I.2 Matrix-Matrix Multiplication AB

Inner products (*rows times columns*) produce each of the numbers in $AB = C$:

$$\begin{array}{c}\textbf{row 2 of } \boldsymbol{A}\\ \textbf{column 3 of } \boldsymbol{B}\\ \textbf{give } \boldsymbol{c_{23}} \textbf{ in } \boldsymbol{C}\end{array} \quad \begin{bmatrix} \cdot & \cdot & \cdot \\ a_{21} & a_{22} & a_{23} \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot & b_{13} \\ \cdot & \cdot & b_{23} \\ \cdot & \cdot & b_{33} \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \boldsymbol{c_{23}} \\ \cdot & \cdot & \cdot \end{bmatrix} \tag{1}$$

That dot product $c_{23} = (\text{row 2 of } A) \cdot (\text{column 3 of } B)$ is a sum of a's times b's :

$$\boldsymbol{c_{23}} = a_{\mathbf{21}}\, b_{\mathbf{13}} + a_{\mathbf{22}}\, b_{\mathbf{23}} + a_{\mathbf{23}}\, b_{\mathbf{33}} = \sum_{k=1}^{3} \boldsymbol{a_{2k}\, b_{k3}} \quad \text{and} \quad \boldsymbol{c_{ij}} = \sum_{k=1}^{n} \boldsymbol{a_{ik}\, b_{kj}}. \tag{2}$$

This is how we usually compute each number in $AB = C$. But there is another way.

The other way to multiply AB is **columns of $\boldsymbol{A}$ times rows of $\boldsymbol{B}$**. We need to see this ! I start with numbers to make two key points : *one column* $\boldsymbol{u}$ times *one row* $\boldsymbol{v}^{\mathrm{T}}$ produces a *matrix*. Concentrate first on that piece of AB. This matrix $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ is especially simple :

$$\textbf{“Outer product”} \quad \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 8 & 12 \\ 6 & 8 & 12 \\ 3 & 4 & 6 \end{bmatrix} = \textbf{“rank one matrix”}$$

An m by 1 matrix (a column $\boldsymbol{u}$) times a 1 by p matrix (a row $\boldsymbol{v}^{\mathrm{T}}$) gives an m by p matrix. Notice what is special about the rank one matrix $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$:

$$\text{All columns of } \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \text{ are multiples of } \boldsymbol{u} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \quad \text{All rows are multiples of } \boldsymbol{v}^{\mathrm{T}} = \begin{bmatrix} 3 & 4 & 6 \end{bmatrix}$$

The column space of $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ is one-dimensional : *the line in the direction of* $\boldsymbol{u}$. The dimension of the column space (the number of independent columns) is the **rank of the matrix**—a key number. **All nonzero matrices $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ have rank one**. They are the perfect building blocks for every matrix.

Notice also : **The row space of $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ is the line through $\boldsymbol{v}$**. By definition, the row space of any matrix A is the column space $\mathbf{C}(A^{\mathrm{T}})$ of its transpose A^{T}. That way we stay with column vectors. In the example, we transpose $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ (**exchange rows with columns**) to get the matrix $\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}$:

$$(\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})^{\mathrm{T}} = \begin{bmatrix} \mathbf{6} & 8 & 12 \\ \mathbf{6} & 8 & 12 \\ \mathbf{3} & 4 & 6 \end{bmatrix}^{\mathbf{T}} = \begin{bmatrix} \mathbf{6} & \mathbf{6} & \mathbf{3} \\ 8 & 8 & 4 \\ 12 & 12 & 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \end{bmatrix} = \boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}.$$

We are seeing the clearest possible example of the first great theorem in linear algebra:

Row rank = Column rank	$\boldsymbol{r}$ **independent columns** $\Leftrightarrow$ $\boldsymbol{r}$ **independent rows**

A nonzero matrix $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ has one independent column and one independent row. All columns are multiples of $\boldsymbol{u}$ and all rows are multiples of $\boldsymbol{v}^{\mathrm{T}}$. The rank is $r = 1$ for this matrix.

$\boldsymbol{AB}$ = Sum of Rank One Matrices

We turn to the full product AB, using columns of A times rows of B. Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ be the n columns of A. Then B must have n rows $\boldsymbol{b}_1^*, \boldsymbol{b}_2^*, \ldots, \boldsymbol{b}_n^*$. The matrix A can multiply the matrix B. **Their product $\boldsymbol{AB}$ is the sum of columns $\boldsymbol{a}_k$ times rows $\boldsymbol{b}_k^*$:**

Column-row multiplication of matrices

$$AB = \begin{bmatrix} | & & | \\ \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \text{---} \; \boldsymbol{b}_1^* \; \text{---} \\ \vdots \\ \text{---} \; \boldsymbol{b}_n^* \; \text{---} \end{bmatrix} = \boldsymbol{a}_1 \boldsymbol{b}_1^* + \boldsymbol{a}_2 \boldsymbol{b}_2^* + \cdots + \boldsymbol{a}_n \boldsymbol{b}_n^*. \quad \textbf{sum of rank 1 matrices} \tag{3}$$

Here is a 2 by 2 example to show the $n = 2$ pieces (column times row) and their sum AB:

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{3} \end{bmatrix} \begin{bmatrix} \mathbf{2} & \mathbf{4} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{5} \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 12 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 17 \end{bmatrix} \tag{4}$$

We can count the multiplications of number times number. Four multiplications to get $2, 4, 6, 12$. Four more to get $0, 0, 0, 5$. A total of $2^3 = 8$ multiplications. Always there are $\boldsymbol{n^3}$ multiplications when A and B are n by n. And $\boldsymbol{mnp}$ multiplications when $AB =$ $(m$ by $n)$ times $(n$ by $p)$: n rank one matrices, each of those matrices is m by p.

The count is the same for the usual inner product way. Row of A times column of B needs n multiplications. We do this for every number in AB: mp dot products when AB is m by p. The total count is again $\boldsymbol{mnp}$ when we multiply $(m$ by $n)$ times $(n$ by $p)$.

rows times columns	$\boldsymbol{mp}$ **inner products,**	$\boldsymbol{n}$ **multiplications each**	$\boldsymbol{mnp}$
columns times rows	$\boldsymbol{n}$ **outer products,**	$\boldsymbol{mp}$ **multiplications each**	$\boldsymbol{mnp}$

When you look closely, they are exactly the same multiplications $a_{ik}\, b_{kj}$ in different orders. Here is the algebra proof that each number c_{ij} in $C = AB$ is the same by outer products in (3) as by inner products in (2):

$$\text{The } i, j \text{ entry of } \boldsymbol{a}_k \boldsymbol{b}_k^* \text{ is } \boldsymbol{a_{ik} b_{kj}}. \text{ Add to find } \boldsymbol{c_{ij}} = \sum_{k=1}^{n} a_{ik}\, b_{kj} = \textbf{row } \boldsymbol{i} \cdot \textbf{ column } \boldsymbol{j}.$$

Insight from Column times Row

Why is the outer product approach essential in data science? The short answer is: *We are looking for the important part of a matrix A.* We don't usually want the biggest number in A (though that could be important). What we want more is the largest piece of A. **And those pieces are rank one matrices $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$.** A dominant theme in applied linear algebra is:

Factor $\boldsymbol{A}$ into $\boldsymbol{CR}$ and look at the pieces $\boldsymbol{c}_k\boldsymbol{r}_k^*$ of $\boldsymbol{A} = \boldsymbol{CR}$.

Factoring A into CR is the reverse of multiplying $CR = A$. Factoring takes longer, especially if the pieces involve *eigenvalues* or *singular values*. But those numbers have inside information about the matrix A. That information is not visible until you factor.

Here are five important factorizations, with the standard choice of letters (usually A) for the original product matrix and then for its factors. This book will explain all five.

$$\boldsymbol{A} = \boldsymbol{LU} \qquad \boldsymbol{A} = \boldsymbol{QR} \qquad \boldsymbol{S} = \boldsymbol{Q\Lambda Q}^{\mathbf{T}} \qquad \boldsymbol{A} = \boldsymbol{X\Lambda X}^{-1} \qquad \boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathbf{T}}$$

At this point we simply list key words and properties for each of these factorizations.

1 $A = LU$ comes from **elimination**. Combinations of rows take A to U and U back to A. The matrix L is lower triangular and U is upper triangular as in equation (4).

2 $A = QR$ comes from **orthogonalizing** the columns $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$ as in "Gram-Schmidt". Q has orthonormal columns ($Q^{\mathrm{T}}Q = I$) and R is upper triangular.

3 $S = Q\Lambda Q^{\mathrm{T}}$ comes from the **eigenvalues** $\lambda_1, \ldots, \lambda_n$ of a symmetric matrix $S = S^{\mathrm{T}}$. Eigenvalues on the diagonal of Λ. **Orthonormal eigenvectors** in the columns of Q.

4 $A = X\Lambda X^{-1}$ is **diagonalization** when A is n by n with n independent eigenvectors. *Eigenvalues* of A on the diagonal of Λ. *Eigenvectors* of A in the columns of X.

5 $A = U\Sigma V^{\mathrm{T}}$ is the **Singular Value Decomposition** of any matrix A (square or not). **Singular values** $\sigma_1, \ldots, \sigma_r$ in Σ. Orthonormal **singular vectors** in U and V.

Let me pick out a favorite (number **3**) to illustrate the idea. This special factorization $Q\Lambda Q^{\mathrm{T}}$ starts with a symmetric matrix S. That matrix has orthogonal unit eigenvectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$. Those perpendicular eigenvectors (dot products $= 0$) go into the columns of Q.

S and Q are the kings and queens of linear algebra:

Symmetric matrix $\boldsymbol{S}$	$\boldsymbol{S}^{\mathbf{T}} = \boldsymbol{S}$	All $\boldsymbol{s}_{ij} = \boldsymbol{s}_{ji}$
Orthogonal matrix $\boldsymbol{Q}$	$\boldsymbol{Q}^{\mathbf{T}} = \boldsymbol{Q}^{-1}$	All $\boldsymbol{q}_i \cdot \boldsymbol{q}_j = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}$

The diagonal matrix Λ contains real eigenvalues λ_1 to λ_n. Every real symmetric matrix S has n orthonormal eigenvectors $\boldsymbol{q}_1$ to $\boldsymbol{q}_n$. When multiplied by S, the eigenvectors keep the same direction. They are just rescaled by the number λ :

$$\boxed{\textbf{Eigenvector } q \textbf{ and eigenvalue } \boldsymbol{\lambda} \qquad S\boldsymbol{q} = \lambda\boldsymbol{q}} \tag{5}$$

Finding λ and $\boldsymbol{q}$ is not easy for a big matrix. But n pairs always exist when S is symmetric. Our purpose here is to see how $\boldsymbol{SQ} = \boldsymbol{Q\Lambda}$ comes column by column from $S\boldsymbol{q} = \lambda\boldsymbol{q}$:

$$\boldsymbol{SQ} = S\begin{bmatrix} \boldsymbol{q}_1 & \dots & \boldsymbol{q}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{q}_1 & \dots & \lambda_n\boldsymbol{q}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \dots & \boldsymbol{q}_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \boldsymbol{Q\Lambda} \tag{6}$$

Multiply $SQ = Q\Lambda$ by $Q^{-1} = Q^{\mathrm{T}}$ to get $\boldsymbol{S = Q\Lambda Q^{\mathrm{T}}}$ = a symmetric matrix. Each eigenvalue λ_k and each eigenvector $\boldsymbol{q}_k$ contribute a rank one piece $\lambda_k\boldsymbol{q}_k\boldsymbol{q}_k^{\mathrm{T}}$ to S.

$$\textbf{Rank one pieces} \qquad S = (Q\Lambda)Q^{\mathrm{T}} = (\lambda_1\boldsymbol{q}_1)\boldsymbol{q}_1^{\mathrm{T}} + (\lambda_2\boldsymbol{q}_2)\boldsymbol{q}_2^{\mathrm{T}} + \cdots + (\lambda_n\boldsymbol{q}_n)\boldsymbol{q}_n^{\mathrm{T}} \tag{7}$$

$$\textbf{All symmetric} \qquad \text{The transpose of } \boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}} \text{ is } \boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}} \tag{8}$$

Please notice that the columns of $Q\Lambda$ are $\lambda_1\boldsymbol{q}_1$ to $\lambda_n\boldsymbol{q}_n$. When you multiply a matrix on the right by the diagonal matrix Λ, you multiply its *columns* by the λ's.

We close with a comment on the proof of this **Spectral Theorem** $\boldsymbol{S = Q\Lambda Q^{\mathrm{T}}}$: Every symmetric S has n real eigenvalues and n orthonormal eigenvectors. Section 1.6 will construct the eigenvalues as the roots of the nth degree polynomial $P_n(\lambda)$ = determinant of $S - \lambda I$. They are real numbers when $S = S^{\mathrm{T}}$. The delicate part of the proof comes when an eigenvalue λ_i is *repeated*— it is a double root or an Mth root from a factor $(\lambda - \lambda_j)^M$. In this case we need to produce M independent eigenvectors. The rank of $S - \lambda_j I$ must be $n - M$. This is true when $S = S^{\mathrm{T}}$. But it requires a proof.

Similarly the Singular Value Decomposition $A = U\Sigma V^{\mathrm{T}}$ requires extra patience when a singular value σ is repeated M times in the diagonal matrix Σ. Again there are M pairs of singular vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ with $A\boldsymbol{v} = \sigma\boldsymbol{u}$. Again this true statement requires proof.

Notation for rows We introduced the symbols $\boldsymbol{b}_1^*, \dots, \boldsymbol{b}_n^*$ for the rows of the second matrix in AB. You might have expected $\boldsymbol{b}_1^{\mathrm{T}}, \dots, \boldsymbol{b}_n^{\mathrm{T}}$ and that was our original choice. But this notation is not entirely clear—it seems to mean the transposes of the columns of B. Since that right hand factor could be U or R or Q^{T} or X^{-1} or V^{T}, it is safer to say definitely : *we want the rows of that matrix*.

Problem Set I.2

1 Suppose $A\boldsymbol{x} = \boldsymbol{0}$ and $A\boldsymbol{y} = \boldsymbol{0}$ (where $\boldsymbol{x}$ and $\boldsymbol{y}$ and $\boldsymbol{0}$ are vectors). Put those two statements together into one matrix equation $AB = C$. What are those matrices B and C ? If the matrix A is m by n, what are the shapes of B and C ?

2 Suppose $\boldsymbol{a}$ and $\boldsymbol{b}$ are column vectors with components $a_1, \ldots, a_m$ and $b_1, \ldots, b_p$. Can you multiply $\boldsymbol{a}$ times $\boldsymbol{b}^{\mathrm{T}}$ (yes or no)? What is the shape of the answer $\boldsymbol{ab}^{\mathrm{T}}$? What number is in row i, column j of $\boldsymbol{ab}^{\mathrm{T}}$? What can you say about $\boldsymbol{aa}^{\mathrm{T}}$?

3 (Extension of Problem **2**: Practice with subscripts) Instead of that one vector $\boldsymbol{a}$, suppose you have n vectors $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$ in the columns of A. Suppose you have n vectors $\boldsymbol{b}_1^{\mathrm{T}}, \ldots, \boldsymbol{b}_n^{\mathrm{T}}$ in the rows of B.

(a) Give a "sum of rank one" formula for the matrix-matrix product AB.

(b) Give a formula for the i, j entry of that matrix-matrix product AB. Use sigma notation to add the i, j entries of each matrix $\boldsymbol{a}_k\boldsymbol{b}_k^{\mathrm{T}}$, found in Problem 2.

4 Suppose B has only one column ($p = 1$). So each row of B just has one number. A has columns $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$ as usual. Write down the column times row formula for AB. In words, the m by 1 column vector AB is a combination of the _____.

5 Start with a matrix B. If we want to take combinations of its rows, we premultiply by A to get AB. If we want to take combinations of its columns, we postmultiply by C to get BC. For this question we will do both.

Row operations then column operations First AB then $(AB)C$

Column operations then row operations First BC then $A(BC)$

The **associative law** says that we get the same final result both ways.

Verify $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$ for $A = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$.

6 If A has columns $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ and $B = I$ is the identity matrix, what are the rank one matrices $\boldsymbol{a}_1\boldsymbol{b}_1^*$ and $\boldsymbol{a}_2\boldsymbol{b}_2^*$ and $\boldsymbol{a}_3\boldsymbol{b}_3^*$? They should add to $AI = A$.

7 *Fact*: The columns of AB are combinations of the columns of A. Then the column *space* of AB is *contained in* the column space of A. Give an example of A and B for which AB has a smaller column space than A.

8 To compute $C = AB = (m \text{ by } n)\,(n \text{ by } p)$, what order of the same three commands leads to columns times rows (outer products)?

Rows times columns	**Columns times rows**
For $i = 1$ to m	For...
For $j = 1$ to p	For...
For $k = 1$ to n	For...
$C(i,j) = C(i,j) + A(i,k) * B(k,j)$	$C =$