Linear Algebra

V.S. Sunder
Institute of Mathematical Sciences
Chennai

Preface

These notes were initially prepared for a summer school in the MTTS programme more than a decade ago. They are intended to introduce an undergraduate student to the basic notions of linear algebra, and to advocate a geometric rather than a coordinate-dependent purely algebraic approach. Thus, the goal is to make the reader see the advantage of thinking of abstract vectors rather than n-tuples of numbers, vector addition as the parallelogram law in action rather than as coordinate-wise algebraic manipulations, abstract inner products rather than the more familiar dot products and most importantly, linear ransformations rather than rectangular arrays of numbers, and compositions of linear tranformations rather than seemingly artificially defined rules of matrix multiplication. For instance, some care is taken to introduce the determinant of a transformation as the signed volume of the image of the unit cube before making contact with (and thus motivating) the usual method of expansion along rows with appropriate signs thrown in. After a fairly easy paced discussion, the notes culminate in the celebrated spectral theorem (in the real as well as complex cases). These notes may be viewed as a 'baby version' of the ultimate book in this genre, viz., *Finite-dimensional vector spaces*, by Halmos. They are quite appropriate for a one semester course to undergraduate students in their first year, which will prepare them for the method of abstraction in modern mathematics.

Contents

Chapter 1

Vector Spaces

1.1 The Euclidean space $I\!R^3$

We begin with a quick review of elementary 'co-ordinate geometry in three dimensions'. The starting point is the observation that - having fixed a system of three mutually orthogonal (oriented) co-ordinate axes - a mathematical model of the three-dimensional space that we inhabit is given by the set

$$I\!R^3 = \{\mathbf{x} = (x_1, x_2, x_3) : x_1, x_2, x_3 \in I\!R\}.$$

We refer to the element $\mathbf{x} = (x_1, x_2, x_3)$ of $I\!R^3$ as the point whose i-th co-ordinate (with respect to the initially chosen 'frame of reference') is the real number x_i; we shall also refer to elements of $I\!R^3$ as vectors (as well as refer to real numbers as 'scalars'). Clearly, the point $\mathbf{0} = (0,0,0)$ denotes the 'origin', i.e., the point of intersection of the three co-ordinate axes.

The set $I\!R^3$ comes naturally equipped with two algebraic operations, both of which have a nice geometric interpretation; they are:

(i) **Scalar multiplication**: if $\mathbf{x} = (x_1, x_2, x_3) \in I\!R^3$ and $\alpha \in I\!R$, define $\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, \alpha x_3)$; geometrically, the point $\alpha\mathbf{x}$ denotes the point which lies on the line joining $\mathbf{x}$ to $\mathbf{0}$, whose distance from $\mathbf{0}$ is $|\alpha|$ times the distance from $\mathbf{x}$ to $\mathbf{0}$, and is such that $\mathbf{x}$ and $\alpha\mathbf{x}$ lie on the same or the opposite side of $\mathbf{0}$ according as the scalar α is positive or negative.

(ii) **Vector addition**: if $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{y} = (y_1, y_2, y_3) \in I\!R^3$, define $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, x_3 + y_3)$; geometrically, the points $\mathbf{0}$, $\mathbf{x}$, $\mathbf{x} + \mathbf{y}$ and $\mathbf{y}$ are successive vertices of a parellelogram in the plane determined by the three points $\mathbf{0}$, $\mathbf{x}$ and $\mathbf{y}$. (As is well-known, this gives a way to determine the resultant of two forces.)

The above algebraic operations allow us to write equations to describe geometric objects: thus, for instance, if $\mathbf{x}$ and $\mathbf{y}$ are vectors in $I\!R^3$, then $\{(1-t)\mathbf{x} + t\mathbf{y} : 0 \leq t \leq 1\}$ is the 'parametric form' of the line segment joining $\mathbf{x}$ to $\mathbf{y}$ in the sense that as the parameter t increases from 0 to 1, the point $(1-t)\mathbf{x} + t\mathbf{y}$ moves from the point $\mathbf{x}$ to the point $\mathbf{y}$ monotonically (at a 'uniform speed') along the line segment joining the two points; in fact, as the parameter t ranges from $-\infty$ to $+\infty$, the point $(1-t)\mathbf{x} + t\mathbf{y}$ sweeps out the entire line determined by the points $\mathbf{x}$ and $\mathbf{y}$, always proceeding in the direction of the vector $\mathbf{y} - \mathbf{x}$.

In addition to the above 'linear structure', there is also a notion of orthogonality in our Euclidean space $I\!R^3$. More precisely, if $\mathbf{x}$ and $\mathbf{y}$ are as above, their **dot-product** is defined by $\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + x_3y_3$, and this 'dot-product' has the following properties :

(i) if we define $||\mathbf{x}||^2 = \mathbf{x} \cdot \mathbf{x}$, then $||\mathbf{x}||$ is just the Euclidean distance from the origin $\mathbf{0}$ to the point $\mathbf{x}$, and we refer to $||\mathbf{x}||$ as the **norm** of the vector $\mathbf{x}$;

(ii) $\mathbf{x} \cdot \mathbf{y} = ||\mathbf{x}|| \, ||\mathbf{y}|| \cos\theta$, where θ denotes the 'oriented angle' subtended at the origin by the line segment joining $\mathbf{x}$ and $\mathbf{y}$; and in particular, the vectors $\mathbf{x}$ and $\mathbf{y}$ are 'orthogonal' or perpendicular - meaning that the angle θ occurring above is an odd multiple of $\pi/2$ - precisely when $\mathbf{x} \cdot \mathbf{y} = 0$.

Clearly, we can use the notion of the dot-product to describe a much greater variety of geometric objects via equations; for instance :

(a) $\{\mathbf{x} \in I\!R^3 : ||\mathbf{x} - \mathbf{y}|| = r\}$ is the sphere with centre at $\mathbf{y}$ and radius r; and

(b) If $\mathbf{u} = (u_1, u_2, u_3) \in I\!R^3$, then $\{\mathbf{x} = (x_1, x_2, x_3) \in I\!R^3 : u_1x_1 + u_2x_2 + u_3x_3 = 0\}$ is precisely the plane through the origin consisting of all vectors which are orthogonal to the vector $\mathbf{u}$.

One often abbreviates the preceding two sentences into such statements as

(a') $||\mathbf{x} - \mathbf{y}|| = r$ is the equation of a sphere; and

(b') $ax + by + cz = 0$ is the equation of the plane perpendicular to (a, b, c).

Thus, for instance, if $a_{ij}, 1 \leq i, j \leq 3$ are known real numbers, the algebraic problem of determining whether the system of (simultaneous) linear equations

$$\begin{array}{rcrcrcl} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & = & 0 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & = & 0 \end{array} \tag{1.1.1}$$

admits a non-trivial solution x_1, x_2, x_3 - i.e., the x_j's are real numbers not all of which are equal to zero - translates into the geometric problem of determining whether three specified planes through the origin share a common point other than the origin.

EXERCISE 1.1.1 *Argue on geometric grounds that the foregoing problem admits a non-trivial solution if and only if the three points* $\mathbf{u}_i = (a_{i1}, a_{i2}, a_{i3}),\ 1 \leq i \leq 3$ *are co-planar (i.e., lie in a plane).*

One of the many successes of the subject called 'linear algebra' is the ability to tackle mathematical problems such as the foregoing one, by bringing to bear the intuition arising from geometric considerations. We shall have more to say later about solving systems of simultaneous linear equations.

Probably **the** central object in linear algebra is the notion of a **linear transformation**. In a sense which can be made precise, a linear transformation on $\mathbb{R}^3$ is essentially just a mapping $T : \mathbb{R}^3 \to \mathbb{R}^3$ which 'preserves collinearity' in the sense that whenever $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are three points in $\mathbb{R}^3$ with the property of lying on a straight line, the images $T\mathbf{x}$, $T\mathbf{y}$, $T\mathbf{z}$ of these points under the mapping T also have the same property. Examples of such mappings are given below.

(i) $T_1(x_1, x_2, x_3) = (2x_1, 2x_2, 2x_3)$, the operation of 'dilation by a factor of two';

(ii) T_2 = the operation of (performing a) rotation by 90^o in the counter-clockwise direction about the z-axis;

(iii) T_3 = the reflection in the xy-plane;

(iv) T_4 = the perpendicular projection onto the xy-plane; and

(v) T_5 = the 'antipodal map' which maps a point to its 'reflection about the origin'.

EXERCISE 1.1.2 *Write an explicit formula of the form* $T_i\mathbf{x} = \cdots$ *for* $2 \leq i \leq 4$. *(For instance,* $T_5(x_1, x_2, x_3) = (-x_1, -x_2, -x_3)$.*)*

More often than not, one is interested in the converse problem - i.e., when a linear transformation is given by an explicit formula, one would like to have a geometric interpretation of the mapping.

Exercise 1.1.3 *Give a geometric interpretation of the linear transformation on $I\!R^3$ defined by the equation*

$$T(x_1, x_2, x_3) \;=\; (\frac{2x_1 - x_2 - x_3}{3}, \frac{-x_1 + 2x_2 - x_3}{3}, \frac{-x_1 - x_2 + 2x_3}{3}).$$

The subject of linear algebra enables one to deal with such problems and, most importantly, equips one with a geometric intuition that is invaluable in tackling any 'linear' problem. In fact, even for a 'non-linear' problem, one usually begins by using linear algebraic methods to attack the best 'linear approximation' to the problem - which is what the calculus makes possible.

Rather than restricting oneself to $I\!R^3$, it will be profitable to consider a more general picture. An obvious first level of generalisation - which is a most important one - is to replace 3 by a more general positive integer and to study $I\!R^n$. The even more useful generalisation - or abstraction - is the one that rids us of having to make an artificial initial choice of co-ordinate axes and to adopt a 'co-ordinate free' perspective towards n-dimensional space, which is what the notion of a vector space is all about.

1.2 Finite-dimensional (real) vector spaces

We begin directly with the definition of an abstract (real) vector space; in order to understand the definition better, the reader will do well to keep the example of $I\!R^3$ in mind. (The reason for our calling these objects 'real' vector spaces will become clear later - in §4.2.)

Definition 1.2.1 *A (real) vector space is a set V equipped with two algebraic operations called scalar multiplication and vector addition which satisfy the following conditions:*

Vector addition : *To every pair of elements u, v in V, there is uniquely associated a third element of V denoted by $u + v$ such that the following conditions are satisfied by all u, v, w in V:*
(A1) (commutativity) $u+v = v+u$;

(A2) (associativity) $(u+v)+w = u+(v+w)$*;*
(A3) (zero) there exists an element in V*, denoted simply by* 0 *, with the property that* $u+0 = u$*;*
(A4) (negative) to each u *in* V *, there is associated another element in* V *denoted by* $-u$ *with the property that* $u + (-u) = 0$*.*

Scalar multiplication : *There exists a mapping from* $\mathbb{R} \times V$ *to* V*, the image under this mapping of the pair* (α, v) *being denoted by* αv *, such that the following conditions are satisfied for all choices of 'vectors'* u, v, w *and all scalars* α, β, γ *- where we refer to elements of* V *as vectors and elements of* $\mathbb{R}$ *as scalars :*
(S1) $\alpha(u + v) = \alpha u + \alpha v$*;*
(S2) $(\alpha + \beta)u = \alpha u + \beta u$*;*
(S3) $\alpha(\beta u) = (\alpha\beta)u$*;*
(S4) $1u = u$*.*

The purpose of the following exercise is to show, through a sample set of illustrations, that the above axioms guarantee that everything that should happen, does.

EXERCISE 1.2.1 *If* V *is a real vector space as above, then prove the validity of the following statements :*
(i) if $u, v, w \in V$ *and if* $u + v = w$*, then* $u = w + (-v)$*;*
(ii) deduce from (i) above that the zero vector and the negative of a vector, which are guaranteed to exist by the axioms of a vector space, are uniquely determined by their defining properties; hence deduce that $-(-u) = u$*;*
(iii) if $u, v \in V$*, then we shall henceforth write* $u - v$ *instead of the more elaborate (but correct) expression* $u + (-v)$*; show that, for any* u, v *in* V *, we have :* $-(u + v) = -u - v = -v - u, -(u - v) = v - u$*.*
(iv) extend the associative and commutative laws to show that if $v_1, \cdots, v_n$ *are any finite number of vectors, then there is only one meaningful way of making sense of the expression* $v_1 + \cdots + v_n$*; we shall also denote this 'sum' by the usual notation* $\sum_{i=1}^n v_i$*;*
(v) if $u \in V$ *and* n *is any positive integer, and if we set* $v_1 = \cdots = v_n = u$ *, then* $\sum_{i=1}^n v_i = nu$*, where the term on the left is defined by vector addition and the term on the right is defined by scalar multiplication; state and prove a similar assertion concerning* nu *when* n *is a negative integer. What if* $n = 0$*?*

Before proceeding any further, we pause to list a number of examples of real vector spaces.

Example 1.2.2 *(1) Fix a positive integer* n *and define* n**-dimensional Euclidean space** $\mathbb{R}^n$ *by*

$$\mathbb{R}^n = \{x = (x_1, \ldots, x_n) : x_1, \ldots x_n \in \mathbb{R}\}$$

and define vector addition and scalar multiplication 'component-wise' - exactly as in the case $n = 3$.

(2) For fixed positive integers m, n, *recall that a real* $m \times n$ **-matrix** *is, by definition, a rectangular array of numbers of the form*

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

The set $M_{m\times n}(\mathbb{R})$ *of all real* $m \times n$*-matrices constitutes a real vector space with respect to entry-wise scalar multiplication and addition of matrices. (This vector space is clearly a (somewhat poorly disguised) version of the vector space* $\mathbb{R}^{mn}$*.) For future reference, we record here that we shall adopt the short-hand of writing* $M_n(\mathbb{R})$ *rather than* $M_{n\times n}(\mathbb{R})$*. We shall call a matrix in* $M_{m\times n}(\mathbb{R})$ *square or rectangular according as* $m = n$ *or* $m \neq n$.

(3) Define $\mathbb{R}^\infty$ *to be the set of all infinite sequences of real numbers, and define vector addition and scalar multiplication component-wise.*

(4) Consider the subset ℓ^∞ *of* $\mathbb{R}^\infty$ *consisting of those sequences which are uniformly bounded; this is also a real vector space with respect to component-wise definitions of vector addition and scalar multiplication.*

(5) If X *is any set, the set* $Fun_{\mathbb{R}}(X)$ *of real-valued functions defined on the set* X *has a natural structure of a real vector space with respect to the operations defined by* $(f + g)(x) = f(x) + g(x)$ *and* $(\alpha f)(x) = \alpha f(x)$*. (In fact, the first three of the foregoing examples can be thought of as special cases of this example, for judicious choices of the underlying set* X*.)*

(6) The set $C[0, 1]$ *, consisting of continuous functions defined on the closed unit interval [0,1], is also a vector space with respect to vector addition and scalar multiplication defined exactly as in the last example.*

(7) The set $C^\infty(0,1)$, consisting of functions defined on the open unit interval (0,1) which are 'infinitely differentiable', is a real vector space with respect to vector addition and scalar multiplication defined exactly as in the past two examples.

(8) Consider the subset D of the set $C^\infty(0,1)$ consisting of those functions f which satisfy the differential equation

$$f''(x) - 2f'(x) + f(x) = 0, \quad \forall\ x \ \in \ (0,1).$$

Then D is a vector space with respect to the same definitions of the vector operations as in the past three examples.

The preceding examples indicate one easy way to manufacture new examples of vector spaces from old, in the manner suggested by the next definition.

Definition 1.2.3 *A subset W of a vector space V is said to be a* **subspace** *of the vector space V if it contains 0 and if it is 'closed under the vector space operations of V' in the sense that whenever $u, v \in W$ and $\alpha \in \mathbb{R}$, then also $u+v \in W$ and $\alpha u \in W$. (The reason for the requirement that W contains 0 is only to rule out the vacuous possibility W is the empty set.)*

It must be clear that a subspace of a vector space is a vector space in its own right with respect to the definitions of scalar multiplication and vector addition in the ambient space restricted to the vectors from the subspace. In the foregoing examples, for instance, the vector spaces ℓ^∞, $C[0,1]$ and D are, respectively, subspaces of the vector spaces $\mathbb{R}^\infty$, $Fun_{\mathbb{R}}([0,1])$ and $C^\infty(0,1)$.

We pause for a couple of exercises whose solution will help clarify this notion.

EXERCISE 1.2.2 *(i) Show that the set of points in $\mathbb{R}^3$ whose co-ordinates add up to 0 is a subspace.*
(ii) Can you describe all the subspaces of $\mathbb{R}^3$?

EXERCISE 1.2.3 *Show that a subset W of a vector space V is a subspace if and only if it satisfies the following condition: whenever $u, v \in W$ and $\alpha, \beta \in \mathbb{R}$, it is also the case that $(\alpha u + \beta v) \in W$.*

The preceding exercise used a notion that we shall see sufficiently often in the sequel to justify introducing a bit of terminology : if $v, v_1, \ldots, v_n$ are vectors in a vector space, we shall say that the vector v is a **linear combination** of the vectors $v_1, \ldots, v_n$ if there exists scalars (i.e., real numbers) $\alpha_1, \ldots, \alpha_n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. In this terminology, we can say that a subset of a vector space is a subspace precisely when it is closed under the formation of linear combinations. (Note: This last statement says something slightly more than the last exercise; convince yourself that it does and that the 'slightly more' that it does say is true.)

In the 'opposite direction', we have the following elementary fact which we single out as a proposition, since it will facilitate easy reference to this fact and since it introduces a bit of very convenient notation and terminology:

Proposition 1.2.4 *If* $S = \{v_1, \ldots\ v_n\ \}$ *is a finite set of vectors in a vector space* V*, define*

$$\bigvee S = \{ \sum_{i=1}^{n} \alpha_i v_i : \alpha_1, \ \ldots, \ \alpha_n \in I\!R \}.$$

Then, $\bigvee S$ *is the smallest subspace containing the set* S *in the sense that:*
(i) it is a vector subspace of V *which contains* S *, and*
(ii) if W *is any subspace of* V *which contains* S *, then* W *necessarily contains* $\bigvee S$*.*

The proof of this proposition and the following generalisation of it are easy and left as exercises to the reader.

EXERCISE 1.2.4 *If* S *is an arbitrary (not necessarily finite) subset of a vector space* V *, and if* $\bigvee S$ *is defined to be the set of vectors expressible as linear combinations of vectors from* S *, then show that* $\bigvee S$ *is the smallest subspace of* V *which contains the set* S *.*

If S and W are related as in the above exercise, we say that W is the subspace **spanned** (or generated) by the set S.

In view of Exercise 1.2.4, we will have an explicit understanding of a subspace if we can find an explicit set of vectors which spans it. For instance, while the vector space D of Example 1.2.2 (8) might look esoteric at first glance, it looks completely harmless after one has succeeded in establishing that $D = \bigvee S$ where S is the set consisting of the two functions $f(x) = e^x$ and $g(x) = xe^x$. (This is something you might have seen in the study of what are called 'linear differential equations with constant coefficients'.)

Obviously the first sentence of the last paragraph is not unconditionally true since one always has the trivial identity $W = \bigvee W$ for any subspace W. Hence, in order for the above point of view to be most effectively employed, we should start with a spanning set for a subspace which is 'as small as possible'.

The next proposition is fundamental in our quest for efficient spanning sets.

Proposition 1.2.5 *Let W be a subspace of a vector space V. Suppose $W = \bigvee S$. Then the following conditions on the spanning set S are equivalent :*
(i) The set S is a minimal spanning set in the sense that if S_0 is any proper subset of S, then S_0 does not span W;
(ii) The set S is **linearly independent***, meaning that there is no non-trivial linear relation among the elements of S; more precisely, if $v_1, \ldots, v_n$ are distinct elements in S, if $\alpha_1, \ldots, \alpha_n \in$ I and if $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$, then necessarily $\alpha_1 = \cdots = \alpha_n = 0$.*

Proof : $(i) \Rightarrow (ii)$: Suppose $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$, where $v_1, \ldots, v_n$ are distinct elements in S and $\alpha_1, \ldots, \alpha_n$ are scalars, not all of which zero; by re-labelling, if necessary, we may assume without loss of generality that $\alpha_n \neq 0$. Then $v_n = \beta_1 v_1 + \cdots + \beta_{n-1} v_{n-1}$, where $\beta_j = -\frac{\alpha_j}{\alpha_n}$ for $1 \leq j < n$, whence we see that v_n is a linear combination of elements of $S \setminus \{v_n\} = S_0$ (say); it follows easily that $\bigvee S_0 = \bigvee S$, which contradicts the assumed minimality property of the set S.

$(ii) \Rightarrow (i)$: Suppose S does not have the asserted minimality property, and that $W = \bigvee S_0$ for some proper subset S_0 of S. Pick v from $S \setminus S_0$; since $v \in W$, this implies the existence of elements $v_1, \ldots v_{n-1}$ in S_0 with the property that $v = \alpha_1 v_1 + \cdots \alpha_{n-1} v_{n-1}$ for some scalars $\alpha_i, 1 \leq i < n$. Writing $\alpha_n = -1 \neq 0, v_n = v$, we find that we have arrived at a contradiction to the assumed linear independence of the set S. $\square$

We pause to remark that a set of vectors is said to be linearly dependent if it is not linearly independent. The following exercises should give the reader some feeling for the notion of linear independence.

EXERCISE 1.2.5 *Determine which of the following sets S of vectors in the vector space V constitute linearly independent sets :*

(i) $S = \{(1,0,0),(1,1,0),(0,0,0)\}, \; V = \mathbb{R}^3$*;*
(ii) $S = \{0\}, \; V$ *arbitrary;*
(iii) $S = \{(1,0,0,0),(1,1,0,0),(1,1,1,0),(0,0,0,1)\}, V = \mathbb{R}^4$*;*
(iv) $S = \{(1,0,1),(1,1,0),(0,1,1)\}, V = \mathbb{R}^3$*;*
(v) $S = \{ x^n : n = 0,1,2,\ldots\}, \; V = C[0,1]$.

The following exercise contains an assertion which is 'dual' to the above proposition.

Exercise 1.2.6 *Let S be a linearly independent subset of a subspace W. The following conditions on the set S are equivalent :*
(i) $\bigvee S = W$*;*
(ii) the set S is a maximal linearly independent set in W - meaning that there exists no linearly independent subset of W which contains S as a proper subset.

The contents of Proposition 1.2.5 and Exercise 1.2.6 may be paraphrased as follows.

Corollary 1.2.6 *The following conditions on a subset B of a subspace W of a vector space are equivalent :*
(i) B is a maximal linearly independent subset of W;
(ii) B is a minimal spanning subset of W;
(iii) B is linearly independent and spans W.

If a set B and a subspace W are related as above, then B is said to be a **basis** *for the subspace W.*

Before proceeding further, we make a definition which imposes a simplifying (and quite severely constrictive) condition on a vector space; we shall be concerned only with such vector spaces for the rest of these notes.

Definition 1.2.7 *A vector space is said to be* **finite-dimensional** *if it admits a finite spanning set.*

The reader should go through the list of examples furnished by Example 1.2.2 and decide precisely which of those examples is finite- dimensional.

Before proceeding any further, we iterate that henceforth **all vector spaces considered in these notes will be finite-dimensional** and unless we explicitly say that a vector-space is not finite-dimensional, we shall always tacitly assume that when we say 'vector space' we shall mean a finite- dimensional one.

Proposition 1.2.8 *Every vector space has a basis.*

Proof : If V is a vector space, then - by our standing convention - there exists a finite set S such that $V = \bigvee S$. If S is a minimal spanning set for V, then it is a basis and we are done. Otherwise, there exists a proper subset S_1 which also spans V. By repeating this process as many times as is necessary, we see - in view of the assumed finiteness of S - that we will eventually arrive at a subset of S which is a minimal spanning set for V, i.e., a basis for V. $\Box$

(Perhaps it should be mentioned here that the theorem is valid even for vector spaces which are not finite-dimensional. The proof of that fact relies upon the axiom of choice (or Zorn's lemma or one of its equivalent forms). The reader to whom the previous sentence sentence did not make any sense need not worry as we shall never need that fact here, and also because the more general fact is never used as effectively as its finite-dimensional version.)

EXERCISE 1.2.7 *(i) If B is a basis for a vector space V, show that every element of V can be expressed in a unique way as a linear combination of elements of B.*
(ii) Exhibit at lease one basis for (a) each of the finite-dimensional vector spaces in the list of examples given in Example 1.2.2, as well as for (b) the subspace of $\mathbb{R}^3$ described in Exercise 1.2.2(i).

1.3 Dimension

The astute reader might have noticed that while we have talked of a vector space being finite-dimensional, we have not yet said anything about what we mean by the dimension of such a space. We now proceed towards correcting that lacuna.

Lemma 1.3.1 *If* $L = \{u_1, \ldots, u_m\}$ *is a linearly independent subset of a vector space* V *and if* S *is a subset of* V *such that* $\bigvee S = V$ *, then there exists a subset* S_0 *of* S *such that* $S \setminus S_0$ *has exactly* m *elements and* $\bigvee(L \bigcup S_0) = V$*. In particular, any spanning set of* V *has at least as many elements as any linearly independent set.*

Proof : The proof is by induction on m. If $m = 1$, then since $u_1 \in V$, it is possible to express u_1 as a linear combination of elements of S - say, $u_1 = \sum_{i=1}^{n} \alpha_i v_i$, for some elements $v_1, \ldots, v_n$ of S. Since $u_1 \neq 0$ (why?), clearly some α_i must be non-zero ; suppose the numbering is so that $\alpha_1 \neq 0$. It is then easy to see that v_1 is expressible as a linear combination of $u_1, v_2, \ldots, v_n$, and to consequently deduce that $S_0 = S \setminus \{v_1\}$ does the job.

Suppose the lemma is true when L has $(m-1)$ elements. Then we can, by induction hypothesis applied to $L \setminus \{u_m\}$ and S, find a subset S_1 of S such that $S \setminus S_1$ has $(m-1)$ elements and $\bigvee (S_1 \bigcup \{u_1, \ldots, u_{m-1}\}) = V$. In particular, since $u_m \in V$, we can find scalars $\alpha_1, \ldots, \alpha_{m-1}$, $\beta_1, \ldots, \beta_n$ and vectors $v_1, \ldots, v_n$ in S_1 such that $u_m = \sum_{i=1}^{m-1} \alpha_i u_i + \sum_{j=1}^{n} \beta_j v_j$. The linear independence of the u_i's, and the fact that $u_m \neq 0$ implies that $\beta_{j_0} \neq 0$ for some j_0. We may deduce from this that $v_{j_0} \in \bigvee(L \bigcup S_0)$ where $S_0 = S_1 \setminus \{v_{j_0}\}$ and that consequently this S_0 does the job. □

Theorem 1.3.2 *(i) Any two bases of a vector space have the same number of elements. This common cardinality of all bases of a vector space* V *is called its* **dimension** *and is denoted by* **dim V**.
(ii) If V *is a vector space with dim* $V = n$*, then any linearly independent set in* V *contains at most* n *elements.*
(iii) Any linearly independent set can be 'extended' to a basis - meaning that if it is not already a basis, it is possible to add some more elements to the set so that the extended set is a basis.

Proof : (i) If B_1 and B_2 are two bases for V, then apply the foregoing lemma twice, first with $L = B_1$, $S = B_2$, and then with $L = B_2$, $S = B_1$.

(ii) If L is a linearly dependent set in V and if S is a basis for V, this assertion follows at once from the lemma.

(iii) This also follows from the lemma, for if L, S are as in the proof of (ii) above, then, in the notation of the lemma, the set obtained by adding the set S_0 to L is indeed a basis for V. (Why ?) $\square$

EXERCISE 1.3.1 *(i) Can you find subspaces* $W_0 \subset W_1 \subset W_2 \subset W_3 \subset I\!R^2$ *such that no two of the subspaces* W_i *are equal ?*
(ii) Determine the dimension of each of the finite-dimensional vector spaces in the list given in Example 1.2.2.
(iii) Define $W = \{A = ((a_{ij})) \in M_n(I\!R) : a_{ij} = 0 \text{ if } j < i\}$. *(Elements of* W *are called 'upper-triangular' matrices.) Verify that* W *is a subspace of* $M_n(I\!R)$ *and compute its dimension.*

We conclude this section with some elementary facts concerning subspaces and their dimensions.

Proposition 1.3.3 *Let* W_1 *and* W_2 *be subspaces of a vector space* V. *Then,*
(i) $W_1 \bigcap W_2$ *is also a subspace of* V*;*
(ii) if we define $W_1 + W_2 = \{v_1 + v_2 : v_1 \in W_1, v_2 \in W_2\}$, *then* $W_1 + W_2$ *is a subspace of* V*;*
(iii) $W_1 + W_2 = \bigvee(W_1 \bigcup W_2)$ *and we have :*

$$dim\ W_1 + dim\ W_2 = dim\ (W_1 + W_2) + dim\ (W_1 \bigcap W_2).$$

Proof : The assertions (i) and (ii), as well as the first assertion in (iii) are easily seen to be true. As for the final assertion in (iii), begin by choosing a basis $\{w_1, \ldots, w_n\}$ for the subspace $W_1 \bigcap W_2$. Apply Theorem 1.3.2 (iii) twice to find vectors $u_1, \ldots, u_l$, (resp., $v_1, \ldots, v_m$) such that $\{w_1, \ldots, w_n, u_1, \ldots, u_l\}$ (resp., $\{w_1, \ldots, w_n, v_1, \ldots, v_m\}$) is a basis for W_1 (resp., W_2). It clearly suffices to prove that the set $\{w_1, \ldots, w_n, u_1, \ldots, u_l, v_1, \ldots, v_m\}$ is a basis for $W_1 + W_2$. It follows easily from the description of $W_1 + W_2$ that the above set does indeed span $W_1 + W_2$. We now verify that the set is indeed linearly independent.

Suppose $\sum_{i=1}^{l} \alpha_i u_i + \sum_{j=1}^{m} \beta_j v_j + \sum_{k=1}^{n} \gamma_k w_k = 0$ for some scalars $\alpha_i, \beta_j, \gamma_k$. Set $x = \sum_{i=1}^{l} \alpha_i u_i$, $y = \sum_{j=1}^{m} \beta_j v_j$ and $z = \sum_{k=1}^{n} \gamma_k w_k$, so that our assumption is that $x + y + z = 0$. Then, by construction, we see that $x \in W_1$, $y \in W_2$ and $z \in (W_1 \bigcap W_2)$. Our assumption that $x + y + z = 0$ is seen to now imply that also $x = -(y+z) \in W_2$, whence $x \in (W_1 \bigcap W_2)$. By the definition of the w_k's, this means that x is expressible as a linear combination of the w_k's. Since the set $\{u_1, \ldots, u_l,\ w_1, \ldots, w_n\}$ is a basis for W_1, it follows from Exercise 1.2.7 (i) that we must have $x = 0$ and $\alpha_i = 0$ for $1 \leq i \leq l$. Hence also $y + z = 0$ and the assumed linear independence of the set $\{w_1, \ldots, w_n,\ v_1, \ldots, v_m\}$ now completes the proof. $\square$

EXERCISE 1.3.2 *(i) Show that any two subspaces of a finite-dimensional vector space have non-zero intersection provided their dimensions add up to more than the dimension of the ambient vector space.*
(ii) What can you say if the sum of the dimensions is equal to the dimension of the ambient space ?

Chapter 2

Linear Transformations and Matrices

2.1 Linear transformations

If a vector space V has dimension n, and if $\{v_1, \ldots, v_n\}$ is a basis for V, then it follows from Exercise 1.2.7 that every vector in V is uniquely expressible in the form $v = \sum_{i=1}^{n} \alpha_i v_i$; in other words, the mapping $T : V \to I\!R^n$ defined by $Tv = (\alpha_1, \cdots, \alpha_n)$ is a bijective correspondence which is seen to 'respect the linear operations' on the two spaces. This mapping T deserves to be called an isomorphism, and we proceed to formalise this nomenclature.

Definition 2.1.1 *(i) If V and W are vector spaces, a mapping $T : V \to W$ is called a* **linear transformation** *if it 'preserves linear combinations' in the sense that $T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 T v_1 + \alpha_2 T v_2$ for all v_1, v_2 in V and all scalars α_1, α_2.*
(ii) The set of all linear transformations from V to W will be denoted by $\mathcal{L}(V, W)$. When $V = W$, we abbreviate $\mathcal{L}(V, V)$ to $\mathcal{L}(V)$.
(iii) A linear transformation $T : V \to W$ is said to be an **isomorphism** *if there exists a linear transformation T^{-1} in $\mathcal{L}(W, V)$ such that $T \circ T^{-1} = id_W$ and $T^{-1} \circ T = id_V$. If there exists such an isomorphism, the vector spaces V and W are said to be* **isomorphic**.

EXERCISE 2.1.1 *Show that two vector spaces are isomorphic if and only if they have the same dimension, and, in particular, that* $\mathbb{R}^n$ *and* $\mathbb{R}^m$ *are isomorphic if and only if* $m = n$.

We list some elementary properties of linear transformations in the following proposition, whose elementary verification is left as an exercise to the reader.

Proposition 2.1.2 *Let* V, W, V_1, V_2, V_3 *denote vector spaces.*
(i) $\mathcal{L}(V, W)$ *is itself a vector space if the vector operations are defined pointwise in the following sense :* $(S + T)(v) = Sv + Tv, (\alpha T)(v) = \alpha(Tv)$*;*
(ii) If $T \in \mathcal{L}(V_1, V_2)$, $S \in \mathcal{L}(V_2, V_3)$ *and if we define* ST *to be the composite map* $S \circ T$, *then* $ST \in \mathcal{L}(V_1, V_3)$.

We conclude this section by introducing a very important pair of subspaces associated to a linear transformation and establishing a fundamental relation between them.

Definition 2.1.3 *Suppose* $T \in \mathcal{L}(V, W)$.
(i) The set **ker T** $= \{v \in V : Tv = 0\}$ *is called the* **kernel** *or the* **null space** *of the transformation* T, *and the dimension of this subspace of* V *is defined to be the* **nullity** *of* T, *and will henceforth be denoted by* $\nu(T)$.
(ii) The set **ran T** $= T(V) = \{Tv : v \in V\}$ *is called the* **range** *of the transformation* T, *and the dimension of this subspace of* W *is defined to be the* **rank** *of* T, *and will henceforth be denoted by* $\rho(T)$.

It should be clear to the reader that the range and kernel of a linear transformation $T : V \to W$ are indeed subspaces of W and V respectively. What may may not be so immediately clear is the following fundamental relation between the rank and nullity of T.

Theorem 2.1.4 *(Rank-Nullity Theorem) Let* $T \in \mathcal{L}(V, W)$. *Then*

$$\nu(T) + \rho(T) = dim\ V.$$

Proof : Suppose $\{w_1, \ldots, w_\rho\}$ is a basis for *ran* T. Then, by definition of the range of T, there exists v_i, $1 \leq i \leq \rho$ such that $Tv_i = w_i$ for $1 \leq i \leq \rho$. Also, let $\{v_j : \rho < j \leq \rho + \nu\}$ denote a basis for *ker* T. (Note that our notation is such that $\rho = \rho(T), \nu = \nu(T)$.) The proof will be complete once we show that $\{v_j : 1 \leq j \leq \rho + \nu\}$ is a basis for V.

We first prove the linear independence of $\{v_i : 1 \leq i \leq \rho + \nu\}$. If $\sum_{i=1}^{\rho+\nu} \alpha_i v_i = 0$, apply T to this equation, note that $Tv_i = 0$ for $\rho < i \leq \rho + \nu$, and find that $0 = \sum_{i=1}^{\rho} \alpha_i w_i$; the assumed linear independence of the w_i's now ensures that each $\alpha_i = 0$.) Hence we have $\sum_{i=\rho+1}^{\rho+\nu} \alpha_i v_i = 0$ and the assumed linear independence of the set $\{v_i : \rho < i \leq \rho + \nu\}$ implies that also $\alpha_i = 0$ for $\rho < i \leq \rho + \nu$.

We finally show that $\{v_i : 1 \leq i \leq \rho + \nu\}$ spans V. To see this, suppose $v \in V$. Then since $Tv \in$ *ran* T, the definition of the w_i's ensures the existence of scalars α_i, $1 \leq i \leq \rho$ such that $Tv = \sum_{i=1}^{\rho} \alpha_i w_i$. Put $u = \sum_{i=1}^{\rho} \alpha_i v_i$ and note that $T(v-u) = Tv - Tu = 0$ whence $v - u \in$ *ker* T; hence there exists suitable scalars α_i, $\rho < i \leq \rho + \nu$ such that $v - u = \sum_{i=\rho+1}^{\rho+\nu} \alpha_i v_i$ and in particular, we find that $v \in \bigvee\{v_i : 1 \leq i \leq \rho + \nu\}$, and the proof of the theorem is complete. □

Corollary 2.1.5 *Let $T \in \mathcal{L}(V)$. Then the following conditions on T are equivalent:*
(i) T is one-to-one - i.e., $Tx = Ty \Rightarrow x = y$.
(ii) ker $T = \{0\}$.
(iii) T maps V onto V.
(iv) ran $T = V$.
(v) T is an isomorphism. (We shall also call such an operator **invertible**.*)*

Proof : Exercise.

EXERCISE 2.1.2 *Compute the rank and nullity of each of the transformations labelled $T_i, 1 \leq i \leq 5$ and considered in Exercise 1.1.2. Also compute the rank and nullity of the transformation considered in Exercise 1.1.3.*

2.2 Matrices

Suppose V and W are finite-dimensional vector spaces of dimension n and m respectively. To be specific, let us fix a basis $B_V = \{v_1, \ldots, v_n\}$ for V and a basis $B_W = \{w_1, \ldots, w_m\}$ for W.

Suppose $T : V \rightarrow W$ is a linear transformation. To start with, the linearity of T shows that the mapping T is completely determined by its restriction to B_V. (Reason : if $v \in V$, then we can write $v = \sum_{j=1}^{n} \alpha_j v_j$ for a uniquely determined set of scalars $\alpha_j, 1 \leq j \leq n$, and it follows that $Tv = \sum_{j=1}^{n} \alpha_j T v_j$.)

Next, any vector u in W is uniquely expressible as a linear combination of the w_i's. Thus, with the notation of the last paragraph, there exist uniquely determined scalars $t_{ij}, 1 \leq j \leq n, 1 \leq i \leq m$ such that

$$Tv_j = \sum_{i=1}^{m} t_{ij} w_i \text{ for } 1 \leq j \leq n. \tag{2.2.1}$$

Thus when one has fixed 'ordered' bases B_V, B_W in the two vector spaces (meaning bases whose elements are written in some fixed order), the linear transformation T gives rise to an $m \times n$ matrix $((t_{ij}))$, which we shall denote by $[T]_{B_V}^{B_W}$ (when we want to explicitly mention the bases under consideration), according to the prescription given by equation 2.2.1.

Explicitly, if the linear transformation is given, then the matrix is defined by the above equation; note that the conclusion of the earlier paragraphs is that the mapping T is completely determined by the matrix $((t_{ij}))$ thus :

$$T(\sum_{j=1}^{n} \alpha_j v_j) = \sum_{i=1}^{m} (\sum_{j=1}^{n} t_{ij} \alpha_j) \; w_i. \tag{2.2.2}$$

Conversely, it is easy to verify that if $((t_{ij}))$ is an arbitrary matrix of scalars, then equation 2.2.2 uniquely defines a linear transformation $T \in \mathcal{L}(V, W)$ such that $[T]_{B_V}^{B_W} = ((t_{ij}))$.

The reader should ensure that (s)he understood the foregoing by going through the details of the solution of the following exercise (which summarises the conclusions of the foregoing considerations).

EXERCISE 2.2.1 *(i) With the preceding notation, verify that the passage $T \mapsto [T]_{B_V}^{B_W}$ defines an isomorphism between the vector spaces $\mathcal{L}(V, W)$ and $M_{m \times n}(\mathbb{R})$.*
(ii) Define the 'standard basis for $M_{m \times n}(\mathbb{R})$' to be the set $\{E_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ where E_{ij} is the matrix whose (i, j)-th entry is 1 and all other entries are 0. Show that this is indeed a basis for $M_{m \times n}(\mathbb{R})$.
(iii) Describe the linear transformation $T_{ij} \in \mathcal{L}(V, W)$ for which $[T]_{B_V}^{B_W} = E_{ij}$ and determine the rank and nullity of T_{ij}.

In the rest of this section, we shall be concerned with $\mathcal{L}(V)$, the main reason for this being that - in view of Proposition 2.1.2 (ii) - the set $\mathcal{L}(V)$ is more than just a vector space: it also has a product (given by composition of mappings) defined on it; thus the set $\mathcal{L}(V)$ is an example of what is referred to as a *unital algebra* in the sense that the vector space operations and the product $\mathcal{L}(V)$ are related as in the following Exercise.

EXERCISE 2.2.2 *Let V be a vector space. If $S, T, T_1, T_2 \in \mathcal{L}(V)$ and $\alpha \in \mathbb{R}$, then the following relations hold:*
(i) $(TT_1)T_2 = T(T_1T_2)$;
(ii) $T(T_1 + T_2) = TT_1 + TT_2$;
(iii) $(T_1 + T_2)T = T_1T + T_2T$;
(iv) $S(\alpha T) = (\alpha S)T = \alpha(ST)$;
(v) there exists a unique linear transformation $I_V \in \mathcal{L}(V)$ such that $I_V T = TI_V = T$ for all $T \in \mathcal{L}(V)$;
(vi) show that multiplication is not necessarily commutative, by considering the pair of linear transformations denoted by T_2 and T_3 in Exercise 1.1.2.

Rather than using the long expression 'T is a linear transformation from V to V', we shall, in the sequel, use the following shorter expression 'T is an operator on V' (to mean exactly the same thing).

When we deal with $\mathcal{L}(V)$, especially because we want to keep track of multiplication of linear transformations, we shall, when representing an operator on V by a matrix as in Exercise 2.2.1, we shall fix only one basis B for V and consider the matrix $[T]_B$ which we denoted by $[T]_B^B$ in the said exercise. To be explicit, if $B = \{v_1, \ldots, v_n\}$ is a basis for V, we write

$$[T]_B = ((t_{ij})) \Leftrightarrow Tv_j = \sum_{i=1}^{n} t_{ij} v_i \ \forall j. \tag{2.2.3}$$

Thus, **the j-th column of the matrix $[T]_B$ is just the column of coefficients needed to express Tv_j as a linear combination of the v_i's.**

Now if S, T are two operators on V, we see that

$$(ST)(v_j) = S(\sum_{k=1}^{n} t_{kj} v_k) = \sum_{i,k=1}^{n} s_{ik} t_{kj} v_i$$

where $[S]_B = ((s_{ij})), [T]_B = ((t_{ij}))$. Thus, if we write $U = ST$ and $[U]_B = ((u_{ij}))$, we find that

$$u_{ij} = \sum_{k=1}^{n} s_{ik} \, t_{kj} \text{ for } 1 \leq i, j \leq n. \tag{2.2.4}$$

We combine Exercise 2.2.1 and Exercise 2.2.2 and state the resulting conclusion in the next exercise.

EXERCISE 2.2.3 *Fix a positive integer n. Define a multiplication in $M_n(\mathbb{R})$ by requiring that if $S = ((s_{ij})), T = ((t_{ij}))$ are arbitrary $n \times n$ matrices, then $ST = ((u_{ij}))$ where the u_{ij}'s are defined by equation 2.2.4. With respect to this multiplication, the vector space $M_n(\mathbb{R})$ acquires the structure of a unital algebra, meaning that if $S, T, T_1, T_2 \in M_n(\mathbb{R})$ and $\alpha \in \mathbb{R}$ are arbitrary, then:*
(i) $(TT_1)T_2 = T(T_1T_2)$;
(ii) $T(T_1 + T_2) = TT_1 + TT_2$;
(iii) $(T_1 + T_2)T = T_1T + T_2T$;
(iv) $S(\alpha T) = (\alpha S)T = \alpha(ST)$;
(v) there exists a unique matrix $I_n \in M_n(\mathbb{R})$ such that $I_nT = TI_n = T$ for all $T \in M_n(\mathbb{R})$; explicitly describe this so-called 'identity' matrix;
(vi) if $\{E_{ij} : 1 \leq i, j \leq n\}$ is the standard basis for $M_n(\mathbb{R})$ -see Exercise 2.2.1- then show that $E_{ij}E_{kl} = \delta_{jk} \, E_{il}$, for $1 \leq i, j, k, l \leq n$, where the symbol δ_{ij} denotes -here and elsewhere in the sequel- the so-called Kronecker delta defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.5}$$

Deduce from the above that $M_n(\mathbb{R})$ (resp., $\mathcal{L}(V)$) is a non- commutative algebra unless $n = 1$ (resp., $\dim V = 1$).
(vii) Call a square matrix $T \in M_n(\mathbb{R})$ **invertible** *if there exists a matrix $T^{-1} \in M_n(\mathbb{R})$ such that $TT^{-1} = T^{-1}T = I_n$. Prove that the set $Gl_n(\mathbb{R})$ of invertible matrices of order n constitute a group with respect to multiplication, meaning :*
(a) $Gl(n, \mathbb{R})$ has a product operation defined on it, which is associative;
(b) there exists an (identity) element $I_n \in GL(n, \mathbb{R})$ such that $I_nT = TI_n = T$ for all T in $Gl(n, \mathbb{R})$;

(c) to each element T in $Gl(n, I\!R)$, there exists an element T^{-1} in $Gl(n, I\!R)$ such that $TT^{-1} = T^{-1}T = I_n$; and
(viii) prove that the inverse (of (vii)(c) above) is unique.

We continue with another set of exercises which lead to what may be called 'two-space versions' of many of the foregoing statements - in the sense that while the foregoing statements were about linear transformations from a vector space to itself, the 'two-space version' concerns linear transformations between two different spaces. The proofs of these statements should pose no problem to the reader who has understood the proofs of the corresponding one-space versions.

EXERCISE 2.2.4 *(i) Suppose V_1, V_2, V_3 are finite-dimensional vector spaces and suppose $B_i = \{v_1^{(i)}, \ldots, v_{n_i}^{(i)}\}$ is a basis for V_i, for $1 \leq i \leq 3$. Suppose $T \in \mathcal{L}(V_1, V_2)$, $S \in \mathcal{L}(V_2, V_3)$ and $U = ST$. If $[S]_{B_2}^{B_3} = ((s_{ij})), [T]_{B_1}^{B_2} = ((t_{ij}))$ and $[U]_{B_1}^{B_3} = ((u_{ij}))$, then show that*

$$u_{ij} = \sum_{k=1}^{n_2} s_{ik}\, t_{kj} \text{ for } 1 \leq i \leq n_3, 1 \leq j \leq n_1. \tag{2.2.6}$$

(ii) Taking a cue from equation 2.2.6, define the product of two rectangular matrices 'with compatible sizes' as follows : if m, n, p are positive integers, if S is an $m \times n$ matrix and if T is an $n \times p$ matrix, the product ST is the $m \times p$ matrix defined by $u_{ij} = \sum_{k=1}^{n} s_{ik}\, t_{kj}$ for $1 \leq i \leq m, 1 \leq j \leq p$. Thus, the conclusion of the first part of this exercise may be written thus:

$$[ST]_{B_1}^{B_3} = [S]_{B_2}^{B_3}[T]_{B_1}^{B_2}. \tag{2.2.7}$$

(iii) State and prove 'rectangular' versions of the statements (i)-(vi) of Exercise 2.2.3 whenever possible.

The reader who had diligently worked out part (iii) of Exercise 2.2.4 would have noticed that if V and W are vector spaces of dimensions n and m respectively, then the space $\mathcal{L}(V, W)$ has dimension mn . This is, in particular, true when $W = I\!R^1 = I\!R$ (and $m = 1$).

Definition 2.2.1 *(i) If V is a vector space, the vector space $\mathcal{L}(V, I\!R)$ is called the* **dual space** *of V and is denoted by V^*.*
(ii) An element of the dual space V^* is called a **linear functional** *on V.*

The reader should have little trouble in settling the following exercises.

EXERCISE 2.2.5 *Let* $B = \{v_1, \ldots, v_n\}$ *be a basis for a vector space* V *.*
(i) Prove that the mapping

$$\phi \ \mapsto \ (\phi(v_1), \phi(v_2), \cdots, \phi(v_n))$$

establishes an isomorphism $T : V^* \to I\!R^n$. *(Notice that an alternate way of describing the mapping* T *is thus:* $T\phi = [\phi]_B^{B_1}$ *, where* $B_1 = \{1\}$ *is the canonical one-element-basis of* $I\!R$ *.)*
(ii) Deduce from (i) that there exists a unique basis $B' = \{\phi_1, \ldots, \phi_n\}$ *for* V^* *such that*

$$\phi_i(v_j) \ = \ \delta_{ij} \ = \ \begin{cases} 1 & if \ i = j \\ 0 & otherwise \end{cases}$$

for $1 \leq i, j \leq n$ *. This basis* B' *is called the* **dual basis** *of* B *; equivalently* B *and* B' *are referred to as a pair of dual bases.*

We end this section by addressing ourselves to the following related questions:
(i) what is the relation between two operators which are represented by the same matrix with respect to different bases; explicitly, if $S, T \in \mathcal{L}(V)$ and if B, B' are a pair of ordered bases for V such that $[T]_B = [S]_{B'}$, how are the operators S and T related ?
(ii) dually, what is the relation between different matrices which represent the same operator with respect to different bases; explicitly, if B, B' are a pair of ordered bases for V and if $T \in \mathcal{L}(V)$, how are the matrices $[T]_B$ and $[T]_{B'}$ related ?

In order to facilitate the statement of the answer to these questions, we introduce some terminology.

Definition 2.2.2 *(i) Two matrices* $T_1, T_2 \in M_n(I\!R)$ *are said to be* **similar** *if there exists an invertible matrix* $S \in GL_n(I\!R)$ *such that* $T_2 = ST_1S^{-1}$.
(ii) Two operators T_1 *and* T_2 *on a vector space* V *are said to be* **similar** *if there exists an invertible operator* $S \in \mathcal{L}(V)$ *such that* $T_2 = ST_1S^{-1}$.

Before proceeding further, note that it is an easy consequence of Exercise 2.2.3 (vii) that similarity is an equivalence relation on $\mathcal{L}(V)$ as well as $M_n(I\!R)$.

Lemma 2.2.3 *Let V be a vector space, and let $T \in \mathcal{L}(V)$. Let $B = \{v_1, \ldots, v_n\}$ and $B' = \{v'_1, \ldots, v'_n\}$ denote a pair of bases for V. Then,*
(i) there exists a unique invertible operator S on V such that $Sv_j = v'_j$ for $1 \leq j \leq n$; and
(ii) if S is as in (i) above, then

$$[T]_{B'} = [S^{-1}TS]_B.$$

Proof The first assertion is clearly true. (The invertibility of S follows from the fact that *ran* S is a subspace of V which contains the basis B', which implies that S maps V onto itself.) As for the second, suppose we write $[T]_B = ((t_{ij})), [T]_{B'} = ((t'_{ij})), [S]_B = ((s_{ij}))$. Then we have, for $1 \leq j \leq n$,

$$\begin{aligned} Tv'_j &= T(Sv_j) \\ &= T\ (\ \sum_{i=1}^{n} s_{ij}\ v_i\ ) \\ &= \sum_{i=1}^{n} s_{ij}(\sum_{k=1}^{n} t_{ki}\ v_k\ ); \end{aligned}$$

on the other hand, we also have

$$\begin{aligned} Tv'_j &= \sum_{i=1}^{n} t'_{ij}\ v'_i \\ &= \sum_{i=1}^{n} t'_{ij}\ Sv_i \\ &= \sum_{i=1}^{n} t'_{ij}(\sum_{k=1}^{n} s_{ki}\ v_k\ ). \end{aligned}$$

Deduce from the linear independence of the v_k's that

$$\sum_{i=1}^{n} t_{ki}s_{ij} = \sum_{i=1}^{n} s_{ki}t'_{ij} \ \text{ for } \ 1 \leq i, j \leq n$$

and thus

$$[T]_B\ [S]_B = [S]_B\ [T]_{B'}.$$

Since the mapping $U \mapsto [U]_B$ is a *unital algebra-isomorphism* of $\mathcal{L}(V)$ onto $M_n(\mathbb{R})$, it is clear that the invertibility of the operator S implies that the matrix $[S]_B$ is invertible and that $[S^{-1}]_B = [S]_B^{-1}$. Hence the last equation implies that

$$\begin{aligned} [T]_{B'} &= [S]_B^{-1}[T]_B[S]_B \\ &= [S^{-1}TS]_B \end{aligned}$$

where we have again used the fact that the mapping $U \mapsto [U]_B$ preserves products; the proof of the lemma is complete. □

In view of the above lemma, the reader should have no trouble in supplying the proof of the following important result.

Theorem 2.2.4 *(i) Two operators T_1, T_2 on a vector space V are similar if and only if there exist bases B_1, B_2 for V such that $[T_1]_{B_1} = [T_2]_{B_2}$.*
(ii) Two $n \times n$ matrices T_1, T_2 are similar if and only if there exists an operator T on an n-dimensional vector space V and a pair of bases B_1, B_2 of V such that $[T]_{B_i} = T_i, i = 1, 2$.

Chapter 3

(Real) Inner Product Spaces

3.1 Inner Products

Now we shall start to also abstract, to the context of a general vector space, the notion of orthogonality in R^3 and related concepts that can be read off from the dot-product. We begin with the relevant definition.

Definition 3.1.1 *(a) An* **inner product** *on a (real) vector space* V *is a mapping from* $V \times V$ *to* $\mathbb{R}$ *, denoted by* $(u, v) \mapsto \ < u, v >$*, which satisfies the following conditions, for all* $u, v, w \in V$ *and* $\alpha \in \mathbb{R}$*:*
(i) $< u + v, w > \ = \ < u, w > + < v, w >$*;*
(ii) $< \alpha u, v > \ = \alpha < u, v >$*:*
(iii) $< u, v > \ = \ < v, u >$*;*
(iv) $< u, u > \ \geq 0$ *, and* $< u, u > \ = 0 \Leftrightarrow u = 0$.
(b) We shall write $||v|| \ =< v, v >^{\frac{1}{2}}$ *for any* v *in* V *and refer to this quantity as the* **norm** *of the vector* v.
(c) A vector space equipped with an inner product is called an **inner product space***.*
(d) Two vectors $u, v \in V$ *are said to be* **orthogonal** *if* $< u, v >= 0$.

A few elementary consequences of the definition are stated in the next exercise.

Exercise 3.1.1 *Let* V *be an inner product space. Then the following conditions are satisfied by all* $u, v, w \in V$ *,* $\alpha \in \mathbb{R}$*:*

(i) $< u, \alpha v > \; = \; \alpha < u, v >$;
(ii) $< u, 0 > \; = \; 0$.

Before proceeding any further, we pause to describe a couple of examples of inner product spaces.

Example 3.1.2 *(i) The fundamental example of an inner product space is provided by* n*-dimensional Euclidean space* $\mathbb{R}^n$*. The inner product there is usually called the 'dot-product' and one customarily writes* $\mathbf{x}\cdot\mathbf{y}$ *rather than* $< \mathbf{x}, \mathbf{y} >$*; to be specific, the dot-product* $\mathbf{x}\cdot\mathbf{y}$ *of two vectors* $\mathbf{x} = (x_1, \cdots, x_n)$, $\mathbf{y} = (y_1, \cdots, y_n)$ *in* $\mathbb{R}^n$ *is defined by* $\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^n x_i y_i$*, and it is trivially verified that this does indeed satisfy the properties of an inner product.*

(ii) This is an infinite-dimensional example, but finite-dimensional subspaces of this vector space will naturally yield finite-dimensional inner product spaces.The vector space $C[0,1]$ *of continuous functions on the closed unit interval* $[0,1]$ *has a natural inner product defined as follows :*

$$< f, g > \; = \; \int_0^1 f(x)g(x) \; dx.$$

(The reader should convince him(her)self that this does indeed define an inner product on $C[0,1]$*.)*

(iii) The set $\mathcal{P}_n$ *consisting of polynomials with degree strictly less than* n *is an* n *-dimensional subspace of the inner product space* $C[0,1]$ *considered in (ii) above, and hence is an* n*-dimensional inner product space.*

As Example 3.1.2 (iii) suggests, it must be noted that if V is an inner product space and if W is a subspace of V, then W is an inner product space in its own right. As in the case of $\mathbb{R}^3$, we would like to be able to think of the inner product $< u, v >$ - even in general inner product spaces - as being given by the expression $||u||||v||cos\ \theta$; in order for this to be possible, we would certainly need the validity of the following fundamental inequality.

Theorem 3.1.3 (Cauchy-Schwarz inequality) *If* V *is an inner product space, and if* $u, v \in V$*, then*

$$| < u, v > | \; \leq ||u|| \; ||v||;$$

further, equality occurs in the above inequality if and only if the vectors u *and* v *are linearly dependent.*

Proof : If $v = 0$, the theorem is trivially valid, so assume $||v|| > 0$. Consider the real-valued function defined by $f(t) = ||u - tv||^2$, and consider the problem of minimising this function as t ranges over the real line. (This corresponds to the geometric problem of finding the 'distance' from the point u to the line spanned by v.) Notice first that

$$0 \leq f(t) = ||u||^2 + t^2||v||^2 - 2t < u, v > .$$

An easy application of the methods of the calculus shows that this function is minimised when $t = \frac{<u,v>}{||v||^2} = t_0$(say). In particular, for $t = t_0$, the above inequality reads as follows:

$$\begin{aligned} 0 &\leq ||u||^2 + \frac{< u, v >^2}{||v||^2} - \frac{2 < u, v >^2}{||v||^2} \\ &= ||u||^2 - \frac{< u, v >^2}{||v||^2}. \end{aligned} \tag{3.1.1}$$

This proves the desired inequality. The final assertion follows from the observation that equality can hold in the inequality if and only if $f(t_0) = 0$, which happens if and only if the distance from u to the line through v is zero, which happens if and only if u is on the line through v. □

Besides other things, the Cauchy-Schwarz inequality shows that the norm, as defined by an inner product, is indeed a measure of the 'size' of a vector. Specifically, we have the following:

Proposition 3.1.4 *The norm on an inner product space -defined as in Definition 3.1.1(b)- satisfies the following properties, for all* u, v *in* V *and* $\alpha \in I\!R$*:*
(i) $||v|| \geq 0$*, and* $||v|| = 0 \Leftrightarrow v = 0$;
(ii) $||\alpha v|| = |\alpha| \ ||v||$;
(iii) (triangle inequality) $||u + v|| \leq ||u|| + ||v||$.

Proof : The first two properties are trivially verified. As for the third (which is called the triangle inequality for reasons that the reader should discover for him(her)self, by drawing a few pictures if necessary),

$$
\begin{aligned}
||u+v||^{\,2} &= \;< u+v, u+v > \\
&= ||u||^{\,2} + ||v||^{\,2} + 2 < u, v > \\
&\leq ||u||^{\,2} + ||v||^{\,2} + 2||u||||v|| \\
&= (||u|| + ||v||)^{\,2}.
\end{aligned}
$$

$\square$

EXERCISE 3.1.2 *If* u, v *are any two vectors in an inner product space, show that:*
(i) $||u+v||^{\,2} \;=\; ||u||^{\,2} \;+\; ||v||^{\,2} \;\Leftrightarrow\; < u, v > \;=\; 0$; *and*
(ii) $||u+v||^{\,2} \;+\; ||u-v||^{\,2} \;=\; 2\;(\;||u||^{\,2} \;+\; ||v||^{\,2}\;)$.
What is the geometric interpretation of these identities?

Once we have a notion of orthogonality, it makes sense for us not to pick any basis to use as a reference frame, but to pick a basis which incorporates this additional structure; to be precise, we shall be interested in bases which are 'orthonormal' in the following sense.

Definition 3.1.5 *A set* $\{e_i \;:\; 1 \leq i \leq m\}$ *of vectors in an inner product space is said to be* **orthonormal** *if the following conditions are satisfied:*
(i) $< e_i, e_j > \;=\; 0$ *if* $1 \leq i < j \leq m$; *i.e., the vectors in the set are pairwise orthogonal; and*
(ii) $||e_i|| \;=\; 1$ *for* $1 \leq i \leq m$; *i.e., the vectors are 'normalised' to have unit norm.*
The two conditions above can be simplified using the Kronecker delta symbol thus:

$$< e_i, e_j > \;=\; \delta_{ij} \;\; \textit{for} \;\; 1 \leq i, j \leq m.$$

The reader will have no difficulty in verifying that the standard basis in $\mathbb{R}^n$ is an orthonormal set. In general, in any inner product space, we will like to deal with a basis which is an orthonormal set. Of course, we first need to prove that such bases exist. A first step is the following lemma.

Before stating the lemma, we pause to state that unless otherwise stated, the symbol V will always denote a finite-dimensional inner product space in the rest of these notes.

Lemma 3.1.6 *(a) Any orthonormal set is linearly independent.*
(b) If $\{e_1, \ldots, e_m\}$ *is an orthonormal set in* V *and if* $W = \bigvee\{e_i : 1 \leq i \leq m\}$*, then the following hold:*
(i) $x \in W \Rightarrow x = \sum_{i=1}^m < x, e_i > e_i$ *;*
(ii) $x, y \in W \Rightarrow < x, y > = \sum_{i=1}^m < x, e_i >< y, e_i >$ *and in particular,*

$$||x||^2 = \sum_{i=1}^m | < x, e_i > |^2.$$

Proof : (a) Suppose $\{e_1, \ldots, e_m\}$ is an orthonormal set. Suppose $\alpha_i, 1 \leq i \leq m$ are some scalars such that $\sum_{i=1}^m \alpha_i e_i = 0$. Fix an arbitrary $j, 1 \leq j \leq m$, take inner product with e_j in the above equation, and use the assumed orthonormality of the e_i's to deduce that indeed $\alpha_j = 0$ for each j as desired.

(b) (i) The assumption that $\{e_i : 1 \leq i \leq m\}$ spans W ensures that $x = \sum_{i=1}^m \alpha_i e_i$ for some scalars $\alpha_1, \ldots, \alpha_m$. Fix an arbitrary j, $1 \leq j \leq m$, take inner product with e_j in the above equation, and use the assumed orthonormality of the e_i's to deduce that indeed $< x, e_j > = \alpha_j$.

(ii) The second assertion is a trivial consequence of (i) and orthonormality. □

The above lemma shows that orthonormal sets in an n-dimensional inner product space cannot have more than n elements. Our next step is to be able to find orthonormal sets with at least n elements.

Proposition 3.1.7 *Let* $\{x_1, \ldots, x_n\}$ *be any linearly independent set in* V *. Then there exists an orthonormal set* $\{e_1, \ldots, e_n\}$ *in* V *such that*

$$\bigvee\{e_k : 1 \leq k \leq m\} = \bigvee\{x_k : 1 \leq k \leq m\}, \quad \text{for } 1 \leq m \leq n. \tag{3.1.2}$$

Proof : The hypothesis ensures that $x_1 \neq 0$; set $e_1 = \frac{x_1}{||x_1||}$.

Inductively, suppose $m < n$, and suppose we have been able to construct an orthonormal set $\{e_1, \ldots, e_m\}$ so that $\bigvee\{e_k : 1 \leq k \leq m\} = \bigvee\{x_k : 1 \leq k \leq m\}$. In particular, if we put $z = x_{m+1} - \sum_{k=1}^m < x_{m+1}, e_k > e_k$, the assumed linear independence of the x_i's and the 'induction hypothesis' of the previous sentence imply that $z \neq 0$. Define $e_{m+1} = \frac{z}{||z||}$. It is an easy matter now to prove that $\{e_k : 1 \leq k \leq (m+1)\}$ is an orthonormal set such that $\bigvee\{e_k : 1 \leq k \leq (m+1)\} = \bigvee\{x_k : 1 \leq k \leq (m+1)\}$. This constructive procedure completes the proof of the proposition. □

The above construction of an orthonormal set from a given linearly independent set is called the **Gram-Schmidt orthogonalisation process**. It is both useful and important, and the purpose of the following exercise is to familiarise the reader with this important construction.

EXERCISE 3.1.3 *(a) Show that the above construction is almost 'forced' in the sense that if $\{\tilde{e}_i : 1 \leq i \leq n\}$ is another orthonormal set in V which also satisfies the condition in equation 3.1.2, then necessarily $\tilde{e}_i = \pm e_i$ for $1 \leq i \leq n$.*

(b) Consider the inner product space denoted by $\mathcal{P}_n$ in Example 3.1.2 (iii). What is the result of applying the Gram-Schmidt process to the basis $\{x^k : 0 \leq k \leq 3\}$ of the subspace $\mathcal{P}_3$? Can you prove that, for general n, the result of applying this process to the set $\{x^k : 0 \leq k \leq n\}$ will always result in a sequence $\{p_k(x) : 0 \leq k \leq n\}$ with the property that $p_k(x)$ is a polynomial of degree exactly equal to k, for $0 \leq k \leq n$?

(c) If $\{x_1, \ldots, x_m\}$ is a linearly independent set in V such that $\{x_1, \ldots, x_k\}$ is an orthonormal set for some $k \leq m$, and if $\{e_1, \ldots, e_m\}$ is the orthonormal set obtained by applying the Gram-Schmidt orthogonalisation process to the set $\{x_1, \ldots, x_m\}$, then show that $e_i = x_i$ for $1 \leq i \leq k$.

All the pieces are now in place for the proof of the following result.

Theorem 3.1.8 *Let V be an n-dimensional inner product space.*
(a) There exists an orthonormal basis for V.
(b) The following conditions on an orthonormal set $\{e_1, \ldots, e_m\}$ in V are equivalent:
(i) $m = n$;
(ii) $\{e_1, \ldots, e_m\}$ is a basis for V;
(iii) If $x \in V$ is arbitrary, then $x = \sum_{i=1}^m < x, e_i > e_i$;
(iv) If $x, y \in V$ are arbitrary, then $< x, y > = \sum_{i=1}^m < x, e_i >< y, e_i >$;
(v) If $x \in V$ is arbitrary, then $||x||^2 = \sum_{i=1}^m | < x, e_i > |^2$.
(vi) $\{e_1, \ldots, e_m\}$ is a maximal orthonormal set in V.

Proof : (a) Let $\{e_i : 1 \leq i \leq n\}$ denote the orthonormal set obtained upon applying the Gram-Schmidt process to some basis of V. Then $\{e_i : 1 \leq i \leq n\}$ is a spanning set of V which has exactly as many elements as $dim\ V$, and which is linearly independent in view of Lemma 3.1.6, and is consequently an orthonormal basis for V.

(b) $(i) \Rightarrow (ii)$ In view of Lemma 3.1.6 (a) and Exercise 1.2.6 it suffices to show that $\{e_i : 1 \leq i \leq n\}$ is a maximal linearly independent set; if it were not, then it would possible to find a linearly independent set in V with $(n+1)$ elements, which contradicts the assumption that $dim\ V = n$.

$(ii) \Rightarrow (iii)$ This follows from Lemma 3.1.6(b)(i).

$(iii) \Rightarrow (iv)$ This is an immediate consequence of orthonormality.

$(iv) \Rightarrow (v)$ Put $y = x$.

$(v) \Rightarrow (vi)$ The condition (v) clearly implies that the only vector in V which is orthogonal to each e_i is the zero vector, which implies the validity of condition (vi).

$(vi) \Rightarrow (i)$ If $m \neq n$, then it follows from Lemma 3.1.6 (a) that we must have $m < n$. Then the linearly independent set $\{e_i : 1 \leq i \leq m\}$ can be extended to a basis, say $\{e_1, \ldots, e_m, x_{m+1}, \ldots, x_n\}$ for V . It follows from Exercise 3.1.3(c) that there exist vectors $e_{m+1}, \ldots, e_n$ such that $\{e_i : 1 \leq i \leq n\}$ is an orthonormal basis for V. Thus the assumption $m \neq n$ has led to a contradiction to the assumption (vi) , and the proof is complete. □

A fact that made its appearance in the proof of the implication $(vi) \Rightarrow (i)$ in (b) of the above theorem is worth isolating as a separate proposition.

Proposition 3.1.9 *Any orthonormal set can be extended to an orthonormal basis.* □

3.2 The Adjoint

In this section, we introduce the vital notion of the adjoint of a linear transformation between inner product spaces. (Actually, it is possible to define the adjoint of a linear transformation between any two abstract vector spaces (which do not necessarily have any inner product structure). We touch upon this general notion in the exercises, but restrict ourselves, in the text, to the context of inner product spaces. The connecting link between the two notions is the so-called Riesz representation theorem.)

We begin with this basic result which identifies the dual space of an inner product space.

Theorem 3.2.1 *(***Riesz Representation Theorem***) Let V be a finite-dimensional inner product space.*
(a) If $y \in V$, define $\phi_y : V \to I\!R$ by $\phi_y(x) = < x, y >$. Then $\phi_y \in V^*$ for all y in V.
(b) If $\phi \in V^*$ is arbitrary, then there exists a unique element y in V such that $\phi = \phi_y$.

Proof : Assertion (a) is clearly true in view of the 'bilinearity' of the inner product. (Note, incidentally, that if $\{e_1, \ldots, e_n\}$ is any orthonormal basis for V, then $y = \sum_{i=1}^n \phi_y(e_i)\ e_i$.)

As for (b), let $B = \{e_1, \ldots, e_n\}$ denote an orthonormal basis for V, define $y = \sum_{i=1}^n \phi_y(e_i)\ e_i$, note that the two linear functionals ϕ_y and ϕ agree on the basis B and must hence be identically equal. □

Exercise 3.2.1 *Show that the association $y \mapsto \phi_y$ given in our statement of the Riesz representation theorem defines an isomorphism $\phi_V : V \to V^*$.*

We next state a consequence of the Riesz representation theorem which leads to the definition of the adjoint of a linear transformation between inner product spaces.

Proposition 3.2.2 *Suppose $T \in \mathcal{L}(V, W)$, where V and W are finite-dimensional inner product spaces. Then there exists a unique linear transformation $T^* : W \to V$ with the property that*

$$< Tv, w > \quad = \quad < v, T^*w > \quad \textit{for all } v \textit{ in } V \textit{ and } w \textit{ in } W. \tag{3.2.3}$$

(The inner product on the left side of the equation refers to the inner product in W while the one on the right refers to the inner product in V.)

Proof : First fix a vector w in W and consider the map $\phi : V \to I\!R$ defined by $\phi(v) = < Tv, w >$. Since ϕ is clearly a linear functional on V, it follows from the Riesz representation theorem that there exists a unique vector - call it T^*w - such that $\phi(v) = < v, T^*w >$ for all v in V.

It is an easy consequence of the uniqueness assertion in (part (b) of the statement of) the Riesz representation theorem that the mapping T^* that we have defined above from W to V is indeed a linear map and the proof of the thoeorem is complete. □

Definition 3.2.3 *If* $T \in \mathcal{L}(V,W)$, $T^* \in \mathcal{L}(W,V)$ *are as in Proposition 3.2.2, then* T^* *is called the* **adjoint** *of* T.

We list some basic properties of the operation of adjunction in the following exercise, whose verification involves nothing more than a repeated use of the uniqueness in the Riesz representation theorem.

EXERCISE 3.2.2 *(a) Let* V, W *denote finite-dimensional inner product spaces. Then the following statements are valid for all* $S, T \in \mathcal{L}(V,W)$ *and* $\alpha \in I\!R$:
(i) $(S+T)^* = S^* + T^*$;
(ii) $(\alpha T)^* = \alpha T^*$;
(iii) $(T^*)^* = T$;
(iv) if I_V *denotes the identity operator on* V *, then* $I_V{}^* = I_V$.

(b) If $V_i, i = 1,2,3$ *are finite-dimensional inner product spaces, if* $T \in \mathcal{L}(V_1,V_2)$ *and* $S \in \mathcal{L}(V_2,V_3)$ *, then* $(ST)^* = T^*S^*$.

Inspired by the successful use, in the proof of the Riesz representation theorem, of representing linear functionals by their matrix with respect to orthonormal bases in the domain and range, and because it is such bases which are natural in the context of inner product spaces, we shall henceforth always choose only orthonormal bases in the various inner product spaces when we wish to represent a linear transformation in terms of a matrix. Our first task is to identify what the operation of adjunction corresponds to at the level of matrices. Before we do so, we pause to mention that if $A = ((a_{ij}))$ denotes an $m \times n$ matrix, we shall use the symbol A^t denote the so-called **transpose matrix** which is, by definition, the $n \times m$ matrix with (i,j)-th entry equal to a_{ji}.

Proposition 3.2.4 *Suppose* $T \in \mathcal{L}(V,W)$ *and suppose* $B_V = \{e_1,\ldots,e_n\}$ *(resp.,* $B_W = \{f_1,\ldots,f_m\}$ *) is an orthonormal basis for* V *(resp.,* W *). Then,*

$$[T^*]_{B_W}^{B_V} = ([T]_{B_V}^{B_W})^t. \tag{3.2.4}$$

Proof : By definition, the left side of equation 3.2.4 is the $n \times m$ matrix with (i,j)-th entry equal to

$$< T^* f_j, e_i > = < f_j, Te_i > = < Te_i, f_j >$$

which is nothing but the (j,i)-th entry of the matrix on the right side of equation 3.2.4. $\Box$

We pause now to introduce a very important class of linear transformations between inner product spaces.

Proposition 3.2.5 *Let V, W be inner product spaces. The following conditions on a linear transformation $T : V \rightarrow W$ are equivalent:*
(i) if $\{e_1, \ldots, e_k\}$ is any orthonormal set in V, then $\{Te_1, \ldots, Te_k\}$ is an orthonormal set in W;
(ii) There exists some orthonormal basis $\{e_1, \ldots, e_n\}$ for V such that $\{Te_1, \ldots, Te_n\}$ is an orthonormal set in W;
(iii) $< Tx, Ty > = < x, y >$ for all x, y in V;
(iv) $T^*T = I_V$ (the identity operator on V);
(v) $||Tx|| = ||x||$ for all x in V.

Proof : $(i) \Rightarrow (ii)$: Trivial.

$(ii) \Rightarrow (iii)$: Let $\{e_1, \ldots, e_n\}$ be as in (ii). The hypothesis implies that $\{Te_1, \ldots, Te_n\}$ is an orthonormal basis for the subspace $W_0 = ran\ T$ of W. The validity of (iii) follows from Theorem3.1.8(b)(iii), thus:

$$\begin{aligned}
< Tx, Ty > &= < T(\sum_{i=1}^n < x, e_i > e_i) \ , \ T(\sum_{j=1}^n < y, e_j > e_j) \ > \\
&= \sum_{i,j=1}^n < x, e_i >< y, e_j > \ < Te_i, Te_j > \\
&= \sum_{i,j=1}^n < x, e_i >< y, e_j > \ \delta_{ij} \\
&= \sum_{i=1}^n < x, e_i >< y, e_i > \\
&= < x, y >
\end{aligned}$$

as desired.

$(iii) \Rightarrow (iv)$: The hypothesis implies that for all x, y in V, we have:

$$< T^*Tx, y > \ = \ < Tx, Ty > \ = \ < x, y > .$$

This means that the vector $(T^*Tx - x)$ is orthogonal to every vector y in V. Putting $y = (T^*Tx - x)$, we see that this implies that $T^*Tx - x = 0$ as desired.

$(iv) \Rightarrow (v)$: If $T^*T = I_V$, then it is clear that for all x in V, we have

$$||Tx||^2 = < Tx, Tx > = < T^*Tx, x > = < x, x > = ||x||^2.$$

$(v) \Rightarrow (i)$: It suffices to prove assertion (i) in the case $k = 2$. (Why?) If e_1 and e_2 are a pair of unit vectors which are orthogonal to one another, the hypothesis ensures that Te_1 and Te_2 are also a pair of unit vectors. It remains only to prove that $c = 0$ where $c = < Te_1, Te_2 >$. If a, b are arbitrary scalars, note that

$$||ae_1 + be_2||^2 = a^2 + b^2$$

while

$$\begin{aligned} ||T(ae_1 + be_2)||^2 &= ||aTe_1 + bTe_2||^2 \\ &= a^2||Te_1||^2 + b^2||Te_2||^2 + 2ab < Te_1, Te_2 > \\ &= a^2 + b^2 + 2abc. \end{aligned}$$

Hence the condition (v) implies that

$$a^2 + b^2 = a^2 + b^2 + 2abc.$$

Since a, b were arbitrary, it must be the case that $c = 0$ and the proof is complete. □

Definition 3.2.6 *A linear transformation T between two inner product spaces V and W which satisfies the equivalent conditions of Proposition 3.2.5 is called an* **isometry**.

The following exercises should help clarify the notion of an isometry to the reader.

EXERCISE 3.2.3 *(a) Let $T : V \to W$ be an isometry. Prove the following assertions:*
(i) The linear transformation T is 1-1, i.e., $x, y \in V$ and $Tx = Ty \Rightarrow x = y$; equivalently, ker $T = \{0\}$.
(ii) dim $V \leq$ dim W.

(b) If V, W are finite-dimensional inner product spaces, prove that the following conditions are equivalent:
(i) dim $V \leq$ dim W ;
(ii) There exists an isometry $T : V \rightarrow W$.

(c) Prove that the following conditions on an isometry $T : V \rightarrow W$ are equivalent:
(i) The linear transformation T maps V onto W , i.e., ran $T = W$.
(ii) dim $V =$ dim W.
(iii) $TT^* = I_W$, i.e., T^* is also an isometry.
(An isometry $T : V \rightarrow W$ which satisfies the equivalent conditions (i)-(iii) above of (c) is said to be **unitary***.)*

(d) If V, W are finite-dimensional inner product spaces, prove that the following conditions are equivalent:
(i) dim $V =$ dim W ;
(ii) There exists a unitary transformation $T : V \rightarrow W$.

(e) If V_0 is an arbitrary vector space and if $B = \{e_1, \ldots, e_n\}$ is any basis for V_0 , show that there exists a unique inner product that one can define on V_0 with respect to which B is an orthonormal basis for V_0.

The next exercise considers the special case $V = W$ and also looks at the matricial analogues of operator-theoretic statements concerning unitary transformations.

EXERCISE 3.2.4 *(a) Given a finite-dimensional vector space V , consider the set $GL(V) = \{T \in$ is invertible $\}$. (See Corollary 2.1.5 for the terminology.) Prove that $GL(V)$ is a group. (See Exercise 2.2.3.)*

(b) Show that the set $U(V) = \{T \in \mathcal{L}(V) : T$ is unitary $\}$ is a 'subgroup' of $GL(V)$ meaning that it is a subset of $GL(V)$ which is closed under the formation of products and inverses (and is consequently a group in its own right).

(b') Show that the set $O(n, I\!\!R) = \{T \in M_n(I\!\!R) : TT^t = I_n\}$ is a subgroup of $Gl(n, I\!\!R)$. (Elements of $O(n, I\!\!R)$ are referred to as **orthogonal** *matrices.)*

(c) Show that the set

$$\left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \right\}$$

is a subgroup of $O(4, I\!R)$.

Just as invertibility is precisely the property that an operator should have in order that it map a basis onto a basis, we have seen that unitariry is precisely the property that an operator on an inner product space should possess in order that it maps an orthonormal basis onto an orthonormal basis. With this in mind, the following definitions are natural : two linear operators T_1, T_2 on a finite-dimensional inner product space V are said to be **unitarily equivalent** if there exists a unitary operator U on V such that $T_2 = UT_1U^*$; and two matrices $T_1, T_2 \in M_n(I\!R)$ are said to be **orthogonally similar** if there exists an orthogonal matrix $U \in O_n(I\!R)$ such that $T_2 = UT_1U^t$.

The reader should have no difficulty in imitating the proofs of Lemma 2.2.3 and Theorem 2.2.4 and proving the following analogues.

EXERCISE 3.2.5 *(i) Two operators T_1, T_2 on a finite-dimensional inner product space V are unitarily equivalent if and only if there exist orthonormal bases B_1, B_2 of V such that* $[T_1]_{B_1} = [T_2]_{B_2}$.
(ii) Two $n \times n$ matrices T_1, T_2 are orthogonally similar if and only if there exists an operator T on an n-dimensional inner product space V and a pair of orthonormal bases B_1, B_2 of V such that $[T]_{B_i} = T_i, i = 1, 2$.

Before concluding this section, we outline, in an exercise, how the notion of an adjoint makes sense even when one works in arbitrary vector spaces (which have no inner product structure prescribed on them) and why the notion we have considered in this section is a special case of that more general notion.

EXERCISE 3.2.6 *If V, W are abstract vector spaces and if $T \in \mathcal{L}(V, W)$, define $T' : W^* \to V^*$ by the prescription*

$$(T'\phi)\ (v) = \phi\ (Tv) \quad \textit{whenever} \quad \phi \in W^*,\ v \in V.$$

The linear transformation T' is called the 'transpose' of the transformation T.

(a) Prove analogues of Exercise 3.2.2 and Proposition 3.2.4 with 'transpose' in place of 'adjoint'.

(b) If V, W are inner product spaces, and if $\phi_V : V \rightarrow V^*$ denotes the isomorphism guaranteed by the Riesz representation theorem - see Exercise 3.2.1 - show that

$$T' = \phi_V \circ T^* \circ \phi_W^{-1}.$$

(c) Use (b) above to give an alternate solution to (a) above.

3.3 Orthogonal Complements and Projections

This section is devoted to the important notion of the so-called orthogonal complement ${V_0}^{\perp}$ of a subspace of a finite-dimensional inner product space V and to the (probably even more important notion) of the 'orthogonal projection of V onto V_0'.

Lemma 3.3.1 *Suppose V_0 is a subspace of a finite-dimensional inner product space V. Let $\{e_i : 1 \leq i \leq m\}$ be any orthonormal basis for V_0 and let $\{f_i : 1 \leq i \leq (n-m)\}$ be any set of vectors such that $\{e_1, \ldots, e_m,\ f_1, \ldots, f_{n-m}\}$ is an orthonormal basis for V. Then the following conditions on a vector y in V are equivalent:*
(i) $< y, x > \ = \ 0$ for all x in V_0;
(ii) $y \ \in \ \bigvee\{f_1, \ldots, f_{n-m}\}$, i.e., $y \ = \ \sum_{j=1}^{n-m} < y, f_j > \ f_j$.

Proof: $(i) \Rightarrow (ii)$: On the one hand, the assumption that $\{e_1, \ldots, e_m,\ f_1, \ldots, f_{n-m}\}$ is an orthonormal basis for V implies that

$$y \ = \ \sum_{i=1}^{m} < y, e_i > \ e_i \ + \ \sum_{j=1}^{n-m} < y, f_j > \ f_j;$$

the assumption (i) implies that the first sum above is zero, thus ensuring the validity of (ii).

$(ii) \Rightarrow (i)$: If $x \in V_0$, the assumption about $\{e_i : 1 \leq i \leq m\}$ implies that $x \ = \ \sum_{i=1}^{m} < x, e_i > \ e_i$; while the assumption (ii) implies that $y \ = \ \sum_{j=1}^{n-m} < y, f_j > \ f_j$; since $< e_i, f_j > \ = \ 0$, it follows at once that $< y, x > \ = \ 0$. Since x was an arbitrary element of V_0 , this proves (i). $\square$

The next result is fundamental and introduces a very important notion.

Theorem 3.3.2 *Let V_0 denote an arbitrary subspace of a finite-dimensional inner product space V. Define*

$$V_0{}^{\perp} \;=\; \{\; y \in V \;:\; < y, x > \;=\; 0 \;\; \text{for all} \;\; x \;\; \text{in} \;\; V_0 \;\} \;. \tag{3.3.5}$$

Then,

(a) $V_0{}^{\perp}$ is a subspace of V with

$$dim\ V_0 \;+\; dim\ V_0{}^{\perp} \;=\; dim\ V.$$

(b) If $B_0 \cup B_1$ is any orthonormal basis for V such that B_0 is an orthonormal basis for V_0, then B_1 is an orthonormal basis for $V_0{}^{\perp}$; and conversely, if B_0 (resp., B_1) is any orthonormal basis for V_0 (resp., $V_0{}^{\perp}$), then $B_0 \cup B_1$ is an orthonormal basis for V.

(c) $(V_0{}^{\perp})^{\perp} \;=\; V_0$.

(d) Every vector $z \in V$ is uniquely expressible in the form $z \;=\; x \;+\; y$, where $x \in V_0,\ y \in V_0{}^{\perp}$.

Proof: If $B_0 = \{e_1, \ldots, e_m\}$ is an orthonormal basis for V_0 and if $B_1 = \{f_1, \ldots, f_{n-m}\}$ is such that $B_0 \cup B_1$ is an orthonormal basis for V, it follows at once from Lemma 3.3.1 that $V_0{}^{\perp} \;=\; \bigvee \{f_1, \ldots, f_{n-m}\}$ and the truth of (a) and the first half of (b) follow immediately; the second half of (b) follows from the fact that $dim\ V_0 \;+\; dim\ V_0{}^{\perp} \;=\; dim\ V$. The truth of assertions (c) and (d) follow immediately from (b). □

Corollary 3.3.3 *Let V_0 be a subspace of a finite-dimensional vector space V.*

(a) Define $P : V \to V$ as follows : if $z = x + y$ is the canonical decomposition of a typical vector z in V as in Theorem 3.3.2 (d), then $Pz = x$. (This transformation is called the **orthogonal projection onto the subspace** *V_0.) Then,*

(i) $Pz \;=\; \sum_{i=1}^{m} < z, e_i > \; e_i$ where $\{e_i \;:\; 1 \leq i \leq m\}$ is any orthonormal basis for V_0;
(ii) $P \in \mathcal{L}(V)$;
(iii) $P = P^* = P^2$;
(iv) $ran\ P = V_0$; in fact, $x \in V_0 \Leftrightarrow Px = x$; and
(v) $ker\ P \;=\; V_0^{\perp}$.

(b) Conversely, if $P \in \mathcal{L}(V)$ *and if* $P = P^* = P^2$ *, then* P *is the orthogonal projection onto the subspace* $V_0 = ran\ P$.

(c) If P *is the orthogonal projection onto a subspace* V_0 *, then* $I_V - P$ *is the orthogonal projection onto* $V_0^\perp$.

Proof : (a) Assertion (i) is immediate from the previous Proposition, and assertion (ii) is an immediate consequence of (i).

(iii) If $x, y \in V$, and if $\{e_1, \ldots, e_m\}$ is any orthonormal basis for V_0 , it follows immediately from (i) that

$$\begin{aligned} < Px, y > &= < (\sum_{i=1}^{m} < x, e_i > \ e_i\ )\ ,\ y > \\ &= \sum_{i=1}^{m} < x, e_i > \ < e_i, y > \ , \end{aligned}$$

from which it follows immediately that

$$< Px, y > \ = \ < Px, Py > \ = \ < x, Py > \quad \text{for all } x, y \in V. \tag{3.3.6}$$

This proves that P is **self-adjoint** - i.e., it is its own adjoint. The fact that P is **idempotent** - i.e., it is equal to its square - follows immediately from property (i) above.

(iv) The assertion that if $x \in V$, then $x \in V_0 \Leftrightarrow Px = x$, is also an immediate consequence of property (i). This statement clearly proves the equality $ran\ P = V_0$.

Assertion (v) is also an immediate consequence of (i).

(b) Suppose $P \in \mathcal{L}(V)$ and suppose P is self-adjoint and idempotent. Define $V_0 = ran\ P$. Then V_0 is clearly a subspace of V . The idempotence of P implies that $Px = x$ for all $x \in V_0$, while the self-adjointness of P implies that if $y \in V_0^\perp$, then

$$||Py||^2 = < Py, Py > = < y, P^2 y > = 0$$

as $P^2 y \in ran\ P = V_0$, whence $Py = 0$ for all $y \in V_0{}^\perp$. It follows that if $z = x + y$ is the canonical decomposition of an arbitrary element z of V as in Theorem 3.3.2(d), then $Pz = x$. In other words, the operator P is precisely the orthogonal projection onto V_0.

(c) This is an immediate consequence of the definitions. □

EXERCISE 3.3.1 *(a) Compute the matrix* $[P]_B$ *, where* P *is the orthogonal projection of* $I\!R^3$ *onto the subspace* $\{(x,y,z) \in I\!R^3 : x+y+z = 0\}$ *and* B *denotes the standard basis for* $I\!R^3$.

(b) If $B = B_0 \cup B_1$ *is an orthonormal basis for an arbitrary vector space* V *, and if* P *denotes the orthogonal projection of* V *onto the subspace* $V_0 = \bigvee B_0$ *, compute* $[P]_B$.

EXERCISE 3.3.2 *(a) Suppose* P *is the orthogonal projection of a vector space* V *onto a subspace* V_0 *. Consider the operator (of 'reflection' in the subspace* V_0 *) defined by* $U = 2P - I_V$. *Show that* $U = U^*$ *and* $U^2 = I_V$; *and deduce in particular that* U *is a unitary operator on* V.

(b) Conversely, if U *is a self-adjoint unitary operator on a finite- dimensional vector space* V *, show that* U *is the operator of reflection in some subspace. (Hint: In (a) above, express* P *and* V_0 *in terms of* U*.)*

The next result states an important relationship between a linear transformation and its adjoint.

Proposition 3.3.4 *Suppose* $T \in \mathcal{L}(V,W)$ *, where* V *and* W *are finite-dimensional inner product spaces. Then,*

$$(ran\ T)^{\perp} = ker\ T^* \tag{3.3.7}$$

and in particular,

$$\rho\ (T) = \rho\ (T^*) \tag{3.3.8}$$

where of course the symbol ρ *denotes the rank.*

Proof : If $w \in W$, then we find that

$$\begin{aligned} w \in (ran\ T)^{\perp} &\Leftrightarrow\ < Tv, w > = 0 \text{ for all } v \in V \\ &\Leftrightarrow\ < v, T^*w > = 0 \text{ for all } v \in V \\ &\Leftrightarrow\ T^*w = 0 \end{aligned}$$

thus proving the first assertion.

As for the second, it follows from the first assertion and the rank-nullity theorem that

$$\rho(T) = dim\ W - dim(ran\ T)^{\perp} = dim\ W - \nu(T^*) = \rho(T^*)$$

as desired. $\square$

We conclude this section with some remarks concerning the implications that equation 3.3.8 has for matrices. To start with, we shall identify $M_{m\times n}(\mathbb{R})$ with $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ by identifying an $m \times n$ matrix T with the linear transformation $\tilde{T}$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ for which $[\tilde{T}]_{B_n}^{B_m} = T$, where we use the symbol B_k to denote the standard basis for $\mathbb{R}^k$. Furthermore, it will be convenient to identify R^n with $M_{n\times 1}(\mathbb{R})$ in the natural manner. Then, the columns of T are, by definition, nothing but the images of the standard basis under the linear transformation $\tilde{T}$ and consequently the columns of T span $ran\ \tilde{T}$. It follows easily that there is some subset of the set of columns of T which is (a minimal spanning set and consequently) a basis for $ran\ \tilde{T}$. Thus, we find that $\rho(\tilde{T})$ is nothing but 'the maximum number of linearly independent columns of the matrix T '. Some books refer to the expression within quotes in the last sentence as the **column-rank** of the matrix T. In analogy, if we define the **row-rank** of the matrix T to be the 'maximum number of linearly independent rows of the matrix T , a moment's thought reveals that the row-rank of the matrix T is nothing but the column-rank of the matrix T^t . It is now a consequence of Proposition 3.2.4 and equation 3.3.8 that we have

$$\begin{aligned}\text{row-rank of } T &= \rho(\tilde{T}^{\,*}) \\ &= \rho(\tilde{T}) \\ &= \text{column-rank of } T\end{aligned}$$

and the column value of these quantities is simply called the rank of the matrix T .

The following exercise is a nice consequence of the foregoing remarks.

EXERCISE 3.3.3 *By definition, a* **submatrix** *of a matrix is the matrix obtained by 'deleting some rows and columns'. Show that the rank of a rectangular matrix is the largest size among all invertible (necessarily sqaure) submatrices.*

3.4 Determinants

In this section, we shall introduce a very important scalar that is associated with every operator on a finite-dimensional vector space. Actually we shall define the determinant of a square matrix and define the determinant of the operator T as the determinant of the matrix $[T]_B$ where B is any basis for V, and use Theorem 2.2.4 to verify that this definition depends only on the operator T and is independent of the basis B.

Suppose T is an $n \times n$ matrix; let us write $\mathbf{v}_j$ for the j-th column of the matrix T, thought of as a vector in $I\!R^n$. The determinant of T will be defined as a number whose absolute value equals the 'n-dimensional volume' of the 'n-dimensional parallelopiped with sides $\mathbf{v}_1, \ldots, \mathbf{v}_n$'. The sign of the determinant will indicate the 'orientation of the frame determined by the above vectors'. In order to make all this precise, let us first consider the cases $n = 2$ and $n = 3$ where our geometric intuition is strong and then 'extrapolate' to higher n.

The case $\mathbf{n = 2}$:

Suppose

$$T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

Put $\mathbf{u} = (a, b), \mathbf{v} = (c, d)$. We wish to compute the 'signed area' of the parallelogram determined by these vectors. Notice that the vector $\mathbf{n} = \frac{1}{||(a,b)||}(-b, a)$ is a unit vector which is perpendicular to $\mathbf{u}$. Thus it is seen that the parallelogram determined by $\mathbf{u}, \mathbf{v}$ has 'base' $= ||\mathbf{u}||$ and 'height' $= |\mathbf{v} \cdot \mathbf{n}|$; thus the area of this parallelogram is equal to $|(c, d) \cdot (-b, a)| = |(ad - bc)|$.

It is also not hard to see that if we write P_1, P_2, Q for the points in the plane represented by the vectors $\mathbf{u}, \mathbf{v}, \mathbf{n}$ respectively, then the expression $ad - bc$ is positive or negative according as the angle P_2OP_1 is less than or greater than 180^o. In mathematics, it is customary to consider the counter-clockwise (rather than the clockwise) direction as the 'positive one'.

Thus, we see that the expression $(ad - bc)$ has the following features :

(i) its magnitude is the area of the parallelogram spanned by the vectors $\mathbf{u} = (a, b)$ and $\mathbf{v} = (c, d)$;

(ii) its sign is positive precisely when the following happens : if we move in the mathematically positive direction from $\mathbf{u}$, we will come to $\mathbf{v}$ before we come to $-\mathbf{v}$.

Hence - with T as above - if we define $det\ T = ad - bc$, we find that this definition meets the requirements of defining the 'signed area' of the parallelogram spanned by the columns of T. We shall also write

$$T = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc. \tag{3.4.9}$$

Notice, in particular, that two vectors $(a, c), (b, d)$ in $I\!R^2$ are linearly dependent if and only if $(ad - bc) = 0$.

The case n = 3 :

Suppose now that

$$T = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & k \end{bmatrix}. \tag{3.4.10}$$

We first consider the volume of the parallelopiped spanned by the three vectors $\mathbf{u} = (a, b, c)$, $\mathbf{v} = (d, e, f)$, $\mathbf{w} = (g, h, k)$ in $I\!R^3$. A moment's thought must show that this volume is given by $Vol = B \times h$ where B denotes the (two-dimensional) area of the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$, and h equals the magnitude of the orthogonal projection of $\mathbf{w}$ onto the line perpendicular to the plane $\bigvee\{\mathbf{u}\ , \mathbf{v}\}$. (Notice that the volume must be zero if all the three vectors lie in a plane, i.e., if the set $\{\mathbf{u}\ , \mathbf{v}\ , \mathbf{w}\}$ is linearly dependent; so we assume in the following that these vectors are indeed linearly independent.)

In order to compute these quantities, we first need to be able to find
(a) the two-dimensional area B of the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$ (in the plane - i.e., the two-dimensional subspace - spanned by $\mathbf{u}$ and $\mathbf{v}$); and
(b) the vector $\mathbf{n}$ orthogonal to $\mathbf{u}$ and $\mathbf{v}$ - which is unique up to scaling by a constant.

As for (a), it must be clear, after a moment's deliberation, that if θ is the angle subtended at the origin by the vectors $\mathbf{u}$ and $\mathbf{v}$, then $B = ||\mathbf{u}||\ ||\mathbf{v}||\ |sin\ \theta|$; since $||\mathbf{u}||\ ||\mathbf{v}||\ cos\ \theta = \mathbf{u} \cdot \mathbf{v}$, it follows at once that

$$B^2 = ||\mathbf{u}||^2\ ||\mathbf{v}||^2\ (\ 1 - cos^2\ \theta)$$

$$
\begin{aligned}
&= ||\mathbf{u}||^{\,2}||\mathbf{v}||^{\,2} \;-\; (\mathbf{u}\cdot\mathbf{v})^{\,2} \\
&= (a^{\,2}+b^{\,2}+c^{\,2})(d^{\,2}+e^{\,2}+f^{\,2}) \;-\; (ad+be+cf)^{\,2} \\
&= a^{\,2}(e^{\,2}+f^{\,2})+b^{\,2}(d^{\,2}+f^{\,2})+c^{\,2}(d^{\,2}+e^{\,2}) \;-2(adbe+adcf+becf) \\
&= (ae-bd)^{\,2} \;+\; (af-cd)^{\,2} \;+\; (bf-ce)^{\,2}
\end{aligned}
$$

and hence,

$$
B = || \left(\pm \begin{vmatrix} b & c \\ e & f \end{vmatrix}, \; \pm \begin{vmatrix} a & c \\ d & f \end{vmatrix}, \; \pm \begin{vmatrix} a & b \\ d & e \end{vmatrix} \right) || \,. \tag{3.4.11}
$$

In particular, if $\mathbf{u}$ and $\mathbf{v}$ are linearly independent, then, the vector displayed in the above equation is non-zero.

We now proceed to (b). Notice that the 3×2 submatrix of T obtained by deleting the third column has column-rank equal to 2 (in view of the assumed linear independence of $\mathbf{u}$ and $\mathbf{v}$). It follows from Exercise 3.3.3 and the already discussed case ($n=2$) of the determinant that one of the three determinants

$$
\begin{vmatrix} a & b \\ d & e \end{vmatrix}, \; \begin{vmatrix} a & c \\ d & f \end{vmatrix}, \; \begin{vmatrix} b & c \\ e & f \end{vmatrix}
$$

must be non-zero. Assume (without loss of generality) that

$$
\begin{vmatrix} a & b \\ d & e \end{vmatrix} \neq 0 \,. \tag{3.4.12}
$$

Now, the vector $\mathbf{n} = (x,y,z)$ is perpendicular to the plane $\bigvee\{\mathbf{u}\,,\mathbf{v}\,\}$ if and only if the following equations are satisfied:

$$
\begin{aligned}
ax+by+cz &= 0 \qquad (3.4.13)\\
dx+ey+fz &= 0. \qquad (3.4.14)
\end{aligned}
$$

Eliminating y from these equations, we find that

$$
(ae-bd)x \;+\; (ce-bf)z \;=\; 0.
$$

Thanks to our assumption that equation 3.4.12 is satisfied, it follows from the last equation that the vector (x, z) must be proportional to the vector

$$\left(\begin{vmatrix} b & c \\ e & f \end{vmatrix}, \begin{vmatrix} a & b \\ d & e \end{vmatrix} \right).$$

Similarly, by eliminating x from equation 3.4, we find that the vector (z, y) must be proportional to the vector

$$\left(\begin{vmatrix} a & b \\ d & e \end{vmatrix}, - \begin{vmatrix} a & c \\ d & f \end{vmatrix} \right).$$

We conclude from the foregoing that in order that the vector (x, y, z) be perpendicular to both the vectors (a, b, c) and (d, e, f), it is necessary and sufficient that (x, y, z) be a multiple of the vector

$$\left(\begin{vmatrix} b & c \\ e & f \end{vmatrix}, - \begin{vmatrix} a & c \\ d & f \end{vmatrix}, \begin{vmatrix} a & b \\ d & e \end{vmatrix} \right).$$

In some books on vector calculus, the reader might have seen the vector displayed above being referred to as the so-called **cross-product** of the vectors (a, b, c) and (d, e, f). Thus, given vectors $\mathbf{u} = (a, b, c), \mathbf{v} = (d, e, f)$, their cross-product is the vector displayed above- and usually denoted by $\mathbf{u} \times \mathbf{v}$. The foregoing analysis shows that this vector is perpendicular to the plane spanned by $\mathbf{u}$ and $\mathbf{v}$, and has magnitude (i.e., norm) equal to the area of the parallelogram spanned by the vectors $\mathbf{u}$ and $\mathbf{v}$.

Finally, if we write $\mathbf{n} = \frac{\mathbf{u} \times \mathbf{v}}{||\mathbf{u} \times \mathbf{v}||}$, it follows that the desired volume of the parallelopiped spanned by the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$is given by

$$Vol = ||\mathbf{u} \times \mathbf{v}|| \; |\mathbf{w} \cdot \mathbf{n}| = |\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})| \, .$$

Furthermore, it is not too hard to verify that the scalar $\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})$ is positive or negative according as whether the co-ordinate frame $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is ‘right-handed’ or ‘left-handed’; we thus find that all the requirements of the determinant, as we set them out to be, are satisfied by the definition

$$\begin{aligned} det\ T &= \begin{vmatrix} a & d & g \\ b & e & h \\ c & f & k \end{vmatrix} \\ &= g(bf - ec) \ - h(af - dc) \ + k(ae - db) \\ &= aek \ + \ bfg \ + \ cdh \ - \ afh \ - \ bdk \ - \ ceg \ . \end{aligned}$$

EXERCISE 3.4.1 *If* $\mathbf{u}$,$\mathbf{v}$ *and* $\mathbf{w}$ *are any three vectors in* $I\!R^3$ *, show that*

(i) $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \mathbf{v} \cdot (\mathbf{w} \times \mathbf{u}) = \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})$, *and hence deduce that the determinant of a* 3×3 *matrix is unaffected by a cyclic permutation of the columns.*

(ii) More generally, show that if $\tilde{T}$ *is the matrix obtained by interchanging two columns of a '*3×3 *matrix* T *, then* $det\ \tilde{T} = -\ det\ T$ *.*

(iii) Even more generally than (ii), if σ *is any permutation of the set* $\{1, 2, 3\}$ *and if* T_σ *denotes the matrix obtained by permuting the columns of a* 3×3 *matrix* T *according to the permutation* σ *, then show that* $det\ T_\sigma = \epsilon_\sigma\ det\ T$ *, where* ϵ_σ *is the so-called 'sign' of the permutation defined to be equal to* ± 1 *according as* σ *is an odd or an even permutation.*

(iv) If $S, T \in M_3(I\!R)$ *, show that* $det(ST) = (det\ S)(det\ T)$*.*

(v) Can you see that (i) is an immediate consequence of (ii), that (ii) is a special case of (iii) and that (iii) is a special case of (iv)?

(vi) Deduce from (iv) that any two 3×3 *matrices which are similar must necessarily have the same determinant.*

(vii) Show that $det\ T = det\ T^t$ *for all* T *in* $M_3(I\!R)$*.*

It is worth explicitly stating a couple of consequences of the preceding exercises.

The first concerns the possibility of 'expanding a determinant along any row'. To be explicit, if $T = ((t_{ij}))$ is any 3×3 matrix, and if $1 \leq i, j \leq 3$, let us write $T_{(ij)}$ for the '2×2 submatrix obtained by deleting the i-th row and j-th column of T ; then it follows from Exercise 3.4.1 (i) that

$$det\ T = \sum_{i=1}^{3} (-1)^{i+j} t_{ij}\ det\ T_{(ij)} \ \text{ for } \ 1 \leq j \leq 3; \qquad (3.4.15)$$

it follows from Exercise 3.4.1(vii) that we also have 'expansion of a determinant along rows', i.e.,

$$det\ T \;=\; \sum_{j=1}^{3}(-1)^{i+j}t_{ij}\ det\ T_{(ij)} \ \text{ for } \ 1 \leq i \leq 3. \tag{3.4.16}$$

The second concerns the following 'closed-form' expansion for the determinant which follows easily from Exercise 3.4.1 (iii) : if $T = ((t_{ij}))$ is any 3×3 matrix, then

$$det\ T \;=\; \sum_{\sigma \in S_3} \epsilon_\sigma \left(\prod_{i=1}^{3} t_{i\sigma(i)}\right) \tag{3.4.17}$$

where we have used the notation S_3 for the group of all permutations of the set $\{1,2,3\}$.

The case of a general n :

Taking our cue from the case $n = 3$, we define the determinant of an arbitrary $n \times n$ matrix $T = ((t_{ij}))$ by the formula

$$det\ T \;=\; |((t_{ij}))| \;=\; \sum_{\sigma \in S_n} \epsilon_\sigma \left(\prod_{i=1}^{n} t_{i\sigma(i)}\right) \tag{3.4.18}$$

where we have used the notation S_n for the group of all permutations of the set $\{1,\ldots,n\}$.

We shall not go into any proofs here, but just content ourselves with the remark that all the properties for determinants listed in Exercise 3.4.1 (with the exception of (i) which is valid only for odd n) and the subsequent remarks are valid with 3 replaced by n . Most importantly, the magnitude of the determinant of T continues to equal the 'n-dimensional volume' of the parallelopiped spanned by the columns of T and the sign of the determinant describes the 'orientation' of the frame given by $\{Te_1,\ldots,Te_n\}$.

We list the basic facts concerning determinants as a theorem.

Theorem 3.4.1 *(a) The assignment $T \rightarrow det\ T$, as a mapping from $M_n(I\!R)$ to $I\!R$, satisfies the following conditions:*
(i) (expansion along rows and columns)

$$det\ T \;=\; \sum_{j=1}^{n}(-1)^{i+j}t_{ij}\ det\ T_{(ij)} \ \textit{ for } \ 1 \leq i \leq n$$

$$= \sum_{j=1}^{n} (-1)^{i+j} t_{ij} \ det \ T_{(ij)} \quad for \ \ 1 \leq i \leq n$$

where, as before, the symbol $T_{(ij)}$ *determines the submatrix obtained by deleting the* i*-th row and* j*-th column of* T*.*

(ii) The determinant function is multilinear as a function of the columns of the matrix, in the sense that if $S, T \in M_n(I\!R)$ *satisfy* $Se_i = Te_i$ *for all* $i \neq j$ *, for some* j *, where* $B = \{e_1, \ldots, e_n\}$ *denotes the standard basis for* $I\!R^n$ *,then* $det \ (\alpha S + \beta T) = \alpha(det \ S) + \beta(det \ T) \ \forall \ \alpha, \beta \in I\!R$*.*

(iii) The determinant function is an 'alternating function' in the sense that if $\sigma \in S_n$ *and if, for* T *in* $M_n(I\!R)$ *, we let* T_σ *denote the matrix obtained by permuting the columns of* T *according to* σ *- i.e.,* $(T_\sigma)_{ij} = (T)_{i\sigma(j)}$ *- then* $det \ T_\sigma = \epsilon_\sigma \ det \ T$ *.*

(iv) If $S, T \in M_n(I\!R)$ *, then* $det \ (ST) = (det \ S)(det \ T)$ *, and hence similar matrices have the same determinant.*

(b) In view of (a)(iv) above, we see that if $T \in \mathcal{L}(V)$ *, where* V *is a finite-dimensional space, then the value of* $det \ [T]_B$ *does not depend upon the choice of the basis* B *of* V *and we define this common value as* $det \ T$*.*

(i) If $T \in \mathcal{L}(V)$ *, then a necessary and sufficient condition for the invertibility of* T *is that* $det \ T = 0$*.*

$\square$

Chapter 4

The Spectral Theorem

4.1 The real case

This chapter is devoted to probably the most fundamental result in linear algebra. In order to get a feeling for what the spectral theorem says, recall Exercises 1.1.3 and 3.3.1(a). In the former, one was confronted with a seemingly complicated-looking matrix and wished to obtain some insight into what the operator represented by that matrix 'looked like'; whereas, in the latter, one was given a perfectly decent looking operator and was presented with the somewhat tame problem of computing its matrix; at the end of the computation, what one saw was the complicated looking matrix of the former exercise. Rather than leaving the solution of problems of the former sort to such a fortuitous coincidence, the spectral theorem points a way to understanding fairly general matrices by geometric means.

Actually, part (b) of Exercise 3.3.1 indicates the kind of thing that happens. Explicitly, the purpose of that exercise was to convince the reader that if P is a projection operator on some finite-dimensional vector space V - which can be recognised by the easily verified algebraic conditions $P = P^* = P^2$ - then there exists an orthonormal basis B for V such that the matrix $[P]_B$ is particularly simple : this matrix has zero entries off the main diagonal, and on the main diagonal there is a string of 1's followed by a string of 0's, with the number of 1's being exactly equal to the rank of P (or equivalently the dimension of the subspace onto which P projects).

The (real case of the) spectral theorem says something entirely similar. It says that if

T is a self-adjoint operator - i.e., $T^* = T$ - on a finite-dimensional (real) inner product space V, then there exists an orthonormal basis $B = \{e_1, \ldots, e_n\}$ for V such that $[T]_B$ is a **diagonal matrix** , i.e., has only zero entries off the main diagonal. An equivalent formulation is the following.

Theorem 4.1.1 (The (real) spectral theorem) *If T is a self-adjoint operator on a finite-dimensional (real) inner product space, then there exists an orthonormal basis $\{e_1, \ldots, e_n\}$ for V and real numbers $\lambda_1, \ldots, \lambda_n$ such that $Te_j = \lambda_j e_j$ for $1 \leq j \leq n$.*

□

There are two ingredients in the proof, which we establish as separate lemmas.

Lemma 4.1.2 *Let V_0 be a subspace of a finite-dimensional inner product space V. The following conditions on an operator T on V are equivalent:*
(i) $T(V_0) \subseteq V_0$;
(ii) $T^*({V_0}^\perp) \subseteq {V_0}^\perp$.

Proof : $(i) \Rightarrow (ii)$: Let $y \in {V_0}^\perp$, $x \in V_0$ be arbitrary. The assumption (i) implies that $Tx \in V_0$ and consequently,

$$< T^*y, x > \; = \; < y, Tx > \; = \; 0 \; .$$

Since x was an arbitrary element of V_0, this shows that $T^*y \in {V_0}^\perp$. Since y was an arbitrary element of ${V_0}^\perp$, we have indeed shown that $(i) \Rightarrow (ii)$.

$(i) \Rightarrow (ii)$: This follows from the implication $(i) \Rightarrow (ii)$ applied with the roles of (V_0, T) and $({V_0}^\perp, T^*)$ interchanged. □

Bsfore proceeding further, we establish some terminology and elementary properties concerning the notion discussed in the above lemma.

Exercise 4.1.1 *Let $T \in \mathcal{L}(V)$, with V a finite-dimensional inner product space, let V_0 be a subspace of V, and let P denote the orthogonal projection onto V_0.*

(a) Prove that the following conditions are equivalent:
(i) $T(V_0) \subseteq V_0$;

(ii) $TP = PTP$ *.*

(When these equivalent conditions are satisfied, we say that V_0 *is* **invariant** *or stable under* T*, or that* V_0 *is an invariant subspace for* T*.)*

(b) Use (a) twice to give an alternate proof of Lemma 4.1.2.

(c) If V_0 *is one-dimensional, show that* V_0 *is invariant under* T *if and only if there exists a scalar* λ *such that* $Tx = \lambda x$ *for all* x *in* V_0*.*

(d) What are all the invariant subspaces of the following matrix (which is, naturally, viewed as the operator on $\mathbb{R}^3$ *which it represents with respect to the standard basis) :*

$$N = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad ?$$

(Hint: Look at N^2 *and* N^3*.)*

Before proceeding any further, it will help matters if we introduce some notation for a notion which has already appeared more than once. If $T \in \mathcal{L}(V)$ and if $x \in V$ satisfies $x \neq 0$ and $Tx = \lambda x$ for some scalar λ , then x is called an **eigenvector** of T corresponding to the **eigenvalue** λ. We shall also find it convenient to have to use one of these expressions in the absence of the other : thus, we shall say that a scalar λ (resp. a non-zero vector x) is an eigenvalue (resp., eigenvector) of an operator T if there exists a non-zero vector x (resp. a scalar λ) such that $Tx = \lambda x$. The (solution to the) following 'trivial' exercise might help clarify some of these notions.

EXERCISE 4.1.2 *What are the eigenvalues and eigenvectors of the operator* N *of Exercise 4.1.1 (d) ?*

We prepare the ground for the second ingredient in the proof of the spectral theorem, with another exercise.

EXERCISE 4.1.3 *Consider the operator* T *on* $\mathbb{R}^3$ *represented (with respect to the standard basis) by the matrix*

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad .$$

(a) What are the eigenvalues of T?
(b) Compute the number

$$\lambda \ = \ sup \ \{ \ < Tx, x > \ : \ x \ \in \ I\!\!R^{\,3}, ||x|| \ = \ 1 \ \}$$

and describe the set $\{x \ \in \ I\!\!R^{\,3} \ : \ < Tx, x > \ = \ \lambda \ \ and \ \ ||x|| \ = \ 1\}$.

Lemma 4.1.3 *Let T be a self-adjoint operator on a finite-dimensional (real) inner product space V. Then T has an eigenvector.*

Proof : We shall prove that the number

$$\alpha \ = \ sup \ \{ \ < Tx, x > \ : \ x \ \in \ V \ , \ ||x|| \ = \ 1 \ \}$$

is an eigenvalue of T.

First of all, this number is actually a maximum - i.e., the supremum is attained at some point - and consequently finite. Reason : if $((t_{ij}))$ is the matrix representing the operator T with respect to any orthonormal basis, then

$$\alpha \ = \ sup \ \{ \ \sum_{i,j=1}^{n} t_{ij}\xi_i\xi_j \ : \ \xi_1, \ldots, \xi_n \ \in \ I\!\!R, \ \sum_{i=1}^{n} {\xi_i}^{\,2} \ = 1 \ \}$$

and thus α is nothing but the supremum of a quadratic polynomial over the unit sphere in $I\!\!R^n$; and any continuous function on a compact set is bounded and attains its bounds.

The rest of the proof consists in showing that if x_1 is any unit vector in V such that $< Tx_1, x_1 > \ = \ \alpha$, then in fact $Tx_1 \ = \ \alpha x_1$.

We first argue that the problem faced in proving the assertion of the last paragraph is a two-dimensional one, as follows: Suppose, to the contrary, that there exists a unit vector $x_1 \in V$ such that $< Tx_1, x_1 > \ = \ \alpha$ and that the set $\{x_1, \ Tx_1\}$ is linearly independent. Let $V_0 \ = \ \bigvee \ \{x_1, \ Tx_1\}$, let P denote the orthogonal projection onto V_0 and consider the operator T_0 on the 2-dimensional space V_0 defined by $T_0x \ = \ PTx$ for all x in V_0. It is an easy matter to verify that :
(i) $< T_0x, x > \ = \ < Tx, x >$ for all x in V_0; and consequently,
(ii) $\alpha \ = \ sup \ \{ \ < T_0x, x > \ : \ x_0 \ \in \ V \ , \ ||x|| \ = \ 1 \ \}$ (since $x_1 \ \in \ V_0$).

Let $\{x_1,\ x_2\}$ be the result of applying the Gram-Schmidt process to $\{x_1,\ Tx_1\}$ and let $B_0 = \{x_1, x_2\}$; it follows from $< Tx_1, x_1 > = \alpha$ that

$$[T_0]_{B_0} = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} ,$$

for some scalars β, γ , the symmetry of the matrix being a consequence of the assumed self-adjointness of T. Our assumption that x_1 is not an eigenvector of T implies that $\beta \neq 0$ and we shall deduce a contradiction from this.

Since any unit vector in V_0 is of the form $(cos\ \theta)x_1 + (sin\ \theta)x_2$ for some θ , the assumptions on α imply that

$$\alpha \geq \alpha\ cos\ ^2\theta + 2\beta\ cos\ \theta sin\ \theta + \gamma\ sin\ ^2\theta \quad \text{for all } \theta.$$

It follows that

$$(\alpha - \gamma) \geq 2\beta\ cot\ \theta \quad \text{whenever } 0 < \theta < \pi$$

which is clearly possible only if $\beta = 0$, and the proof is complete. □

Proof of the Spectral Theorem : The proof is by induction on the dimension of the underlying vector space V. The theorem is trivial if V is one-dimensional; so assume the theorem is known to hold for all $(n-1)$- dimensional vector spaces, and suppose V is n-dimensional. Lemma 4.1.3 tells us that there exists at least one 1-dimensional invariant subspace for T , say V_0. Appeal now to Lemma 4.1.2 and the self-adjointness of T to conclude that $V_0{}^\perp$ is an $(n-1)-$dimensional invariant subspace for T. Since the restriction of a self-adjoint operator to an invariant subspace is again self-adjoint - Verify ! - we may appeal to the induction hypothesis to find an orthonormal basis B_2 for $V_0{}^\perp$ consisting of eigenvectors for (the restriction to V_0 of T and hence for) T. Set $B = B_1 \cup B_2$, where B_2 is one of the two possible orthonormal bases for V_0 . Then B is an orthonormal basis for V which **diagonalises** T - in the sense that $[T]_B$ is a diagonal matrix. □

EXERCISE 4.1.4 *Let T be a self-adjoint operator on a finite-dimensional inner-product space.*

(a) Show that $\sup \{ < Tx, x > \ : x \in V , ||x|| = 1 \}$ *is the largest eigenvalue of* T.
(b) Show that $\inf \{ < Tx, x > \ : x \in V , ||x|| = 1 \}$ *is the smallest eigenvalue of* T.
(c) Can you think of such a description for the 'intermediate' eigenvalues ?

Before proceeding further, note that an obvious reformulation of the spectral theorem in terms of matrices is as follows. (In the theorem, we use the adjective **symmetric** to describe a matrix which is equal to its own transpose.)

Theorem 4.1.4 *Any (square real) symmetric matrix is orthogonally similar to a diagonal matrix.*

□

As an application of the spectral theorem, we single out a particularly important special class of self-adjoint operators.

Proposition 4.1.5 *The following conditions on an operator* T *on a finite- dimensional inner product space* V *are equivalent :*
(i) T *is self-adjoint and all its eigenvalues are non-negative.*
(ii) There exists a self-adjoint operator S *on* V *such that* $T = S^2$.
(iii) There exists a linear transformation $S : V \rightarrow W$ *, with* W *some inner product space, such that* $T = S^*S$.
(iv) T *is self-adjoint and* $< Tx, x > \ \geq 0$ *for all* x *in* V.

An operator T *satisfying the above conditions is said to be* **positive** *and we write* $T \geq 0$.

Proof : $(i) \Rightarrow (ii)$: The spectral theorem says that there exists an orthonormal basis $B = \{e_1, \ldots, e_n\}$ and scalars $\alpha_1, \ldots, \alpha_n$ such that $Te_j = \alpha_j e_j$ for all j. Condition (i) ensures that $\alpha_j \geq 0$ for all j. Define S to be the unique operator for which $Se_j = \sqrt{\alpha_j}\, e_j$ for $1 \leq j \leq n$. The operator S is self-adjoint since $[S]_B$ is a (diagonal matrix and hence) symmetric matrix, and clearly $S^2 = T$.

$(ii) \Rightarrow (iii)$: Obvious.

$(iii) \Rightarrow (iv)$: If $T = S^*S$, then, $T^* = S^*(S^*)^* = S^*S = T$ and if $x \in V$, then $< Tx, x > \ = \ < S^*Sx, x > \ = ||Sx||^2$.

$(iv) \Rightarrow (i)$: $Tx = \lambda x$ and $x \neq 0 \Rightarrow \lambda ||x||^2 = < Tx, x > \geq 0 \Rightarrow \lambda \geq 0$. $\Box$

Note that the 'square root' S of a positive operator T, whose existence was proved above, was constructed so that in fact $S \geq 0$. It is a fact that *the positive square root of a positive operator is unique.*

EXERCISE 4.1.5 *(a) Show that every self-adjoint operator T admits a unique decomposition $T = T_+ - T_-$ such that (i) $T_\pm \geq 0$, and (ii) $T_+ T_- = 0$.*

(b) What is this decomposition for the (operator on $I\!R^3$ represented, with respect to standard basis, by the) matrix given by

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

4.2 Complex scalars

What was not so clear from the last section was that while a diagonal entry of a diagonal matrix is clearly an eigenvalue of the matrix, actually every eigenvalue of a self-adjoint operator must appear as a diagonal entry in any diagonal matrix which represents the operator in some orthonormal basis. In fact, it is not even clear why there can be only finitely many eigenvalues of a self-adjoint operator.

Secondly, although our proof of the spectral theorem made strong use of self-adjointness, it is natural to ask if one can, by exhibiting sufficient cleverness, come up with a proof which does not use self-adjointness; in other words, can it be that the spectral theorem is actually true for all operators ?

The answer to these questions can be found in a closer analysis of the notion of an eigenvalue. On the one hand, a real number λ is an eigenvalue of an operator $T \in \mathcal{L}(V)$ precisely when $(T - \lambda I_V)$ has a non- zero kernel, which is equivalent to the non-invertibility of $(T - \lambda I_V)$. On the other hand, an operator S fails to be invertible precisely when $det\ S = 0$, where of course $det\ S$ denotes the determinant of any matrix representing S with respect to any basis for V . This leads naturally to the following definition.

Definition 4.2.1 *Let $T \in \mathcal{L}(V)$, where V is any finite-dimensional vector space, and let B denote any basis for V . Let $[T]_B = ((t_{ij}))$. Then the function $p_T(\lambda)$ defined by*

$$p_T(\lambda) \;=\; det(T - \lambda I_V) \;=\; \begin{vmatrix} t_{11}-\lambda & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22}-\lambda & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn}-\lambda \end{vmatrix} \tag{4.2.1}$$

is called the **characteristic polynomial** *of the operator T (resp., the matrix $((t_{ij}))$).*

It must be observed at the outset that :
(i) the function $p_T(\lambda)$ is indeed a polynomial of degree n ($=$ $dim\ V$), with leading coefficient (= coefficient of λ^n) equal to $(-1)^n$; and
(ii) the value of the characteristic polynomial is independent of the basis chosen to represent the operator as a matrix.

Having said all that, the proof of the following statement must be clear.

Proposition 4.2.2 *Let $T \in \mathcal{L}(V)$ and let $\lambda_0 \in I\!R$. The following conditions are equivalent :*
(i) λ_0 is an eigenvalue of T.
(ii) $p_T(\lambda_0) = 0$.

□

Now we are faced with the following situation. A real polynomial may not have any real roots. Thus, for instance, if T is the operator on $I\!R^2$ given by the matrix

$$T \;=\; \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad , \tag{4.2.2}$$

we see that

$$p_T(\lambda) \;=\; \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} \;=\; \lambda^2 + 1 \quad .$$

Thus this operator has no real eigenvalues - which must have been evident on geometric grounds. (Why?) In particular, the conclusion of Lemma 4.1.3 (and consequently of Theorem 4.1.1) is false for this non-self-adjoint operator T.

One way out of this impasse is to 'allow complex scalars'. The way to make this precise is to note that the definition of a vector space which we have given - Definition 1.2.1 - makes perfect sense if we replace every occurrence of the word 'real' (resp., the set $\mathbb{R}$) by the word 'complex' (resp., the set $\mathbb{C}$). As in the real case, one can show that every n-dimensional complex vector space is isomorphic to the vector space

$$\mathbb{C}^n = \{\mathbf{z} = (z_1, \ldots, z_n) : z_1, \ldots, z_n \in \mathbb{C}\}$$

(where the vector operations are co-ordinatewise).

In fact, all the statements in the first two chapters continue to remain valid for complex (rather than just real) vector spaces, and the proof of their validity is exactly as in the real case.

It is only when we come to inner products that there is a minor variation. This is because the natural norm on $\mathbb{C}^n$ is given by $||\mathbf{z}||^2 = \sum_{i=1}^n |z_i|^2$ where $|\lambda|$ is the absolute value of the complex number λ which satisfies $|\lambda|^2 = \lambda\bar{\lambda}$, where $\bar{\lambda}$ is the complex conjugate of λ. (Recall that if $\lambda = a+ib$ where $a, b \in \mathbb{R}$, then $\bar{\lambda} = a-ib$.) Consequently, the inner (or dot-)product on $\mathbb{C}^n$ is defined by

$$\mathbf{z} \cdot \mathbf{w} = \sum_{i=1}^n z_i \bar{w}_i \quad .$$

Hence, an inner product on a complex vector space V is defined to be a mapping $< \cdot, \cdot > : V \times V \rightarrow \mathbb{C}$ satisfying, for all $u, v, w \in V$ and $\alpha \in \mathbb{C}$:

(i) $< u+v, w > = < u, w > + < v, w >$;

(ii) $< \alpha u, v > = \alpha < u, v >$:

(iii) $< u, v > = \overline{< v, u >}$;

(iv) $< u, u > \geq 0$, and $< u, u > = 0 \Leftrightarrow u = 0$.

With this minor modification, all the results of Chapter 3 also go through in the complex case with only slight changes; for instance, we have :

(i) If $\{e_1, \ldots, e_n\}$ is an orthonormal basis for a complex inner product space, then

$$< x, y > = \sum_{i=1}^n < x, e_i >< e_i, y > \quad \forall\, x, y \;;$$

(ii) the mapping $y \to \phi_y$ (cf. Exercise 3.2.1) defines, not an isomorphism, but an **anti-isomorphism** - i.e., a bijective mapping Φ such that $\Phi(\alpha y + z) = \bar{\alpha}\Phi(y) + \Phi(z)$ - of V onto V^*; and

(iii) If $T \in \mathcal{L}(V, W)$ is a linear transformation between finite-dimensional complex inner product spaces, and if B_V, B_W are orthonormal bases for these spaces, then the matrix $[T^*]_{B_W}^{B_V}$ is not simply the transpose, but rather the **conjugate-transpose** - i.e., the entry-wise complex conjugate of the transpose - of the matrix $[T]_{B_V}^{B_W}$.

The reason for having to go to complex scalars lies in Proposition 4.2.2 and the following fundamental fact about the set of complex numbers (which is sometimes expressed by the statement that *the set of complex numbers is algebraicially closed*).

Theorem 4.2.3 (The Fundamental Theorem of Algebra) *If*

$$p(z) = \alpha_n z^n + \alpha_{n-1} z^{n-1} + \cdots + \alpha_0$$

is a complex polynomial of degree n *- i.e., each* α_i *is a complex number and the leading coefficient* α_n *is non-zero - then there exist (not necessarily distinct) complex numbers* $\lambda_1, \ldots, \lambda_n$ *such that*

$$p(z) = \alpha_n (z - \lambda_1) \cdots (z - \lambda_n)$$

for all z *in* $\mathbb{C}$.

□

We will not prove this theorem here, but accept its validity and proceed. The first bonus of working with complex scalars is the following analogue of Lemma 4.1.3, which is valid for all (not necessarily self-adjoint) operators on complex inner product spaces.

Lemma 4.2.4 *Every operator on a finite-dimensional complex inner product space has an eigenvector.*

Proof : First notice that Proposition 4.2.2 is valid with $\mathbb{R}$ replaced by $\mathbb{C}$. According to the Fundamental Theorem of Algebra, there exists a complex number λ such that $p_T(\lambda) = 0$. Then, the last sentence shows that λ is indeed an eigenvalue of T, and where there is an eigenvalue, there must be an eigenvector. □

The natural consequence that this lemma has for a general (non-self-adjoint) operator is not diagonalisability but **triangulisability**.

Theorem 4.2.5 *If T is an operator on a finite-dimensional complex inner product space V, there exists an orthonormal basis B for V which 'triangulises' T in the sense that $[T]_B$ is* **upper triangular** *- i.e., if $[T]_B = ((t_{ij}))$, then $t_{ij} = 0$ whenever $i > j$.*

Equivalently, there exists a chain

$$\{0\} = V_0 \subset V_1 \subset \cdots \subset V_{n-1} \subset V_n = V \tag{4.2.3}$$

of subspaces of V such that $dim\ V_i = i,\ 0 \leq i \leq n$ and such that each V_i is an invariant subspace for T. (A chain of subspaces as in equation 4.2.3 is also called a **flag** *of subspaces.)*

Proof : The two formulations are indeed equivalent. For if T has an upper triangular matrix with respect to an orthonormal basis $\{e_1, \ldots, e_n\}$ for V, and if we define $V_i = \bigvee\{e_k : 1 \leq k \leq i\}, 1 \leq i \leq n$, then the V_i's are easily seen to give a flag of invariant subspaces of T; and conversely, given a flag $\{V_i\}$ of invariant subspaces of T, simply choose an orthonormal basis $\{e_1, \ldots, e_n\}$ for V such that $V_i = \bigvee\{e_k : 1 \leq k \leq i\}$ for $1 \leq i \leq n$ and observe that any such basis will triangulise T. (Verify !)

As in the case of Theorem 4.1.1, the proof is by induction on $dim\ V$. The theorem is trivial when $dim\ V = 1$. So assume that $dim\ V = n$ and that the theorem is known to be valid for all spaces of dimension less than n.

Apply Lemma 4.2.4 to the operator T^* to find a unit vector x_n and a scalar λ_n such that $T^*x_n = \bar{\lambda}_n x_n$. Since the one-dimensional space spanned by x_n is clearly invariant under T^*, it follows from Lemma 4.1.2 that the subspace $V_{n-1} = \{x_n\}^\perp$ is invariant under T. Apply the induction hypothesis to the operator T_0 which is the restriction to V_{n-1} of T to find a flag $\{0\} = V_0 \subset V_1 \subset \cdots \subset V_{n-1}$ of invariant subspaces for T_0; note that then each V_i is necessarily also invariant under T and deduce that if we put $V_n = V$, then $\{V_i : 0 \leq i \leq n\}$ is indeed a flag of invariant subspaces for T. □

Before concluding this section, we pause to make some natural observations about complex matrices.

EXERCISE 4.2.1 *(a) Show that the set* $M_n(\mathbb{C})$ *has a natural structure of a 'unital algebra over the field of complex numbers'; show further that the operation defined in* $M_n(\mathbb{C})$ *by* $T \mapsto T^*$ *(where the* (i,j)*-th entry of* T^* *is, by definition, equal to the complex conjugate of the* (j,i)*-th entry of the matrix* T *) satisfies the following conditions :*
(i) $(S+T)^* = S^* + T^*$ *;*
(ii) $(\alpha T)^* = \bar{\alpha}\, T^*$ *;*
(iii) $(T^*)^* = T$;
(iv) $(ST)^* = T^* S^*$.

(b) Show that the set $Gl(n,\mathbb{C}) = \{T \in M_n(\mathbb{C}) : det\ T \neq 0\}$ *is a group with respect to the multiplication in* $M_n(\mathbb{C})$.

(c) Show that the set $U(n,\mathbb{C}) = \{T \in M_n(\mathbb{C}) : T^*T = I_n\}$ *is a subgroup of* $Gl(n,\mathbb{C})$ *, the elements of which are referred to as unitary matrices.*

(d) Show that if $T \in M_n(\mathbb{C})$ *, then there exists* $U \in U(n,\mathbb{C})$ *such that* UTU^* *is an upper triangular matrix.*

EXERCISE 4.2.2 *Go through every statement in these notes until §4.1, consider the statement obtained by replacing each occurrence of* $\mathbb{R}$ *with* $\mathbb{C}$ *, and either (a) prove it, or (b) prove a modified version of it, or (c) show that there is no reasonable statement of it which can be valid. How many of these statements fall into category (c) ?*

4.3 The spectral theorem for normal operators

We begin this section with an elementary, nevertheless very useful fact.

Proposition 4.3.1 *(a)* **(Polarisation identity)** *If* x, y *are any two vectors in a complex inner product space* V *and if* $T \in \mathcal{L}(V)$ *, then*

$$\begin{aligned} 4 < Tx, y > \quad &= \quad < T(x+y), (x+y) > \; - \; < T(x-y), (x-y) > \\ & \quad + i \; < T(x+iy), (x+iy) > \; - i \; < T(x-iy), (x-iy) > \; . \end{aligned}$$

(b) The following conditions on an operator $T \in \mathcal{L}(V)$ *are equivalent :*
(i) $T = 0$.
(ii) $< Tx, x > = 0$ *for all* x.
(iii) $< Tx, y > = 0$ *for all* x, y.

Proof : The proof of the polarisation identity - is a brutal computation. Thus :

$$< T(x+y),(x+y) > \ = \ < Tx,x > + < Ty,y > \ + \ < Tx,y > + < Ty,x > \ . \quad (4.3.4)$$

Replace y by $-y$ in equation 4.3.4 to obtain :

$$< T(x-y),(x-y) > \ = \ (\ < Tx,x > + < Ty,y > \ ) - (\ < Tx,y > + < Ty,x > \ ) \ . \quad (4.3.5)$$

Take the difference of the two preceding equations to get :

$$< T(x+y), x+y > \ - \ < T(x-y), x-y > \ = \ 2 \ (\ < Tx,y > + < Ty,x > \ ) \ . \quad (4.3.6)$$

Replace y by iy in equation 4.3.6 to find that :

$$< T(x+iy), x+iy > \ - \ < T(x-iy), x-iy > \ = \ 2i \ (- < Tx,y > + < Ty,x > \ ) \ . \quad (4.3.7)$$

Multiply equation 4.3.7 by i , add the result to equation 4.3.5, and the polarisation identity results.

As for (b), the implication $(i) \Rightarrow (ii)$ is trivial, while the implication $(ii) \Rightarrow (iii)$ is an immediate consequence of the Polarisation identity. As for the implication $(iii) \Rightarrow (i)$, if $< Tx,y > \ = \ 0$ for all vectors x and y , then in particular, setting $y = Tx$, we see that $Tx \ = \ 0$ for all x . □

The operator T of equation 4.2.2 is an operator on a real inner product space which satisfies (ii) but not (i) or (iii) of Proposition 4.3.1 (b), and hence the requirement that we are 'working over complex numbers' is essential for the validity of this proposition.

Corollary 4.3.2 *Let* T *be an operator on a finite-dimensional complex inner product space. The following conditions on* T *are equivalent :*
(i) T *is self-adjoint.*
(ii) $< Tx,x > \ \in \ I\!R$ *for all* x *in* V*.*

Proof : $(i) \Rightarrow (ii)$:

$$< Tx,x > \ = \ < x, T^*x > \ = \ < x, Tx > \ = \ \overline{< Tx,x >}$$

and a complex number is real precisely when it equals its complex conjugate.

$(ii) \Rightarrow (i)$: It follows essentially from the preceding computation that the assumption $< Tx, x > \ \in \mathbb{R}$ implies that $< (T - T^*)x, x > \ = 0$. The validity of this equation for all x implies, in view of Proposition 4.3.1 (b), that $T = T^*$. □

Now we address ourselves to the most general form of the spectral theorem one can hope for, in the context of complex inner product spaces. The clue is furnished by the simple observation that (while multiplication in $M_n(\mathbb{C})$ is not, in general, commutative) any two diagonal matrices commute - i.e., their product is independent of the order in which they are multiplied - and in particular, any diagonal matrix commutes with its adjoint. This leads us to the following crucial definition.

Definition 4.3.3 *An operator (resp., a square matrix)* T *is said to be* **normal** *if it commutes with its adjoint - i.e.,* $TT^* = T^*T$.

EXERCISE 4.3.1 *(a) A necessary condition for an operator on a finite-dimensional inner product space to be diagonalisable is that it is normal.*
(b) The operator N *of Exercise 4.1.1 (d) is not normal and is consequently not diagonalisable.*
(c) Argue alternatively that the only eigenvalue of N *is zero, and as* $N \neq 0$ *, it follows that* N *cannot be diagonalisable.*

We shall find the following simple facts concerning normal operators useful.

Proposition 4.3.4 *(a) An operator* T *on a finite-dimensional complex inner product space is normal if and only if*

$$||Tx|| \ = \ ||T^*x|| \ \text{ for all } x \ .$$

(b) If T *is normal, then*

$$ker\ T \ = \ ker\ T^* \ .$$

(c) If T *is normal, then* $ran\ T$ *and* $ker\ T$ *are orthogonal complements of one another.*

Proof : (a) In view of Proposition 4.3.1, the operator T is normal if and only if $< (T^*T - TT^*)x, x > \ = 0$ for all x.

(b) This is an immediate consequence of (a) above.

(c) This is an immediate consequence of (b) and Proposition 3.3.4. $\square$

We state an immediate corollary of the above proposition as an exercise.

Exercise 4.3.2 *(a) Show that if T is a normal operator on a finite-dimensional complex inner product space V and if $\lambda \in \mathbb{C}$, then also $(T - \lambda I_V)$ is normal.*

(b) If x is an eigenvector of a normal operator corresponding to an eigenvalue λ, then show that x is an eigenvector of T^* corresponding to the eigenvalue $\bar{\lambda}$.

(c) Show, by example, that the assertion (b) above is false if the assumption of normality is dropped.

Now we are ready for the spectral theorem.

Theorem 4.3.5 (The Spectral theorem) *The following conditions on an operator T on a finite-dimensional complex inner product space are equivalent :*
(i) T is diagonalisable -i.e., there exists an orthonormal basis $B = \{e_1, \ldots, e_n\}$ for V and complex numbers $\lambda_1, \ldots, \lambda_n$ such that $Te_j = \lambda_j e_j$ for all j.
(ii) T is normal.

Proof : $(i) \Rightarrow (ii)$: This has already been observed.

$(ii) \Rightarrow (i)$: We shall show that if an upper triangular matrix is normal, then it must be a diagonal matrix. (In view of Theorem 4.2.5, this will prove the theorem.)

So suppose $T = ((t_{ij}))$ is a normal matrix such that $t_{ij} = 0$ whenever $i > j$. This says, in particular, that if $\{e_1, \ldots, e_n\}$ denotes the standard basis for $\mathbb{C}^n$, then $Te_1 = t_{11}e_1$. It follows from Exercise 4.3.2 (b) that $T^*e_1 = \overline{t_{11}}e_1$. On the other hand, we see, by the form of T^*, that we must have $T^*e_1 = \sum_{j=1}^n \overline{t_{1j}} e_j$. The linear dependence of the e_i's now forces $t_{12} = \cdots = t_{1n} = 0$.

This shows then that $Te_2 = t_{22}e_2$ and arguing exactly as above, we see that also $t_{2j} = 0$ whenever $j \neq 2$. Proceeding thus, we see that indeed we must have $t_{ij} = 0$ whenever $j \neq i$ for every i and the proof is complete. $\square$

As in the real case, note that an equivalent (matricial) formulation of the spectral theorem is the following :

If an $n \times n$ matrix T is normal, then there exists a unitary matrix $U \in M_n(\mathbb{C})$ such that UTU^* is a diagonal matrix.

EXERCISE 4.3.3 *(a) Consider the 3×3 matrix*

$$P_3 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} .$$

Show that P_3 is a unitary matrix and explicitly exhibit a unitary matrix U such that UP_3U^* is diagonal.

(b) Can you generalise (a) from 3 to a general n ?

Before reaping some of the consequences of the spectral theorem, it will be useful to introduce some terminology.

Definition 4.3.6 *Let T denote an operator on a finite-dimensional complex vector space or an $n \times n$ matrix. The* **spectrum of T** *is the set, denoted by $\sigma(T)$, defined by*

$$\sigma(T) = \{\lambda \in \mathbb{C} : \lambda \text{ is an eigenvalue of } T\}.$$

The following corollary lists some important special classes of normal operators.

Corollary 4.3.7 *Let T be an operator on a finite-dimensional complex inner product space.*

(a) The following conditions on T are equivalent:

(i) T is self-adjoint;

(ii) T is normal and $\sigma(T) \subset \mathbb{R}$;

(iii) $< Tx, x > \in \mathbb{R}$ for all x in V.

(b) The following conditions on T are equivalent:

(i) T is unitary - i.e., $T^*T = TT^* = I$;

(ii) T is normal and $\sigma(T) \subset \mathbf{T} = \{z \in \mathbb{C} : |z| = 1\}$.

(c) The following conditions on T are equivalent:

(i) T admits a factorisation $T = S^*S$;

(ii) T *admits a self-adjoint square root - i.e., there exists a self-adjoint* S *such that* $S^2 = T$;
(iii) T *is normal and* $\sigma(T) \subset [0, \infty)$;
(iv) $< Tx, x > \in [0, \infty)$ *for all* x *in* V.

Proof : (a) $(i) \Rightarrow (ii)$: A self-adjoint operator is normal since it clearly commutes with itself. The fact that every eigenvalue of a self-adjoint matrix is real follows at once from Exercise 4.3.2 (b).

$(ii) \Rightarrow (iii)$: The spectral theorem guarantees the existence of an orthonormal basis $\{e_1, \ldots, e_n\}$ and complex numbers $\lambda_1, \ldots, \lambda_n$ such that $Te_j = \lambda_j e_j$ for all j. The assumption (ii) says that each λ_i is non-negative; hence

$$\begin{aligned} < Tx, x > &= < T\,(\sum_{i=1}^{n} < x, e_i > e_i\,),\ (\sum_{j=1}^{n} < x, e_j > e_j\,)\ > \\ &= < \sum_{i=1}^{n} < x, e_i > \ \lambda_i e_i\ ,\ \sum_{j=1}^{n} < x, e_j > e_j \ > \\ &= \sum_{i=1}^{n} | < x, e_i > |^{\,2}\ \lambda_i \qquad (4.3.8) \\ &\geq 0\ . \end{aligned}$$

$(iii) \Rightarrow (i)$: This has already been proved in Corollary 4.3.2.

(b) $(i) \Rightarrow (ii)$: It is obvious that unitarity implies normality. If U is unitary and $Ux = \lambda x$, it follows from Exercise 4.3.2 (b) that also $U^*x = \bar{\lambda} x$ so that the assumed unitarity says that $x = U^*Ux = |\lambda|^{\,2} x$ and hence $\lambda \in \mathbf{T}$.

$(ii) \Rightarrow (i)$: If U is a normal operator, appeal to the spectral theorem to find an orthonormal basis $\{e_1, \ldots, e_n\}$ and complex numbers $\lambda_1, \ldots, \lambda_n$ - which have modulus 1 according to the condition (ii) - such that $Ue_i = \lambda e_i$ for all i. Deduce as before that also $U^*x = \bar{\lambda} x$ so that we find that, for each basis vector e_i , we have $U^*Ue_i = UU^*e_i = e_i$. This implies that U must be unitary.

(c) This is proved exactly like Proposition 4.1.5, the only additional point to be made being that, in view of (a) above, the assumption that $< Tx, x > \ \geq 0\ \forall\, x$ already implies the self-adjointness of T. (Note that this part of the argument relies heavily on the fact that we are working in a complex inner product space.) □

We pause to mention a fact that is a consequence of equation 4.3.8; the proof is left as an exercise to the reader. (Recall that a subset Σ of a (real or complex) vector space is said to be convex if it contains the line segment joining any two of its points - equivalently, $v_0, v_1 \in \Sigma \Rightarrow (1-t)v_0 + tv_1 \in \Sigma \, \forall \, t \in [0,1]$.)

Corollary 4.3.8 *If T is a normal operator on a finite-dimensional complex inner product space, the so-called* **numerical range of T** *which is defined to be the set*

$$W(T) = \{ < Tx, x > \; : \; x \in V \;\; and \;\; ||x|| = 1\} \tag{4.3.9}$$

coincides with the **convex hull** *of the spectrum $\sigma(T)$ of T which is, by definition, the smallest convex set containing $\sigma(T)$.*

Exercise 4.3.4 *(a) Show that any operator T on a complex inner product space (resp., any square matrix T) is uniquely expressible as $T = T_1 + iT_2$ where each T_i is self-adjoint. (In analogy with the one-dimensional case, this is referred to as the* **Cartesian decomposition** *of T and we define $Re\, T = T_1$, $Im\, T = T_2$ and call these the real and imaginary parts of T.)*

(b) Show that an operator (resp., square matrix) T is normal if and only if its real and imaginary parts commute.

The preceding exercise, as well as Exercise 4.1.5 and Corollary 4.3.7 should convey to the reader that there is a parallel between (real and complex) numbers and operators. This is a useful parallel to bear in mind; it sometimes suggests results that one might hope to prove about operators. Thus, complex numbers admit Cartesian decompositions, as do operators. Complex numbers also possess a 'polar decomposition'.

Exercise 4.3.5 *(a) State and prove a (left as well as a right) version of the polar decomposition for operators on finite-dimensional complex inner product spaces.*

(b) Prove the following alternative form of the spectral theorem : if T is a normal operator on a finite-dimensional inner product spave V , and if $\sigma(T) = \{\lambda_1, \ldots, \lambda_k\}$, then $T = \sum_{i=1}^{k} \lambda_i P_i$, where P_i denotes the orthogonal projection onto $V_i = ker\ (T - \lambda_i I_V$, for $1 \leq i \leq k$. (The subspace W_i is also called the **eigenspace of** *T corresponding to the eigenvalue T.)*

Suggested Reading

Linear algebra is one of the most basic tools of all mathematics. Consequently, it should come as no surprise that there is a sea of literature on the subject. Rather than giving an interminably long list of possible texts on the subject, we will content ourselves with four references, each of which presents the subject from a different point of view, with a different goal in mind.

(1) P.R. Halmos, *Finite-dimensional Vector Spaces*, Van Nostrand, Princeton, New Jersey 1958.

This book has a flavour that is probably somewhat similar to these notes, in the sense that it stresses the 'geometric' or 'operator-theoretic' point of view. It is written with a view to the possible extensions of the theory to its infinite-dimensional context of 'operator theory on Hilbert space'.

(2) M.W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra*, Academic Press, Florida, 1974.

This book is written with the goal of understanding the rudiments of dynamical systems. It is a beautiful exposition of the interplay between linear algebra and the theory of solutions of differential equations; the reader will do well to browse through this book, if at least to realise the power and applicability of the spectral theorem.

(3) K. Hoffman and R. Kunze, *Linear Algebra*, 2nd. ed. Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

This book stresses the matricial point of view. It will equip the reader well with the skills to actually sit down and tackle problems such as: how does one actually compute the rank of a, say, 5×5 matrix, or invert it if it turns out to be invertible? This point of view is a necessary complement to the geometric point of view espoused by Halmos or these notes.

(4) G. Strang, *Linear Algebra and its Applications*, Academic Press, New York, 1976.

This book, the most recent of the four volumes discussed here, reflects its modernity in its point of view. It is concerned with serious problems of computation with matrices, such as : what is a workable algorithm for tackling a specific computation using matrices? which of these methods are fast from the point of view of machine time or number of computations involved? what sort of control does one have in rounding-off errors in each of the available algorithms? and so on. Besides the computational point of view that is evident here, the reader will find plenty of useful 'pure mathematics' here.