

LINEAR TRANSFORMATIONS AND THEIR REPRESENTING MATRICES

DAVID WEBB

CONTENTS

1. Linear transformations	1
2. The representing matrix of a linear transformation	3
3. An application: reflections in the plane	6
4. The algebra of linear transformations	8

1. LINEAR TRANSFORMATIONS

1.1. DEFINITION. A function $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is *linear* if it satisfies two properties:

(1) For any vectors \mathbf{v} and \mathbf{w} in \mathbb{R}^n ,

$$(1.2) \quad T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) \text{ (“compatibility with addition”);}$$

(2) For any vector $\mathbf{v} \in \mathbb{R}^n$ and any $\lambda \in \mathbb{R}$,

$$(1.3) \quad T(\lambda \mathbf{v}) = \lambda T(\mathbf{v}) \text{ (“compatibility with scalar multiplication”).}$$

Often a linear mapping $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called a *linear transformation* or *linear map*. Some authors reserve the word “function” for a real-valued function, and call a vector-valued mapping a “transformation” or a “mapping.” In the special case when $n = m$ above, so the domain and target spaces are the same, we call a linear map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ a *linear operator*.

From the two properties (1) and (2) defining linearity above, a number of familiar facts follow. For instance:

1.4. PROPOSITION. A linear transformation carries the origin to the origin. That is, if $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is linear and $\mathbf{0}_k \in \mathbb{R}^k$ denotes the zero vector in \mathbb{R}^k , then $T(\mathbf{0}_n) = \mathbf{0}_m$.

Proof. Consider the equation $\mathbf{0}_n + \mathbf{0}_n = \mathbf{0}_n$ in \mathbb{R}^n . Apply T to both sides: then in \mathbb{R}^m we have the equation $T(\mathbf{0}_n + \mathbf{0}_n) = T(\mathbf{0}_n)$. By the compatibility of T with addition (the first property above), this becomes $T(\mathbf{0}_n) + T(\mathbf{0}_n) = T(\mathbf{0}_n)$. Now add $-T(\mathbf{0}_n)$ to both sides to obtain $T(\mathbf{0}_n) = \mathbf{0}_m$. \square

1.5. EXAMPLES. We consider some examples – and some non-examples – of linear transformations.

- (1) The identity map $\text{Id} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ defined by $\text{Id}(\mathbf{v}) = \mathbf{v}$ is linear (check this).
- (2) Define $R : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ by

$$R \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) = \begin{bmatrix} x \\ y \\ -z \end{bmatrix},$$

reflection through the xy -plane. Then R is linear. Indeed,

$$\begin{aligned} R \left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \right) &= R \left(\begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ v_3 + w_3 \end{bmatrix} \right) = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ -(v_3 + w_3) \end{bmatrix} \\ &= \begin{bmatrix} v_1 \\ v_2 \\ -v_3 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ -w_3 \end{bmatrix} = R \left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \right) + R \left(\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \right), \end{aligned}$$

so (1.2) holds, and one checks (1.3) similarly.

- (3) The mapping $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by $T \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} x \\ x \end{bmatrix}$ is linear (check this).
- (4) The mapping $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ given by $T \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} x-1 \\ y-1 \end{bmatrix}$ is *not* linear: indeed, $T \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$, but we know by Proposition (1.4) that if T were linear, then we would have $T(\mathbf{0}) = \mathbf{0}$.
- (5) The mapping $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ given by $T \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$ is not linear (why not?).
- (6) Any linear mapping $T : \mathbb{R} \rightarrow \mathbb{R}$ is given by $T(x) = mx$ for some $m \in \mathbb{R}$; i.e., its graph in \mathbb{R}^2 is a line through the origin. To see this, let $m = T(1)$. Then by (1.3), $T(x) = T(x \cdot 1) = xT(1) = xm = mx$.

Warning: Note that a linear map $f : \mathbb{R} \rightarrow \mathbb{R}$ in the sense of Definition (1.1) is *not* what you probably called a linear function in single-variable calculus; the latter was a function of the form $f(x) = mx + b$ whose graph need not pass through the origin. The proper term for a function like $f(x) = mx + b$ is *affine*. (Some authors call such a function “affine linear,” but the terminology is misleading — it incorrectly suggests that an affine linear map is a special kind of linear map, when in fact as noted above, an affine map is usually not linear at all! We will use the term “affine” in order to avoid this potential confusion. More generally, an *affine map* $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a mapping of the form $A(\mathbf{x}) = M(\mathbf{x}) + \mathbf{b}$ for $\mathbf{x} \in \mathbb{R}^n$, where $M : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear map and $\mathbf{b} \in \mathbb{R}^m$ is some fixed vector in the target space.)

- (7) **(Very important example)** Let A be any $m \times n$ matrix. Then A defines a linear map $L_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (“left-multiplication by A ”) defined by

$$(1.6) \quad L_A(\mathbf{v}) = A\mathbf{v}$$

for $\mathbf{v} \in \mathbb{R}^n$. Note that this makes sense: the matrix product of A (an $n \times m$ matrix) and \mathbf{v} (a column vector in \mathbb{R}^n , i.e., an $n \times 1$ matrix) is an $m \times 1$ matrix, i.e., a column vector in \mathbb{R}^m . That L_A is linear is clear from some basic properties of matrix multiplication. For example, to check that (1.2) holds, note that for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$,

$$L_A(\mathbf{v} + \mathbf{w}) = A(\mathbf{v} + \mathbf{w}) = A\mathbf{v} + A\mathbf{w} = L_A(\mathbf{v}) + L_A(\mathbf{w}),$$

and (1.3) is checked similarly.

As a special case of this last example, let $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. Then

$$L_A \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ x \end{bmatrix},$$

so in this case L_A is just the linear function T of the third example above.

2. THE REPRESENTING MATRIX OF A LINEAR TRANSFORMATION

The remarkable fact that matrix calculus so useful is that *every* linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is of the form L_A for some suitable $m \times n$ matrix A ; this matrix A is called the *representing matrix* of T , and we may denote it by $[T]$.

(2.1) **Thus $A = [T]$ is just another way of writing $T = L_A$.**

At the risk of belaboring the obvious, note that saying that a matrix A is the representing matrix of a linear transformation T just amounts to saying that for any vector \mathbf{v} in the domain of T ,

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

(the product on the right side is matrix multiplication), since $T(\mathbf{v}) = L_A(\mathbf{v}) = A\mathbf{v}$ by definition of L_A .

Proof. To see that every linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ has the form L_A for some $m \times n$ matrix A , let's consider the effect of T on an arbitrary vector

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

in the domain \mathbb{R}^n . Let

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

be the standard coordinate unit vectors (in \mathbb{R}^3 , these vectors are traditionally denoted $\mathbf{i}, \mathbf{j}, \mathbf{k}$). Then

$$\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \dots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \dots + x_n \mathbf{e}_n,$$

so

$$T(\mathbf{v}) = T(x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \dots + x_n \mathbf{e}_n) = T(x_1 \mathbf{e}_1) + T(x_2 \mathbf{e}_2) + \dots + T(x_n \mathbf{e}_n)$$

$$(2.2) \quad = x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + \dots + x_n T(\mathbf{e}_n).$$

Thus, to know what $T(\mathbf{v})$ is for any \mathbf{v} , all we need to know is the vectors $T(\mathbf{e}_1), T(\mathbf{e}_2), \dots, T(\mathbf{e}_n)$. Each of these is a vector in \mathbb{R}^m . Let's write them as

$$T(\mathbf{e}_1) = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, T(\mathbf{e}_2) = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \dots, T(\mathbf{e}_n) = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix};$$

thus a_{ij} is the i th component of the vector $T(\mathbf{e}_j) \in \mathbb{R}^m$. By (2.2),

$$\begin{aligned} T(\mathbf{v}) &= x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + \dots + T(\mathbf{e}_n) = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \dots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{v} = L_A(\mathbf{v}), \end{aligned}$$

where

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

□

Thus we have shown that T is just L_A , where A is the $m \times n$ matrix whose columns are the vectors $T(\mathbf{e}_1), T(\mathbf{e}_2), \dots, T(\mathbf{e}_n)$; equivalently:

(2.3)

The representing matrix $[T]$ is the matrix whose columns are the vectors $T(\mathbf{e}_1), T(\mathbf{e}_2), \dots, T(\mathbf{e}_n)$.

We might write this observation schematically as

$$[T] = \begin{bmatrix} \uparrow & \uparrow & \dots & \uparrow \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \dots & T(\mathbf{e}_n) \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}.$$

2.4. EXAMPLE. Let us determine the representing matrix $[\text{Id}]$ of the identity map $\text{Id} : \mathbb{R}^n \rightarrow \mathbb{R}^n$. By (2.3), the columns of $[\text{Id}]$ are $\text{Id}(\mathbf{e}_1), \text{Id}(\mathbf{e}_2), \dots, \text{Id}(\mathbf{e}_n)$, i.e., $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$. But the matrix whose columns are $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ is just

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

the $n \times n$ identity matrix I . Thus $[\text{Id}] = I$.

Note that two linear transformations $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $U : \mathbb{R}^n \rightarrow \mathbb{R}^m$ that have the same representing matrix must in fact be the same transformation: indeed, if $[T] = A = [U]$, then by (2.1), $T = L_A = U$.

Finally, there is a wonderfully convenient fact that helps to organize many seemingly complicated computations very simply; for example, as we will see later, it permits a very simple and easily remembered statement of the multivariable Chain Rule. Suppose that $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $U : \mathbb{R}^m \rightarrow \mathbb{R}^p$ are linear transformations. Then there is a composite linear transformation $U \circ T : \mathbb{R}^n \rightarrow \mathbb{R}^p$ given by $(U \circ T)(\mathbf{v}) = U(T(\mathbf{v}))$ for $\mathbf{v} \in \mathbb{R}^n$.

$$\begin{array}{ccccc} \mathbb{R}^n & \xrightarrow{T} & \mathbb{R}^m & \xrightarrow{U} & \mathbb{R}^p \\ & \searrow & & \nearrow & \\ & & U \circ T & & \end{array}$$

2.5. THEOREM. $[U \circ T] = [U][T]$. That is, the representing matrix of the composite of two linear transformations is the product of their representing matrices. (Note that this makes sense: $[T]$ is an $m \times n$ matrix while U is a $p \times m$ matrix, so the matrix product $[U][T]$ is a $p \times n$ matrix, as $[U \circ T]$ should be if it represents a linear transformation $\mathbb{R}^n \rightarrow \mathbb{R}^p$.)

Proof. To see why this is true, let $A = [T]$ and $B = [U]$. By (2.1), this is just another way of saying that $T = L_A$ and $U = L_B$. Then for any vector $\mathbf{v} \in \mathbb{R}^n$,

$$(U \circ T)(\mathbf{v}) = U(T(\mathbf{v})) = L_B(L_A(\mathbf{v})) = B(A\mathbf{v}) = (BA)\mathbf{v} = L_{BA}(\mathbf{v}),$$

where in the fourth equality we have used the associativity of matrix multiplication. Thus $U \circ T = L_{BA}$, which by (2.1) is just another way of saying that $[U \circ T] = BA$, i.e., that $[U \circ T] = [U][T]$. \square

In fact, this theorem is the *reason* that matrix multiplication is defined the way it is!

2.6. EXAMPLE. Consider the linear transformation R_θ that rotates the plane \mathbb{R}^2 counterclockwise by an angle θ . Clearly,

$$R_\theta(\mathbf{e}_1) = R_\theta\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix},$$

by definition of the sine and cosine functions. Now \mathbf{e}_2 is perpendicular to \mathbf{e}_1 , so $R_\theta(\mathbf{e}_2)$ should be a unit vector perpendicular to $R_\theta(\mathbf{e}_1) = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$, i.e., $R_\theta(\mathbf{e}_2)$ is one of the vectors

$$\begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix}, \begin{bmatrix} \sin \theta \\ -\cos \theta \end{bmatrix}.$$

From the picture it is clear that

$$R_\theta(\mathbf{e}_2) = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix},$$

since the other possibility would also involve a reflection. By (2.3), the representing matrix $[R_\theta]$ of R_θ is the matrix whose columns are $R_\theta(\mathbf{e}_1)$ and $R_\theta(\mathbf{e}_2)$; thus we have

$$(2.7) \quad [R_\theta] = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

Consider another rotation R_φ by an angle φ ; its representing matrix is given by

$$[R_\varphi] = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}.$$

Similarly, the representing matrix of the rotation $R_{\varphi+\theta}$ through the angle $\varphi + \theta$ is given by

$$[R_{\varphi+\theta}] = \begin{bmatrix} \cos(\varphi + \theta) & -\sin(\varphi + \theta) \\ \sin(\varphi + \theta) & \cos(\varphi + \theta) \end{bmatrix}.$$

Now the effect of rotating by an angle θ and then by an angle φ should be simply rotation by the angle $\varphi + \theta$, i.e.,

$$(2.8) \quad R_{\varphi+\theta} = R_\varphi \circ R_\theta.$$

Thus

$$[R_{\varphi+\theta}] = [R_{\varphi} \circ R_{\theta}].$$

By Theorem (2.5), this becomes

$$[R_{\varphi+\theta}] = [R_{\varphi}] [R_{\theta}],$$

i.e.,

$$\begin{bmatrix} \cos(\varphi + \theta) & -\sin(\varphi + \theta) \\ \sin(\varphi + \theta) & \cos(\varphi + \theta) \end{bmatrix} = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

After multiplying out the right hand side, this becomes

$$\begin{bmatrix} \cos(\varphi + \theta) & -\sin(\varphi + \theta) \\ \sin(\varphi + \theta) & \cos(\varphi + \theta) \end{bmatrix} = \begin{bmatrix} \cos \varphi \cos \theta - \sin \varphi \sin \theta & -\cos \varphi \sin \theta - \sin \varphi \cos \theta \\ \sin \varphi \cos \theta + \cos \varphi \sin \theta & -\sin \varphi \sin \theta + \cos \varphi \cos \theta \end{bmatrix}.$$

Comparing the entries in the first column of each matrix, we obtain

$$(2.9) \quad \cos(\varphi + \theta) = \cos \varphi \cos \theta - \sin \varphi \sin \theta,$$

$$(2.10) \quad \sin(\varphi + \theta) = \sin \varphi \cos \theta + \cos \varphi \sin \theta.$$

Thus from (2.8) along with Theorem (2.5) we recover the trigonometric sum identities.

3. AN APPLICATION: REFLECTIONS IN THE PLANE

Let l_{θ} be a line through the origin in the plane; say its equation is $y = mx$, where $m = \tan \theta$; thus θ is the angle of that the line l_{θ} makes with the x -axis. Consider the linear transformation $F_{\theta} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ called “reflection through l_{θ} ” defined geometrically as follows: intuitively, we view the line l_{θ} as a mirror; then for any vector $\mathbf{v} \in \mathbb{R}^2$ (viewed as the position vector of a point P in the plane), $F_{\theta}(\mathbf{v})$ is the position vector of the “mirror image” point of P on the other side of the line l_{θ} . More precisely, write \mathbf{v} as a sum of two vectors:

$$(3.1) \quad \mathbf{v} = \mathbf{v}^{\parallel} + \mathbf{v}^{\perp},$$

where \mathbf{v}^{\parallel} is a vector along the line l_{θ} and \mathbf{v}^{\perp} is perpendicular to l_{θ} . Then $F_{\theta}(\mathbf{v}) = \mathbf{v}^{\parallel} - \mathbf{v}^{\perp}$.

We wish to determine an explicit formula for F_{θ} . This can be done using elementary plane geometry and familiar facts about the dot product, but we seek here to outline a method that works in high dimensions as well, when it is not so easy to visualize the map geometrically.

To this end, note that we already know a formula for F_{θ} when $\theta = 0$: indeed, in that case, the line l_{θ} is just the x -axis, and the reflection F_0 is given by

$$F_0 \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} x \\ -y \end{bmatrix}.$$

For our purposes, though, a more useful (but equivalent) way of saying this is via the representing matrix $[F_0]$, which we now compute. The representing matrix $[F_0]$ is the matrix whose columns are $F_0(\mathbf{e}_1)$ and $F_0(\mathbf{e}_2)$. It is obvious that $F_0(\mathbf{e}_1) = \mathbf{e}_1$ and $F_0(\mathbf{e}_2) = -\mathbf{e}_2$, but let’s check it anyway. If we take $\mathbf{v} = \mathbf{e}_1$ in equation (3.1), we see that $\mathbf{v}^{\parallel} = \mathbf{e}_1$ ($\mathbf{v} = \mathbf{e}_1$ is on the line l_0 , the x -axis), so $\mathbf{v}^{\perp} = \mathbf{0}$, the zero vector; thus $F_0(\mathbf{e}_1) = \mathbf{v}^{\parallel} - \mathbf{v}^{\perp} = \mathbf{e}_1 - \mathbf{0} = \mathbf{e}_1$. Similarly, if we take $\mathbf{v} = \mathbf{e}_2$ in equation (3.1), we find $\mathbf{v}^{\parallel} = \mathbf{0}$ while $\mathbf{v}^{\perp} = \mathbf{e}_2$ ($\mathbf{v} = \mathbf{e}_2$ is itself perpendicular to the x -axis), so $F_0(\mathbf{e}_2) = \mathbf{v}^{\parallel} - \mathbf{v}^{\perp} = \mathbf{0} - \mathbf{e}_2 = -\mathbf{e}_2$. Thus $[F_0]$ is the matrix whose columns are \mathbf{e}_1 and $-\mathbf{e}_2$, i.e.,

$$(3.2) \quad [F_0] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

To use this seemingly trivial observation to compute F_θ , we note that F_θ can be written as a composition of three linear transformations:

$$(3.3) \quad F_\theta = R_\theta \circ F_0 \circ R_{-\theta},$$

where R_θ is the “rotation by θ ” operator whose representing matrix we already determined in equation (2.7). That is, in order to reflect through the line l_θ , we can proceed as follows: first, rotate the whole plane through an angle $-\theta$; this operation moves the mirror line l_θ to the x -axis. We now reflect the plane through the x -axis, which we know how to do by equation (3.2). Finally, we rotate the whole plane back through an angle of θ in order to “undo” the effect of our original rotation through an angle of $-\theta$; this operation restores the line in which we are interested (the “mirror”) to its original position, making an angle θ with the x -axis.

By Theorem (2.5), we can now compute $[F_\theta]$ easily:

$$[F_\theta] = [R_\theta \circ F_0 \circ R_{-\theta}] = [R_\theta][F_0][R_{-\theta}],$$

i.e.,

$$[F_\theta] = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix};$$

multiplying out the matrices yields

$$[F_\theta] = \begin{bmatrix} \cos^2 \theta - \sin^2 \theta & 2 \cos \theta \sin \theta \\ 2 \cos \theta \sin \theta & \sin^2 \theta - \cos^2 \theta \end{bmatrix}.$$

Using the identities (2.9) with $\varphi = \theta$, we finally conclude that

$$(3.4) \quad [F_\theta] = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}.$$

To see that this makes some sense geometrically, note that the first column of the matrix $[F_\theta]$ is the same as the first column of $[R_{2\theta}]$, the representing matrix of the rotation through an angle 2θ . But for *any* linear map $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, the first column of $[T]$ is just the vector $T(\mathbf{e}_1)$. Thus the equality of the first columns of $[F_\theta]$ and $[R_{2\theta}]$ just means that $F_\theta(\mathbf{e}_1) = R_{2\theta}(\mathbf{e}_1)$, which makes geometric sense: if you reflect the horizontal vector \mathbf{e}_1 through the line l_θ , the resulting vector should surely be a unit vector making an angle θ with the line l_θ , hence making an angle 2θ with the x -axis.

Although we will not make serious use of the striking power of the technique illustrated above, we conclude by showing how the algebra of linear transformations, once codified via their representing matrices, can lead to illuminating geometric insights. We consider a very simple example to illustrate. Let us rewrite equation (3.4):

$$[F_\theta] = [F_\theta] = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix} = \begin{bmatrix} \cos(2\theta) & -\sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

the right hand side of which we immediately recognize as $[R_{2\theta}][F_0]$. But by Theorem (2.5), the latter is just $[R_{2\theta} \circ F_0]$. Thus $[F_\theta] = [R_{2\theta} \circ F_0]$. Since two linear transformations with the same representing matrix are the same, it follows that $F_\theta = R_{2\theta} \circ F_0$. Now compose on the right by F_0 and use associativity of function composition:

$$(3.5) \quad F_\theta \circ F_0 = (R_{2\theta} \circ F_0) \circ F_0 = R_{2\theta} \circ (F_0 \circ F_0).$$

Finally, observe that $F_0 \circ F_0 = \text{Id}$, the identity transformation. (This is obvious geometrically, but we can prove it carefully by using the representing matrix: by Theorem (2.5) and equation (3.2),

$$[F_0 \circ F_0] = [F_0][F_0] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I = [\text{Id}],$$

so $F_0 \circ F_0 = \text{Id}$.) Then equation (3.5) becomes

$$(3.6) \quad F_\theta \circ F_0 = R_{2\theta}.$$

Thus we have proved:

3.7. THEOREM. *The effect of reflection of the plane through the x -axis followed by reflection of the plane through the line l_θ making an angle θ with the x -axis is just that of rotation of the plane through the angle 2θ .*

You should draw pictures and convince yourself of this fact.

In fact, more is true: for any line λ in the plane, let $F_\lambda : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ denote reflection through the line λ .

3.8. THEOREM. *For any two lines α and β in the plane, the composite $F_\beta \circ F_\alpha$ is the rotation through double the angle between the lines α and β .*

The proof is immediate: we merely set up our coordinate axes so that the x -axis is the line α ; then we have reduced to the case of Theorem (3.7).

4. THE ALGEBRA OF LINEAR TRANSFORMATIONS

We will denote the collection of all linear maps $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ by $\mathbf{L}(\mathbb{R}^n, \mathbb{R}^m)$. We can add two elements of $\mathbf{L}(\mathbb{R}^n, \mathbb{R}^m)$ in an obvious way:

4.1. DEFINITION. If $T, U \in \mathbf{L}(\mathbb{R}^n, \mathbb{R}^m)$, then $T + U$ is the mapping given by $(T + U)(\mathbf{v}) = T(\mathbf{v}) + U(\mathbf{v})$ for all vectors $\mathbf{v} \in \mathbb{R}^n$.

You should check that $T + U$ is indeed linear.

4.2. THEOREM. *Let $T, U \in \mathbf{L}(\mathbb{R}^n, \mathbb{R}^m)$. Then*

$$[T + U] = [T] + [U].$$

In words, the representing matrix of the sum of two linear transformations is the sum of their representing matrices. The proof is a simple exercise using the definition of the representing matrix and the definition of matrix addition.

4.3. DEFINITION. Given a linear map $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and a scalar $c \in \mathbb{R}$, we define a linear map cT from \mathbb{R}^n to \mathbb{R}^m (called scalar multiplication of T by c) by $(cT)(\mathbf{v}) = c T(\mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^n$.

As above, you should check that cT is linear, and prove the following theorem:

4.4. THEOREM. *Let $T \in \mathbf{L}(\mathbb{R}^n, \mathbb{R}^m)$, and let $c \in \mathbb{R}$. Then*

$$[cT] = c[T].$$

Note that the two definitions above endow the set $\mathbf{L}(\mathbb{R}^n, \mathbb{R}^m)$ with the structure of a vector space. The definition of matrix addition and scalar multiplication endow the space $M_{m \times n}(\mathbb{R})$ of m by n matrices (with real entries) with the structure of a vector space. There is a mapping $[-] : \mathbf{L}(\mathbb{R}^n, \mathbb{R}^m) \rightarrow M_{m \times n}(\mathbb{R})$ that sends a linear transformation T to its representing matrix $[T]$. The two simple theorems above merely say that this mapping $[-]$ is *itself* a linear map, using the vector space structures on its domain and target that we

defined above! But $[_]$ is even better than just a linear map: it has an inverse $L_{(_)} : M_{m \times n}(\mathbb{R}) \rightarrow \mathbf{L}(\mathbb{R}^n, \mathbb{R}^m)$ that sends a matrix $A \in M_{m \times n}(\mathbb{R})$ to the linear transformation $L_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ given by $L_A(\mathbf{v}) = A\mathbf{v}$. The fact that $[_]$ and $L_{(_)}$ are inverses of each other is just the observation (2.1). Since $[_] : \mathbf{L}(\mathbb{R}^n, \mathbb{R}^m) \rightarrow M_{m \times n}$ is linear, any “linear” question about linear maps from \mathbb{R}^n to \mathbb{R}^m (i.e., any question involving addition of such linear maps or multiplication by a scalar) can be translated via $[_]$ to a question about matrices, which are much more concrete and better suited to computation than abstract entities like linear transformations; the fact that $[_]$ has a linear inverse $L_{(_)}$ means that nothing is lost in the translation.

The mapping $[_]$ is a classic example of an *isomorphism*. Formally, an isomorphism of two vector spaces V and W is just a linear map $V \rightarrow W$ with a linear inverse $W \rightarrow V$. The idea is that V and W are to all intents and purposes “the same,” at least as far as linear algebra is concerned. An isomorphism of an abstract vector space V with a more concrete and readily understood vector space W makes V just as easy to understand as W : we simply use the isomorphism to translate a question about V to an equivalent question about W , solve our problem in W (where it is easy), then use the inverse of the isomorphism to translate our solution back to V .

While we will not make serious use of the idea of isomorphism (at least not explicitly), there is one important example that will use, and in any case the idea is so important throughout mathematics that it is worth making it explicit at this point. We will show that the space of linear maps $\mathbf{L}(\mathbb{R}, \mathbb{R}^m)$ of linear maps from the real line to \mathbb{R}^m is “the same” as the space \mathbb{R}^m itself. In fact, any linear map $T : \mathbb{R} \rightarrow \mathbb{R}^m$ has a representing matrix $[T] \in M_{m \times 1}$ in the space of m by 1 matrices; but an m by 1 matrix (m rows, one column) is just a column vector of length m , i.e., an element of \mathbb{R}^m . What does this representing matrix $[T]$ look like? Recall that its columns are just the values that the linear map T takes on the standard coordinate vectors. But in \mathbb{R} , there is only *one* standard coordinate vector: $\mathbf{e}_1 = 1$, the real number 1. Thus the first (and only!) column of $[T]$ is just the vector $T(1) \in \mathbb{R}^m$.

We can summarize as follows:

(4.5)

We can view any linear map $T : \mathbb{R} \rightarrow \mathbb{R}^m$ as a column vector in \mathbb{R}^m ; we do so by associating to the linear map T the vector $T(1) \in \mathbb{R}^m$ — which, when viewed as an m by 1 matrix, is just the representing matrix $[T]$.

Another way of arriving at the same conclusion is the following. If $T : \mathbb{R} \rightarrow \mathbb{R}^m$ is any linear map from the real line to \mathbb{R}^m , note that for any $x \in \mathbb{R}$, $T(x) = T(x \cdot 1) = xT(1)$, by (1.3); equivalently, $T(x) = T(1)x$, which just says that $T = L_{T(1)}$, i.e. (by (2.1)), $[T] = T(1)$.