

## Chapter 7

# The Singular Value Decomposition (SVD)

- 1 The SVD produces **orthonormal bases** of  $v$ 's and  $u$ 's for the four fundamental subspaces.
- 2 Using those bases,  $A$  becomes a diagonal matrix  $\Sigma$  and  $Av_i = \sigma_i u_i$  :  $\sigma_i =$  **singular value**.
- 3 The two-bases diagonalization  $A = U\Sigma V^T$  often has more information than  $A = X\Lambda X^{-1}$ .
- 4  $U\Sigma V^T$  separates  $A$  into rank-1 matrices  $\sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$ .  $\sigma_1 u_1 v_1^T$  is the largest!

### 7.1 Bases and Matrices in the SVD

The Singular Value Decomposition is a highlight of linear algebra.  $A$  is any  $m$  by  $n$  matrix, square or rectangular. Its rank is  $r$ . We will diagonalize this  $A$ , but not by  $X^{-1}AX$ . The eigenvectors in  $X$  have three big problems: They are usually not orthogonal, there are not always enough eigenvectors, and  $Ax = \lambda x$  requires  $A$  to be a square matrix. The **singular vectors** of  $A$  solve all those problems in a perfect way.

Let me describe what we want from the SVD : **the right bases for the four subspaces**. Then I will write about the steps to find those bases **in order of importance**.

The price we pay is to have **two sets of singular vectors**,  $u$ 's and  $v$ 's. The  $u$ 's are in  $\mathbf{R}^m$  and the  $v$ 's are in  $\mathbf{R}^n$ . They will be the columns of an  $m$  by  $m$  matrix  $U$  and an  $n$  by  $n$  matrix  $V$ . I will first describe the SVD in terms of those basis vectors. Then I can also describe the SVD in terms of the orthogonal matrices  $U$  and  $V$ .

(using vectors) The  $u$ 's and  $v$ 's give bases for the four fundamental subspaces :

$u_1, \dots, u_r$  is an orthonormal basis for the **column space**  
 $u_{r+1}, \dots, u_m$  is an orthonormal basis for the **left nullspace**  $N(A^T)$   
 $v_1, \dots, v_r$  is an orthonormal basis for the **row space**  
 $v_{r+1}, \dots, v_n$  is an orthonormal basis for the **nullspace**  $N(A)$ .

More than just orthogonality, these basis vectors diagonalize the matrix  $A$ :

$$\text{"A is diagonalized"} \quad Av_1 = \sigma_1 u_1 \quad Av_2 = \sigma_2 u_2 \quad \dots \quad Av_r = \sigma_r u_r \quad (1)$$

Those **singular values**  $\sigma_1$  to  $\sigma_r$  will be positive numbers:  $\sigma_i$  is the length of  $Av_i$ . The  $\sigma$ 's go into a diagonal matrix that is otherwise zero. That matrix is  $\Sigma$ .

(using matrices) Since the  $u$ 's are orthonormal, the matrix  $U$  with those  $r$  columns has  $U^T U = I$ . Since the  $v$ 's are orthonormal, the matrix  $V$  has  $V^T V = I$ . Then the equations  $Av_i = \sigma_i u_i$  tell us column by column that  $AV_r = U_r \Sigma_r$ :

$$\begin{matrix} (m \text{ by } n) & (n \text{ by } r) \\ (m \text{ by } r) & (r \text{ by } r) \end{matrix} \quad A \begin{bmatrix} v_1 & \dots & v_r \end{bmatrix} = \begin{bmatrix} u_1 & \dots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}. \quad (2)$$

This is the heart of the SVD, but there is more. Those  $v$ 's and  $u$ 's account for the row space and column space of  $A$ . We have  $n - r$  more  $v$ 's and  $m - r$  more  $u$ 's, from the nullspace  $N(A)$  and the left nullspace  $N(A^T)$ . They are automatically orthogonal to the first  $v$ 's and  $u$ 's (because the whole nullspaces are orthogonal). We now include all the  $v$ 's and  $u$ 's in  $V$  and  $U$ , so these matrices become *square*. **We still have**  $AV = U\Sigma$ .

$$\begin{matrix} (m \text{ by } n) & (n \text{ by } n) \\ (m \text{ by } m) & (m \text{ by } n) \end{matrix} \quad A \begin{bmatrix} v_1 & \dots & v_r & \dots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \dots & u_r & \dots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}. \quad (3)$$

The new  $\Sigma$  is  $m$  by  $n$ . It is just the  $r$  by  $r$  matrix in equation (2) with  $m - r$  extra zero rows and  $n - r$  new zero columns. The real change is in the shapes of  $U$  and  $V$ . Those are square orthogonal matrices. So  $AV = U\Sigma$  can become  $A = U\Sigma V^T$ . This is the **Singular Value Decomposition**. I can multiply columns  $u_i \sigma_i$  from  $U\Sigma$  by rows of  $V^T$ :

$$\text{SVD} \quad A = U\Sigma V^T = u_1 \sigma_1 v_1^T + \dots + u_r \sigma_r v_r^T. \quad (4)$$

Equation (2) was a "reduced SVD" with bases for the row space and column space. Equation (3) is the full SVD with nullspaces included. They both split up  $A$  into the same  $r$  matrices  $u_i \sigma_i v_i^T$  of rank one: column times row.

We will see that each  $\sigma_i^2$  is an eigenvalue of  $A^T A$  and also  $AA^T$ . When we put the singular values in descending order,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ , the splitting in equation (4) gives the  $r$  rank-one pieces of  $A$  in **order of importance**. This is crucial.

**Example 1** When is  $\Lambda = U\Sigma V^T$  (singular values) the same as  $X\Lambda X^{-1}$  (eigenvalues)?

**Solution**  $A$  needs orthonormal eigenvectors to allow  $X = U = V$ .  $A$  also needs eigenvalues  $\lambda \geq 0$  if  $\Lambda = \Sigma$ . So  $A$  must be a **positive semidefinite (or definite) symmetric matrix**. Only then will  $A = X\Lambda X^{-1}$  which is also  $Q\Lambda Q^T$  coincide with  $A = U\Sigma V^T$ .

**Example 2** If  $A = \mathbf{x}\mathbf{y}^T$  (rank 1) with unit vectors  $\mathbf{x}$  and  $\mathbf{y}$ , what is the SVD of  $A$ ?

**Solution** The reduced SVD in (2) is exactly  $\mathbf{x}\mathbf{y}^T$ , with rank  $r = 1$ . It has  $\mathbf{u}_1 = \mathbf{x}$  and  $\mathbf{v}_1 = \mathbf{y}$  and  $\sigma_1 = 1$ . For the full SVD, complete  $\mathbf{u}_1 = \mathbf{x}$  to an orthonormal basis of  $\mathbf{u}$ 's, and complete  $\mathbf{v}_1 = \mathbf{y}$  to an orthonormal basis of  $\mathbf{v}$ 's. No new  $\sigma$ 's, only  $\sigma_1 = 1$ .

### Proof of the SVD

We need to show how those amazing  $\mathbf{u}$ 's and  $\mathbf{v}$ 's can be constructed. The  $\mathbf{v}$ 's will be **orthonormal eigenvectors of  $A^T A$** . This must be true because we are aiming for

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T \Sigma V^T. \quad (5)$$

On the right you see the eigenvector matrix  $V$  for the symmetric positive (semi) definite matrix  $A^T A$ . And  $(\Sigma^T \Sigma)$  must be the eigenvalue matrix of  $(A^T A)$ : Each  $\sigma^2$  is  $\lambda(A^T A)$ !

Now  $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$  tells us the unit vectors  $\mathbf{u}_1$  to  $\mathbf{u}_r$ . This is the key equation (1). The essential point—the whole reason that the SVD succeeds—is that those unit vectors  $\mathbf{u}_1$  to  $\mathbf{u}_r$  are automatically orthogonal to each other (*because the  $\mathbf{v}$ 's are orthogonal*):

$$\text{Key step} \quad \mathbf{u}_i^T \mathbf{u}_j = \left( \frac{A\mathbf{v}_i}{\sigma_i} \right)^T \left( \frac{A\mathbf{v}_j}{\sigma_j} \right) = \frac{\mathbf{v}_i^T A^T A \mathbf{v}_j}{\sigma_i \sigma_j} = \frac{\sigma_i^2}{\sigma_i \sigma_j} \mathbf{v}_i^T \mathbf{v}_j = \mathbf{zero}. \quad (6)$$

The  $\mathbf{v}$ 's are eigenvectors of  $A^T A$  (symmetric). They are orthogonal and now the  $\mathbf{u}$ 's are also orthogonal. *Actually those  $\mathbf{u}$ 's will be eigenvectors of  $AA^T$ .*

Finally we complete the  $\mathbf{v}$ 's and  $\mathbf{u}$ 's to  $n$   $\mathbf{v}$ 's and  $m$   $\mathbf{u}$ 's with any orthonormal bases for the nullspaces  $\mathcal{N}(A)$  and  $\mathcal{N}(A^T)$ . We have found  $V$  and  $\Sigma$  and  $U$  in  $A = U\Sigma V^T$ .

### An Example of the SVD

Here is an example to show the computation of three matrices in  $A = U\Sigma V^T$ .

**Example 3** Find the matrices  $U, \Sigma, V$  for  $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ . The rank is  $r = 2$ .

With rank 2, this  $A$  has positive singular values  $\sigma_1$  and  $\sigma_2$ . We will see that  $\sigma_1$  is larger than  $\lambda_{\max} = 5$ , and  $\sigma_2$  is smaller than  $\lambda_{\min} = 3$ . Begin with  $A^T A$  and  $AA^T$ :

$$A^T A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \quad AA^T = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$$

Those have the same trace (50) and the same eigenvalues  $\sigma_1^2 = 45$  and  $\sigma_2^2 = 5$ . The square roots are  $\sigma_1 = \sqrt{45}$  and  $\sigma_2 = \sqrt{5}$ . Then  $\sigma_1 \sigma_2 = 15$  and this is the determinant of  $A$ .

A key step is to find the eigenvectors of  $A^T A$  (with eigenvalues 45 and 5):

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 45 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 5 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Then  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are those (orthogonal!) eigenvectors rescaled to length 1.

**Right singular vectors**  $v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$   $v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ .  $u_i =$  **left singular vectors**.

Now compute  $Av_1$  and  $Av_2$  which will be  $\sigma_1 u_1 = \sqrt{45} u_1$  and  $\sigma_2 u_2 = \sqrt{5} u_2$ :

$$Av_1 = \frac{3}{\sqrt{2}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sqrt{45} \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sigma_1 u_1$$

$$Av_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \sqrt{5} \frac{1}{\sqrt{10}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \sigma_2 u_2$$

The division by  $\sqrt{10}$  makes  $u_1$  and  $u_2$  orthonormal. Then  $\sigma_1 = \sqrt{45}$  and  $\sigma_2 = \sqrt{5}$  as expected. The Singular Value Decomposition is  $A = U\Sigma V^T$ :

$$U = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sqrt{45} & \\ & \sqrt{5} \end{bmatrix} \quad V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \quad (7)$$

$U$  and  $V$  contain orthonormal bases for the column space and the row space (both spaces are just  $\mathbf{R}^2$ ). The real achievement is that those two bases diagonalize  $A$ :  $AV$  equals  $U\Sigma$ . Then **the matrix  $U^T AV = \Sigma$  is diagonal**.

The matrix  $A$  splits into a combination of two rank-one matrices, columns times rows:

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \frac{\sqrt{45}}{\sqrt{20}} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} + \frac{\sqrt{5}}{\sqrt{20}} \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = A.$$

### An Extreme Matrix

Here is a larger example, when the  $u$ 's and the  $v$ 's are just columns of the identity matrix. So the computations are easy, but keep your eye on the *order of the columns*. The matrix  $A$  is badly lopsided (strictly triangular). All its eigenvalues are zero.  $AA^T$  is not close to  $A^T A$ . The matrices  $U$  and  $V$  will be permutations that fix these problems properly.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{array}{l} \text{eigenvalues } \lambda = 0, 0, 0, 0 \text{ all zero!} \\ \text{only one eigenvector } (1, 0, 0, 0) \\ \text{singular values } \sigma = 3, 2, 1 \\ \text{singular vectors are columns of } I \end{array}$$

We always start with  $A^T A$  and  $AA^T$ . They are diagonal (with easy  $v$ 's and  $u$ 's):

$$A^T A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix} \quad AA^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Their eigenvectors ( $u$ 's for  $AA^T$  and  $v$ 's for  $A^T A$ ) go in decreasing order  $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$  of the eigenvalues. These eigenvalues  $\sigma^2 = 9, 4, 1$  are not zero!

$$U = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 3 & & & \\ & 2 & & \\ & & 1 & \\ & & & 0 \end{bmatrix} \quad V = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Those first columns  $u_1$  and  $v_1$  have 1's in positions 3 and 4. Then  $u_1 \sigma_1 v_1^T$  picks out the biggest number  $A_{34} = 3$  in the original matrix  $A$ . The three rank-one matrices in the SVD come exactly from the numbers 3, 2, 1 in  $A$ .

$$A = U \Sigma V^T = 3u_1 v_1^T + 2u_2 v_2^T + 1u_3 v_3^T.$$

*Note* Suppose I remove the last row of  $A$  (all zeros). Then  $A$  is a 3 by 4 matrix and  $AA^T$  is 3 by 3—its fourth row and column will disappear. We still have eigenvalues  $\lambda = 1, 4, 9$  in  $A^T A$  and  $AA^T$ , producing the same singular values  $\sigma = 3, 2, 1$  in  $\Sigma$ .

Removing the zero row of  $A$  (now  $3 \times 4$ ) just removes the last row of  $\Sigma$  together with the last row and column of  $U$ . Then  $(3 \times 4) = (3 \times 3)(3 \times 4)(4 \times 4)$ . The SVD is totally adapted to rectangular matrices.

A good thing, because the rows and columns of a data matrix  $A$  often have completely different meanings (like a spreadsheet). If we have the grades for all courses, there would be a column for each student and a row for each course: The entry  $a_{ij}$  would be the grade. Then  $\sigma_1 u_1 v_1^T$  could have  $u_1 =$  **combination course** and  $v_1 =$  **combination student**. And  $\sigma_1$  would be the grade for those combinations: the highest grade.

The matrix  $A$  could count the frequency of key words in a journal: A different article for each column of  $A$  and a different word for each row. The whole journal is indexed by the matrix  $A$  and the most important information is in  $\sigma_1 u_1 v_1^T$ . Then  $\sigma_1$  is the largest frequency for a hyperword (the word combination  $u_1$ ) in the hyperarticle  $v_1$ .

I will soon show pictures for a different problem: *A photo broken into SVD pieces.*

### Singular Value Stability versus Eigenvalue Instability

The 4 by 4 example  $A$  provides an example (an extreme case) of the instability of eigenvalues. **Suppose the 4,1 entry barely changes** from zero to  $1/60,000$ . The rank is now 4.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ \frac{1}{60,000} & 0 & 0 & 0 \end{bmatrix}$$

That change by only  $1/60,000$  produces a much bigger jump in the eigenvalues of  $A$   
 $\lambda = 0, 0, 0, 0$  to  $\lambda = \frac{1}{10}, \frac{i}{10}, \frac{-1}{10}, \frac{-i}{10}$

The four eigenvalues moved from zero onto a circle around zero. The circle has radius  $\frac{1}{10}$  when the new entry is only  $1/60,000$ . This shows serious instability of eigenvalues when  $AA^T$  is far from  $A^T A$ . At the other extreme, if  $A^T A = AA^T$  (a “normal matrix”) the eigenvectors of  $A$  are orthogonal and the eigenvalues of  $A$  are totally stable.

By contrast, **the singular values of any matrix are stable**. They don’t change more than the change in  $A$ . In this example, the new singular values are **3, 2, 1, and  $1/60,000$** . The matrices  $U$  and  $V$  stay the same. The new fourth piece of  $A$  is  $\sigma_4 u_4 v_4^T$ , with fifteen zeros and that small entry  $\sigma_4 = 1/60,000$ .

### Singular Vectors of $A$ and Eigenvectors of $S = A^T A$

Equations (5–6) “proved” the SVD *all at once*. The singular vectors  $v_i$  are the eigenvectors  $q_i$  of  $S = A^T A$ . The eigenvalues  $\lambda_i$  of  $S$  are the same as  $\sigma_i^2$  for  $A$ . The rank  $r$  of  $S$  equals the rank  $r$  of  $A$ . The all-important rank-one expansions (from columns times rows) were perfectly parallel:

**Symmetric  $S$**   
**SVD of  $A$**

$$\begin{aligned} S &= Q\Lambda Q^T &= \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots + \lambda_r q_r q_r^T \\ A &= U\Sigma V^T &= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T \end{aligned}$$

The  $q$ ’s are orthonormal, the  $u$ ’s are orthonormal, the  $v$ ’s are orthonormal. Beautiful.

But I want to look again, for two good reasons. One is to fix a weak point in the eigenvalue part, where Chapter 6 was not complete. If  $\lambda$  is a *double* eigenvalue of  $S$ , we can and must find *two* orthonormal eigenvectors. The other reason is to see how the SVD picks off the largest term  $\sigma_1 u_1 v_1^T$  before  $\sigma_2 u_2 v_2^T$ . We want to understand the eigenvalues  $\lambda$  (of  $S$ ) and singular values  $\sigma$  (of  $A$ ) **one at a time instead of all at once**.

Start with the largest eigenvalue  $\lambda_1$  of  $S$ . It solves this problem:

$$\lambda_1 = \text{maximum ratio } \frac{x^T S x}{x^T x}. \quad \text{The winning vector is } x = q_1 \text{ with } S q_1 = \lambda_1 q_1. \quad (8)$$

Compare with the largest singular value  $\sigma_1$  of  $A$ . It solves this problem:

$$\sigma_1 = \text{maximum ratio } \frac{\|Ax\|}{\|x\|}. \quad \text{The winning vector is } x = v_1 \text{ with } A v_1 = \sigma_1 u_1. \quad (9)$$

This “one at a time approach” applies also to  $\lambda_2$  and  $\sigma_2$ . But not all  $x$ ’s are allowed:

$$\lambda_2 = \text{maximum ratio } \frac{x^T S x}{x^T x} \text{ among all } x\text{'s with } q_1^T x = 0. \text{ The winning } x \text{ is } q_2. \quad (10)$$

$$\sigma_2 = \text{maximum ratio } \frac{\|Ax\|}{\|x\|} \text{ among all } x\text{'s with } v_1^T x = 0. \text{ The winning } x \text{ is } v_2. \quad (11)$$

When  $S = A^T A$  we find  $\lambda_1 = \sigma_1^2$  and  $\lambda_2 = \sigma_2^2$ . Why does this approach succeed?

Start with the ratio  $r(x) = x^T S x / x^T x$ . This is called the *Rayleigh quotient*. To maximize  $r(x)$ , set its partial derivatives to zero:  $\partial r / \partial x_i = 0$  for  $i = 1, \dots, n$ . Those derivatives are messy and here is the result: one vector equation for the winning  $x$ :

$$\text{The derivatives of } r(x) = \frac{x^T S x}{x^T x} \text{ are zero when } Sx = r(x)x. \quad (12)$$

So the winning  $x$  is an eigenvector of  $S$ . The maximum ratio  $r(x)$  is the largest eigenvalue  $\lambda_1$  of  $S$ . All good. Now turn to  $A$ —and notice the connection to  $S = A^T A$ !

$$\text{Maximizing } \frac{\|Ax\|}{\|x\|} \text{ also maximizes } \left( \frac{\|Ax\|}{\|x\|} \right)^2 = \frac{x^T A^T A x}{x^T x} = \frac{x^T S x}{x^T x}.$$

So the winning  $x = v_1$  in (9) is the top eigenvector  $q_1$  of  $S = A^T A$  in (8).

Now I have to explain why  $q_2$  and  $v_2$  are the winning vectors in (10) and (11). We know they are orthogonal to  $q_1$  and  $v_1$ , so they are allowed in those competitions. These paragraphs can be omitted by readers who aim to see the SVD in action (Section 7.2).

Start with any orthogonal matrix  $Q_1$  that has  $q_1$  in its first column. The other  $n - 1$  orthonormal columns just have to be orthogonal to  $q_1$ . Then use  $Sq_1 = \lambda_1 q_1$ :

$$SQ_1 = S \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & w^T \\ 0 & S_{n-1} \end{bmatrix} = Q_1 \begin{bmatrix} \lambda_1 & w^T \\ 0 & S_{n-1} \end{bmatrix}. \quad (13)$$

Multiply by  $Q_1^T$ , remember  $Q_1^T Q_1 = I$ , and recognize that  $Q_1^T S Q_1$  is symmetric like  $S$ :

$$\text{The symmetry of } Q_1^T S Q_1 = \begin{bmatrix} \lambda_1 & w^T \\ 0 & S_{n-1} \end{bmatrix} \text{ forces } w = 0 \text{ and } S_{n-1}^T = S_{n-1}.$$

The requirement  $q_1^T x = 0$  has reduced the maximum problem (10) to size  $n - 1$ . The largest eigenvalue of  $S_{n-1}$  will be the *second largest* for  $S$ . **It is  $\lambda_2$ .** The winning vector in (10) will be the eigenvector  $q_2$  with  $Sq_2 = \lambda_2 q_2$ .

We just keep going—or use the magic word *induction*—to produce all the eigenvectors  $q_1, \dots, q_n$  and their eigenvalues  $\lambda_1, \dots, \lambda_n$ . The Spectral Theorem  $S = Q\Lambda Q^T$  is proved even with repeated eigenvalues. All symmetric matrices can be diagonalized.

Similarly the SVD is found one step at a time from (9) and (11) and onwards. Section 7.2 will show the geometry—we are finding the axes of an ellipse. Here I ask a different question: **How are the  $\lambda$ ’s and  $\sigma$ ’s actually computed?**

### Computing the Eigenvalues of $S$ and the SVD of $A$

The singular values  $\sigma_i$  of  $A$  are the square roots of the eigenvalues  $\lambda_i$  of  $S = A^T A$ . This connects the SVD to the *symmetric eigenvalue problem* (symmetry is good). In the end we don't want to multiply  $A^T$  times  $A$  (squaring is time-consuming; not good). But the same ideas govern both problems. How to compute the  $\lambda$ 's for  $S$  and singular values  $\sigma$  for  $A$ ?

The first idea is to *produce zeros in  $A$  and  $S$  without changing the  $\sigma$ 's and the  $\lambda$ 's*. Singular vectors and eigenvectors will change—no problem. The similar matrix  $Q^{-1}SQ$  has the **same  $\lambda$ 's as  $S$** . If  $Q$  is orthogonal, this matrix is  $Q^T SQ$  and still symmetric. Section 11.3 will show how to build  $Q$  from 2 by 2 rotations so that  $Q^T SQ$  is **symmetric and tridiagonal** (many zeros). We can't get all the way to a diagonal matrix  $\Lambda$ —which would show all the eigenvalues of  $S$ —without a new idea and more work in Chapter 11.

For the SVD, what is the parallel to  $Q^{-1}SQ$ ? Now we don't want to change any singular values of  $A$ . Natural answer: You can multiply  $A$  by *two different orthogonal matrices*  $Q_1$  and  $Q_2$ . Use them to produce zeros in  $Q_1^T A Q_2$ . The  $\sigma$ 's and  $\lambda$ 's don't change:

$$(Q_1^T A Q_2)^T (Q_1^T A Q_2) = Q_2^T A^T A Q_2 = Q_2^T S Q_2 \text{ gives the same } \sigma(A) \text{ from the same } \lambda(S).$$

The freedom of two  $Q$ 's allows us to reach  $Q_1^T A Q_2 =$  **bidiagonal matrix** (2 diagonals). This compares perfectly to  $Q^T S Q = 3$  diagonals. It is nice to notice the connection between them:  $(\text{bidiagonal})^T (\text{bidiagonal}) = \text{tridiagonal}$ .

The final steps to a *diagonal*  $\Lambda$  and a *diagonal*  $\Sigma$  need more ideas. This problem can't be easy, because underneath we are solving  $\det(S - \lambda I) = 0$  for polynomials of degree  $n = 100$  or 1000 or more. The favorite way to find  $\lambda$ 's and  $\sigma$ 's uses simple orthogonal matrices to approach  $Q^T S Q = \Lambda$  and  $U^T A V = \Sigma$ . **We stop when very close to  $\Lambda$  and  $\Sigma$ .**

This 2-step approach is built into the commands `eig(S)` and `svd(A)`.

### ■ REVIEW OF THE KEY IDEAS ■

1. The SVD factors  $A$  into  $U \Sigma V^T$ , with  $r$  singular values  $\sigma_1 \geq \dots \geq \sigma_r > 0$ .
2. The numbers  $\sigma_1^2, \dots, \sigma_r^2$  are the nonzero eigenvalues of  $AA^T$  and  $A^T A$ .
3. The orthonormal columns of  $U$  and  $V$  are eigenvectors of  $AA^T$  and  $A^T A$ .
4. Those columns hold orthonormal bases for the four fundamental subspaces of  $A$ .
5. Those bases diagonalize the matrix:  $Av_i = \sigma_i u_i$  for  $i \leq r$ . This is  $AV = U\Sigma$ .
6.  $A = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$  and  $\sigma_1$  is the maximum of the ratio  $\|Ax\| / \|x\|$ .



### ■ WORKED EXAMPLES ■

**7.1.A** Identify by name these decompositions of  $A$  into a sum of columns times rows:

1. *Orthogonal* columns  $u_1\sigma_1, \dots, u_r\sigma_r$  times *orthonormal* rows  $v_1^T, \dots, v_r^T$ .
  2. *Orthonormal* columns  $q_1, \dots, q_r$  times *triangular* rows  $r_1^T, \dots, r_r^T$ .
  3. *Triangular* columns  $l_1, \dots, l_r$  times *triangular* rows  $u_1^T, \dots, u_r^T$ .
- Where do the rank and the pivots and the singular values of  $A$  come into this picture?

**Solution** These three factorizations are basic to linear algebra, pure or applied:

1. **Singular Value Decomposition**  $A = U\Sigma V^T$
2. **Gram-Schmidt Orthogonalization**  $A = QR$
3. **Gaussian Elimination**  $A = LU$

You might prefer to separate out singular values  $\sigma_i$  and heights  $h_i$  and pivots  $d_i$ :

1.  $A = U\Sigma V^T$  with unit vectors in  $U$  and  $V$ . **The singular values  $\sigma_i$  are in  $\Sigma$ .**
2.  $A = QHR$  with unit vectors in  $Q$  and diagonal 1's in  $R$ . **The heights  $h_i$  are in  $H$ .**
3.  $A = LDU$  with diagonal 1's in  $L$  and  $U$ . **The pivots  $d_i$  are in  $D$ .**

Each  $h_i$  tells the height of column  $i$  above the plane of columns 1 to  $i-1$ . The volume of the full  $n$ -dimensional box ( $r = m = n$ ) comes from  $A = U\Sigma V^T = LDU = QHR$ :

$$|\det A| = |\text{product of } \sigma\text{'s}| = |\text{product of } d\text{'s}| = |\text{product of } h\text{'s}|.$$

**7.1.B** Show that  $\sigma_1 \geq |\lambda|_{\max}$ . **The largest singular value dominates all eigenvalues.**

**Solution** Start from  $A = U\Sigma V^T$ . Remember that multiplying by an orthogonal matrix does not change length:  $\|Qx\| = \|x\|$  because  $\|Qx\|^2 = x^T Q^T Q x = x^T x = \|x\|^2$ . This applies to  $Q = U$  and  $Q = V^T$ . In between is the diagonal matrix  $\Sigma$ .

$$\|Ax\| = \|U\Sigma V^T x\| = \|\Sigma V^T x\| \leq \sigma_1 \|V^T x\| = \sigma_1 \|x\|. \quad (14)$$

An eigenvector has  $\|Ax\| = |\lambda| \|x\|$ . So (14) says that  $|\lambda| \|x\| \leq \sigma_1 \|x\|$ . Then  $|\lambda| \leq \sigma_1$ .

Apply also to the unit vector  $x = (1, 0, \dots, 0)$ . Now  $Ax$  is the first column of  $A$ . Then by inequality (14), this column has length  $\leq \sigma_1$ . Every entry must have  $|a_{ij}| \leq \sigma_1$ .

Equation (14) shows again that **the maximum value of  $\|Ax\|/\|x\|$  equals  $\sigma_1$** .

Section 11.2 will explain how the ratio  $\sigma_{\max}/\sigma_{\min}$  governs the roundoff error in solving  $Ax = b$ . MATLAB warns you if this “condition number” is large. Then  $x$  is unreliable.