1 VECTOR SPACES AND SUBSPACES

What is a vector? Many are familiar with the concept of a vector as:

- Something which has magnitude and direction.

- an ordered pair or triple.

- a description for quantities such as Force, velocity and acceleration.

Such vectors belong to the foundation vector space - $\mathbb{R}^n$ - of all vector spaces. The properties of general vector spaces are based on the properties of $\mathbb{R}^n$. It is therefore helpful to consider briefly the nature of $\mathbb{R}^n$.

1.1 The Vector Space $\mathbb{R}^n$

Definitions

- If n is a positive integer, then an **ordered n-tuple** is a sequence of n real numbers $(a_1, a_2, \ldots, a_n)$. The set of all ordered n-tuples is called **n-space** and is denoted by $\mathbb{R}^n$.

When $n = 1$ each ordered n-tuple consists of one real number, and so $\mathbb{R}$ may be viewed as the set of real numbers. Take $n = 2$ and one has the set of all 2-tuples which are more commonly known as **ordered pairs**. This set has the geometrical interpretation of describing all points and directed line segments in the Cartesian $x-y$ plane. The vector space $\mathbb{R}^3$, likewise is the set of **ordered triples**, which describe all points and directed line segments in 3-D space.

In the study of 3-space, the symbol (a_1, a_2, a_3) has two different geometric interpretations: it can be interpreted as a point, in which case a_1, a_2 and a_3 are the coordinates, or it can be interpreted as a vector, in which case a_1, a_2 and a_3 are the components. It follows, therefore, that an ordered n-tuple $(a_1, a_2, \ldots, a_n)$ can be

viewed as a "generalized point" or a "generalized vector" - the distinction is mathematically unimportant. Thus, we can describe the 5-tuple $(1, 2, 3, 4, 5)$ either as a point or a vector in $\mathbb{R}^5$.

Definitions

- Two vectors $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in $\mathbb{R}^n$ are called **equal** if

$$u_1 = v_1, u_2 = v_2, \ldots, u_n = v_n$$

- The **sum** $u + v$ is defined by

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$$

- Let k be any scalar, then the **scalar multiple** $k\mathbf{u}$ is defined by

$$k\mathbf{u} = (ku_1, ku_2, \ldots, ku_n)$$

- These two operations of addition and scalar multiplication are called the **standard operations** on $\mathbb{R}^n$.

- The **zero vector** in $\mathbb{R}^n$ is denoted by $\mathbf{0}$ and is defined to be the vector

$$\mathbf{0} = (0, 0, \ldots, 0)$$

- The **negative** (or **additive inverse**) of u is denoted by -u and is defined by

$$-\mathbf{u} = (-u_1, -u_2, \ldots, -u_n)$$

- The **difference** of vectors in $\mathbb{R}^n$ is defined by

$$\mathbf{v} - \mathbf{u} = \mathbf{v} + (-\mathbf{u})$$

The most important arithmetic properties of addition and scalar multiplication of vectors in $\mathbb{R}^n$ are listed in the following theorem. This theorem enabes us to manipulate vectors in $\mathbb{R}^n$ without expressing the vectors in terms of componenets.

Theorem 1.1. *If* $\mathbf{u} = (u_1, u_2, \ldots, u_n)$, $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, *and* $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ *are vectors in* $\mathbb{R}^n$ *and* k *and* l *are scalars, then:*

1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

2. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$

3. $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

4. $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$; *that is,* $\mathbf{u} - \mathbf{u} = \mathbf{0}$

5. $k(l\mathbf{u}) = (kl)\mathbf{u}$

6. $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$

7. $(k + l)\mathbf{u} = k\mathbf{u} + l\mathbf{u}$

8. $1\mathbf{u} = \mathbf{u}$

1.2   Generalized Vector Spaces

The time has now come to generalize the concept of a vector. In this section a set of axioms are stated, which if satisfied by a class of objects, entitles those objects to be called "vectors". The axioms were chosen by abstracting the most important properties (theorem 1.1). of vectors in $\mathbb{R}^n$; as a consequence, vectors in $\mathbb{R}^n$ automatically satisfy these axioms. Thus, the new concept of a vector, includes many new kinds of vector without excluding the "common vector". The new types of vectors include, among other things, various kinds of matrices and functions.

Definition

A *vector space* V over a field $\mathbb{F}$ is a nonempty set on which two operations are defined - addition and scalar multiplication. Addition is a rule for associating with each pair of objects $\mathbf{u}$ and $\mathbf{v}$ in V an object $\mathbf{u} + \mathbf{v}$, and scalar multiplication is a rule for associating with each scalar $k \in \mathbb{F}$ and each object $\mathbf{u}$ in V an object $k\mathbf{u}$ such that

1. If $\mathbf{u}, \mathbf{v} \in V$, then $\mathbf{u} + \mathbf{v} \in V$.

2. If $\mathbf{u} \in V$ and $k \in \mathbb{F}$, then $k\mathbf{u} \in V$.

3. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

4. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$

5. There is an object $\mathbf{0}$ in V, called a **zero vector** for V, such that $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$ for all $\mathbf{u}$ in V.

6. For each $\mathbf{u}$ in V, there is an object -$\mathbf{u}$ in V, called the **additive inverse** of $\mathbf{u}$, such that $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$;

7. $k(l\mathbf{u}) = (kl)\mathbf{u}$

8. $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$

9. $(k + l)\mathbf{u} = k\mathbf{u} + l\mathbf{u}$

10. $1\mathbf{u} = \mathbf{u}$

Remark The elements of the underlying field $\mathbb{F}$ are called scalars and the elements of the vector space are called vectors. Note also that we often restrict our attention to the case when $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Examples of Vector Spaces

A wide variety of vector spaces are possible under the above definition as illustrated by the following examples. In each example we specify a nonempty set of objects V. We must then define two operations - addition and scalar multiplication, and as an exercise we will demonstrate that all the axioms are satisfied, hence entitling V with the specified operations, to be called a vector space.

1. The set of all n-tuples with entries in the field $\mathbb{F}$, denoted $\mathbb{F}^n$ (especially note $\mathbb{R}^n$ and $\mathbb{C}^n$).

2. The set of all $m \times n$ matrices with entries from the field $\mathbb{F}$, denoted $M_{m \times n}(\mathbb{F})$.

3. The set of all real-valued functions defined on the real line $(-\infty, \infty)$.

4. The set of polynomials with coefficients from the field $\mathbb{F}$, denoted $P(\mathbb{F})$.

5. (Counter example) Let $V = \mathbb{R}^2$ and define addition and scalar multiplication oparations as follows: If $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$, then define

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2)$$

and if k is any real number, then define

$$k\mathbf{u} = (ku_1, 0).$$

1.2.1 Some Properties of Vectors

It is important to realise that the following results hold for all vector spaces. They provide a useful set of vector properties.

Theorem 1.2. *If $u, v, w \in V$ (a vector space) such that $u + w = v + w$, then $u = v$.*

Corollary 1.1. *The zero vector and the additive inverse vector (for each vector) are unique.*

Theorem 1.3. *Let V be a vector space over the field $\mathbb{F}$, $\mathbf{u} \in V$, and $k \in \mathbb{F}$. Then the following statement are true:*

(a) $0\mathbf{u} = \mathbf{0}$

(b) $k\mathbf{0} = \mathbf{0}$

(c) $(-k)\mathbf{u} = -(k\mathbf{u}) = k(-\mathbf{u})$

(d) If $k\mathbf{u} = \mathbf{0}$, then $k = 0$ or $\mathbf{u} = 0$.

1.2.2 Quiz

True or false?

(a) Every vector space contains a zero vector.

(b) A vector space may have more than one zero vector.

(c) In any vector space, $a\mathbf{u} = b\mathbf{u}$ implies $a = b$.

(d) In any vector space, $a\mathbf{u} = a\mathbf{v}$ implies $\mathbf{u} = \mathbf{v}$.

1.3 Subspaces

It is possible for one vector space to be contained within a larger vector space. This section will look closely at this important concept.

Definitions

- A subset W of a vector space V is called a **subspace** of V if W is itself a vector space under the addition and scalar multiplication defined on V.

In general, all ten vector space axioms must be verified to show that a set W with addition and scalar multiplication forms a vector space. However, if W is part of a larget set V that is already known to be a vector space, then certain axioms need not be verified for W because they are inherited from V. For example, there is no need to check that $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (axiom 3) for W because this holds for all vectors in V and consequently holds for all vectors in W. Likewise, axioms 4, 7, 8, 9 and 10 are inherited by W from V. Thus to show that W is a subspace of a vector space V (and hence that W is a vector space), only axioms 1, 2, 5 and 6 need to be verified. The following theorem reduces this list even further by showing that even axioms 5 and 6 can be dispensed with.

Theorem 1.4. *If W is a set of one or more vectors from a vector space V, then W is a subspace of V if and only if the following conditions hold.*

(a) If $\mathbf{u}$ and $\mathbf{v}$ are vectors in W, then $\mathbf{u} + \mathbf{v}$ is in W.

(b) If k is any scalar and $\mathbf{u}$ is any vector in W, then $k\mathbf{u}$ is in W.

Proof. If W is a subspace of V, then all the vector space axioms are satisfied; in particular, axioms 1 and 2 hold. These are precisely conditions (a) and (b).

Conversely, assume conditions (a) and (b) hold. Since these conditions are vector space axioms 1 and 2, it only remains to be shown that W satisfies the remaining eight axioms. Axioms 3, 4, 7, 8, 9 and 10 are automatically satisfied by the vectors in W since they are satisfied by all vectors in V. Therefore, to complete the proof, we need only verify that Axioms 5 and 6 are satisfied by vectors in W.

Let $\mathbf{u}$ be any vector in W. By condition (b), $k\mathbf{u}$ is in W for every scalar k. Setting $k = 0$, it follows from theorem 1.3 that $0\mathbf{u} = \mathbf{0}$ is in W, and setting $k = -1$, it follows that $(-1)\mathbf{u} = -\mathbf{u}$ is in W. $\qquad\square$

Remarks

- Note that a consequence of (b) is that $\mathbf{0}$ is an element of W.

- A set W of one or more vectors from a vector space V is said to be **closed under addition** if condition (a) in theorem 1.4 holds and **closed under scalar multiplication** if condition (b) holds. Thus, theorem 1.4 states that W is *a subspace of V if and only if W is closed under addition and closed under scalar multiplication.*

Examples of Subspaces

1. A plane through the origin of $\mathbb{R}^3$ forms a subspace of $\mathbb{R}^3$. This is evident geometrically as follows: Let W be any plane through the origin and let $\mathbf{u}$ and $\mathbf{v}$ be any vectors in W other than the zero vector. Then $\mathbf{u} + \mathbf{v}$ must lie in W because it is the diagonal of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$, and $k\mathbf{u}$ must lie in W for any scalar k because $k\mathbf{u}$ lies on a line through $\mathbf{u}$. Thus, W is closed under addition and scalar multiplication, so it is a subspace of $\mathbb{R}^3$.

2. A line through the origin of $\mathbb{R}^3$ is also a subspace of $\mathbb{R}^3$. It is evident geometrically that the sum of two vectors on this line also lies on the line and that a scalar multiple of a vector on the line is on the line as well. Thus, W is closed under addition and scalar multiplication, so it is a subspace of $\mathbb{R}^3$.

3. Let n be a positive integer, and let W consist of all functions expressible in the form

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

where $a_0, \ldots, a_n$ belong to some field $\mathbb{F}$. Thus, W consists of the zero function together with all polynomials in $\mathbb{F}$ of degree n or less. The set W is a subspace of $P(\mathbb{F})$ (example 4 on page 5), and if $\mathbb{F} = \mathbb{R}$ it is also a subspace of the vector space of all real-valued functions (discussed in example 3 on page 5).

To see this, let $\mathbf{p}$ and $\mathbf{q}$ be the polynomials

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

and

$$q(x) = b_0 + b_1 x + \cdots + b_n x^n$$

Then

$$(\mathbf{p} + \mathbf{q})(x) = p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n$$

and

$$(k\mathbf{p})(x) = kp(x) = (ka_0) + (ka_1)x + \cdots + (ka_n)x^n$$

These functions have the form given above, so $\mathbf{p} + \mathbf{q}$ and $k\mathbf{p}$ lie in W. This vector space W is denoted $P_n(\mathbb{F})$.

4. The *transpose* A^T of an $m \times n$ matrix A is the $n \times m$ matrix obtained from A by interchanging rows and columns. A *symmetric matrix* is a square matrix A such that $A^T = A$. The set of all symmetric matrices in $M_{n \times n}(\mathbb{F})$ is a subspace of $M_{n \times n}(\mathbb{F})$.

5. The trace of an $n \times n$ matrix A, denoted $\mathrm{tr}(A)$, is the sum of the diagonal entries of A. The set of $n \times n$ matrices having trace equal to zero is a subspace of $M_{n \times n}(\mathbb{F})$.

1.3.1 Operations on Vector Spaces

Definitions

- The addition of two subsets U and V of a vector space is defined by:

$$U + V = \{\mathbf{u} + \mathbf{v} | \mathbf{u} \in U, \mathbf{v} \in V\}$$

- The intersection $\cap$ of two subsets U and V of a vector space is defined by:

$$U \cap V = \{\mathbf{w} | \mathbf{w} \in U \ and \ \mathbf{w} \in V\}$$

- A vector space W is called the direct sum of U and V, denoted $U \oplus V$, if U and V are subspaces of W with $U \cap V = \{0\}$ and $U + V = W$.

The following theorem shows how we can form a new subspace from other ones.

Theorem 1.5. *Any intersection or sum of subspaces of a vector space V is also a subspace of V.*

1.3.2 Quiz

True or false?

(a) If V is a vector space and W is a subset of V that is also a vector space, then W is a subspace of V.

(b) The empty set is a subspace of every vector space.

(c) If V is a vector space other than the zero vector space, then V contains a subspace W such that $W \neq V$.

(d) The intersection of any two subsets of V is a subspace of V.

(e) Any union of subspaces of a vector space V is a subspace of V.

1.4 Linear Combinations of Vectors and Systems of Linear Equations

Have m linear equations in n variables:

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

Write in matrix form: $A\boldsymbol{x} = \boldsymbol{b}$.

$A = [a_{ij}]$ is the $m \times n$ coefficient matrix.

$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is the column vector of unknowns, and $\boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ is the column vector of RHS.

Note: $a_{ij}, b_j \in \mathbb{R}$ or $\mathbb{C}$.

1.4.1 Gaussian Elimination

To solve $A\boldsymbol{x} = \boldsymbol{b}$:

write *augmented matrix*: $[A|\boldsymbol{b}]$.

1. Find the left-most non-zero column, say column j.

2. Interchange top row with another row if necessary, so top element of column j is non-zero. (The **pivot**.)

3. Subtract multiples of row 1 from all other rows so all entries in column j below the top are then 0.

4. Cover top row; repeat **1** above on rest of rows.

Continue until all rows are covered, or until only $00\ldots0$ rows remain.

Result is a triangular system, easily solved by *back substitution*: solve the last equation first, then 2nd last equation and so on.

1.4.2 Example

Use Gaussian elimination to solve:

$$
\begin{aligned}
x_3 - x_4 &= 2 \\
-9x_1 - 2x_2 + 6x_3 - 12x_4 &= -7 \\
3x_1 + x_2 - 2x_3 + 4x_4 &= 2 \\
2x_3 &= 6
\end{aligned}
$$

1.4.3 Definition (row echelon form)

A matrix is in *row echelon form* (r.e.f.) if each row after the first starts with *more* zeros than the previous row (or else rows at bottom of matrix are all zeros).

The Gauss algorithm converts any matrix to one in row echelon form. The 2 matrices are *equivalent*, that is, they have the same solution set.

1.4.4 Elementary row operations

1. $r_i \leftrightarrow r_j$: swap rows i and j.

2. $r_i \rightarrow r_i - cr_j$: replace row i with

(row i minus c times row j).

3. $r_i \rightarrow cr_i$:

replace row i with c times row i, where $c \neq 0$.

The Gauss algorithm uses only **1** and **2**.

1.4.5 Possible solutions for $A\boldsymbol{x} = \boldsymbol{b}$

Consider the r.e.f. of $[A|\boldsymbol{b}]$. Then we have three possibilities:

(1) *Exactly one* solution; here the r.e.f. gives each variable a single value, so the number of variables, n, equals the number of non-zero rows in the r.e.f.

(2) *No* solution; when one row of r.e.f. is $(0\,0\,\ldots\,d)$ with $d \neq 0$. We can't solve $0x_1 + 0x_2 + \cdots + 0x_m = d$ if $d \neq 0$; it says $0 = d$. In this case the system is said to be *inconsistent*.

(3) *Infinitely many* solutions; here the number of rows of the r.e.f. is *less* than the number of variables.

Note that a *homogeneous* system has $\boldsymbol{b} = \boldsymbol{0}$, i.e., all zero RHS. Then we always have at least the trivial solution, $x_i = 0$, $1 \leq i \leq n$.

1.4.6 Examples

$$
\begin{aligned}
x_1 + x_2 - x_3 &= 0 \\
2x_1 - x_2 \phantom{{}+ 2x_3} &= 0 \\
4x_1 + x_2 - 2x_3 &= 0
\end{aligned}
$$

$$
\begin{aligned}
x_2 - 2x_3 + 4x_4 &= 2 \\
2x_2 - 3x_3 + 7x_4 &= 6 \\
x_3 - x_4 &= 2
\end{aligned}
$$

1.4.7 Different right hand sides

To solve $A\boldsymbol{x} = \boldsymbol{b}_j$, for $j = 1, \ldots, r$, for r different sets of right hand sides $\boldsymbol{b}_j$:

Form a *big* augmented matrix $[A|\boldsymbol{b}_1\boldsymbol{b}_2\ldots\boldsymbol{b}_r]$ and find its r.e.f. $[U|\boldsymbol{b}_1'\boldsymbol{b}_2'\ldots\boldsymbol{b}_r']$. So U will be a r.e.f. corresponding to A. Then solve each of the systems $U\boldsymbol{x} = \boldsymbol{b}_j'$, $j = 1, 2, \ldots, r$, by back substitution.

1.4.8 Special case: finding A^{-1} (if it exists)

If A is $n \times n$ and it has an inverse, then solving $A\boldsymbol{x} = \boldsymbol{e}_j$ (where $\boldsymbol{e}_j$ is the $n \times 1$ column with 1 in jth place and 0 elsewhere) gives jth column of A^{-1}.

So we find r.e.f. of $[A|\boldsymbol{e}_1\boldsymbol{e}_2 \ldots \boldsymbol{e}_n]$, i.e., determine the r.e.f. of $[A|I]$ where I is $n \times n$ identity matrix.

Once we have found the r.e.f. of $[A|I]$ to be $[U|*]$, we then use row operations to convert it to $[I|D]$, so $D = A^{-1}$.

If the last row of U is all zeros, A has no inverse.

Note that if A and I are square, $AC = I$ implies $CA = I$ and conversely.

If such a matrix C exists, it is unique. We write $C = A^{-1}$, and we say A is *non-singular* or *invertible*.

1.4.9 Example

Does $A = \begin{pmatrix} 1 & -1 & 4 \\ 1 & 0 & -2 \\ 2 & -2 & 10 \end{pmatrix}$ have an inverse?
If so, find it.

1.4.10 Linear combinations

Definitions

- A vector $\mathbf{w}$ is called a **linear combination** of the vectors $v_1, v_2, \ldots, v_r$ if it can be expressed in the form

$$\mathbf{w} = k_1 v_1 + k_2 v_2 + \cdots + k_r v_r$$

  where $k_1, k_2, \ldots, k_r$ are scalars.

Example

1. Consider the vectors $\mathbf{u} = (1, 2, -1)$ and $\mathbf{v} = (6, 4, 2)$ in $\mathbb{R}^3$. Show that $\mathbf{w} = (9, 2, 7)$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$ and that $\mathbf{w}' = (4, -1, 8)$ is not a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

1.4.11 Spanning

If $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ are vectors in a vector space V, then generally some vectors in V may be linear combinations of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ and others may not. The following theorem shows that if a set W is constructed consisting of all those vectors that are expressible as linear combinations of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$, then W forms a subspace of V.

Theorem 1.6. *If $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ are vectors in a vector space V, then:*

(a) *The set W of all linear combinations of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ is a subspace of V.*

(b) *W is the smallest subspace of V that contains $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ every other subspace of V that contains $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ must contain W*

Proof. (a) To show that W is a subspace of V, it must be proven that it is closed under addition and scalar multiplication. There is at least one vector in W, namely, $\mathbf{0}$, since $\mathbf{0} = 0\mathbf{v_1} + 0\mathbf{v_2} + \cdots + 0\mathbf{v_r}$. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in W, then

$$\mathbf{u} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_r\mathbf{v_r}$$

and

$$\mathbf{v} = k_1\mathbf{v_1} + k_2\mathbf{v_2} + \cdots + k_r\mathbf{v_r}$$

where $c_1, c_2, \ldots, c_r, k_1, k_2, \ldots, k_r$ are scalars. Therefore

$$\mathbf{u} + \mathbf{v} = (c_1 + k_1)\mathbf{v_1} + (c_2 + k_2)\mathbf{v_2} + \cdots + (c_r + k_r)\mathbf{v_r}$$

and, for any scalar k,

$$k\mathbf{u} = (kc_1)\mathbf{v_1} + (kc_2)\mathbf{v_2} + \cdots + (kc_r)\mathbf{v_r}$$

Thus, $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ are linear combinations of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ and consequently lie in W. Therefore, W is closed under addition and scalar multiplication.

(b) Each vector $\mathbf{v_i}$ is a linear combination of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ since we can write

$$\mathbf{v_i} = 0\mathbf{v_1} + 0\mathbf{v_2} + \cdots + 1\mathbf{v_i} + \cdots + 0\mathbf{v_r}$$

Therefore, the subspace W contains each of the vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. Let W' be any other subspace that contains $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. Since W' is closed under addition and scalar multiplication, it must contain all linear combinations of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$. Thus W' contains each vector of W.

$\square$

Definitions

- If $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}\}$ is a set of vectors in a vector space V, then the subspace W of V consisting of all linear combinations of the vectors in S is called the **space spanned** by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$, and it is said that the vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ **span** W. To indicate that W is the space spanned by the vectors in the set $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}\}$ the below notation is used.

$$W = span(S) \quad or \quad W = span\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}\}$$

Examples The polynomials $1, x, x^2, \ldots, x^n$ span the vector space P_n defined previously since each polynomial $\mathbf{p}$ in P_n can be written as

$$\mathbf{p} = a_0 + a_1 x + \cdots + a_n x^n$$

which is a linear combination of $1, x, x^2, \ldots, x^n$. This can be denoted by writing

$$P_n = span\{1, x, x^2, \ldots, x^n\}$$

Spanning sets are not unique. For example, any two noncolinear vectors that lie in the $x - y$ plane will span the $x - y$ plane. Also, any nonzero vector on a line will span the same line.

Theorem 1.7. *Let* $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}\}$ *and* $S' = \{\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_k}\}$ *be two sets of vectors in a vector space* V. *Then*

$$span(S) = span(S')$$

if and only if each vector in S *is a linear combination of those in* S' *and (conversely) each vector in* S' *is a linear combination of those in* S.

Proof. If each vector in S is a linear combination of those in S' then

$$span(S) \subseteq span(S')$$

and if each vector in S' is a linear combination of those in S then

$$span(S') \subseteq span(S)$$

and therefore

$$span(S) = span(S').$$

If

$$\mathbf{v_i} \neq a_1 \mathbf{w_1} + a_2 \mathbf{w_2} + \cdots + a_n \mathbf{w_n}$$

for all possible $a_1, a_2, \ldots, a_n$ then

$$\mathbf{v_i} \in span(S) \quad but \quad \mathbf{v_i} \notin span(S')$$

therefore

$$span(S) \neq span(S')$$

and vice versa. $\qquad\qquad\square$

1.4.12   Quiz

True or false?

(a) $\mathbf{0}$ is a linear combination of any non-empty set of vectors.

(b) If $S \subseteq V$ (vector space V), then $span(S)$ equals the intersection of all subspaces of V that contain S.

1.5 Linear Independence

In the previous section it was stated that a set of vectors S spans a given vector space V if every vector in V is expressible as a linear combination of the vectors in S. In general, it is possible that there may be more than one way to express a vector in V as a linear combination of vectors in a spanning set. This section will focus on the conditions under which each vector in V is expressible as a unique linear combination of the spanning vectors. Spanning sets with this property play a fundamental role in the study of vector spaces.

Definitions If $S = \{v_1, v_2, \ldots, v_r\}$ is a nonempty set of vectors, then the vector equation

$$k_1 \mathbf{v_1} + k_2 \mathbf{v_2} + \cdots + k_r \mathbf{v_r} = \mathbf{0}$$

has at least one solution, namely

$$k_1 = 0, k_2 = 0, \ldots, k_r = 0$$

If this is the only solution, then S is called a **linearly independent** set. If there are other solutions, then S is called a **linearly dependent** set.

Examples

1. If $\mathbf{v_1} = (2, -1, 0, 3), v_2 = (1, 2, 5, -1)$ and $v_3 = (7, -1, 5, 8)$, then the set of vectors $S = \{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ is linearly dependent, since $3\mathbf{v_1} + \mathbf{v_2} - \mathbf{v_3} = \mathbf{0}$.

2. The polynomials

$$\mathbf{p_1} = 1 - x, \ \mathbf{p_2} = 5 + 3x - 2x^2, \ \mathbf{p_3} = 1 + 3x - x^2$$

   form a linearly dependent set in P_2 since $3\mathbf{p_1} - \mathbf{p_2} + 2\mathbf{p_3} = \mathbf{0}$

3. Consider the vectors $\mathbf{i} = (1, 0, 0), \mathbf{j} = (0, 1, 0), \mathbf{k} = (0, 0, 1)$ in $\mathbb{R}^3$. In terms of components the vector equation

$$k_1 \mathbf{i} + k_2 \mathbf{j} + k_3 \mathbf{k} = \mathbf{0}$$

17

becomes

$$k_1(1,0,0) + k_2(0,1,0) + k_3(0,0,1) = (0,0,0)$$

or equivalently,

$$(k_1, k_2, k_3) = (0,0,0)$$

Thus the set $S = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is linearly independent. A similar argument can be used to extend S to a linear independent set in $\mathbb{R}^n$.

4. In $M_{2\times3}(\mathbb{R})$, the set

$$\left\{ \begin{pmatrix} 1 & -3 & 2 \\ -4 & 0 & 5 \end{pmatrix}, \begin{pmatrix} -3 & 7 & 4 \\ 6 & -2 & -7 \end{pmatrix}, \begin{pmatrix} -2 & 3 & 11 \\ -1 & -3 & 2 \end{pmatrix} \right\}$$

is linearly dependent since

$$5\begin{pmatrix} 1 & -3 & 2 \\ -4 & 0 & 5 \end{pmatrix} + 3\begin{pmatrix} -3 & 7 & 4 \\ 6 & -2 & -7 \end{pmatrix} - 2\begin{pmatrix} -2 & 3 & 11 \\ -1 & -3 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The following two theorems follow quite simply from the definition of linear independence and linear dependence.

Theorem 1.8. *A set S with two or more vectors is:*

(a) *Linearly dependent if and only if at least one of the vectors in S is expressible as a linear combination of the other vectors in S.*

(b) *Linearly independent if and only if no vector in S is expressible as a linear combination of the other vectors in S.*

Example

1. Recall that the vectors

$$\mathbf{v_1} = (2, -1, 0, 3), \ \mathbf{v_2} = (1, 2, 5, -1), \ \mathbf{v_3} = (7, -1, 5, 8)$$

were linear dependent because

$$3\mathbf{v_1} + \mathbf{v_2} - \mathbf{v_3} = \mathbf{0}.$$

It is obvious from the equation that

$$\mathbf{v_1} = \frac{-1}{3}\mathbf{v_2} + \frac{1}{3}\mathbf{v_3}, \quad \mathbf{v_2} = -3\mathbf{v_1} + 1\mathbf{v_3}, \quad \mathbf{v_3} = 3\mathbf{v_1} + 1\mathbf{v_2}$$

Theorem 1.9. *(a) A finite set of vectors that contains the zero vector is linearly dependent.*

(b) A set with exactly two vectors is linearly independent if and only if neither vector is a scalar multiple of the other.

2 BASIS AND DIMENSION

A line is thought of as 1-Dimensional, a plane 2-Dimensional, and surrounding space as 3-Dimensional. This section will attempt to make this intuitive notion of dimension precise and extend it to general vector spaces.

2.1 Coordinate systems of General Vector Spaces

A line is thought of as 1-Dimensional because every point on that line can be specified by 1 coordinate. In the same way a plane is thought of as 2 Dimensional because every point on that plane can be specified by 2 coordinates and so on. What defines this coordinate system? The most common form of defining a coordinate system is the use of coordinate axes. In the case of the plane the x and y axes are used most frequently. But there is also a way of specifying the coordinate system with vectors. This can be done by replacing each axis with a vector of length one that points in the positive direction of the axis. In the case of the $x - y$ plane the x and y-axes are replaced by the well known unit vectors $\mathbf{i}$ and $\mathbf{j}$ respectively. Let O be the origin of the system and P be any point in the plane. The point P can be specified by the vector $\overline{OP}$. Every vector, $\overline{OP}$ can be written as a linear combination of $\mathbf{i}$ and $\mathbf{j}$:

$$\overline{OP} = a\mathbf{i} + b\mathbf{j}$$

The coordinates of P, corresponding to this coordinate system, are (a, b).

Informally stated, vectors such as $\mathbf{i}$ and $\mathbf{j}$ that specify a coordinate system are called "basis vectors" for that system. Although in the preceding discussion our basis vectors were chosen to be of unit length and mutually perpendicular this is not essential. As long as linear combinations of the vectors chosen are capable of specifying all points in the plane. In our example this only requires that the two vectors are not colinear. Different basis vectors however do change the coordinates of a point, as the following example demonstrates.

Example Let $S = \{\mathbf{i}, \mathbf{j}\}$, $U = \{\mathbf{i}, 2\mathbf{j}\}$ and $V = \{\mathbf{i} + \mathbf{j}, \mathbf{j}\}$. Let the sets S, U and V be three sets of basis vectors. Let P be the point $\mathbf{i} + 2\mathbf{j}$. The coordinates of P relative to each set of basis vectors is:

$$S \to (1, 2)$$

$$U \to (1, 1)$$

$$T \to (1, 1)$$

The following definition makes the preceding ideas more precise and enables the extension of a coordinate system to general vector spaces.

Definition

- If V is any vector space and $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ is a set of vectors in V, then S is called a **basis** for V if the following two conditions hold:

  (a) S is linearly independent

  (b) S spans V

A basis is the vector space generalization of a coordinate system in 2-space and 3-space. The following theorem will aid in understanding how this is so.

Theorem 2.1. *If $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ is a basis for a vector space V, then every vector $\mathbf{v}$ in V can be expressed in the form $\mathbf{v} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_n\mathbf{v_n}$ in exactly one way.*

Proof. Since S spans V, it follows from the definition of a spanning set that every vector in V is expressible as a linear combination of the vectors in S. To see that there is only one way to express a vector as a linear combination of the vectors in S, suppose that some vector $\mathbf{v}$ can be written as

$$\mathbf{v} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_n\mathbf{v_n}$$

and also as

$$\mathbf{v} = k_1\mathbf{v_1} + k_2\mathbf{v_2} + \cdots + k_n\mathbf{v_n}$$

Subtracting the second equation from the first gives

$$\mathbf{0} = (c_1 - k_1)\mathbf{v_1} + (c_2 - k_2)\mathbf{v_2} + \cdots + (c_n - k_n)\mathbf{v_n}$$

Since the right side of this equation is a linear combination of vectors in S, the linear independence of S implies that

$$(c_1 - k_1) = 0, (c_2 - k_2) = 0, \ldots, (c_n - k_n)$$

That is

$$c_1 = k_1, c_2 = k_2, \ldots, c_n = k_n$$

Thus the two expressions for $\mathbf{v}$ are the same. $\square$

Definitions

- If $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ is a basis for a vector space V, and

$$\mathbf{v} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_n\mathbf{v_n}$$

  is the expression for a vector $\mathbf{v}$ in terms of the basis S, then the scalars $c_1, c_2, \ldots, c_n$ are called the **coordinates** of $\mathbf{v}$ relative to the basis S. The vector $(c_1, c_2, \ldots, c_n)$ in $\mathbb{F}^n$ constructed from these coordinates is called the **coordinate vector of v relative to** S; it is denoted by

$$[\mathbf{v}]_S = (c_1, c_2, \ldots, c_n)$$

- If $\mathbf{v} = [\mathbf{v}]_S$ then S is called the **standard basis**.

Remark It should be noted that coordinate vectors depend not only on the basis S but also on the order in which the basis vectors are written; a change in the order of the basis vectors results in a corresponding change of order for the entries in the coordinate vectors.

Examples

1. In example 3 of Section 1.5 it was shown that if

$$\mathbf{i} = (1, 0, 0), \quad \mathbf{j} = (0, 1, 0), \quad \mathbf{k} = (0, 0, 1)$$

then $S = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is a linearly independent set in $\mathbb{R}^3$. This set also spans $\mathbb{R}^3$ since any vector $\mathbf{v} = (a, b, c)$ can be written as

$$\mathbf{v} = (a, b, c) = a(1, 0, 0) + b(0, 1, 0) + c(0, 1, 1) = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

Thus, S is a basis for $\mathbb{R}^3$. It is in fact a **standard basis** for $\mathbb{R}^3$. Looking at the coefficients of $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ above, it follows that the coordinates of $\mathbf{v}$ relative to the standard basis are a, b and c, so

$$[\mathbf{v}]_S = (a, b, c)$$

and so we have

$$[\mathbf{v}]_S = \mathbf{v}.$$

2.2 Dimension of General Vector Spaces

Definition

- A nonzero vector space V is called **finite-dimensional** if it contains a finite set of vectors $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ that forms a basis. If no such set exists, V is called **infinite-dimensional**. In addition, the zero vector space is regarded as finite-dimensional.

Examples

- The vector spaces $\mathbb{F}^n$ and P_n are both finite-dimensional.

- The vector space of all real valued functions defined on $(-\infty, \infty)$ is infinite-dimensional.

Theorem 2.2. *If V is a finite-dimensional vector space and $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ is any basis, then:*

(a) Every set with more than n vectors is linearly dependent.

(b) No set with fewer than n vectors spans V.

Proof. (a) Let $S' = \{\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_m}\}$ be any set of m vectors in V, where $m > n$. It remains to be shown that S' is linearly dependent. Since $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ is a basis for V, each $\mathbf{w_i}$ can be expressed as a linear combination of the vectors in S, say:

$$\mathbf{w_1} = a_{11}\mathbf{v_1} + a_{21}\mathbf{v_2} + \cdots + a_{n1}\mathbf{v_n}$$

$$\mathbf{w_2} = a_{12}\mathbf{v_1} + a_{22}\mathbf{v_2} + \cdots + a_{n2}\mathbf{v_n}$$

$$\vdots$$

$$\mathbf{w_m} = a_{1m}\mathbf{v_1} + a_{2m}\mathbf{v_2} + \cdots + a_{nm}\mathbf{v_n}$$

To show that S' is linearly dependent, scalars $k_1, k_2, \ldots, k_n$ must be found, not all zero, such that

$$k_1\mathbf{w_1} + k_2\mathbf{w_2} + \cdots + k_m\mathbf{w_m} = \mathbf{0}$$

combining the above 2 systems of equations gives

$$(k_1 a_{11} + k_2 a_{12} + \cdots + k_m a_{1m})\mathbf{v_1}$$
$$+ (k_1 a_{21} + k_2 a_{22} + \cdots + k_m a_{2m})\mathbf{v_2}$$
$$\ddots$$
$$+ (k_1 a_{n1} + k_2 a_{n2} + \cdots + k_m a_{nm})\mathbf{v_n} = \mathbf{0}$$

Thus, from the linear independence of S, the problem of proving that S' is a linearly dependent set reduces to showing there are scalars $k_1, k_2, \ldots, k_m$, not all zero, that satisfy

$$a_{11}k_1 + a_{12}k_2 + \cdots + a_{1m}k_m = 0$$

$$a_{21}k_1 + a_{22}k_2 + \cdots + a_{2m}k_m = 0$$

$$\vdots$$

24

$$a_{n1}k_1 + a_{n2}k_2 + \cdots + a_{nm}k_m = 0$$

As the system is homogenous and there are more unknowns than equations $(m > n)$, we have an infinite number of solutions, or in other words there are non trivial solutions such that $k_1, k_2, \ldots, k_m$ are not all zero.

(b) Let $S' = \{\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_m}\}$ be any set of m vectors in V, where $m < n$. It remains to be shown that S' does not span V. The proof is by contradiction: assume S' spans V. This leads to a contradiction of the linear dependence of the basis $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ of V.

If S' spans V, then every vector in V is a linear combination of the vectors in S'. In particular, each basis vector $\mathbf{v_i}$ is a linear combination of the vectors in S', say

$$\mathbf{v_1} = a_{11}\mathbf{w_1} + a_{21}\mathbf{w_2} + \cdots + a_{n1}\mathbf{w_m}$$

$$\mathbf{v_2} = a_{12}\mathbf{w_1} + a_{22}\mathbf{w_2} + \cdots + a_{n2}\mathbf{w_m}$$

$$\vdots$$

$$\mathbf{v_n} = a_{1n}\mathbf{w_1} + a_{2n}\mathbf{w_2} + \cdots + a_{mn}\mathbf{w_m}$$

To obtain the contradiction it will be shown that there exist scalars $k_1, k_2, \ldots, k_n$ not all zero, such that

$$k_1\mathbf{v_1} + k_2\mathbf{v_2} + \cdots + k_n\mathbf{v_n} = \mathbf{0}$$

Observe the similarity to the above two systems compared with those given in the proof of (a). It can be seen that they are identical except that the $\mathbf{w}$'s and the $\mathbf{v}$'s and the m's and n's have been interchanged. Thus the above system in the same way again reduces to the problem of finding $k_1, k_2, \ldots, k_n$ not all zero, that satisfy

$$a_{11}k_1 + a_{12}k_2 + \cdots + a_{1m}k_n = 0$$

$$a_{21}k_1 + a_{22}k_2 + \cdots + a_{2m}k_n = 0$$

25

$$\vdots$$

$$a_{m1}k_1 + a_{m2}k_2 + \cdots + a_{mn}k_n = 0$$

As the system is homogenous and there are more unknowns than equations $(n > m)$, we have an infinite number of solutions, or in other words there exist non trivial solutions such that $k_1, k_2, \ldots, k_m$ are not all zero. Hence the contradiction.

$\square$

The last theorem essentially states the following. Let S be a set with n vectors which forms a basis for the vector space V. Let S' be another set of vectors in V consisting of m vectors. If m is greater than n, S' cannot form a basis for V as the vectors in S' cannot be linearly independent. If m is less than n, S' cannot form a basis for V because it does not span V. Thus, theorem 2.2 leads directly into one of the most important theorems in linear algebra.

Theorem 2.3. *All bases for a finite-dimenstional vector space have the same number of vectors.*

And thus the concept of dimension is almost complete. All that is needed is a definition.

Definition

- The **dimension** of a finite-dimensional vector space V, denoted by $dim(V)$, is defined to be the number of vectors in a basis for V. In addition, the zero vector space has dimension zero.

Examples

1. The dimensions of some common vector spaces are given below:

$$dim(\mathbb{F}^n) = n$$
$$dim(P_n) = n + 1$$
$$dim(M_{n \times n}(\mathbb{F})) = mn$$

26

2. Determine a basis (and hence dimension) for the solution space of the homogenous system:

$$
\begin{aligned}
2x_1 + 2x_2 - x_3 + x_5 &= 0 \\
-x_1 - x_2 + 2x_3 - 3x_4 + x_5 &= 0 \\
x_1 + x_2 - 2x_3 - x_5 &= 0 \\
x_3 + x_4 + x_5 &= 0
\end{aligned}
$$

2.3   Related Theorems

The remaining part of this section states theorems which illustrate the subtle relationships among the concepts of spanning, linear independence, basis and dimension. In many ways these theorems form the building blocks of other results in linear algebra.

Theorem 2.4. Plus/Minus Theorem. *Let S be a nonempty set of vectors in a vector space V.*

(a) *If S is a linearly independent set, and if $\mathbf{v}$ is a vector in V that is outside of the span(S), then the set $S \cup \{\mathbf{v}\}$ that results by inserting $\mathbf{v}$ is still linearly independent.*

(b) *If $\mathbf{v}$ is a vector in S that is expressible as a linear combination of other vectors in S, and if $S - \{\mathbf{v}\}$ denotes the set obtained by removing $\mathbf{v}$ from S, then S and $S - \{\mathbf{v}\}$ span the same space: that is,*

$$span(S) = span(S - \{\mathbf{v}\})$$

A proof will not be included, but the theorem can be visualised in $\mathbb{R}^3$ as follows.

(a) Consider two linearly independent vectors in $\mathbb{R}^3$. These two vectors span a plane. If you add a third vector to them that is not in the plane, then the three vectors are still linearly independent and they span the entire domain of $\mathbb{R}^3$.

(b) Consider three non-colinear vectors in a plane that form a set S. The set S spans the plane. If any one of the vectors is removed from S to give S' it is clear that S' still spans the plane. That is $span(S) = span(S')$.

Theorem 2.5. *If V is an n-dimensional vector space and if S is a set in V with exactly n vectors, then S is a basis for V if either S spans V or S is linearly independent.*

Proof. Assume that S has exactly n vectors and spans V. To prove that S is a basis it must be shown that S is a linearly independent set. But if this is not so, then some vector $\mathbf{v}$ in S is a linear combination of the remaining vectors. If this vector is removed from S, then it follows from the theorem 2.4(b) that the remaining set of n-1 vectors still spans V. But this is impossible, since it follows from theorem 2.2(b), that no set with fewer than n vectors can span an n-dimensional vector space. Thus, S is linearly independent.

Assume S has exactly n vectors and is a linearly independent set. To prove that S is a basis it must be shown that S spans V. But if this is not so, then there is some vector $\mathbf{v}$ in V that is not in span(S). If this vector is inserted in S, then it follows from the theorem 2.4(a) that this set of n+1 vectors is still linearly independent. But this is impossible because it follows from theorem 2.2(a) that no set with more than n vectors in an n-dimensional vector space can be linearly independent. Thus S spans V. $\square$

Examples

- $\mathbf{v_1} = (-3, 8)$ and $\mathbf{v_2} = (1, 1)$ form a basis for $\mathbb{R}^2$ because $\mathbb{R}^2$ has dimension two and v_1 and v_2 are linearly independent.

Theorem 2.6. *Let S be a finite set of vectors in a finite-dimensional vector space V.*

(a) *If S spans V but is not a basis for V, then S can be reduced to a basis for V by removing appropriate vectors from S.*

(b) If S is a linearly independent set that is not already a basis for V, then S can be enlarged to a basis for V by inserting appropriate vectors into S.

Proof.   (a) The proof is constructive and is called the **left to right algorithm**.

Let $\mathbf{v_{c_1}}$ be the first nonzero vector in the set S. Choose the next vector in the list which is not a linear combination of $\mathbf{v_{c_1}}$ and call it $\mathbf{v_{c_2}}$. Find the next vector in the list which is not a linear combination of $\mathbf{v_{c_1}}$ and $\mathbf{v_{c_2}}$ and call it $\mathbf{v_{c_3}}$. Continue in such a way until the number of vectors chosen equals $dim(V)$.

(b) This proof is also constructive.

Let V be a vector space. Begin with $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_r}$ which form a linearly independent family in V. Let $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ be a basis for V. Now it is necessary and important that $r < n$. To extend the basis, simply apply the left to right algorithm to the set (note that this set spans V because it contains a basis within it)

$$\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_r}, \mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$$

This will select a basis for V that commences with $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_r}$

$\square$

Theorem 2.7. *If W is a subspace of a finite-dimensional vector space V, then $dim(W) \leq dim(V)$; moreover, if $dim(W) = dim(V)$, then $W = V$*

Proof. Let $S = \{\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_m}\}$ be a basis for W. Either S is also a basis for V or it is not. If it is, then $dim(W) = dim(V) = m$. If it is not, then by the previous theorem, vectors can be added to the linearly independent set S to make it into a basis for V, so $dim(W) < dim(V)$. Thus, $dim(W) \leq dim(V)$ in all cases. If $dim(W) = dim(V)$, then S is a set of m linearly independent vectors in the m-dimensional vector space V; hence by theorem 2.5, S is a basis for V. Therefore $W = V$. $\square$

2.3.1 Quiz

True or false?

(a) The zero vector space has no basis.

(b) Every vector space that is spanned by a finite set has a basis.

(c) Every vector space has a finite basis.

(d) A vector space cannot have more than one basis.

(e) If a vector space has a finite basis, then the number of vectors in every basis is the same.

(f) Suppose that V is a finite dimensional vector space, S_1 is a linear independent subset of V, and S_2 is a subset of V that spans V. Then S_1 cannot contain more vectors than S_2.

(g) If S spans the vector space V, then every vector in V can be written as a linear combination of vectors in S in only one way.

(h) Every subspace of a finite dimensional vector space is finite dimensional.

(i) If V is an n dimensional vector space, then V has exactly one subspace with dimension 0 and one with dimension n.

(j) If V is an n dimensional vector space, and if S is a subset of V with n vectors, then S is linearly independent if and only if S spans V.

3 INNER PRODUCT SPACES AND ORTHONOR-MAL BASES

In many applications of vector spaces, we are concerned with the notion of measurement. In this section we introduce the idea of length through the structure of inner product spaces. We only consider $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Definition

Let V be a vector space over $\mathbb{F}$. We define an inner product $\langle , \rangle$ on V to be a function that assigns a scalar $\langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{F}$ to every pair of ordered vectors $\mathbf{u}, \mathbf{v} \in V$ such that the following properties hold for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\alpha \in \mathbb{F}$:

(a) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

(b) $\langle \alpha \mathbf{u}, \mathbf{v} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle$

(c) $\overline{\langle \mathbf{u}, \mathbf{v} \rangle} = \langle \mathbf{v}, \mathbf{u} \rangle$

(d) $\langle \mathbf{u}, \mathbf{u} \rangle > 0$ if $\mathbf{u} \neq \mathbf{0}$.

The main example is when $V = \mathbb{F}^n$. In this case we often use the notation $\langle \mathbf{u}, \mathbf{v} \rangle \equiv \mathbf{u} \cdot \mathbf{v}$ which is determined by

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i \overline{v_i}$$

where $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$.

Definitions

- A vector space V over $\mathbb{F}$ endowed with a specific inner product is called an inner product space. If $\mathbb{F} = \mathbb{R}$ then V is said to be a real inner product space, whereas if $\mathbb{F} = \mathbb{C}$ we call V a complex inner product space.

- The norm (or length, or magnitude) of a vector $\mathbf{u}$ is given by $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$.

- Two vectors $\mathbf{u}, \mathbf{v}$ in an inner product space are said to be **orthogonal** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

- If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors and both $\mathbf{u}$ and $\mathbf{v}$ have a magnitude of one (with respect to $\langle, \rangle$), then $\mathbf{u}$ and $\mathbf{v}$ are said to be **orthonormal**.

- A set of vectors in an inner product space is called an **orthogonal set** if all pairs of distinct vectors in the set are orthogonal. An orthogonal set in which each vector has a magnitude of one is called an **orthonormal set**.

The following additional properties follow easily from the axioms:

Theorem 3.1. *Let V be an inner product space, $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $c \in \mathbb{F}$.*

(a) $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$.

(b) $\langle \mathbf{x}, c\mathbf{y} \rangle = \bar{c}\langle \mathbf{x}, \mathbf{y} \rangle$.

(c) $\langle \mathbf{x}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{x} \rangle = 0$.

(d) $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

(e) If $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle$ for all $\mathbf{x} \in V$, then $\mathbf{y} = \mathbf{z}$.

Proof. (a) - (d) exercises

(e) By part (a) and (b), $\langle \mathbf{x}, \mathbf{y} - \mathbf{z} \rangle = 0$ for all $\mathbf{x} \in V$. Since this is true for all $\mathbf{x}$, it is true for $\mathbf{x} = \mathbf{y} - \mathbf{z}$, thus $\langle \mathbf{y} - \mathbf{z}, \mathbf{y} - \mathbf{z} \rangle = 0$. By (d) this implies that $\mathbf{y} = \mathbf{z}$. $\square$

Now that the groundwork has been laid the following theorem can be stated. The proof of this result is extremely important, since it makes use of an algorithm, or method, for converting an arbitary basis into an orthonormal basis.

Theorem 3.2. *Every non-zero finite dimensional inner product space V has an orthonormal basis.*

Proof. Let $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_m}\}$ be any basis for V. It suffices to show that V has an orthogonal basis, since the vectors in the orthogonal basis can be normalized to produce an orthonormal basis for V. The following sequence of steps will produce an orthogonal basis $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_m}\}$ for V.

Step 1 Let $\mathbf{v_1} = \mathbf{u_1}$.

Step 2 Obtain a vector $\mathbf{v_2}$ that is orthogonal to $\mathbf{v_1}$ by computing the component of $\mathbf{u_2}$ that is orthogonal to the space W_1 spanned by $\mathbf{v_1}$. This can be done using the formula:

$$\mathbf{v_2} = \mathbf{u_2} - \left(\frac{\langle \mathbf{u_2}, \mathbf{v_1} \rangle}{\langle \mathbf{v_1}, \mathbf{v_1} \rangle} \right) \mathbf{v_1}$$

Of course, if $\mathbf{v_2} = \mathbf{0}$, then $\mathbf{v_2}$ is not a basis vector. But this cannot happen, since it would then follow from the preceding formula for $\mathbf{v_2}$ that

$$\mathbf{u_2} = \left(\frac{\langle \mathbf{u_2}, \mathbf{v_1} \rangle}{\langle \mathbf{v_1}, \mathbf{v_1} \rangle} \right) \mathbf{v_1} = \left(\frac{\langle \mathbf{u_2}, \mathbf{v_1} \rangle}{\langle \mathbf{u_1}, \mathbf{u_1} \rangle} \right) \mathbf{u_1}$$

which says that $\mathbf{u_2}$ is a multiple of $\mathbf{u_1}$, contradicting the linear independence of the basis $S = \{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$.

Step 3 To construct a vector $\mathbf{v_3}$ that is orthogonal to both $\mathbf{v_1}$ and $\mathbf{v_2}$, compute the component of $\mathbf{u_3}$ orthogonal to the space W_2 spanned by $\mathbf{v_1}$ and $\mathbf{v_2}$ using the formula:

$$\mathbf{v_3} = \mathbf{u_3} - \left(\frac{\langle \mathbf{u_3}, \mathbf{v_1} \rangle}{\langle \mathbf{v_1}, \mathbf{v_1} \rangle} \right) \mathbf{v_1} - \left(\frac{\langle \mathbf{u_3}, \mathbf{v_2} \rangle}{\langle \mathbf{v_2}, \mathbf{v_2} \rangle} \right) \mathbf{v_2}$$

As in step 2, the linear independence of $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ ensures that $\mathbf{v_3} \neq \mathbf{0}$. The remaining details are left as an exercise.

Step 4 To determine a vector $\mathbf{v_4}$ that is orthogonal to $\mathbf{v_1}, \mathbf{v_2}$ and $\mathbf{v_3}$, compute the component of $\mathbf{u_4}$ orthogonal to the space W_3 spanned by $\mathbf{v_1}, \mathbf{v_2}$ and $\mathbf{v_3}$ using the formula

$$\mathbf{v_4} = \mathbf{u_4} - \left(\frac{\langle \mathbf{u_4}, \mathbf{v_1} \rangle}{\langle \mathbf{v_1}, \mathbf{v_1} \rangle} \right) \mathbf{v_1} - \left(\frac{\langle \mathbf{u_4}, \mathbf{v_2} \rangle}{\langle \mathbf{v_2}, \mathbf{v_2} \rangle} \right) \mathbf{v_2} - \left(\frac{\langle \mathbf{u_4}, \mathbf{v_3} \rangle}{\langle \mathbf{v_3}, \mathbf{v_3} \rangle} \right) \mathbf{v_3}$$

Continuing in this way, an orthogonal set of vectors, $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_m}\}$, will be obtained after m steps. Since V is an m-dimensional vector space and every orthogonal set is linearly independent, the set $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ is an orthogonal basis for V. $\quad\square$

This preceding step-by-step construction for converting an arbitary basis into an orthogonal basis is called the **Gram-Schmidt process**.

Examples: *THE GRAM-SCHMIDT PROCESS*

1. Consider the vector space $\mathbb{R}^3$ with the Euclidean inner product. Apply the Gram-Schmidt process to transform the basis vectors $\mathbf{u_1} = (1, 1, 1)$, $\mathbf{u_2} = (0, 1, 1)$, $\mathbf{u_3} = (0, 0, 1)$ into an orthogonal basis $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$; then normalize the orthogonal basis vectors to obtain an orthonormal basis $\{\mathbf{q_1}, \mathbf{q_2}, \mathbf{q_3}\}$.

   Step 1

   $$\mathbf{v_1} = \mathbf{u_1} = (1, 1, 1)$$

   Step 2

   $$
   \begin{aligned}
   \mathbf{v_2} &= \mathbf{u_2} - \left(\frac{\mathbf{u_2} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}} \right) \mathbf{v_1} \\
   &= (0, 1, 1) - \frac{2}{3}(1, 1, 1) = \left(\frac{-2}{3}, \frac{1}{3}, \frac{1}{3} \right)
   \end{aligned}
   $$

   Step 3

   $$
   \begin{aligned}
   \mathbf{v_3} &= \mathbf{u_3} - \left(\frac{\mathbf{u_3} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}} \right) \mathbf{v_1} - \left(\frac{\mathbf{u_3} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}} \right) \mathbf{v_2} \\
   &= (0, 0, 1) - \frac{1}{3}(1, 1, 1) - \frac{1/3}{2/3} \left(\frac{-2}{3}, \frac{1}{3}, \frac{1}{3} \right) \\
   &= \left(0, -\frac{1}{2}, \frac{1}{2} \right)
   \end{aligned}
   $$

   Thus,

   $$\mathbf{v_1} = (1, 1, 1), \quad \mathbf{v_2} = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3} \right), \quad \mathbf{v_3} = \left(0, -\frac{1}{2}, \frac{1}{2} \right)$$

   form an orthogonal basis for $\mathbb{R}^3$. The norms of these vectors are

   $$\|\mathbf{v_1}\| = \sqrt{3}, \quad \|\mathbf{v_2}\| = \frac{\sqrt{6}}{3}, \quad \|\mathbf{v_3}\| = \frac{1}{\sqrt{2}}$$

   so an orthonormal basis for $\mathbb{R}^3$ is

   $$\mathbf{q_1} = \frac{\mathbf{v_1}}{\|\mathbf{v_1}\|} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right), \quad \mathbf{q_2} = \frac{\mathbf{v_2}}{\|\mathbf{v_2}\|} = \left(\frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right)$$

$$\mathbf{q_3} = \frac{\mathbf{v_3}}{\|\mathbf{v_3}\|} = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

The Gramm-Schmidt process with subsequent normalization not only converts an arbitary basis $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ into an orthonormal basis $\{\mathbf{q_1}, \mathbf{q_2}, \ldots, \mathbf{q_n}\}$, but it does it in such a way that for $k \geq 2$ the following relationships hold:

- $\{\mathbf{q_1}, \mathbf{q_2}, \ldots, \mathbf{q_k}\}$ is an orthonormal basis for the space spanned by $\{\mathbf{u_1}, \ldots, \mathbf{u_k}\}$.

- $\mathbf{q_k}$ is orthogonal to $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_{k-1}}\}$.

The proofs are omitted but these facts should become evident after some thoughtful examination of the proof of Theorem 3.1.

3.1   Quiz

True or false?

- An inner product is a scalar-valued function on the set of ordered pairs of vectors.

- An inner product space must be over the field of real or complex numbers.

- An inner product is linear in both components.

- If x, y and z are vectors in an inner product space such that $\langle x, y \rangle = \langle x, z \rangle$, then $y = z$.

- If $\langle x, y \rangle = 0$ for all x in an inner product space, then $y = 0$.

4 LINEAR TRANSFORMATIONS AND MATRICES

Definitions

- Let V, W be vector spaces over a field $\mathbb{F}$. A function that maps V into W, $T : V \to W$, is called a **linear transformation** from V to W if for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V and all scalars $c \in \mathbb{F}$

  (a) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

  (b) $T(c\mathbf{u}) = cT(\mathbf{u})$

- In the special case where $V = W$, the linear transformation $T : V \to V$ is called a **linear operator** on V.

- Let A be an $m \times n$ matrix and let $T : \mathbb{F}^n \to \mathbb{F}^m$ be the linear transformation defined by $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$. Then as a matter of notational convention it is said that T is the linear transformation T_A.

4.0.1 Basic Properties of Linear Transformations

Theorem 4.1. *If $T : V \to W$ is a linear transformation, then:*

(a) If T is linear, then $T(\mathbf{0}) = \mathbf{0}$

(b) T is linear if and only if $T(a\mathbf{v} + \mathbf{w}) = aT(\mathbf{v}) + T(\mathbf{w})$ for all $\mathbf{v}, \mathbf{w}$ in V and $a \in \mathbb{F}$.

(c) $T(\mathbf{v} - \mathbf{w}) = T(\mathbf{v}) - T(\mathbf{w})$ for all $\mathbf{v}$ and $\mathbf{w}$ in V.

Part (a) of the above theorem states that a linear transformation maps $\mathbf{0}$ into $\mathbf{0}$. This property is useful for identifying transformations that are not linear. Part (b) is usually used to show that a transformation is linear.

Examples

1. T_A is a linear transformation. Let A be an $m \times n$ matrix and let $T : \mathbb{F}^n \to \mathbb{F}^m$ be the linear transformation defined by $T_A(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{F}^n$. Let $\mathbf{u}$ and $\mathbf{v} \in \mathbb{F}^n$, then

$$T(\lambda\mathbf{u} + \mathbf{v}) = A(\lambda\mathbf{u} + \mathbf{v})$$

$$= \lambda A\mathbf{u} + A\mathbf{v}$$

$$= \lambda T_A(\mathbf{u}) + T_A(\mathbf{v})$$

and thus T_A is a linear transformation.

2. If I is the $n \times n$ identity matrix, then for every vector $\mathbf{x}$ in $\mathbb{F}^n$

$$T_I(\mathbf{x}) = I\mathbf{x} = \mathbf{x}$$

so multiplication by I maps every vector in $\mathbb{F}^n$ into itself. $T_I(\mathbf{x})$ is called the **identity operator** on $\mathbb{F}^n$.

3. Let A, B and X be $n \times n$ matrices. Then $Y = AX - XB$ is also $n \times n$.

Let $V = M_{n \times n}(\mathbb{F})$ be the vector space of all $n \times n$ matrices. Then $Y = AX - XB$ defines a transformation $T : V \to V$. The transformation is linear since

$$
\begin{aligned}
T(\lambda X_1 + X_2) &= A(\lambda X_1 + X_2) - (\lambda X_1 + X_2)B \\
&= \lambda AX_1 + AX_2 - \lambda X_1 b - X_2 B \\
&= \lambda(AX_1 - X_1 B) + AX_2 - X_2 B \\
&= \lambda T(X_1) + T(X_2)
\end{aligned}
$$

Theorem 4.2. *If $T : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation, then there exists an $m \times n$ matrix A such that $T = T_A$.*

Example

1. Find the 2×2 matrix A such that $T = T_A$ has the property that

$$T\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \text{ and } T\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

4.1 Geometric Transformations in $\mathbb{R}^2$

This section consists of various different transformations of the form T_A that have a geometrical interpretation. Such transformations form the building blocks for understanding linear transformations.

Examples of Geometric Transformations

- Operators on $\mathbb{R}^2$ and $\mathbb{R}^3$ that map each vector into its symmetric image about some line or plane are called **reflection operators**. Such operators are of the form T_A and are thus linear. There are three main reflections in $\mathbb{R}^2$. These are summarised below. Considering the transformation from the coordinates (x, y) to (w_1, w_2) the properties of the operator are as follows.

    1. **Reflection about the y-axis**: The equations for this transformation are

    $$\begin{aligned} w_1 &= -x \\ w_2 &= y \end{aligned}$$

    The standard matrix for the transformation is clearly

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

    To demonstrate the reflection, consider the example below.

    $$\text{Let } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

    $$\text{therefore } T_A(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

    2. **Reflection about the x-axis**: The equations for this transformation are

    $$\begin{aligned} w_1 &= x \\ w_2 &= -y \end{aligned}$$

38

The standard matrix for the transformation is clearly

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

To demonstrate the reflection, consider the example below.

$$\text{Let } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\text{therefore } T_A(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

3. **Reflection about the line** $y = x$: The equations for this transformation are

$$\begin{aligned} w_1 &= y \\ w_2 &= x \end{aligned}$$

The standard matrix for the transformation is clearly

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

To demonstrate the reflection, consider the example below.

$$\text{Let } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\text{therefore } T_A(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

- Operators on $\mathbb{R}^2$ and $\mathbb{R}^3$ that map each vector into its orthogonal projection on a line or plane through the origin are called **orthogonal projection operators**.

39

Such operators are of the form T_A and are thus linear. There are two main projections in $\mathbb{R}^2$. These are summarised below. Considering the transformation from the coordinates (x, y) to (w_1, w_2) the properties of the operator are as follows.

1. **Orthogonal projection onto the x-axis**: The equations for this transformation are

$$
\begin{aligned}
w_1 &= x \\
w_2 &= 0
\end{aligned}
$$

The standard matrix for the transformation is clearly

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

To demonstrate the projection, consider the example below.

$$
\text{Let } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}
$$

$$
\text{therefore } T_A(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}
$$

2. **Orthogonal projection on the y-axis**: The equations for this transformation are

$$
\begin{aligned}
w_1 &= 0 \\
w_2 &= y
\end{aligned}
$$

The standard matrix for the transformation is clearly

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

To demonstrate the projection, consider the example below.

$$\text{Let } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\text{therefore } T_A(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

- An operator that rotates each vector in $\mathbb{R}^2$, through a fixed angle θ is called a **rotation operator** on $\mathbb{R}^2$. Such operators are of the form T_A and are thus linear. There is only one rotation in $\mathbb{R}^2$, due to the generality of the formula. This rotation is summarised below. Considering the transformation from the coordinates (x, y) to (w_1, w_2) the properties of the operator are as follows.

  1. **Rotation through an angle** θ: The equations for this transformation are

$$
\begin{aligned}
w_1 &= x \cos \theta - y \sin \theta \\
w_2 &= x \sin \theta + y \cos \theta
\end{aligned}
$$

  The standard matrix for the transformation is clearly

$$A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

  To demonstrate the projection, consider the example below.

$$\text{Let } \theta = 30° \text{ and let } \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{therefore } T_A(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}$$

- If k is a nonnegative scalar, then the operator $T(\mathbf{x}) = k\mathbf{x}$ on $\mathbb{R}^2$ and $\mathbb{R}^3$ is called a **contraction with factor** k if $0 \leq k \leq 1$, and a **dilation with factor** k if $k \geq 1$. Such operators are of the form T_A and are thus linear. The contraction and the dilation operators are summarised below. Considering the transformation from the coordinates (x, y) to (w_1, w_2) the properties of the operator are as follows.

  1. **Contraction with factor k on $\mathbb{R}^2$, $(0 \leq k \leq 1)$:** The equations for this transformation are.

  $$
  \begin{aligned}
  w_1 &= kx \\
  w_2 &= ky
  \end{aligned}
  $$

  The standard matrix for the transformation is clearly

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

  To demonstrate the contraction, consider the example below.

  $$
  \text{Let } k = \frac{1}{2} \text{ and let } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}
  $$

  $$
  \text{therefore } T_A(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix}
  $$

  2. **Dilation with factor k on $\mathbb{R}^2$, $(k \geq 1)$:** The equations for this transformation are

  $$
  \begin{aligned}
  w_1 &= kx \\
  w_2 &= ky
  \end{aligned}
  $$

  The standard matrix for the transformation is clearly

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

To demonstrate the dilation, consider the example below.

$$\text{Let } k = 2 \text{ and let } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\text{therefore } T_A(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

4.2 Product of Linear Transformations

Definition

- If $T_1 : U \rightarrow V$ and $T_2 : V \rightarrow W$ are linear transformations, the **composite of** T_2 **with** T_1 denoted by $T_2 \circ T_1$, is the function defined by the formula

$$(T_2 \circ T_1)(\mathbf{u}) = T_2(T_1(\mathbf{u}))$$

where $\mathbf{u}$ is a vector in U.

Remark: Observe that this definition requires the domain of T_2 (which is V) to contain the range of T_1; this is essential for the formula $T_2(T_1(\mathbf{u}))$ to make sense.

The next result shows that the composition of two linear transformations is itself a linear transformation.

Theorem 4.3. *If* $T_1 : U \rightarrow V$ *and* $T_2 : V \rightarrow W$ *are linear transformations, then* $(T_2 \circ T_1) : U \rightarrow W$ *is also a linear transformation.*

Proof. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in U and $s \in \mathbb{F}$, then it follows from the definition of a composite transformation and from the linearity of T_1 and T_2 that

$$\begin{aligned} T_2 \circ T_1(s\mathbf{u} + \mathbf{v}) &= T_2(T_1(s\mathbf{u} + \mathbf{v})) \\ &= T_2(sT_1(\mathbf{u}) + T_1(\mathbf{v})) \\ &= sT_2(T_1(\mathbf{u})) + T_2(T_1(\mathbf{v})) \\ &= sT_2 \circ T_1(\mathbf{u}) + T_2 \circ T_1(\mathbf{v}) \end{aligned}$$

and thus the proof is complete. $\qquad\square$

Examples

1. Let A be an $m \times n$ matrix, and B be an $n \times p$ matrix, then AB is an $m \times p$ matrix. Also $T_A : \mathbb{F}^n \to \mathbb{F}^m$, and $T_B : \mathbb{F}^p \to \mathbb{F}^n$ are both linear transformations. Then

$$
\begin{aligned}
T_A \circ T_B &= T_A(T_B(\mathbf{x})) \\
&= AB\mathbf{x} \\
&= (AB)\mathbf{x} \\
&= T_{AB}(\mathbf{x})
\end{aligned}
$$

where $\mathbf{x} \in \mathbb{F}^p$. And therefore $T_A \circ T_B = T_{AB} : \mathbb{F}^p \to \mathbb{F}^m$.

2. If V has a basis $\beta = \{\mathbf{v_1}, \mathbf{v_2}\}$ and $T : V \to V$ is a linear transformation given by

$$
\begin{aligned}
T(\mathbf{v_1}) &= 2\mathbf{v_1} + 3\mathbf{v_2} \\
T(\mathbf{v_2}) &= -7\mathbf{v_1} + 8\mathbf{v_2}
\end{aligned}
$$

To find $T \circ T(-\mathbf{v_1} + 3\mathbf{v_2})$ takes two steps as shown below.

$$
\begin{aligned}
T(-\mathbf{v_1} + 3\mathbf{v_2}) &= -T(\mathbf{v_1}) + 3T(\mathbf{v_2}) \\
&= -2\mathbf{v_1} - 3\mathbf{v_2} + 3(-7\mathbf{v_1} + 8\mathbf{v_2}) \\
&= -23\mathbf{v_1} + 21\mathbf{v_2}
\end{aligned}
$$

Hence

$$
\begin{aligned}
T \circ T(-\mathbf{v_1} + 3\mathbf{v_2}) &= T(-23\mathbf{v_1} + 21\mathbf{v_2}) \\
&= -23T(\mathbf{v_1}) + 21T(\mathbf{v_2}) \\
&= -23(2\mathbf{v_1} + 3\mathbf{v_2}) + 21(-7\mathbf{v_1} + 8\mathbf{v_2}) \\
&= -193\mathbf{v_1} + 99\mathbf{v_2}
\end{aligned}
$$

4.3   Kernel and Image

Definitions

- If $T : V \to W$ is a linear transformation, then the set of vectors in V that T maps into $\mathbf{0}$ is called the **kernel** of T. It is denoted by $ker(T)$. In mathematical notation:

$$ker(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}$$

- If $T : V \to W$ is a linear transformation, then the set of all vectors in W that are images under T of at least one vector in V is called the **Image** (or range in some texts) of T; it is denoted by $Im(T)$. In mathematical notation:

$$Im(T) = \{\mathbf{w} \in W \mid \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \in V\}$$

Examples

1. Let $I : V \to V$ be the identity operator. Since $I\mathbf{v} = \mathbf{v}$ for all vectors in V, every vector in V is the image of some vector (namely, itself); thus, $Im(I) = V$. Since the *only* vector that I maps into $\mathbf{0}$ is $\mathbf{0}$, it follows that $ker(I) = \{\mathbf{0}\}$.

2. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the orthogonal projection on the $x - y$ plane. The kernel of T is the set of points that T maps into $\mathbf{0} = (0, 0, 0)$; these are the points on the z-axis. Since T maps every point in $\mathbb{R}^3$ into the $x - y$ plane, the image of T must be some subset of this plane. But every point $(x_0, y_0, 0)$ in the $x - y$ plane is the image under T of some point; in fact, it is the image of all points on the vertical line that passes through $(x_0, y_0, 0)$. Thus $Im(T)$ is the entire $x - y$ plane.

3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator that rotates each vector in the $x-y$ plane through the angle θ. Since *every* vector in the $x - y$ plane can be obtaine by rotating some vector through the angle θ, one obtains $Im(T) = \mathbb{R}^2$. Moreover, the only vector that rotates into $\mathbf{0}$ is $\mathbf{0}$, so $ker(T) = \{\mathbf{0}\}$.

4. Find the kernel and image of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$T(\begin{bmatrix} x \\ y \\ z \end{bmatrix}) = \begin{bmatrix} x - y \\ 2z \end{bmatrix}$$

In all of the preceding examples, $ker(T)$ and $Im(T)$ turned out to be *subspaces*. This is no accident as the following theorem points out.

Theorem 4.4. *If $T : V \to W$ is a linear transformation, then:*

(a) *The kernel of T is a subspace of V.*

(b) *The range of T is a subspace of W.*

Proof. (a) To show that $ker(T)$ is a subspace, it must be shown that it contains at least one vector and is closed under addition and scalar multiplication. By part (a) of Theorem 4.1, the vector $\mathbf{0}$ is in $ker(T)$, so this set contains at least one vector. Let $\mathbf{v_1}$ and $\mathbf{v_2}$ be vectors in $ker(T)$, and let k be any scalar. Then

$$T(\mathbf{v_1} + \mathbf{v_2}) = T(\mathbf{v_1}) + T(\mathbf{v_2}) = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

so that $\mathbf{v_1}$ and $\mathbf{v_2}$ is in $ker(T)$. Also,

$$T(k\mathbf{v_1}) = kT(\mathbf{v_1}) = k\mathbf{0} = \mathbf{0}$$

so that $k\mathbf{v_1}$ is in $ker(T)$.

(b) Since $T(\mathbf{0}) = \mathbf{0}$, there is at least one vector in $Im(T)$. Let $\mathbf{w_1}$ and $\mathbf{w_2}$ be vectors in the range of T, and let k be any scalar. To prove this part it must be shown that $\mathbf{w_1} + \mathbf{w_2}$ and $k\mathbf{w_1}$ are in the range of T; that is, vectors $\mathbf{a}$ and $\mathbf{b}$ must be found in V such that $T(\mathbf{a}) = \mathbf{w_1} + \mathbf{w_2}$ and $T(\mathbf{b}) = k\mathbf{w_1}$.

Since $\mathbf{w_1}$ and $\mathbf{w_2}$ are in the range of T, there are vectors $\mathbf{a_1}$ and $\mathbf{a_2}$ in V such that $T(\mathbf{a_1}) = \mathbf{w_1}$ and $T(\mathbf{a_2}) = \mathbf{w_2}$. Let $\mathbf{a} = \mathbf{a_1} + \mathbf{a_2}$ and $\mathbf{b} = k\mathbf{a_1}$. Then

$$T(\mathbf{a}) = T(\mathbf{a_1} + \mathbf{a_2}) = T(\mathbf{a_1}) + T(\mathbf{a_2}) = \mathbf{w_1} + \mathbf{w_2}$$

46

and

$$T(\mathbf{b}) = T(k\mathbf{a_1}) = kT(\mathbf{a_1}) = k\mathbf{w_1}$$

which completes the proof.

$\square$

Theorem 4.5. *If $T : U \rightarrow V$ is a linear transformation and $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ forms a basis for U, then $Im(T) = span(T(\mathbf{u_1}), T(\mathbf{u_2}), \ldots, T(\mathbf{u_n}))$*

This theorem is best demonstrated by a simple example.

Example

Let A be $m \times n$ and let $T = T_A$. Then $T_A : \mathbb{F}^n \to \mathbb{F}^m$. Let $\{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}\}$ be the standard basis for $\mathbb{F}^n$. Then by the previous theorem it can be stated

$$
\begin{aligned}
Im(T_A) &= span(T_A(\mathbf{e_1}), T_A(\mathbf{e_2}), \ldots, T_A(\mathbf{e_n})) \\
&= span(A\mathbf{e_1}, A\mathbf{e_2}, \ldots, A\mathbf{e_n}) \\
&= span(col_1(A), col_2(A), \ldots, col_n(A))
\end{aligned}
$$

4.4 Rank and Nullity

Definitons If $T : U \to V$ is a linear transformation,

- the dimension of the image of T is called the **rank of** T and is denoted by rank(T),

- the dimension of the kernel is called the **nullity of** T and is denoted by nullity(T).

Example

- Let U be a vector space of dimension n, with basis $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$, and let $T : U \to U$ be a linear transformation defined by

$$T(\mathbf{u_1}) = \mathbf{u_2}, T(\mathbf{u_2}) = \mathbf{u_3}, \cdots, T(\mathbf{u_{n-1}}) = \mathbf{u_n} \text{ and } T(\mathbf{u_n}) = \mathbf{0}$$

Find bases for $ker(T)$ and $Im(T)$ and determine rank(T) and nullity(T).

Theorem 4.6. *If $T : U \to V$ is a linear transformation from an n-dimensional vector space U to a vector space V, then*

$$rank(T) + nullity(T) = dim(U) = n$$

Proof. The proof is divided up into two cases.

Case 1 Let U be the zero vector space. Then due to theorem 4.1 it is known that $T(\mathbf{0}) = \mathbf{0}$. Therefore it can be stated that

$$Im(T) = \{\mathbf{0}\} \text{ and } ker(T) = \{\mathbf{0}\}$$

therefore

$$\text{rank}(T) + \text{nullity}(T) = 0 + 0 = 0 = \dim(U)$$

Case 2 Let U be an n-dimensional vector space with the basis $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$. Then the proof can be divided into three parts.

(a) Consider the case where $ker(T) = \{\mathbf{0}\}$. Let $\mathbf{u} \in ker(T)$. As $\mathbf{u} \in U$ it can be expressed as

$$\mathbf{u} = x_1\mathbf{u_1} + x_2\mathbf{u_2} + \cdots + x_n\mathbf{u_n} \tag{1}$$

As $\mathbf{u} \in ker(T)$ it can be stated that

$$\mathbf{0} = T(\mathbf{u}) = x_1 T(\mathbf{u_1}) + x_2 T(\mathbf{u_2}) + \cdots + x_n T(\mathbf{u_n}) \tag{2}$$

Due to the fact that $ker(T) = \{\mathbf{0}\}$, $\mathbf{u} = \mathbf{0}$. Due to the linear independence of $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}$ it follows from equation (1) that $x_1, x_2, \ldots, x_n = 0$. It then also follows from equation (2) that $T(\mathbf{u_1}), T(\mathbf{u_2}), \ldots, T(\mathbf{u_n})$ are linearly independent. It is known from Theorem 4.5 that $Im(T) = \text{span}(T(\mathbf{u_1}), T(\mathbf{u_2}), \ldots, T(\mathbf{u_n}))$. As $T(\mathbf{u_1}), T(\mathbf{u_2}), \ldots, T(\mathbf{u_n})$ are linearly independent they form a basis for $Im(T)$. It can therefore be stated that

$$\text{rank}(T) + \text{nullity}(T) = n + 0 = n = \dim(U)$$

(b) Consider the case where $ker(T) = U$. Theorem 4.5 states: $Im(T) = \text{span}(T(\mathbf{u_1}), T(\mathbf{u_2}), \ldots, T(\mathbf{u_n}))$. However $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n} \in ker(T)$. Therefore $T(\mathbf{u_1}), T(\mathbf{u_2}), \ldots, T(\mathbf{u_n}) = \mathbf{0}$. So it can be stated that $Im(T) = \text{span}(\mathbf{0}) = \{\mathbf{0}\}$. Therefore

$$\text{rank}(T) + \text{nullity}(T) = 0 + n = n = \dim(U)$$

(c) Consider the case where $1 \leq \text{nullity}(T) < n$. Assume that the nullity$(T) = r$, and let $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_r}$ be a basis for the kernel. Since $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_r}\}$ form a linearly independent set, theorem 2.6(b) states that there are $n - r$ vectors, $\mathbf{u_{r+1}}, \mathbf{u_{r+2}}, \ldots, \mathbf{u_n}$, such that $\{\mathbf{u_1}, \ldots, \mathbf{u_r}, \mathbf{u_{r+1}}, \ldots, \mathbf{u_n}\}$ is a basis for U. To complete the proof it shall be shown that the $n - r$ vectors in the set $S = \{T(\mathbf{u_{r+1}}), \ldots, T(\mathbf{u_n})\}$ form a basis for the image of T. It then follows that

$$\text{rank}(T) + \text{nullity}(T) = n - r + r = n = \dim(U)$$

First it shall be shown that S spans the image of T. If $\mathbf{b}$ is any vector in $Im(T)$, then $\mathbf{b} = T(\mathbf{u})$ for some vector $\mathbf{u}$ in U. Since $\{\mathbf{u_1}, \ldots, \mathbf{u_r}, \mathbf{u_{r+1}}, \ldots, \mathbf{u_n}\}$ is a basis for U, the vector $\mathbf{u}$ can be written in the form

$$\mathbf{u} = c_1 \mathbf{u_1} + \cdots + c_r \mathbf{u_r} + c_{r+1} \mathbf{u_{r+1}} + \cdots + c_n \mathbf{u_n}$$

since $\mathbf{u_1}, \ldots, \mathbf{u_r}$ lie in the kernel of T, it is clear that $T(\mathbf{u_1}), \ldots, T(\mathbf{u_r}) = \mathbf{0}$, so that

$$\mathbf{b} = T(\mathbf{u}) = c_{r+1} T(\mathbf{u_{r+1}}) + \cdots + c_n T(\mathbf{u_n})$$

Thus, S spans the image of T.

Finally, it shall be shown that S is a linearly independent set and consequently forms a basis for $Im(T)$. Suppose that some linear combination of the vectors in S is zero; that is,

$$k_{r+1} T(\mathbf{u_{r+1}}) + \cdots + k_n T(\mathbf{u_n}) = \mathbf{0} \tag{3}$$

It must be shown that $k_{r+1} = \cdots = k_n = 0$. Since T is linear, equation (3) can be rewritten as

$$T(k_{r+1} \mathbf{u_{r+1}} + \cdots + k_n \mathbf{u_n}) = \mathbf{0}$$

50

which says that $k_{r+1}\mathbf{u_{r+1}} + \cdots + k_n\mathbf{u_n}$ is in the kernel of T. This vector can therefore be written as a linear combination of the basis vectors $\{\mathbf{u_1}, \ldots, \mathbf{u_r}\}$, say

$$k_{r+1}\mathbf{u_{r+1}} + \cdots + k_n\mathbf{u_n} = k_1\mathbf{u_1} + \cdots + k_r\mathbf{u_r}$$

Thus,

$$k_1\mathbf{u_1} + \cdots + k_r\mathbf{u_r} - k_{r+1}\mathbf{u_{r+1}} - \cdots - k_n\mathbf{u_n} = \mathbf{0}$$

Since $\{\mathbf{u_1}, \ldots, \mathbf{u_n}\}$ is linearly independent, all of the k's are zero; in particular $k_{r+1} = \cdots = k_n = 0$, which completes the proof.

$\square$

Examples Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator that rotates each vector in the $x - y$ plane through an angle of θ. It was shown previously that $ker(T) = \{\mathbf{0}\}$ and $Im(T) = \mathbb{R}^2$. Thus,

$$\text{rank}(T) + \text{nullity}(T) = 2 + 0 = 2 = \dim(U)$$

which is consistent with the fact that the domain of T is two-dimensional.

4.5   Matrix of a Linear Transformation

In this section it shall be shown that if U and V are finite-dimensional vector spaces, then with a little ingenuity any linear transformation $T : U \to V$ can be regarded as a matrix transformation. The basic idea is to work with coordinate matrices of the vectors rather than with the vectors themselves.

Definition

- Suppose that U is an n-dimensional vector space and V an m-dimensional vector space. Let $T : U \to V$ be a linear transformation. Let β and γ be bases for U and V respectively, then for each $\mathbf{x}$ in U, the coordinate vector $[\mathbf{x}]_\beta$ will be a

vector in $\mathbb{F}^n$, and the coordinate vector $[T(\mathbf{x})]_\gamma$ will be a vector in $\mathbb{F}^m$. If there exists an $m \times n$ matrix A, such that

$$A[\mathbf{x}]_\beta = [T(\mathbf{x})]_\gamma \tag{4}$$

then A is called **the matrix of the transformation relative to bases β and γ** and it is written

$$A = [T]_\beta^\gamma$$

Theorem 4.7. *Let $\beta = \{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ and $\gamma = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_m}\}$ be bases for the vector spaces U and V respectively, and let $\mathbf{x} \in U$. If $T : U \to V$ is a linear transformation then*

(a) *the matrix of the transformation relative to bases β and γ always exists. That is to say, there always exists a matrix $A = [T]_\beta^\gamma$ such that*

$$A[\mathbf{x}]_\beta = [T(\mathbf{x})]_\gamma$$

(b) *The matrix of the transformation relative to basis β and γ has the form*

$$[T]_\beta^\gamma = [[T(\mathbf{u_1})]_\gamma \mid [T(\mathbf{u_2})]_\gamma \mid \cdots \mid [T(\mathbf{u_n})]_\gamma]$$

Proof. Let $\beta = \{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ be a basis for the n-dimensional space U and let $\gamma = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_m}\}$ be a basis for the m-dimensional space V. Then the matrix $[T]_\beta^\gamma = A$ must have the form

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

such that (4) holds for all vectors $\mathbf{x}$ in U. In particular, this equation must hold for the basis vectors $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}$; that is,

$$A[\mathbf{u_1}]_\beta = [T(\mathbf{u_1})]_\gamma, A[\mathbf{u_2}]_\beta = [T(\mathbf{u_2})]_\gamma, \ldots, A[\mathbf{u_n}]_\beta = [T(\mathbf{u_n})]_\gamma \tag{5}$$

But

$$[\mathbf{u_1}]_\beta = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, [\mathbf{u_2}]_\beta = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, [\mathbf{u_n}]_\beta = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

so

$$A[\mathbf{u_1}]_\beta = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

$$A[\mathbf{u_2}]_\beta = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}$$

$$\vdots$$

$$A[\mathbf{u_n}]_\beta = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

Substituting these results into equation (5) yields

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} = [T(\mathbf{u_1})]_\gamma, \quad \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} = [T(\mathbf{u_2})]_\gamma, \ldots, \quad \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = [T(\mathbf{u_n})]_\gamma$$

which shows that the successive columns of A are coordinate vectors of

$$T(\mathbf{u_1}), T(\mathbf{u_2}), \ldots, T(\mathbf{u_n})$$

with respect to the basis γ. Thus the matrix for T with respect to the bases β and γ is

$$[T]_\beta^\gamma = [[T(\mathbf{u_1})]_\gamma \mid [T(\mathbf{u_2})]_\gamma \mid \cdots \mid [T(\mathbf{u_n})]_\gamma]$$

Thus the proof is complete. $\qquad\square$

Examples

1. Let $T_B : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation defined by $T(X) = BX$ where B is an $m \times n$ matrix. Let $\beta = \{\mathbf{E_1}, \mathbf{E_2}, \ldots, \mathbf{E_n}\}$ be the standard basis for $\mathbb{F}^n$ and let $\gamma = \{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_m}\}$ be the standard basis for $\mathbb{F}^m$. Then it is known from the previous theorem that $[T]_\beta^\gamma$ is the following matrix

   $$[T_B]_\beta^\gamma = [[T_B(\mathbf{E_1})]_\gamma \mid [T_B(\mathbf{E_2})]_\gamma \mid \cdots \mid [T_B(\mathbf{E_n})]_\gamma]$$

   In general, for $1 \le j \le n$ it follows from the definition of the transformation that

   $$T_B(\mathbf{E_j}) = B\mathbf{E_j} = col_j(B) = b_{1j}\mathbf{e_1} + b_{2j}\mathbf{e_2} + \cdots + b_{mj}\mathbf{e_m}$$

   therefore

   $$[T_B(\mathbf{E_j})]_\gamma = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj} \end{bmatrix} = col_j(B)$$

and thus it is clear that

$$[T_B]_\beta^\gamma = B$$

2. Let U have the basis $\beta = \{\mathbf{u_1}, \mathbf{u_2}, \mathbf{u_3}\}$ and let V have the basis $\gamma = \{\mathbf{v_1}, \mathbf{v_2}\}$. Let T be the linear transformation defined by

$$T(\mathbf{u_1}) = 2\mathbf{v_1} + \mathbf{v_2}, \; T(\mathbf{u_2}) = \mathbf{v_1} - \mathbf{v_2}, \; T(\mathbf{u_3}) = 2\mathbf{v_2}$$

Then clearly

$$[T]_\beta^\gamma = \begin{bmatrix} 2 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}$$

3. Let $V = M_{2\times 2}(\mathbb{R})$ and let $T : V \to V$ be the linear transformation given by $T(X) = BX - XB$ where $X \in V$ and

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

Let $\beta = \{E_{11}, E_{12}, E_{21}, E_{22}\}$ be the standard basis for V where

$$E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \; E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \; E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \; E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

To find $[T]_\beta^\beta$ it is necessary to do the following calculations:

$$T(E_{11}) = BE_{11} - E_{11}B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -b \\ c & 0 \end{bmatrix} = 0E_{11} + -bE_{12} + cE_{21} + 0E_{22}$$

$$T(E_{12}) = BE_{12} - E_{12}B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$= \begin{bmatrix} -c & a-d \\ 0 & c \end{bmatrix} = -cE_{11} + (a-d)E_{12} + 0E_{21} + cE_{22}$$

$$T(E_{21}) = BE_{21} - E_{21}B = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$= \begin{bmatrix} b & 0 \\ d-a & -b \end{bmatrix} = bE_{11} + 0E_{12} + (d-a)E_{21} + -bE_{22}$$

$$T(E_{22}) = BE_{22} - E_{22}B = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$= \begin{bmatrix} 0 & b \\ -c & 0 \end{bmatrix} = 0E_{11} + bE_{12} + -cE_{21} + 0E_{22}$$

Therefore it follows that

$$[T]_\beta^\beta = \begin{bmatrix} 0 & -c & b & 0 \\ -b & a-d & 0 & b \\ c & 0 & d-a & -c \\ 0 & c & -b & 0 \end{bmatrix}$$

The following theorem follows directly from the definition of the matrix of a linear transformation.

Theorem 4.8. *Let $T : U \to V$ be a linear transformation, and let β and γ be bases for U and V respectively. Then if $\mathbf{u} \in U$*

$$[T(\mathbf{u})]_\gamma = [T]_\beta^\gamma [\mathbf{u}]_\beta$$

Examples

1. Let U have the basis $\beta = \{\mathbf{u_1}, \mathbf{u_2}, \mathbf{u_3}\}$ and let V have the basis $\gamma = \{\mathbf{v_1}, \mathbf{v_2}\}$. Let T be the linear transformation defined by

$$T(\mathbf{u_1}) = 2\mathbf{v_1} + \mathbf{v_2}, \ T(\mathbf{u_2}) = \mathbf{v_1} - \mathbf{v_2}, \ T(\mathbf{u_3}) = 2\mathbf{v_2}$$

Given that $\mathbf{u} = 3\mathbf{u_1} + -2\mathbf{u_2} + 7\mathbf{u_3}$

$$[\mathbf{u}]_\beta = \begin{bmatrix} 3 \\ -2 \\ 7 \end{bmatrix}, \text{ and } [T]_\beta^\gamma = \begin{bmatrix} 2 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix}$$

Hence

$$[T(\mathbf{u})]_\gamma = [T]_\beta^\gamma [\mathbf{u}]_\beta = \begin{bmatrix} 2 & 1 & 0 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \\ 7 \end{bmatrix} = \begin{bmatrix} 4 \\ 19 \end{bmatrix}$$

Hence $T(\mathbf{u}) = 4\mathbf{v_1} + 19\mathbf{v_2}$.

Example

1. Let $T : V \to W$ be a linear transformation, and let $\beta = \{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ and $\gamma = \{\mathbf{w_1}, \mathbf{w_2}, \mathbf{w_3}\}$ be bases for V and W respectively. T is the linear transformation given by

$$\begin{aligned} T(\mathbf{v_1}) &= \mathbf{w_1} + \mathbf{w_2} + -\mathbf{w_3} \\ T(\mathbf{v_2}) &= 2\mathbf{w_1} + -3\mathbf{w_2} \\ T(\mathbf{v_3}) &= 3\mathbf{w_1} + -2\mathbf{w_2} + -\mathbf{w_3} \end{aligned}$$

Find $ker(T)$ and $Im(T)$.

The following theorem gives a recipe for finding bases for $ker(T)$ and the $Im(T)$ where possible.

Theorem 4.9. *Let A be an $m \times n$ matrix such that $A = [T]_\beta^\gamma$, where $T : V \to W$ is a linear transformation and $\beta = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ and $\gamma = \{\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_m}\}$ are bases for V and W respectively. Let $s = nullity(A)$ and $r = rank(A)$. Then suppose that*

$$\mathbf{x_j} = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{bmatrix}$$

for $1 \le j \le s$, form a basis for $N(A)$, while $col_{c_1}(A), col_{c_2}(A), \ldots, col_{c_t}(A)$ form a basis for $C(A)$ Then

1. (a) the vectors $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_s}$ defined by

$$\mathbf{u_j} = x_{1j}\mathbf{v_1} + x_{2j}\mathbf{v_2} + \cdots + x_{nj}\mathbf{v_n}$$

   will be a basis for the kernel of T.

   (b) the vectors $T(\mathbf{v_{c_1}}), T(\mathbf{v_{c_2}}), \ldots, T(\mathbf{v_{c_r}})$ form a basis for the image of T.

2. If $N(A) = \{\mathbf{0}\}$, then $ker(T) = \{\mathbf{0}\}$

   If $C(A) = \{\mathbf{0}\}$, then $Im(T) = \{\mathbf{0}\}$

3. Thus it follows that $rank(T) = rank(A)$ and $nullity(T) = nullity(A)$.

Quiz: True or false?

Let U and V be vector spaces of dimension n and m respectively over a field $\mathbb{F}$, and let T be a linear transformation from U to V. Let $\beta = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis for U and let $\gamma = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be a basis for V. Let $\mathbf{u} \in U$.

(a) For any $a_1, \ldots, a_n \in \mathbb{F}$,

$$T(\sum_{i=1}^{n} a_i \mathbf{u}_i) = \sum_{i=1}^{n} a_i T(\mathbf{u}_i).$$

(b) $\{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_n)\}$ is a basis for $\text{Im}(T)$.

(c) If $nullity(T) = 0$ then $m = n$.

(d) $rank(T) + nullity(T) = n$

(e) If α is a basis for $\ker(T)$ and $\alpha \subseteq \beta$, then $\beta \setminus \alpha$ is a basis for $\text{Im}(T)$.

(f) If $U = \mathbb{F}^n$, then $[\mathbf{u}]_\beta = \mathbf{u}$.

(g) $[\mathbf{u}_i]_\beta = \mathbf{e}_i$.

(h) $[\mathbf{u}]_\beta$ will be a column vector if and only if $U = \mathbb{F}^n$.

(i) $[\mathbf{u}]_\beta$ depends on the order of β.

(j) If $A = [T]_\beta^\gamma$, then $col_i A = [T(\mathbf{u}_i)]_\beta$.

(k) If $m = n$ and $\mathbf{u}_i = \mathbf{v}_i$ for all i, then $[T]_\beta^\gamma = I$.

(l) To get $[T]_\beta^\gamma$, we need to calculate $T(\mathbf{u}_i)$ for each i, and express the answer as a linear combination of vectors from γ.

4.6 T-invariant subspaces

Definition

Let $T : V \to V$ be a linear operator. A subspace W of V is called T-invariant if $T(w) \in W, \forall w \in W$.

Examples

The are many examples of T-invariant subspaces. Veriy that the following are all T-invariant:

- $\{0\}$

- V

- $ker(T)$

- $Im(T)$

- E_λ which is the space spanned by linearly independent eigenvectors of T corresponding to eigenvalue λ. (*)

- A T-cyclic subspace of V generated by $v \in V$, given by $span\{v, T(v), T^2(v), \ldots\}$

4.7 Vector space of linear transformations

If $f, g : V \to W$ are functions and V, W are vector spaces over $\mathbb{F}$ we have seen that we can define addition and scalar multiplication by $(f + g)(v) = f(v) + g(v)$ and $(af)(v) = af(v)$ with $a \in \mathbb{F}$ and $v \in V$.

Using the above definition, it is easily verified that if T_1, T_2 are linear transformations, then the linear combination $aT_1 + T_2$ is also a linear transformation. In fact, the set of all linear transformations from V to W is itself a vector space, denoted $\ell(V, W)$. In the case $V = W$ we often write $\ell(V)$.

In fact we have the relationships

$$[T_1 + T_2]_\beta^\gamma = [T_1]_\beta^\gamma + [T_2]_\beta^\gamma,$$

$$[aT]_\beta^\gamma = a[T]_\beta^\gamma.$$

This is leading up to the notion of associating the vector space $\ell(V, W)$ with $M_{m \times n}$ in the case V and W are of dimension n and m respectively. Before making this identification we should investigate the concept of isomorphic vector spaces.

4.8   Isomorphisms and inverses of transformations

A linear transformation $T : V \to W$ is said to be invertible if there exists a unique transformation $T^{-1} : W \to V$ such that $T \circ T^{-1} = I_W$ and $T^{-1} \circ T = I_V$. We call T^{-1} the inverse of T.

We have the following facts regarding inverses:

- A function is invertible if and only if it is one-to-one and onto.

- T^{-1} is linear.

- T is invertible if and only if $rank(T) = \dim(V) = \dim(W)$.

- $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

We say that V is **isomorphic** to W if there exists an invertible linear transformation $T : V \to W$. We write $V \cong W$ to indicate that V is isomorphic to W. Such a T is called an isomorphism.

The main result of this section is the following:

If V and W are finite dimensional vector spaces over the same field, then $V \cong W$ if and only if $\dim(V) = \dim(W)$.

Examples

(a) $T : \mathbb{F}^2 \to P_2(\mathbb{F})$ via $T(a, b) = a + bx$.

(b) $T : P_3(\mathbb{F}) \to M_2(\mathbb{F})$ via $T(a + bx + cx^2 + dx^3) = \begin{pmatrix} a + b & b + c \\ c + d & d \end{pmatrix}$.

A corollary to our result is that for a vector space V over $\mathbb{F}$, V is isomorphic to $\mathbb{F}^n$ if and only if $\dim(V) = n$.

This formalises our association of n dimensional vector spaces with $\mathbb{F}^n$ as I hinted at when we looked at standard bases.

Another consequence is that we can associate $\ell(V, W)$ with $M_{m \times n}$.

4.9   Change of Basis

A basis of a vector space is a set of vectors that specify the coordinate system. A vector space may have an infinite number of bases but each basis contains the same number of vectors. The number of vectors in the basis is called the dimension of the vector space. The coordinate vector or coordinate matrix of a point changes with any change in the basis used. If the basis for a vector space is changed from some old bases β to some new bases γ, how is the old coordinate vector $[\mathbf{v}]_\beta$ of a vector $\mathbf{v}$ related to the new coordinate vector $[\mathbf{v}]_\gamma$? The following theorem answers that question.

Theorem 4.10. *If the basis for a vector space is changed from some old basis $\beta = \{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ to some new basis $\gamma = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$, then the old coordinate vector $[\mathbf{w}]_\beta$ is related to the new coordinate vector $[\mathbf{w}]_\gamma$ of the same vector $\mathbf{w}$ by the equation*

$$[\mathbf{w}]_\gamma = P[\mathbf{w}]_\beta$$

where the columns of P are the coordinate vectors of the old basis vectors relative to the new basis; that is, the column vectors of P are

$$[\mathbf{u_1}]_\gamma, [\mathbf{u_2}]_\gamma, \ldots, [\mathbf{u_n}]_\gamma$$

P is called the **change of basis matrix** *or the* **change of coordinate matrix**.

Proof. Let V be a vector space with a basis $\beta = \{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ and a new basis

$\gamma = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$. Let $\mathbf{w} \in V$. Therefore $\mathbf{w}$ can be expressed as:

$$\mathbf{w} = a_1\mathbf{u_1} + a_2\mathbf{u_2} + \cdots + a_n\mathbf{u_n}$$

Thus we have

$$[\mathbf{w}]_\beta = (a_1, a_2, \ldots, a_n)$$

As γ is also a basis of V the elements of β can be expressed as follows

$$\mathbf{u_1} = p_{11}\mathbf{v_1} + p_{21}\mathbf{v_2} + \cdots + p_{n1}\mathbf{v_n}$$

$$\mathbf{u_2} = p_{12}\mathbf{v_1} + p_{22}\mathbf{v_2} + \cdots + p_{n2}\mathbf{v_n}$$

$$\vdots$$

$$\mathbf{u_n} = p_{1n}\mathbf{v_1} + p_{2n}\mathbf{v_2} + \cdots + p_{nn}\mathbf{v_n}$$

Combining this system of equations with the above expression for $\mathbf{w}$ gives

$$\mathbf{w} = (p_{11}a_1 + p_{12}a_2 + \cdots + p_{1n}a_n)\mathbf{v_1}$$

$$+(p_{21}a_1 + p_{22}a_2 + \cdots + p_{2n}a_n)\mathbf{v_2}+$$

$$\vdots$$

$$+(p_{n1}a_1 + p_{n2}a_2 + \cdots + p_{nn}a_n)\mathbf{v_n}+$$

and thus it can be seen that

$$[\mathbf{w}]_\gamma = \begin{bmatrix} p_{11}a_1 + p_{12}a_2 + \cdots + p_{1n}a_n \\ p_{21}a_1 + p_{22}a_2 + \cdots + p_{2n}a_n \\ \vdots \\ p_{n1}a_1 + p_{n2}a_2 + \cdots + p_{nn}a_n \end{bmatrix}$$

which can be written as

$$[\mathbf{w}]_\gamma = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

from which it can be seen

$$[\mathbf{w}]_\gamma = P[\mathbf{w}]_\beta$$

where P's columns are

$$[\mathbf{u_1}]_\gamma, [\mathbf{u_2}]_\gamma, \ldots, [\mathbf{u_n}]_\gamma$$

□

Example

1. Consider the bases $\gamma = \{\mathbf{v_1}, \mathbf{v_2}\}$ and $\beta = \{\mathbf{u_1}, \mathbf{u_2}\}$ for $\mathbb{R}^2$, where

$$\mathbf{v_1} = (1, 0); \quad \mathbf{v_2} = (0, 1); \quad \mathbf{u_1} = (1, 1); \quad \mathbf{u_2} = (2, 1)$$

(a) Find the transition matrix from β to γ. First the coordinate vectors of the old basis vectors $\mathbf{u_1}$ and $\mathbf{u_2}$ must be found relative to the new basis β. By inspection:

$$\mathbf{u_1} = \mathbf{v_1} + \mathbf{v_2}$$

$$\mathbf{u_2} = 2\mathbf{v_1} + \mathbf{v_2}$$

so that

$$[\mathbf{u_1}]_\gamma = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad and \quad [\mathbf{u_2}]_\gamma = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Thus the transition matrix from β to γ

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$$

(b) Use the transition matrix to find $[\mathbf{v}]_\gamma$ if

$$[\mathbf{v}]_\beta = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$$

It is known from the above change of basis theorem 4.10 that

$$[\mathbf{v}]_\gamma = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 7 \\ 2 \end{bmatrix}$$

As a check it should be possible to recover the vector $\mathbf{v}$ either from $[\mathbf{v}]_\beta$ or $[\mathbf{v}]_\gamma$. It is left for the student to show that $-3\mathbf{u_1}+5\mathbf{u_2} = 7\mathbf{v_1}+2\mathbf{v_2} = (7, 2)$.

4.10 Similar Matrices

The matrix of a linear operator $T : V \to V$ depends on the basis selected for V. One of the fundamental problems of linear algebra is to choose a basis for V that makes the matrix for T as simple as possible - diagonal or triangular, for example. This section is devoted to the study of this problem

To demonstrate that certain bases produce a much simpler matrix of transformation than others, consider the following example.

Example

1. Standard bases do not necessarily produce the simplest matrices for linear operators. For example, consider the linear operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ -2x_1 + 4x_2 \end{bmatrix}$$

and the standard basis $\beta = \{\mathbf{e_1}, \mathbf{e_2}\}$ for $\mathbb{R}^2$, where

$$\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

By theorem 4.7, the matrix for T with respect to this basis is the standard matrix for T; that is,

$$[T]_\beta^\beta = [T(\mathbf{e_1}) \mid T(\mathbf{e_2})]$$

From the definition of the linear transformation T,

$$T(\mathbf{e_1}) = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad T(\mathbf{e_2}) = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

so

$$[T]_\beta^\beta = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$$

In comparison, consider the basis $\gamma = \{\mathbf{u_1}, \mathbf{u_2}\}$, where

$$\mathbf{u_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{u_2} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

By theorem 4.7, the matrix for T with respect to the basis γ is

$$[T]_\gamma^\gamma = [[T(\mathbf{u_1})]_\gamma \mid [T(\mathbf{u_2})]_\gamma$$

From the definiton of the linear transformation T,

$$T(\mathbf{u_1}) = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2\mathbf{u_1}, \quad T(\mathbf{u_2}) = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3\mathbf{u_2}$$

Hence

$$[T(\mathbf{u_1})]_\gamma = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad [T(\mathbf{e_2})]_\gamma = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

So

$$[T]_\gamma^\gamma = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

This matrix is 'simpler' in the sense that diagonal matrices enjoy special properties that more general matrices do not.

Much research has been devoted to determining the "simplest possible form" that can be obtained for the matrix of a linear operator $T : V \to V$, by choosing the basis appropriately. This problem can be attacked by first finding a matrix for T relative to *any* basis, say a standard basis, where applicable, then changing the basis in a manner that simplifies the matrix. Before pursuing this idea further, it is necessary to grasp the theorem below. It gives a useful alternative viewpoint about change of basis matrices; it shows that the transition matrix form a basis β to γ can be regarded as the matrix of transformation of the identity operator.

Theorem 4.11. *If β and γ are bases for a finite-dimensional vector space V, and if $I : V \to V$ is the identity operator, then $[I]_\beta^\gamma$ is the transition matrix form β to γ.*

Proof. Suppose that $\beta = \{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ and $\gamma = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ are bases for V. Using the fact that $I(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in V$, it follows that

$$
\begin{aligned}
[I]_\beta^\gamma &= [[I(\mathbf{u_1})]_\gamma \mid [I(\mathbf{u_2})]_\gamma \mid [I(\mathbf{u_n})]_\gamma] \\
&= [[\mathbf{u_1}]_\gamma \mid [\mathbf{u_2}]_\gamma \mid [\mathbf{u_n}]_\gamma]
\end{aligned}
$$

which is the change of basis matrix from β to γ. $\qquad\square$

The ground work has been laid to consider the main problem in this section.

Problem: If β and γ are two bases for a finite-dimensional vector space V, and if $T : V \to V$ is a linear operator, what relationship, if any, exists between the matrices $[T]_\beta^\beta$ and $[T]_\gamma^\gamma$?

The answer to this question can be obtained by considering the composition of three linear operators. Consider a vector $\mathbf{v} \in V$. Let the vector $\mathbf{v}$ be mapped into itself by the identity operator, then let $\mathbf{v}$ be mapped into $T(\mathbf{v})$ by T, then let $T(\mathbf{v})$ be mapped into itself by the identity operator. All four vector spaces involved in the composition are the same (namely V); however, the bases for the spaces vary. Since the starting vector is $\mathbf{v}$ and the final vector is $T(\mathbf{v})$, the composition is the same as T; that is,

$$T = I \circ T \circ I$$

If the first and last vector spaces are assigned the basis γ and the middle two spaces are assigned the basis β, then it follows from the previous statement $T = I \circ T \circ I$, that

$$[T]_\gamma^\gamma = [I \circ T \circ I]_\gamma^\gamma = [I]_\beta^\gamma [T]_\beta^\beta [I]_\gamma^\beta$$

But $[I]_\gamma^\beta$ is the change of basis matrix from γ to β and consequently $[I]_\beta^\gamma$ is the change of basis matrix from β to γ. Thus, let $P = [I]_\gamma^\beta$, then $P^{-1} = [I]_\beta^\gamma$ and hence it can be written that

$$[T]_\gamma^\gamma = P^{-1}[T]_\beta^\beta P$$

This is all summarised in the following theorem.

Theorem 4.12. *Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V, and let β and γ be bases for V. Then*

$$[T]_\gamma^\gamma = P^{-1}[T]_\beta^\beta P \tag{6}$$

where P is the change of basis matrix from γ to β.

Remark: When applying theorem 4.12, it is easy to forget whether P is the change of basis matrix from β to γ or the change of basis matrix from γ to β. Just remember that in order for $[T]_\beta^\beta$ to operate succesfully on a vector $\mathbf{v}$, $\mathbf{v}$ must be in the basis β. Therefore, due to P's positioning in the formula, it must be the change of basis matrix from γ to β.

Example

1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ -2x_1 + 4x_2 \end{bmatrix}$$

Find the matrix of T with respect to the standard basis $\beta = \{\mathbf{e_1}, \mathbf{e_2}\}$ for $\mathbb{R}^2$, then use theorem 4.12 to find the matrix of T with respect to the basis $\gamma = \{\mathbf{u_1}, \mathbf{u_2}\}$, where

$$\mathbf{u_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{u_2} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

It was shown earlier that

$$[T]_\beta^\beta = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$$

To find $[T]_\gamma^\gamma$ from (6), requires the change of basis matrix P, where

$$P = [I]_\gamma^\beta = [[\mathbf{u_1}]_\beta \mid [\mathbf{u_2}]_\beta]$$

By inspection

$$\mathbf{u_1} = \mathbf{e_1} + \mathbf{e_2}$$
$$\mathbf{u_2} = \mathbf{e_1} + 2\mathbf{e_2}$$

so that

$$[\mathbf{u_1}]_\beta = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad [\mathbf{u_2}]_\beta = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Thus the transition matrix from γ to β is

$$P = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

It is clear that

$$P^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

so that by theorem 4.12 the matrix of T relative to the basis γ is

$$[T]_\gamma^\gamma = P^{-1}[T]_\beta^\beta P = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

which agrees with the previous result.

The relationship in (6) is of such importance that there is some terminology associated with it.

Definition

- If A and B are square matrices, it is said that B **is similar to** A if there is an invertible matrix P such that $B = P^{-1}AP$.