4.10. Matrices as linear transformations. Properties.

Linear transformations and matrices

Definition A function T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is called a "linear transformation" if it has the form

$T(x_1, \ldots, x_n) = (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n,\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n,\ \ldots,\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n),$

for some numbers a_{ij}.

They are also called "linear maps".

Note: Every linear transformation can be written using a matrix:

$$T(x_1, \ldots, x_n) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Conversely, every $m \times n$ matrix A defines a transformation $T(\vec{x}) = A\vec{x}$.

Notation:

- Given a matrix A, T_A is the linear transformation defined by $T_A(\vec{x}) = A\vec{x}$.
- Given a linear transformation T, we denote by $[T]$ the matrix such that $[T]\vec{x} = T(\vec{x})$.

Example: Let $T(x_1, x_2) = (3x_1 + 2x_2,\ 2x_1 - x_2,\ x_1 + x_2)$.

Then

$$[T] = \begin{bmatrix} 3 & 2 \\ 2 & -1 \\ 1 & 1 \end{bmatrix}. \text{ because } \begin{bmatrix} 3 & 2 \\ 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 + 2x_2 \\ 2x_1 - x_2 \\ x_1 + x_2 \end{bmatrix}$$

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}.$$

Then

$$T_A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ 4x_1 + 5x_2 + 6x_3 \end{bmatrix}.$$

Note that T_A has the following properties:

- $T_A(\vec{0}) = \vec{0}$
- $T_A(k\vec{x}) = kT_A(\vec{x})$
- $T_A(\vec{u} + \vec{v}) = T_A(\vec{u}) + T_A(\vec{v})$
- $T_A(\vec{x}) = \vec{0} \iff A\vec{x} = \vec{0} \iff \vec{x} \in \text{Null}(A)$
- $\vec{b} = T_A(\vec{x})$ for some $\vec{x} \iff A\vec{x} = \vec{b}$ for some $\vec{x} \iff \vec{b} \in \text{Col}(A)$

Kernel and range

Definition:

- The *kernel* of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, denoted $\ker(T)$, is the set

$$\ker(T) = \{\vec{x} \in \mathbb{R}^n : T(\vec{x}) = \vec{0}\}.$$

- The *range* of T, denoted $R(T)$ (or $\mathrm{Im}(T)$), is the set

$$R(T) = \{\vec{b} \in \mathbb{R}^m : \vec{b} = T(\vec{x}) \text{ for some } \vec{x} \in \mathbb{R}^n\}.$$

Note that, therefore,

$$\ker(T_A) = \mathrm{Null}(A) = \text{solution set of } A\vec{x} = \vec{b}$$

$$R(T_A) = \mathrm{Col}(A) = \text{set of vectors } \vec{b} \text{ for which } A\vec{x} = \vec{b} \text{ is consistent}$$

This gives **three different ways** to see the same set: transformation view, matrix view, and system of equations view.

Now recall: A function is one-to-one or injective if $f(x)$ and $f(y)$ are different whenever x and y are different. Or

$$f(x) = f(y) \implies x = y.$$

Theorem: Let A be an $m \times n$ matrix. Then T_A is one-to-one if and only if

$$\mathrm{Null}(A) = \{\vec{0}\}.$$

Why? Because if T_A is one-to-one then, if $T_A(\vec{x}) = \vec{0}$ then, since $T_A(\vec{0}) = \vec{0}$ as well, we must have $\vec{x} = \vec{0}$.

Therefore, the only element in $\ker(T)$ is $\vec{0}$, or $\ker(T) = \{\vec{0}\}$, and $\mathrm{Null}(T) = \ker(T) = \{\vec{0}\}$.

Now remember: For an $n \times n$ matrix A,

$$A \text{ invertible} \iff A\vec{x} = \vec{0} \text{ has } \vec{0} \text{ as the only solution}$$

$$\iff A\vec{x} = \vec{b} \text{ is always consistent.}$$

Therefore we have:

Theorem Let A be an $n \times n$ matrix. Then the following are equivalent:

- A is invertible
- T_A is one-to-one
- The kernel of T_A is $\{\vec{0}\}$
- The range of T_A is $\mathbb{R}^n$

In other words, T_A transforms all of $\mathbb{R}^n$ into all of $\mathbb{R}^n$, in such a way that every vector $\vec{b}$ in $\mathbb{R}^n$ is $T_A(\vec{x})$ for some $\vec{x} \in \mathbb{R}^n$.

Another way: T_A takes all the "information" and transforms it, but it does not lose any of it.

In this context, the dimension theorem would read: If $T_A : \mathbb{R}^n \to \mathbb{R}^m$

$$\dim(\ker T_A) + \dim(R(T_A)) = n.$$

We can think of $R(T_A)$ as the information preserved and $\ker(T_A)$ as the information lost (because all these vectors go to $\vec{0}$).

So: info lost + info preserved = original info.

Composition

Now recall: When you have two functions T and S so that the domain of T contains the range of S, we can compose them:

$$T \circ S(\vec{x}) = T(S(\vec{x})).$$

If $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^p$, then

$$T \circ S : \mathbb{R}^n \to \mathbb{R}^p.$$

So, if B is an $m \times n$ matrix and A is a $p \times m$ matrix, we have

$$T_A : \mathbb{R}^m \to \mathbb{R}^p, \quad T_A(\vec{x}) = A\vec{x},$$

$$T_B : \mathbb{R}^n \to \mathbb{R}^m, \quad T_B(\vec{x}) = B\vec{x},$$

and

$$T_A \circ T_B : \mathbb{R}^n \to \mathbb{R}^p, \quad \text{with}$$

$$T_A \circ T_B(\vec{x}) = T_A(T_B(\vec{x})) = T_A(B\vec{x}) = A(B\vec{x}) = AB\vec{x}.$$

In other words:

$$T_A \circ T_B = T_{AB}.$$

In words: composition of linear transformations corresponds to multiplication of matrices.

Recall: The function S is the inverse of a function T if

$$T \circ S(\vec{x}) = \vec{x} \quad \text{for all } \vec{x}$$

and

$$S \circ T(\vec{y}) = \vec{y} \quad \text{for all } \vec{y}.$$

Also, if T has an inverse, then it is unique, and is denoted by T^{-1}.

Also, recall from precalculus that for a function to be invertible, it has to be one-to-one and onto.

Thus, if T_A is invertible, then A is invertible, and

$$(T_A)^{-1} = T_{A^{-1}}, \quad \text{because}$$

$$T_A \circ T_{A^{-1}}(\vec{x}) = AA^{-1}\vec{x} = \vec{x}$$

and

$$T_{A^{-1}} \circ T_A(\vec{y}) = A^{-1}A\vec{x} = \vec{x}.$$

We have, therefore, a complete equivalence between $m \times n$ matrices and linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, with matrix multiplication corresponding to composition of transformations, and matrix inverses corresponding to inverse of transformations. In particular we have the following equivalent statements:

Theorem: For an $n \times n$ matrix A, the following statements are equivalent:

(a) A is invertible.

(b) $A\vec{x} = \vec{0}0$ has only the trivial solution.

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

(e) $A\vec{x} = \vec{b}$ is consistent for every $n \times 1$ matrix b.

(f) $A\vec{x} = \vec{b}$ has exactly one solution for every $n \times 1$ matrix b.

(g) $\det(A) \neq 0$.

(h) The column vectors of A are linearly independent.

(i) The row vectors of A are linearly independent.

(j) The column vectors of A span $\mathbb{R}^n$.

(k) The row vectors of A span $\mathbb{R}^n$.

(l) The column vectors of A form a basis for $\mathbb{R}^n$.

(m) The row vectors of A form a basis for $\mathbb{R}^n$.

(n) A has rank n.

(o) A has nullity 0.

(p) The orthogonal complement of the null space of A is $\mathbb{R}^n$.

(q) The orthogonal complement of the row space of A is $\{0\}$.

(r) The kernel of T_A is $\{0\}$.

(s) The range of T_A is $\mathbb{R}^n$.

(t) T_A is one-to-one.