05. *Matrix multiplication–the Concept*

0.1 Start Small

$$\text{“} \quad \begin{pmatrix}1&2\\3&4\end{pmatrix}\begin{pmatrix}5\\7\end{pmatrix} \equiv \begin{pmatrix}1\cdot5+2\cdot7\\3\cdot5+4\cdot7\end{pmatrix} = \begin{pmatrix}19\\43\end{pmatrix}. \text{”}$$

Consider a single linear equation in n variables:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

We can rewrite this equation by encoding the coefficients $\{a_i\}$ in a row and the unknowns $\{x_i\}$ in a column:

$$\begin{pmatrix}a_1 & a_2 & \cdots a_n\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix} = b\,.$$

This is the first instance of matrix multiplication: we multiply a $1\times n$ matrix (in the shape of a row) by an $n\times 1$ matrix (in the shape of a column) to obtain a scalar, or more formally, a 1×1 matrix.

We can extend this notion to matrices with multiple rows and columns by “stacking” individual row-column products. For example, take 2×2 matrices A, B and multiply:

$$A\cdot B \equiv \begin{pmatrix}1&2\\3&4\end{pmatrix}\begin{pmatrix}5&6\\7&8\end{pmatrix} \equiv \begin{pmatrix}19&22\\43&50\end{pmatrix}.$$

Where does this come from, and what does it mean? Well, break the leftmost matrix into row and the next matrix into columns, and multiply:

$$\begin{pmatrix}1&2\\3&4\end{pmatrix}\left(\begin{pmatrix}5\\7\end{pmatrix}\begin{pmatrix}6\\8\end{pmatrix}\right) \equiv \begin{pmatrix}\begin{pmatrix}1&2\end{pmatrix}\begin{pmatrix}5\\7\end{pmatrix} & \begin{pmatrix}1&2\end{pmatrix}\begin{pmatrix}6\\8\end{pmatrix}\\ \begin{pmatrix}3&4\end{pmatrix}\begin{pmatrix}5\\7\end{pmatrix} & \begin{pmatrix}3&4\end{pmatrix}\begin{pmatrix}6\\8\end{pmatrix}\end{pmatrix} \tag{tl}$$

$$= \begin{pmatrix}1\cdot5+2\cdot7 & 1\cdot6+2\cdot8\\3\cdot5+4\cdot7 & 3\cdot6+4\cdot8\end{pmatrix}$$

$$= \begin{pmatrix}19&22\\43&50\end{pmatrix}.$$

Before we leave this matrix product example, let's view it from another perspective. Focus on the first column of the right matrix in (tl): we can also view it as 5 times the left column of A plus 7 times the right column of A, i.e., the “linear combination” $5\binom{1}{3}+7\binom{2}{4}$. Similarly, the right column of the right matrix in (tl) can be written as a linear combination of the columns of A: $6\binom{1}{3}+8\binom{2}{4}$.

$$\begin{pmatrix}1&2\\3&4\end{pmatrix}\begin{pmatrix}5&6\\7&8\end{pmatrix} \equiv \left(5\begin{pmatrix}1\\3\end{pmatrix}+7\begin{pmatrix}2\\4\end{pmatrix} \quad 6\begin{pmatrix}1\\3\end{pmatrix}+8\begin{pmatrix}2\\4\end{pmatrix}\right)$$

$$= \left(\begin{pmatrix}5\cdot1+7\cdot2\\5\cdot3+7\cdot4\end{pmatrix} \quad \begin{pmatrix}6\cdot1+8\cdot2\\6\cdot3+8\cdot4\end{pmatrix}\right)$$

$$= \left(\begin{pmatrix}19\\43\end{pmatrix} \quad \begin{pmatrix}22\\50\end{pmatrix}\right).$$

We write:

- *The columns of AB are linear combinations of the columns of A, with scalars provided by columns of B.*
- *The rows of AB are linear combinations of the rows of B, with scalars provided by rows of A.*

We haven't yet earned the right to the second bubble point, but it's entirely analogous to the first:

$$\begin{pmatrix}1&2\\3&4\end{pmatrix}\begin{pmatrix}5&6\\7&8\end{pmatrix} \equiv \begin{pmatrix}1\cdot\begin{pmatrix}5&6\end{pmatrix}+2\cdot\begin{pmatrix}7&8\end{pmatrix}\\3\cdot\begin{pmatrix}5&6\end{pmatrix}+4\cdot\begin{pmatrix}7&8\end{pmatrix}\end{pmatrix}.$$

This illustrates the computational side of matrix multiplication. It's important to link that with conceptual aspects.

0.2 Many equations in Numerous unknowns

Matrix multiplication may be interpreted in several ways, all of which are technically equivalent. However, each one can contribute insight in its own way. The reader is encouraged to explore and contemplate as many of these perspectives as possible. Here we will begin with the general form of a linear system of m equations in n unknowns:

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1\\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2\\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= b_3\\ &\vdots\\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= b_m\end{aligned} \tag{1}$$

We summarize this in the compactified notation

$$A\vec{x} = \vec{b}. \tag{2}$$

Here, A is the $m\times n$ matrix whose (i,j) entry (the entry on the i^{th} row, at position j from left) is a_{ij}. Thus A is the *coefficient matrix* of the system (1). The expression $\vec{x}$ encodes the unknowns, or variables, $x_1,\dots x_n$, stored in the $n\times 1$ column vector below:

$$\begin{pmatrix}x_1\\\vdots\\x_n\end{pmatrix};$$

similarly, $\vec{b}$ encodes the m right hand sides of the system (1), stored in the $m\times 1$ column vector with entries $b_1,\dots,b_m$ (top to bottom, by convention).

©elg Copyrighted Material

0.3 Functional Understanding

We can view the matrix A as a *function* that takes an $n \times 1$ column vector $\vec{x}$ as input, and produces an $m \times 1$ column vector $A\vec{x}$ as output. Sometimes we'll write $A : \vec{x} \mapsto A\vec{x}$. For example, take matrices C, D as follows:

$$C \equiv \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \quad ; \quad D \equiv \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The reader can verify that the matrices C, D, viewed as the functions $\left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right) \mapsto C\left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right) \mapsto D\left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right)$, respectively, can be represented as follows:

$$C : \left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right) \mapsto \left(\begin{smallmatrix} u-w \\ u-w \end{smallmatrix}\right) \quad ; \quad D : \left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right) \mapsto \left(\begin{smallmatrix} u+w \\ u+w \end{smallmatrix}\right).$$

Now view the expression DC as the composition of functions

$$DC\left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right) \equiv D\left(C\left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right)\right).$$

Evidently, $DC\left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right) = D\left(\begin{smallmatrix} u-w \\ u-w \end{smallmatrix}\right) = \left(\begin{smallmatrix} 2u-2w \\ 2u-2w \end{smallmatrix}\right)$, after minor algebra. Thus, the function DC takes a 2×1 vector $\left(\begin{smallmatrix} u \\ w \end{smallmatrix}\right)$ and produces the 2×1 vector $\left(\begin{smallmatrix} 2u-2w \\ 2u-2w \end{smallmatrix}\right)$.

This job can also be done by the matrix $\left(\begin{smallmatrix} 2 & -2 \\ 2 & -2 \end{smallmatrix}\right)$, viewed as a function; hence DC must *be* the 2×2 matrix $\left(\begin{smallmatrix} 2 & -2 \\ 2 & -2 \end{smallmatrix}\right)$, by the following principle.

Proposition. *"The Function Determines the Matrix"*
If E, F are two $p \times q$ matrices which are the same as functions, that is, $E\vec{x} = F\vec{x}$ for every $q \times 1$ column vector $\vec{x}$, then E and F are equal as matrices.

The reader is invited to verify this result, which simply depends on unraveling definitions. Thus we can write

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ 2 & -2 \end{pmatrix}.$$

Working in the opposite order, the reader will easily compute the composition of functions and verify that

$$\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

This makes salient some contrasting features of matrix multiplication, e.g., the product of two non-zero matrices can be zero (the zero matrix, of appropriate size, that is); the order of multiplication matters in the matrix domain. (Matrix multiplication is *non-commutative*, that is, *order matters*).

Sometimes matrix multiplication is introduced by means of a formula. The general situation is an extension of the most basic one: the product of a $1 \times p$ row vector, say $\vec{a}$, and a $p \times 1$ column vector, say $\vec{b}$:

$$\begin{pmatrix} a_1 & a_2 & \cdots a_p \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} \equiv a_1 b_1 + a_2 b_2 + \ldots + a_p b_p, \qquad (3)$$

which can also be written as $\sum_{i=1}^{p} a_i b_i$. This product is regarded as a *scalar*, or, if you insist, as a 1×1 matrix. In some contexts it is called *the scalar product* or *the dot product* of the respective vectors $\vec{a}, \vec{b}$. Can we justify this using our *function* approach? (And can we justify the print space needed to discuss it?) We can. To save a little space, the $p\times1$ column matrix $\vec{b}$, and other column matrices like it, will be denoted as $\begin{pmatrix} b_1 & b_2 & \cdots & b_p \end{pmatrix}^T$. The operation $(\cdot)^T$ is called *transpose*; it turns rows into columns and columns into rows, by rotation, say. *Transpose* is more than just a space-saving device. It is quite profound, actually, and its simplicity betrays its significance. Much more on that will be presented later. To make even better use of space, we'll move to an appendix at the end of this reading the discussion of the row-column matrix product from the function perspective.

©elg Copyrighted Material

0.4 Matrix Multiplication--General Picture

The multiplication of two general matrices, GH, where G has exactly as many columns as H has rows, is manifested from the basic row-column multiplication rule (3) by simply "bunching up" the appropriate row-column products, as in (2). Say G is $m \times r$ and H is $r \times n$. Denote the i^{th} row of G as $(\, g_{i,1}\; g_{i,2}\; \cdots\; g_{i,r}\,)$ and denote the j^{th} column of H as $(\, h_{1,j} \cdots\; h_{2,j}\; \cdots\; h_{r,j}\,)^T$, where we have used the *transpose* notation above to denote a column vector with horizontal printing. Then we can describe the product as follows:

$$GH = \begin{pmatrix} & \vdots & \\ g_{i,1} & g_{i,2} \cdots & g_{i,r} \\ & \vdots & \end{pmatrix} \begin{pmatrix} & h_{1,j} & \\ \cdots & h_{2,j} & \cdots \\ & \vdots & \\ & h_{r,j} & \end{pmatrix} = \begin{pmatrix} & \vdots & \\ \cdots & p_{i,j} & \cdots \\ & \vdots & \end{pmatrix},$$

where

$$p_{i,j} = g_{i,1} h_{1,j} + g_{i,2} h_{2,j} + \cdots + g_{i,r} h_{r,j} = \sum_{k=1}^{r} g_{i,k} h_{k,j},$$

i.e., the ij^{th} entry of the product matrix GH is the dot product of the i^{th} row of G with the j^{th} column of H.

Appendix

(You may wish to skip this on a first reading)

Row-Column matrix product:
the function perspective

(Not everyone finds this useful)

How can we view the $p \times 1$ column vector $\begin{pmatrix} b_1 & b_2 & \cdots & b_p \end{pmatrix}^T$ as a *function*? Well, it can operate on the 1×1 vector x to yield the $p \times 1$ column vector $x\vec{b}$, i.e., $\begin{pmatrix} b_1 \cdot x & b_2 \cdot x & \cdots & b_p \cdot x \end{pmatrix}^T$.

How can we view the $1 \times p$ row vector $\vec{a} = \begin{pmatrix} a_1 & a_2 & \cdots a_p \end{pmatrix}$ as a function? Well, it can take the column vector $\vec{y} = \begin{pmatrix} y_1 & y_2 & \cdots & y_p \end{pmatrix}^T$ to the 1×1 matrix $\left(\sum_{i=1}^{p} a_i y_i\right)$. (This comes from a system of one linear equation in p unknowns.) Composing the row vector $\vec{a}$, viewed as a $1 \times p$ matrix, and as a function, with the column matrix $\vec{b}$, viewed as a $p \times 1$ matrix and also a function, we get a new function, taking a 1×1 vector x (which is both a column and a row vector at the same time) to an output value as follows:

$$x \mapsto (a_1 b_1 + a_2 b_2 + \cdots + a_p b_p)x = \left(\sum_{i+1}^{p} a_i b_i \right) x,$$

hence it may be viewed as the scalar (or 1×1 matrix, if you insist) $\left(\sum_{i+1}^{p} a_i b_i\right)$.

We leave it to the reader to decide if this function approach to the row-column product is worthy. In the opinion of this writer the function perspective is more satisfying when viewed in the context of larger-sized matrices, starting with 2×2.

©elg Copyrighted Material