Math 115A - Week 3
Textbook sections: 2.1-2.3
Topics covered:

- Null spaces and nullity of linear transformations
- Range and rank of linear transformations
- The Dimension theorem
- Linear transformations and bases
- Co-ordinate bases
- Matrix representation of linear transformations
- Sum, scalar multiplication, and composition of linear transformations

* * * * *

Review of linear transformations

- A linear transformation is any map $T : V \to W$ from one vector space V to another W such that T preserves vector addition (i.e. $T(v+v') = Tv + Tv'$ for all $v, v' \in V$) and T preserves scalar multiplication (i.e. $T(cv) = cTv$ for all scalars c and all $v \in V$).
- A map which preserves vector addition is sometimes called *additive*; a map which preserves scalar multiplication is sometimes called *homogeneous*.
- We gave several examples of linear transformations in the previous notes; here are a couple more.
- **Sampling as a linear transformation** Recall that $\mathcal{F}(\mathbf{R}, \mathbf{R})$ is the space of all functions from $\mathbf{R}$ to $\mathbf{R}$. This vector space might be used to represent, for instance, sound signals $f(t)$. In practice, a measuring device cannot capture all the information in a signal (which contains an infinite amount of data); instead it only samples a finite amount, at some fixed times. For instance, a measuring device might only sample

$f(t)$ for $t = 1, 2, 3, 4, 5$ (this would correspond to sampling at 1Hz for five seconds). This operation can be described by a linear transformation $S : \mathcal{F}(\mathbf{R}, \mathbf{R}) \to \mathbf{R}^5$, defined by

$$Sf := (f(1), f(2), f(3), f(4), f(5));$$

i.e. S transforms a signal $f(t)$ into a five-dimensional vector, consisting of f sampled at five times. For instance,

$$S(x^2) = (1, 4, 9, 16, 25)$$

$$S(\sqrt{x}) = (\sqrt{1}, \sqrt{2}, \sqrt{3}, \sqrt{4}, \sqrt{5})$$

etc. (Why is this map linear?)

- One can similarly sample polynomial spaces. For instance, the map $S : P_2(\mathbf{R}) \to \mathbf{R}^3$ defined by

  $$Sf := (f(0), f(1), f(2))$$

  is linear.

- **Interpolation as a linear transformation** Interpolation can be viewed as the reverse of sampling. For instance, given three numbers y_1, y_2, y_3, the Lagrange interpolation formula gives us a polynomial $f \in P_2(\mathbf{R})$ such that $f(0) = y_1$, $f(1) = y_2$, and $f(2) = y_3$:

  $$f(x) = y_1 \frac{(x-1)(x-2)}{(0-1)(0-2)} + y_2 \frac{(x-0)(x-2)}{(1-0)(1-2)} + y_3 \frac{(x-0)(x-1)}{(2-0)(2-1)}.$$

  One can view this as a linear transformation $S : \mathbf{R}^3 \to P_2(\mathbf{R})$ defined by $S(y_1, y_2, y_3) := f$, e.g.

  $$S(3,4,7) = y_1 \frac{(x-1)(x-2)}{(0-1)(0-2)} + y_2 \frac{(x-0)(x-2)}{(1-0)(1-2)} + y_3 \frac{(x-0)(x-1)}{(2-0)(2-1)}.$$

  (Why is this linear?). This is the *inverse* of the transformation S defined in the previous paragraph - but more on that later.

- **Linear combinations as a linear transformation** Let V be a vector space, and let $v_1, \ldots, v_n$ be a set of vectors in V. Then the transformation $T : \mathbf{R}^n \to V$ defined by

$$T(a_1, \ldots, a_n) := a_1 v_1 + \ldots + a_n v_n$$

is a linear transformation (why?). Also, one can express many of the statements from previous notes in terms of this transformation T. For instance, $\operatorname{span}(\{v_1, \ldots, v_n\})$ is the same thing as the image $T(\mathbf{R}^n)$ of T; thus $\{v_1, \ldots, v_n\}$ spans V if and only if T is onto. On the other hand, T is one-to-one if and only if $\{v_1, \ldots, v_n\}$ is linearly independent (more on this later). Thus T is a bijection if and only if $\{v_1, \ldots, v_n\}$ is a basis.

* * * * *

Null spaces and nullity

- A note on notation: in this week's notes, we shall often be dealing with two different vector spaces V and W, so we have two different types of vectors. We will try to reserve the letter v to denote vectors in V, and w to denote vectors in W, in what follows.

- Not all linear transformations are alike; for instance, the zero transformation $T : V \to W$ defined by $Tv := 0$ behaves rather differently from, say, the identity transformation $T : V \to V$ defined by $Tv := v$. Now we introduce some characteristics of linear transformations to start telling them apart.

- **Definition** Let $T : V \to W$ be a linear transformation. The *null space* of T, called $N(T)$, is defined to be the set

$$N(T) := \{v \in V : Tv = 0\}.$$

- In other words, the null space consists of all the stuff that T sends to zero (this is the zero vector 0_W of W, not the zero vector 0_V of V): $N(T) = T^{-1}(\{0\})$. Some examples: if $T : V \to W$ is the zero transformation $Tv := 0$, then the null space $N(T) = V$. If instead $T : V \to V$ is the identity transformation $Tv := v$, then $N(T) = \{0\}$. If $T : \mathbf{R}^3 \to \mathbf{R}$ is the linear transformation $T(x, y, z) = x + y + z$, then $N(T)$ is the plane $\{(x, y, z) \in \mathbf{R}^3 : x + y + z = 0\}$.

- The null space of T is sometimes also called the *kernel* of T, and is sometimes denoted $\ker(T)$; but we will use the notation $N(T)$ throughout this course.

- The null space $N(T)$ is always a subspace of V; this is an exercise. Intuitively, the larger the null space, the more T resembles the 0 transformation. The null space also measures the extent to which T fails to be one-to-one:

- **Lemma 1.** Let $T : V \to W$ be a linear transformation. Then T is one-to-one if and only if $N(T) = \{0\}$.

- **Proof.** First suppose that T is one-to-one; we have to show that $N(T) = \{0\}$. First of all, it is clear that $0 \in N(T)$, because $T0 = 0$. Now we show that no other element is in $N(T)$. Suppose for contradiction that there was a non-zero vector $v \in V$ such that $v \in N(T)$, i.e. that $Tv = 0$. Then $Tv = T0$. But T is one-to-one, so this forces $v = 0$, contradiction.

- Now suppose that $N(T) = \{0\}$; we have to show that T is one-to-one. In other words, we need to show that whenever $Tv = Tv'$, then we must have $v = v'$. So suppose that $Tv = Tv'$. Then $Tv - Tv' = 0$, so that $T(v - v') = 0$. Thus $v - v' \in N(T)$, which means by hypothesis that $v - v' = 0$, so $v = v'$, as desired. $\square$

- Example: Take the transformation $T : \mathbf{R}^n \to V$ defined by

  $$T(a_1, \ldots, a_n) := a_1 v_1 + \ldots + a_n v_n$$

  which we discussed earlier. If $\{v_1, \ldots, v_n\}$ is linearly dependent, then there is a non-zero n-tuple $(a_1, \ldots, a_n)$ such that $0 = a_1 v_1 + \ldots + a_n v_n$; i.e. $N(T)$ will consist of more than just the 0 vector. Conversely, if $N(T) \neq \{0\}$, then $\{v_1, \ldots, v_n\}$ is linearly dependent. Thus by Lemma 1, T is injective if and only if $\{v_1, \ldots, v_n\}$ is linearly independent.

- Since $N(T)$ is a vector space, it has a dimension. We define the *nullity* of T to be the dimension of $N(T)$; this may be infinite, if $N(T)$ is infinite dimensional. The nullity of T will be denoted $\text{nullity}(T)$, thus $\text{nullity}(T) = \dim(N(T))$.

- Example: let $\pi : \mathbf{R}^5 \to \mathbf{R}^5$ be the operator

$$\pi(x_1, x_2, x_3, x_4, x_5) := (x_1, x_2, x_3, 0, 0)$$

  (Why is this linear?). Then

$$N(\pi) = \{(0, 0, 0, x_4, x_5) : x_4, x_5 \in \mathbf{R}\}$$

  (why?); this is a two-dimensional space (it has a basis consisting of $(0, 0, 0, 1, 0)$ and $(0, 0, 0, 0, 1)$ and so nullity$(\pi) = 2$.

- Example: By Lemma 1, a transformation is injective if and only if it has a nullity of 0.

- The nullity of T measures how much information (or degrees of freedom) is lost when applying T. For instance, in the above projection, two degrees of freedom are lost: the freedom to vary the x_4 and x_5 co-ordinates are lost after applying π. An injective transformation does not lose any information (if you know Tv, then you can reconstruct v).

* * * * *

Range and rank

- You may have noticed that many concepts in this field seem to come in complementary pairs: spanning set versus linearly independent set, one-to-one versus onto, etc. Another such pair is null space and range, or nullity and rank.

- **Definition** The *range* $R(T)$ of a linear transformation $T : V \to W$ is defined to be the set

$$R(T) := \{Tv : v \in V\}.$$

- In other words, $R(T)$ is all the stuff that T maps into: $R(T) = T(V)$. (Unfortunately, the space W is also sometimes called the range of T; to avoid confusion we will try to refer to W instead as the *target space* for T; V is the *initial space* or *domain* of T.)

- Just as the null space $N(T)$ is always a subspace of V, it can be shown that $R(T)$ is a subspace of W (this is part of an exercise).

- Examples: If $T : V \to W$ is the zero transformation $Tv := 0$, then $R(T) = \{0\}$. If $T : V \to V$ is the identity transformation $Tv := v$, then $R(T) = V$. If $T : \mathbf{R}^n \to V$ is the transformation

$$T(a_1, \ldots, a_n) := a_1 v_1 + \ldots + a_n v_n$$

discussed earlier, then $R(T) = \text{span}(\{v_1, \ldots, v_n\})$.

- Example: A map $T : V \to W$ is onto if and only if $R(T) = W$.

- **Definition** The *rank* $\text{rank}(T)$ of a linear transformation $T : V \to W$ is defined to be the dimension of $R(T)$, thus $\text{rank}(T) = \dim(R(T))$.

- Examples: The zero transformation has rank 0 (and indeed these are the only transformations with rank 0). The transformation

$$\pi(x_1, x_2, x_3, x_4, x_5) := (x_1, x_2, x_3, 0, 0)$$

defined earlier has range

$$R(\pi) = \{(x_1, x_2, x_3, 0, 0) : x_1, x_2, x_3 \in \mathbf{R}\}$$

(why?), and so has rank 3.

- The rank measures how much information (or degrees of freedom) is *retained* by the transformation T. For instance, with the example of π above, even though two degrees of freedom have been lost, three degrees of freedom remain.

* * * * *

The dimension theorem

- Let $T : V \to W$ be a linear transformation. Intuitively, $\text{nullity}(T)$ measures how many degrees of freedom are lost when applying T; $\text{rank}(T)$ measures how many degrees of freedom are retained. Since the initial space V originally has $\dim(V)$ degrees of freedom, the following theorem should not be too surprising.

- **Dimension theorem** Let V be a finite-dimensional space, and let $T : V \to W$ be a linear transformation. Then

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

- The proof here will involve a lot of shuttling back and forth between V and W using T; and is an instructive example as to how to analyze linear transformations.

- **Proof.** By hypothesis, $\dim(V)$ is finite; let's define $n := \dim(V)$. Since $N(T)$ is a subspace of V, it must also be finite-dimensional; let's call $k := \dim(N(T)) = \text{nullity}(T)$. Then we have $0 \leq k \leq n$. Our task is to show that $k + \text{rank}(T) = n$, or in other words that $\dim(R(T)) = n - k$.

- By definition of dimension, the space $N(T)$ must have a basis $\{v_1, \ldots, v_k\}$ of k elements. (Probably it has many such bases, but we just need one such for this argument). This set of k elements lies in $N(T)$, and thus in V, and is linearly independent; thus by part (f) of Corollary 1 of last week's notes, it must be part of a basis of V, which must then have $n = \dim(V)$ elements (by part (c) of Corollary 1). Thus we may add $n - k$ extra elements $v_{k+1}, \ldots, v_n$ to our $N(T)$-basis to form an V-basis $\{v_1, \ldots, v_n\}$.

- Since $v_{k+1}, \ldots, v_n$ lie in V, the elements $Tv_{k+1}, \ldots, Tv_n$ lie in $R(T)$. We now claim that $\{Tv_{k+1}, \ldots, Tv_n\}$ are a basis for $R(T)$; this will imply that $R(T)$ has dimension $n - k$, as desired.

- To verify that $\{Tv_{k+1}, \ldots, Tv_n\}$ form a basis, we must show that they span $R(T)$ and that they are linearly independent. First let's show they span $R(T)$. This means that every vector in $R(T)$ is a linear combination of $Tv_{k+1}, \ldots, Tv_n$. So let's pick a typical vector w in $R(T)$; our job is to show that w is a linear combination of $Tv_{k+1}, \ldots, Tv_n$. By definition of $R(T)$, w must equal Tv for some v in V.

- On the other hand, we know that $\{v_1, \ldots, v_n\}$ spans V, thus we must have

  $$v = a_1 v_1 + \ldots + a_n v_n$$

  for some scalars $a_1, \ldots, a_n$. Applying T to both sides and using the fact that T is linear, we obtain

  $$Tv = a_1 T v_1 + \ldots + a_n T v_n.$$

- Now we use the fact that $v_1, \ldots, v_k$ lie in $N(T)$, so $Tv_1 = \ldots = Tv_k = 0$. Thus
$$Tv = a_{k+1}Tv_{k+1} + \ldots + a_nTv_n.$$
Thus $w = Tv$ is a linear combination of $Tv_{k+1}, \ldots, Tv_n$, as dsired.

- Now we show that $\{Tv_{k+1}, \ldots, Tv_n\}$ is linearly independent. Suppose for contradiction that this set was linearly dependent, thus
$$a_{k+1}Tv_{k+1} + \ldots + a_nTv_n = 0$$
for some scalars $a_{k+1}, \ldots, a_n$ which were not all zero. Then by the linearity of T again, we have
$$T(a_{k+1}v_{k+1} + \ldots + a_nv_n) = 0$$
and thus by definition of null space
$$a_{k+1}v_{k+1} + \ldots + a_nv_n \in N(T).$$
Since $N(T)$ is spanned by $\{v_1, \ldots, v_k\}$, we thus have
$$a_{k+1}v_{k+1} + \ldots + a_nv_n = a_1v_1 + \ldots a_kv_k$$
for some scalars $a_1, \ldots, a_k$. We can rearrange this as
$$-a_1v_1 - \ldots - a_kv_k + a_{k+1}v_{k+1} + \ldots + a_nv_n = 0.$$
But the set $\{v_1, \ldots, v_n\}$ is linearly independent, which means that all the a's must then be zero. But that contradicts our hypothesis that not all of the $a_{k+1}, \ldots, a_n$ were zero. Thus $\{Tv_{k+1}, \ldots, Tv_n\}$ must have been linearly independent, and we are done. $\square$.

- **Example** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ denote the linear transformation
$$T(x, y) := (x + y, 2x + 2y).$$
The null space of this transformation is
$$N(T) = \{(x, y) \in \mathbf{R}^2 : x + y = 0\}$$

(why?); this is a line in $\mathbf{R}^2$, and thus has dimension 1 (for instance, it has $\{(1,-1)\}$ as a basis). The range of this transformation is

$$R(T) = \{(t, 2t) : t \in \mathbf{R}\}$$

(why?); this is another line in $\mathbf{R}^2$ and has dimension 1. Since 1+1=2, the Dimension theorem is verified in this case.

- **Example** For the zero transformation $Tx := 0$, we have nullity$(T) = \dim(X)$ and rank$(T) = 0$ (so all the degrees of freedom are lost); while for the identity transformation $Tx := x$ we have nullity$(T) = 0$ and rank$(T) = \dim(X)$ (so all the degrees of freedom are retained). In both cases we see that the Dimension theorem is verified.

- One important use of the Dimension theorem is that it allows us to discover facts about the range of T just from knowing the null space of T, and vice versa. For instance:

- **Example** Let $T : P_5(\mathbf{R}) \to P_4(\mathbf{R})$ denote the differentiation map

  $$Tf := f';$$

  thus for instance $T(x^3+2x) = 3x^2+2$. The null space of T consists of all polynomials f in $P_5(\mathbf{R})$ for which $f' = 0$; i.e. the constant polynomials

  $$N(T) = \{c : c \in \mathbf{R}\} = P_0(\mathbf{R}).$$

  Thus $N(T)$ has dimension 1 (it has $\{1\}$ as a basis). Since $P_5(\mathbf{R})$ has dimension 6, we thus see from the dimension theorem that $R(T)$ must have dimension 5. But $R(T)$ is a subspace of $P_4(\mathbf{R})$, and $P_4(\mathbf{R})$ has dimension 5. Thus $R(T)$ must equal all of $P_4(\mathbf{R})$. In other words, every polynomial of degree at most 4 is the derivative of some polynomial of degree at most 5. (This is of course easy to check by integration, but the amazing fact was that we could deduce this fact purely from linear algebra - using only a very small amount of calculus).

- Here is another example:

- **Lemma 2** Let V and W be finite-dimensional vector spaces of the same dimension ($\dim(V) = \dim(W)$), and let $T : V \to W$ be a linear transformation from V to W. Then T is one-to-one if and only if T is onto.

- **Proof** If T is one-to-one, then nullity$(T) = 0$, which by the dimension theorem implies that rank$(T) = \dim(V)$. Since $\dim(V) = \dim(W)$, we thus have $\dim R(T) = \dim(W)$. But $R(T)$ is a subspace of W, thus $R(T) = W$, i.e. T is onto. The reverse implication then follows by reversing the above steps (we leave as an exercise to verify that all the steps are indeed reversible). $\square$

- Exercise: re-interpret Corollary 1(de) from last week's notes using this Lemma, and the linear transformation

$$T(a_1, \ldots, a_n) := a_1 v_1 + \ldots + a_n v_n$$

discussed earlier.

Linear transformations and bases

- Let $T : V \to W$ be a linear transformation, and let $\{v_1, \ldots, v_n\}$ be a collection of vectors in V. Then $\{Tv_1, \ldots, Tv_n\}$ is a collection of vectors in W. We now study how similar these two collections are; for instance, if one is a basis, does this mean the other one is also a basis?

- **Theorem 3** If $T : V \to W$ is a linear transformation, and $\{v_1, \ldots, v_n\}$ spans V, then $\{Tv_1, \ldots, Tv_n\}$ spans $R(T)$.

- **Proof**. Let w be any vector in $R(T)$; our job is to show that w is a linear combination of $Tv_1, \ldots, Tv_n$. But by definition of $R(T)$, $w = Tv$ for some $v \in V$. Since $\{v_1, \ldots, v_n\}$ spans V, we thus have $v = a_1 v_1 + \ldots + a_n v_n$ for some scalars $a_1, \ldots, a_n$. Applying T to both sides, we obtain $Tv = a_1 Tv_1 + \ldots + a_n Tv_n$. Thus we can write $w = Tv$ as a linear combination of $Tv_1, \ldots, Tv_n$, as desired. $\square$

- **Theorem 4** If $T : V \to W$ is a linear transformation which is one-to-one, and $\{v_1, \ldots, v_n\}$ is linearly independent, then $\{Tv_1, \ldots, Tv_n\}$ is also linearly independent.

- **Proof** Suppose we can write 0 as a linear combination of $\{Tv_1, \ldots, Tv_n\}$:

$$0 = a_1 Tv_1 + \ldots + a_n Tv_n.$$

Our job is to show that the $a_1, \ldots, a_n$ must all be zero. Using the linearity of T, we obtain

$$0 = T(a_1 v_1 + \ldots + a_n v_n).$$

Since T is one-to-one, $N(T) = \{0\}$, and thus

$$0 = a_1 v_1 + \ldots + a_n v_n.$$

But since $\{v_1, \ldots, v_n\}$ is linearly independent, this means that $a_1, \ldots, a_n$ are all zero, as desired. □

- **Corollary 5** If $T : V \to W$ is both one-to-one and onto, and $\{v_1, \ldots, v_n\}$ is a basis for V, then $\{Tv_1, \ldots, Tv_n\}$ is a basis for W. (In particular, we see that $\dim(V) = \dim(W)$).

- **Proof** Since $\{v_1, \ldots, v_n\}$ is a basis for V, it spans V; and hence, by Theorem 3, $\{Tv_1, \ldots, Tv_n\}$ spans $R(T)$. But $R(T) = W$ since T is onto. Next, since $\{v_1, \ldots, v_n\}$ is linearly independent and T is one-to-one, we see from Theorem 4 that $\{Tv_1, \ldots, Tv_n\}$ is also linearly independent. Combining these facts we see that $\{Tv_1, \ldots, Tv_n\}$ is a basis for W. □

- The converse is also true: if $T : V \to W$ is one-to-one, and $\{Tv_1, \ldots, Tv_n\}$ is a basis, then $\{v_1, \ldots, v_n\}$ is also a basis; we leave this as an exercise (it's very similar to the previous arguments).

- **Example** The map $T : P_3(\mathbf{R}) \to \mathbf{R}^4$ defined by

$$T(ax^3 + bx^2 + cx + d) := (a, b, c, d)$$

is both one-to-one and onto (why?), and is also linear (why?). Thus we can convert every basis of $P_3(\mathbf{R})$ to a basis of $\mathbf{R}^4$ and vice versa. For instance, the standard basis $\{1, x, x^2, x^3\}$ of $P_3(\mathbf{R})$ can be converted to the basis $\{(0,0,0,1), (0,0,1,0), (0,1,0,0), (1,0,0,0)\}$ of $\mathbf{R}^4$. In principle, this allows one to convert many problems about the vector space $P_3(\mathbf{R})$ into one about $\mathbf{R}^4$, or vice versa. (The formal way of saying this is that $P_3(\mathbf{R})$ and $\mathbf{R}^4$ are *isomorphic*; more about this later).

* * * * *

Using a basis to specify a linear transformation

- In this section we discuss one of the fundamental reasons why bases are important; one can use them to describe linear transformations in a compact way.

- In general, to specify a function $f : X \to Y$, one needs to describe the value $f(x)$ for every point x in X; for instance, if $f : \{1, 2, 3, 4, 5\} \to \mathbf{R}$, then one needs to specify $f(1)$, $f(2)$, $f(3)$, $f(4)$, $f(5)$ in order to completely describe the function. Thus, when X gets large, the amount of data needed to specify a function can get quite large; for instance, to specify a function $f : \mathbf{R}^2 \to \mathbf{R}^3$, one needs to specify a vector $f(x) \in \mathbf{R}^3$ for every single point x in $\mathbf{R}^2$ - and there are infinitely many such points! The remarkable thing, though, is that if f is *linear*, then one does not need to specify f at every single point - one just needs to specify f on a basis and this will determine the rest of the function.

- **Theorem 6** Let V be a finite-dimensional vector space, and let $\{v_1, \ldots, v_n\}$ be a basis for V. Let W be another vector space, and let $w_1, \ldots, w_n$ be some vectors in W. Then there exists exactly one linear transformation $T : V \to W$ such that $Tv_j = w_j$ for each $j = 1, 2, \ldots, n$.

- **Proof** We need to show two things: firstly, that there exists a linear transformation T with the desired properties, and secondly that there is at most one such transformation.

- Let's first show that there is at most one transformation. Suppose for contradiction that we had two different linear transformations $T : V \to W$ and $U : V \to W$ such that $Tv_j = w_j$ and $Uv_j = w_j$ for each $j = 1, \ldots, n$. Now take any vector $v \in V$, and consider Tv and Uv.

- Since $\{v_1, \ldots, v_n\}$ is a basis of V, we have a unique representation

$$v = a_1 v_1 + a_2 v_2 + \ldots + a_n v_n$$

where $a_1, \ldots, a_n$ are scalars. Thus, since T is linear

$$Tv = a_1 T v_1 + a_2 T v_2 + \ldots + a_n T v_n$$

but since $Tv_j = w_j$, we have

$$Tv = a_1 w_1 + a_2 w_2 + \ldots + a_n w_n.$$

Arguing similarly with U instead of T, we have

$$Uv = a_1 w_1 + a_2 w_2 + \ldots + a_n w_n.$$

so in particular $Tv = Uv$ for all vectors v. Thus T and U are exactly the same linear transformation, a contradiction. Thus there is at most one linear transformation.

- Now we need to show that there is at least one linear transformation T for which $Tv_j = w_j$. To do this, we need to specify Tv for every vector $v \in V$, and then verify that T is linear. Well, guided by our previous arguments, we know how to find Tv: we first decompose v as a linear combination of $v_1, \ldots, v_n$

$$v = a_1 v_1 + \ldots + a_n v_n$$

and then *define* Tv by the formula above:

$$Tv := a_1 w_1 + \ldots + a_n w_n.$$

This is a well-defined construction, since the scalars $a_1, \ldots, a_n$ are unique (see the Lemma on page 36 of week 1 notes). To check that $Tv_j = w_j$, note that

$$v_j = 0v_1 + \ldots + 0v_{j-1} + 1v_j + 0v_{j+1} + \ldots + 0v_n$$

and thus by definition of T

$$Tv_j = 0w_1 + \ldots + 0w_{j-1} + 1w_j + 0v_{j+1} + \ldots + 0w_n = w_j$$

as desired.

- It remains to verify that T is linear; i.e. that $T(v + v') = Tv + Tv'$ and that $T(cv) = cTv$ for all vectors $v, v' \in V$ and scalars c.

- We'll just verify that $T(v + v') = Tv + Tv'$, and leave $T(cv) = cTv$ as an exercise. Fix any $v, v' \in V$. We can decompose

$$v = a_1 v_1 + \ldots + a_n v_n$$

and

$$v' = b_1 v_1 + \ldots + b_n v_n$$

for some scalars $a_1, \ldots, a_n, b_1, \ldots, b_n$. Thus, by definition of T,

$$Tv = a_1 w_1 + \ldots + a_n w_n$$

and

$$Tv' = b_1 w_1 + \ldots + b_n w_n$$

and thus

$$Tv + Tv' = (a_1 + b_1) w_1 + \ldots + (a_n + b_n) w_n.$$

On the other hand, adding our representations of v and v' we have

$$v + v' = (a_1 + b_1) v_1 + \ldots + (a_n + b_n) v_n$$

and thus by the definition of T again

$$T(v + v') = (a_1 + b_1) w_1 + \ldots + (a_n + b_n) w_n$$

and so $T(v + v') = Tv + Tv'$ as desired. The derivation of $T(cv) = cTv$ is similar and is left as an exercise. This completes the construction of T and the verification of the desired properties. □

- Example: We know that $\mathbf{R}^2$ has $\{(1,0),(0,1)\}$ as a basis. Thus, by Theorem 6, for any vector space W and any vectors w_1, w_2 in W, there is exactly one linear transform $T : \mathbf{R}^2 \to W$ such that $T(1,0) = w_1$ and $T(0,1) = w_2$. Indeed, this transformation is given by

  $$T(x,y) := x w_1 + y w_2$$

  (why is this transformation linear, and why does it have the desired properties?).

- Example: Let θ be an angle. Suppose we want to understand the operation $Rot_\theta : \mathbf{R}^2 \to \mathbf{R}^2$ of anti-clockwise rotation of $\mathbf{R}^2$ by θ. From elementary geometry one can see that this is a linear transformation. By some elementary trigonometry we see that $Rot_\theta(1,0) = (\cos\theta, \sin\theta)$ and $Rot_\theta(0,1) = (-\sin\theta, \cos\theta)$. Thus from the previous example, we see that

  $$Rot_\theta(x,y) = x(\cos\theta, \sin\theta) + y(-\sin\theta, \cos\theta).$$

* * * * *

Co-ordinate bases

- Of all the vector spaces, $\mathbf{R}^n$ is the easiest to work with; every vector v consists of nothing more than n separate scalars - the n *co-ordinates* of the vector. Vectors from other vector spaces - polynomials, matrices, etc. - seem to be more complicated to work with. Fortunately, by using *co-ordinate bases*, one can convert every (finite-dimensional) vector space into a space just like $\mathbf{R}^n$.

- **Definition.** Let V be a finite dimensional vector space. An *ordered basis* of V is an *ordered* sequence $(v_1, \ldots, v_n)$ of vectors in V such that the set $\{v_1, \ldots, v_n\}$ is a basis.

- **Example** The sequence $((1,0,0),(0,1,0),(0,0,1))$ is an ordered basis of $\mathbf{R}^3$; the sequence $((0,1,0),(1,0,0),(0,0,1))$ is a different ordered basis of $\mathbf{R}^3$. (Thus sequences are different from sets; rearranging the elements of a set does not affect the set).

- More generally, if we work in $\mathbf{R}^n$, and we let e_j be the vector with j^{th} co-ordinate 1 and all other co-ordinates 0, then the sequence $(e_1, e_2, \ldots, e_n)$ is an ordered basis of $\mathbf{R}^n$, and is known as the *standard ordered basis* for $\mathbf{R}^n$. In a similar spirit, $(1, x, x^2, \ldots, x^n)$ is known as the *standard ordered basis* for $P_n(\mathbf{R})$.

- Ordered bases are also called *co-ordinate bases*; we shall often give bases names such as β. The reason why we need ordered bases is so that we can refer to the *first* basis vector, *second* basis vector, etc. (In a set, which is unordered, one cannot refer to the first element, second element, etc. - they are all jumbled together and are just plain elements).

- Let $\beta = (v_1, \ldots, v_n)$ be an ordered basis for V, and let v be a vector in V. From the Lemma on page 36 of Week 1 notes, we know that v has a unique representation of the form

$$v = a_1 v_1 + \ldots + a_n v_n.$$

The scalars $a_1, \ldots, a_n$ will be referred to as the *co-ordinates of v with respect to β*, and we define the *co-ordinate vector of v relative to β*, denoted $[v]^\beta$, by

$$[v]^\beta := \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

(In the textbook, $[v]_\beta$ is used instead of $[v]^\beta$. I believe this is a mistake - there is a convention that superscripts should refer to column vectors and subscripts to row vectors - although this distinction is of course very minor. This convention becomes very useful in physics, especially when one begins to study *tensors* - a generalization of vectors and matrices - but for this course, please don't worry too much about whether an index should be a subscript or superscript.)

- **Example** Let's work in $\mathbf{R}^3$, and let $v := (3, 4, 5)$. If β is the standard ordered basis $\beta := ((1, 0, 0), (0, 1, 0), (0, 0, 1))$, then

$$[v]^\beta = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$$

since

$$(3, 4, 5) = 3(1, 0, 0) + 4(0, 1, 0) + 5(0, 0, 1).$$

On the other hand, if we use the ordered basis $\beta' := ((0, 1, 0), (1, 0, 0), (0, 0, 1))$, then

$$[v]^{\beta'} = \begin{pmatrix} 4 \\ 3 \\ 5 \end{pmatrix}$$

since

$$(3, 4, 5) = 4(0, 1, 0) + 3(1, 0, 0) + 5(0, 0, 1).$$

If instead we use the basis $\beta'' := ((3, 4, 5), (0, 1, 0), (0, 0, 1))$, then

$$[v]^{\beta'} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

since

$$(3,4,5) = 1(3,4,5) + 0(0,1,0) + 0(0,0,1).$$

For more general bases, one would probably have to do some Gaussian elimination to work out exactly what the co-ordinate vector is (similar to what we did in Week 1).

- **Example** Now let's work in $P_2(\mathbf{R})$, and let $f = 3x^2 + 4x + 6$. If β is the standard ordered basis $\beta := (1, x, x^2)$, then

$$[f]^\beta = \begin{pmatrix} 6 \\ 4 \\ 3 \end{pmatrix}$$

since

$$f = 6 \times 1 + 4 \times x + 3 \times x^2.$$

Or using the reverse standard ordered basis $\beta' := (x^2, x, 1)$, we have

$$[f]^{\beta'} = \begin{pmatrix} 3 \\ 4 \\ 6 \end{pmatrix}$$

since

$$f = 3 \times x^2 + 4 \times x + 6 \times 1.$$

Note that while $\begin{pmatrix} 6 \\ 4 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} 3 \\ 4 \\ 6 \end{pmatrix}$ are clearly different column vectors, they both came from the same object f. It's like how one person may perceive a pole as being 12 feet long and another may perceive it as being 4 yards long; both are correct, even though 12 is not equal to 4. It's just that one person is using feet as a basis for length and the other is using yards as a basis for length. (Units of measurement are to scalars as bases are to vectors. To be pedantic, the space V of all possible lengths is a one-dimensional vector space, and both $(yard)$ and $(foot)$ are bases. A length v might be equal to 4 yards, so that $[v]^{(yard)} = (4)$, while also being equal to 12 feet, so $[v]^{(foot)} = 12$).

- Given any vector v and any ordered basis β, we can construct the co-ordinate vector $[v]^\beta$. Conversely, given the co-ordinate vector

$$[v]^\beta = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

and the ordered basis $\beta = (v_1, \ldots, v_n)$, one can reconstruct v by the formula

$$v = a_1 v_1 + \ldots + a_n v_n.$$

Thus, for any fixed basis, one can go back and forth between vectors v and column vectors $[v]^\beta$ without any difficulty.

- Thus, the use of co-ordinate vectors gives us a way to represent any vector as a familiar column vector, provided that we supply a basis β. The above examples show that the choice of basis β is important; different bases give different co-ordinate vectors.

- A philosophical point: This flexibility in choosing bases underlies a basic fact about the standard Cartesian grid structure, with its x and y axes, etc: it is artificial! (though of course very convenient for computations). The plane $\mathbf{R}^2$ is a very natural object, but our Cartesian grid is not (the ancient Greeks were working with the plane back in 300 BC, but Descartes only introduced the grid in the 1700s). Why couldn't we make, for instance, the x-axis point northwest and the y-axis point northeast? This would correspond to a different basis (for instance, using $((1,1),(1,-1))$ instead of $((1,0),(0,1))$ but one could still do all of geometry, calculus, etc. perfectly well with this grid.

- (The way mathematicians describe this is: the plane is *canonical*, but the Cartesian co-ordinate system is *non-canonical*. Canonical means that there is a natural way to define this object uniquely, without recourse to any artificial convention.)

- As we will see later, it does make sense every now and then to shift one's co-ordinate system to suit the situation - for instance, the above basis $((1,1),(1,-1))$ might be useful in dealing with shapes which were always at 45 degree angles to the horizontal (i.e. diamond-shaped objects). But in the majority of cases, the standard basis suffices, if for no reason other than tradition.

- The very operation of sending a vector v to its co-ordinate vector $[v]^\beta$ is itself a linear transformation, from V to $\mathbf{R}^n$: see this week's homework.

* * * * *

The matrix representation of linear transformations

- We have just seen that by using an ordered basis of V, we can represent vectors in V as column vectors. Now we show that by using an ordered basis of V and another ordered basis of W, we can represent *linear transformations* from V to W as *matrices*. This is a very fundamental observation in this course; it means that from now on, we can study linear transformations by focusing on matrices, which is exactly what we will be doing for the rest of this course.

- Specifically, let V and W be finite-dimensional vector spaces, and let $\beta := (v_1, \ldots, v_n)$ and $\gamma = (w_1, \ldots, w_m)$ be ordered bases for V and W respectively; thus $\{v_1, \ldots, v_n\}$ is a basis for V and $\{w_1, \ldots, w_m\}$ is a basis for W, so that V is n-dimensional and W is m-dimensional. Let T be a linear transformation from V to W.

- **Example** Let $V = P_3(\mathbf{R})$, $W = P_2(\mathbf{R})$, and $T : V \to W$ be the differentiation map $Tf := f'$. We use the standard ordered basis $\beta := (1, x, x^2, x^3)$ for V, and the standard ordered basis $\gamma := (1, x, x^2)$ for W. We shall continue with this example later.

- Returning now to the general situation, let us take a vector v in V and try to compute Tv using our bases. Since v is in V, it has a co-ordinate representation

$$[v]^\beta = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

with respect to β. Similarly, since Tv is in W, it has a co-ordinate representation

$$[Tv]^\gamma = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

with respect to γ. Our question is now: how are the column vectors $[v]^\beta$ and $[Tv]^\gamma$ related? More precisely, if we know the column vector $[v]^\beta$, can we work out what $[Tv]^\gamma$ will be? Of course, the answer will depend on T; but as we shall see, we can quantify this more precisely, by saying that the answer will depend on a certain *matrix representation* of T with respect to β and γ.

- **Example** Continuing our previous example, let's pick a $v \in P_3(\mathbf{R})$ at random, say $v := 3x^2 + 7x + 5$, so that

$$[v]^\beta = \begin{pmatrix} 5 \\ 7 \\ 3 \\ 0 \end{pmatrix}.$$

Then we have $Tv = 6x + 7$, so that

$$[Tv]^\gamma = \begin{pmatrix} 7 \\ 6 \\ 0 \end{pmatrix}.$$

The question here is this: starting from the column vector $\begin{pmatrix} 5 \\ 7 \\ 3 \\ 0 \end{pmatrix}$ for v, how does one work out the column vector $\begin{pmatrix} 7 \\ 6 \\ 0 \end{pmatrix}$ for Tv?

- Return now to the general case. From our formula for $[v]_\beta$, we have

$$v = x_1 v_1 + x_2 v_2 + \ldots + x_n v_n,$$

so if we apply T to both sides we obtain

$$Tv = x_1 Tv_1 + x_2 Tv_2 + \ldots + x_n Tv_n. \tag{0.1}$$

while from our formula for $[Tv]^\gamma$ we have

$$Tv = y_1 w_1 + y_2 w_2 + \ldots + y_m w_m. \tag{0.2}$$

Now to connect the two formulae. The vectors $Tv_1, \ldots, Tv_n$ lie in W, and so they are linear combinations of $w_1, \ldots, w_m$:

$$\begin{aligned} Tv_1 &= a_{11} w_1 + a_{21} w_2 + \ldots + a_{m1} w_m \\ Tv_2 &= a_{12} w_1 + a_{22} w_2 + \ldots + a_{m2} w_m \\ &\vdots \\ Tv_n &= a_{1n} w_1 + a_{2n} w_2 + \ldots + a_{mn} w_m; \end{aligned}$$

note that the numbers $a_{11}, \ldots, a_{nm}$ are scalars that only depend on T, β, and γ (the vector v is only relevant for computing the x's and y's).

Substituting the above formulae into (0.1) we obtain

$$\begin{aligned} Tv = \; & x_1(a_{11} w_1 + \ldots + a_{m1} w_m) \\ & + x_2(a_{12} w_1 + \ldots + a_{m2} w_m) \\ & \vdots \\ & + x_n(a_{1n} w_1 + \ldots + a_{mn} w_m) \end{aligned}$$

Collecting coefficients and comparing this with (0.2) (remembering that $\{w_1, \ldots, w_m\}$ is a basis, so there is only one way to write Tv as a linear combination of $w_1, \ldots, w_m$) - we obtain

$$\begin{aligned} y_1 &= a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n \\ y_2 &= a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n \\ &\vdots \\ y_m &= a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n. \end{aligned}$$

This may look like a mess, but it becomes cleaner in matrix form:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

- Thus, if we define $[T]_\beta^\gamma$ to be the matrix

$$[T]_\beta^\gamma := \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

  then we have answered our question of how to link $[v]^\beta$ with $[Tv]^\gamma$:

$$[Tv]^\gamma = [T]_\beta^\gamma [v]^\beta.$$

  (If you like, the β subscript on the T has "cancelled" the β superscript on the v. This is part of a more general rule, known as the *Eisenstein summation convention*, which you might encounter in advanced physics courses when dealing with things called *tensors*).

- It is no co-incidence that matrices were so conveniently suitable for this problem; in fact matrices were initially invented for the express purpose of understanding linear transformations in co-ordinates.

- **Example.** We return to our previous example. Note

$$Tv_1 = T1 = 0 = 0w_1 + 0w_2 + 0w_3$$

$$Tv_2 = Tx = 1 = 1w_1 + 0w_2 + 0w_3$$

$$Tv_3 = Tx^2 = 2x = 0w_1 + 2w_2 + 0w_3$$

$$Tv_4 = Tx^3 = 3x^2 = 0w_1 + 0w_2 + 3w_3$$

  and hence

$$[T]_\beta^\gamma := \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

  Thus $[v]^\beta$ and $[Tv]^\gamma$ are linked by the equation

$$[Tv]^\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} [v]^\beta,$$

thus for instance returning to our previous example

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

- The matrix $[T]_\beta^\gamma$ is called the *matrix representation of T with respect to the bases β and γ.* Notice that the j^{th} column of $[T]_\beta^\gamma$ is just the co-ordinate vector of Tv_j with respect to γ:

$$[T]_\beta^\gamma = ([Tv_1]^\gamma \quad [Tv_2]^\gamma \quad \ldots \quad [Tv_n]^\gamma)$$

(see for instance the previous Example).

- In many cases, V will be equal to W, and β equal to γ; in that case we may abbreviate $[T]_\beta^\gamma$ as $[T]_\beta$.

- Just like a vector v can be reconstructed from its co-ordinate vector $[v]^\beta$ and vice versa (provided one knows what β is, of course), a linear transformation T can be reconstructed from its co-ordinate matrix $[T]_\beta^\gamma$ and vice versa (provided β and γ are given). Indeed, if one knows $[T]_\beta^\gamma$, then one can work out the rule to get from v to Tv as follows: first write v in terms of β, obtaining the co-ordinate vector $[v]^\beta$; multiply this column vector by $[T]_\beta^\gamma$ to obtain $[Tv]^\gamma$, and then use γ to convert this back into the vector Tv.

- **The scalar case** All this stuff may seem very abstract and foreign, but it is just the vector equivalent of something you are already familiar with in the scalar case: conversion of units. Let's give an example. Suppose a car is travelling in a straight line at a steady speed T for a period v of time (yes, the letters are strange, but this is deliberate). Then the distance that this car traverses is of course Tv. Easy enough, but now let's do everything with units.

- Let's say that the period of time v was half an hour, or thirty minutes. It is not quite accurate to say that $v = 1/2$ or $v = 30$; the precise statement (in our notation) is that $[v]^{(hour)} = (1/2)$, or $[v]^{(minute)} = (30)$. (Note that (*hour*) and (*minute*) are both ordered bases for time, which is a one-dimensional vector space). Since our bases just have one element, our column vector has only one row, which makes it a rather silly vector in our case.

- Now suppose that the speed T was twenty miles an hour. Again, it is not quite accurate to say that $T = 20$; the correct statement is that

$$[T]^{(mile)}_{(hour)} = (20)$$

since we clearly have

$$T(1 \times hour) = 20 \times mile.$$

We can also represent T in other units:

$$[T]^{(mile)}_{(minute)} = (1/3)$$

$$[T]^{(kilometer)}_{(hour)} = (32)$$

etc. In this case our "matrices" are simply 1×1 matrices - pretty boring!

Now we can work out Tv in miles or kilometers:

$$[Tv]^{(mile)} = [T]^{(mile)}_{(hour)}[v]^{(hour)} = (20)(1/2) = (10)$$

or to do things another way

$$[Tv]^{(mile)} = [T]^{(mile)}_{(minute)}[v]^{(minute)} = (1/3)(30) = (10).$$

Thus the car travels for 10 miles - which was of course obvious from the problem. The point here is that these strange matrices and bases are not alien objects - they are simply the vector versions of things that you have seen even back in elementary school mathematics.

- **A matrix example** Remember the car company example from Week 1? Let's run an example similar to that. Suppose the car company needs money and labor to make cars. To keep things very simple, let's suppose that the car company only makes exteriors - doors and wheels. Let's say that there are two types of cars: *coupes*, which have two doors and four wheels, and *sedans*, which have four doors and four wheels. Let's say that a wheel requires 2 units of money and 3 units of labor, while a door requires 4 units of money and 5 units of labor.

- We're going to have two vector spaces. The first vector space, V, is the space of *orders* - the car company may have an order v of 2 coupes and 3 sedans, which translates to 16 doors and 20 wheels. Thus

$$[v]^{(coupe,sedan)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

and

$$[v]^{(door,wheel)} = \begin{pmatrix} 16 \\ 20 \end{pmatrix};$$

both $(coupe, sedan)$ and $(door, wheel)$ are ordered bases for V. (One could also make other bases, such as $(coupe, wheel)$, although those are rather strange).

- The second vector space, W, is the space of *resources* - in this case, just money and labor. We're only going to use one ordered basis here: $(money, labor)$.

- There is an obvious linear transformation T from V to W - the *cost* (actually, *price* is a more accurate name for T; *cost* should really refer to Tv). Thus, for any order v in V, Tv is the amount of resources required to create v. By our hypotheses,

$$T(door) = 4 \times money + 5 \times labor$$

and

$$T(wheel) = 2 \times money + 3 \times labor$$

so

$$[T]^{(money,labor)}_{(door,wheel)} = \begin{pmatrix} 4 & 2 \\ 5 & 3 \end{pmatrix}.$$

You may also check that

$$[T]^{(money,labor)}_{(coupe,sedan)} = \begin{pmatrix} 16 & 24 \\ 22 & 32 \end{pmatrix}.$$

Thus, for our order v, the cost to make v can be computed as

$$[Tv]^{(money,labor)} = [T]^{(money,labor)}_{(door,wheel)}[v]^{(door,wheel)} = \begin{pmatrix} 4 & 2 \\ 5 & 3 \end{pmatrix}\begin{pmatrix} 16 \\ 20 \end{pmatrix} = \begin{pmatrix} 104 \\ 140 \end{pmatrix}$$

or equivalently as

$$[Tv]^{(money,labor)} = [T]^{(money,labor)}_{(coupe,sedan)}[v]^{(coupe,sedan)} = \begin{pmatrix} 16 & 24 \\ 22 & 32 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 104 \\ 140 \end{pmatrix} -$$

i.e. one needs 104 units of money and 140 units of labor to complete the order. (Can you explain why these two apparently distinct computations gave exactly the same answer, and why this answer is actually the correct cost of this order?). Note how the different bases $(coupe, sedan)$ and $(door, wheel)$ have different advantages and disadvantages; the $(coupe, sedan)$ basis makes the co-ordinate vector for v nice and simple, while the $(door, wheel)$ basis makes the co-ordinate matrix for T nice and simple.

* * * * *

Things to do with linear transformations

- We know that certain operations can be performed on vectors; they can be added together, or multiplied with a scalar. Now we will observe that there are similar operations on *linear transformations*; they can also be added together and multiplied by a scalar, but also (under certain conditions) can also be multiplied with each other.

- **Definition.** Let V and W be vector spaces, and let $S : V \to W$ and $T : V \to W$ be two linear transformations from V to W. We define the *sum* $S + T$ of these transformations to be a third transformation $S + T : V \to W$, defined by

$$(S + T)(v) := Sv + Tv.$$

- **Example.** Let $S : \mathbf{R}^2 \to R^2$ be the doubling transformation, defined by $Sv := 2v$. Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the identity transformation, defined by $Tv := v$. Then $S + T$ is the tripling transformation

$$(S + T)v = Sv + Tv = 2v + v = 3v.$$

- **Lemma 7** The sum of two linear transformations is again a linear transformation.

- **Proof** Let $S : V \to W$, $T : V \to W$ be linear transformations. We need to show that $S + T : V \to W$ is also linear; i.e. it preserves addition and preserves scalar multiplication. Let's just show that it preserves scalar multiplication, i.e. for any $v \in V$ and scalar c, we have to show that

  $$(S + T)(cv) = c(S + T)v.$$

  But the left-hand side, by definition, is

  $$S(cv) + T(cv) = cSv + cTv$$

  since S, T are linear. Similarly, the right-hand side is

  $$c(Sv + Tv) = cSv + cTv$$

  by the axioms of vector spaces. Thus the two are equal. The proof that $S + T$ preserves addition is similar and is left as an exercise. □

- Note that we can only add two linear transformations S, T if they have the same domain and target space; for instance it is not permitted to add the identity transformation on $\mathbf{R}^2$ to the identity transformation on $\mathbf{R}^3$. This is similar to how vectors can only be added if they belong to the same space; a vector in $\mathbf{R}^2$ cannot be added to a vector in $\mathbf{R}^3$.

- **Definition.** Let $T : V \to W$ be a linear transformation, and let c be a scalar. We define the *scalar multiplication* cT of c and T to be the transformation $cT : V \to W$, defined by

  $$(cT)(v) = c(Tv).$$

- It is easy to verify that cT is also a linear transformation; we leave this as an exercise.

- **Example** Let $S : \mathbf{R}^2 \to \mathbf{R}^2$ be the doubling transformation, defined by $Sv := 2v$. Then $2S : \mathbf{R}^2 \to \mathbf{R}^2$ is the quadrupling transformation, defined by $2Sv := 4v$.

- **Definition** Let $\mathcal{L}(V, W)$ be the space of linear transformations from V to W.

- **Example** In the examples above, the transformations S, T, $S+T$, and $2S$ all belonged to $\mathcal{L}(\mathbf{R}^2, \mathbf{R}^2)$.

- **Lemma 8** The space $\mathcal{L}(V, W)$ is a subspace of $\mathcal{F}(V, W)$, the space of all functions from V to W. In particular, $\mathcal{L}(V, W)$ is a vector space.

- **Proof** Clearly $\mathcal{L}(V, W)$ is a subset of $\mathcal{F}(V, W)$, since every linear transformation is a transformation. Also, we have seen that the space $\mathcal{L}(V, W)$ of linear transformations from V to W is closed under addition and scalar multiplication. Hence, it is a subspace of the vector space $\mathcal{F}(V, W)$, and is hence itself a vector space. (Alternatively, one could verify each of the vector space axioms (I-VIII) in turn for $\mathcal{L}(V, W)$; this is a tedious but not very difficult exercise). $\square$

- The next basic operation is that of *multiplying* or *composing* two linear transformations.

- **Definition** Let U, V, W be vector spaces. Let $S : V \to W$ be a linear transformation from V to W, and let $T : U \to V$ be a linear transformation from U to V. Then we define the *product* or *composition* $ST : U \to W$ to be the transformation

$$ST(u) := S(T(u)).$$

- **Example** Let $U : \mathbf{R}^\infty \to \mathbf{R}^\infty$ be the right shift operator

$$U(x_1, x_2, \ldots) := (0, x_1, x_2, \ldots).$$

Then the operator $UU = U^2$ is given by

$$U^2(x_1, x_2, \ldots) := U(U(x_1, x_2, \ldots)) = U(0, x_1, x_2, \ldots) = (0, 0, x_1, x_2, \ldots),$$

i.e. the double right shift.

- **Example** Let $U^* : \mathbf{R}^\infty \to \mathbf{R}^\infty$ be the left-shift operator

$$U^*(x_1, x_2, \ldots) := (x_2, x_3, \ldots).$$

Then U^*U is the identity map:

$$U^*U(x_1, x_2, \ldots) = U^*(0, x_1, x_2, \ldots) = (x_1, x_2, \ldots)$$

but UU^* is not:

$$UU^*(x_1, x_2, \ldots) = U(x_2, \ldots) = (0, x_2, \ldots).$$

Thus multiplication of operators is *not* commutative.

- Note that in order for ST to be defined, the target space of T has to match the initial space of S. (This is very closely related to the fact that in order for matrix multiplication AB to be well defined, the number of columns of A must equal the number of rows of B).

* * * * *

Addition and multiplication of matrices

- We have just defined addition, scalar multiplication, and composition of linear transformations. On the other hand, we also know how to add, scalar multiply, and multiply matrices. Since linear transformations can be represented (via bases) as matrices, it is thus a natural question as to whether the linear transform notions of addition, scalar multiplication, and composition are in fact compatible with the matrix notions of addition, scalar multiplication, and multiplication. This is indeed the case; we will now show this.

- **Lemma 9** Let V, W be finite-dimensional spaces with ordered bases β, γ respectively. Let $S : V \to W$ and $T : V \to W$ be linear transformations from V to W, and let c be a scalar. Then

$$[S + T]_\beta^\gamma = [S]_\beta^\gamma + [T]_\beta^\gamma$$

and

$$[cT]_\beta^\gamma = c[T]_\beta^\gamma.$$

- **Proof.** We'll just prove the second statement, and leave the first as an exercise. Let's write $\beta = (v_1, \ldots, v_n)$ and $\gamma = (w_1, \ldots, w_n)$, and denote the matrix $[T]_\beta^\gamma$ by

$$[T]_\beta^\gamma := \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

Thus

$$\begin{aligned}
Tv_1 &= a_{11}w_1 + a_{21}w_2 + \ldots + a_{m1}w_m \\
Tv_2 &= a_{12}w_1 + a_{22}w_2 + \ldots + a_{m2}w_m \\
&\vdots \\
Tv_n &= a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{mn}w_m;
\end{aligned}$$

Multiplying by c, we obtain

$$\begin{aligned}
(cT)v_1 &= ca_{11}w_1 + ca_{21}w_2 + \ldots + ca_{m1}w_m \\
(cT)v_2 &= ca_{12}w_1 + ca_{22}w_2 + \ldots + ca_{m2}w_m \\
&\vdots \\
(cT)v_n &= ca_{1n}w_1 + ca_{2n}w_2 + \ldots + ca_{mn}w_m
\end{aligned}$$

and thus

$$[cT]_\beta^\gamma := \begin{matrix} ca_{11} & ca_{12} & \ldots & ca_{1n} \\ ca_{21} & ca_{22} & \ldots & ca_{2n} \\ & & \vdots & \\ ca_{m1} & ca_{m2} & \ldots & ca_{mn} \end{matrix},$$

i.e. $[cT]_\beta^\gamma = c[T]_\beta^\gamma$ as desired. □

- We'll leave the corresponding statement connecting composition of linear transformations with matrix multiplication for next week's notes.