Chapter 9

Linear Transformations

9.1 The vector space $\mathbb{R}^n$

As we saw in Chapter 2, once we have chosen an origin and unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, we can assign a *position vector* $\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ to each point in 3-space. From now on we shall *represent* this position vector by the *column vector* of coefficients of $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, that is to say the column vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, though we shall continue to represent the *points* of 3-space by triples (*row* vectors) (x, y, z).

For any given $n \in \mathbb{N}$, we let $\mathbb{R}^n$ denote the set of all column n-vectors ($n \times 1$ matrices):

$$\mathbb{R}^n = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} : a_1, a_2, \ldots, a_n \in \mathbb{R} \right\}.$$

The same notation $\mathbb{R}^n$ is also often used to denote the set of all n-tuples (*row* vectors) $(a_1, a_2, \ldots, a_n)$, but as most of our computations from now on will involve column vectors, for the rest of this module we shall reserve the notation $\mathbb{R}^n$ for these. For the special case when $n = 0$, we note that $\mathbb{R}^0$ has just a single vector (which is necessarily the zero vector). We denote by $\mathbf{0}_n$ the column vector which has n entries, all of which are 0; this is the *zero vector* of $\mathbb{R}^n$.

Addition and scalar multiplication of vectors in $\mathbb{R}^n$ is componentwise, entirely analogous to the situation for $\mathbb{R}^3$ as detailed in Section 2.1, and also for matrices, as given in Sections 7.1 and 7.3. For all $\lambda \in \mathbb{R}$ we have

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix} \quad \text{and} \quad \lambda \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix}.$$

From the properties of addition and multiplication by scalars that we have already proved for matrices, we observe that for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and all $\alpha, \beta \in \mathbb{R}$ we have the following ten properties.

(a.0) $\mathbf{u} + \mathbf{v} \in \mathbb{R}^n$.
(a.1) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
(a.2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
(a.3) There exists $\mathbf{e} \in \mathbb{R}^n$ such that $\mathbf{v} + \mathbf{e} = \mathbf{v} = \mathbf{e} + \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. [$\mathbf{e} = \mathbf{0}_n$.]
(a.4) For every $\mathbf{v} \in \mathbb{R}^n$ there is a $-\mathbf{v} \in \mathbb{R}^n$ such that $(-\mathbf{v}) + \mathbf{v} = \mathbf{v} + (-\mathbf{v}) = \mathbf{e}$.
(m.0) $\alpha\mathbf{v} \in \mathbb{R}^n$.
(m.1) $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$.
(m.2) $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$.
(m.3) $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$.
(m.4) $1\mathbf{v} = \mathbf{v}$.

This shows that $\mathbb{R}^n$ satisfies the rules to be an algebraic structure called a *vector space* (over $\mathbb{R}$). These are studied in detail in the module MTH5112: Linear Algebra I. You will come across many other examples of vector spaces, for example the set of all $m \times n$ matrices for a given m and n, as well as some more exotic examples such as the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

Properties (a.0)–(a.4) are for vector addition, while properties (m.0)–(m.4) are those for scalar multiplication. We call (a.0) and (m.0) *closure* properties, with the axioms of a vector space regarded as being (a.1)–(a.4) and (m.1)–(m.4).[1] I have told you the names of Properties (a.1)–(a.4), (m.1) and (m.2) before. Can you remember them? The symbol $\mathbf{e}$ in (a.3) and (a.4) is usually wriiten as $\mathbf{0}$ in this context.[2] Note also that the element $\mathbf{e}$ of (a.3) is unique, and that for any $\mathbf{v}$, the element $-\mathbf{v}$ of (a.4) is unique.

Vector subtraction is defined as you might expect (see earlier chapters), namely that $\mathbf{u} - \mathbf{v}$ is defined to be $\mathbf{u} + (-\mathbf{v})$. It is also possible to define the division of a vector by a nonzero scalar by $\mathbf{v}/\lambda := (1/\lambda)\mathbf{v}$ where $\mathbf{v}$ is a vector and $\lambda \neq 0$ is a scalar. However, this notation is rather ugly and should be avoided.

9.2 Linear transformations

Definition 9.1. A function $t : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* if for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and all $\alpha \in \mathbb{R}$ we have:

(i) $t(\mathbf{u} + \mathbf{v}) = t(\mathbf{u}) + t(\mathbf{v})$, and
(ii) $t(\alpha\mathbf{u}) = \alpha t(\mathbf{u})$.

If $m = n$ we call t a *linear transformation of* $\mathbb{R}^n$. Linear transformations are also called *linear maps*.

---

[1] Any vector space must by definition also satisfy (a.0) and (m.0), but the definition of a vector space is usually stated in such a way that they are not regarded as axioms.

[2] Apparently the letter 'e' comes from a German word, but nobody seems to be quite sure which one. Possibilities include *Eins* meaning *1 (the digit)*, *Einheit* meaning *unit(y)*, or a compound word such as *Einselement* or *Einheitselement*.

Example 1. Let $t : \mathbb{R}^2 \to \mathbb{R}^2$ be the function

$$t\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ -b \end{pmatrix}.$$

If $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$, and $\alpha \in \mathbb{R}$, we have

$$t(\mathbf{u}+\mathbf{v}) = t\begin{pmatrix} u_1+v_1 \\ u_2+v_2 \end{pmatrix} = \begin{pmatrix} u_1+v_1 \\ -u_2-v_2 \end{pmatrix} = \begin{pmatrix} u_1 \\ -u_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ -v_2 \end{pmatrix} = t(\mathbf{u}) + t(\mathbf{v})$$

and

$$t(\alpha\mathbf{u}) = t\begin{pmatrix} \alpha u_1 \\ u\alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha u_1 \\ -\alpha u_2 \end{pmatrix} = \alpha\begin{pmatrix} u_1 \\ -u_2 \end{pmatrix} = \alpha t(\mathbf{u}).$$

For each point (a, b) in the plane, t maps its position vector $\begin{pmatrix} a \\ b \end{pmatrix}$ to $\begin{pmatrix} a \\ -b \end{pmatrix}$, the position vector of $(a, -b)$. Geometrically, t is a reflexion in the x-axis.

Example 2. The function $t : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$t\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b+1 \end{pmatrix}$$

is *not* a linear transformation, since, for example,

$$t\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = t\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

but

$$t\begin{pmatrix} 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

9.3 Properties of linear transformations

Theorem 9.2. Let $t : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. then for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and for all scalars α, β we have:

(i) $t(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha t(\mathbf{u}) + \beta t(\mathbf{v})$;

(ii) $t(\mathbf{0}_n) = \mathbf{0}_m$;

(iii) $t(-\mathbf{u}) = -t(\mathbf{u})$.

Proof. (i). $t(\alpha\mathbf{u}+\beta\mathbf{v}) = t(\alpha\mathbf{u})+t(\beta\mathbf{v}) = \alpha t(\mathbf{u})+\beta t(\mathbf{v})$ (since t is a linear transformation).
(ii). $t(\mathbf{0}_n) = t(\mathbf{0}_n + \mathbf{0}_n) = t(\mathbf{0}_n) + t(\mathbf{0}_n)$. Adding $-t(\mathbf{0}_n)$ to both sides gives $\mathbf{0}_m = t(\mathbf{0}_n)$.
(iii). $t(-\mathbf{u}) = t((-1)\mathbf{u}) = (-1)t(\mathbf{u}) = -t(\mathbf{u})$. □

Notice that Part (ii) of this theorem gives us an alternative proof that the map t in Example 2 (above) is not a linear transformation, since it tells us that every linear transformation sends the origin in $\mathbb{R}^n$ to the origin in $\mathbb{R}^m$. Notice also that Part (i) of the theorem tells us that if t is a linear transformation then t maps every straight line $\{\,\lambda\mathbf{u} + (1-\lambda)\mathbf{v} : \lambda \in \mathbb{R}\,\}$ in $\mathbb{R}^n$ to either a straight line (if $t(\mathbf{u}) \neq t(\mathbf{v})$) or a point (if $t(\mathbf{u}) = t(\mathbf{v})$) in $\mathbb{R}^m$. In either case, the image of the straight line under t is $\{\,\lambda t(\mathbf{u}) + (1-\lambda)t(\mathbf{v}) : \lambda \in \mathbb{R}\,\}$.

Notice also, that Theorem 9.2 (i) combines both conditions (as per Definition 9.1) for t to be a linear transformation into a single condition (with more variables). It is also possible for a map $t : \mathbb{R}^n \to \mathbb{R}^n$ to satisfy both Parts (ii) and (iii) of Theorem 9.2 and still *not* be a linear transformation. One example is given by the map from $\mathbb{R}^2$ to itself given by

$$t\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a^3 \\ b^3 \end{pmatrix}.$$

9.4 Matrices and linear transformations

Let A be an $m \times n$ matrix. We define t_A to be the function

$$t_A : \mathbb{R}^n \to \mathbb{R}^m, \quad \mathbf{u} \mapsto t_A(\mathbf{u}) := A\mathbf{u}.$$

By properties of multiplication of matrices we have proved earlier, we know that for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and for all $\alpha \in \mathbb{R}$, we have:

$$\begin{aligned} t_A(\mathbf{u}+\mathbf{v}) &= A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = t_A(\mathbf{u}) + t_A(\mathbf{v}) \\ \text{and} \quad t_A(\alpha\mathbf{u}) &= A(\alpha\mathbf{u}) = \alpha(A\mathbf{u}) = \alpha t_A(\mathbf{u}). \end{aligned}$$

So t_A is a linear transformation: we call t_A the linear transformation *represented by A*.

Example. Let A be the 3×3 matrix

$$\begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_3 \\ u_3 & v_3 & w_3 \end{pmatrix}.$$

Then $t_A : \mathbb{R}^3 \to \mathbb{R}^3$, defined by $t_A(\mathbf{u}) = A\mathbf{u}$ is a linear transformation. It has

$$t_A\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad t_A\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad \text{and} \quad t_A\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}.$$

Is every linear transformation $t : \mathbb{R}^n \to \mathbb{R}^m$ represented by some matrix? The answer is yes. Rather than write out a formal proof for general m and n, let us think about the case of a linear transformation $t : \mathbb{R}^3 \to \mathbb{R}^3$. Suppose that

$$t\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad t\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad \text{and} \quad t\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}.$$

But then, since t is a linear transformation, we know that for any $\begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathbb{R}^3$ we have

$$t\begin{pmatrix} a \\ b \\ c \end{pmatrix} = t\left(a\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right) = at\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + bt\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + ct\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$
$$= a\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + b\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} + c\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

and so the linear transformation t is represented by the matrix which has as its columns

$$t\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad t\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad t\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

9.5 Composition of linear transformations and multiplication of matrices

Suppose that $s : \mathbb{R}^n \to \mathbb{R}^m$ and $t : \mathbb{R}^p \to \mathbb{R}^n$ are linear transformations. Define their *composition* $s \circ t : \mathbb{R}^p \to \mathbb{R}^m$ by:

$$(s \circ t)(\mathbf{u}) = s(t(\mathbf{u})) \text{ for all } \mathbf{u} \in \mathbb{R}^p.$$

Now suppose s is represented by the $m \times n$ matrix A and t is represented by the $n \times p$ matrix B. Then for all $\mathbf{u} \in \mathbb{R}^p$ we have

$$(s \circ t)(\mathbf{u}) = s(t(\mathbf{u})) = A(B\mathbf{u}) = (AB)\mathbf{u},$$

where the last equality follows from the associativity of matrix multiplication. We conclude the $s \circ t$ is a linear transformation, and that it is represented by the $m \times p$ matrix AB, which explains why we define matrix multiplication the way that we do!

Example. Let $s : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation which has

$$s\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad s\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix} \quad \text{and} \quad s\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix},$$

and let $t : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformation which has

$$t\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad t\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Then s is represented by $A = \begin{pmatrix} 1 & 1 & -1 \\ 2 & -3 & 0 \end{pmatrix}$, and t is represented by $B = \begin{pmatrix} 2 & 1 \\ 0 & 2 \\ 1 & 3 \end{pmatrix}$.

Moreover, $s \circ t : \mathbb{R}^2 \to \mathbb{R}^2$ is represented by $AB = \begin{pmatrix} 1 & 0 \\ 4 & -4 \end{pmatrix}$.

Thus if $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, then

$$s\left(t\begin{pmatrix} x \\ y \end{pmatrix}\right) = (s \circ t)\begin{pmatrix} x \\ y \end{pmatrix} = (AB)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 4 & -4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 4x - 4y \end{pmatrix}.$$

9.6 Rotations and reflexions of the plane

Let r_θ denote a rotation of the plane, keeping the origin fixed, through an angle θ. If $\theta > 0$ this is an *anticlockwise* rotation. (In mathematics the convention is that anticlockwise is the positive direction when it comes to measuring angles.) It is relatively easy to prove that r_θ is a linear transformation of the plane, as follows:

(i) Consider the parallelogram defining the sum $\mathbf{u} + \mathbf{v}$ of the position vectors $\mathbf{u}$ and $\mathbf{v}$ of any two points in the plane $\mathbb{R}^2$. If we rotate this parallelogram through an angle θ we obtain a new parallelogram, congruent to the original one. This new parallelogram has vertices with position vectors $\mathbf{0}, r_\theta(\mathbf{u}), r_\theta(\mathbf{v}), r_\theta(\mathbf{u}+\mathbf{v})$, and the very fact that it is a parallelogram tells us that $r_\theta(\mathbf{u}+\mathbf{v}) = r_\theta(\mathbf{u}) + r_\theta(\mathbf{v})$.

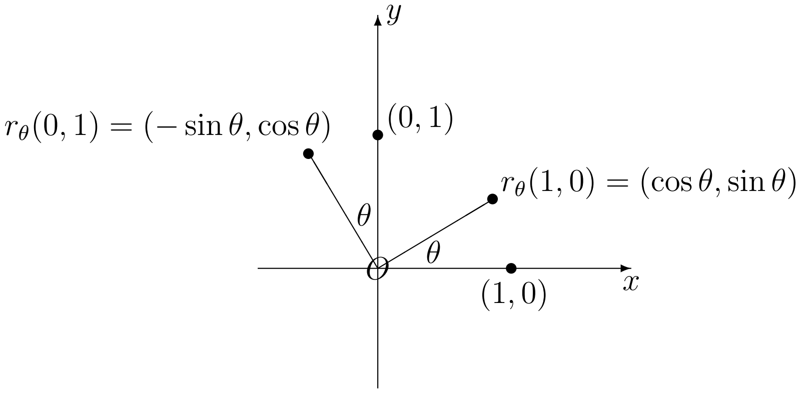
(ii) Given any vector $\mathbf{u}$ and scalar $\lambda \in \mathbb{R}$, we must show that $r_\theta(\lambda\mathbf{u}) = \lambda r_\theta(\mathbf{u})$. If $\lambda = 0$ or $\mathbf{u} = \mathbf{0}$ then $r_\theta(\lambda\mathbf{u}) = \mathbf{0} = \lambda r_\theta(\mathbf{u})$, as required. Now suppose that $\mathbf{u} \neq 0$ and $\lambda \neq 0$, and consider the line ℓ through the origin in the direction of $\mathbf{u}$. Let P be the point on this line with position vector $\mathbf{u}$ and Q be the point with position vector $\lambda\mathbf{u}$. Now r_θ sends O, P, Q to the points $O' = O, P', Q'$ (respectively) on $r_\theta(\ell)$, which is also a straight line. So we get that $|OQ'|/|OP'| = |OQ|/|OP| = |\lambda|$, since rotation preserves distances.

Rotations also preserve betweenness: that is if X, Y, Z are points on ℓ such that Y is between X and Z then $r_\theta(Y)$ is between $r_\theta(X)$ and $r_\theta(Z)$. Thus $O = O'$ is (strictly) between P' and Q' if and only if O is between P and Q, which is if and only if $\lambda < 0$. Hence $r_\theta(\lambda\mathbf{u}) = \lambda r_\theta(\mathbf{u})$ for all $\lambda \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^2$.

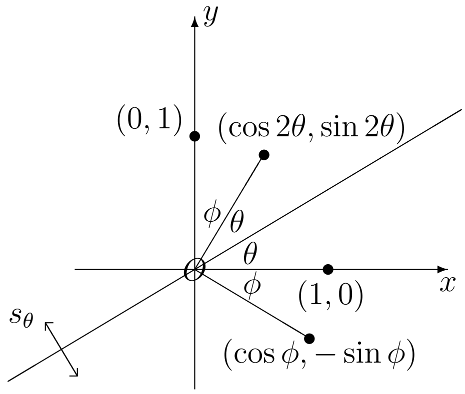
Since r_θ sends $(1, 0)$ to $(\cos\theta, \sin\theta)$ and sends $(0, 1)$ to $(-\sin\theta, \cos\theta)$ (see the picture overleaf, or draw your own picture), it is the linear transformation represented by the matrix:

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

It is easily checked that $(R_\theta)^{-1} = R_{-\theta}$ and that $R_{\phi+\theta}$ is the product of the matrices R_ϕ and R_θ. (For those of you who know about polar coördinates, the transformation r_θ maps the point with polar coördinates (r, ψ) to $(r, \psi + \theta)$.)



For this year [as I was away], we denote by s_θ the linear transformation which reflects the (x, y)-plane, with mirror the line through the origin at (anticlockwise) angle θ to the x-axis. Just as with a rotation r_θ, it is easy to prove geometrically that s_θ is a linear transformation (the details are left as an exercise for you).



From the illustration above it is apparent that $s_\theta(1,0) = (\cos 2\theta, \sin 2\theta)$ and that $s_\theta(0,1) = (\cos\phi, -\sin\phi)$, where ϕ is the angle shown. But $\phi + 2\theta = \frac{\pi}{2}$, so $\cos\phi = \sin 2\theta$ and $\sin\phi = \cos 2\theta$. Hence $s_\theta(0,1) = (\sin 2\theta, -\cos 2\theta)$, and we deduce that the reflexion s_θ is represented by the matrix

$$S_\theta = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}.$$

(In terms of polar coördinates, s_θ maps (r, ψ) to $(r, 2\theta - \psi)$.)

Exercise. Check that $(S_\theta)^{-1} = S_\theta$ by showing that $(S_\theta)^2 = \mathrm{I}_2$. (This verifies that applying the same reflexion twice brings every point of the (x, y)-plane back to itself.)

Remark. My preferred notation is that the mirror of the reflexion s_θ should have angle $\theta/2$ with the x-axis. The matrix of s_θ would then be the same as $S_{\theta/2}$ above, but would now be denoted S_θ, and we would have

$$S_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

In terms of polar coördinates, this version of s_θ maps (r, ψ) to $(r, \theta - \psi)$.

Remark. Rotations and reflexions are examples of *orthogonal* linear transformations of the plane, that is to say linear transformations s that have the property that for every $\mathbf{u}$ and $\mathbf{v}$ the dot products $\mathbf{u}\cdot\mathbf{v}$ and $s(\mathbf{u})\cdot s(\mathbf{v})$ are equal. In geometric terms this means that s preserves lengths of vectors and angles between vectors: in other words s is a 'rigid motion' of the plane (with the origin a fixed point). It can be shown that a linear transformation s of $\mathbb{R}^n$ is orthogonal if and only the matrix S that represents it has the property that $SS^\mathsf{T} = \mathrm{I}_n = S^\mathsf{T}S$ (where S^T is the transpose of S). In the module MTH5112: Linear Algebra I you will study orthogonal matrices, but we remark here that every 2×2 matrix which is orthogonal is the matrix either of a rotation or of a reflexion (depending on whether its determinant is $+1$ or -1).

9.7 Other linear transformations of the plane

We consider some examples. Other examples may obtained by composing linear transformations already listed. (This new linear transformation need not be one of the types we have listed.) For example, the composition $s_\phi \circ s_\theta$ of two reflexions is a rotation r_ψ, as you can check by multiplying the corresponding matrices (as an exercise, compute the angle ψ in terms of θ and ϕ). The composition $r_\phi \circ t_{2\mathrm{I}_2}$ is not in general one of listed types of linear transformation of $\mathbb{R}^2$.

9.7.1 Linear 'stretches' of axes and dilations

The linear transformation t_A of the plane represented by the matrix

$$A = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$$

sends the point $(x, 0)$ on the x-axis to the point $(ax, 0)$ on the x-axis, so it sends the x axis to itself, stretching it by a factor a (or contracting it if $|a| < 1$, and reflecting it if $a < 0$). We say that the x-axis is a *fixed line* of t_A, since every point on this line is mapped to a point on the same line. Similarly the y-axis is a fixed line for this t_A: it is stretched by a factor d. If $a \neq d$, the x-axis and the y-axis are the *only* fixed lines for this t_A. When $a = d$, *every* line through the origin is a fixed line and the effect of t_A is to apply the same stretching factor in all directions: such a t_A is called a *dilation* when $a = d > 0$. More generally, one can have stretches which fix some pair of lines other than the axes.

9.7.2 Shears (or transvections)

The linear transformation t_B represented by

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is an example of a *shear*. In some areas of Mathematics (such abstract algebra) shears are known as *transvections* instead. Here each point (x, y) is mapped to the point $(x+y, y)$. So each line $y = c$ (constant) is a fixed line for the transformation. Each point on the line $y = c$ is translated to the right by a distance c (which of course is negative when c is negative). Notice that the x-axis is fixed *pointwise* (every point on this axis is mapped to itself), and the y-axis is not even a fixed line: it is mapped to the line with equation $x = y$ which passes through the origin and has angle $\frac{\pi}{4}$ with the x-axis. More generally, for any nonzero constants b and c

$$B = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$$

are shears. The most general form of a shear in 2 dimensions is the matrix

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where $a+d = 2$ and $ad-bc = 1$, except that (by convention) the case $T = \mathrm{I}_2$ is excluded. There is some line that T fixes pointwise; this line is unique (since $T \neq \mathrm{I}_2$), and need not be one of the axes.

Shears of $\mathbb{R}^3$ fix a plane Π pointwise and send a (particular) vector $\mathbf{x} \notin \Pi$ to $\mathbf{x} + \mathbf{u}$ where $\mathbf{0} \neq \mathbf{u} \in \Pi$. The general vector $\mathbf{v} \in \mathbb{R}^3$ has form $\mathbf{v} = \lambda\mathbf{x} + \mathbf{w}$ for some (unique) $\lambda \in \mathbb{R}$ and $\mathbf{w} \in \Pi$. Since shears are linear maps, the shear must send $\mathbf{v}$ to $\mathbf{v} + \lambda\mathbf{u}$, and I leave it as an exercise for you to verify this. (I write $\mathbf{w} \in \Pi$ to mean that the position vector of $\mathbf{w}$ is in Π.) Shears of $\mathbb{R}^n$ are like shears of $\mathbb{R}^3$, except that Π is now a hyperplane that is fixed pointwise.

9.7.3 Singular transformations

The $n \times n$ matrix A is called *non-singular* if $\det A \neq 0$ and *singular* if $\det A = 0$. So a non-singular matrix is invertible, and a singular matrix is not. We call the linear transformation $t_A : \mathbb{R}^n \to \mathbb{R}^n$ *non-singular* if A is a non-singular matrix, and we call it *singular* if A is a singular matrix. So far all our examples of linear transformations of the plane have been non-singular. What about the case that A is singular?

The simplest singular case is when A is the zero 2×2 matrix (or zero $n \times n$ matrix for any $n \geqslant 1$). There is not much to say about this case geometrically, just that the corresponding linear transformation sends every point of the plane to the origin. But what about the case when

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $ad - bc = 0$ but a, b, c, d not all zero? Suppose for example that the first column of A is nonzero. Then $ad - bc = 0$ implies that $b = (d/c)a$, $d = (d/c)c$ if $c \neq 0$ and $b = (b/a)a$, $d = (b/a)c$ if $a \neq 0$. So the second column of A is a scalar multiple of the first column. Writing $\mathbf{u}$ for first column of A (so that $\mathbf{u} \neq \mathbf{0} = \mathbf{0}_2$) and $\lambda\mathbf{u}$ for the second column of A we see that

$$A\mathbf{i} = A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mathbf{u} \quad \text{and} \quad A\mathbf{j} = A\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \lambda\mathbf{u}$$

and so

$$A\begin{pmatrix} x \\ y \end{pmatrix} = A\left(x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = (x + \lambda y)\mathbf{u}$$

for every point (x, y) of the plane. Thus we find that

$$A(x\mathbf{v}) = A\begin{pmatrix} \lambda x \\ -x \end{pmatrix} = \mathbf{0} \text{ for all } x \in \mathbb{R}, \text{ where } \mathbf{v} = \begin{pmatrix} \lambda \\ -1 \end{pmatrix}.$$

So the whole of the plane is mapped onto the straight line through the origin which consists of the set of all scalar multiples of $\mathbf{u}$. If we take any point $\mathbf{c} = \alpha\mathbf{u}$ on this line then the equations $A\mathbf{x} = \mathbf{c}$ have infinitely many solutions (all the points on the straight line $\mathbf{r} = \alpha\mathbf{i} + \mu\mathbf{v}$ through $\alpha\mathbf{i}$ parallel to $\mathbf{v}$), and if we take any point $\mathbf{c}$ not on this line [$\mathbf{r} = \mu\mathbf{u}$] then the equations $A\mathbf{x} = \mathbf{c}$ have no solution. (Note that in general $\mathbf{u}$ and $\mathbf{v}$ are neither perpendicular nor parallel, though of course they may be sometimes.)

A similar situation arises when $A : \mathbb{R}^3 \to \mathbb{R}^3$ is a singular matrix. Then it can be shown that A maps the whole of $\mathbb{R}^3$ onto a plane through the origin, a line through the origin, or a single point (the origin).

9.7.4 Translations and affine maps

In general, these are *not* linear maps, but are a slight generalisation thereof. An *affine (linear) transformation* or *affine map* of $\mathbb{R}^n$ is a map from $\mathbb{R}^n$ to itself of the form:

$$f(A, \mathbf{b}) : \mathbf{r} \mapsto A\mathbf{r} + \mathbf{b},$$

where A is an $n \times n$ real matrix and $\mathbf{b}, \mathbf{r} \in \mathbb{R}^n$ with $\mathbf{b}$ fixed. Despite the use of the word 'linear' in the terminology, the map $f(A, \mathbf{b})$ is linear only when $\mathbf{b} = \mathbf{0}[= \mathbf{0}_n]$. Composing two affine maps gives us an affine map, for we can calculate that $f(A, \mathbf{b}) \circ f(C, \mathbf{d}) = f(AC, A\mathbf{d} + \mathbf{b})$, which I leave it to you to check. The map $f(A, \mathbf{b})$ is invertible if and only if A is, and in that case $f(A, \mathbf{b}) = f(A^{-1}, -A^{-1}\mathbf{b})$.

A special case of affine maps occurs when $A = \mathrm{I}_n$. Such maps are *translations* of $\mathbb{R}^n$. The most general form of a translation is $T_\mathbf{b} : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T_\mathbf{b} : \mathbf{r} \mapsto \mathbf{r} + \mathbf{b}$, that is $T_\mathbf{b}$ has the effect of adding $\mathbf{b}$ to each vector of $\mathbb{R}^n$. So $T_\mathbf{b} = f(\mathrm{I}_n, \mathbf{b})$. It is linear if and only if $\mathbf{b} = \mathbf{0}$. Composing two translations gives a translation: we have $T_\mathbf{b} \circ T_\mathbf{c} = T_\mathbf{c} \circ T_\mathbf{b} = T_{\mathbf{b}+\mathbf{c}}$. Also, $(T_\mathbf{b})^{-1} = T_{-\mathbf{b}}$ for all $\mathbf{b} \in \mathbb{R}^n$.

In fact, translations and affine maps of $\mathbb{R}^n$ can be regarded as linear maps of $\mathbb{R}^{n+1}$. A standard way in which this can achieved is to let $f(A, \mathbf{b})$ correspond to the following $(n+1) \times (n+1)$ real matrix:

$$\begin{pmatrix} & A & & \mathbf{b} \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$