§2 Linear Transformations and Matrices

§2.1 Linear Transformations

Definition Let V_1 and V_2 be vector spaces. A *linear transformation* is a function

$$T : V_1 \to V_2$$

with the following properties:

1. For any $v, w \in V_1$ we have $T(v + w) = T(v) + T(w)$.
2. For any $v \in V_1, r \in \mathbb{R}$ we have $T(rv) = rT(v)$.

In other words, a linear transformation is a function between vector spaces which is compatible with addition and scalar multiplication. In many cases, the parentheses in the notation will be dropped and we simply write Tv for $T(v)$. The following proposition is easy but very useful:

Proposition Let $T : V_1 \to V_2$ be a linear transformation and let 0_1 and 0_2 be the zero vectors in V_1 and V_2. Then we have $T(0_1) = 0_2$.

proof Multiplying the scalar 0 by any vector yields the zero vector. Consequently, we have

$$T(0_1) = T(00_1) = 0T(0_1) = 0_2$$

where the second equality follows from the second part of the definition of linearity. □

Definition A bijective linear transformation $T : V_1 \to V_2$ is called a linear *isomorphism*. (We shall usually suppress the adjective 'linear'.)

Proposition Let $T : V_1 \to V_2$ be a linear isomorphism and let $T^{-1} : V_1 \to V_2$ be its inverse. Then T^{-1} is also a linear transformation.

proof By the definition of the inverse $T \circ T^{-1} = \mathrm{Id}_{V_2}$ and $T^{-1} \circ T = \mathrm{Id}_{V_1}$. Let $v, w \in V_2$; we may write $v = T(T^{-1}(v))$ and $w = T(T^{-1}(w))$. We have

$$\begin{aligned} T^{-1}(v + w) &= T^{-1}(T(T^{-1}(v)) + T(T^{-1}(w))) \\ &= (T^{-1} \circ T)[T^{-1}(v) + T^{-1}(w)] \\ &= T^{-1}(v) + T^{-1}(w) \end{aligned}$$

where the second equality follows from the fact that T is linear. If we let $r \in \mathbb{R}$ be a scalar, then

$$T^{-1}(rv) = T^{-1}(rT(T^{-1}(v))) = (T^{-1} \circ T)[rT^{-1}(v)] = rT^{-1}(v).$$

Again, the second equality follows from the fact that T is linear. □

§2.2 Examples of Linear Transformations

We begin by giving an example of a function which is *not* a linear transformation. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function given by the rule $f(x, y) = x^2 + y + 1$. Since $f(0, 0) = 1 \neq 0$ we conclude by the proposition above that f is not linear.

We now give some examples of linear transformations. Let $T : \mathbb{R} \to \mathbb{R}$ be given by the rule

$$T(x) = ax$$

where $a \in \mathbb{R}$. To verify that T is a linear transformation, we must first check that for any $x, y \in \mathbb{R}$

$$T(x + y) = a(x + y) = ax + ay = T(x) + T(y).$$

Furthermore, we must also check that

$$T(rx) = a(rx) = r(ax) = rT(x)$$

for any scalar r.

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by the rule

$$T(x, y) = (ax + by, cx + dy)$$

where $a, b, c, d \in \mathbb{R}$. Then T is also a linear transformation (the proof is left as an exercise).

Now we give a more exotic example. Let $\mathcal{S}$ be the vector space of sequences described above, and define $T : \mathcal{S} \to \mathcal{S}$ by the rule

$$T((s_1.s_2, s_3, \dots)) = (s_2, s_1, s_3, \dots).$$

In other words, we exchange the first two terms of the sequence. We verify that T is a linear transformation. Given sequences $s = (s_n), t = (t_n) \in \mathcal{S}$, we have

$$\begin{aligned} T(s+t) &= T((s_1 + t_1, s_2 + t_2, \dots)) \\ &= (s_2 + t_2, s_1 + t_1, \dots) \\ &= (s_2, s_1, \dots) + (t_2, t_1, \dots) = T(s) + T(t). \end{aligned}$$

For any scalar r we also have

$$T(rs) = (rs_2, rs_1, rs_3, \dots) = r(s_2, s_1, \dots) = rT(s).$$

Thus T is a linear transformation.

§2.3 Basic Properties of Matrices

Definition A *matrix* A is a rectangular array of real numbers

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

The real numbers in the matrix are called its *entries*. The *rows* of the matrix are the horizontal sequences of entries and the *columns* are the vertical sequences of entries. We will use a_{ij} to denote the entry of the matrix A in the ith row and the jth column. (Notice the convention that the row number precedes the column number). A matrix with m rows and n columns is called an $m \times n$ matrix.

Consider the matrix
$$A = \begin{pmatrix} 3 & 2 & 1 \\ 0 & 2 & -1 \end{pmatrix}.$$
We have

$$\begin{aligned} a_{11} = 3 \quad a_{12} = 2 \quad a_{13} = 1 \\ a_{21} = 0 \quad a_{22} = 2 \quad a_{23} = -1 \end{aligned}$$

The matrix has two rows

$$\begin{pmatrix} 3 & 2 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 2 & -1 \end{pmatrix}$$

and three columns
$$\begin{pmatrix} 3 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 2 \\ 2 \end{pmatrix} \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$
Thus A is a 2×3 matrix.

An $m \times 1$ matrix is called a *column vector* and a $1 \times n$ matrix is called a *row vector*. An $n \times n$ matrix is called a *square* matrix.

We now describe some operations on matrices. Let $M(m, n)$ denote the set of all $m \times n$ matrices, and let $A, B \in M(m, n)$ so that

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & \dots b_{mn} \end{pmatrix}.$$

Then $A + B \in M(m, n)$ is defined as the matrix with entries $a_{ij} + b_{ij}$, i.e.

$$A + B = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{pmatrix}.$$

If $r \in \mathbb{R}$ is a scalar, then we define $rA \in M(m, n)$ to be the matrix with entries ra_{ij}
$$rA = \begin{pmatrix} ra_{11} & \dots & ra_{1n} \\ \vdots & & \vdots \\ ra_{m1} & \dots & ra_{mn} \end{pmatrix}.$$
Taking these operations to be addition and scalar multiplication, we find that $M(m, n)$ has the structure of a vector space:

Theorem The set $M(m, n)$ of all $m \times n$ matrices forms a vector space with respect to the operations defined above.

The proof is left as an exercise. $\square$

Remark Note that the space $M(1, n)$ of row vectors is the same as $\mathbb{R}^n$. Furthermore, the space $M(m, 1)$ of column vectors is just $\mathbb{R}^m$ 'written vertically'.

We will often take the liberty of writing elements of $\mathbb{R}^m$ as column vectors in $M(m,1)$.

§2.4 Matrices and Linear Transformations

The most important aspect of matrices is that they can be used to construct linear transformations. We describe how they may be used to transform vectors in $\mathbb{R}^n$ into vectors in $\mathbb{R}^m$. We start with an illustrative example.

Let A be the 3×2 matrix

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 0 \\ 5 & -2 \end{pmatrix}.$$

This matrix can be used to construct a function $T_A : \mathbb{R}^2 \to \mathbb{R}^3$, given by the rule

$$T_A(x_1, x_2) = \begin{pmatrix} 2x_1 + x_2 \\ 3x_1 \\ 5x_1 - 2x_2 \end{pmatrix}.$$

Note that we are writing the elements of $\mathbb{R}^3$ as column vectors. We encourage the reader to verify that T_A is a linear transformation. Notice that $T_A(1,0)$ equals the first column of A and $T_A(0,1)$ equals the second column.

We generalize this now to arbitrary matrices. Let A be an $m \times n$ matrix. We define a function $T_A : \mathbb{R}^n \to \mathbb{R}^m$ by the rule

$$T_A(x) = T_A((x_1, \ldots, x_n)) = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{pmatrix}$$

Again, we write the elements of $\mathbb{R}^m$ as *column* vectors.

We prove that T_A is linear. For any $x, x' \in \mathbb{R}^n$ we have

$$\begin{aligned} T_A(x + x') &= T_A((x_1 + x_1', \ldots, x_n + x_n')) \\ &= \begin{pmatrix} a_{11}(x_1 + x_1') + \ldots + a_{1n}(x_n + x_n') \\ \vdots \\ a_{m1}(x_1 + x_1') + \ldots + a_{mn}(x_n + x_n') \end{pmatrix} \\ &= \begin{pmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \end{pmatrix} + \begin{pmatrix} a_{11}x_1' + \ldots + a_{1n}x_n' \\ \vdots \\ a_{m1}x_1' + \ldots + a_{mn}x_n' \end{pmatrix} \\ &= T_A(x) + T_A(x'). \end{aligned}$$

Furthermore, for any scalar r we have

$$\begin{aligned}
T_A(rx) &= T_A((rx_1, \dots, rx_n)) \\
&= \begin{pmatrix} a_{11}rx_1 + \ldots + a_{1n}rx_n \\ \vdots \\ a_{m1}rx_1 + \ldots + a_{mn}rx_n \end{pmatrix} \\
&= r\begin{pmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \end{pmatrix} \\
&= rT_A(x).
\end{aligned}$$

We have proven the following theorem:

Theorem Let A be an $m \times n$ matrix. Then the function $T_A : \mathbb{R}^n \to \mathbb{R}^m$ defined above is a linear transformation.

The function T_A is called the *linear transformation associated to the matrix* A.

Our next goal is to classify linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$. We have already seen that $m \times n$ matrices give many examples of such linear transformations. We shall prove that these are the only examples.

First we introduce some notation. The vectors $e_1 = (1, 0, \dots, 0), e_2 = (0, 1, 0 \dots, 0), \dots, e_n = (0, \dots, 0, 1)$ in $\mathbb{R}^n$ will be called the *standard basis vectors*. Generally, all the components of e_i are zero except for the ith component which is one. Any vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ may be written

$$x = x_1e_1 + x_2e_2 \ldots + x_ne_n.$$

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation then we have

$$T(x) = T(x_1e_1 + \ldots + x_ne_n) = x_1T(e_1) + \ldots + x_nT(e_n).$$

Consequently T is entirely determined by its values on the standard basis vectors, which we may write

$$T(e_1) = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} \ldots T(e_n) = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Combining this with the previous formula, we obtain

$$\begin{aligned}
T(x) &= x_1\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + \ldots + x_n\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{pmatrix}
\end{aligned}$$

Consequently, $T = T_A$ for the matrix

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}.$$

We have proved the following theorem:

Theorem Let $T : \mathbb{R}^n \to \mathbb{R}^{>}$ be a linear transformation. Then $T = T_A$ for some $A \in M(m, n)$. The jth column of A is equal to $T(e_j)$ for $j = 1, \dots, n$.

Putting the previous two theorems together, we obtain the following

Corollary There is a one-to-one correspondence between linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ and $m \times n$ matrices.

Remark There is a more sophisticated version of this result. In the exercises, we shall see that the set of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ forms a vector space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. The ideas above may be used to show that $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is isomorphic to $M(m, n)$.

§2.5 Composition and Matrix Multiplication

Theorem Let $S : V_1 \to V_2$ and $T : V_2 \to V_3$ be linear transformations. Then

$$T \circ S : V_1 \to V_3$$

is also a linear transformation.

proof Let $v, w \in V_1$. We have

$$\begin{aligned} (T \circ S)(v + w) &= T(S(v + w)) = T(S(v) + T(w)) = T(S(v)) + T(S(w)) \\ &= (T \circ S)(v) + (T \circ S)(w) \end{aligned}$$

(the second and third equalities rely on the linearity of S and T). If r is any scalar then

$$(T \circ S)(rv) = T(S(rv)) = T(rS(v)) = rT(S(v)) = r(T \circ S)(v)$$

(again, the second and third equalities use the linearity of S and T). Therefore, $T \circ S$ is linear. $\square$

Let A be an $n \times k$ matrix and B an $m \times n$ matrix. Let $T_A : \mathbb{R}^k \to \mathbb{R}^n$ and $T_B : \mathbb{R}^n \to \mathbb{R}^m$ denote the linear transformations associated to A and B. The theorem above implies that $T_B \circ T_A$ is linear. Using the results from §2.4, we conclude that

$$T_B \circ T_A = T_C$$

for some $m \times k$ matrix C. How do we compute C from A and B?

We derive a formula for C in terms of A and B. To prevent possible confusion, we now use $e_1, \dots, e_k$ to denote the standard basis for $\mathbb{R}^k$ and $f_1, \dots, f_n$ to denote the standard basis for $\mathbb{R}^n$. The key to the formula is the fact that the columns of C are equal to the vectors $T_C(e_1), \dots, T_C(e_k)$. Let's compute $T_C(e_1)$ to see the general pattern:

$$\begin{aligned}
T_C(e_1) &= (T_B \circ T_A)(e_1) = T_B(T_A(e_1)) \\
&= T_B \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} \\
&= a_{11}T_B(f_1) + a_{21}T_B(f_2) + \ldots a_{n1}T_B(f_n) \\
&= a_{11} \begin{pmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{m1} \end{pmatrix} + a_{21} \begin{pmatrix} b_{12} \\ b_{22} \\ \vdots \\ b_{m2} \end{pmatrix} + \ldots a_{n1} \begin{pmatrix} b_{1n} \\ b_{2n} \\ \vdots \\ b_{mn} \end{pmatrix} \\
&= \begin{pmatrix} b_{11}a_{11} + b_{12}a_{21} + \ldots + b_{1n}a_{n1} \\ b_{21}a_{11} + b_{22}a_{21} + \ldots + b_{2n}a_{n1} \\ b_{m1}a_{11} + b_{m2}a_{21} + \ldots + b_{mn}a_{n1} \end{pmatrix}.
\end{aligned}$$

The same computation shows that

$$T_C(e_j) = \begin{pmatrix} b_{11}a_{1j} + b_{12}a_{2j} + \ldots + b_{1n}a_{nj} \\ b_{21}a_{1j} + b_{22}a_{2j} + \ldots + b_{2n}a_{nj} \\ b_{m1}a_{1j} + b_{m2}a_{2j} + \ldots + b_{mn}a_{nj} \end{pmatrix}.$$

We summarize this computation in the following proposition:

Proposition Let A be an $n \times k$ matrix and B am $m \times n$ matrix. Let C be the matrix for the composite linear transformation $T_B \circ T_A : \mathbb{R}^k \to \mathbb{R}^m$. This is a $m \times k$ matrix with entry

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots a_{in}b_{nj}$$

in the ith row and jth column.

Exercises

Transposes

1)Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by the rule $T(x, y) = (x + y, x + 2y)$.

a)Show that T is injective.

b)Show that T is surjective.

c)Compute a formula for the inverse of T.

2)Let V and W be fixed vector spaces. Let $\mathcal{L}(V, W)$ be the set of all linear transformations from V to W. Given linear transformations T_1 and T_2 in $\mathcal{L}(V, W)$ then we can define $T_1 + T_2$ to be the linear transformation given by $(T_1 + T_2)(v) = T_1(v) + T_2(v)$ for all $v \in V$. Also if $r \in \mathbb{R}$, then we can define rT to be the linear transformation given by $(rT)(v) = rT(v)$ for all $v \in V$. Show that with the above definitions of addition and scalar multiplication, $\mathcal{L}(V, W)$ is a real vector space.

3)Let V and W be fixed vector spaces. Let $T : V \to W$ be a linear transformation. Let K be the subset of V given by

$$K = \{v \in V : T(v) = 0\}$$

Show that K is also a vector space. This vector space is called the *kernel* of T and is denoted $\ker(T)$.

Let L be the subset of W given by

$$L = \{w \in W : \exists v \in V \ni T(v) = w\}$$

Show that L is a vector space. This vector space is often called the *image* of T and is denoted $\mathrm{im}(T)$.

Hint: For these two questions, you basically need to check that these subsets are closed under addition and scalar multiplication.