

$$A = LU$$

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 7 \end{bmatrix}, \text{ 将 } A \text{ 作初等行变换.}$$

$$\begin{bmatrix} 2 & 3 \\ 4 & 7 \end{bmatrix} \xrightarrow{\text{第1行} \times (-2) + \text{第2行}} \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$$

$A \qquad \qquad U$

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

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

$\uparrow \qquad \qquad \uparrow$   
 $L \qquad \qquad U$

求  $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$  的逆, 即

如何将  $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$  变回  $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

即第1行  $\times 2$  加到第2行.  $\therefore \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\text{故 } \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$$

## I.2 Matrix-Matrix Multiplication $AB$

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

$$\begin{array}{l} \text{row 2 of } A \\ \text{column 3 of } B \\ \text{give } c_{23} \text{ in } C \end{array} \quad \begin{bmatrix} \cdot & \cdot & \cdot \\ a_{21} & a_{22} & a_{23} \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot & b_{13} \\ \cdot & \cdot & b_{23} \\ \cdot & \cdot & b_{33} \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & c_{23} \\ \cdot & \cdot & \cdot \end{bmatrix} \quad (1)$$

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:

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

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  $A$  times rows of  $B$** . We need to see this! I start with numbers to make two key points: *one column  $u$  times one row  $v^T$  produces a matrix*. Concentrate first on that piece of  $AB$ . This matrix  $uv^T$  is especially simple:

$$\begin{array}{c} \text{"Outer} \\ \text{product"} \end{array} \quad uv^T = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 8 & 12 \\ 6 & 8 & 12 \\ 3 & 4 & 6 \end{bmatrix} = \begin{array}{c} \text{"rank one} \\ \text{matrix"} \end{array}$$

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

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

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

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

$$(uv^T)^T = \begin{bmatrix} 6 & 8 & 12 \\ 6 & 8 & 12 \\ 3 & 4 & 6 \end{bmatrix}^T = \begin{bmatrix} 6 & 6 & 3 \\ 8 & 8 & 4 \\ 12 & 12 & 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \end{bmatrix} = vu^T.$$

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

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

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

### $AB = \text{Sum of Rank One Matrices}$

We turn to the full product  $AB$ , using columns of  $A$  times rows of  $B$ . Let  $a_1, a_2, \dots, a_n$  be the  $n$  columns of  $A$ . Then  $B$  must have  $n$  rows  $b_1^*, b_2^*, \dots, b_n^*$ . The matrix  $A$  can multiply the matrix  $B$ . **Their product  $AB$  is the sum of columns  $a_k$  times rows  $b_k^*$ :**

#### Column-row multiplication of matrices

important  $\rightarrow$

★

columns  $\times$  rows  $\Rightarrow$

$$AB = \left[ \begin{array}{c|ccc|c} & & & & \\ \hline & a_1 & \dots & a_n & \\ \hline & & & & \end{array} \right] \left[ \begin{array}{c} \text{--- } b_1^* \text{ ---} \\ \vdots \\ \text{--- } b_n^* \text{ ---} \end{array} \right] = \underbrace{a_1 b_1^* + a_2 b_2^* + \dots + a_n b_n^*}_{\text{sum of rank 1 matrices}} \quad (3)$$

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

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 12 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 17 \end{bmatrix} \quad (4)$$

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  $n^3$  multiplications when  $A$  and  $B$  are  $n$  by  $n$ . And  $mnp$  multiplications when  $AB = (m \text{ by } n) \text{ times } (n \text{ by } p)$ :  $n$  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  $mnp$  when we multiply  $(m \text{ by } n)$  times  $(n \text{ by } p)$ .

rows times columns	<b><math>mp</math> inner products,</b>	<b><math>n</math> multiplications each</b>	<b><math>mnp</math></b>
columns times rows	<b><math>n</math> outer products,</b>	<b><math>mp</math> multiplications each</b>	<b><math>mnp</math></b>

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):

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

### 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  $uv^T$ .** A dominant theme in applied linear algebra is:

**Factor  $A$  into  $CR$  and look at the pieces  $c_k r_k^*$  of  $A = CR$ .**

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.

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

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  $a_1$  to  $a_n$  as in "Gram-Schmidt".  $Q$  has orthonormal columns ( $Q^T Q = I$ ) and  $R$  is upper triangular.
- 3  $S = Q\Lambda Q^T$  comes from the **eigenvalues**  $\lambda_1, \dots, \lambda_n$  of a symmetric matrix  $S = S^T$ . Eigenvalues on the diagonal of  $\Lambda$ . **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  $\Lambda$ . Eigenvectors of  $A$  in the columns of  $X$ .*
- 5  $A = U\Sigma V^T$  is the **Singular Value Decomposition** of any matrix  $A$  (square or not). **Singular values**  $\sigma_1, \dots, \sigma_r$  in  $\Sigma$ . Orthonormal **singular vectors** in  $U$  and  $V$ .

Let me pick out a favorite (number 3) to illustrate the idea. This special factorization  $Q\Lambda Q^T$  starts with a symmetric matrix  $S$ . That matrix has orthogonal unit eigenvectors  $q_1, \dots, q_n$ . Those perpendicular eigenvectors (dot products = 0) go into the columns of  $Q$ .  $S$  and  $Q$  are the kings and queens of linear algebra:

<b>Symmetric matrix <math>S</math></b>	$S^T = S$	All $s_{ij} = s_{ji}$
<b>Orthogonal matrix <math>Q</math></b>	$Q^T = Q^{-1}$	All $q_i \cdot q_j = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}$

对A矩阵分解的好处

The diagonal matrix  $\Lambda$  contains real eigenvalues  $\lambda_1$  to  $\lambda_n$ . Every real symmetric matrix  $S$  has  $n$  orthonormal eigenvectors  $\mathbf{q}_1$  to  $\mathbf{q}_n$ . When multiplied by  $S$ , the eigenvectors keep the same direction. They are just rescaled by the number  $\lambda$ :

<b>Eigenvector <math>\mathbf{q}</math> and eigenvalue <math>\lambda</math></b> $S\mathbf{q} = \lambda\mathbf{q}$
--

(5)

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

$$S\mathbf{Q} = S \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \mathbf{q}_1 & \cdots & \lambda_n \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \mathbf{Q}\Lambda \quad (6)$$

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

**Rank one pieces**      $S = (\mathbf{Q}\Lambda)\mathbf{Q}^T = (\lambda_1 \mathbf{q}_1) \mathbf{q}_1^T + (\lambda_2 \mathbf{q}_2) \mathbf{q}_2^T + \cdots + (\lambda_n \mathbf{q}_n) \mathbf{q}_n^T \quad (7)$

**All symmetric**     The transpose of  $\mathbf{q}_i \mathbf{q}_i^T$  is  $\mathbf{q}_i \mathbf{q}_i^T$  (8)

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

We close with a comment on the proof of this **Spectral Theorem**  $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^T$ : Every symmetric  $S$  has  $n$  real eigenvalues and  $n$  orthonormal eigenvectors. Section 1.6 will construct the eigenvalues as the roots of the  $n$ th degree polynomial  $P_n(\lambda) = \det(S - \lambda I)$ . They are real numbers when  $S = S^T$ . The delicate part of the proof comes when an eigenvalue  $\lambda_i$  is *repeated*—it is a double root or an  $M$ th 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  $S = S^T$ . But it requires a proof.

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

*Notation for rows* We introduced the symbols  $\mathbf{b}_1^*, \dots, \mathbf{b}_n^*$  for the rows of the second matrix in  $AB$ . You might have expected  $\mathbf{b}_1^T, \dots, \mathbf{b}_n^T$  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  $\mathbf{Q}^T$  or  $X^{-1}$  or  $V^T$ , it is safer to say definitely: *we want the rows of that matrix*.

1.  $A \begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$  the shape of  $B$  is  $m$  by  $2$   
 $\uparrow \quad \uparrow$   
 $B \quad C$  the shape of  $C$  is  $m$  by  $2$

2. yes.  $A$  times  $b^T$  I.2. Matrix-Matrix Multiplication  $AB$

13

will get a matrix.

## Problem Set I.2

and its shape is  $m$  by  $p$ . The number of

$ab^T[i][j]$  is  $a_i b_j$ .

$aa^T$  is a symmetric matrix, which means  $(aa^T)^T = aa^T$

1 Suppose  $Ax = 0$  and  $Ay = 0$  (where  $x$  and  $y$  and  $0$  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  $a$  and  $b$  are column vectors with components  $a_1, \dots, a_m$  and  $b_1, \dots, b_p$ . Can you multiply  $a$  times  $b^T$  (yes or no)? What is the shape of the answer  $ab^T$ ? What number is in row  $i$ , column  $j$  of  $ab^T$ ? What can you say about  $aa^T$ ?

3 (Extension of Problem 2: Practice with subscripts) Instead of that one vector  $a$ , suppose you have  $n$  vectors  $a_1$  to  $a_n$  in the columns of  $A$ . Suppose you have  $n$  vectors  $b_1^T, \dots, b_n^T$  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  $a_k b_k^T$ , found in Problem 2.

Suppose  $B$  has only one column ( $p = 1$ ). So each row of  $B$  just has one number.  $A$  has columns  $a_1$  to  $a_n$  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

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

6 If  $A$  has columns  $a_1, a_2, a_3$  and  $B = I$  is the identity matrix, what are the rank one matrices  $a_1 b_1^*$  and  $a_2 b_2^*$  and  $a_3 b_3^*$ ? 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**

For  $i = 1$  to  $m$

For  $j = 1$  to  $p$

For  $k = 1$  to  $n$

$C(i, j) = C(i, j) + A(i, k) * B(k, j)$

**Columns times rows**

For...

For...

For...

$C =$

$B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix}$

b.