1. Linear transformations

Linear transformations are the bread and butter of Linear Algebra. You have already encountered them in Geometry I. Roughly speaking a linear transformation is a mapping between two vector spaces that preserves the linear structure of the underlying spaces. To be precise:

Definition 1.1. Let V and W be two vector spaces. A mapping $L : V \to W$ is called a *linear transformation* or a *linear mapping*, or simply, a *linear map*, if it satisfies the following two conditions:

(i) $L(\mathbf{v}+\mathbf{w}) = L(\mathbf{v}) + L(\mathbf{w})$ for all $\mathbf{v}$ and $\mathbf{w}$ in V;
(ii) $L(\alpha\mathbf{v}) = \alpha L(\mathbf{v})$ for all $\mathbf{v}$ in V and all scalars α.

Example 1.2. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$L(\mathbf{x}) = 2\mathbf{x}\,.$$

Then L is linear since, if $\mathbf{x}$ and $\mathbf{y}$ are arbitrary vectors in $\mathbb{R}^2$ and α is an arbitrary real number, then

(i) $L(\mathbf{x}+\mathbf{y}) = 2(\mathbf{x}+\mathbf{y}) = 2\mathbf{x}+2\mathbf{y} = L(\mathbf{x}) + L(\mathbf{y})$;
(ii) $L(\alpha\mathbf{x}) = 2(\alpha\mathbf{x}) = \alpha(2\mathbf{x}) = \alpha L(\mathbf{x})$.

Example 1.3. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$L(\mathbf{x}) = x_1\mathbf{e}_1\,, \qquad \text{where} \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Then L is linear. In order to see this suppose that $\mathbf{x}$ and $\mathbf{y}$ are arbitrary vectors in $\mathbb{R}^2$ with

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Notice that, if α is an arbitrary real number, then

$$\mathbf{x}+\mathbf{y} = \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \end{pmatrix} \qquad \text{and} \qquad \alpha\mathbf{x} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \end{pmatrix}.$$

Thus

(i) $L(\mathbf{x}+\mathbf{y}) = (x_1+y_1)\mathbf{e}_1 = x_1\mathbf{e}_1 + y_1\mathbf{e}_1 = L(\mathbf{x}) + L(\mathbf{y})$;
(ii) $L(\alpha\mathbf{x}) = (\alpha x_1)\mathbf{e}_1 = \alpha(x_1\mathbf{e}_1) = \alpha L(\mathbf{x})$.

Hence L is linear, as claimed.

In order to shorten statements of theorems and examples let us introduce the following convention:

> If $\mathbf{x}$ is a vector in $\mathbb{R}^n$, we shall henceforth denote its i-th entry by x_i, and similarly for vectors in $\mathbb{R}^n$ denoted by other bold symbols. So, for example, if $\mathbf{y} = (1,4,2,7)^T \in \mathbb{R}^4$, then $y_3 = 2$.

Example 1.4. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$L(\mathbf{x}) = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} .$$

L is linear, since, if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$, then

(i) $L(\mathbf{x}+\mathbf{y}) = \begin{pmatrix} -(x_2+y_2) \\ x_1+y_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} + \begin{pmatrix} -y_2 \\ y_1 \end{pmatrix} = L(\mathbf{x}) + L(\mathbf{y})$;

(ii) $L(\alpha\mathbf{x}) = \begin{pmatrix} -\alpha x_2 \\ \alpha x_1 \end{pmatrix} = \alpha \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} = \alpha L(\mathbf{x})$.

Example 1.5. The mapping $M : \mathbb{R}^2 \to \mathbb{R}^1$ defined by

$$M(\mathbf{x}) = \sqrt{x_1^2 + x_2^2}$$

is not linear. Indeed, $M((1,0)^T) = \sqrt{1^2} = 1$ while $M(-(1,0)^T) = M((-1,0)^T) = \sqrt{(-1)^2} = 1$. Thus

$$M(-(1,0)^T) = 1 \neq -1 = -M((1,0)^T) .$$

Any $m \times n$ matrix A induces a linear transformation $L_A : \mathbb{R}^n \to \mathbb{R}^m$ given by

$$L_A(\mathbf{x}) = A\mathbf{x} \quad \text{for each } \mathbf{x} \in \mathbb{R}^n .$$

The transformation L_A is linear, since, if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, then

(i) $L_A(\mathbf{x}+\mathbf{y}) = A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} = L_A(\mathbf{x}) + L_A(\mathbf{y})$;

(ii) $L_A(\alpha\mathbf{x}) = A(\alpha\mathbf{x}) = \alpha A\mathbf{x} = \alpha L_A(\mathbf{x})$.

In other words, every $m \times n$ matrix gives rise to a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. We shall see shortly that, conversely, *every* linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ arises from an $m \times n$ matrix.

Theorem 1.6. *If V and W are vector spaces and $L : V \to W$ is a linear transformation, then*

(a) $L(\mathbf{0}) = \mathbf{0}$*;*

(b) $L(-\mathbf{v}) = -L(\mathbf{v})$ *for any* $\mathbf{v} \in V$*;*

(c) $L(\sum_{i=1}^n \alpha_i \mathbf{v}_i) = \sum_{i=1}^n \alpha_i L(\mathbf{v}_i)$ *for any* $\mathbf{v}_i \in V$ *and any scalars* α_i *where* $i = 1, \ldots, n$*.*

Proof.

(a) $L(\mathbf{0}) = L(0\mathbf{0}) = 0L(\mathbf{0}) = \mathbf{0}$;

(b) $L(-\mathbf{v}) = L((-1)\mathbf{v}) = (-1)L(\mathbf{v}) = -L(\mathbf{v})$;

(c) follows by repeated application of the defining properties (i) and (ii) of linear transformations.

□

Let's look at some examples, which should convince you that linear transformations arise naturally in other areas of Mathematics.

Example 1.7. Let $L : C[a,b] \to \mathbb{R}^1$ be defined by

$$L(\mathbf{f}) = \int_a^b \mathbf{f}(t)\,dt\,.$$

L is linear since, if $\mathbf{f}, \mathbf{g} \in C[a,b]$ and $\alpha \in \mathbb{R}$, then

(i) $L(\mathbf{f}+\mathbf{g}) = \displaystyle\int_a^b (\mathbf{f}(t)+\mathbf{g}(t))\,dt = \int_a^b \mathbf{f}(t)\,dt + \int_a^b \mathbf{g}(t)\,dt = L(\mathbf{f}) + L(\mathbf{g})$;

(ii) $L(\alpha\mathbf{f}) = \displaystyle\int_a^b (\alpha\mathbf{f}(t))\,dt = \alpha \int_a^b \mathbf{f}(t)\,dt = \alpha L(\mathbf{f})$.

In other words, integration is a linear transformation.

Example 1.8. Let $C^1(a,b)$ be the real vector space of real continuously differentiable functions on the open interval (a,b) in $\mathbb{R}$. Let $D : C^1(a,b) \to C(a,b)$ be defined to be the transformation that sends an $\mathbf{f} \in C^1(a,b)$ to its derivative $\mathbf{f}' \in C(a,b)$, that is,

$$D(\mathbf{f}) = \mathbf{f}'\,.$$

Then D is linear since, if $\mathbf{f}, \mathbf{g} \in C^1(a,b)$ and $\alpha \in \mathbb{R}$, then

(i) $D(\mathbf{f}+\mathbf{g}) = (\mathbf{f}+\mathbf{g})' = \mathbf{f}' + \mathbf{g}' = D(\mathbf{f}) + D(\mathbf{g})$;
(ii) $D(\alpha\mathbf{f}) = (\alpha\mathbf{f})' = \alpha\mathbf{f}' = \alpha D(\mathbf{f})$.

In other words, differentiation is a linear transformation.

Example 1.9. Let V be a vector space and let $Id : V \to V$ denote the *identity map* on V, that is,

$$Id(\mathbf{v}) = \mathbf{v} \quad \text{for all } \mathbf{v} \in V.$$

The transformation Id is linear, since, if $\mathbf{v}, \mathbf{w} \in V$ and α is a scalar, then

(i) $Id(\mathbf{v}+\mathbf{w}) = \mathbf{v}+\mathbf{w} = Id(\mathbf{v}) + Id(\mathbf{w})$;
(ii) $Id(\alpha\mathbf{v}) = \alpha\mathbf{v} = \alpha Id(\mathbf{v})$.

2. Image and Kernel

Definition 2.1. Let V and W be vector spaces, and let $L : V \to W$ be a linear transformation. The *kernel* of L, denoted by $\ker(L)$, is the subset of V given by

$$\ker(L) = \{\, \mathbf{v} \in V \mid L(\mathbf{v}) = \mathbf{0} \,\}\,.$$

Example 2.2. If $A \in \mathbb{R}^{m\times n}$ and L_A is the corresponding linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, then

$$\ker(L_A) = N(A)\,,$$

that is, the kernel of L_A is the nullspace of A.

The previous example shows that the kernel of a linear transformation is the natural generalisation of the nullspace of a matrix.

Definition 2.3. Let V and W be vector spaces. Let $L : V \to W$ be a linear transformation and let H be a subspace of V. The *image* of H (under L), denoted by $L(H)$, is the subset of W given by

$$L(H) = \{ \mathbf{w} \in W \mid \mathbf{w} = L(\mathbf{v}) \text{ for some } \mathbf{v} \in H \} .$$

The image $L(V)$ of the entire vector space V under L is called the *range* of L.

Example 2.4. If $A \in \mathbb{R}^{m \times n}$ and L_A is the corresponding linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, then

$$L_A(\mathbb{R}^n) = \mathrm{col}(A) ,$$

that is, the range of L_A is the column space of A.

The previous example shows that the range of a linear transformation is the natural generalisation of the column space of a matrix.

We saw previously that the nullspace and the column space of an $m \times n$ matrix are subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. The same is true for the abstract analogues introduced above.

Theorem 2.5. *Let V and W and be vector spaces. If $L : V \to W$ is a linear transformation and H is a subspace of V, then*

(a) $\ker(L)$ *is a subspace of V;*
(b) $L(H)$ *is a subspace of W.*

Proof.

(a) First observe that $\ker(L)$ is not empty since $\mathbf{0} \in \ker(L)$ by Theorem 1.6. Suppose now that $\mathbf{v}_1, \mathbf{v}_2 \in \ker(L)$. Then

$$L(\mathbf{v}_1 + \mathbf{v}_2) = L(\mathbf{v}_1) + L(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0} ,$$

so $\mathbf{v}_1 + \mathbf{v}_2 \in \ker(L)$. Moreover, if $\mathbf{v} \in \ker(L)$ and α is a scalar, then

$$L(\alpha\mathbf{v}) = \alpha L(\mathbf{v}) = \alpha\mathbf{0} = \mathbf{0} ,$$

so $\alpha\mathbf{v} \in \ker(L)$. Thus, as $\ker(L)$ is closed under addition and scalar multiplication, it is a subspace of V as claimed.

(b) First observe that $L(H)$ is not empty since $\mathbf{0} \in L(H)$ by Theorem 1.6. Suppose now that $\mathbf{w}_1, \mathbf{w}_2 \in L(H)$. Then there are $\mathbf{v}_1, \mathbf{v}_2 \in H$ such that $L(\mathbf{v}_1) = \mathbf{w}_1$ and $L(\mathbf{v}_2) = \mathbf{w}_2$ and so

$$\mathbf{w}_1 + \mathbf{w}_2 = L(\mathbf{v}_1) + L(\mathbf{v}_2) = L(\mathbf{v}_1 + \mathbf{v}_2) .$$

But $\mathbf{v}_1 + \mathbf{v}_2 \in H$, because H is a subspace, so $\mathbf{w}_1 + \mathbf{w}_2 \in L(H)$. Moreover, if $\mathbf{w} \in L(H)$ and α is a scalar, then there is $\mathbf{v} \in H$ such that $L(\mathbf{v}) = \mathbf{w}$ and so

$$\alpha\mathbf{w} = \alpha L(\mathbf{v}) = L(\alpha\mathbf{v}) .$$

But $\alpha\mathbf{v} \in H$, because H is a subspace, so $\alpha\mathbf{w} \in L(H)$. Thus, as $L(H)$ is closed under addition and scalar multiplication, it is a subspace of W as claimed.

□

Example 2.6. Let $D : P_3 \to P_3$ be the differentiation transformation given by

$$D(\mathbf{p}) = \mathbf{p}' .$$

Find $\ker(D)$ and $D(P_3)$.

Solution. The derivative of a polynomial $\mathbf{p} \in P_3$ is the zero polynomial if and only if $\mathbf{p}$ is a constant. Thus

$$\ker(D) = P_0 .$$

Since differentiation lowers the degree of a polynomial by 1, we see that $D(P_3)$ is a subspace of P_2. However, any polynomial in P_2 has an antiderivative in P_3, so every polynomial in P_2 will be the image of a polynomial in P_3 under D. Thus

$$D(P_3) = P_2 .$$

□

Example 2.7. Let $L : C^1(1, 1) \to C(-1, 1)$ be the linear transformation given by

$$L(\mathbf{f}) = \mathbf{f} + \mathbf{f}' .$$

Find the kernel and range of L.

Solution. To determine the kernel of L we need to find all $\mathbf{f} \in C^1(-1, 1)$ such that

$$\mathbf{f} + \mathbf{f}' = \mathbf{0} . \tag{1}$$

This is a first order homogeneous differential equation with integrating factor e^t. Thus, a function $\mathbf{f}$ satisfies (1) if and only if

$$\frac{d}{dt}(e^t \mathbf{f}(t)) = 0 ,$$

so

$$e^t \mathbf{f}(t) = \alpha ,$$

for some $\alpha \in \mathbb{R}$, and hence

$$\mathbf{f}(t) = \alpha e^{-t} .$$

Thus

$$\ker(L) = \mathrm{Span}\,(\mathbf{h}) ,$$

where $\mathbf{h}(t) = e^{-t}$.

To determine the range of L, notice that L clearly sends continuously differentiable functions to continuous functions. The question is whether *every* continuous function arises as an image of some $\mathbf{f} \in C^1(-1, 1)$ under L. The answer is yes! To see this, fix $\mathbf{g} \in C(-1, 1)$. We need to show that there is an $\mathbf{f} \in C^1(-1, 1)$ such that $L(\mathbf{f}) = \mathbf{g}$, that is,

$$\mathbf{f}' + \mathbf{f} = \mathbf{g} . \tag{2}$$

This is a first order linear inhomogeneous differential equation for $\mathbf{f}$. Using the integrating factor e^t we find that (2) is equivalent to

$$\frac{d}{dt}(e^t\mathbf{f}(t)) = e^t\mathbf{g}(t)\,.$$

But the right hand side of the equation above has an antiderivative, say $\mathbf{H}$, that is,

$$\frac{d}{dt}\mathbf{H}(t) = e^t\mathbf{g}(t)$$

so

$$e^t\mathbf{f}(t) = \mathbf{H}(t) + \alpha\,,$$

for some $\alpha \in \mathbb{R}$, hence

$$\mathbf{f}(t) = e^{-t}\mathbf{H}(t) + \alpha e^{-t}\,.$$

Notice that the $\mathbf{f}$ just found is clearly continuously differentiable.

□

3. Matrix representations of linear transformations

Let $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis of a vector space V. Each vector $\mathbf{v} \in V$ can be written uniquely as a linear combination of vectors in B:

$$v = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$$

where $\alpha_1, \dots, \alpha_n$ are scalars. We call $(\alpha_1, \dots, \alpha_n)$ the *coordinate row vector* of $\mathbf{v}$ with respect to B. The *coordinate column vector* of $\mathbf{v}$ w.r.t. B is denoted by

$$[\mathbf{v}]_B = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = (\alpha_1, \dots, \alpha_n)^T.$$

Theorem 3.1. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then there is an $m \times n$ matrix $A \in \mathbb{R}^{m\times n}$ such that $L = L_A$, that is,*

$$L(\mathbf{x}) = L_A(\mathbf{x}) = A\mathbf{x} \quad \textit{for each } \mathbf{x} \in \mathbb{R}^n.$$

Proof. Let $B_{\mathbb{R}^n} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{R}^n$. For $j = 1, \dots, n$, we have

$$L(\mathbf{e}_j) = \begin{pmatrix} a_{ij} \\ \vdots \\ a_{nj} \end{pmatrix} \in \mathbb{R}^m.$$

Let $A \in \mathbb{R}^{m\times n}$ be the matrix whose j-th column is $L(\mathbf{e}_i)$:

$$A = (L(\mathbf{e}_1) \cdots L(\mathbf{e}_n)).$$

Then it can be easily verified that $L(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. □

Let $B_{\mathbb{R}^m}$ be the standard basis of $\mathbb{R}^m$. In the above theorem, the coordinate column vector $[L(\mathbf{e}_j)]_{B_{\mathbb{R}^m}}$ of $L(\mathbf{e}_j)$ with respect to the standard basis $B_{\mathbb{R}^m}$ is itself:

$$\begin{pmatrix} a_{ij} \\ \vdots \\ a_{nj} \end{pmatrix}.$$

The above matrix A is call a *representing matrix* of L and, as seen below, it is *the matrix representing L with respect to the standard bases $B_{\mathbb{R}^n}$ and $B_{\mathbb{R}^m}$.*

Example 3.2. Let $L : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$L(\mathbf{x}) = \begin{pmatrix} x_1 - x_2 \\ x_2 + 2x_3 \end{pmatrix}.$$

The transformation L is easily seen to be linear. Now

$$L(\mathbf{e}_1) = \begin{pmatrix} 1 & - & 0 \\ 0 & + & 2\cdot 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$L(\mathbf{e}_2) = \begin{pmatrix} 0 & - & 1 \\ 1 & + & 2\cdot 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$
$$L(\mathbf{e}_3) = \begin{pmatrix} 0 & - & 0 \\ 0 & + & 2\cdot 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

so if we set

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix},$$

then indeed

$$A\mathbf{x} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 - x_2 \\ x_2 + 2x_3 \end{pmatrix} = L(\mathbf{x}).$$

We now extend the matrix representation of a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ in the above theorem to any linear map L between finite dimensional vector spaces.

Theorem 3.3 (Matrix Representation Theorem)**.** *Let $B_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $B_W = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be bases of vector spaces V and W respectively, and let $L : V \to W$ be a linear map. Then there is an $m \times n$ matrix A such that*

$$[L(\mathbf{v})]_{B_W} = A[\mathbf{v}]_{B_V} \quad \textit{for each } \mathbf{v} \in V.$$

Proof. The construction of the matrix A is the same as before. For each $j = 1, \ldots n$, apply the map L to the basis vector $\mathbf{v}_j$ and get the coordinate column vector $[L(\mathbf{v}_j)]_{B_W}$ with respect to the basis B_W:

$$[L(\mathbf{v}_j)]_{B_W} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Form the matrix A with these column vectors:

$$A = ([L(\mathbf{v}_1)]_{B_W} \cdots [L(\mathbf{v}_n)]_{B_W}).$$

Then we have

$$[L(\mathbf{v})]_{B_W} = A[\mathbf{v}]_{B_V}.$$

□

Definition 3.4. Given vector spaces V and W with corresponding bases B_V and B_W, and a linear transformation $L : V \to W$, we call the matrix A constructed in the theorem above the **matrix representation of L with respect to B_V and B_W**, and denote it by $[L, B_v, B_W]$. Thus, for any $\mathbf{v} \in V$ we have

$$[L(\mathbf{v})]_{B_W} = [L, B_V, B_W][\mathbf{v}]_{B_V}.$$

Example 3.5. Let $L : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$L(\mathbf{x}) = x_1\mathbf{b}_1 + (x_2 + x_3)\mathbf{b}_2,$$

where

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Find the matrix representation of L with respect to the standard basis $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and the basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$.

Solution. Since

$$\begin{array}{rcrcr} L(\mathbf{e}_1) & = & 1\mathbf{b}_1 & + & 0\mathbf{b}_2 \\ L(\mathbf{e}_2) & = & 0\mathbf{b}_1 & + & 1\mathbf{b}_2 \\ L(\mathbf{e}_3) & = & 0\mathbf{b}_1 & + & 1\mathbf{b}_2 \end{array}$$

we see that

$$[L(\mathbf{e}_1)]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad [L(\mathbf{e}_2)]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad [L(\mathbf{e}_3)]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

so

$$[L, \mathcal{E}, \mathcal{B}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

□

Example 3.6. Consider the linear transformation $D : P_2 \to P_1$ given by

$$(D(\mathbf{p}))(t) = \mathbf{p}'(t).$$

Define

$$\mathbf{p}_1(t) = 1, \quad \mathbf{p}_2(t) = t, \quad \mathbf{p}_3(t) = t^2,$$

and let $\mathcal{P}_2 = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ and $\mathcal{P}_1 = \{\mathbf{p}_1, \mathbf{p}_2\}$ be bases for P_2 and P_1 respectively. Since

$$(D(\mathbf{p}_1))(t) = \mathbf{p}_1'(t) = 0$$
$$(D(\mathbf{p}_2))(t) = \mathbf{p}_2'(t) = 1$$
$$(D(\mathbf{p}_3))(t) = \mathbf{p}_3'(t) = 2t$$

we have

$$\begin{array}{rcrcr} D(\mathbf{p}_1) & = & 0\mathbf{p}_1 & + & 0\mathbf{p}_2 \\ D(\mathbf{p}_2) & = & 1\mathbf{p}_1 & + & 0\mathbf{p}_2 \\ D(\mathbf{p}_3) & = & 0\mathbf{p}_1 & + & 2\mathbf{p}_2 \end{array}$$

so

$$[D, \mathcal{P}_2, \mathcal{P}_1] = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} .$$

Suppose now that $\mathbf{p} \in P_2$ is given by

$$\mathbf{p}(t) = a + bt + ct^2 .$$

We want to find $D(\mathbf{p})$. Of course we could do this working directly from the definition of D, but we can also use the Matrix Representation Theorem: since

$$\mathbf{p} = a\mathbf{p}_1 + b\mathbf{p}_2 + c\mathbf{p}_3 ,$$

we have, by the Matrix Representation Theorem,

$$[D(\mathbf{p})]_{\mathcal{P}_1} = [D, \mathcal{P}_2, \mathcal{P}_1][\mathbf{p}]_{\mathcal{P}_2} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} b \\ 2c \end{pmatrix} ,$$

so

$$D(\mathbf{p}) = b\mathbf{p}_1 + 2c\mathbf{p}_2 ,$$

that is,

$$\mathbf{p}'(t) = b + 2ct ,$$

as expected.

Example 3.7. Let $A \in \mathbb{R}^{m\times n}$ and let L_A be the corresponding linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Since $L_A(\mathbf{e}_j)$ is just the j-th column of A, we see that the matrix representation of L_A with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ is just A itself.

4. Composition of linear transformations

Suppose that U, V and W are vector spaces and that we are given two linear transformations

$$T : U \to V ,$$
$$S : V \to W .$$

We can then form a new transformation $S \circ T : U \to W$ by defining

$$(S \circ T)(\mathbf{u}) = S(T(\mathbf{u})) \quad \text{for each } \mathbf{u} \in U.$$

The transformation $S \circ T$ is called the *composite* of S and T. Observe that $S \circ T$ is linear as well. In order to see this, let $\mathbf{u}_1, \mathbf{u}_2 \in U$ and α_1, α_2 be scalars. Then

$$\begin{aligned}(S \circ T)(\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2) &= S(T(\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2)) \\ &= S(\alpha_1 T(\mathbf{u}_1) + \alpha_2 T(\mathbf{u}_2)) \\ &= \alpha_1 S(T(\mathbf{u}_1)) + \alpha_2 S(T(\mathbf{u}_2)) \\ &= \alpha_1 (S \circ T)(\mathbf{u}_1) + \alpha_2 (S \circ T)(\mathbf{u}_2)\,.\end{aligned}$$

Choosing $\alpha_1 = \alpha_2 = 1$ in the above equality gives

$$(S \circ T)(\mathbf{u}_1 + \mathbf{u}_2) = (S \circ T)(\mathbf{u}_1) + (S \circ T)(\mathbf{u}_2)\,,$$

while choosing $\alpha_1 = 1$ and $\alpha_2 = 0$ gives

$$(S \circ T)(\alpha_1 \mathbf{u}_1) = \alpha_1 (S \circ T)(\mathbf{u}_1)\,,$$

so $S \circ T$ is linear, as claimed.

Example 4.1. Suppose that $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times r}$. Let $L_A : \mathbb{R}^n \to \mathbb{R}^m$ and $L_B : \mathbb{R}^r \to \mathbb{R}^n$ be the corresponding linear transformations. Then $L_A \circ L_B : \mathbb{R}^r \to \mathbb{R}^m$ is the linear transformation given by

$$(L_A \circ L_B)(\mathbf{x}) = L_A(L_B(\mathbf{x})) = L_A(B\mathbf{x}) = AB\mathbf{x}\,,$$

so

$$L_A \circ L_B = L_{AB}\,.$$

In other words, the composite of L_A and L_B is the linear transformation arising from the product AB.

5. Change of basis

In this section, we shall consider the problem of how the matrix representation of a given linear transformation changes when the bases of the underlying vector spaces are changed.

Question.

Let V be a finite dimensional vector space with bases $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $B' = \{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$.

Given $\mathbf{v} \in V$ and its B-coordinate vector $[\mathbf{v}]_B$, how is the B'-coordinate vector $[\mathbf{v}]_{B'}$ of $\mathbf{v}$ related to $[\mathbf{v}]_B$?

The answer is as follows. We have

$$\begin{array}{rcl}\mathbf{v}_1 &=& c_{11}\mathbf{v}'_1 + c_{12}\mathbf{v}'_2 + \cdots + c_{1n}\mathbf{v}'_n \\ \cdot & \cdot & \cdots \\ \mathbf{v}_n &=& c_{n1}\mathbf{v}'_1 + c_{n2}\mathbf{v}'_2 + \cdots + c_{nn}\mathbf{v}'_n.\end{array}$$

Let $M_B^{B'}$ be the matrix formed by the coordinate vectors $[\mathbf{v}_1]_{B'}, \cdots, [\mathbf{v}_n]_{B'}$:

$$M_B^{B'} = ([\mathbf{v}_1]_{B'} \cdots [\mathbf{v}_n]_{B'}) = \begin{pmatrix} c_{11} & \cdots & c_{n1} \\ c_{12} & \cdots & c_{n2} \\ \vdots & & \vdots \\ c_{1n} & \cdots & c_{nn} \end{pmatrix}$$

which is called the *change-of-basis matrix* or *transistion matrix from B to B'*. We note that

$$\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_n \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{n1} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix} \begin{pmatrix} \mathbf{v}'_1 \\ \vdots \\ \mathbf{v}'_n \end{pmatrix} = (M_B^{B'})^T \begin{pmatrix} \mathbf{v}'_1 \\ \vdots \\ \mathbf{v}'_n \end{pmatrix}.$$

For any vector $\mathbf{v} \in V$, we have

$$[\mathbf{v}]_{B'} = M_B^{B'} [\mathbf{v}]_B.$$

Note that the matrix $M_B^{B'}$ is invertible with inverse $M_{B'}^B$ and also

$$[\mathbf{v}]_B = M_{B'}^B [\mathbf{v}]_{B'}.$$

Theorem 5.1. *Let V be a vector space with bases B_V and B'_V. Let W be a vector space with bases B_W and B'_W. Let $L : V \to W$ be a linear map with two matrix representations $[L, B_V, B_W]$ and $[L, B'_V, B'_W]$. Then they are related by the change-of-basis matrices as follows:*

$$[L, B'_V, B'_W] = M_{B_W}^{B'_W} [L, B_V, B_W] M_{B'_V}^{B_V}.$$

In particular, if $W = V$, we have

$$[L, B'_V, B'_V] = M_{B_V}^{B'_V} [L, B_V, B_V] M_{B'_V}^{B_V} = (M_{B'_V}^{B_V})^{-1} [L, B_V, B_V] M_{B'_V}^{B_V}.$$

.

Definition 5.2. Let A and B be two $n \times n$ matrices. The matrix B is said to be **similar to** A if there is an $n \times n$ invertible matrix S such that

$$B = S^{-1} A S.$$

Notice that if B is similar to A, then A is similar to B, because if $R = S^{-1}$, then

$$A = SBS^{-1} = R^{-1} B R.$$

Thus we may simply say that A and B are similar matrices.

Letting $A = [L, B'_V, B'_V]$ and $B = [L, B_V, B_V]$ in Theorem 5.1, the content of the theorem can now be rephrased as follows: if A and B are two matrix representations of the same linear transformation $L : V \to V$ on a vector space V, then A and B are similar.

6. Inner product and orthogonality in $\mathbb{R}^n$

We now return to the concrete vector space $\mathbb{R}^n$ and introduce two new concepts, namely, the notion of an inner product and orthogonality. The latter extends our intuitive notion of perpendicularity in $\mathbb{R}^2$ and $\mathbb{R}^3$ to $\mathbb{R}^n$. We regard a 1×1 matrix as a scalar.

Definition 6.1. Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in $\mathbb{R}^n$. The scalar $\mathbf{x}^T\mathbf{y}$ is called the *standard inner product*, or *scalar product*, or *dot product*, of $\mathbf{x}$ and $\mathbf{y}$ and is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$ or $\mathbf{x}\cdot\mathbf{y}$. Thus, if

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

then

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}\cdot\mathbf{y} = \mathbf{x}^T\mathbf{y} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n\,.$$

Example 6.2. If

$$\mathbf{x} = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix},$$

then

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y} = \begin{pmatrix} 2 & -3 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = 2 \cdot 4 + (-3) \cdot 5 + 1 \cdot 6 = 8 - 15 + 6 = -1\,.$$

Having had a second look at the example above it should be clear why $\mathbf{x}\cdot\mathbf{y} = \mathbf{y}\cdot\mathbf{x}$. In fact, this is true in general. The following further properties of the dot product follow easily from properties of the transpose operation:

Theorem 6.3. *Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ be vectors in $\mathbb{R}^n$, and let α be a scalar. Then*

(a) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$;
(b) $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$;
(c) $\langle \alpha\mathbf{x}, \mathbf{y} \rangle = \alpha\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \alpha\mathbf{y} \rangle$;
(d) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, *and* $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

We call the vector space $\mathbb{R}^n$, equipped with the standard inner product $\langle \cdot, \cdot \rangle$, the *$n$-dimensional Euclidean space.*

Definition 6.4. If $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, the **length** or *norm* of $\mathbf{x}$ is the nonnegative scalar $\|\mathbf{x}\|$ defined by

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{x_1^2 + \cdots + x_n^2}\,.$$

A vector whose length is 1 is called a *unit vector* .

Example 6.5. If $\mathbf{x} = (a, b)^T \in \mathbb{R}^2$, then

$$\|\mathbf{x}\| = \sqrt{a^2 + b^2}\,.$$

The above example should convince you that in $\mathbb{R}^2$ and $\mathbb{R}^3$ the definition of the length of a vector $\mathbf{x}$ coincides with the standard notion of the length of the line segment from the origin to $\mathbf{x}$.

Note that if $\mathbf{x} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ then

$$\|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|\,,$$

because $\|\alpha\mathbf{x}\|^2 = (\alpha\mathbf{x})\cdot(\alpha\mathbf{x}) = \alpha^2(\mathbf{x}\cdot\mathbf{x}) = \alpha^2\|\mathbf{x}\|^2$. Thus, if $\mathbf{x} \neq \mathbf{0}$, we can always find a unit vector $\mathbf{y}$ in the same direction as $\mathbf{x}$ by setting

$$\mathbf{y} = \frac{1}{\|\mathbf{x}\|}\mathbf{x}\,.$$

The process of creating a unit vector $\mathbf{y}$ from $\mathbf{x}$ is called *normalising* $\mathbf{x}$.

Definition 6.6. For $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, the Euclidean **distance between $\mathbf{x}$ and $\mathbf{y}$**, written $\operatorname{dist}(\mathbf{x}, \mathbf{y})$, is the length of $\mathbf{x} - \mathbf{y}$, that is,

$$\operatorname{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|\,.$$

Definition 6.7. Two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

Note that the zero vector is orthogonal to every other vector in $\mathbb{R}^n$.

Theorem 6.8 (Pythagorean Theorem)**.** *Two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ are orthogonal if and only if*

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\,.$$

Proof. We have

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= \langle \mathbf{x} + \mathbf{y},\, \mathbf{x} + \mathbf{y} \rangle \\ &= \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \\ &= \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2. \end{aligned}$$

Hence $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ if and only if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. □

Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in $\mathbb{R}^2$ forming an angle $\theta \leq \pi$. By the cosine law, we have

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$$

which gives

$$\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta.$$

Definition 6.9. A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ in $\mathbb{R}^n$ is said to be an *orthogonal set* if each pair of distinct vectors is orthogonal, that is, if

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0 \quad \text{whenever } i \neq j\,.$$

Example 6.10. If

$$\mathbf{u}_1 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} -1 \\ -4 \\ 7 \end{pmatrix},$$

then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set since

$$\begin{aligned} \mathbf{u}_1 \cdot \mathbf{u}_2 &= 3 \cdot (-1) + 1 \cdot 2 + 1 \cdot 1 = 0 \\ \mathbf{u}_1 \cdot \mathbf{u}_3 &= 3 \cdot (-1) + 1 \cdot (-4) + 1 \cdot 7 = 0 \\ \mathbf{u}_2 \cdot \mathbf{u}_3 &= (-1) \cdot (-1) + 2 \cdot (-4) + 1 \cdot 7 = 0 \end{aligned}$$

Theorem 6.11. *If* $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ *is an orthogonal set of nonzero vectors, then the vectors* $\mathbf{u}_1, \dots, \mathbf{u}_r$ *are linearly independent.*

Proof. Suppose that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_r\mathbf{u}_r = \mathbf{0}\,.$$

Then

$$\begin{aligned} 0 &= \mathbf{0} \cdot \mathbf{u}_1 \\ &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_r\mathbf{u}_r) \cdot \mathbf{u}_1 \\ &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_1) + \cdots + c_r(\mathbf{u}_r \cdot \mathbf{u}_1) \\ &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)\,, \end{aligned}$$

since $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \dots, \mathbf{u}_r$. But since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1 \cdot \mathbf{u}_1$ is nonzero, so $c_1 = 0$. Similarly, $c_2, \dots, c_r$ must be zero, and the assertion follows. □

Definition 6.12. An *orthogonal basis* for a subspace H of $\mathbb{R}^n$ is a basis of H that is also an orthogonal set.

Definition 6.13. A set $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ of vectors in $\mathbb{R}^n$ is called an *orthonormal set* if it is an orthogonal set of unit vectors. In other words, $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ is an orthonormal set if and only if

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij} \quad \text{for } i, j = 1, \dots, r\,,$$

where

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

An *orthonormal basis* of a subspace $H \subset \mathbb{R}^n$ is a basis $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ of H which is an orthonormal set.

Example 6.14. The standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of $\mathbb{R}^n$ is an orthonormal set (and also an orthonormal basis for $\mathbb{R}^n$). Moreover, any nonempty subset of $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is an orthonormal set.

Here is a less trivial example:

Example 6.15. If

$$\mathbf{u}_1 = \begin{pmatrix} 2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} -1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix},$$

then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal set, since

$$\begin{aligned} \mathbf{u}_1 \cdot \mathbf{u}_2 &= -2/\sqrt{18} + 1/\sqrt{18} + 1/\sqrt{18} = 0 \\ \mathbf{u}_1 \cdot \mathbf{u}_3 &= 0/\sqrt{12} - 1/\sqrt{12} + 1/\sqrt{12} = 0 \\ \mathbf{u}_2 \cdot \mathbf{u}_3 &= 0/\sqrt{6} - 1/\sqrt{6} + 1/\sqrt{6} = 0 \end{aligned}$$

and

$$\begin{aligned} \mathbf{u}_1 \cdot \mathbf{u}_1 &= 4/6 + 1/6 + 1/6 = 1 \\ \mathbf{u}_2 \cdot \mathbf{u}_2 &= 1/3 + 1/3 + 1/3 = 1 \\ \mathbf{u}_3 \cdot \mathbf{u}_3 &= 0/2 + 1/2 + 1/2 = 1 \end{aligned}$$

Moreover, since by Theorem 6.11 the vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are linearly independent and $\dim \mathbb{R}^3 = 3$, the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a basis for $\mathbb{R}^3$. Thus $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis for $\mathbb{R}^3$.

7. Orthogonal complements

Definition 7.1. Let Y be a subset of $\mathbb{R}^n$. A vector $\mathbf{x} \in \mathbb{R}^n$ is said to be **orthogonal to** Y if $\mathbf{x}$ is orthogonal to every vector in Y. The set of all vectors in $\mathbb{R}^n$ that are orthogonal to Y is called the **orthogonal complement of** Y and is denoted by $Y^\perp$ (pronounced 'Y perpendicular' or 'Y perp' for short). Thus

$$Y^\perp = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{y} = 0 \text{ for all } \mathbf{y} \in Y \,\} .$$

Example 7.2. Let W be a plane through the origin in $\mathbb{R}^3$ and let L be the line through the origin and perpendicular to W. By construction, each vector in W is orthogonal to every vector in L, and each vector in L is orthogonal to every vector in W. Hence

$$L^\perp = W \quad \text{and} \quad W^\perp = L .$$

Theorem 7.3. (a) *Let Y be a subset of $\mathbb{R}^n$. Then $Y^\perp$ is a subspace of $\mathbb{R}^n$.*

(b) *Let Y be a subspace of $\mathbb{R}^n$. Then a vector $\mathbf{x}$ belongs to $Y^\perp$ if and only if $\mathbf{x}$ is orthogonal to every vector in any spanning set of Y.*

Theorem 7.4 (Fundamental Subspace Theorem)**.** *Let $A \in \mathbb{R}^{m\times n}$. Then:*

(a) $N(A) = \mathrm{col}(A^T)^\perp \subset \mathbb{R}^m$.
(b) $N(A^T) = \mathrm{col}(A)^\perp \subset \mathbb{R}^n$.

Proof. In this proof we shall identify the rows of A (which are strictly speaking $1 \times n$ matrices) with vectors in $\mathbb{R}^n$.

(a) Let $\mathbf{x} \in \mathbb{R}^n$. Then

$$\begin{aligned} \mathbf{x} \in N(A) &\iff A\mathbf{x} = \mathbf{0} \\ &\iff \mathbf{x} \text{ is orthogonal to every row of } A \\ &\iff \mathbf{x} \text{ is orthogonal to every column of } A^T \\ &\iff \mathbf{x} \in \mathrm{col}(A^T)^\perp, \end{aligned}$$

so $N(A) = \mathrm{col}(A^T)^\perp$.

(b) Apply (a) to A^T.

□

8. Gram-Schmidt orthogonalisation process

Theorem 8.1 (Gram Schmidt process)**.** *Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ of a subspace H of $\mathbb{R}^n$, define*

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{x}_1 \\ \mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \boldsymbol{\cdot} \mathbf{v}_1}{\mathbf{v}_1 \boldsymbol{\cdot} \mathbf{v}_1}\mathbf{v}_1 \\ \mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \boldsymbol{\cdot} \mathbf{v}_1}{\mathbf{v}_1 \boldsymbol{\cdot} \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \boldsymbol{\cdot} \mathbf{v}_2}{\mathbf{v}_2 \boldsymbol{\cdot} \mathbf{v}_2}\mathbf{v}_2 \\ &\vdots \\ \mathbf{v}_r &= \mathbf{x}_r - \frac{\mathbf{x}_r \boldsymbol{\cdot} \mathbf{v}_1}{\mathbf{v}_1 \boldsymbol{\cdot} \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_r \boldsymbol{\cdot} \mathbf{v}_2}{\mathbf{v}_2 \boldsymbol{\cdot} \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_r \boldsymbol{\cdot} \mathbf{v}_{r-1}}{\mathbf{v}_{r-1} \boldsymbol{\cdot} \mathbf{v}_{r-1}}\mathbf{v}_{r-1} \end{aligned}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthogonal basis for H.

Proof. Consider $\mathbf{v}_k$ for $k = 1, \ldots, r$. We show they are orthogonal by induction on k. Evidently $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$. Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}$ are orthogonal. Then, for $j = 1, \ldots, k-1$, we have

$$\begin{aligned} \langle \mathbf{v}_k, \mathbf{v}_j \rangle &= \left\langle \mathbf{x}_k - \sum_{i=1}^{k-1} \frac{\langle \mathbf{x}_k, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle}\mathbf{v}_i,\ \mathbf{v}_j \right\rangle \\ &= \langle \mathbf{x}_k, \mathbf{v}_j \rangle - \sum_{i=1}^{k-1} \frac{\langle \mathbf{x}_k, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle}\langle \mathbf{v}_i,\ \mathbf{v}_j \rangle \\ &= \langle \mathbf{x}_k, \mathbf{v}_j \rangle - \frac{\langle \mathbf{x}_k, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}\langle \mathbf{v}_j,\ \mathbf{v}_j \rangle = 0. \end{aligned}$$

Hence $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_k\}$ is orthogonal. It follows that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is orthogonal and is a basis for H since it must be a linearly independent set, as shown before. □

Example 8.2. Let $H = \text{Span}\,(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ where

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 6 \end{pmatrix}.$$

Clearly $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is a basis of H. Construct an orthogonal basis of H.

Solution. We start by setting

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

The vector $\mathbf{v}_2$ is constructed by subtracting the orthogonal projection of $\mathbf{x}_2$ onto Span $(\mathbf{v}_1)$ from $\mathbf{x}_2$, that is,

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \mathbf{x}_2 - \frac{4}{4} \mathbf{v}_1 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

The vector $\mathbf{v}_3$ is constructed by subtracting the orthogonal projection of $\mathbf{x}_3$ onto Span $(\mathbf{v}_1, \mathbf{v}_2)$ from $\mathbf{x}_3$, that is,

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 = \mathbf{x}_3 - \frac{8}{4} \mathbf{v}_1 - \frac{6}{2} \mathbf{v}_2 = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \end{pmatrix},$$

producing the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for H. □

9. Orthogonal projections

Let H be a subspace of $\mathbb{R}^n$ and let $\mathbf{y}$ be a vector in $\mathbb{R}^n$. By the Gram-Schmidt process, we can fined an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ for H. We define

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_r}{\mathbf{u}_r \cdot \mathbf{u}_r} \mathbf{u}_r \,. \tag{3}$$

We call $\hat{\mathbf{y}}$ the **orthogonal projection of y onto** H , and is written

$$\hat{\mathbf{y}} = \text{proj}_H \mathbf{y} \,.$$

Theorem 9.1 (Orthogonal Decomposition Theorem)**.** *Let H be a subspace of $\mathbb{R}^n$. Then $\mathbb{R}^n = H \oplus H^\perp$. Hence* $\dim H + \dim H^\perp = n$.

Proof. It is easy to see that $H \cap H^{\perp} = \{\mathbf{0}\}$. We need to show that $\mathbb{R}^n = H + H^{\perp}$. Let $\mathbf{y} \in \mathbb{R}^n$ and let $\hat{\mathbf{y}}$ be its projection onto H, given by (3). Since $\hat{\mathbf{y}}$ is a linear combination of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$, we have $\hat{\mathbf{y}} \in H$. Let $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$. Then

$$\begin{aligned} \mathbf{z}\cdot\mathbf{u}_1 &= (\mathbf{y} - \hat{\mathbf{y}})\cdot\mathbf{u}_1 \\ &= \mathbf{y}\cdot\mathbf{u}_1 - \left(\frac{\mathbf{u}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)(\mathbf{u}_1\cdot\mathbf{u}_1) - 0 - \cdots - 0 \\ &= \mathbf{y}\cdot\mathbf{u}_1 - \mathbf{y}\cdot\mathbf{u}_1 \\ &= 0\,, \end{aligned}$$

so $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$. Similarly, we see that $\mathbf{z}$ is orthogonal to $\mathbf{u}_j$ for $j = 2, \ldots, r$, so $\mathbf{z} \in H^{\perp}$ by Theorem 7.3 (b). Therefore we have

$$\mathbf{y} = \hat{\mathbf{y}} + (\mathbf{y} - \hat{\mathbf{y}}) \in H + H^{\perp}.$$

□

One of the reasons why orthogonal projections play an important role in Linear Algebra, and indeed in other branches of Mathematics, is made plain in the following theorem:

Theorem 9.2 (Best Approximation Theorem)**.** *Let H be a subspace of $\mathbb{R}^n$, $\mathbf{y}$ any vector in $\mathbb{R}^n$, and $\hat{\mathbf{y}} = \mathrm{proj}_H\mathbf{y}$. Then $\hat{\mathbf{y}}$ is the closest point in H to $\mathbf{y}$, in the sense that*

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \tag{4}$$

for all $\mathbf{v} \in H$ distinct from $\hat{\mathbf{y}}$.

Proof. Take $\mathbf{v} \in H$ distinct from $\hat{\mathbf{y}}$. Then $\hat{\mathbf{y}} - \mathbf{v} \in H$. By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to H, so $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$.

Since

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})\,,$$

the Pythagorean Theorem (Theorem 6.8) gives

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2\,.$$

But $\|\hat{\mathbf{y}} - \mathbf{v}\|^2 > 0$, since $\hat{\mathbf{y}} \neq \mathbf{v}$, so the desired inequality (4) holds. □

The theorem above is the reason why the orthogonal projection of $\mathbf{y}$ onto H is often called the **best approximation of $\mathbf{y}$ by elements in H**.