Matrices characterize linear transformations

Matrices as functions

In the context of matrix-vector multiplication, we can think of a matrix as a function between vectors spaces. That is, if we hold a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ as fixed, this matrix maps vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$. Making this more explicit, we can define a function $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ as:

$$T(\mathbf{x}) := \mathbf{A}\mathbf{x}$$

where T uses the matrix $\mathbf{A}$ to performing the mapping.

Common matrix-defined functions

The identity matrix defines the identity function

Recall an identity function f for a set S is the function $f(x) := x$ for all $x \in S$. In the context of a function T over a vector space $\mathbb{R}^n$, the identity function $T(\mathbf{x}) := \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$ is defined using the **identity matrix** for $\mathbb{R}^n$. The identity matrix for $\mathbb{R}^n$, denoted $\mathbf{I}_n$ (or simply $\mathbf{I}$ if the dimensionality is implied by the context), is a square matrix of all zeros except for ones along the diagonal (Definition 1). For example, the identity matrix for $\mathbb{R}^3$ is defined as

$$\mathbf{I}_3 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

It can easily be shown that applying matrix-vector multiplication using an identity matrix $\mathbf{I}_n$ with any vector $\mathbf{x} \in \mathbb{R}^n$ will result in the same vector $\mathbf{x}$. Thus, a function $T(\mathbf{x}) := \mathbf{I}\mathbf{x}$ is the identity function for $\mathbf{R}^n$.

Definition 1 *Each real-valued, Euclidean vector space $\mathbb{R}^n$ is associated with an **identity matrix**, denoted $\mathbf{I}_{n \times n}$ (or simply $\mathbf{I}$ if the dimensionality is implied by the context), which is a square matrix consisting of zeros in the off-diagonal entries and ones along the diagonal.*

The zero matrix defines the zero function

Recall a zero-function f for a set S is the function $f(x) := 0$ for all $x \in S$. In the context of a function T over a vector space $\mathbb{R}^n$, the zero function $T(\mathbf{x}) := \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}^n$ is defined using the **zero matrix** for $\mathbb{R}^n$. The zero matrix for $\mathbb{R}^n$, denoted $\mathbf{0}_n$ is a square matrix of all zeros (Definition 2). For example, the zero matrix for $\mathbb{R}^3$ is defined as

$$\mathbf{0}_3 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

It can easily be shown that applying matrix-vector multiplication using an identity matrix $\mathbf{0}_n$ with any vector $\mathbf{x} \in \mathbb{R}^n$ will result in the zero vector $\mathbf{0}$. Thus, a function $T(\mathbf{x}) := \mathbf{0}_n\mathbf{x}$ is the zero function for $\mathbb{R}^n$.

Definition 2 *Each real-valued, Euclidean vector space $\mathbb{R}^n$ is associated with a **zero matrix**, denoted $\mathbf{0}_{n \times n}$, which is a square matrix consisting of all zeros.*

A matrix characterizes a linear transformation between coordinate vector spaces

In fact, any function T in the form of $T(\mathbf{x}) := \mathbf{A}\mathbf{x}$ is a linear function (Theorem 1). Thus, a matrix in $\mathbb{R}^{m \times n}$ defines a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Moreover, given a linear transformation between coordinate vector spaces, the matrix that performs this linear transformation is unique. For a given linear transformation T, that matrix is called the **standard matrix** (Definition 3). Thus, each matrix uniquely defines a certain linear transformation between coordinate vector spaces and every linear transformation between coordinate vector spaces is defined by some unique matrix (Theorem 2).

Theorem 1 *The function $T(\mathbf{x}) := \mathbf{A}\mathbf{x}$ is linear.*

Proof:

We show that for all $\mathbf{u}, \mathbf{v}$ in the domain of T and for all scalars c, the following conditions hold:

$$\begin{aligned} &\text{a.)} \quad \mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{A}\mathbf{u} + \mathbf{A}\mathbf{v} \\ &\text{b.)} \quad \mathbf{A}(c\mathbf{u}) = c(\mathbf{A}\mathbf{u}) \end{aligned}$$

a.)

$$\begin{aligned}\mathbf{A}(\mathbf{u}+\mathbf{v}) &= \mathbf{a}_{*,1}(u_1+v_1)+\mathbf{a}_{*,2}(u_2+v_2)+\cdots+\mathbf{a}_{*,n}(u_n+v_n)\\ &= \mathbf{a}_{*,1}u_1+\mathbf{a}_{*,1}v_1+\mathbf{a}_{*,2}u_2+\mathbf{a}_{*,2}v_2+\cdots+\mathbf{a}_{*,n}u_n+\mathbf{a}_{*,n}v_n\\ &= (\mathbf{a}_{*,1}u_1+\mathbf{a}_{*,2}u_2+\cdots+\mathbf{a}_{*,n}u_n)+(\mathbf{a}_{*,1}v_1+\mathbf{a}_{*,2}v_2+\cdots+\mathbf{a}_{*,n}v_n)\\ &= \mathbf{A}\mathbf{u}+\mathbf{A}\mathbf{v}\end{aligned}$$

b.)

$$\begin{aligned}\mathbf{A}(c\mathbf{u}) &= \mathbf{a}_{*,1}(cu_1)+\mathbf{a}_{*,2}(cu_2)+\cdots+\mathbf{a}_{*,n}(cu_n)\\ &= c\mathbf{a}_{*,1}(u_1)+c\mathbf{a}_{*,2}(u_2)+\cdots+c\mathbf{a}_{*,n}(u_n)\\ &= c(\mathbf{a}_{*,1}u_1+\mathbf{a}_{*,2}u_2+\cdots+\mathbf{a}_{*,n}u_n)\\ &= c(\mathbf{A}\mathbf{u})\end{aligned}$$

□

Definition 3 *Given a linear transformation T defined as*

$$T(\mathbf{x}) := \mathbf{A}\mathbf{x}$$

*where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{x} \in \mathbb{R}^n$, the matrix $\mathbf{A}$ is called the **standard matrix** of T.*

Theorem 2 *Given a linear transformation T, this transformation is characterized by a unique standard matrix $\mathbf{A}$. Furthermore, $\mathbf{A}$ is defined as follows:*

$$\mathbf{A} := \left[T(\mathbf{i}_{*,1}), T(\mathbf{i}_{*,2}), \dots, T(\mathbf{i}_{*,n})\right]$$

where $\mathbf{i}_{*,i}$ is the ith column of the identity matrix $\mathbf{I}_n$.

Proof:

$$
\begin{aligned}
\mathbf{x} &= \mathbf{I}\mathbf{x} \\
&= \mathbf{i}_{*,1}x_1, + \mathbf{i}_{*,2}x_2 + \cdots + \mathbf{i}_{*,n}x_n \\
\implies T(\mathbf{x}) &= T(\mathbf{i}_{*,1}x_1 + \mathbf{i}_{*,2}x_2 + \cdots + \mathbf{i}_{*,n}x_n) \\
&= T(\mathbf{i}_{*,1}x_1) + T(\mathbf{i}_{*,2}x_2) + \cdots + T(\mathbf{i}_{*,n}x_n) && T \text{ is linear} \\
&= \left[T(\mathbf{i}_{*,1}), T(\mathbf{i}_{*,2}), \ldots, T(\mathbf{i}_{*,n})\right]\mathbf{x}
\end{aligned}
$$

This implies that the linear transformation T operates on a vector $\mathbf{x}$ through the matrix

$$A := \left[T(\mathbf{i}_{*,1}), T(\mathbf{i}_{*,2}), \ldots, T(\mathbf{i}_{*,n})\right]$$

That is, the standard matrix $\mathbf{A}$ is formed by transforming each column of the identity matrix with T and using each resultant vector as a column of $\mathbf{A}$.

□