Rules of Matrix Multiplication: A Computational Derivation

Max T.

January 2023

Chapter 1

Introduction

1.1 Definition of Matrix Multiplication

For any standard Linear Algebra course, students are taught how to multiply matrices. For an $m \times n$ matrix A and a $n \times p$ matrix B, AB is defined to be

$$AB := \begin{bmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{n1} & \cdots & a_{11}b_{1p} + \cdots + a_{1n}b_{np} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{n1} & \cdots & a_{m1}b_{1p} + \cdots + a_{mn}b_{np} \end{bmatrix} \tag{1.1}$$

where we can write A and B as

$$A := \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \text{ and } B := \begin{bmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{np} \end{bmatrix}$$

for $a_{ij}, b_{jk} \in \mathbb{R}$. Here we assume that matrices A and B contain real-valued entries for convenience. However, all work that goes into the derivation of matrix multiplication can be extended to complex-valued entries.

1.2 Motivation

When matrix multiplication is typically introduced, students are taught that matrix multiplication is a computational method of composing two linear transformations. So, for example, If T and L are linear transformations that have standard matrices A and B respectively, then supposing $T \circ L$ is well-defined, $T \circ L$ has a standard matrix of AB.

Asides from a quick word on compositions of transformations, the rules of matrix multiplication are not typically derived in detail. Understanding matrix multiplication does not require many prerequisites. However, we review the core ideas of Linear Algebra in this chapter as a refresher to the reader.

We then derive the rules of matrix multiplication through a lens of composing linear transformations. It's in our hopes that this paper will clear any mysticism that currently exists regarding the rules of matrix multiplication.

Chapter 2

Prerequisites

2.1 Summation Notation

Let $s_i \in S$ where S is an arbitrary set of ordered elements s_i. In other words, S is a sequence. Assuming that there's a notion of addition in S (an algebra exists on S), we can notate the sum of elements in S to be

$$\sum_{i=n}^{k} s_i = s_n + \cdots + s_k$$

for $n \leq k \leq |S|$ where $|S|$ is the cardinality (size) of S and where $n, k \in \mathbb{N}$. We call n the lower limit of summation and k the upper limit of summation.

2.2 Vector Spaces

For a set V with two binary operations $\cdot$ and $+$ with elements $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ over a field $\mathbb{F}$, the following properties are satisfied.

$$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$$

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$

$$\text{There exists some } \mathbf{0} \in V \text{ such that } \mathbf{u} + \mathbf{0} = \mathbf{u} \text{ for all } \mathbf{u} \in V$$

$$\text{For all } \mathbf{u} \in V \text{ there exists } -\mathbf{u} \in V \text{ such that } \mathbf{u} + (-\mathbf{u}) = \mathbf{0}$$

$$\text{For } a, b \in \mathbb{F},\ a(b\mathbf{u}) = (ab)\mathbf{u}$$

$$\text{There exists an identity element } 1 \in \mathbb{F} \text{ such that } 1\mathbf{u} = \mathbf{u}$$

$$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$$

$$(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$$

Note that we denote scalar multiplication as $a\mathbf{u}$ instead of $a \cdot \mathbf{u}$. We only use $\cdot$ to denote scalar multiplication by itself, not when the operation acts on two elements. The operation itself is implicit hence its omission. Moreover, vector addition takes in $\mathbf{u}, \mathbf{v} \in V$ to produce another vector $\mathbf{v} + \mathbf{u} \in V$ while scalar multiplication takes in $a \in \mathbb{F}$ and $\mathbf{u} \in V$ to produce another vector in V.

For simplicity, we deal with the vector space of $\mathbb{R}^n$ — that is a vector space of real-valued n-tuples — for $n \in \mathbb{N}$ where our field is $\mathbb{R}$. Additionally, vectors in $\mathbb{R}^n$ will be written with arrowheads instead of bold-faced font; $\vec{v}$ instead of $\mathbf{v}$. We use this notation due to convenient visualizations of vectors of $\mathbb{R}^n$ as "arrows" in space. We then call $\vec{0}$ the 0-vector and 1 the multiplicative identity of $\mathbb{R}$.

2.3 Standard Basis Vectors of $\mathbb{R}^n$

For $\mathbb{R}^n$, we define the standard basis to be

$$\mathfrak{B} := \{\vec{e}_1, \ldots, \vec{e}_n\} \tag{2.1}$$

Note that $|\mathfrak{B}| = n$. By being a basis, we meant that for $\vec{v} \in \mathbb{R}^n$, and its components $v_i \in \mathbb{R}$, we can write $\vec{v}$ as

$$\vec{v} := \sum_{i=1}^{n} v_i \vec{e}_i.$$

Here, the intuition is that a vector is defined by "how much" the basis vectors are scaled and combined with addition.

2.4 Inner Product Spaces

An inner product is used to measure the *distance* between two elements of a vector space, *lengths*, and other geometric constructs. For $\vec{x}, \vec{y} \in \mathbb{R}^n$, we define the inner product to be

$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{n} x_i y_i. \tag{2.2}$$

From the inner product, we then define the norm of a vector $\vec{x}$ as

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}.$$

For $\mathbb{R}^n$, we denote the inner product as $\vec{x} \cdot \vec{y}$ and call it the dot-product. Note that each vector space's inner product can be arbitrarily chosen. Unlike the axioms of vector spaces, various properties of the dot-product can be proven rather than assumed. So, we omit properties of the dot-product and leave them as an exercise to the reader to be proven. One property that is particularly nice is that for $\mathbb{R}^2$, $\vec{x} \cdot \vec{y} = \|\vec{x}\|\|\vec{y}\| \cos(\theta)$ where θ is the angle between $\vec{x}$ and $\vec{y}$. This property can be proven from the law of cosines and can be extended into $\mathbb{R}^n$ to define the notion of the angle between two vectors. Used in this expression

2.5 Linear Transformations

A linear transformation $T : V \to W$ — where $\mathbf{u}, \mathbf{v} \in V$ and V and W are vector spaces — follows the following properties:

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$
$$T(k\mathbf{u}) = kT(\mathbf{u})$$
$$T(\mathbf{0}_V) = \mathbf{0}_W$$

where $k \in \mathbb{F}$ where $\mathbb{F}$ is the field of vector-space V and where $\mathbf{0}_V$ and $\mathbf{0}_W$ are the additive identities of vector-spaces V and W respectively.

2.6 The Standard Matrix of a Transformation

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by $\vec{x} \mapsto A\vec{x}$. We know

$$T(\vec{x}) = T\left(\sum_{i=1}^{n} x_1 \vec{e}_i\right) = \sum_{i=1}^{n} x_i T(\vec{e}_i)$$

for $\vec{x} \in \mathbb{R}^n$ and $x_i \in \mathbb{R}$. If we let

$$A := \begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix}$$

then

$$T(\vec{x}) = A\vec{x} = \sum_{i=1}^{n} x_i T(\vec{e}_i)$$

So, to multiply a matrix by a vector, we take a linear combination of the column vectors of the matrix multiplied by the corresponding components of the vector. So, for a matrix

$$A := \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix},$$

then if we define

$$\vec{a}_k := \begin{bmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{bmatrix} = \sum_{i=1}^{m} a_{ik}\vec{e}_i$$

then we can define

$$A\vec{x} = \sum_{k=1}^{n}\sum_{i=1}^{m} x_i\vec{a}_k = \sum_{k=1}^{n}\sum_{i=1}^{m} x_k a_{ik}\vec{e}_i$$

where we can think about this process as summing all of the the column vector-vector products.

Chapter 3

Derivation

3.1 Composition of Two Linear Transformations

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ given by $\vec{y} \mapsto A\vec{y}$ and $L : \mathbb{R}^p \to \mathbb{R}^n$ given by $\vec{x} \mapsto B\vec{x}$. We know that $T \circ L : \mathbb{R}^p \to \mathbb{R}^m$, however, our problem is how do we find the rule of the composition of transformations? Let $\vec{x} \in \mathbb{R}^p$. We know that $(T \circ L)(\vec{x}) = T(L(\vec{x}))$. Because $L(\vec{x}) \in \mathbb{R}^n$, let $\vec{y} := L(\vec{x})$. Then, notice that $L(\vec{x}) = B\vec{x}$, and so $T(\vec{y}) = A\vec{y} = AB\vec{x}$.

Using the ideas presented above, we then say that AB is the standard matrix of $T \circ L$. Because it's computationally inefficient to two matrix-vector products every time we want to compute the mapping, we want to compute AB. How do we do so?

We first posit there exists a matrix $C \in \mathbb{R}^{m \times p}$ such that

$$AB\vec{x} = C\vec{x}.$$

We will first compute the LHS. We first start by defining A and B in terms of their column vectors. Let $\vec{a}_i = \sum_{j=1}^{m} a_{ji}\vec{e}_j$ and $\vec{b}_k = \sum_{j=1}^{n} b_{jk}\vec{e}_j$. We then see that $AB\vec{x} = C\vec{x}$ can be written as

$$\begin{aligned}
\begin{bmatrix}\vec{a}_1 & \cdots & \vec{a}_n\end{bmatrix}\begin{bmatrix}\vec{b}_1 & \cdots & \vec{b}_p\end{bmatrix}\vec{x} &= C\vec{x} \\
\begin{bmatrix}\vec{a}_1 & \cdots & \vec{a}_n\end{bmatrix}\sum_{k=1}^{p} x_k\vec{b}_k &= C\vec{x} \\
\sum_{k=1}^{p} x_k A\vec{b}_k &= C\vec{x} \\
\sum_{k=1}^{p}\sum_{i=1}^{n} x_k(\vec{b}_k \cdot \vec{e}_k)\vec{a}_i &= C\vec{x} \\
\implies \sum_{k=1}^{p}\sum_{i=1}^{n} x_k(\vec{b}_k \cdot \vec{e}_k)\vec{a}_i &= \sum_{k=1}^{p} x_k\vec{c}_k
\end{aligned}$$

for $\vec{c}_k = \sum_{j=1}^{m} c_{jk}\vec{e}_j$. Or, using matrices,

$$\begin{bmatrix}
\sum_{j=1}^{n} a_{1i}b_{i1} & \cdots & \sum_{j=1}^{n} a_{1i}b_{ip} \\
\vdots & \ddots & \vdots \\
\sum_{j=1}^{n} a_{mi}b_{i1} & \cdots & \sum_{j=1}^{n} a_{mi}b_{ip}
\end{bmatrix} = C$$

It's here that we conclude the derivation.