Math 13 — W 2000 —Handout 1
Linear functions and representing matrices

Definition. A function $T:\mathbf{R}^n \to \mathbf{R}^m$ is *linear* if it satisfies two properties:

(1) $T(\mathbf{v}+\mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$ for any vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbf{R}^n$; ("compatibility with addition")

(2) $T(\lambda\mathbf{v}) = \lambda\, T(\mathbf{v})$ for any vector $\mathbf{v}$ in $\mathbf{R}^n$ and any real number λ ("compatibility with scalar multiplication").

From these properties, a number of familiar facts follow. For instance:

Proposition. A linear transformation must carry the origin to the origin. More precisely, if $T:\mathbf{R}^n \to \mathbf{R}^m$ is linear and $\mathbf{0}_k$ denotes the zero vector in $\mathbf{R}^k$, then $T(\mathbf{0}_n) = \mathbf{0}_m$.

Proof: Consider the equation $\mathbf{0}_n + \mathbf{0}_n = \mathbf{0}_n$ in $\mathbf{R}^n$. Apply T to both sides: then in $\mathbf{R}^m$ we have $T(\mathbf{0}_n + \mathbf{0}_n) = T(\mathbf{0}_n)$. Using (1) 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 $T(\mathbf{0}_n) = \mathbf{0}_m$.

Examples.

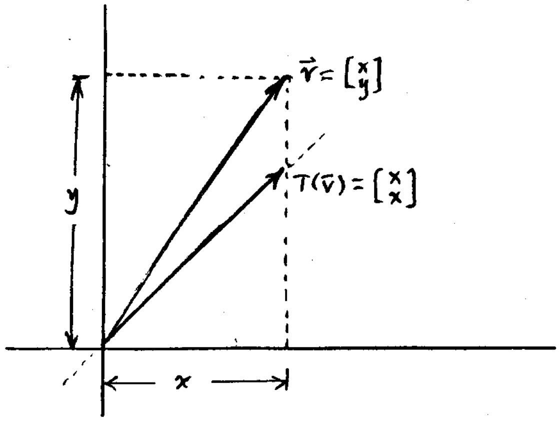
(i) The identity map $Id:\mathbf{R}^n \to \mathbf{R}^n$ defined by $Id(\mathbf{v}) = \mathbf{v}$ is linear (check this).

(ii) Define $R:\mathbf{R}^3 \to \mathbf{R}^3$ 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. In fact,

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

(1) holds, and one checks (2) similarly.

(iii) The function $T:\mathbf{R}^2 \to \mathbf{R}^2$ defined by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ x \end{bmatrix}$ is linear (check this).



(iv) The function $T:\mathbf{R}^2 \to \mathbf{R}^2$ defined by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x-1 \\ x-1 \end{bmatrix}$ is *not* linear: indeed,

$T(\mathbf{0}) = T\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$, but we know by the Proposition that if T were linear, then we would have $T(\mathbf{0}) = \mathbf{0}$.

(v) The function $T:\mathbf{R}^2 \to \mathbf{R}^2$ defined by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$ is not linear (why not?).

(vi) **(Important example)** Let A be any $m \times n$ matrix. Then A defines a linear map $M_A : \mathbf{R}^n \to \mathbf{R}^m$ ("multiplication by A") defined by $M_A(\mathbf{v}) = A\mathbf{v}$ for $\mathbf{v} \in \mathbf{R}^n$. Note that this makes sense: the matrix product of A (an $m \times n$ matrix) with $\mathbf{v}$ (a column vector in $\mathbf{R}^n$, i.e., an $n \times 1$ matrix) is an $m \times 1$ matrix, i.e., a column vector in $\mathbf{R}^m$. That M_A is linear is clear from some basic properties of matrix multiplication; for example, to check that (1) holds, note that for $\mathbf{v}, \mathbf{w} \in \mathbf{R}^n$, $M_A(\mathbf{v}+\mathbf{w}) = A(\mathbf{v}+\mathbf{w}) = A\mathbf{v} + A\mathbf{w} = M_A(\mathbf{v}) + M_A(\mathbf{w})$, and (2) is checked similarly.

[Matrix multiplication is distributive]

As a special case of Example (vi), let $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. Then $M_A\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) =$

$A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ x \end{bmatrix}$, so in this case M_A is just the linear function T of Example (iii) above.

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

Thus $A = [T]$ is just another way of writing $T = M_A$.

To see that every linear transformation $T:\mathbf{R}^n \to \mathbf{R}^m$ has the form M_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 $\mathbf{R}^n$. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$ be the standard coordinate basis vectors.

Then $\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \ldots + x_n \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n$, so $T(\mathbf{v}) =$

[By (1) in the definition of linearity] [By (2) in the definition of linearity.]

$T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n) = T(x_1\mathbf{e}_1) + T(x_2\mathbf{e}_2) + \ldots + T(x_n\mathbf{e}_n) =$
$x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \ldots + x_nT(\mathbf{e}_n)$.

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

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

thus a_{ij} is the ith component of the vector $T(\mathbf{e}_j)$. By the above, $T(\mathbf{v}) =$

$$x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \ldots + x_nT(\mathbf{e}_n) = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \ldots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} =$$

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{v} = M_A(\mathbf{v}),$$

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

Thus, we have shown that

T is just M_A, where A is the $m \times n$ matrix whose columns are the vectors $T(\mathbf{e}_1)$, $T(\mathbf{e}_2), \ldots, T(\mathbf{e}_n)$. Equivalently, $[T]$ is the $m \times n$ matrix whose columns are the vectors $T(\mathbf{e}_1)$, $T(\mathbf{e}_2), \ldots, T(\mathbf{e}_n)$.

Finally, there is a wonderful fact that makes many difficult-looking computations completely routine. Suppose that $T : \mathbf{R}^n \to \mathbf{R}^m$ and $U : \mathbf{R}^m \to \mathbf{R}^p$ are linear functions. Then we have a composite linear function $U \circ T : \mathbf{R}^n \to \mathbf{R}^p$ given by

$$\begin{tikzcd} \mathbf{R}^n \arrow[r, "T"] \arrow[rr, bend right, "U \circ T"'] & \mathbf{R}^m \arrow[r, "U"] & \mathbf{R}^p; \end{tikzcd}$$

thus for $\mathbf{v} \in \mathbf{R}^n$, $(U \circ T)(\mathbf{v}) = U(T(\mathbf{v}))$.

Theorem. $[U \circ T] = [U][T]$. That is, the representing matrix of the composite function is the product of the representing matrices. (Note that this makes sense: $[T]$ is an $m \times n$ matrix, $[U]$ is a $p \times m$ matrix, so $[U][T]$ is a $p \times n$ matrix, as $[U \circ T]$ should be if it is to represent a linear function $\mathbf{R}^n \to \mathbf{R}^p$.)

Proof. To see why this is true, let $A = [T]$, $B = [U]$. This is just another way of saying that $T = M_A$ and $U = M_B$. (See the boxed assertion on page 2.) Then for any vector $\mathbf{v} \in \mathbf{R}^n$,

$(U \circ T)(\mathbf{v}) = U(T(\mathbf{v})) = M_B(M_A(\mathbf{v})) = B(A\mathbf{v}) = (BA)\mathbf{v} = M_{BA}(\mathbf{v})$. i.e., $U \circ T = M_{BA}$, which is just another way of saying that $[U \circ T] = BA$, i.e., that $[U \circ T] = [U][T]$.

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