LINEAR TRANSFORMATIONS

1. Matrices and linear transformations

Before defining the notion of linear transformation, we begin with some familiar examples ((i) and (ii) below) and also an example ((iii) below) that is probably not familiar to you.

1.1. Examples. (i) A linear transformation $L : \mathbf{R} \to \mathbf{R}$ is any function of the form $L(x) = mx$ where m is a constant. The graph of L is a line through the origin.

(ii) A linear function $L : \mathbf{R}^2 \to \mathbf{R}$ is any function of the form $L(x, y) = ax + by$ where a and b are constants. The graph of L is a plane through the origin. Denoting elements of $\mathbf{R}^2$ as column matrices

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

, we have

$$L(x, y) = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = A\mathbf{x}$$

where A is the matrix $\begin{bmatrix} a & b \end{bmatrix}$.

(iii) Let

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 5 & 6 & 1 \end{bmatrix}.$$

Writing

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbf{R}^3$$

define

$$L(x, y, z) = A\mathbf{x} = \begin{bmatrix} 2 & 1 & 4 \\ 5 & 6 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2x + y + 4z \\ 5x + 6y + z \end{bmatrix}.$$

We can also write

$$L(x, y, z) = (2x + y + 4z, 5x + 6y + z).$$

Note that the expression $A\mathbf{x}$ in the third example makes sense because A is a 2×3 matrix and $\mathbf{x}$ is a 3×1 matrix. The resulting expression is a 2×1 matrix (a column) representing an element of $\mathbf{R}^2$.

Definition 1.1. A *linear transformation* L from $\mathbf{R}^n$ to $\mathbf{R}^m$ is any function of the form $L(\mathbf{x}) = A\mathbf{x}$ where A is an $m \times n$ matrix. (Here $\mathbf{x}$ denotes elements of $\mathbf{R}^n$, i.e., $n \times 1$ column matrices.) The matrix A is called the *representing matrix* of L.

1.2. Examples. (i). Let A be a 1×1 matrix, e.g., $A = [5]$. The corresponding linear transformation is given by $L(x) = [5][x] = [5x]$. In other words, L is the function $L(x) = 5x$. (Compare with example 1.1(i).)

(ii) Let $A = \begin{bmatrix} 3 & 4 \end{bmatrix}$. Then the corresponding linear transformation L has domain $\mathbf{R}^2$ and image $\mathbf{R}$. L is given by

$$L(x, y) = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [3x + 4y].$$

I.e., $L(x, y) = 3x + 4y$. Note that the graph of L, given by $z = 3x + 4y$, is a plane through the origin in $\mathbf{R}^3$. More generally, the graph of any linear transformation $L : \mathbf{R}^2 \to \mathbf{R}$ is a plane through the origin in $\mathbf{R}^3$. (Compare with example 1.1(ii).)

(iii) Let

$$A = \begin{bmatrix} 2 & 3 \\ 1 & -2 \end{bmatrix}$$

The corresponding linear transformation has domain $\mathbf{R}^2$ (since A has 2 columns) and range $\mathbf{R}^2$ (since A has two rows). It is given by

$$L(x, y) = \begin{bmatrix} 2 & 3 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x + 3y \\ x - 2y \end{bmatrix}$$

In other words, $L(x, y) = (2x + 3y, x - 2y)$.

(Aside: We can't graph L since the domain and range are both 2-dimensional. The graph would lie in $\mathbf{R}^4$, which we are unable to draw. However, if we could draw in four dimensions, the graph would be a two-dimensional plane sitting in $\mathbf{R}^4$, in the same way that a one-dimensional line sits in $\mathbf{R}^3$.)

Note that each of the component functions $2x + 3y$ and $x - 2y$ is given by multiplying each variable by a constant and adding. (I.e., they are first degree polynomials without constant terms.) This is the characteristic feature of linear transformations.

1.3. Examples. (i) $L(x, y) = (x^2 + y, xy)$ is not linear, because it has an x^2 term and also because it has a term xy.

(ii) $L(x, y) = (2 + x - y, x + y)$ is not linear because of the constant term 2.

(iii) $L(x, y) = (2x + 5y, x - y)$ is linear. Its representing matrix is given by

$$A = \begin{bmatrix} 2 & 5 \\ 1 & -1 \end{bmatrix}$$

1.4. Example. Let $H : \mathbf{R}^2 \to \mathbf{R}$ be given by $H(u, v) = 2u + 3v$, i.e.,

$$H(x, y) = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix}$$

where $A = \begin{bmatrix} 2 & 3 \end{bmatrix}$. Let $L : \mathbf{R}^2 \to \mathbf{R}^2$ be given by $L(x, y) = (x + y, x - y)$, i.e.,

$$L(x, y) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = B \begin{bmatrix} x \\ y \end{bmatrix}$$

where $B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Note that the composition $H \circ L : \mathbf{R}^2 \to \mathbf{R}$ is given by

$$H \circ L(x, y) = H(L(x, y)) = A(B \begin{bmatrix} x \\ y \end{bmatrix}) = (AB) \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

(In the second equality, we used the fact that matrix multiplication is associative, even though it's not commutative.)

More generally:

1.5. Theorem. *Suppose $H : \mathbf{R}^m \to \mathbf{R}^p$ and $L : \mathbf{R}^n \to \mathbf{R}^m$ are linear transformations with representing matrices A and B, respectively. Then $H \circ L$ is a linear transformation with representing matrix AB.*

1.6. Exercise. Let $H : \mathbf{R}^3 \to \mathbf{R}^2$ be given by $H(u, v, w) = (3u + 2v + w, u + 4v + w)$ and let $L : \mathbf{R}^2 \to \mathbf{R}^3$ be given by $L(x, y) = x + y, 5x + y, x + 2y)$.

(1) Write down the representing matrix for H.
(2) Write down the representing matrix for L.
(3) Use the theorem to find the representing matrix for $H \circ L$.
(4) Write down the expression for $H \circ L$ in the form $H \circ L(x, y) = (\ ,\ )$.

1.7. Tangent approximations of functions. A primary reason that the derivative at a point x_0 of a function $f : \mathbf{R} \to \mathbf{R}$ is so useful is that it enables us to approximate f near x_0 by its tangent line $y - y_0 = m(x - x_0)$ where $m = f'(x_0)$. This line is of course a translate of the line through the origin given by $y = mx$, which is the graph of the linear function $L(x) = mx$. As we saw in the document "Derivatives as matrices", the derivative a function in higher dimensions is a matrix A. Let L be the linear transformation with matrix A. The tangent approximation of f near $\mathbf{x}_0$ is given by $\mathbf{y} - \mathbf{y}_0 = A(\mathbf{x} - \mathbf{x}_0)$, which is just a translate of the graph of the linear transformation L. Thus an understanding of the geometric behavior of linear functions enables us to get an approximate understanding of more general functions. We will discuss the geometry of linear transformations in the next section.

1.8. Example. Let $L : \mathbf{R}^2 \to \mathbf{R}^2$ be given by $L(x, y) = (2x + 3y, 4x + y)$. Note that L is a linear transformation. Computing the derivative of L, we get

$$L'(x, y) = \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix}.$$

Note that this is precisely the representing matrix of the linear transformation L! More generally, the derivative of any linear transformation L at an arbitrary point is always the representing matrix of L. This is consistent with the idea of the derivative giving the best linear approximation (the tangent approximation) to a function near a point. If the function is already linear, the tangent "approximation" is exact.

2. Geometry of linear transformations

In this section we will consider for illustration only linear transformations $L : \mathbf{R}^2 \to \mathbf{R}^2$. (The behavior is similar in other dimensions but this is all we'll look at.) Before reading this section, you should read the discussion of more general transformations $T : \mathbf{R}^2 \to \mathbf{R}^2$ in Stewart: last paragraph of 1064 and all of page 1065. As in Stewart, we will use u, v for the independent variables and x, y for the dependent variables. I.e., we write $(x, y) = L(u, v)$. You're more used to using x, y for the independent variables. The reason for switching the roles is that it will be helpful when we talk about change of variables in integrals in Section 15.10.

Recall that a linear transformation $L : \mathbf{R}^2 \to \mathbf{R}^2$ is expressed as multiplication by a 2×2 matrix

$$L(u, v) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

It's useful to compute

$$L(1, 0) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ c \end{bmatrix}$$

and

$$L(0, 1) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}.$$

Thus L carries the coordinate vectors $(1, 0)$ and $(0, 1)$ to the vectors given by the two columns of the representing matrix. Moreover, the unit square is mapped to the parallelogram spanned by the two column vectors.

2.1. Example. Let $L(u, v) = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$. Writing $\begin{bmatrix} x \\ y \end{bmatrix} = L(u, v)$, we have

$$x = 2u \quad y = 3v.$$

This transformation expands horizontal distances by a factor of 2 and expands vertical distances by a factor of 3. So, for example, squares in the (u, v) plane are mapped to rectangles in (x, y) plane. The area of the image rectangle is 6 times as big as that of the starting square. (You'll notice that the determinant of the representing matrix of L is also 6.)

2.2. Example. Let $L(u, v) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$. Writing $\begin{bmatrix} x \\ y \end{bmatrix} = L(u, v)$, we have

$$x = u + v, \quad y = v.$$

Check that the unit square $0 \le u \le 1, 0 \le v \le 1$ is mapped to the parallelogram in the (x, y)-plane spanned by the vectors $\langle 1, 0 \rangle$ and $\langle 1, 1 \rangle$. This parallelogram has the same area (=1) as the unit square; it's just skewed. (Compute the determinant of the representing matrix, and you'll begin noticing a pattern.)

2.3. Theorem. *Let $L : \mathbf{R}^2 \to \mathbf{R}^2$ be a linear transformation and assume that the determinant of the representing matrix is nonzero. Then L maps lines in the (u, v) plane to lines in the (x, y) plane. Moreover, parallel lines go to parallel lines.*

We illustrate the theorem with an example.

2.4. EXAMPLE. $L(u,v) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$. Writing $\begin{bmatrix} x \\ y \end{bmatrix} = L(u,v)$, we have

$$x = u + 2v \quad y = 3u + 4v.$$

Consider the line in the (u,v) plane through the point (u_0, v_0) with direction vector $\langle a, b \rangle$. (Here a and b are constants. The slope of the line is $\frac{b}{a}$ if $a \neq 0$.) This line is given parametrically by

$$u = u_0 + at, \quad v = v_0 + bt.$$

The image is then given by

$$x = u + 2v = (u_0 + at) + (2v_0 + 2bt) = (u_0 + 2v_0) + (a + 2b)t.$$

$$y = 3u + 4v = (3u_0 + 3at) + (4v_0 + 4bt) = (3u_0 + 4v_0) + (3a + 4b)t.$$

I.e., in vector notation

$$\langle x, y \rangle = \langle u_0 + 2v_0, 3u_0 + 4v_0 \rangle + t\langle a + 2b, 3a + 4b \rangle.$$

Thus all lines in the direction $\langle a, b \rangle$ get mapped to lines with direction vector $\langle a + 2b, 3a + 4b \rangle$ E.g., the line $u = 2 + t, \ v = 1 + 5t$ is mapped to the line $x = 4 + 11t, \ y = 10 + 23t$.

Linear transformations $L(u,v) = A \begin{bmatrix} u \\ v \end{bmatrix}$ have many special properties that we list below. We are assuming that the determinant of the representing matrix A is non-zero. In cases where the determinant is zero, the entire image collapses to a line in the (x,y) plane (or to the single point 0 if all entries of A are zero).

- Parallel lines in the (u,v) plane go to parallel lines in (x,y) plane (as already noted).
- All parallelograms in the (u,v) plane are mapped to parallelograms in the (x,y) plane. (Moreover as already noted: the unit square $0 \leq u \leq 1, 0 \leq v \leq 1$ is mapped to the parallelogram spanned by the two vectors making up the columns of the matrix A.)
- Circles go to ellipses.
- All areas are multiplied by $|det(A)|$. E.g., if $det(A) = -5$, then every region in the (u,v) plane is mapped to some region in the (x,y) plane that is 5 times as large.

The last property above will be especially important when we discuss change of variables in multiple integrals.