem.

Theorem 5.6: Let V be of finite dimension n. Then:

(i) Any set of $n+1$ or more vectors is linearly dependent.

(ii) Any linearly independent set is part of a basis, i.e. can be extended to a basis.

(iii) A linearly independent set with n elements is a basis.

Basis & Dimension…

DIMENSION AND SUBSPACES

The following theorems give basic relationships between the dimension of a vector space and the dimension of a subspace.

Theorem 5.7: Let W be a subspace of an n-dimension vector space V. Then $\dim W \leq n$. In particular if $\dim W = n$, then $W = V$.

Example 5.8: Let W be a subspace of the real space $\mathbf{R}^3$. Now $\dim \mathbf{R}^3 = 3$; hence by the preceding theorem the dimension of W can only be 0, 1, 2 or 3. The following cases apply:

(i) $\dim W = 0$, then $W = \{0\}$, a point;
(ii) $\dim W = 1$, then W is a line through the origin;
(iii) $\dim W = 2$, then W is a plane through the origin;
(iv) $\dim W = 3$, then W is the entire space $\mathbf{R}^3$.

Theorem 5.8: Let U and W be finite-dimensional subspaces of a vector space V. Then $U + W$ has finite dimension and

$$\dim(U+V) \quad = \quad \dim U + \dim W - \dim(U \cap W)$$

Note that if V is the direct sum of U and W, i.e. $V = U \oplus W$, then $\dim V = \dim U + \dim W$ (Problem 5.48).

Basis & Dimension…

RANK OF A MATRIX

Let A be an arbitrary $m \times n$ matrix over a field K. Recall that the row space of A is the subspace of K^n generated by its rows, and the column space of A is the subspace of K^m generated by its columns. The dimensions of the row space and of the column space of A are called, respectively, the *row rank* and the *column rank* of A.

Theorem 5.9: The row rank and the column rank of the matrix A are equal.

Definition: The *rank* of the matrix A, written $\operatorname{rank}(A)$, is the common value of its row rank and column rank.

Thus the rank of a matrix gives the maximum number of independent rows, and also the maximum number of independent columns. We can obtain the rank of a matrix as follows.

Basis & Dimension…

APPLICATIONS TO LINEAR EQUATIONS

Consider a system of m linear equations in n unknowns $x_1, \ldots, x_n$ over a field K:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

or the equivalent matrix equation

$$AX = B$$

where $A = (a_{ij})$ is the coefficient matrix, and $X = (x_i)$ and $B = (b_i)$ are the column vectors consisting of the unknowns and of the constants, respectively. Recall that the *augmented matrix* of the system is defined to be the matrix

$$(A, B) = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{pmatrix}$$

Basis & Dimension…

Remark 1. The above linear equations are said to be dependent or independent according as the corresponding vectors, i.e. the rows of the augmented matrix, are dependent or independent.

Remark 2. Two systems of linear equations are equivalent if and only if the corresponding augmented matrices are row equivalent, i.e. have the same row space.

Remark 3. We can always replace a system of equations by a system of independent equations, such as a system in echelon form. The number of independent equations will always be equal to the rank of the augmented matrix.

Observe that the above system is also equivalent to the vector equation

$$x_1\begin{pmatrix} a_{11} \\ a_{21} \\ \cdot\cdot \\ a_{m1} \end{pmatrix} + x_2\begin{pmatrix} a_{12} \\ a_{22} \\ \cdot\cdot \\ a_{m2} \end{pmatrix} + \cdots + x_n\begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdot\cdot \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \cdot\cdot \\ b_m \end{pmatrix}$$

Thus the system $AX = B$ has a solution if and only if the column vector B is a linear combination of the columns of the matrix A, i.e. belongs to the column space of A. This gives us the following basic existence theorem.

Theorem 5.10: The system of linear equations $AX = B$ has a solution if and only if the coefficient matrix A and the augmented matrix (A, B) have the same rank.

Basis & Dimension…

COORDINATES

Let $\{e_1, \ldots, e_n\}$ be a basis of an n-dimensional vector space V over a field K, and let v be any vector in V. Since $\{e_i\}$ generates V, v is a linear combination of the e_i:

$$v = a_1e_1 + a_2e_2 + \cdots + a_ne_n, \qquad a_i \in K$$

Since the e_i are independent, such a representation is unique (Problem 5.7), i.e. the n scalars $a_1, \ldots, a_n$ are completely determined by the vector v and the basis $\{e_i\}$. We call these scalars the *coordinates* of v in $\{e_i\}$, and we call the n-tuple $(a_1, \ldots, a_n)$ the *coordinate vector* of v relative to $\{e_i\}$ and denote it by $[v]_e$ or simply $[v]$:

$$[v]_e = (a_1, a_2, \ldots, a_n)$$

Basis & Dimension…

MAPPINGS

Let A and B be arbitrary sets. Suppose to each $a \in A$ there is assigned a unique element of B; the collection, f, of such assignments is called a *function* or *mapping* (or: *map*) from A into B, and is written

$$f: A \rightarrow B \qquad \text{or} \qquad A \xrightarrow{f} B$$

We write $f(a)$, read "f of a", for the element of B that f assigns to $a \in A$; it is called the *value* of f at a or the *image* of a under f. If A' is any subset of A, then $f(A')$ denotes the set of images of elements of A'; and if B' is any subset of B, then $f^{-1}(B')$ denotes the set of elements of A each of whose image lies in B':

$$f(A') = \{f(a) : a \in A'\} \qquad \text{and} \qquad f^{-1}(B') = \{a \in A : f(a) \in B'\}$$

We call $f(A')$ the *image* of A' and $f^{-1}(B')$ the *inverse image* or *preimage* of B'. In particular, the set of all images, i.e. $f(A)$, is called the *image* (or: *range*) of f. Furthermore, A is called the *domain* of the mapping $f: A \rightarrow B$, and B is called its *co-domain*.

To each mapping $f: A \rightarrow B$ there corresponds the subset of $A \times B$ given by $\{(a, f(a)) : a \in A\}$. We call this set the *graph* of f. Two mappings $f: A \rightarrow B$ and $g: A \rightarrow B$ are defined to be *equal*, written $f = g$, if $f(a) = g(a)$ for every $a \in A$, that is, if they have the same graph. Thus we do not distinguish between a function and its graph. The negation of $f = g$ is written $f \neq g$ and is the statement: there exists an $a \in A$ for which $f(a) \neq g(a)$.

Basis & Dimension…

Theorem 5.12: Let V be an n-dimensional vector space over a field K. Then V and K^n are isomorphic.

Remark: Every $m \times n$ matrix A over a field K determines the mapping $T: K^n \to K^m$ defined by

$$v \mapsto Av$$

where the vectors in K^n and K^m are written as column vectors. For convenience we shall usually denote the above mapping by A, the same symbol used for the matrix.

Linear Mappings

Definition: A mapping $f: A \to B$ is said to be *one-to-one* (or one-one or 1-1) or *injective* if different elements of A have distinct images; that is,

if $a \neq a'$ implies $f(a) \neq f(a')$

or, equivalently, if $f(a) = f(a')$ implies $a = a'$

Definition: A mapping $f: A \to B$ is said to be *onto* (or: f maps A *onto* B) or *surjective* if every $b \in B$ is the image of at least one $a \in A$.

A mapping which is both one-one and onto is said to be *bijective*.

Linear Mappings…

LINEAR MAPPINGS

Let V and U be vector spaces over the same field K. A mapping $F: V \to U$ is called a *linear mapping* (or *linear transformation* or *vector space homomorphism*) if it satisfies the following two conditions:

(1) For any $v, w \in V$, $F(v+w) = F(v) + F(w)$.

(2) For any $k \in K$ and any $v \in V$, $F(kv) = kF(v)$.

In other words, $F: V \to U$ is linear if it "preserves" the two basic operations of a vector space, that of vector addition and that of scalar multiplication.

Substituting $k = 0$ into (2) we obtain $F(0) = 0$. That is, every linear mapping takes the zero vector into the zero vector.

Linear Mappings…

Now for any scalars $a, b \in K$ and any vectors $v, w \in V$ we obtain, by applying both conditions of linearity,

$$F(av + bw) = F(av) + F(bw) = aF(v) + bF(w)$$

More generally, for any scalars $a_i \in K$ and any vectors $v_i \in V$ we obtain the basic property of linear mappings:

$$F(a_1v_1 + a_2v_2 + \cdots + a_nv_n) = a_1F(v_1) + a_2F(v_2) + \cdots + a_nF(v_n)$$

We remark that the condition $F(av + bw) = aF(v) + bF(w)$ completely characterizes linear mappings and is sometimes used as its definition.

Definition: A linear mapping $F: V \to U$ is called an *isomorphism* if it is one-to-one. The vector spaces V, U are said to be *isomorphic* if there is an isomorphism of V onto U.

Linear Mappings…

Theorem 6.2: Let V and U be vector spaces over a field K. Let $\{v_1, v_2, \ldots, v_n\}$ be a basis of V and let $u_1, u_2, \ldots, u_n$ be any vectors in U. Then there exists a unique linear mapping $F: V \to U$ such that $F(v_1) = u_1$, $F(v_2) = u_2$, $\ldots$, $F(v_n) = u_n$.

We emphasize that the vectors $u_1, \ldots, u_n$ in the preceding theorem are completely arbitrary; they may be linearly dependent or they may even be equal to each other.

KERNEL AND IMAGE OF A LINEAR MAPPING

We begin by defining two concepts.

Definition: Let $F: V \to U$ be a linear mapping. The *image* of F, written $\operatorname{Im} F$, is the set of image points in U:

$$\operatorname{Im} F = \{u \in U : F(v) = u \text{ for some } v \in V\}$$

The *kernel* of F, written $\operatorname{Ker} F$, is the set of elements in V which map into $0 \in U$:

$$\operatorname{Ker} F = \{v \in V : F(v) = 0\}$$

The following theorem is easily proven (Problem 6.22).

Theorem 6.3: Let $F: V \to U$ be a linear mapping. Then the image of F is a subspace of U and the kernel of F is a subspace of V.

Linear Mappings…

OPERATIONS WITH LINEAR MAPPINGS

We are able to combine linear mappings in various ways to obtain new linear mappings. These operations are very important and shall be used throughout the text.

Suppose $F: V \to U$ and $G: V \to U$ are linear mappings of vector spaces over a field K. We define the sum $F + G$ to be the mapping from V into U which assigns $F(v) + G(v)$ to $v \in V$:

$$(F+G)(v) = F(v) + G(v)$$

Furthermore, for any scalar $k \in K$, we define the product kF to be the mapping from V into U which assigns $kF(v)$ to $v \in V$:

$$(kF)(v) = kF(v)$$

We show that if F and G are linear, then $F + G$ and kF are also linear. We have, for any vectors $v, w \in V$ and any scalars $a, b \in K$,

$$\begin{aligned}(F+G)(av+bw) &= F(av+bw) + G(av+bw) \\ &= aF(v) + bF(w) + aG(v) + bG(w) \\ &= a(F(v)+G(v)) + b(F(w)+G(w)) \\ &= a(F+G)(v) + b(F+G)(w)\end{aligned}$$

and

$$\begin{aligned}(kF)(av+bw) &= kF(av+bw) = k(aF(v)+bF(w)) \\ &= akF(v) + bkF(w) = a(kF)(v) + b(kF)(w)\end{aligned}$$

Thus $F + G$ and kF are linear.

The following theorem applies.

Linear Mappings…

Theorem 6.6: Let V and U be vector spaces over a field K. Then the collection of all linear mappings from V into U with the above operations of addition and scalar multiplication form a vector space over K.

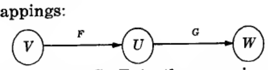
The space in the above theorem is usually denoted by

$$\mathrm{Hom}\,(V, U)$$

Here Hom comes from the word homomorphism. In the case that V and U are of finite dimension, we have the following theorem.

Theorem 6.7: Suppose $\dim V = m$ and $\dim U = n$. Then $\dim \mathrm{Hom}(V, U) = mn$.

Now suppose that V, U and W are vector spaces over the same field K, and that $F: V \to U$ and $G: U \to W$ are linear mappings:



Recall that the composition function $G \circ F$ is the mapping from V into W defined by $(G \circ F)(v) = G(F(v))$. We show that $G \circ F$ is linear whenever F and G are linear. We have, for any vectors $v, w \in V$ and any scalars $a, b \in K$,

$$\begin{aligned}(G \circ F)(av + bw) &= G(F(av + bw)) = G(aF(v) + bF(w)) \\ &= aG(F(v)) + bG(F(w)) = a(G \circ F)(v) + b(G \circ F)(w)\end{aligned}$$

That is, $G \circ F$ is linear.

Linear Mappings…

Theorem 6.8: Let V, U and W be vector spaces over K. Let F, F' be linear mappings from V into U and G, G' linear mappings from U into W, and let $k \in K$. Then:

(i) $G \circ (F + F') = G \circ F + G \circ F'$

(ii) $(G + G') \circ F = G \circ F + G' \circ F$

(iii) $k(G \circ F) = (kG) \circ F = G \circ (kF)$.

Linear Mappings…

ALGEBRA OF LINEAR OPERATORS

Let V be a vector space over a field K. We now consider the special case of linear mappings $T: V \to V$, i.e. from V into itself. They are also called *linear operators* or *linear transformations* on V. We will write $A(V)$, instead of $\text{Hom}(V, V)$, for the space of all such mappings.

By Theorem 6.6, $A(V)$ is a vector space over K; it is of dimension n^2 if V is of dimension n. Now if $T, S \in A(V)$, then the composition $S \circ T$ exists and is also a linear mapping from V into itself, i.e. $S \circ T \in A(V)$. Thus we have a "multiplication" defined in $A(V)$. (We shall write ST for $S \circ T$ in the space $A(V)$.)

We remark that an *algebra* A over a field K is a vector space over K in which an operation of multiplication is defined satisfying, for every $F, G, H \in A$ and every $k \in K$,

(i) $F(G+H) = FG + FH$

(ii) $(G+H)F = GF + HF$

(iii) $k(GF) = (kG)F = G(kF)$.

Linear Mappings…

If the associative law also holds for the multiplication, i.e. if for every $F, G, H \in A$,

(iv) $(FG)H = F(GH)$

then the algebra A is said to be *associative*. Thus by Theorems 6.8 and 6.1, $A(V)$ is an associative algebra over K with respect to composition of mappings; hence it is frequently called the *algebra of linear operators* on V.

Observe that the identity mapping $I: V \to V$ belongs to $A(V)$. Also, for any $T \in A(V)$, we have $TI = IT = T$. We note that we can also form "powers" of T; we use the notation $T^2 = T \circ T$, $T^3 = T \circ T \circ T$, ... Furthermore, for any polynomial

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n, \qquad a_i \in K$$

we can form the operator $p(T)$ defined by

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_n T^n$$

(For a scalar $k \in K$, the operator kI is frequently denoted by simply k.) In particular, if $p(T) = 0$, the zero mapping, then T is said to be a *zero* of the polynomial $p(x)$.

Linear Mappings…

INVERTIBLE OPERATORS

A linear operator $T: V \to V$ is said to be *invertible* if it has an inverse, i.e. if there exists $T^{-1} \in A(V)$ such that $TT^{-1} = T^{-1}T = I$.

Now T is invertible if and only if it is one-one and onto. Thus in particular, if T is invertible then only $0 \in V$ can map into itself, i.e. T is nonsingular. On the other hand, suppose T is nonsingular, i.e. $\operatorname{Ker} T = \{0\}$. Recall (page 127) that T is also one-one. Moreover, assuming V has finite dimension, we have, by Theorem 6.4,

$$\begin{aligned} \dim V &= \dim(\operatorname{Im} T) + \dim(\operatorname{Ker} T) = \dim(\operatorname{Im} T) + \dim(\{0\}) \\ &= \dim(\operatorname{Im} T) + 0 = \dim(\operatorname{Im} T) \end{aligned}$$

Then $\operatorname{Im} T = V$, i.e. the image of T is V; thus T is onto. Hence T is both one-one and onto and so is invertible. We have just proven

Theorem 6.9: A linear operator $T: V \to V$ on a vector space of finite dimension is invertible if and only if it is nonsingular.

Linear Mappings…

We now give an important application of the above theorem to systems of linear equations over K. Consider a system with the same number of equations as unknowns, say n. We can represent this system by the matrix equation

$$Ax = b \qquad (*)$$

where A is an n-square matrix over K which we view as a linear operator on K^n. Suppose the matrix A is *nonsingular*, i.e. the matrix equation $Ax = 0$ has only the zero solution. Then, by Theorem 6.9, the linear mapping A is one-to-one and onto. This means that the system (*) has a unique solution for any $b \in K^n$. On the other hand, suppose the matrix A is *singular*, i.e. the matrix equation $Ax = 0$ has a nonzero solution. Then the linear mapping A is not onto. This means that there exist $b \in K^n$ for which (*) does not have a solution. Furthermore, if a solution exists it is not unique. Thus we have proven the following fundamental result:

Theorem 6.10: Consider the following system of linear equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

Linear Mappings…

(i) If the corresponding homogeneous system has only the zero solution, then the above system has a unique solution for any values of the b_i.

(ii) If the corresponding homogeneous system has a nonzero solution, then: (i) there are values for the b_i for which the above system does not have a solution; (ii) whenever a solution of the above system exists, it is not unique.

Matrices & Linear Operators

Matrices and Linear Operators

INTRODUCTION

Suppose $\{e_1, \ldots, e_n\}$ is a basis of a vector space V over a field K and, for $v \in V$, suppose $v = a_1e_1 + a_2e_2 + \cdots + a_ne_n$. Then the coordinate vector of v relative to $\{e_i\}$, which we write as a column vector unless otherwise specified or implied, is

$$[v]_e = \begin{pmatrix} a_1 \\ a_2 \\ \cdot\ \cdot \\ a_n \end{pmatrix}$$

Recall that the mapping $v \mapsto [v]_e$, determined by the basis $\{e_i\}$, is an isomorphism from V onto the space K^n.

In this chapter we show that there is also an isomorphism, determined by the basis $\{e_i\}$, from the algebra $A(V)$ of linear operators on V onto the algebra $\mathcal{A}$ of n-square matrices over K.

A similar result also holds for linear mappings $F: V \to U$, from one space into another.

Matrices & Linear Operators…

MATRIX REPRESENTATION OF A LINEAR OPERATOR

Let T be a linear operator on a vector space V over a field K and suppose $\{e_1, \ldots, e_n\}$ is a basis of V. Now $T(e_1), \ldots, T(e_n)$ are vectors in V and so each is a linear combination of the elements of the basis $\{e_i\}$:

$$\begin{aligned} T(e_1) &= a_{11}e_1 + a_{12}e_2 + \cdots + a_{1n}e_n \\ T(e_2) &= a_{21}e_1 + a_{22}e_2 + \cdots + a_{2n}e_n \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ T(e_n) &= a_{n1}e_1 + a_{n2}e_2 + \cdots + a_{nn}e_n \end{aligned}$$

The following definition applies.

Definition: The transpose of the above matrix of coefficients, denoted by $[T]_e$ or $[T]$, is called the *matrix representation of T relative to the basis $\{e_i\}$* or simply the *matrix of T in the basis $\{e_i\}$*:

$$[T]_e = \begin{pmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ a_{12} & a_{22} & \ldots & a_{n2} \\ \ldots & \ldots & \ldots & \ldots \\ a_{1n} & a_{2n} & \ldots & a_{nn} \end{pmatrix}$$

Matrices & Linear Operators…

Remark: Recall that any n-square matrix A over K defines a linear operator on K^n by the map $v \mapsto Av$ (where v is written as a column vector). We show (Problem 7.7) that the matrix representation of this operator is precisely the matrix A if we use the usual basis of K^n.

Our first theorem tells us that the "action" of an operator T on a vector v is preserved by its matrix representation:

Theorem 7.1: Let $\{e_1, \ldots, e_n\}$ be a basis of V and let T be any operator on V. Then, for any vector $v \in V$, $[T]_e [v]_e = [T(v)]_e$.

That is, if we multiply the coordinate vector of v by the matrix representation of T, then we obtain the coordinate vector of $T(v)$.

Matrices & Linear Operators…

Now we have associated a matrix $[T]_e$ to each T in $A(V)$, the algebra of linear ope on V. By our first theorem the action of an individual operator T is preserved b representation. The next two theorems tell us that the three basic operations with operators

(i) addition, (ii) scalar multiplication, (iii) composition

are also preserved.

Theorem 7.2: Let $\{e_1, \ldots, e_n\}$ be a basis of V over K, and let $\mathcal{A}$ be the algel n-square matrices over K. Then the mapping $T \mapsto [T]_e$ is a vector isomorphism from $A(V)$ onto $\mathcal{A}$. That is, the mapping is one-one an and, for any $S, T \in A(V)$ and any $k \in K$,

$$[T+S]_e = [T]_e + [S]_e \quad \text{and} \quad [kT]_e = k[T]_e$$

Theorem 7.3: For any operators $S, T \in A(V)$, $[ST]_e = [S]_e [T]_e$.

We illustrate the above theorems in the case $\dim V = 2$. Suppose $\{e_1, e_2\}$ is a bε V, and T and S are operators on V for which

$$T(e_1) = a_1e_1 + a_2e_2 \qquad S(e_1) = c_1e_1 + c_2e_2$$
$$T(e_2) = b_1e_1 + b_2e_2\,, \qquad S(e_2) = d_1e_1 + d_2e_2$$

Then $$[T]_e = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \quad \text{and} \quad [S]_e = \begin{pmatrix} c_1 & d_1 \\ c_2 & d_2 \end{pmatrix}$$

Matrices & Linear Operators…

Now we have
$$(T+S)(e_1) = T(e_1) + S(e_1) = a_1e_1 + a_2e_2 + c_1e_1 + c_2e_2 = (a_1+c_1)e_1 + (a_2+c_2)e_2$$
$$(T+S)(e_2) = T(e_2) + S(e_2) = b_1e_1 + b_2e_2 + d_1e_1 + d_2e_2 = (b_1+d_1)e_1 + (b_2+d_2)e_2$$

Thus
$$[T+S]_e = \begin{pmatrix} a_1+c_1 & b_1+d_1 \\ a_2+c_2 & b_2+d_2 \end{pmatrix} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} + \begin{pmatrix} c_1 & d_1 \\ c_2 & d_2 \end{pmatrix} = [T]_e + [S]_e$$

Also, for $k \in K$, we have
$$(kT)(e_1) = kT(e_1) = k(a_1e_1 + a_2e_2) = ka_1e_1 + ka_2e_2$$
$$(kT)(e_2) = kT(e_2) = k(b_1e_1 + b_2e_2) = kb_1e_1 + kb_2e_2$$

Hence $$[kT]_e = \begin{pmatrix} ka_1 & kb_1 \\ ka_2 & kb_2 \end{pmatrix} = k\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = k[T]_e$$

Finally, we have
$$(ST)(e_1) = S(T(e_1)) = S(a_1e_1 + a_2e_2) = a_1S(e_1) + a_2S(e_2)$$
$$= a_1(c_1e_1 + c_2e_2) + a_2(d_1e_1 + d_2e_2)$$
$$= (a_1c_1 + a_2d_1)e_1 + (a_1c_2 + a_2d_2)e_2$$
$$(ST)(e_2) = S(T(e_2)) = S(b_1e_1 + b_2e_2) = b_1S(e_1) + b_2S(e_2)$$
$$= b_1(c_1e_1 + c_2e_2) + b_2(d_1e_1 + d_2e_2)$$
$$= (b_1c_1 + b_2d_1)e_1 + (b_1c_2 + b_2d_2)e_2$$

Accordingly,
$$[ST]_e = \begin{pmatrix} a_1c_1 + a_2d_1 & b_1c_1 + b_2d_1 \\ a_1c_2 + a_2d_2 & b_1c_2 + b_2d_2 \end{pmatrix} = \begin{pmatrix} c_1 & d_1 \\ c_2 & d_2 \end{pmatrix}\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = [S]_e[T]_e$$

Matrices & Linear Operators…

Theorem 7.4: Let P be the transition matrix from a basis $\{e_i\}$ to a basis $\{f_i\}$ in a vector space V. Then, for any vector $v \in V$, $P[v]_f = [v]_e$. Hence $[v]_f = P^{-1}[v]_e$.

We emphasize that even though P is called the transition matrix from the old basis $\{e_i\}$ to the new basis $\{f_i\}$, its effect is to transform the coordinates of a vector in the new basis $\{f_i\}$ back to the coordinates in the old basis $\{e_i\}$.

Matrices & Linear Operators…

We illustrate the above theorem in the case $\dim V = 3$. Suppose P is the transition matrix from a basis $\{e_1, e_2, e_3\}$ of V to a basis $\{f_1, f_2, f_3\}$ of V; say,

$$\begin{aligned} f_1 &= a_1e_1 + a_2e_2 + a_3e_3 \\ f_2 &= b_1e_1 + b_2e_2 + b_3e_3\,. \\ f_3 &= c_1e_1 + c_2e_2 + c_3e_3 \end{aligned} \qquad \text{Hence} \quad P = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

Now suppose $v \in V$ and, say, $v = k_1f_1 + k_2f_2 + k_3f_3$. Then, substituting for the f_i from above, we obtain

$$\begin{aligned} v &= k_1(a_1e_1 + a_2e_2 + a_3e_3) + k_2(b_1e_1 + b_2e_2 + b_3e_3) + k_3(c_1e_1 + c_2e_2 + c_3e_3) \\ &= (a_1k_1 + b_1k_2 + c_1k_3)e_1 + (a_2k_1 + b_2k_2 + c_2k_3)e_2 + (a_3k_1 + b_3k_2 + c_3k_3)e_3 \end{aligned}$$

Thus

$$[v]_f = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} \qquad \text{and} \qquad [v]_e = \begin{pmatrix} a_1k_1 + b_1k_2 + c_1k_3 \\ a_2k_1 + b_2k_2 + c_2k_3 \\ a_3k_1 + b_3k_2 + c_3k_3 \end{pmatrix}$$

Accordingly,

$$P[v]_f = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}\begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = \begin{pmatrix} a_1k_1 + b_1k_2 + c_1k_3 \\ a_2k_1 + b_2k_2 + c_2k_3 \\ a_3k_1 + b_3k_2 + c_3k_3 \end{pmatrix} = [v]_e$$

Also, multiplying the above equation by P^{-1}, we have

$$P^{-1}[v]_e = P^{-1}P[v]_f = I[v]_f = [v]_f$$

Matrices & Linear Operators…

Theorem 7.5: Let P be the transition matrix from a basis $\{e_i\}$ to a basis $\{f_i\}$ in a vector space V. Then for any linear operator T on V, $[T]_f = P^{-1}[T]_e P$.

Example 7.7: Let T be the linear operator on $\mathbf{R}^2$ defined by $T(x, y) = (4x - 2y, 2x + y)$. Then for the bases of $\mathbf{R}^2$ in Example 7.5, we have

$$T(e_1) = T(1,0) = (4,2) = 4(1,0) + 2(0,1) = 4e_1 + 2e_2$$
$$T(e_2) = T(0,1) = (-2,1) = -2(1,0) + (0,1) = -2e_1 + e_2$$

Accordingly, $$[T]_e = \begin{pmatrix} 4 & -2 \\ 2 & 1 \end{pmatrix}$$

Matrices & Linear Operators…

CHANGE OF BASIS

We have shown that we can represent vectors by n-tuples (column vectors) and linear operators by matrices once we have selected a basis. We ask the following natural question: How does our representation change if we select another basis? In order to answer this question, we first need a definition.

Definition: Let $\{e_1, \ldots, e_n\}$ be a basis of V and let $\{f_1, \ldots, f_n\}$ be another basis. Suppose

$$f_1 = a_{11}e_1 + a_{12}e_2 + \cdots + a_{1n}e_n$$
$$f_2 = a_{21}e_1 + a_{22}e_2 + \cdots + a_{2n}e_n$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$f_n = a_{n1}e_1 + a_{n2}e_2 + \cdots + a_{nn}e_n$$

Then the transpose P of the above matrix of coefficients is termed the *transition matrix* from the "old" basis $\{e_i\}$ to the "new" basis $\{f_i\}$:

$$P = \begin{pmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ a_{12} & a_{22} & \ldots & a_{n2} \\ \ldots & \ldots & \ldots & \ldots \\ a_{1n} & a_{2n} & \ldots & a_{nn} \end{pmatrix}$$

We comment that since the vectors $f_1, \ldots, f_n$ are linearly independent, the matrix P is invertible (Problem 5.47). In fact, its inverse P^{-1} is the transition matrix from the basis $\{f_i\}$ back to the basis $\{e_i\}$.

Matrices & Linear Operators…

SIMILARITY

Suppose A and B are square matrices for which there exists an invertible matrix P such that $B = P^{-1}AP$. Then B is said to be *similar* to A or is said to be obtained from A by a *similarity transformation*. We show (Problem 7.22) that similarity of matrices is an equivalence relation. Thus by Theorem 7.5 and the above remark, we have the following basic result.

Theorem 7.6: Two matrices A and B represent the same linear operator T if and only if they are similar to each other.

That is, all the matrix representations of the linear operator T form an equivalence class of similar matrices.

A linear operator T is said to be *diagonalizable* if for some basis $\{e_i\}$ it is represented by a diagonal matrix; the basis $\{e_i\}$ is then said to *diagonalize* T. The preceding theorem gives us the following result.

Theorem 7.7: Let A be a matrix representation of a linear operator T. Then T is diagonalizable if and only if there exists an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix.

That is, T is diagonalizable if and only if its matrix representation can be diagonalized by a similarity transformation.

Matrices & Linear Operators…

We emphasize that not every operator is diagonalizable. However, we will show (Chapter 10) that every operator T can be represented by certain "standard" matrices called its *normal* or *canonical* forms. We comment now that that discussion will require some theory of fields, polynomials and determinants.

Now suppose f is a function on square matrices which assigns the same value to similar matrices; that is, $f(A) = f(B)$ whenever A is similar to B. Then f induces a function, also denoted by f, on linear operators T in the following natural way: $f(T) = f([T]_e)$, where $\{e_i\}$ is any basis. The function is well-defined by the preceding theorem.

The *determinant* is perhaps the most important example of the above type of functions. Another important example follows.

Example 7.8: The *trace* of a square matrix $A = (a_{ij})$, written $\operatorname{tr}(A)$, is defined to be the sum of its diagonal elements:

$$\operatorname{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

We show (Problem 7.22) that similar matrices have the same trace. Thus we can speak of the trace of a linear operator T; it is the trace of any one of its matrix representations: $\operatorname{tr}(T) = \operatorname{tr}([T]_e)$.

Matrices & Linear Operators…

MATRICES AND LINEAR MAPPINGS

We now consider the general case of linear mappings from one space into another. Let V and U be vector spaces over the same field K and, say, $\dim V = m$ and $\dim U = n$. Furthermore, let $\{e_1, \ldots, e_m\}$ and $\{f_1, \ldots, f_n\}$ be arbitrary but fixed bases of V and U respectively.

Suppose $F: V \to U$ is a linear mapping. Then the vectors $F(e_1), \ldots, F(e_m)$ belong to U and so each is a linear combination of the f_i:

$$\begin{aligned} F(e_1) &= a_{11}f_1 + a_{12}f_2 + \cdots + a_{1n}f_n \\ F(e_2) &= a_{21}f_1 + a_{22}f_2 + \cdots + a_{2n}f_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ F(e_m) &= a_{m1}f_1 + a_{m2}f_2 + \cdots + a_{mn}f_n \end{aligned}$$

The transpose of the above matrix of coefficients, denoted by $[F]_e^f$ is called the *matrix representation* of F relative to the bases $\{e_i\}$ and $\{f_i\}$, or the matrix of F in the bases $\{e_i\}$ and $\{f_i\}$:

$$[F]_e^f = \begin{pmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \ldots & \ldots & \ldots & \ldots \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{pmatrix}$$

The following theorems apply.

Theorem 7.8: For any vector $v \in V$, $[F]_e^f [v]_e = [F(v)]_f$.

That is, multiplying the coordinate vector of v in the basis $\{e_i\}$ by the matrix $[F]_e^f$, we obtain the coordinate vector of $F(v)$ in the basis $\{f_i\}$.

Matrices & Linear Operators…

Theorem 7.9: The mapping $F \mapsto [F]_e^f$ is an isomorphism from $\operatorname{Hom}(V, U)$ onto the vector space of $n \times m$ matrices over K. That is, the mapping is one-one and onto and, for any $F, G \in \operatorname{Hom}(V, U)$ and any $k \in K$,

$$[F+G]_e^f = [F]_e^f + [G]_e^f \quad \text{and} \quad [kF]_e^f = k[F]_e^f$$

Remark: Recall that any $n \times m$ matrix A over K has been identified with the linear mapping from K^m into K^n given by $v \mapsto Av$. Now suppose V and U are vector spaces over K of dimensions m and n respectively, and suppose $\{e_i\}$ is a basis of V and $\{f_i\}$ is a basis of U. Then in view of the preceding theorem, we shall also identify A with the linear mapping $F: V \to U$ given by $[F(v)]_f = A[v]_e$. We comment that if other bases of V and U are given, then A is identified with another linear mapping from V into U.

Theorem 7.10: Let $\{e_i\}$, $\{f_i\}$ and $\{g_i\}$ be bases of V, U and W respectively. Let $F: V \to U$ and $G: U \to W$ be linear mappings. Then

$$[G \circ F]_e^g = [G]_f^g [F]_e^f$$

That is, relative to the appropriate bases, the matrix representation of the composition of two linear mappings is equal to the product of the matrix representations of the individual mappings.

We lastly show how the matrix representation of a linear mapping $F: V \to U$ is affected when new bases are selected.

Theorem 7.11: Let P be the transition matrix from a basis $\{e_i\}$ to a basis $\{e_i'\}$ in V, and let Q be the transition matrix from a basis $\{f_i\}$ to a basis $\{f_i'\}$ in U. Then for any linear mapping $F: V \to U$,

$$[F]_{e'}^{f'} = Q^{-1} [F]_e^f P$$

Matrices & Linear Operators…

Thus in particular,

$$[F]^{f'}_{e} = Q^{-1}[F]^{f}_{e}$$

i.e. when the change of basis only takes place in U; and

$$[F]^{f}_{e'} = [F]^{f}_{e} P$$

i.e. when the change of basis only takes place in V.

Note that Theorems 7.1, 7.2, 7.3 and 7.5 are special cases of Theorems 7.8, 7.9, 7.10 and 7.11 respectively.

The next theorem shows that every linear mapping from one space into another can be represented by a very simple matrix.

Theorem 7.12: Let $F: V \to U$ be linear and, say, rank $F = r$. Then there exist bases of V and of U such that the matrix representation of F has the form

$$A = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

where I is the r-square identity matrix. We call A the *normal* or *canonical* form of F.

Matrices & Linear Operators…

WARNING

As noted previously, some texts write the operator symbol T to the right of the vector v on which it acts, that is,

$$vT \quad \text{instead of} \quad T(v)$$

In such texts, vectors and operators are represented by n-tuples and matrices which are the transposes of those appearing here. That is, if

$$v = k_1e_1 + k_2e_2 + \cdots + k_ne_n$$

then they write

$$[v]_e = (k_1, k_2, \ldots, k_n) \quad \text{instead of} \quad [v]_e = \begin{pmatrix} k_1 \\ k_2 \\ \cdot\cdot \\ k_n \end{pmatrix}$$

And if

$$T(e_1) = a_1e_1 + a_2e_2 + \cdots + a_ne_n$$
$$T(e_2) = b_1e_1 + b_2e_2 + \cdots + b_ne_n$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$T(e_n) = c_1e_1 + c_2e_2 + \cdots + c_ne_n$$

then they write

$$[T]_e = \begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ b_1 & b_2 & \ldots & b_n \\ \ldots & \ldots & \ldots & \ldots \\ c_1 & c_2 & \ldots & c_n \end{pmatrix} \quad \text{instead of} \quad [T]_e = \begin{pmatrix} a_1 & b_1 & \ldots & c_1 \\ a_2 & b_2 & \ldots & c_2 \\ \ldots & \ldots & \ldots & \ldots \\ a_n & b_n & \ldots & c_n \end{pmatrix}$$

This is also true for the transition matrix from one basis to another and for matrix representations of linear mappings $F: V \to U$. We comment that such texts have theorems which are analogous to the ones appearing here.

Inner Product Spaces & Operators

INNER PRODUCT SPACES

We begin with a definition.

Definition: Let V be a (real or complex) vector space over K. Suppose to each pair of vectors $u, v \in V$ there is assigned a scalar $\langle u, v\rangle \in K$. This mapping is called an *inner product* in V if it satisfies the following axioms:

$[I_1]$ $\langle au_1 + bu_2, v\rangle = a\langle u_1, v\rangle + b\langle u_2, v\rangle$

$[I_2]$ $\langle u, v\rangle = \overline{\langle v, u\rangle}$

$[I_3]$ $\langle u, u\rangle \geqq 0$; and $\langle u, u\rangle = 0$ if and only if $u = 0$.

The vector space V with an inner product is called an *inner product space*.

Observe that $\langle u, u\rangle$ is always real by $[I_2]$, and so the inequality relation in $[I_3]$ makes sense. We also use the notation

$$||u|| = \sqrt{\langle u, u\rangle}$$

This nonnegative real number $||u||$ is called the *norm* or *length* of u. Also, using $[I_1]$ and $[I_2]$ we obtain (Problem 13.1) the relation

$$\langle u, av_1 + bv_2\rangle = \bar{a}\langle u, v_1\rangle + \bar{b}\langle u, v_2\rangle$$

Inner Product Spaces & Operators…

Remark 1: If $||v|| = 1$, i.e. if $\langle v, v\rangle = 1$, then v is called a *unit* vector or is said to be *normalized*. We note that every nonzero vector $u \in V$ can be *normalized* by setting $v = u/||u||$.

Remark 2: The nonnegative real number $d(u, v) = ||v - u||$ is called the *distance* between u and v; this function does satisfy the axioms of a metric space (see Problem 13.51).

Inner Product Spaces & Operators…

CAUCHY-SCHWARZ INEQUALITY

The following formula, called the Cauchy-Schwarz inequality, is used in many branches of mathematics.

Theorem 13.1: (Cauchy-Schwarz): For any vectors $u, v \in V$,

$$|\langle u, v\rangle| \leq \|u\|\,\|v\|$$

Next we examine this inequality in specific cases.

Example 13.6: Consider any complex numbers $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbf{C}$. Then by the Cauchy-Schwarz inequality,

$$(a_1\bar{b}_1 + \cdots + a_n\bar{b}_n)^2 \leq (|a_1|^2 + \cdots + |a_n|^2)(|b_1|^2 + \cdots + |b_n|^2)$$

that is, $$(u \cdot v)^2 \leq \|u\|^2\,\|v\|^2$$

where $u = (a_i)$ and $v = (b_i)$.

Inner Product Spaces & Operators…

ORTHOGONALITY

Let V be an inner product space. The vectors $u, v \in V$ are said to be *orthogonal* if $\langle u, v\rangle = 0$. The relation is clearly symmetric; that is, if u is orthogonal to v, then $\langle v, u\rangle = \overline{\langle u, v\rangle} = \bar{0} = 0$ and so v is orthogonal to u. We note that $0 \in V$ is orthogonal to every $v \in V$ for

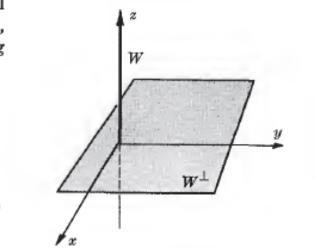
$$\langle 0, v\rangle = \langle 0v, v\rangle = 0\langle v, v\rangle = 0$$

Conversely, if u is orthogonal to every $v \in V$, then $\langle u, u\rangle = 0$ and hence $u = 0$ by $[I_3]$.

Inner Product Spaces & Operators…

Theorem 13.2: Let W be a subspace of V. Then V is the direct sum of W and $W^\perp$, i.e. $V = W \oplus W^\perp$.

Now if W is a subspace of V, then $V = W \oplus W^\perp$ by the above theorem; hence there is a unique projection $E_W: V \to V$ with image W and kernel $W^\perp$. That is, if $v \in V$ and $v = w + w'$, where $w \in W$, $w' \in W^\perp$, then E_W is defined by $E_W(v) = w$. This mapping E_W is called the *orthogonal projection* of V onto W.



Example 13.8: Let W be the z axis in $\mathbf{R}^3$, i.e.

$$W = \{(0,0,c) : c \in \mathbf{R}\}$$

Then $W^\perp$ is the xy plane, i.e.

$$W^\perp = \{(a,b,0) : a, b \in \mathbf{R}\}$$

As noted previously, $\mathbf{R}^3 = W \oplus W^\perp$. The orthogonal projection E of $\mathbf{R}^3$ onto W is given by $E(x,y,z) = (0,0,z)$.

Inner Product Spaces & Operators…

ORTHONORMAL SETS

A set $\{u_i\}$ of vectors in V is said to be *orthogonal* if its distinct elements are orthogonal, i.e. if $\langle u_i, u_j \rangle = 0$ for $i \neq j$. In particular, the set $\{u_i\}$ is said to be *orthonormal* if it is orthogonal and if each u_i has length 1, that is, if

$$\langle u_i, u_j \rangle = \delta_{ij} = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}$$

An orthonormal set can always be obtained from an orthogonal set of nonzero vectors by normalizing each vector.

Example 13.10: Consider the usual basis of Euclidean 3-space $\mathbf{R}^3$:

$$\{e_1 = (1,0,0),\ e_2 = (0,1,0),\ e_3 = (0,0,1)\}$$

It is clear that

$$\langle e_1, e_1 \rangle = \langle e_2, e_2 \rangle = \langle e_3, e_3 \rangle = 1 \quad \text{and} \quad \langle e_i, e_j \rangle = 0 \ \text{ for } i \neq j$$

That is, $\{e_1, e_2, e_3\}$ is an orthonormal basis of $\mathbf{R}^3$. More generally, the usual basis of $\mathbf{R}^n$ or of $\mathbf{C}^n$ is orthonormal for every n.

Example 13.11: Let V be the vector space of real continuous functions on the interval $-\pi \leq t \leq \pi$ with inner product defined by $\langle f, g \rangle = \int_{-\pi}^{\pi} f(t)\, g(t)\, dt$. The following is a classical example of an orthogonal subset of V:

$$\{1, \cos t, \cos 2t, \ldots, \sin t, \sin 2t, \ldots\}$$

The above orthogonal set plays a fundamental role in the theory of Fourier series.

Inner Product Spaces & Operators…

GRAM-SCHMIDT ORTHOGONALIZATION PROCESS

Orthonormal bases play an important role in inner product spaces. The next theorem shows that such a basis always exists; its proof uses the celebrated Gram-Schmidt orthogonalization process.

Theorem 13.4: Let $\{v_1, \ldots, v_n\}$ be an arbitrary basis of an inner product space V. Then there exists an orthonormal basis $\{u_1, \ldots, u_n\}$ of V such that the transition matrix from $\{v_i\}$ to $\{u_i\}$ is triangular; that is, for $i = 1, \ldots, n$,

$$u_i = a_{i1}v_1 + a_{i2}v_2 + \cdots + a_{ii}v_i$$

Inner Product Spaces & Operators…

Proof. We set $u_1 = v_1/\|v_1\|$; then $\{u_1\}$ is orthonormal. We next set

$$w_2 = v_2 - \langle v_2, u_1\rangle u_1 \quad \text{and} \quad u_2 = w_2/\|w_2\|$$

By Lemma 13.3, w_2 (and hence u_2) is orthogonal to u_1; then $\{u_1, u_2\}$ is orthonormal. We next set

$$w_3 = v_3 - \langle v_3, u_1\rangle u_1 - \langle v_3, u_2\rangle u_2 \quad \text{and} \quad u_3 = w_3/\|w_3\|$$

Again, by Lemma 13.3, w_3 (and hence u_3) is orthogonal to u_1 and u_2; then $\{u_1, u_2, u_3\}$ is orthonormal. In general, after obtaining $\{u_1, \ldots, u_i\}$ we set

$$w_{i+1} = v_{i+1} - \langle v_{i+1}, u_1\rangle u_1 - \cdots - \langle v_{i+1}, u_i\rangle u_i \quad \text{and} \quad u_{i+1} = w_{i+1}/\|w_{i+1}\|$$

(Note that $w_{i+1} \neq 0$ because $v_{i+1} \notin L(v_1, \ldots, v_i) = L(u_1, \ldots, u_i)$.) As above, $\{u_1, \ldots, u_{i+1}\}$ is also orthonormal. By induction we obtain an orthonormal set $\{u_1, \ldots, u_n\}$ which is independent and hence a basis of V. The specific construction guarantees that the transition matrix is indeed triangular.

Inner Product Spaces & Operators…

Example 13.12: Consider the following basis of Euclidean space $\mathbf{R}^3$:

$$\{v_1 = (1,1,1),\ v_2 = (0,1,1),\ v_3 = (0,0,1)\}$$

We use the Gram-Schmidt orthogonalization process to transform $\{v_i\}$ into an orthonormal basis $\{u_i\}$. First we normalize v_1, i.e. we set

$$u_1 = \frac{v_1}{\|v_1\|} = \frac{(1,1,1)}{\sqrt{3}} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$$

Next we set

$$w_2 = v_2 - \langle v_2, u_1\rangle u_1 = (0,1,1) - \frac{2}{\sqrt{3}}\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

and then we normalize w_2, i.e. we set

$$u_2 = \frac{w_2}{\|w_2\|} = \left(-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)$$

Finally we set

$$\begin{aligned} w_3 &= v_3 - \langle v_3, u_1\rangle u_1 - \langle v_3, u_2\rangle u_2 \\ &= (0,0,1) - \frac{1}{\sqrt{3}}\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) - \frac{1}{\sqrt{6}}\left(-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right) = \left(0, -\frac{1}{2}, \frac{1}{2}\right) \end{aligned}$$

and then we normalize w_3:

$$u_3 = \frac{w_3}{\|w_3\|} = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

The required orthonormal basis of $\mathbf{R}^3$ is

$$\left\{u_1 = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right),\ u_2 = \left(-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right),\ u_3 = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\right\}$$

Inner Product Spaces & Operators…

ORTHOGONAL AND UNITARY OPERATORS

Let U be a linear operator on a finite dimensional inner product space V. As defined above, if

$$U^* = U^{-1} \quad \text{or equivalently} \quad UU^* = U^*U = I$$

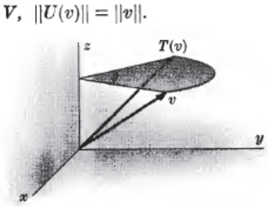
then U is said to be *orthogonal* or *unitary* according as the underlying field is real or complex. The next theorem gives alternate characterizations of these operators.

Theorem 13.9: The following conditions on an operator U are equivalent:
(i) $U^* = U^{-1}$, that is, $UU^* = U^*U = I$.
(ii) *U preserves inner products*, i.e. for every $v, w \in V$, $\langle U(v), U(w)\rangle = \langle v, w\rangle$
(iii) *U preserves lengths*, i.e. for every $v \in V$, $\|U(v)\| = \|v\|$.

Example 13.14: Let $T: \mathbf{R}^3 \to \mathbf{R}^3$ be the linear operator which rotates each vector about the z axis by a fixed angle θ:

$$T(x,y,z) = (x\cos\theta - y\sin\theta,\ x\sin\theta + y\cos\theta,\ z)$$

Observe that lengths (distances from the origin) are preserved under T. Thus T is an orthogonal operator.



Example 13.15: Let V be the l_2-space of Example 13.5. Let $T: V \to V$ be the linear operator defined by $T(a_1, a_2, \ldots) = (0, a_1, a_2, \ldots)$. Clearly, T preserves inner products and lengths. However, T is not surjective since, for example, $(1,0,0,\ldots)$ does not belong to the image of T; hence T is not invertible. Thus we see that Theorem 13.9 is not valid for spaces of infinite dimension.

Inner Product Spaces & Operators…

CHANGE OF ORTHONORMAL BASIS

In view of the special role of orthonormal bases in the theory of inner product spaces, we are naturally interested in the properties of the transition matrix from one such basis to another. The following theorem applies.

Theorem 13.12: Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of an inner product space V. Then the transition matrix from $\{e_i\}$ into another orthonormal basis is unitary (orthogonal). Conversely, if $P = (a_{ij})$ is a unitary (orthogonal) matrix, then the following is an orthonormal basis:

$$\{e_i' = a_{1i}e_1 + a_{2i}e_2 + \cdots + a_{ni}e_n : \ i = 1, \ldots, n\}$$

Inner Product Spaces & Operators…

DIAGONALIZATION AND CANONICAL FORMS IN EUCLIDEAN SPACES

Let T be a linear operator on a finite dimensional inner product space V over K. Representing T by a diagonal matrix depends upon the eigenvectors and eigenvalues of T, and hence upon the roots of the characteristic polynomial $\Delta(t)$ of T (Theorem 9.6). Now $\Delta(t)$ always factors into linear polynomials over the complex field $\mathbf{C}$, but may not have any linear polynomials over the real field $\mathbf{R}$. Thus the situation for Euclidean spaces (where $K = \mathbf{R}$) is inherently different than that for unitary spaces (where $K = \mathbf{C}$); hence we treat them separately. We investigate Euclidean spaces below, and unitary spaces in the next section.

Theorem 13.14: Let T be a symmetric (self-adjoint) operator on a real finite dimensional inner product space V. Then there exists an orthonormal basis of V consisting of eigenvectors of T; that is, T can be represented by a diagonal matrix relative to an orthonormal basis.

We give the corresponding statement for matrices.

Inner Product Spaces & Operators…

Alternate Form of Theorem 13.14: Let A be a real symmetric matrix. Then there exists an orthogonal matrix P such that $B = P^{-1}AP = P^tAP$ is diagonal.

We can choose the columns of the above matrix P to be normalized orthogonal eigenvectors of A; then the diagonal entries of B are the corresponding eigenvalues.

Inner Product Spaces & Operators…

Example 13.18: Let $A = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}$. We find an orthogonal matrix P such that P^tAP is diagonal. The characteristic polynomial $\Delta(t)$ of A is

$$\Delta(t) = |tI - A| = \begin{vmatrix} t-2 & 2 \\ 2 & t-5 \end{vmatrix} = (t-6)(t-1)$$

The eigenvalues of A are 6 and 1. Substitute $t = 6$ into the matrix $tI - A$ to obtain the corresponding homogeneous system of linear equations

$$4x + 2y = 0, \quad 2x + y = 0$$

A nonzero solution is $v_1 = (1, -2)$. Next substitute $t = 1$ into the matrix $tI - A$ to find the corresponding homogeneous system

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

A nonzero solution is $(2, 1)$. As expected by Problem 13.31, v_1 and v_2 are orthogonal. Normalize v_1 and v_2 to obtain the orthonormal basis

$$\{u_1 = (1/\sqrt{5}, -2/\sqrt{5}), \; u_2 = (2/\sqrt{5}, 1/\sqrt{5})\}$$

Finally let P be the matrix whose columns are u_1 and u_2 respectively. Then

$$P = \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{pmatrix} \quad \text{and} \quad P^{-1}AP = P^tAP = \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}$$

As expected, the diagonal entries of P^tAP are the eigenvalues corresponding to the columns of P.

Inner Product Spaces & Operators…

Theorem 13.15: Let T be an orthogonal operator on a real inner product space V. Then there is an orthonormal basis with respect to which T has the following form:

Inner Product Spaces & Operators…

DIAGONALIZATION AND CANONICAL FORMS IN UNITARY SPACES

We now present the fundamental diagonalization theorem for complex inner product spaces, i.e. for unitary spaces. Recall that an operator T is said to be *normal* if it commutes with its adjoint, i.e. if $TT^* = T^*T$. Analogously, a complex matrix A is said to be *normal* if it commutes with its conjugate transpose, i.e. if $AA^* = A^*A$.

Example 13.20: Let $A = \begin{pmatrix} 1 & 1 \\ i & 3+2i \end{pmatrix}$. Then

$$AA^* = \begin{pmatrix} 1 & 1 \\ i & 3+2i \end{pmatrix}\begin{pmatrix} 1 & -i \\ 1 & 3-2i \end{pmatrix} = \begin{pmatrix} 2 & 3-3i \\ 3+3i & 14 \end{pmatrix}$$

$$A^*A = \begin{pmatrix} 1 & -i \\ 1 & 3-2i \end{pmatrix}\begin{pmatrix} 1 & 1 \\ i & 3+2i \end{pmatrix} = \begin{pmatrix} 2 & 3-3i \\ 3+3i & 14 \end{pmatrix}$$

Thus A is a normal matrix.

The following theorem applies.

Inner Product Spaces & Operators…

Theorem 13.16: Let T be a normal operator on a complex finite dimensional inner product space V. Then there exists an orthonormal basis of V consisting of eigenvectors of T; that is, T can be represented by a diagonal matrix relative to an orthonormal basis.

We give the corresponding statement for matrices.

Alternate Form of Theorem 13.16: Let A be a normal matrix. Then there exists a unitary matrix P such that $B = P^{-1}AP = P^*AP$ is diagonal.

The next theorem shows that even non-normal operators on unitary spaces have a relatively simple form.

Inner Product Spaces & Operators…

Theorem 13.17: Let T be an arbitrary operator on a complex finite dimensional inner product space V. Then T can be represented by a triangular matrix relative to an orthonormal basis of V.

Alternate Form of Theorem 13.17: Let A be an arbitrary complex matrix. Then there exists a unitary matrix P such that $B = P^{-1}AP = P^*AP$ is triangular.

Inner Product Spaces & Operators…

SPECTRAL THEOREM

The Spectral Theorem is a reformulation of the diagonalization Theorems 13.14 and 13.16.

Theorem 13.18 (Spectral Theorem): Let T be a normal (symmetric) operator on a complex (real) finite dimensional inner product space V. Then there exist orthogonal projections $E_1, \ldots, E_r$ on V and scalars $\lambda_1, \ldots, \lambda_r$ such that

(i) $T = \lambda_1 E_1 + \lambda_2 E_2 + \cdots + \lambda_r E_r$

(ii) $E_1 + E_2 + \cdots + E_r = I$

(iii) $E_i E_j = 0$ for $i \neq j$.

The next example shows the relationship between a diagonal matrix representation and the corresponding orthogonal projections.

Example 13.21: Consider a diagonal matrix, say $A = \begin{pmatrix} 2 & & & \\ & 3 & & \\ & & 3 & \\ & & & 5 \end{pmatrix}$. Let

$$E_1 = \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 0 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & 1 \end{pmatrix}$$

The reader can verify that the E_i are projections, i.e. $E_i^2 = E_i$, and that

(i) $A = 2E_1 + 3E_2 + 5E_3$, (ii) $E_1 + E_2 + E_3 = I$, (iii) $E_i E_j = 0$ for $i \neq j$