Matrix-vector multiplication

Matrices are often used for manipulating vectors using an operation called **matrix-vector multiplication** (Definition 1). This operation uses a matrix and a vector in order to produce a new vector. We note that a vector can only be multiplied by a matrix that has the same number of columns as there are elements in the vector. The number of elements in the output vector is the number of rows in the matrix.

Definition 1 *Given a matrix* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and vector* $\mathbf{x} \in \mathbb{R}^n$ *the* ***matrix-vector multiplication*** *of* $\mathbf{A}$ *and* $\mathbf{x}$ *is defined as*

$$\mathbf{A}\mathbf{x} := x_1\mathbf{a}_{*,1} + x_2\mathbf{a}_{*,2} + \cdots + x_n\mathbf{a}_{*,n}$$

where $\mathbf{a}_{*,i}$ *is the ith column vector of* $\mathbf{A}$.

Intuition

Matrix-vector multiplication can be conceptualized in several useful ways, at various levels of abstraction. These views come in handy when we attempt to conceptualize the various ways in which we utilize matrix-vector multiplication to model real-world problems. Below are three ways that I find useful for conceptualizing matrix-vector multiplication ordered from most to least abstract:

1. Matrix-vector multiplication allows a matrix to define a mapping between two vector spaces

2. Matrix-vector multiplication is the process of taking a linear combination of the column-space of a matrix using the elements of a vector as the coefficients

3. Matrix-vector multiplication defines a process for creating a new vector using an existing vector: each element of the new vector is "generated" by taking a weighted sum of each row of the matrix using the elements of a vector as coefficients

We discuss each of these views in more detail in the following sections:

1. Matrix-vector multiplication allows matrices to define mappings between vector spaces

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 $T : \mathbb{R}^n \to \mathbb{R}^m$ as:

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

where T uses the matrix $\mathbf{A}$ to performing the mapping (Figure 1). In fact, as we show in a later section, such a matrix-defined function is a linear function. Furthermore, any linear function between vector spaces is uniquely defined by a matrix.

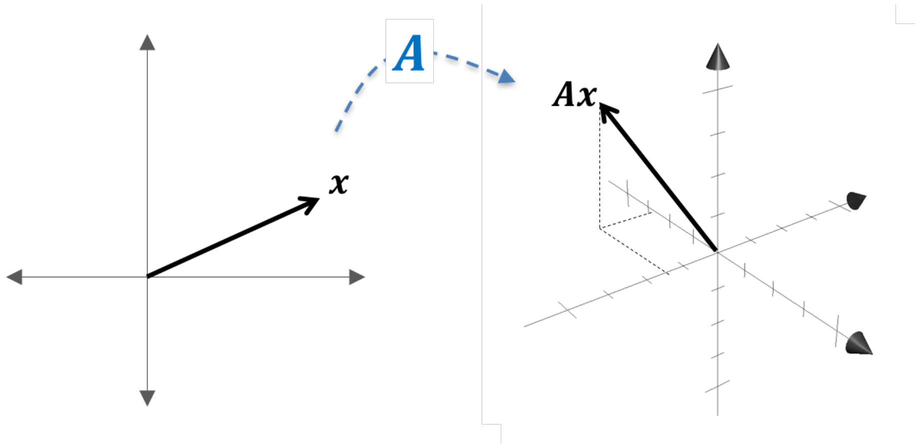

Figure 1: A matrix $\mathbf{A} \in \mathbb{R}^{3\times 2}$ maps each vector in $\mathbb{R}^2$ to a vector in $\mathbb{R}^3$.

2. Matrix-vector multiplication performs a linear combination

Matrix-vector multiplication between a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ and vector $\mathbf{x} \in \mathbb{R}^m$ can be understood as taking a linear combination of the column vectors of $\mathbf{A}$ using the elements of $\mathbf{x}$ as the coefficients. This perspective of matrix-vector multiplication also provides some insight into how a matrix-defined function $T(\mathbf{x}) := \mathbf{A}\mathbf{x}$ operates – it uses the column-vectors of $\mathbf{A}$ to form a set of basis vectors for the vector space that represents T's range. Given an input vector $\mathbf{x}$, T constructs the output vector in $\mathbb{R}^m$ by taking a linear combination of the basis vectors using the elements of $\mathbf{x}$ as the coefficients. This is illustrated in Figure 2 and This process is illustrated schematically in Figure 3.A.

3. Matrix-vector multiplication as a "row-wise, vector-generating" process

We can also view matrix-vector multiplication as a sort of "process" that constructs each element of the output vector. This process involves taking the vector and computing the dot product of that vector with each row in the matrix thereby forming each element of

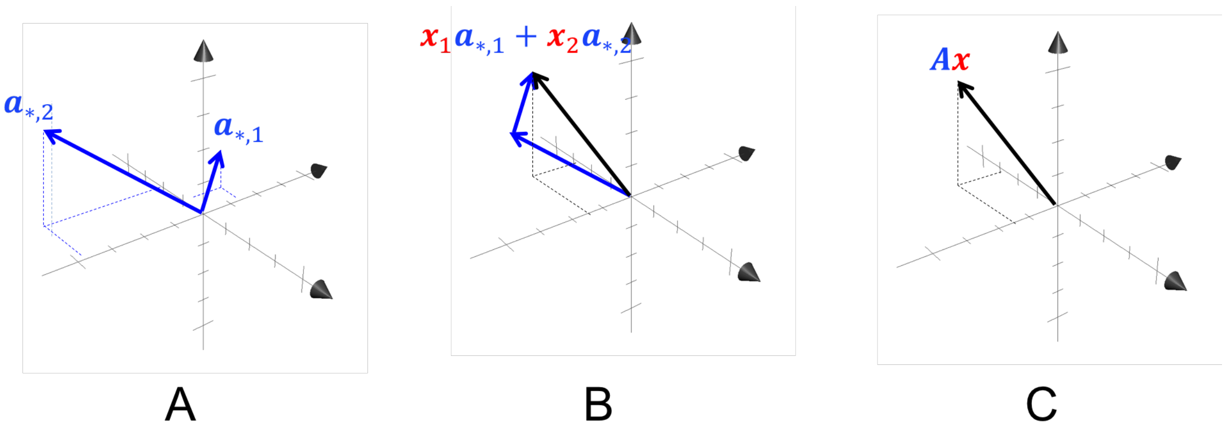

Figure 2: A matrix $\mathbf{A} \in \mathbb{R}^{3 \times 2}$ maps each vector in $\mathbb{R}^2$ to a vector in $\mathbb{R}^3$ by taking the linear combination of $\mathbf{a}_{*,1}$ and $\mathbf{a}_{2,*}$ using the vector $\mathbf{x}$'s elements as coefficients. (A) The column vectors of $\mathbf{A}$. (B) Taking a linear combination of the columns of $\mathbf{A}$ using the elements of $\mathbf{x}$ as coefficients. (C) The resultant vector.

the output vector. That is, Theorem 1 shows that

$$\mathbf{Ax} = \begin{bmatrix} \mathbf{a}_{1,*} \cdot \mathbf{x} \\ \mathbf{a}_{2,*} \cdot \mathbf{x} \\ \vdots \\ \mathbf{a}_{m,*} \cdot \mathbf{x} \end{bmatrix}$$

where $\mathbf{a}_{i,*}$ is the ith row-vector in $\mathbf{A}$. This process is illustrated schematically in Figure 3.B.

Theorem 1 *Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and vector $\mathbf{x} \in \mathbb{R}^n$, follows that*

$$\mathbf{Ax} = \begin{bmatrix} \mathbf{a}_{1,*} \cdot \mathbf{x} \\ \mathbf{a}_{2,*} \cdot \mathbf{x} \\ \vdots \\ \mathbf{a}_{m,*} \cdot \mathbf{x} \end{bmatrix}$$

where $\mathbf{a}_{i,*}$ is the ith row-vector in $\mathbf{A}$

Proof:

$$
\begin{aligned}
\mathbf{Ax} &:= x_1\mathbf{a}_{*,1} + x_2\mathbf{a}_{*,2} + \dots + x_n\mathbf{a}_{*,n} \\
&= x_1 \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{bmatrix} + x_2 \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{bmatrix} + \dots + \begin{bmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{bmatrix} \\
&= \begin{bmatrix} x_1a_{1,1} \\ x_1a_{2,1} \\ \vdots \\ x_1a_{m,1} \end{bmatrix} + \begin{bmatrix} x_2a_{1,2} \\ x_2a_{2,2} \\ \vdots \\ x_2a_{m,2} \end{bmatrix} + \dots + \begin{bmatrix} x_na_{1,n} \\ x_na_{2,n} \\ \vdots \\ x_na_{m,n} \end{bmatrix} \\
&= \begin{bmatrix} x_1a_{1,1} + x_2a_{1,2} + \dots, x_na_{1,n} \\ x_1a_{2,1} + x_2a_{2,2} + \dots, x_na_{2,n} \\ \vdots \\ x_1a_{m,1} + x_2a_{m,2} + \dots, x_na_{m,n} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n a_{1,i}x_i \\ \sum_{i=1}^n a_{2,i}x_i \\ \vdots \\ \sum_{i=1}^n a_{m,i}x_i \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{a}_{1,*} \cdot \mathbf{x} \\ \mathbf{a}_{2,*} \cdot \mathbf{x} \\ \vdots \\ \mathbf{a}_{m,*} \cdot \mathbf{x} \end{bmatrix}
\end{aligned}
$$

$\square$

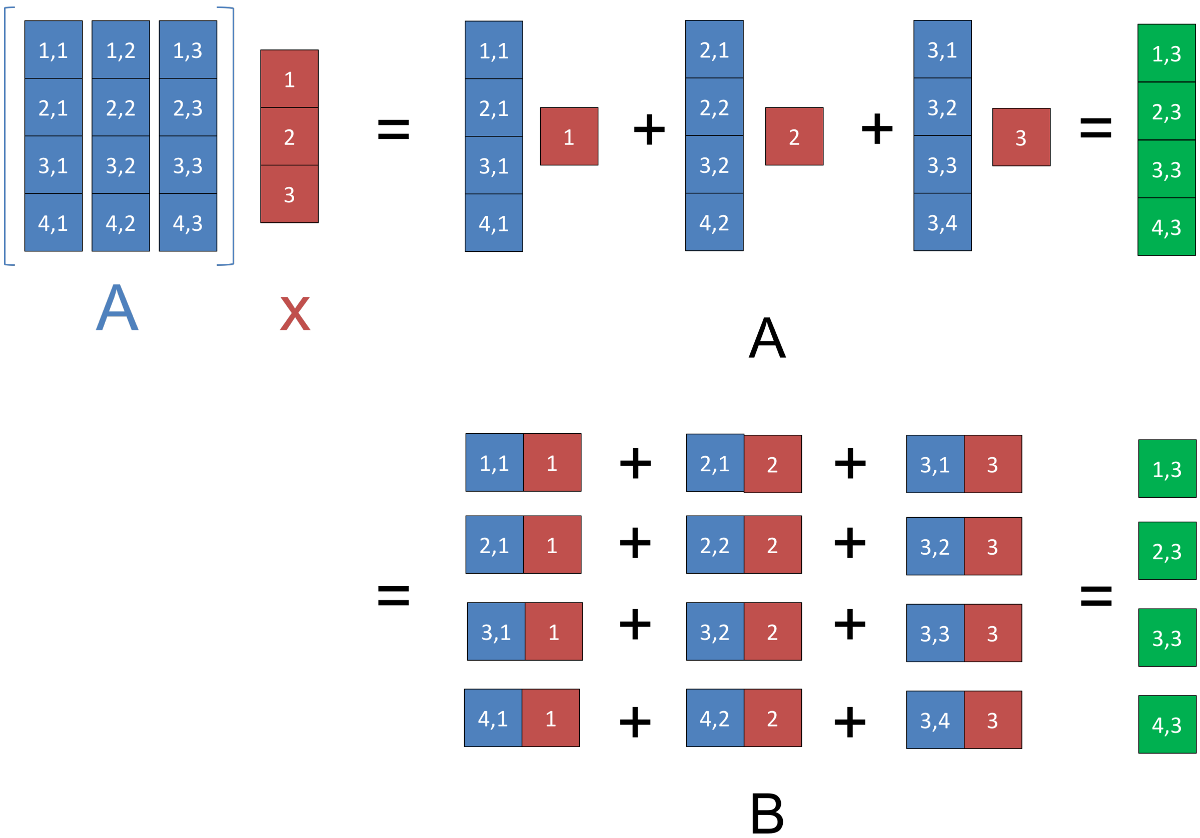

Figure 3: (A) Viewing matrix-vector multiplication as a linear combination of the columns of **A** (B) Viewing matrix-vector multiplication as a the dot product of **x** with each row of the matrix.