Further Mathematical Methods (Linear Algebra) 2002

Lecture 3: Linear Transformations

In this hand-out we are going to look at linear transformations: what they are, what properties they have, and how they can be represented by a matrix. We shall also investigate how bases are actually used and see ways of 'changing' the basis we want to work with. The Learning Objectives associated with this hand-out are given at the end.

3.1 What Are Linear Transformations?

Consider two vector spaces V and W, a *transformation* (or *mapping*) from V to W is a function, $T : V \to W$, which takes vectors $\mathbf{v} \in V$ and gives us corresponding vectors $T(\mathbf{v}) \in W$. In this hand-out we will be concerned with a special class of transformations, namely those which are *linear* — that is, where the transformation of a linear combination of vectors in V will give us a linear combination of transformed vectors in W. Or, in symbols, transformations, $T : V \to W$, where given vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$,

$$T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1 T(\mathbf{v}_1) + \alpha_2 T(\mathbf{v}_2)$$

where $T(\mathbf{v}_1), T(\mathbf{v}_2) \in W$. So, formally, we have:

Definition 3.1 *Let V and W be vector spaces. The function, $T : V \to W$, is a linear transformation from V to W if, for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ and all scalars α_1, α_2,*

$$T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1 T(\mathbf{v}_1) + \alpha_2 T(\mathbf{v}_2)$$

Further, the vector spaces V and W are called the domain and co-domain of T respectively.

The action of a linear transformation on a vector $\mathbf{v} \in V$ is illustrated in Figure 3.1.

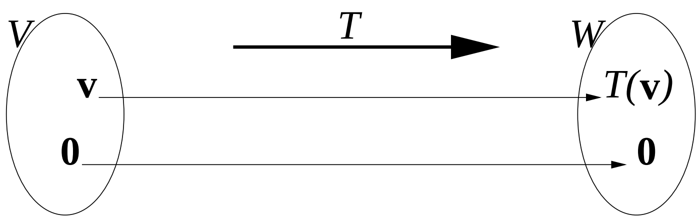

Figure 3.1: *A schematic representation of the linear transformation $T : V \to W$. Every point in the ellipses labelled V and W is supposed to represent a vector in the vector spaces V and W respectively. The vector $\mathbf{v} \in V$ is 'transformed' to the vector $T(\mathbf{v}) \in W$ under the action of the linear transformation $T : V \to W$. (Notice that under such transformations, the null vector in V always maps to the null vector in W — see Theorem 3.3.)*

Now, although the definition only stipulates how a linear combination of two vectors in V is mapped to W under a linear transformation, we should expect that a more general linear combination should be transformed in much the same way. To this end, we note that

Theorem 3.2 *Let V and W be vector spaces. If $T : V \to W$ is a linear transformation, then*

$$T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_m\mathbf{v}_m) = \alpha_1 T(\mathbf{v}_1) + \alpha_2 T(\mathbf{v}_2) + \cdots + \alpha_m T(\mathbf{v}_m)$$

for vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in V$ and scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$. This represents the fact that linear transformations preserve linear combinations.

Proof: Let V and W be vector spaces, and consider the linear transformation from V to W. Clearly, as the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are in the vector space V and $\alpha_1, \alpha_2, \ldots, \alpha_m$ are scalars, the linear combination $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_m\mathbf{v}_m$ is in V too. Consequently, to prove the Theorem, all we have to do is apply Definition 3.1 to the vector represented by this linear combination. That is, using the definition we see that:

$$\begin{aligned} T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 + \cdots + \alpha_m\mathbf{v}_m) &= \alpha_1 T(\mathbf{v}_1) + T(\alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 + \cdots + \alpha_m\mathbf{v}_m) \\ &= \alpha_1 T(\mathbf{v}_1) + \alpha_2 T(\mathbf{v}_2) + T(\alpha_3\mathbf{v}_3 + \cdots + \alpha_m\mathbf{v}_m) \end{aligned}$$

and on the $(m-1)$th application of the definition we find that

$$T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 + \cdots + \alpha_m\mathbf{v}_m) = \alpha_1 T(\mathbf{v}_1) + \alpha_2 T(\mathbf{v}_2) + \alpha_3 T(\mathbf{v}_3) + \cdots + \alpha_m T(\mathbf{v}_m)$$

which is a vector in W. ♠

Indeed, several useful consequences of Definition 3.1 are given in the following theorem:

Theorem 3.3 *Let V and W be vector spaces. If $T : V \to W$ is a linear transformation, then*

1. $T(\mathbf{0}) = \mathbf{0}$.
2. $T(-\mathbf{v}) = -T(\mathbf{v})$.
3. $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$.

for all vectors $\mathbf{u}, \mathbf{v} \in V$.

The proof of this theorem will be discussed in Problem Sheet 2.

Lastly, we consider what happens when you find the *composition* of two linear transformations, namely:

Definition 3.4 *Let U, V and W be vector spaces. If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, then the composition of T_2 with T_1, denoted by $T_2 \circ T_1 : U \to W$, is the function given by the formula*

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

for any vector $\mathbf{u} \in U$.

It should be clear that in order for such a composition to make sense, we require that for any $\mathbf{u} \in U$, the vector $T_1(\mathbf{u}) \in V$ must lie in the domain of T_2. (Otherwise, there would be vectors $\mathbf{u} \in U$ that could not be mapped to the vector space W by the composition $(T_2 \circ T_1)(\mathbf{u})$ contrary to Definition 3.4.) The action of a composition of two linear transformations on a vector $\mathbf{u} \in U$ is illustrated in Figure 3.2. We also note that, as may be expected, the composition of two linear

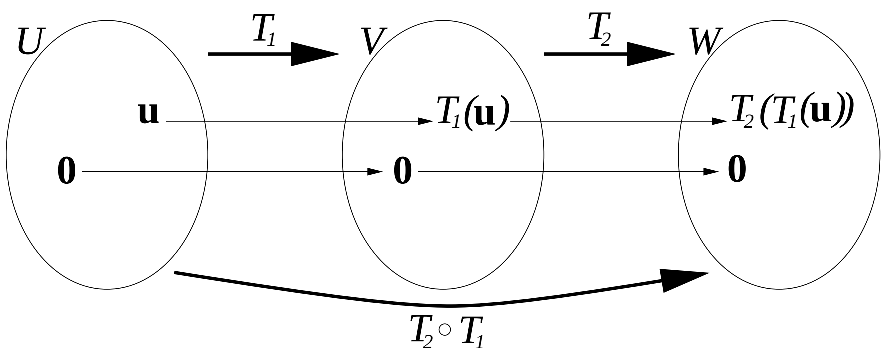

Figure 3.2: *A schematic representation of the composition, $T_2 \circ T_1 : U \to W$, of two linear transformations $T_1 : U \to V$ and $T_2 : V \to W$. The vector $\mathbf{u} \in U$ is 'transformed' to the vector $T_1(\mathbf{u}) \in V$ and this in turn is 'transformed' to the vector $T_2(T_1(\mathbf{u})) \in W$ under the action of the composition $T_2 \circ T_1$. Alternatively, using $(T_2 \circ T_1)(\mathbf{u})$ we 'transform' $\mathbf{u} \in U$ to $T_2(T_1(\mathbf{u})) \in W$ directly. (Notice that under all of these transformations, the null vector in the domain always maps to the null vector in the co-domain — see Theorem 3.3.)*

transformations is itself a linear transformation, i.e.

Theorem 3.5 *If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, then $(T_2 \circ T_1) : U \to W$ is also a linear transformation.*

Proof: Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be two general vectors in U and let α_1 and α_2 be two general scalars. By Definition 3.4, this means that

$$\begin{aligned}(T_2 \circ T_1)(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2) &= T_2(T_1(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2))\\ &= T_2(\alpha_1 T_1(\mathbf{u}_1) + \alpha_2 T_1(\mathbf{u}_1))\\ &= \alpha_1 T_2(T_1(\mathbf{u}_1)) + \alpha_2 T_2(T_1(\mathbf{u}_1))\\ \therefore (T_2 \circ T_1)(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2) &= \alpha_1(T_2 \circ T_1)(\mathbf{u}_1) + \alpha_2(T_2 \circ T_1)(\mathbf{u}_1)\end{aligned}$$

and so, by Definition 3.1, $T_2 \circ T_1$ is a linear transformation. ♠

Let us now look at some examples.

Example: Which of the following three transformations are linear?

1. $T_1 : V \to V$ given by $T_1(\mathbf{v}) = \mathbf{0}$.
2. $T_2 : V \to V$ given by $T_2(\mathbf{v}) = k\mathbf{v}$ for some non-zero fixed scalar k.
3. $T_3 : V \to V$ given by $T_3(\mathbf{v}) = \mathbf{v} + \mathbf{v}_0$ for some fixed non-zero vector $\mathbf{v}_0 \in V$.

Further, find the compositions $T_1 \circ T_2$ and $T_1 \circ T_2$.

Solution: To see whether these transformations are linear, we have to see whether they satisfy Definition 3.1, i.e. does

$$T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1 T(\mathbf{v}_1) + \alpha_2 T(\mathbf{v}_2)$$

for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ and all scalars α_1, α_2?

(1) Consider two *general* vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ and two *general* scalars α_1, α_2. We observe that for the transformation $T_1 : V \to V$ given by $T_1(\mathbf{v}) = \mathbf{0}$, we have

$$T_1(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \mathbf{0}, \;\; T_1(\mathbf{v}_1) = \mathbf{0} \;\text{ and }\; T_1(\mathbf{v}_2) = \mathbf{0}$$

as the vectors $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2$, $\mathbf{v}_1$ and $\mathbf{v}_2$ are all in V. Consequently, we can see that

$$T_1(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \mathbf{0} \;\text{ and }\; \alpha_1 T_1(\mathbf{v}_1) + \alpha_2 T_1(\mathbf{v}_2) = \alpha_1\mathbf{0} + \alpha_2\mathbf{0} = \mathbf{0}$$

These two expressions are equal for the *general* vectors and scalars considered and so $T_1(\mathbf{v})$ is a linear transformation. (Note that this linear transformation is sometimes referred to as the *null* (or *zero*) *transformation*.)

(2) Consider two *general* vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ and two *general* scalars α_1, α_2. We observe that for the transformation $T_2 : V \to V$ given by $T_2(\mathbf{v}) = k\mathbf{v}$ where k is some non-zero fixed scalar, we have

$$T_2(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = k(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2), \;\; T_2(\mathbf{v}_1) = k\mathbf{v}_1 \;\text{ and }\; T_2(\mathbf{v}_2) = k\mathbf{v}_2$$

as the vectors $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2$, $\mathbf{v}_1$ and $\mathbf{v}_2$ are all in V. Consequently, we can see that

$$T_2(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = k(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1 k\mathbf{v}_1 + \alpha_2 k\mathbf{v}_2 = \alpha_1 T_2(\mathbf{v}_1) + \alpha_2 T_2(\mathbf{v}_2)$$

and so, $T_2(\mathbf{v})$ is a linear transformation. (Note that this linear transformation is sometimes referred to as a *contraction of V with factor k* if $0 < k < 1$ (or a *dilation of V with factor k* if $k > 1$) as it 'compresses' (or 'stretches') each vector in V by a factor of k.)

(3) Consider two *general* vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ and two *general* scalars α_1, α_2. We observe that for the transformation $T_3 : V \to V$ given by $T_3(\mathbf{v}) = \mathbf{v} + \mathbf{v}_0$ where $\mathbf{v}_0 \in V$ is a fixed non-zero vector, we have

$$T_3(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \mathbf{v}_0, \;\; T_3(\mathbf{v}_1) = \mathbf{v}_1 + \mathbf{v}_0 \;\text{ and }\; T_3(\mathbf{v}_2) = \mathbf{v}_2 + \mathbf{v}_0$$

as the vectors $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2$, $\mathbf{v}_1$ and $\mathbf{v}_2$ are all in V. Consequently, we can see that as

$$\alpha_1 T_2(\mathbf{v}_1) + \alpha_2 T_2(\mathbf{v}_2) = \alpha_1(\mathbf{v}_1 + \mathbf{v}_0) + \alpha_2(\mathbf{v}_2 + \mathbf{v}_0) = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + (\alpha_1 + \alpha_2)\mathbf{v}_0$$

is not equal to

$$T_3(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \mathbf{v}_0$$

for the *general* vectors and scalars considered, $T_3(\mathbf{v})$ is *not* a linear transformation.[1] (Note that this transformation, although not linear, is very useful in linear algebra and we shall look at why it is important later.)

Lastly, we note that the compositions $T_1 \circ T_2 : V \to V$ and $T_2 \circ T_1 : V \to V$ are given by

$$T_1 \circ T_2(\mathbf{v}) = T_1(T_2(\mathbf{v})) = T_1(k\mathbf{v}) = \mathbf{0}$$

and

$$T_2 \circ T_1(\mathbf{v}) = T_2(T_1(\mathbf{v})) = T_2(\mathbf{0}) = k\mathbf{0} = \mathbf{0}$$

for any vector $\mathbf{v} \in V$ and some fixed non-zero scalar k. (Notice that both compositions map every vector $\mathbf{v} \in V$ to the null vector $\mathbf{0}$ in this case.) ♣

You should also note that within this general framework, we can look at linear transformations which map between real function spaces too. We now turn to an illustration of this.

Example: Recall that $\mathbb{P}_n^{\mathbb{R}}$ is the subspace of $\mathbb{F}^{\mathbb{R}}$ that contains all polynomials of degree at most n (where n is a non-negative integer).[2] That is, every vector $\mathbf{p} \in \mathbb{P}_n^{\mathbb{R}}$ is of the form

$$\mathbf{p} = a_0 \cdot \mathbf{1} + a_1 \cdot \mathbf{x} + \cdots + a_n \cdot \mathbf{x^n}$$

where $\mathbf{p} : x \to p(x)$ for all $x \in \mathbb{R}$ and

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

for all $x \in \mathbb{R}$. Which of the following three transformations are linear?

1. $T_1 : \mathbb{P}_2^{\mathbb{R}} \to \mathbb{P}_2^{\mathbb{R}}$ given by $T_1(\mathbf{p}) : x \to p(x+k)$ for all $x \in \mathbb{R}$ and some fixed real number k.
2. $T_2 : \mathbb{P}_2^{\mathbb{R}} \to \mathbb{P}_{2+m}^{\mathbb{R}}$ given by $T_2(\mathbf{p}) : x \to x^m p(x)$ for all $x \in \mathbb{R}$ and some fixed integer $m \geq 0$.
3. $T_3 : \mathbb{P}_2^{\mathbb{R}} \to \mathbb{P}_2^{\mathbb{R}}$ given by $T_3(\mathbf{p}) : x \to p(x) + k$ for all $x \in \mathbb{R}$ and some fixed non-zero scalar k.

Further, find the compositions $T_1 \circ T_2$ and $T_1 \circ T_2$.

Solution: As in the previous example, to see whether these transformations are linear, we have to see whether they satisfy Definition 3.1, i.e. does

$$T(\alpha_1\mathbf{p}_1 + \alpha_2\mathbf{p}_2) = \alpha_1 T(\mathbf{p}_1) + \alpha_2 T(\mathbf{p}_2)$$

for all vectors $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{P}_2^{\mathbb{R}}$ and all scalars α_1, α_2?

(1) Consider two *general* vectors $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{P}_2^{\mathbb{R}}$ and two *general* scalars α_1, α_2. We observe that for the transformation $T_1 : \mathbb{P}_2^{\mathbb{R}} \to \mathbb{P}_2^{\mathbb{R}}$ given by $T_1(\mathbf{p}) : x \to p(x+k)$ for all $x \in \mathbb{R}$ and some fixed real number k, we have

$$T_1(\alpha_1\mathbf{p}_1 + \alpha_2\mathbf{p}_2) : x \to \alpha_1 p_1(x+k) + \alpha_2 p_2(x+k),\ \forall x \in \mathbb{R}$$

because the vector $\alpha_1\mathbf{p}_1 + \alpha_2\mathbf{p}_2 \in \mathbb{P}_2^{\mathbb{R}}$ is defined using *point-wise* operations and so $[\alpha_1\mathbf{p}_1 + \alpha_2\mathbf{p}_2](x) = \alpha_1 p_1(x) + \alpha_2 p_2(x)$ for all $x \in \mathbb{R}$.[3] Similarly, we have:

$$T_1(\mathbf{p}_1) : x \to p_1(x+k) \text{ and } T_1(\mathbf{p}_2) : x \to p_2(x+k)$$

[1] But, this should be obvious as $T_3(\mathbf{0}) = \mathbf{0} + \mathbf{v}_0 = \mathbf{v}_0 \neq \mathbf{0}$ contrary to Theorem 3.3(1).

[2] This subspace of $\mathbb{F}^{\mathbb{R}}$ was first introduced in the Example on p. 19 of the hand-out for Lecture 2.

[3] See Section 1.3.3.

for all $x \in \mathbb{R}$. Consequently, we can see that as

$$\alpha_1 T_1(\mathbf{p}_1) + \alpha_2 T_1(\mathbf{p}_2) : x \to \alpha_1 p_1(x+k) + \alpha_2 p_2(x+k)$$

for all $x \in \mathbb{R}$, it is the case that

$$T_1(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2) = \alpha_1 T_1(\mathbf{p}_1) + \alpha_2 T_1(\mathbf{p}_2)$$

and so as these expressions are equal for the *general* vectors and scalars considered, $T_1(\mathbf{p})$ is a linear transformation. (Note that when this linear transformation is applied to a [quadratic] polynomial $p(x)$, it 'shifts' it in the negative x-direction by k units.)

(2) Consider two *general* vectors $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{P}_2^{\mathbb{R}}$ and two *general* scalars α_1, α_2. We observe that for the transformation $T_2 : \mathbb{P}_2^{\mathbb{R}} \to \mathbb{P}_{m+2}^{\mathbb{R}}$ given by $T_2(\mathbf{p}) : x \to x^m p(x)$ for all $x \in \mathbb{R}$ and some fixed integer $m \geq 0$, we have

$$T_2(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2) : x \to \alpha_1 x^m p_1(x) + \alpha_2 x^m p_2(x),\ \forall x \in \mathbb{R}$$

because the vector $\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 \in \mathbb{P}_2^{\mathbb{R}}$ is defined using *point-wise* operations and so $[\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2](x) = \alpha_1 p_1(x) + \alpha_2 p_2(x)$ for all $x \in \mathbb{R}$. Similarly, we have:

$$T_2(\mathbf{p}_1) : x \to x^m p_1(x) \text{ and } T_2(\mathbf{p}_2) : x \to x^m p_2(x)$$

for all $x \in \mathbb{R}$. Consequently, we can see that as

$$\alpha_1 T_2(\mathbf{p}_1) + \alpha_2 T_2(\mathbf{p}_2) : x \to \alpha_1 x^m p_1(x) + \alpha_2 x^m p_2(x)$$

for all $x \in \mathbb{R}$, it is the case that

$$T_2(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2) = \alpha_1 T_1(\mathbf{p}_1) + \alpha_2 T_1(\mathbf{p}_2)$$

and so as these expressions are equal for the *general* vectors and scalars considered, $T_2(\mathbf{p})$ is a linear transformation. (Note that when this linear transformation is applied to a [quadratic] polynomial $p(x)$, it turns it into another polynomial $x^m p(x)$ [which is of degree $2+m$].)

(3) Consider two *general* vectors $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{P}_2^{\mathbb{R}}$ and two *general* scalars α_1, α_2. We observe that for the transformation $T_3 : \mathbb{P}_2^{\mathbb{R}} \to \mathbb{P}_2^{\mathbb{R}}$ given by $T_3(\mathbf{p}) : x \to p(x) + k$ for all $x \in \mathbb{R}$ and some fixed non-zero scalar k, we have

$$T_3(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2) : x \to \alpha_1 p_1(x) + \alpha_2 p_2(x) + k,\ \forall x \in \mathbb{R}$$

because the vector $\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 \in \mathbb{P}_2^{\mathbb{R}}$ is defined using *point-wise* operations and so $[\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2](x) = \alpha_1 p_1(x) + \alpha_2 p_2(x)$ for all $x \in \mathbb{R}$. Similarly, we have:

$$T_3(\mathbf{p}_1) : x \to p_1(x) + k \text{ and } T_3(\mathbf{p}_2) : x \to p_2(x) + k$$

for all $x \in \mathbb{R}$. Consequently, we can see that as, for all $x \in \mathbb{R}$,

$$\alpha_1 T_3(\mathbf{p}_1) + \alpha_2 T_3(\mathbf{p}_2) : x \to \alpha_1(p_1(x) + k) + \alpha_2(p_2(x) + k) = \alpha_1 p_1(x) + \alpha_2 p_2(x) + (\alpha_1 + \alpha_2)k$$

is not equal to

$$T_3(\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2) : x \to \alpha_1 p_1(x) + \alpha_2 p_2(x) + k,$$

for the *general* vectors and scalars considered, $T_3(\mathbf{p})$ is *not* a linear transformation.[4] (Note that this transformation (which we could have written as $T_3(\mathbf{p}) = \mathbf{p} + k \cdot \mathbf{1}$), although not linear, is very useful in linear algebra and we shall look at why it is important later.)

[4] But, this should be obvious as $T_3(\mathbf{0}) = \mathbf{0} + k \cdot \mathbf{1} = k \cdot \mathbf{1} \neq 0$ contrary to Theorem 3.3(1). (Recall that here, $\mathbf{0} \in \mathbb{P}_n^{\mathbb{R}}$ corresponds to the zero function (or, indeed, the *null polynomial*) — see Section 1.3.3.)

Lastly, we note that the composition $T_1 \circ T_2$ is only defined if $m = 0$ as we require that the co-domain of T_2 (i.e. $\mathbb{P}^{\mathbb{R}}_{2+m}$) is the same as the domain of T_1 (i.e. $\mathbb{P}^{\mathbb{R}}_2$). So, *in this case*, $T_1 \circ T_2 : \mathbb{P}^{\mathbb{R}}_2 \to \mathbb{P}^{\mathbb{R}}_2$ is given by

$$T_1 \circ T_2(\mathbf{p}) = T_1(T_2(\mathbf{p})) = T_1(\mathbf{p}) : x \to p(x+k),\ \forall x \in \mathbb{R}$$

and $T_2 \circ T_1 : \mathbb{P}^{\mathbb{R}}_2 \to \mathbb{P}^{\mathbb{R}}_{2+m}$ is

$$T_2 \circ T_1(\mathbf{p}(x)) = T_2(T_1(\mathbf{p}(x))) = T_2(\mathbf{p}(x+k)) : x \to x^m p(x+k),\ \forall x \in \mathbb{R}$$

for any vector $\mathbf{p} \in \mathbb{P}^{\mathbb{R}}_2$ such that $\mathbf{p}(x) = p(x)$ and some fixed non-zero real number k and fixed integer $m \geq 0$. ♣

3.2 Ranges, Null Spaces And The Rank-Nullity Theorem

Associated with each linear transformation $T : V \to W$ are two special vector spaces which will be very useful in our study of linear algebra. One of them is the *null space* of T and this is a subspace of V which contains all of the vectors in V which T maps to the *null vector* (or additive identity) in W. The other is the *range* of T and this is a subspace of W which contains all of the vectors in W that are mapped to from vectors in V by T. So, formally, we have:

Definition 3.6 *Let V and W be vector spaces and let $T : V \to W$ be a linear transformation. The null space (or kernel) of T, denoted by $N(T)$, is the subset*

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

of V, and the range of T, denoted by $R(T)$, is the subset

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

of W.

The relationship between the null-space and the range of a linear transformation $T : V \to W$ is illustrated in Figure 3.3. Of course, we still have to justify the assertion that the null space and range of T are not just sub*sets*, but also sub*spaces*, of V and W respectively. This fact is established in the following theorem:

Theorem 3.7 *Let V and W be vector spaces. If $T : V \to W$ is a linear transformation, then the null space of T is a subspace of V and the range of T is a subspace of W.*

The proof of this theorem will be discussed in Problem Sheet 2.

Indeed, as the range and null space of a linear transformation are vector spaces, they will have a dimension associated with them. That is, we can define:

Definition 3.8 *Let V and W be vector spaces. If $T : V \to W$ is a linear transformation, then the dimension of the null space of T is called the nullity of T and is denoted by $\eta(T)$, and the dimension of the range of T is called the rank of T and is denoted by $\rho(T)$.*

Rather surprisingly, despite the fact that the null space and range are subspaces of *different* vector spaces (i.e. V and W respectively), there *is* a relationship between their dimensions as they are related by the linear transformation T. This relationship is given by the *rank-nullity theorem*, i.e.

Theorem 3.9 *Let V and W be vector spaces. If $T : V \to W$ is a linear transformation, then $\rho(T) + \eta(T) = \dim(V)$.*

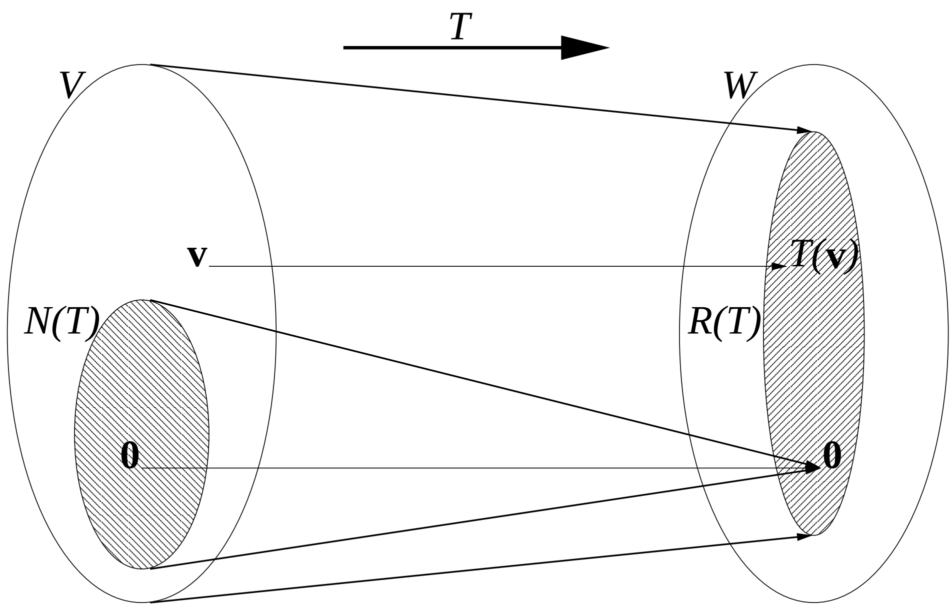

Figure 3.3: *A schematic representation of the relationship between the null space $N(T)$ and range $R(T)$ of a linear transformation $T : V \to W$. Notice that all vectors in V are mapped into $R(T) \subseteq W$, and all vectors in $N(T) \subseteq V$ are mapped to the null vector in W, under the action of T. (Notice that as V, W, $N(T)$ and $R(T)$ are all vector spaces, they all contain the [appropriate] null vector $\mathbf{0}$.)*

Proof: Let V and W be vector spaces, and further, assume that $\dim(V) = n$. We start by establishing that the result holds when $1 \leq \eta(T) \leq n-1$ and treat the special cases where $\eta(T)$ is 0 or n later.[5] So, let us assume that $\eta(T) = r$ where $1 \leq r \leq n-1$, and let the set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a basis of the null space of T, $N(T) \subseteq V$. Now, as S is a basis, it is a linearly independent set of vectors and so we can find $n-r$ vectors $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n$ such that the expanded set $S' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n\}$ is a basis for V.[6]

But what about the range of T, $R(T) \subset W$? The proof now proceeds by establishing that the vectors needed to turn S into S' (i.e. the vectors $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n$), when mapped into W by T, form a set $S'' = \{T(\mathbf{v}_{r+1}), T(\mathbf{v}_{r+2}), \ldots, T(\mathbf{v}_n)\} \subseteq W$ which is a basis for $R(T)$. Of course, to show this, we only need to establish that the set S'' both spans $R(T)$ and is linearly independent (recall Theorem 2.10):

- **To show that S'' spans $R(T)$:** Let $\mathbf{w}$ be any vector in the range of T, that is, by Definition 3.6, there is some vector $\mathbf{v} \in V$ such that $\mathbf{w} = T(\mathbf{v}) \in W$. Now, since the set S' is a basis for V, we can write the vector $\mathbf{v}$ as a linear combination of the vectors in this set, i.e.

$$\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_r\mathbf{v}_r + \alpha_{r+1}\mathbf{v}_{r+1} + \alpha_{r+2}\mathbf{v}_{r+2} + \cdots + \alpha_n\mathbf{v}_n$$

We now note that as the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are in the null space of T, by Definition 3.6, we have $T(\mathbf{v}_1) = T(\mathbf{v}_2) = \cdots = T(\mathbf{v}_r) = \mathbf{0}$, and consequently because T is a linear transformation,[7] we have

$$\mathbf{w} = T(\mathbf{v}) = \mathbf{0} + \alpha_{r+1}T(\mathbf{v}_{r+1}) + \alpha_{r+2}T(\mathbf{v}_{r+2}) + \cdots + \alpha_n T(\mathbf{v}_n)$$

Thus, any vector $\mathbf{w} \in R(T)$ can be written as a linear combination of the vectors in the set $S'' = \{T(\mathbf{v}_{r+1}), T(\mathbf{v}_{r+2}), \ldots, T(\mathbf{v}_n)\}$, and so S'' spans $R(T)$ by Definition 2.2 (as required).

[5] It should be obvious that if U is a subspace of a [finite dimensional] vector space V, then $0 \leq \dim(U) \leq \dim(V)$. Clearly, $\dim(U) \geq 0$ holds because the smallest possible subspace of V is $\{\mathbf{0}\}$ and this has a dimension of zero. Further, the fact that $\dim(U) \leq \dim(V)$ was established in the Harder Problems of Problem Sheet 1.

[6] This is a direct consequence of Theorem 2.9, that is, we could prove a generalisation of this result which says that: If $S \subseteq V$ is a linearly independent set of vectors that is not already a basis for V, then the set S can be expanded to form a new set S' which is a basis for V by adding appropriate vectors to S. If you are suspicious of this, you will have the chance to prove this result (and another related theorem) in the Harder Problems of Problem Sheet 2.

[7] Note the implicit use of Theorem 3.2 here.

- **To show that S'' is linearly independent:** To show that the vectors in S'' are linearly independent, consider the vector equation

$$\alpha_{r+1}T(\mathbf{v}_{r+1}) + \alpha_{r+2}T(\mathbf{v}_{r+2}) + \cdots + \alpha_n T(\mathbf{v}_n) = \mathbf{0}$$

as per Definition 2.6. Since T is a linear transformation, this implies that[8]

$$T(\alpha_{r+1}\mathbf{v}_{r+1} + \alpha_{r+2}\mathbf{v}_{r+2} + \cdots + \alpha_n \mathbf{v}_n) = \mathbf{0}$$

and so, by Definition 3.6, the vector $\alpha_{r+1}\mathbf{v}_{r+1} + \alpha_{r+2}\mathbf{v}_{r+2} + \cdots + \alpha_n \mathbf{v}_n$ is in the null space of T. Consequently, we can write this vector as a linear combination of the vectors in S as this set is a basis for the null space of T, i.e. there are scalars $\alpha_1, \alpha_2, \ldots, \alpha_r$ such that

$$\alpha_{r+1}\mathbf{v}_{r+1} + \alpha_{r+2}\mathbf{v}_{r+2} + \cdots + \alpha_n \mathbf{v}_n = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_r\mathbf{v}_r$$

which on re-arranging gives

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_r\mathbf{v}_r - \alpha_{r+1}\mathbf{v}_{r+1} - \alpha_{r+2}\mathbf{v}_{r+2} - \cdots - \alpha_n \mathbf{v}_n = \mathbf{0}$$

But, since the set S' is a basis, and hence linearly independent, all of the coefficients in this vector equation must be zero. Hence, in particular, $\alpha_{r+1} = a_{r+2} = \cdots = \alpha_n = 0$, and the vectors in S'' are linearly independent (as required).

Consequently, we have shown that the set $S'' = \{T(\mathbf{v}_{r+1}), T(\mathbf{v}_{r+2}), \ldots, T(\mathbf{v}_n)\}$ is a basis of $R(T)$, and further the rank of T, i.e. $\rho(T)$, must equal $n - r$ as there are this many vectors in S''.

The required result now follows from the fact that $\eta(T) = r$ and $\rho(T) = n - r$, i.e.

$$\eta(T) + \rho(T) = r + (n - r) = n = \dim(V)$$

as we assumed that V was an n-dimensional vector space in the proof. (The special cases where $\eta(T) = 0$ or n will be considered in Problem Sheet 2.) ♠

Let us now look at some examples of this.

For example: We saw in the previous section that $T_1(\mathbf{v}) = \mathbf{0}$ and $T_2(\mathbf{v}) = k\mathbf{v}$ [for some non-zero fixed scalar k] are linear transformations from V to V. Let us illustrate these new concepts by applying them to these two simple examples.

As T_1 maps *all* vectors in its domain to the null vector in its co-domain, the null space of T_1 is V and its range is $\{\mathbf{0}\}$. Consequently, the nullity of T_1 is $\dim(V)$, its rank is zero and as $\dim(V) + 0 = \dim(V)$ the rank-nullity theorem is satisfied.

As T_2 only maps the null vector in the domain to the null vector in the co-domain, the null space of T_2 is $\{\mathbf{0}\}$. Further, the range of T_2 is V (as, for every vector in the co-domain, there is a vector in the domain which will map to it under T_2). Consequently, the nullity of T_2 is zero, its rank is $\dim(V)$ and as $0 + \dim(V) = \dim(V)$ the rank-nullity theorem is [again] satisfied. ♣

But finding the range and null-space of a linear transformation may not always be so easy as we shall see in Problem Sheet 2.

3.3 Representing Linear Transformations By Matrices

However, despite the utility of this abstract way of looking at linear transformations when trying to prove theorems, it is not always the most convenient way of dealing with them. In this section we shall look at a linear transformation $T : V \to W$ where V and W are n and m-dimensional spaces respectively to see how it is often easier to represent the transformation as a *matrix* instead of a function. As the development of this idea can be quite confusing, we shall split the discussion into four parts.

[8] Note the implicit use of Theorems 3.2 and 3.3(1) here.

3.3.1 Coordinates and coordinate vectors

To see how this is done, we start by considering that any vector $\mathbf{v} \in V$ is just an object in this n-dimensional vector space and its representation depends on the basis which we choose. For instance if the set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V, we can represent $\mathbf{v}$ as a linear combination of the vectors in this set, i.e.

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

where the coefficients $x_1, x_2, \ldots, x_n$ are scalars. Now, if all of these scalars are real numbers, we can regard the column vector $[x_1, x_2, \ldots, x_n]^t$ in $\mathbb{R}^n$ as another way of representing the vector $\mathbf{v}$ *with respect to* the basis S.[9]

Of course, although I say that this is a 'new representation' of the vector $\mathbf{v} \in V$, you have seen and used it before, even in this course. The column vector in question, namely $[x_1, x_2, \ldots, x_n]^t$, has a standard geometrical interpretation. That is, in the case where V is $\mathbb{R}^n$, we can think of $\mathbf{v}$ as the *position vector* of some point in the space, and this column vector is nothing more than the *coordinate vector* which tells us where this point is *relative* to the basis vectors that lie along the axes being used. Indeed, this is why it is important that we know *which* basis is being used when doing calculations with coordinate vectors — if we change the basis vectors (i.e. the directions of the axes[10]), then we will change the coordinate vector (i.e. the coordinates) of the point we are looking at!

So, formally, we have:

Definition 3.10 *Let the set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of the vector space V. The real numbers $x_1, x_2, \ldots, x_n$ are the [unique] coordinates of a vector $\mathbf{v}$ in the n-dimensional space V with respect to the basis S iff*

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

Further, we define $[\mathbf{v}]_S$, where

$$[\mathbf{v}]_S = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_S$$

to be the coordinate vector of $\mathbf{v}$ relative to the basis S.

Notice that by Definition 2.1 and Theorem 2.12, the coordinates of a vector with respect to a certain basis are just the coefficients of the unique linear combination of the basis vectors that is used to express it.

Example: Consider the sets of vectors

$$E = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \text{ and } S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\},$$

which are both bases of the vector space $\mathbb{R}^3$ (verify this!). The basis given by E is called the *standard basis* of $\mathbb{R}^3$, and so, unless otherwise stated, vectors will be given in terms of it, i.e. the vector $[6, 5, 2]^t$ (say) is implicitly given in terms of this basis. That is, explicitly,

$$\begin{bmatrix} 6 \\ 5 \\ 2 \end{bmatrix}_E = 6\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Thus, the vector $[6, 5, 2]^t$ has coordinates 6, 5, 2 with respect to the basis E [as one would expect!]. However, when referring to a 'non'-standard basis, such as S, we note that we can write the vector

[9] Further, this new representation will be *unique* as the linear combination which gives rise to the coefficients used in this column vector is unique (recall Theorem 2.12).

[10] Strictly, the scales on the axes can change too.

$[6,5,2]^t$ as a linear combination of the vectors in S, i.e.

$$\begin{bmatrix} 6 \\ 5 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

So, it is clear that $[2,1,3]_S^t$ is the coordinate vector of the vector $[6,5,2]^t$ relative to the basis S, and that 2, 1, 3 are the coordinates of this vector with respect to this basis. One of the many applications of the matrix representation of a linear transformation is that it enables us to calculate how the coordinates of a vector change when we switch between bases (see the Appendix at the end of this hand-out).

3.3.2 Why can matrices represent linear transformations?

Now, consider again the set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ which is a basis of the n-dimensional vector space V and the vector $\mathbf{v} \in V$ such that

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

where the coefficients $x_1, x_2, \ldots, x_n$ are scalars. Let us also introduce the set of vectors $S' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ to be a basis of the m-dimensional vector space W. If we now consider our linear transformation, $T : V \to W$, it should be clear that the vectors $T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)$ will be in W and so, as S' is a basis for W, they can be written as linear combinations of the vectors in S', i.e.

$$\begin{array}{ccccccccc} T(\mathbf{v}_1) & = & a_{11}\mathbf{w}_1 & + & a_{21}\mathbf{w}_2 & + & \cdots & + & a_{m1}\mathbf{w}_m \\ T(\mathbf{v}_2) & = & a_{12}\mathbf{w}_1 & + & a_{22}\mathbf{w}_2 & + & \cdots & + & a_{m2}\mathbf{w}_m \\ \vdots & & \vdots & & \vdots & & & & \vdots \\ T(\mathbf{v}_n) & = & a_{1n}\mathbf{w}_1 & + & a_{2n}\mathbf{w}_2 & + & \cdots & + & a_{mn}\mathbf{w}_m \end{array}$$

We also know that linear combinations are preserved under linear transformations (recall Theorem 3.2), and so

$$T(\mathbf{v}) = T(x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n) = x_1T(\mathbf{v}_1) + x_2T(\mathbf{v}_2) + \cdots + x_nT(\mathbf{v}_n)$$

Consequently, substituting for the vectors $T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)$ using the linear combinations above we get

$$T(\mathbf{v}) = x_1(a_{11}\mathbf{w}_1 + a_{21}\mathbf{w}_2 + \cdots + a_{m1}\mathbf{w}_m) + \cdots + x_n(a_{1n}\mathbf{w}_1 + a_{2n}\mathbf{w}_2 + \cdots + a_{mn}\mathbf{w}_m)$$

which on re-arranging gives

$$T(\mathbf{v}) = (x_1a_{11} + x_2a_{12} + \cdots + x_na_{1n})\mathbf{w}_1 + \cdots + (x_1a_{m1} + x_2a_{m2} + \cdots + x_na_{mn})\mathbf{w}_m$$

Now, if we write

$$T(\mathbf{v}) = y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_m\mathbf{w}_m$$

where the coefficients $y_1, y_2, \ldots, y_m$ are scalars and then compare the two expressions for $T(\mathbf{v})$, we get the matrix equation

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where the ith row represents the coefficient of the vector $\mathbf{w}_i$.

But, what does this tell us? To interpret this matrix equation, let us assume for simplicity that the vector spaces V and W are real, i.e. the scalars $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ are all real numbers.[11] This means that the vectors in our matrix equation are such that $[x_1, x_2, \ldots, x_n]^t \in \mathbb{R}^n$ and $[y_1, y_2, \ldots, y_m]^t \in \mathbb{R}^m$, indeed using the notation introduced in Definition 3.10 we can use them to *represent* the vectors $\mathbf{v} \in V$ and $T(\mathbf{v}) \in W$ respectively, i.e.

- The real numbers $x_1, x_2, \ldots, x_n$ are the coordinates of the vector $\mathbf{v} \in V$ with respect to the basis S (as $\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$) and so we can write $[\mathbf{v}]_S = [x_1, x_2, \ldots, x_n]^t_S$.[12]
- The real numbers $y_1, y_2, \ldots, y_m$ are the coordinates of the vector $T(\mathbf{v}) \in W$ with respect to the basis S' (as $T(\mathbf{v}) = y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_m\mathbf{w}_m$) and so we can write $[T(\mathbf{v})]_{S'} = [y_1, y_2, \ldots, y_m]^t_{S'}$.[13]

Thus, we can see that the matrix equation above can be re-written as

$$[T(\mathbf{v})]_{S'} = \mathsf{A}[\mathbf{v}]_S$$

where A is the $m \times n$ matrix $(a_{i,j})$ given above. Consequently, we can represent the linear transformation $T : V \to W$ as a matrix equation, and formally, we can say that[14]

Definition 3.11 *Given a linear transformation $T : V \to W$, the matrix A_T is called the matrix for T with respect to the bases S and S' if*

$$[T(\mathbf{v})]_{S'} = \mathsf{A}_T[\mathbf{v}]_S$$

for all vectors $\mathbf{v} \in V$.

Indeed, the analysis presented above can be used to establish that:

Theorem 3.12 *Let V and W be n and m-dimensional vector spaces respectively. If $T : V \to W$ is a linear transformation, then there is an $m \times n$ matrix A such that $T(\mathbf{v}) = \mathsf{A}\mathbf{v}$.*

or, conversely,

Theorem 3.13 *Let V and W be n and m-dimensional vector spaces respectively. If A is an $m \times n$ matrix, then there is a linear transformation $T : V \to W$ such that $T(\mathbf{v}) = \mathsf{A}\mathbf{v}$.*

The proofs of these two theorems will be considered in Problem Sheet 2.

We now go on to consider how all of this theory works in practice by looking at how we can find A_T given T and *vice versa.*

3.3.3 Given T, how do you find A_T?

So far, we have only considered examples of simple linear transformations (i.e. those like $T(\mathbf{v}) = \mathbf{0}$ and $T(\mathbf{v}) = k\mathbf{v}$ in the Example of Section 1.1) which can be expressed in terms of a vector $\mathbf{v}$ without reference to a basis. However, we want to be in a position to deal with more complicated examples such as, say, the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by the formula

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x + y \\ y + z \end{bmatrix}$$

where clearly, the vectors in $\mathbb{R}^3$ that are being transformed are now being represented *relative to some basis* by the vector $[x, y, z]^t \in \mathbb{R}^3$. For simplicity, we shall follow the convention introduced in

[11] The only other possibility that we will consider in this course is that the vector spaces V and W are complex. In this case, the scalars would be complex numbers and the coordinate vectors would be elements of $\mathbb{C}^n$. The analysis that follows can easily be altered to take this into account.

[12] That is, the vector $[x_1, x_2, \ldots, x_n]^t \in \mathbb{R}^n$ can be interpreted as the coordinate vector $[x_1, x_2, \ldots, x_n]^t_S$ of $\mathbf{v} \in V$ relative to the basis $S \subseteq V$.

[13] That is, the vector $[y_1, y_2, \ldots, y_m]^t \in \mathbb{R}^m$ can be interpreted as the coordinate vector $[y_1, y_2, \ldots, y_m]^t_{S'}$ of $T(\mathbf{v}) \in W$ relative to the basis $S' \subseteq W$.

[14] Notice that Definition 3.11 and Theorems 3.12 and 3.13 are completely general and consequently hold for complex vector spaces too.

the Example of Section 3.3.1 and assume that if *no* basis is specified, then the bases being used are the *standard bases* of the vector spaces in question.

So, in general, we will have a linear transformation $T : V \to W$ given by some formula and we will want to represent it by some matrix A_T such that

$$\mathsf{A}_T[\mathbf{v}]_S = [T(\mathbf{v})]_{S'}$$

where on the left-hand-side we have A_T acting on the coordinate vector of $\mathbf{v}$ relative to some basis S (i.e. we have A_T multiplied by $[\mathbf{v}]_S$) which gives us the *vector* on the right-hand-side which is the result of the transformation — i.e. $T([\mathbf{v}])$ — expressed as a coordinate vector relative to the basis S'. We now recall that $[\mathbf{v}]_S$ is just the vector $[x_1, x_2, \ldots, x_n]^t$ of scalars $x_1, x_2, \ldots, x_n$ such that

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

where $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V and similarly, $[T(\mathbf{v})]_{S'}$ is the vector $[y_1, y_2, \ldots, y_n]^t$ of scalars $y_1, y_2, \ldots, y_n$ such that

$$T(\mathbf{v}) = y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_m\mathbf{w}_m$$

where $S' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is a basis for W. Indeed, using this notation, we found earlier that the matrix which we seek is contained within the matrix equation

$$[T(\mathbf{v})]_{S'} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}_{S'} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_S = \mathsf{A}_T[\mathbf{v}]_S$$

Now, the basis vector $\mathbf{v}_1 \in S$ is represented by the coordinate vector $[\mathbf{v}_1]_S = [1, 0, 0, \ldots, 0]^t_S$ relative to the basis S, $\mathbf{v}_2 \in S$ by $[\mathbf{v}_2]_S = [0, 1, 0, \ldots, 0]^t_S$, *et cetera* and so we can see that

$$[T(\mathbf{v}_1)]_{S'} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

as $x_1 = 1$ is the only non-zero component in $[\mathbf{v}_1]_S = [1, 0, 0, \ldots, 0]^t_S$,

$$[T(\mathbf{v}_2)]_{S'} = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}$$

as $x_2 = 1$ is the only non-zero component in $[\mathbf{v}_2]_S = [0, 1, 0, \ldots, 0]^t_S$, *et cetera.* Consequently, we can see that the ith column of the matrix A is just the coordinate vector $[T(\mathbf{v}_i)]_{S'}$ gained from transforming the ith basis vector in S, that is:

$$\mathsf{A}_T = \begin{bmatrix} | & | & & | \\ [T(\mathbf{v}_1)]_{S'} & [T(\mathbf{v}_2)]_{S'} & \cdots & [T(\mathbf{v}_n)]_{S'} \\ | & | & & | \end{bmatrix}$$

So, given the formula for some linear transformation $T : V \to W$ we now have a method for calculating a matrix A_T such that

$$[T(\mathbf{v})]_{S'} = \mathsf{A}_T[\mathbf{v}]_S$$

which is the sought after matrix for T with respect to the bases S and S' (as defined in Definition 3.11). Luckily, using this method is much simpler than the derivation and to illustrate this, let us look at some examples.

Example: Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation where

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ y+z \end{bmatrix}$$

Find a matrix A_T such that $T(\mathbf{x}) = \mathsf{A}_T\mathbf{x}$.

Solution: This is particularly simple if we take the bases S and S' to be the standard basis for $\mathbb{R}^3$ and $\mathbb{R}^2$ respectively. As you know, this means that

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad S' = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

So, calculating the values of $T(\mathbf{v})$ for the elements of S we find that

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

where the vectors on the left-hand-sides of these expressions are automatically of the required form (i.e. $[T(\mathbf{v})]_{S'}$) as S' is the standard basis of $\mathbb{R}^2$. Thus, in this case,

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

is the required matrix for the bases S and S' given above.

Alternatively, we could make life a little more difficult by specifying that the bases which we want to use are

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\} \quad \text{and} \quad S' = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

where S is the basis of $\mathbb{R}^3$ which we considered in Section 3.3.1 and S' is a basis for $\mathbb{R}^2$ (verify this!). So, as before, we start by calculating $T(\mathbf{v}_i)$ for the basis vectors $\mathbf{v}_i \in S$ and we find that

$$T(\mathbf{v}_1) = T\left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad T(\mathbf{v}_2) = T\left(\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \text{and} \quad T(\mathbf{v}_3) = T\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

where the vectors on the left-hand-sides are written in terms of the standard basis of $\mathbb{R}^2$. However, to find A_T in this case, we need them to be written in terms of the basis $S' = \{\mathbf{w}_1, \mathbf{w}_2\}$ where $\mathbf{w}_1 = [1, 2]^t$ and $\mathbf{w}_1 = [0, -1]^t$, i.e. we need to find $[T(\mathbf{v})]_{S'}$ for each of these vectors. To do this we note that the first expression gives

$$T(\mathbf{v}_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \mathbf{w}_1 + \mathbf{w}_2 \implies [T(\mathbf{v}_1)]_{S'} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{S'}$$

the second expression gives

$$T(\mathbf{v}_2) = \begin{bmatrix} 3 \\ 2 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 4\begin{bmatrix} 0 \\ -1 \end{bmatrix} = 3\mathbf{w}_1 + 4\mathbf{w}_2 \implies [T(\mathbf{v}_2)]_{S'} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}_{S'}$$

and the third expression gives

$$T(\mathbf{v}_3) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 3\begin{bmatrix} 0 \\ -1 \end{bmatrix} = 2\mathbf{w}_1 + 3\mathbf{w}_2 \implies [T(\mathbf{v}_3)]_{S'} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}_{S'}$$

Consequently,

$$\mathsf{A}_T = \begin{bmatrix} 1 & 3 & 2 \\ 1 & 4 & 3 \end{bmatrix}$$

is the required matrix for the bases S and S' given above. (Notice that, as expected, if we change the bases involved we change the matrix.) ♣

However, the power of this method is that the vector spaces V and W do not need to be Euclidean spaces (i.e. spaces of the form $\mathbb{R}^n$ or $\mathbb{C}^n$). Despite the fact that the matrix representation relies on the *coordinate vectors* of $\mathbf{v}$ and $T(\mathbf{v})$ (i.e. $[\mathbf{v}]_S$ and $[T(\mathbf{v})]_{S'}$) being elements of a Euclidean space, the vectors $\mathbf{v}$ and $T(\mathbf{v})$ *themselves* need not be. Let us consider an illustration of this fact.

Example: Let $T : \mathbb{P}_2^{\mathbb{R}} \to \mathbb{P}_1^{\mathbb{R}}$ be the linear transformation where

$$T(a_0 \cdot \mathbf{1} + a_1 \cdot \mathbf{x} + a_2 \cdot \mathbf{x^2}) = a_1 \cdot \mathbf{1} + 2a_2 \cdot \mathbf{x}$$

Find a matrix A such that $T(\mathbf{p}) = \mathsf{A}\mathbf{p}$ for $\mathbf{p} \in \mathbb{P}_2^{\mathbb{R}}$.

Note: Although we haven't come across it yet, the standard basis for $\mathbb{P}_n^{\mathbb{R}}$ is the set of vectors $\{\mathbf{1}, \mathbf{x}, \ldots, \mathbf{x^n}\}$.

Solution: This is particularly simple if we take the bases S and S' to be the standard basis for $\mathbb{P}_2^{\mathbb{R}}$ and $\mathbb{P}_1^{\mathbb{R}}$ respectively. This means that

$$S = \{\mathbf{1}, \mathbf{x}, \mathbf{x^2}\} \quad \text{and} \quad S' = \{\mathbf{1}, \mathbf{x}\}$$

So, calculating the values of $T(\mathbf{p})$ for the elements of S we find that

$$T(\mathbf{1}) = \mathbf{0}, \ \ T(\mathbf{x}) = \mathbf{1} \ \text{ and } \ T(\mathbf{x^2}) = 2 \cdot \mathbf{x}$$

where the vectors on the left-hand-sides of these expressions are automatically of the required form (i.e. $[T(\mathbf{p})]_{S'}$) as S' is the standard basis of $\mathbb{P}_1^{\mathbb{R}}$. Thus, in this case,

$$\mathsf{A}_T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

is the required matrix for the bases S and S' given above.

Alternatively, we could [again] make life a little more difficult by specifying that the bases which we want to use are

$$S = \{4 \cdot \mathbf{x} + \mathbf{1}, \mathbf{x^2} + 2 \cdot \mathbf{x}, \mathbf{x^2} + \mathbf{1}\} \quad \text{and} \quad S' = \{\mathbf{x} + \mathbf{1}, \mathbf{x} - \mathbf{1}\}$$

So, as before, we start by calculating $T(\mathbf{v}_i)$ for the basis vectors $\mathbf{v}_i \in S$ and we find that

$$T(4 \cdot \mathbf{x} + \mathbf{1}) = 4 \cdot \mathbf{1}, \ \ T(\mathbf{x^2} + 2 \cdot \mathbf{x}) = 2 \cdot \mathbf{x} + 2 \cdot \mathbf{1} \ \text{ and } \ T(\mathbf{x^2} + \mathbf{1}) = 2 \cdot \mathbf{x}$$

where the vectors on the left-hand-sides are written in terms of the standard basis of $\mathbb{P}_1^{\mathbb{R}}$. However, to find A_T in this case, we need them to be written in terms of the basis $S' = \{\mathbf{w}_1, \mathbf{w}_2\}$ where $\mathbf{w}_1 = \mathbf{x} + \mathbf{1}$ and $\mathbf{w}_2 = \mathbf{x} - \mathbf{1}$, i.e. we need to find $[T(\mathbf{v})]_{S'}$ for each of these vectors. To do this we note that the first expression gives

$$T(\mathbf{v}_1) = 4 \cdot \mathbf{1} = 2 \cdot (\mathbf{w}_1 - \mathbf{w}_2) = 2 \cdot \mathbf{w}_1 - 2 \cdot \mathbf{w}_2 \implies [T(\mathbf{v}_1)]_{S'} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}_{S'}$$

the second expression gives

$$T(\mathbf{v}_2) = 2 \cdot \mathbf{x} + 2 \cdot \mathbf{1} = 2 \cdot \mathbf{w}_1 \implies [T(\mathbf{v}_2)]_{S'} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}_{S'}$$

and the third expression gives

$$T(\mathbf{v}_3) = 2 \cdot \mathbf{x} = \mathbf{w}_1 + \mathbf{w}_2 \implies [T(\mathbf{v}_3)]_{S'} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{S'}$$

Consequently,

$$\mathsf{A}_T = \begin{bmatrix} 2 & 2 & 1 \\ -2 & 0 & 1 \end{bmatrix}$$

is the required matrix for the bases S and S' given above. (Notice that, as expected, if we change the bases involved we change the matrix.) ♣

3.3.4 Given A_T, how do you find T?

What happens if you are given a linear transformation T in terms of a matrix A_T and you want to find its null space or range? Well, luckily, a matrix also has a range and a null space associated with it, and these are — relative to the bases involved — the same as the range and null space of the transformation. You learnt how to calculate the range and null space of a matrix in MA100 (using row operations *et cetera*) and so we won't go over this again here. However, an alternative method would be to try and recover the linear transformation from the matrix and use this to calculate the range and null space.

So, we start with a matrix equation such as

$$[T(\mathbf{v})]_{S'} = \mathsf{A}_T[\mathbf{v}]_S$$

where $[\mathbf{v}]_S \in V$ and $[T(\mathbf{v})]_{S'} \in W$, and we want to find the linear transformation $T : V \to W$ which corresponds to this, i.e.

$$\mathbf{w} = T(\mathbf{v})$$

where $\mathbf{w} \in W$ and $\mathbf{v} \in V$ are written relative to the standard basis. To see how this is done, we start by re-writing the given matrix equation as

$$[T(\mathbf{v})]_{S'} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}_{S'} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_S = \mathsf{A}_T[\mathbf{v}]_S$$

using the notation introduced earlier. As we want to recover this linear transformation in terms of the standard bases of V and W, it would be good if we could find its matrix representation with respect to these bases. That is, we want to convert the coordinate vectors $[\mathbf{v}]_S$ and $[T(\mathbf{v})]_{S'}$ relative to the bases S and S' of V and W respectively into coordinate vectors $[\mathbf{v}]$ and $[T(\mathbf{v})]$ relative to the standard bases of V and W respectively. To do this, we just 'change basis' (see the Appendix at the end of this hand-out if you don't know how to do this!), i.e. we note that

$$\mathsf{V}[\mathbf{v}]_S = [\mathbf{v}] \quad \text{and} \quad \mathsf{W}[T(\mathbf{v})]_{S'} = [T(\mathbf{v})]$$

where the columns of the matrices V and W are the basis vectors in S and S' respectively. Consequently, on re-arranging these expressions and substituting we have

$$[T(\mathbf{v})] = \mathsf{W}\mathsf{A}_T\mathsf{V}^{-1}[\mathbf{v}]$$

where $\mathsf{W}\mathsf{A}_T\mathsf{V}^{-1}$ is the matrix representing T with respect to the standard bases of S and S'. Now, if we let $[\mathbf{v}]$ be some vector $[z_1, z_2, \ldots, z_n]^t$ and multiply it by $\mathsf{W}\mathsf{A}_T\mathsf{V}^{-1}$, we will find the effect that T has on vectors in V — which is what we were after.

Again, the method is easier to apply in practice, and so let us look at an example.

Example: Given that a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ is represented by the matrix

$$\mathsf{A}_T = \begin{bmatrix} 1 & 3 & 2 \\ 1 & 4 & 3 \end{bmatrix}$$

where the bases being used in $\mathbb{R}^3$ and $\mathbb{R}^2$ are

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\} \quad \text{and} \quad S' = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

find the formula for $T(\mathbf{v})$.

Solution: As per the method above, we start by finding the matrices which will change the bases being used in $\mathbb{R}^3$ and $\mathbb{R}^2$ from S and S' to the appropriate standard basis. As these matrices just have the basis vectors as their columns we get

$$\mathsf{V} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathsf{W} = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$$

So, as indicated above, we then calculate the matrix product

$$\mathsf{W}\mathsf{A}_T\mathsf{V}^{-1} = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ -1 & 1 & 1 \\ 2 & -1 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

and multiply this matrix by a vector $[z_1, z_2, z_3]^t$ representing $[\mathbf{v}]$, to get

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} z_1 + z_2 \\ z_2 + z_3 \end{bmatrix}$$

which tells us that the formula for $T(\mathbf{v})$ is

$$T\left(\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \right) = \begin{bmatrix} z_1 + z_2 \\ z_2 + z_3 \end{bmatrix}$$

as you may have expected. ♣

3.4 Appendix: How To Change Basis

Let us consider a vector $\mathbf{x}$ in a vector space V which is given by the coordinate vectors

- $[\mathbf{x}]_S = [a_1, a_2, \ldots, a_n]_S^t$ relative to a basis $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\} \subseteq V$
- $[\mathbf{x}]_{S'} = [b_1, b_2, \ldots, b_n]_{S'}^t$ relative to some other basis $S' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subseteq V$

Now, as both of these coordinate vectors represent the *same* vector, it should be clear that when they are written out *in full*, they must be equal, i.e.

$$\mathbf{x} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n$$

But, this can be re-written in terms of matrices, i.e.

$$\begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

or, better still,

$$\mathsf{U}[\mathbf{x}]_S = \mathsf{V}[\mathbf{x}]_{S'}$$

where the columns of the matrices U and V are the basis vectors in S and S' respectively.

Further, notice that as S and S' are just different bases of the same n-dimensional vector space, U and V will both be $n \times n$ matrices, and as bases are linearly independent, they will both be invertible too. This means that we can find $[\mathbf{x}]_S$ from $[\mathbf{x}]_{S'}$ (say) by using

$$[\mathbf{x}]_S = \mathsf{U}^{-1}\mathsf{V}[\mathbf{x}]_{S'}$$

and so, it should be clear that changing basis is yet another example of a linear transformation (in this case from V to V).

3.5 Learning Objectives

At the end of this hand-out you should:

- Understand what a linear transformation is and the properties it possesses as given in the Theorems of Section 3.1.
- Understand that the null space and range of a linear transformation are particular subspaces of certain vector spaces and that their dimensions are related by the rank-nullity theorem as given in Section 3.2. (Although, detailed knowledge of the proof of the rank-nullity theorem is not required for this course.)
- Be able to represent a linear transformation as a matrix using given bases and be able to recover the linear transformation from such a matrix as described in Section 3.3. (Although a detailed justification of why this is possible is not required for this course.)
- Be able to construct a matrix which will allow you to change between bases as described in Section 3.4.

This material will be developed in Problem Sheet 2.