Matrix notation and multiplication

COMS 3251 Fall 2022 (Daniel Hsu)

1 Matrices

Much of linear algebra is about developing useful "algebraic" notations for expressing the central concepts of the subject. These notations will help us see connections between those concepts. But it takes a bit of time and practice to "get used to" the notations.

A *matrix* (which has the plural form *matrices*) is a list of vectors. We visually represent a matrix using a table, listing the vectors one after another from left to right. For example, suppose we have the vectors $\mathbf{a}_1 = (1, 2, 3)$, $\mathbf{a}_2 = (2, 4, 6)$, $\mathbf{a}_3 = (3, 4, 5)$, and $\mathbf{a}_4 = (4, 4, 4)$. Then, the matrix A with these vectors, in order, is

$$A = \begin{bmatrix} \uparrow & \uparrow & \uparrow & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix}. \tag{1}$$

(The arrows here are printed just for visual effect; they don't mean anything.) The jth *column* of A is $\mathbf{a}_j$. The ith *row* of A lists ith components of all of the vectors, in order. Note that a more literal depiction of the ordered list is

$$\left[\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} \right].$$

But that is aesthetically displeasing, and it is better to remove the clutter. It will also turn out that the rows of matrices are interesting, even though we have defined the matrices column-wise.[1]

The matrix A shown above has 3 rows and 4 columns. A matrix with m rows and n columns is said to have *shape* (a.k.a. *dimensions*) $m \times n$. It is important to remember this ordering: # rows by # columns.

[1] We could have alternatively adopted a different convention to always write vectors as rows, and then define a matrix to be a stack of vectors, listed from top to bottom. All of linear algebra would work exactly the same way, with the obvious interchanges of rows and columns. See also, e.g., Section 6.

We can add two matrices together (provided they have the same shape), because this just corresponds to adding together the corresponding vectors in the two lists of vectors:

$$\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \downarrow & & \downarrow \end{bmatrix} + \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{b}_1 & \cdots & \mathbf{b}_n \\ \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{a}_1 + \mathbf{b}_1 & \cdots & \mathbf{a}_n + \mathbf{b}_n \\ \downarrow & & \downarrow \end{bmatrix}.$$

And we can scale a matrix by a real number, because this just corresponds to scaling each of the vectors in the list by the real number:

$$c \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & & \uparrow \\ c\,\mathbf{a}_1 & \cdots & c\,\mathbf{a}_n \\ \downarrow & & \downarrow \end{bmatrix}.$$

2 Matrix-vector multiplication

The *(matrix-vector) multiplication* (a.k.a. *product*) of an $m \times n$ matrix A by an n-vector $\mathbf{x}$, written $A\mathbf{x}$, is defined to yield the m-vector

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \sum_{k=1}^{n} x_k\mathbf{a}_k,$$

where $\mathbf{a}_k$ is the kth column of A, and $\mathbf{x} = (x_1, \ldots, x_n)$. In other words, it is just notation for writing a particular linear combination of the columns of A, as specified by $\mathbf{x}$. The *column space* of a $m \times n$ matrix A is $\mathsf{CS}(A) = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\}$, the set of all vectors $A\mathbf{x}$ as $\mathbf{x}$ ranges over all n-vectors.[2] Notice that this is the same as the span of the columns of A.

Example. If A is the matrix from (1), and $\mathbf{x} = (-2, 0, 2, 0)$, then

$$A\mathbf{x} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 2 \\ 0 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 0 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + 0 \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}.$$

[2] Some texts refer to the column space of A by $\mathrm{C}(A)$, while others use $\mathrm{R}(A)$.

And if $\mathbf{y} = (-2, 0, 2, -1)$, then

$$A\mathbf{y} \;=\; -2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 0 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + -1 \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} \;=\; \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \;=\; \mathbf{0}. \qquad \blacksquare$$

Note. You may have heard of another way to multiply a matrix by a vector, which computes one component of the answer at a time. If we write $A_{i,j}$ for the (i, j)th component of the matrix A—i.e., the value in ith row and jth column of A—then the ith component of $A\mathbf{x}$ is

$$A_{i,1}x_1 + \cdots + A_{i,n}x_n \;=\; \sum_{j=1}^{n} A_{i,j}x_j,$$

where $\mathbf{x} = (x_1, \ldots, x_n)$. So, again in our example with A from (1) and $\mathbf{x} = (-2, 0, 2, 0)$, the first component of the answer is

$$1 \times (-2) + 2 \times 0 + 3 \times 2 + 4 \times 0 \;=\; 4,$$

and the third component of the answer is

$$3 \times (-2) + 6 \times 0 + 5 \times 2 + 4 \times 0 \;=\; 4.$$

This, of course, is an equivalent definition of $A\mathbf{x}$, and in some cases, it is a more convenient way to think about $A\mathbf{x}$. (For instance, if you need to compute $A\mathbf{x}$ by hand, this is usually the way to go.) ∎

Example. Does the following system of linear equations (in terms of the variables x and y) have a solution?

$$\begin{cases} x + 3y = 4 \\ 2x + 4y = 4 \\ 3x + 5y = 4 \end{cases}$$

This is equivalent to asking if the vector $(4, 4, 4)$ is in the column space of the matrix

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

We may thus write the system of linear equations using matrix-vector multiplication:

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}.$$

▮

We have defined matrix-vector multiplication $A\mathbf{x}$ in a way that treats the vector $\mathbf{x}$ as the action being performed on the matrix A—it linearly combines the columns of A:

$$A = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \downarrow & & \downarrow \end{bmatrix} \xmapsto{\mathbf{x} = (x_1,\dotsc,x_n)} x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = A\mathbf{x}.$$

But matrix-vector multiplication $A\mathbf{x}$ can also be viewed in a way where A is the action being performed on the vector $\mathbf{x}$—a transformation of $\mathbf{x}$ into another vector:

$$\mathbf{x} \xmapsto{A} A\mathbf{x}.$$

This transformation has a very special property: linearity.

3 Linearity

As asserted above, a very important property of matrix-vector multiplication is linearity, which (as you may recall) is really two separate properties put together.

Additivity: $A(\mathbf{x}+\mathbf{y}) = (A\mathbf{x}) + (A\mathbf{y})$ for all vectors $\mathbf{x}$ and $\mathbf{y}$.

Left-hand side: Add the $\mathbf{x}$ and $\mathbf{y}$, then multiply A by the result.

Right-hand side: Multiply A by $\mathbf{x}$ and also by $\mathbf{y}$, then add the results.

Homogeneity: $A(c\mathbf{x}) = c(A\mathbf{x})$ for all vectors $\mathbf{x}$ and scalars c.

Left-hand side: Scale the vector $\mathbf{x}$ by c, then multiply A by the result.

Right-hand side: Multiply A by $\mathbf{x}$, then scale the result by c.

(Above, A is an $m \times n$ matrix, $\mathbf{x}$ and $\mathbf{y}$ are n-vectors, and c is a scalar.)

Put together, for all vectors $\mathbf{x}$ and $\mathbf{y}$ and all scalars c, we have

$$A(c\mathbf{x}+\mathbf{y}) = c(A\mathbf{x}) + (A\mathbf{y}). \tag{2}$$

Example. Suppose A is the matrix from (1), $\mathbf{x} = (0, 1, 0, 1)$, $\mathbf{y} = (2, -1, 1, 0)$, and $c = 2$. Then $c\mathbf{x} + \mathbf{y} = (0, 2, 0, 2) + (2, -1, 1, 0) = (2, 1, 1, 2)$, so we have

$$A(c\mathbf{x} + \mathbf{y}) \;=\; \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix} \;=\; \begin{bmatrix} 15 \\ 20 \\ 25 \end{bmatrix}.$$

Moreover, $A\mathbf{x} = (6, 8, 10)$ and $A\mathbf{y} = (3, 4, 5)$, so

$$c(A\mathbf{x}) + (A\mathbf{y}) \;=\; 2 \begin{bmatrix} 6 \\ 8 \\ 10 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \;=\; \begin{bmatrix} 15 \\ 20 \\ 25 \end{bmatrix}.$$ ∎

In general, if the columns of A are the m-vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$, and $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$, then (2) says

$$(cx_1 + y_1)\mathbf{a}_1 + \cdots + (cx_n + y_n)\mathbf{a}_n \;=\; c(x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n) + (y_1\mathbf{a}_1 + \cdots + y_n\mathbf{a}_n).$$

Suppose $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. We claim that there is an $m \times n$ matrix A_T such that $T(\mathbf{x}) = A_T\mathbf{x}$ for all n-vectors $\mathbf{x}$. The matrix A_T is given by

$$A_T \;=\; \begin{bmatrix} \uparrow & & \uparrow \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ \downarrow & & \downarrow \end{bmatrix},$$

where $\mathbf{e}_i$ is the n-vector whose jth component is 1 if and only if $i = j$. Note that every n-vector $\mathbf{x} = (x_1, \dots, x_n)$ can be written as a linear combination $x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$ of $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. So, by linearity of T, we have

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) \\ &= x_1\,T(\mathbf{e}_1) + \cdots + x_n\,T(\mathbf{e}_n) \\ &= A_T\mathbf{x} \end{aligned}$$

for the matrix A_T we have just defined.

The following theorem summarizes the facts of this section.

Theorem 1. *For any linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$, there is an $m \times n$ matrix A_T such that $T(\mathbf{x}) = A_T\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. Conversely, for any $m \times n$ matrix A, the transformation $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ given by $T_A(\mathbf{x}) = A\mathbf{x}$ is linear.*

Theorem 1 is what makes matrix notation so powerful: it lets us describe every linear transformation in a very "algebraic" fashion.

There is one more way in which linearity arises with matrix-vector multiplication. Specifically, if $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ and $B = [\mathbf{b}_1, \dots, \mathbf{b}_n]$ are both $m \times n$ matrices, and c is a scalar, then for any n-vector $\mathbf{x} = (x_1, \dots, x_n)$,

$$(c\,A + B)\mathbf{x} \;=\; c\,A\mathbf{x} + B\mathbf{x}.$$

Indeed, we can either first linearly combine corresponding columns of A and B before multiplying by $\mathbf{x}$, or we can first multiply each of A and B by $\mathbf{x}$, and then linearly combine the results:

$$\begin{aligned}
(c\,A + B)\mathbf{x} \;&=\; \left(c \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \downarrow & & \downarrow \end{bmatrix} + \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{b}_1 & \cdots & \mathbf{b}_n \\ \downarrow & & \downarrow \end{bmatrix} \right) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&=\; \begin{bmatrix} \uparrow & & \uparrow \\ c\,\mathbf{a}_1 + \mathbf{b}_1 & \cdots & c\,\mathbf{a}_n + \mathbf{b}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&=\; x_1\,(c\,\mathbf{a}_1 + \mathbf{b}_1) + \cdots + x_n\,(c\,\mathbf{a}_n + \mathbf{b}_n) \\
&=\; c\,(x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n) + (x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n) \\
&=\; c \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{b}_1 & \cdots & \mathbf{b}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&=\; c\,A\mathbf{x} + B\mathbf{x}.
\end{aligned}$$

4 Matrix multiplication

We now come to the main event: matrix-matrix multiplication, usually shortened to matrix multiplication.

Motivation. Suppose you have linear transformations $G \colon \mathbb{R}^n \to \mathbb{R}^m$ and $F \colon \mathbb{R}^p \to \mathbb{R}^n$. The *composition* of G and F is another transformation, written $G \circ F$ (and read aloud as "G on F" or "G composed with F"), that applies G to the result of applying F:

$$(G \circ F)(\mathbf{x}) = G(F(\mathbf{x})).$$

The input space for the composition is $\mathbb{R}^p$ (which is the input space of F), and output space is $\mathbb{R}^m$ (which is the output space of G). Of course, it is important that the input space of G is the same as the output space of F, or else it would not be possible to apply G to the output of F. The composition of linear transformations is also a linear transformation: for any p-vectors $\mathbf{x}$ and $\mathbf{y}$, and any scalar c,

$$\begin{aligned} G(F(c\,\mathbf{x}+\mathbf{y})) &= G(c\,F(\mathbf{x})+F(\mathbf{y})) && \text{(by linearity of } F\text{)} \\ &= c\,G(F(\mathbf{x}))+G(F(\mathbf{y})) && \text{(by linearity of } G\text{).} \end{aligned}$$

Each of G, F, and the composition $G \circ F$ are linear transformations. So Theorem 1 implies that for each of them, there is a matrix such that the linear transformation is the same as multiplying the input by that matrix. If A is the matrix for G, and B is the matrix for F, then what is the matrix for $G \circ F$? Answer: the matrix multiplication of A and B.[3] ∎

The *(matrix) multiplication* (a.k.a. *product*) of an $m \times n$ matrix A by an $n \times p$ matrix B, written AB, is defined to yield the $m \times p$ matrix Z whose kth column is

$$\mathbf{z}_k \;=\; A\mathbf{b}_k,$$

where $Z = [\mathbf{z}_1, \dots, \mathbf{z}_p]$ and $B = [\mathbf{b}_1, \dots, \mathbf{b}_p]$. So each column of Z is a linear combination of the columns of A; there are p such linear combinations, each being specified by a column of B:

$$Z \;=\; \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{z}_1 & \cdots & \mathbf{z}_p \\ \downarrow & & \downarrow \end{bmatrix} \;=\; \begin{bmatrix} \uparrow & & \uparrow \\ A\mathbf{b}_1 & \cdots & A\mathbf{b}_p \\ \downarrow & & \downarrow \end{bmatrix}.$$

This strictly generalizes matrix-vector multiplication, which is recovered in the special case where $p = 1$ (where B has just one column).

Example. Let A be the matrix from (1), and let B be the matrix whose columns are $(0, 1, 0, 1)$ and $(2, -1, 1, 0)$. Then

$$AB \;=\; \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \;=\; \begin{bmatrix} 6 & 3 \\ 8 & 4 \\ 10 & 5 \end{bmatrix}.$$ ∎

[3] The reason this is the answer should become clear after we establish the associativity property.

Another example. Again, A is the matrix from (1), and B is the matrix whose columns are $(-2, 0, 2, 0)$ and $(2, -1, 0, 1)$. Then

$$AB = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ 0 & -1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 4 \\ 4 & 4 \end{bmatrix}.$$ ∎

Caution. Matrix multiplication is unforgiving of "type errors".

- Can we multiply a 3×4 matrix by a 3×4 matrix?

  No, this doesn't make sense: The first matrix is a list of 4 column vectors, while each column of the second matrix is a 3-vector.

- What about multiplying a 1×3 matrix by a 3×1 matrix?

  Yes, this is fine!

- What about multiplying a 3×1 matrix by a 1×3 matrix?

  Also fine!

- Rule: Multiplying an $m \times n$ matrix by a $p \times q$ matrix requires $n = p$. ∎

Finally, we now see the "standard" (but equivalent) definition of matrix multiplication: If $A_{i,j}$ is the (i, j)th component of A, and $B_{j,k}$ is the (j, k)th component of B, then the (i, k)th component of AB is

$$A_{i,1}B_{1,k} + \cdots + A_{i,n}B_{n,k} = \sum_{j=1}^{n} A_{i,j}B_{j,k}. \tag{3}$$

5 Properties of matrix multiplication

5.1 Associativity

An essential property of matrix multiplication is *associativity*: For matrices A, B, and C (such that AB and BC both make sense), we have

$$A(BC) = (AB)C. \tag{4}$$

The reason this holds is really the same as the linearity of matrix-vector multiplication. For now, just consider the case where C has just a single column, $\mathbf{c}$, so (4) becomes

$$A(B\mathbf{c}) \;=\; (AB)\mathbf{c}. \tag{5}$$

Multiplying another matrix by $\mathbf{c}$ results in a linear combination of the columns of that other matrix. In the left-hand side of (5), we first linearly combine the columns of B before multiplying A by the result. In the right-hand side of (5), we first multiply A by each column of B, and then linearly combine the results. By the linearity of matrix-vector multiplication (in the step below marked by $*$), these two approaches give the same result:

$$\begin{aligned}
A(B\mathbf{c}) &= A\left(\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{b}_1 & \cdots & \mathbf{b}_p \\ \downarrow & & \downarrow \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}\right) \\
&= A(c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p) \\
&\overset{*}{=} c_1(A\mathbf{b}_1) + \cdots + c_p(A\mathbf{b}_p) \\
&= \begin{bmatrix} \uparrow & & \uparrow \\ A\mathbf{b}_1 & \cdots & A\mathbf{b}_p \\ \downarrow & & \downarrow \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \\
&= \left(A\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{b}_1 & \cdots & \mathbf{b}_p \\ \downarrow & & \downarrow \end{bmatrix}\right)\begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \;=\; (AB)\mathbf{c}.
\end{aligned}$$

Example.

$$A(B\mathbf{c}) \;=\; \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix}\left(\begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) \;=\; \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix} \;=\; \begin{bmatrix} 15 \\ 20 \\ 25 \end{bmatrix},$$

and

$$(AB)\mathbf{c} \;=\; \left(\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix}\begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\right)\begin{bmatrix} 2 \\ 1 \end{bmatrix} \;=\; \begin{bmatrix} 6 & 3 \\ 8 & 4 \\ 10 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} \;=\; \begin{bmatrix} 15 \\ 20 \\ 25 \end{bmatrix}. \qquad \blacksquare$$

If C has more than one column, the reasoning given above just needs to be applied separately for each column of C.

On account of this associativity property, when we write $A(BC)$ or $(AB)C$, we can remove the parentheses and simply write ABC.

5.2 Non-commutativity

A matrix with the same number of rows and columns is a *square matrix*. If A and B are square matrices of the same shape, then both AB and BA are always valid matrix multiplications.

Matrix multiplication is not generally *commutative*. That is, it is possible that $AB \neq BA$ for matrices A and B, even if they are square. Of course, there are some exceptions where $AB = BA$ does hold, but this is more the exception rather than the rule. (Some exceptions are discussed below.) Thus, matrix multiplication is said to be *non-commutative*.

Example.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 7 & 7 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 4 & 6 \end{bmatrix}.$$

The non-commutativity should not be surprising if you think of the first matrix as specifying a list of vectors, and the second matrix as specifying how you want to linearly combine the vectors from the first matrix. These matrices have very different roles!

5.3 Identity and diagonal matrices

A special exception to the non-commutativity rule is multiplication by the $n \times n$ *identity matrix* I_n, defined by

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

(We'll drop the subscript n when it is clear from context.) The identity matrix has value 1 along the diagonal entries (i.e., the (j, j)th entry of I is 1 for all $j \in \{1, \dots, n\}$) and value 0 in all other entries. For any $n \times n$ matrix A, we have $AI = IA = A$.

Another important class of square matrices are those for which all non-diagonal entries have value 0: these are called *diagonal matrices*. The identity matrix is an example of a diagonal matrix. If D and S are both $n \times n$ diagonal matrices, then $DS = SD$, i.e., D and S *commute*! The result DS is also an $n \times n$ diagonal matrix; its (i, i)th entry is the product of the (i, i)th entry of D and the (i, i)th entry of S. Commutativity holds because scalar multiplication is commutative ($xy = yx$ for any scalars x and y).

Diagonal matrices are also "nice" in that computing the matrix-vector product $D\mathbf{x}$ for a $n \times n$ diagonal matrix D only requires n scalar multiplications, whereas computing $A\mathbf{x}$ for a general $n \times n$ matrix typically requires n^2 scalar multiplications.

6 Row-oriented matrices

We have hinted that the rows of matrices, even when defined column-wise, will be of interest to us. We aren't quite ready to see why just yet, but to prepare for it, let us temporarily enter the alternate universe where matrices are defined row-wise and interpret matrix multiplication from that perspective.

Let the rows of the $m \times n$ matrix A be the *row vectors* $\mathbf{a}_1^{\mathsf{T}}, \dots, \mathbf{a}_m^{\mathsf{T}}$. The symbol in the superscript, "T", is the *transpose* symbol. For now, we only use it to indicate that we want to change a column to a row: $\mathbf{a}_i \in \mathbb{R}^n$ is an n-vector (treated as a column), while $\mathbf{a}_i^{\mathsf{T}}$ is a row vector. And similarly, let the rows of the $n \times p$ matrix B be the row vectors $\mathbf{b}_1^{\mathsf{T}}, \dots, \mathbf{b}_n^{\mathsf{T}}$. So

$$A = \begin{bmatrix} \longleftarrow & \mathbf{a}_1^{\mathsf{T}} & \longrightarrow \\ & \vdots & \\ \longleftarrow & \mathbf{a}_m^{\mathsf{T}} & \longrightarrow \end{bmatrix}, \quad B = \begin{bmatrix} \longleftarrow & \mathbf{b}_1^{\mathsf{T}} & \longrightarrow \\ & \vdots & \\ \longleftarrow & \mathbf{b}_n^{\mathsf{T}} & \longrightarrow \end{bmatrix}.$$

(The arrows here are printed just for visual effect; they don't mean anything.) For an n-vector $\mathbf{x} = (x_1, \dots, x_n)$ and an $n \times p$ matrix B as above, $\mathbf{x}^{\mathsf{T}} B$ is the row vector obtained by linearly combining the rows of B:

$$\mathbf{x}^{\mathsf{T}} B = x_1 \mathbf{b}_1^{\mathsf{T}} + \cdots + x_n \mathbf{b}_n^{\mathsf{T}}.$$

Thus, the multiplication of A by B yields the $m \times p$ matrix C whose ith row $\mathbf{c}_i^\top$ is a linear combination of the rows of B as specified by $\mathbf{a}_i^\top$, written

$$C = \begin{bmatrix} \longleftarrow & \mathbf{c}_1^\top & \longrightarrow \\ & \vdots & \\ \longleftarrow & \mathbf{c}_m^\top & \longrightarrow \end{bmatrix} = \begin{bmatrix} \longleftarrow & \mathbf{a}_1^\top B & \longrightarrow \\ & \vdots & \\ \longleftarrow & \mathbf{a}_m^\top B & \longrightarrow \end{bmatrix}.$$

7 Block-wise matrix multiplication

One advantage of the "standard" definition of matrix multiplication from (3) is that it easily generalizes to *block-wise matrix multiplication*. We'll just give three illustrative examples of this, rather than define it in full generality.

Example: four blocks. Let A be an $m \times n$ matrix, and let B be an $n \times p$ matrix. Suppose each of $\{1, \dots, m\}$, $\{1, \dots, n\}$, and $\{1, \dots, p\}$ is partitioned in two index sets:

- $\{1, \dots, m\} = R \cup S$, where $R = \{1, \dots, d\}$ and $S = \{d+1, \dots, m\}$;
- $\{1, \dots, n\} = T \cup U$, where $T = \{1, \dots, e\}$ and $U = \{e+1, \dots, n\}$;
- $\{1, \dots, p\} = V \cup W$, where $V = \{1, \dots, f\}$ and $W = \{f+1, \dots, p\}$.

So we can write each of A and B in block form as follows:

$$A = \begin{bmatrix} A_{R,T} & A_{R,U} \\ A_{S,T} & A_{S,U} \end{bmatrix}, \quad B = \begin{bmatrix} B_{T,V} & B_{T,W} \\ B_{U,V} & B_{U,W} \end{bmatrix}.$$

Each of A and B can be viewed as a "2×2" matrix of "blocks", where the number of columns in an "A-block" is the same as the number of rows in a corresponding "B-block" (e.g., $A_{R,T}$ and $B_{T,W}$). We can thus use the block-wise generalization of the matrix multiplication rule in (3) to obtain

$$AB = \begin{bmatrix} A_{R,T}B_{T,V} + A_{R,U}B_{U,V} & A_{R,T}B_{T,W} + A_{R,U}B_{U,W} \\ A_{S,T}B_{T,V} + A_{S,U}B_{U,V} & A_{S,T}B_{T,W} + A_{S,U}B_{U,W} \end{bmatrix}.$$

The result, an $m \times p$ matrix, is given as a "2×2" matrix of "blocks". ∎

Example: two blocks. Suppose, in the previous example, that $S = W = \emptyset$, so $R = \{1, \dots, m\}$ and $V = \{1, \dots, p\}$. So each of A and B is partitioned into two blocks:

$$A = \begin{bmatrix} A_{R,T} & A_{R,U} \end{bmatrix}, \quad B = \begin{bmatrix} B_{T,V} \\ B_{U,V} \end{bmatrix}.$$

Then AB just has a single block:

$$AB = A_{R,T} B_{T,V} + A_{R,U} B_{U,V}.$$

∎

Example: sum of outer products. Again, let A be an $m \times n$ matrix, and let B be an $n \times p$ matrix. Now let the columns of A be $\mathbf{a}_1, \dots, \mathbf{a}_n$, and let the rows of B be $\mathbf{b}_1^\intercal, \dots, \mathbf{b}_n^\intercal$:

$$A = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ \downarrow & & \downarrow \end{bmatrix}, \quad B = \begin{bmatrix} \longleftarrow & \mathbf{b}_1^\intercal & \longrightarrow \\ & \vdots & \\ \longleftarrow & \mathbf{b}_n^\intercal & \longrightarrow \end{bmatrix}.$$

Here, we treat A as a single "row" of $n \times 1$ "blocks", and we treat B as a single "column" of $1 \times p$ "blocks". Using the block-wise generalization of the matrix multiplication rule in (3), we write AB as a sum of n matrices:

$$AB = \mathbf{a}_1 \mathbf{b}_1^\intercal + \cdots + \mathbf{a}_n \mathbf{b}_n^\intercal.$$

Each of the terms $\mathbf{a}_i \mathbf{b}_i^\intercal$ in the summation is an $m \times p$ matrix, obtained from multiplying a column vector $\mathbf{a}_i$ by a row vector $\mathbf{b}_i^\intercal$. Such a "product" of two vectors is called an *outer product*. If $\mathbf{x}$ is an m-vector and $\mathbf{y}$ is a p-vector, then we can think of the outer product $\mathbf{x}\mathbf{y}^\intercal$ in two ways:

- a list of p scalings of the vector $\mathbf{x}$, one per entry of $\mathbf{y} = (y_1, \dots, y_p)$,

$$\mathbf{x}\mathbf{y}^\intercal = \begin{bmatrix} \uparrow & & \uparrow \\ y_1\mathbf{x} & \cdots & y_p\mathbf{x} \\ \downarrow & & \downarrow \end{bmatrix};$$

- a stack of m scalings of $\mathbf{y}^\intercal$, one per entry of $\mathbf{x} = (x_1, \dots, x_m)$,

$$\mathbf{x}\mathbf{y}^\intercal = \begin{bmatrix} \longleftarrow & x_1\mathbf{y}^\intercal & \longrightarrow \\ & \vdots & \\ \longleftarrow & x_m\mathbf{y}^\intercal & \longrightarrow \end{bmatrix}.$$

∎