Matrix-Matrix Multiplication

One of the major themes of linear algebra is to build more advanced technology by adapting existing technology is special ways. This is the main message of the fun story of the mathematician and her coffee pot. Based on this strategy, we define matrix-matrix multiplication using the work we did to develop the matrix-vector multiplication operation. Before we do this, let's explore some terminology we can use to described the various parts of each matrix-matrix product.

Let $A \in \mathbb{R}^{m\times p}$ and $X \in \mathbb{R}^{p\times n}$. The *output* of the matrix-matrix multiplication

$$A \cdot X = B \tag{5.1}$$

is the matrix $B \in \mathbb{R}^{m\times n}$, which is also called the *product* of A and X. The matrix-matrix multiplication operation is a map between vector spaces

$$\cdot : \mathbb{R}^{m\times p} \times \mathbb{R}^{p\times n} \longrightarrow \mathbb{R}^{m\times n}.$$

This operation has two inputs and one output. The *left argument* or the *left factor* of the matrix product (5.1) is the matrix $A \in \mathbb{R}^{m\times p}$ on the left side of the multiplication sign. On the other hand, the **right argument** or *right factor* of this product (5.1) is matrix $X \in \mathbb{R}^{p\times n}$ on the right side of the multiplication sign.

We say that *we multiply A on the right by X* if we start with a modeling matrix A in the left argument and place matrix X in the right argument to do algebraic work. To multiply of A on the right by X, the number of columns of the left matrix A must be equal to the number of rows of the right matrix X.

If the column dimension of A equals the row dimension of X, we say that *A is conformable for right multiplication by X*. In other words: "the inner dimensions must agree!" If the number of rows of X does not match the column dimension of A, we say that matrix A is **nonconformable** for matrix-matrix multiplication on the right by X.

EXAMPLE 5.3.1

Consider the following matrices:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \end{bmatrix}, \quad Y = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix}.$$

The column dimension of A and the row dimension of X both equal four. Thus, the inner dimensions agree. We conclude that matrices A and X are conformable to matrix-matrix multiplication. If we define $A \times X = B$, we define the dimensions of B using the "outer dimensions" of A and X. Specifically, the row dimension of B is equal to the row dimension of A while the column dimension of B is equal to that of X. In this case, we see that $B \in \mathbb{R}^{3\times 2}$.

Definition 5.1

Matrix-matrix multiplication via linear combination of columns

Let $A \in \mathbb{R}^{m \times p}$ and $X \in \mathbb{R}^{p \times n}$. If we multiply A on the right by X to form the $m \times n$ matrix $B = A \cdot X$, then

$$\text{Column}_k(B) = A \cdot \text{Column}_k(X),$$

for $k \in \{1, 2, ..., n\}$. In other words, the kth column of B is the matrix A multiplied on the right by the kth column of X. This operation is written using colon notation as

$$B(:, k) = A \cdot X(:, k).$$

The kth column of the product $A \cdot X = B$ is a linear combination of the columns of matrix A with scalar weights defined by the individual entries in the kth column of X. We can write the kth column of the product B as an $m \times 1$ column with

$$B(:, k) = \begin{bmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{mk} \end{bmatrix} = x_{1k} \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_{2k} \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_{pk} \begin{bmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{mp} \end{bmatrix}$$

Using this definition, we execute matrix-matrix multiplication one column at a time to build the output matrix B. Looking back at our work to develop the matrix-vector multiplication operation, we have multiple options for how to execute each matrix-column-vector product used in the more general matrix-matrix multiplication.

This definition 5.1 for matrix-matrix multiplication via linear combination of column vectors suggests a useful framework. We multiply a modeling matrix A on the right by a matrix X when we want to manipulate the columns of matrix A in some way. Through the rest of our work together, we learn many strategies for choosing the matrix X to accomplish the algebraic work of transforming the columns of A. The overarching theme is to apply a series of transformations that results in an equivalent system of equations (written in the form of one of our fundamental problems). If we do our work well, the equivalent system that we construct can be solved using simple algorithms implemented on computers and the solution we calculate is almost exactly the solution we desire.

EXAMPLE 5.3.2

Define a 4×3 matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

Suppose we wish to double column one of A and leave the other columns untouched. To this end, we multiply A on the right-hand side by an appropriately sized dilation matrix:

$$B = A \cdot D_1(2) = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Let's begin by calculating the first column of the product

$$B(:,1) = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} = 2 \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + 0 \cdot \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} + 0 \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix}$$

Next, let's calculate the second column of B, given as a linear combination of the columns of A with scaling coefficients coming from the second column of $D_1(2)$.

$$B(:,2) = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 0 \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + 1 \cdot \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} + 0 \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

Finally, we calculate the last column of B as

$$B(:,3) = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0 \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + 0 \cdot \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} + 1 \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

We have now constructed a column partition of B. We can create the entire matrix B by combining together our three matrix-vector products

$$B = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

We see that right multiplication by $D_1(2)$ scaled the first column of A and left all other columns untouched. We also see that $D_1(2)$ has to be a 3×3 matrix since the column dimension of A is equal to three.

Using the example above, we can extrapolate a more general pattern. Suppose we want to scale the kth column of $A \in \mathbb{R}^{m \times n}$ by the number c. Then we will multiply A on the right by the $n \times n$ dilation matrix $D_k(c)$.

EXAMPLE 5.3.3

Let's use matrix-matrix multiplication to permute the columns 2 and 3 of a matrix. To this end, let $A \in \mathbb{R}^{4\times 4}$. Let P_{23} be the transposition matrix generated by swapping the second and third column of the identity matrix. Then consider

$$B = A \cdot P_{23} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

In this case, let's use the column-partition version to find each column of B. To this end consider, let's begin by finding the first column of B

$$B(:,1) = AP_{23}(:,1) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix} + 0 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} + 0 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{43} \end{bmatrix} + 0 \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \\ a_{44} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix}.$$

Now, let's move onto the second column of B

$$B(:,2) = AP_{23}(:,2) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix} + 0 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} + 1 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{43} \end{bmatrix} + 0 \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \\ a_{44} \end{bmatrix} = \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{43} \end{bmatrix}.$$

To find the third column of our product, we calculate

$$B(:,3) = AP_{23}(:,3) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix} + 1 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} + 0 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{43} \end{bmatrix} + 0 \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \\ a_{44} \end{bmatrix} = \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix}.$$

Finally, the last column of our product is given by

$$B(:,3) = AP_{23}(:,4) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = 0 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix} + 0 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} + 0 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{43} \end{bmatrix} + 1 \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \\ a_{44} \end{bmatrix} = \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \\ a_{44} \end{bmatrix}.$$

Once again, we can extrapolate a more general pattern from the example above. Suppose we want to swap columns i and j in the matrix $A \in \mathbb{R}^{m\times n}$. Then we will multiply A on the right by the $n \times n$ transposition matrix P_{ij}.

The definition for matrix-matrix multiplication via linear combination of column vectors 5.1 provides a useful alternative perspective on calculating the dot product between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n\times 1}$. Specifically, let's consider the entry-by-entry definition of each column vector, given by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{and} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Then, we can consider the matrix-matrix product given by

$$\mathbf{y}^T\,\mathbf{x} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \mathbf{x}\cdot\mathbf{y}.$$

In other words, we can express each dot product as a matrix-matrix multiplication using the transpose. We reached this c we take linear combinations of the columns of the matrix $\mathbf{y}^T$ and scale those columns via the individual entries of the vector $\mathbf{x}$. This results in a matrix-matrix multiplication formula for the dot product given by

$$\mathbf{x}\cdot\mathbf{y} = \mathbf{y}^T\mathbf{x}$$

Sometimes we want to construct a matrix-matrix product by rows instead of by columns. We say that we multiply modeling matrix A on the left by the algebraic worker X when we put A in the right argument and hit it on the left with the matrix X. To multiply A on the left by X, the number of columns of the left matrix X must equal the number of rows of the right matrix A. If the row dimension of A equals the column dimension of X, we say that A *is conformable for left multiplication by* X. Again, the inner dimensions must agree. If the dimensions of A are NOT suitable to multiply on the left by X, we say that matrix A is *nonconformable for multiplication on the left* by X.

Definition 5.2

Matrix-matrix multiplication via linear combination of rows

Let $A \in \mathbb{R}^{p\times n}$ and $X \in \mathbb{R}^{m\times p}$. If we multiply A on the left by X to form the $m \times n$ matrix $B = X \cdot A$, then

$$\mathrm{Row}_i(B) = \mathrm{Row}_i(X) \cdot A,$$

for $i \in \{1, 2, ..., m\}$. In other words, the ith row of B is the matrix A multiplied on the left by the ith row of X. This operation is written using colon notation as

$$B(i, :) = X(i, :) \cdot A$$

The ith row of the product $B = X \cdot A$ is a linear combination of the rows of matrix A with scalar weights defined by the individual entries in the ith row of X. Using this definition, we execute matrix-matrix multiplication one row at a time to build the individual rows of the output matrix B. For the ith row of the product, we calculate

$$\begin{aligned} \begin{bmatrix} b_{i1} & b_{i2} & \cdots & b_{in} \end{bmatrix} &= x_{i1} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{bmatrix} \\ &+ x_{i2} \begin{bmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{bmatrix} \\ &\vdots \\ &+ x_{ip} \begin{bmatrix} a_{p1} & a_{p2} & \cdots & a_{pn} \end{bmatrix} \end{aligned}$$

In such a way we construct the row partitions version of B. We say that we multiply A on the left by X if A is the right argument and X is the left argument. When multiplying A on the left by a matrix, we usually want to manipulate the rows of matrix A in some way. In this case, we think of A as the modeling matrix we start with, and then we choose X to do some special algebraic work. In left multiplication, we use the row version of matrix multiplication to manipulate the rows of A appropriately.

EXAMPLE 5.3.4

Let $A \in \mathbb{R}^{4\times 3}$. Suppose we wish to use matrix-matrix multiplication to multiply row i by 2^{3-i}. To this end, consider the following product:

$$B = D\,A \qquad \text{where} \qquad D = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}$$

Let's begin by finding the first row of the product B:

$$B(1,:) = D(1,:)\,A = 4\begin{bmatrix} a_{11} & a_{12} & a_{13}\end{bmatrix} + 0\begin{bmatrix} a_{21} & a_{22} & a_{23}\end{bmatrix} + 0\begin{bmatrix} a_{31} & a_{32} & a_{33}\end{bmatrix} + 0\begin{bmatrix} a_{41} & a_{42} & a_{43}\end{bmatrix} = \begin{bmatrix} 4a_{11} & 4a_{12} & 4a_{13}\end{bmatrix}$$

In words we see that the first row of the product is 4 times the first row of the matrix A. Let's move onto the second row of our desired output.

$$B(2,:) = D(2,:)\,A = 0\begin{bmatrix} a_{11} & a_{12} & a_{13}\end{bmatrix} + 2\begin{bmatrix} a_{21} & a_{22} & a_{23}\end{bmatrix} + 0\begin{bmatrix} a_{31} & a_{32} & a_{33}\end{bmatrix} + 0\begin{bmatrix} a_{41} & a_{42} & a_{43}\end{bmatrix} = \begin{bmatrix} 2a_{21} & 2a_{22} & 2a_{23}\end{bmatrix}$$

Once again, the second row of the product is two times the second row of A. The third row of our output follows from a similar calculation:

$$B(3,:) = D(3,:)\,A = 0\begin{bmatrix} a_{11} & a_{12} & a_{13}\end{bmatrix} + 0\begin{bmatrix} a_{21} & a_{22} & a_{23}\end{bmatrix} + 1\begin{bmatrix} a_{31} & a_{32} & a_{33}\end{bmatrix} + 0\begin{bmatrix} a_{41} & a_{42} & a_{43}\end{bmatrix} = \begin{bmatrix} a_{31} & a_{32} & a_{33}\end{bmatrix}$$

We finish our work with the final row given by

$$B(4,:) = D(4,:)\,A = 0\begin{bmatrix} a_{11} & a_{12} & a_{13}\end{bmatrix} + 0\begin{bmatrix} a_{21} & a_{22} & a_{23}\end{bmatrix} + 0\begin{bmatrix} a_{31} & a_{32} & a_{33}\end{bmatrix} + \frac{1}{2}\begin{bmatrix} a_{41} & a_{42} & a_{43}\end{bmatrix} = \begin{bmatrix} \frac{1}{2}a_{41} & \frac{1}{2}a_{42} & \frac{1}{2}a_{43}\end{bmatrix}$$

There is something special about the pattern we see here. Once again, the final row of B is just the last row of A multiplied by 0.5. What is the pattern you observe? Can you generalize? Make a conjecture about what is true when multiplying a matrix A on the left by a diagonal matrix. What happens if we multiply A on the right by a diagonal matrix? For readers who enjoy going deeper, take a look at the example above. How many different ways can you prove to yourself that

$$D = D_1(4)\,D_2(2)\,D_3(1)\,D_4(0.5) = D_4(0.5)\,D_3(1)\,D_2(2)\,D_1(4).$$

EXAMPLE 5.3.5

Let $A \in \mathbb{R}^{4\times 4}$ be given as

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ -2 & 0 & 2 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

Let $S_{31}(2)$ be the shear matrix generated by taking the identity matrix $I_4 \in \mathbb{R}^{4\times 4}$ and changing the zero value of the entry in row 3, column 1 into the number 2. Now let's calculate the matrix-matrix produce

$$B = S_{31}(2)\,A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ -2 & 0 & 2 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

We progress row-by-row starting with the first one with

$$B(1,:) = 1\begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix} + 0\begin{bmatrix} -2 & 0 & 2 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}$$

Let's move onto the second row:

$$B(1,:) = 0\begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix} + 0\begin{bmatrix} -2 & 0 & 2 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix}$$

The third row of the product is by far the most interesting:

$$B(3,:) = 2\begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix} + 1\begin{bmatrix} -2 & 0 & 2 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -2 & 2 & 0 \end{bmatrix}$$

This result comes from adding two times row 1 to row 3 and putting that output back into row 3. The final row of the product is given by

$$B(4,:) = 0\begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix} + 0\begin{bmatrix} -2 & 0 & 2 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix}$$

Putting this all together, we see

$$B = S_{31}(2)\,A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ -2 & 0 & 2 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -2 & 2 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

Rows 1, 2, and 4 are unaffected by the product. Row 3 of B comes from adding 2 times row 1 of A to row 3. Just as in our last example, this work highlights a larger pattern. How can you generalize? Make a conjecture about what is true when multiplying a matrix A on the left by the shear matrix $S_{ik}(c)$. What happens if we multiply A on the right by the shear matrix $S_{ik}(c)$?

In our work to develop the definitions of matrix-matrix multiplication via linear combinations of column vectors 5.1 and row vectors 5.2, we construct the output by thinking about the product in terms of vector-valued information. In the case of column vectors, we construct the kth column of $B = AX$ by taking linear combination of the columns of A with scalar multiples coming from the individual entries of the kth column of X. Similarly, for $B = XA$, we build the ith row of the output by taking linear combinations of the rows of A with the scalar coefficients coming from the individual entries of the ith row of X. In both cases, we're thinking about vectorized data. Specifically, we're partitioning the modeling matrix A into vectors in order to execute vector-valued operations to produce our desired output.

One of the major themes of our work together is to develop multiple representations for every idea that we study. In the case of matrix-matrix multiplication, we can shift our gaze from a vector perspective towards a scalar perspective. Specifically, using the dot product operation, we can generate the individual entries of the output B by taking dot products between vectors stored in both A and X. Let's develop a definition that gives yet a third approach to thinking about matrix-matrix multiplication.

Definition 5.3

Matrix-matrix multiplication via dot products

Let $A \in \mathbb{R}^{m\times p}$ and $X \in \mathbb{R}^{p\times n}$. The product $B = A \cdot X$ is the $m \times n$ matrix whose value in the ith row and kth entry is given by

$$b_{ik} = \text{Row}_i(A) \cdot \text{Column}_k(X) = A(i, :) \cdot X(:, k).$$

for $i \in \{1, 2, ..., m\}$ and $k \in \{1, 2, ..., n\}$.

Notice that each entry in B is given by the dot product

$$b_{ij} = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{ip} \end{bmatrix} \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{pk} \end{bmatrix} = a_{i1}x_{1k} + a_{i2}x_{2k} + \cdots + a_{ip}x_{pk} = \sum_{j=1}^{p} a_{ij}x_{jk}.$$

The (i, k)th entry of B is an inner product between the ith row of A (viewed as a column vector) and the kth column of X. The entry-by-entry definition of the matrix product is very efficient for calculating the individual entries of $A \cdot X$ when working on small problems by hand. This is the method of choice for "back of the envelop" calculations required on exams with no calculators. This method does tend to obscure important structural patterns since it focuses on scalar-valued entries rather than vector-valued data.

EXAMPLE 5.3.6

Let's define the matrices

$$X = S_{21}(-3) = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} 2 & 1 & 2 \\ 6 & 2 & 4 \end{bmatrix}.$$

Now let's use dot products to find the output

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 \\ 6 & 2 & 4 \end{bmatrix} = XA.$$

To do this calculation, we generate B entry by entry using the dot product. Using the our definition 5.3, we see:

$$b_{11} = X(1,:) \cdot A(:,1) = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 6 \end{bmatrix} = (1)(2) + (0)(6) = 2,$$

$$b_{12} = X(1,:) \cdot A(:,2) = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = (1)(1) + (0)(2) = 1,$$

$$b_{13} = X(1,:) \cdot A(:,3) = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 4 \end{bmatrix} = (1)(2) + (0)(4) = 1,$$

$$b_{21} = X(2,:) \cdot A(:,1) = \begin{bmatrix} -3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 6 \end{bmatrix} = (-3)(2) + (1)(6) = 0,$$

$$b_{22} = X(2,:) \cdot A(:,2) = \begin{bmatrix} -3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = (-3)(1) + (1)(2) = -1,$$

$$b_{23} = X(2,:) \cdot A(:,3) = \begin{bmatrix} -3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 4 \end{bmatrix} = (-3)(2) + (1)(4) = -2.$$

Combining these size individual scalar-valued calculations together, we form the matrix B with

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

A fun extension to this exercise is to redo these calculations using the other two methods for doing matrix-matrix multiplication. Specifically, confirm this example using linear combinations of both the columns and rows. What patterns do you notice? Of course the output of each method should be identical. However, does that mean that the techniques themselves are the same? Expand on this idea.

EXAMPLE 5.3.7

Let's confirm that the equation

$$AX = \begin{bmatrix} 3 & 0 & -2 \\ 2 & -1 & 2 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 3 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 9 & 1 & 1 \\ 5 & 7 & -2 \end{bmatrix} = B$$

Using definition 5.3 of matrix-matrix multiplication via dot products, we have:

$$b_{11} = A(1,:) \cdot X(:,1) = \begin{bmatrix} 3 & 0 & -2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = (3)(2) + (0)(1) + (-2)(3) = 0,$$

$$b_{21} = A(2,:) \cdot X(:,1) = \begin{bmatrix} 2 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = (2)(2) + (-1)(1) + (2)(3) = 9,$$

$$b_{31} = A(3,:) \cdot X(:,1) = \begin{bmatrix} 0 & 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = (0)(2) + (2)(1) + (1)(3) = 5,$$

$$b_{12} = A(1,:) \cdot X(:,2) = \begin{bmatrix} 3 & 0 & -2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = (3)(1) + (0)(3) + (-2)(1) = 1,$$

$$b_{22} = A(2,:) \cdot X(:,2) = \begin{bmatrix} 2 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = (2)(1) + (-1)(3) + (2)(1) = 1,$$

$$b_{32} = A(3,:) \cdot X(:,2) = \begin{bmatrix} 0 & 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = (0)(1) + (2)(3) + (1)(1) = 7,$$

$$b_{13} = A(1,:) \cdot X(:,3) = \begin{bmatrix} 3 & 0 & -2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = (3)(0) + (0)(-1) + (-2)(0) = 0,$$

$$b_{23} = A(2,:) \cdot X(:,3) = \begin{bmatrix} 2 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = (3)(0) + (-1)(-1) + (2)(0) = 1,$$

$$b_{33} = A(3,:) \cdot X(:,3) = \begin{bmatrix} 0 & 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = (0)(0) + (2)(-1) + (1)(0) = -2.$$

Just as before, go back and redo this example using the other two definitions of matrix-matrix multiplication via linear combinations of the column vectors and row vectors. What do you notice? How is the scalar version using dot products different from the column-vector or row-vector versions?

We are now ready to explore our fourth and final definition for computing matrix-matrix products. In this case, we use a sum of outer product operations between the column vectors of the left factor and the row vectors of the right vector. As we do so, we form the output via a sum of rank one updates.

Definition 5.4

Matrix-matrix multiplication via outer products

Let $A \in \mathbb{R}^{m\times p}$ and $X \in \mathbb{R}^{p\times n}$. The product $B = A \cdot X$ is the $m \times n$ matrix whose value is given by

$$B = A(:,1)X(1,:) + A(:,2)X(2,:) + \cdots + A(:,p)X(p,:)$$

Notice that we are building a version of matrix-matrix multiplication that relies on a sequence of outer products.

Notice that there are exactly p summands that form the output matrix

$$B = \sum_{k=1}^{p} A(:,k)X(k,:)$$

The kth summand in this sequence is an $m \times n$ matrix which results from taking the outer product between the kth column of A and the kth row of X. In other words, we form the output matrix by summing a sequence of rank-1 matrices.

EXAMPLE 5.3.8

Let's return to the our example 5.3.6 where we calculate

$$XA = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 \\ 6 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 2 \\ 0 & -1 & -2 \end{bmatrix} = B.$$

In this case, let's redo our work using the outer product definition 5.4 for matrix-matrix multiplication. To this end, we notice

$$\begin{aligned} XA &= X(:,1)A(1,:) + X(:,2)A(2,:) \\ &= \begin{bmatrix} 1 \\ -3 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 6 & 2 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 1 & 2 \\ -6 & -3 & -6 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 6 & 2 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 1 & 2 \\ 0 & -1 & -2 \end{bmatrix} \end{aligned}$$

In this case, we achieve the same output using a different method. The outer product version of matrix-matrix multiplication focuses on building the output matrix using matrix-sized pieces of data. Each individual summand comes from a rank one outer product operation.

EXAMPLE 5.3.9

To cement our understanding, let's redo example 5.3.7 using the outer product version of matrix-matrix multiplication. Recall that in this example, we have

$$AX = \begin{bmatrix} 3 & 0 & -2 \\ 2 & -1 & 2 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 3 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 9 & 1 & 1 \\ 5 & 7 & -2 \end{bmatrix} = B.$$

Using the outer product definition 5.4, we notice

$$\begin{aligned} AX &= A(:,1)X(1,:) + A(:,2)X(2,:) + A(:,3)X(3,:) \\ &= \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 3 & 0 \\ 4 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ -1 & -3 & 1 \\ 2 & 6 & -2 \end{bmatrix} + \begin{bmatrix} -6 & -2 & 0 \\ 6 & 2 & 0 \\ 3 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 9 & 1 & 1 \\ 5 & 7 & -2 \end{bmatrix} \end{aligned}$$

For the single operation of matrix-matrix multiplication, we have four different ways to calculate the desired output. All the techniques lead to identical entry-by-entry output values but the way we do the calculations differ. By studying multiple ways to get the same result, we increase our power to engage in flexible thinking.

As we continue to dive deeper into linear algebra, we use each of the four different ways of thinking about matrix-matrix multiplication strategically depending on the context and our desires. One of the major themes of solving any of the fundamental problems in linear algebra is to use matrices to do algebraic work. In other words, we start our explorations by generating a modeling matrix A from some real-world context that matters in our life. We spend countless hours transforming that context into the form of one of the fundamental problems. It is often the case, however, that the way we state the problem comes from the way we generate the model and leads to very complex algebraic equations that are impossible to solve by hand.

Using the techniques of applied linear algebra, we leverage modern-day computational power and manipulate our modeling problem using matrix-matrix multiplication with a very special set of matrices. Under the correct conditions and with a lot of thought, we translate our original problem into some equivalent problem whose solution is quite easy to calculate on a computer. The resulting solution then gives us insights into our larger problem.

The entire approach to problem solving depends heavily on leveraging the properties of matrix-matrix multiplication. In this section, we discover the theoretical definitions of this operation. The next challenge is to learn how to implement these definitions as computer code. The state-of-the-art is to optimize such code to be as fast and accurate as possible to enable processing of very large matrices. That is a challenge you might take up in your future career if you are so inclined.

We'll end this section by exploring the algebraic properties of matrix-matrix multiplication. For each property, we might generate at least four separate proofs, one for each definition of matrix-matrix multiplication from this section. In this manuscript, we'll provide four proofs of four separate properties, each proof relying on a unique version of matrix-matrix multiplication. We challenge the reader to create equivalent proofs for themselves.

Theorem 1

Algebraic Properties of Matrix-Matrix Multiplication

Let A, B, and C be matrices of the appropriate sizes so that all of the following operations can be performed. Let $\alpha \in \mathbb{R}$ be any scalar. Then, all of the following are algebraic properties of matrix-matrix multiplication:

i. Associativity: $\Big(A\,B\Big)\,C = A\,\Big(B\,C\Big)$
ii. Left distributivity: $A\,(B \pm C) = A\,B \pm A\,C$
iii. Right distributivity: $(A \pm B)\,C = A\,C \pm B\,C$
iv. Identity Matrix: $A\,I = A = I\,A$
v. Zero Matrix: $A\,0 = 0 = 0\,A$
vi. Transpose: $(A\,B)^T = B^T\,A^T$
vii. Scalar Multiplication: $(\alpha A)\,B = A\,(\alpha B) = \alpha(A\,B)$

Lesson 11: Matrix-Matrix Multiplication Problem Set

1. Let $A \in \mathbb{R}^{4\times 4}$. Execute each of the following operations by multiplying A on the right by the appropriate matrix X. In each case, explicitly write the matrix X and show all steps you used to calculate the matrix vector product $A \cdot X$.

   A. Double column 1

   B. Interchanges columns 1 and 4

   C. Add 2 times column 2 to column 3

   D. Delete column 4 (so that the column dimension is reduced by 1)

2. Let $X \in \mathbb{R}^{4\times 4}$. Execute each of the following operations by multiplying X on the left by the appropriate matrix A. In each case, explicitly write the matrix A and show all steps you used to calculate the matrix vector product $A \cdot X$.

   A. Halve row 3.

   B. Add row 2 to row 4.

   C. Subtract row 2 from each of the other rows

   D. Subtract row 1 from each of the other rows.

   E. Add row 3 to row 1 and also Add row 1 to row 3.

   F. Swap rows 1 and 2 and rows 3 and 4.

   G. Delete rows 1 and 3 (so that the row dimension is reduced by 2).

3. Begin with a modeling matrix $A \in \mathbb{R}^{6\times 5}$. Use matrix multiplication to permute the columns of matrix A using any permutation you can imagine. For example, you might send column 1 to column 3, column 3 to column 4, column 4 to column 2, column 2 to column 5 and column 5 to column 1. To visualize this permutation, we can use Cauchy's two-line notation for a permutation given as:

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

   In this notation, the first line represents the starting indices for each column and the second line represents where the columns end up after the product. Accomplish this permutation with a single matrix-matrix product. Also accomplish the same output with a sequence of matrix multiplications using only transposition matrices. What pattern(s) do you notice?

4. Begin with a modeling matrix $A \in \mathbb{R}^{6\times 5}$. Use matrix multiplication to permute the rows of matrix A using any permutation you can imagine. Suppose you send row 1 to row 2, row 2 to row 3, and row 3 back to row 1. Suppose also you send row 4 to row 6, row 6 to row 5, and row 5 back to row 4. To visualize this permutation, we can use Cauchy's two-line notation for a permutation given as:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 1 & 6 & 4 & 5 \end{pmatrix}.$$

   In this notation, the first line represents the starting indices for each row and the second line represents where the rows end up after the product. Accomplish this permutation with a single matrix-matrix product. Also accomplish the same output with a sequence of matrix multiplications using only transposition matrices. What pattern(s) do you notice?

5. In this section, we saw the matrix-matrix multiplication formula for the dot product given by $\mathbf{x} \cdot \mathbf{y} = \mathbf{y}^T \mathbf{x}$. In this exercise, we generalize this formula. Specifically, let $A \in \mathbb{R}^{n\times n}$ and suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n\times 1}$. For each of the formula found below, create as many different proofs as you can. Also, translate each of these formulas into simple, intuitive (nontechnical) language. Why are these significant? When might you use this information?

   A. $\mathbf{y} \cdot \mathbf{x} = \mathbf{x}^T \mathbf{y}$

   B. $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$

   C. $\mathbf{y} \cdot \mathbf{x} = \mathbf{y}^T \mathbf{x}$

   D. $(A\,\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A^T \mathbf{y})$

   E. $\mathbf{x} \cdot (A\,\mathbf{y}) = (A^T \mathbf{x}) \cdot \mathbf{y}$

6. Take a look at our Theorem statement for the Algebraic Properties of Matrix-Matrix Multiplication. For each item i. - vii. in the theorem statement, come up with as many unique ways to prove that item to be true as you can. For example, you might use the four different definitions of matrix-matrix multiplication to prove each individual point. Or, if that seems too daunting, alternate between which definition you use to prove each item. The point here is to show yourself that you can use the various versions of matrix-matrix multiplication to substantiate the algebraic relationships stated in this theorem. Each approach depends on different algebraic facts but the end results so that you establish the desired results using unique approaches.