Chapter 7

The Singular Value Decomposition (SVD)

1 The SVD produces **orthonormal bases** of $\boldsymbol{v}$'s and $\boldsymbol{u}$'s for the four fundamental subspaces.

2 Using those bases, A becomes a diagonal matrix Σ and $\boldsymbol{Av_i = \sigma_i u_i}$ **:** $\boldsymbol{\sigma_i} =$ **singular value**.

3 The two-bases diagonalization $\boldsymbol{A = U\Sigma V^{\mathrm{T}}}$ often has more information than $A = X\Lambda X^{-1}$.

4 $U\Sigma V^{\mathrm{T}}$ separates A into rank-1 matrices $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$. $\quad \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$ 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 $A\boldsymbol{x} = \lambda \boldsymbol{x}$ 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**, $\boldsymbol{u}$'s and $\boldsymbol{v}$'s. The $\boldsymbol{u}$'s are in $\mathbf{R}^m$ and the $\boldsymbol{v}$'s are in $\mathbf{R}^n$. They will be the columns of an m by m matrix $\boldsymbol{U}$ and an n by n matrix $\boldsymbol{V}$. 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 $\boldsymbol{u}$'s and $\boldsymbol{v}$'s give bases for the four fundamental subspaces :

$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$	is an orthonormal basis for the **column space**
$\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m$	is an orthonormal basis for the **left nullspace** $\boldsymbol{N}(A^{\mathrm{T}})$
$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$	is an orthonormal basis for the **row space**
$\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n$	is an orthonormal basis for the **nullspace** $\boldsymbol{N}(A)$.

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

"A is diagonalized" $\quad \boxed{A\boldsymbol{v}_1 = \sigma_1 \boldsymbol{u}_1} \quad \boxed{A\boldsymbol{v}_2 = \sigma_2 \boldsymbol{u}_2} \quad \ldots \quad \boxed{A\boldsymbol{v}_r = \sigma_r \boldsymbol{u}_r} \qquad (1)$

Those **singular values $\boldsymbol{\sigma_1}$ to $\boldsymbol{\sigma_r}$** will be positive numbers: *σ_i is the length of $A\boldsymbol{v}_i$*. The σ's go into a diagonal matrix that is otherwise zero. That matrix is Σ.

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

$$\begin{array}{c}(m \text{ by } n)(n \text{ by } r) \\ \boldsymbol{AV_r = U_r\Sigma_r} \\ (m \text{ by } r)(r \text{ by } r)\end{array} \quad A\begin{bmatrix} & & \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r \\ & & \end{bmatrix} = \begin{bmatrix} & & \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \\ & & \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}. \qquad (2)$$

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

$$\begin{array}{c}(m \text{ by } n)(n \text{ by } n) \\ \boldsymbol{Av} \text{ equals } \boldsymbol{U\Sigma} \\ (m \text{ by } m)(m \text{ by } n)\end{array} \quad A\begin{bmatrix} & & & & \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r & \cdots & \boldsymbol{v}_n \\ & & & & \end{bmatrix} = \begin{bmatrix} & & & & \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r & \cdots & \boldsymbol{u}_m \\ & & & & \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}. \qquad (3)$$

The new Σ 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 $\boldsymbol{A = U\Sigma V^{\mathrm{T}}}$. This is the ***Singular Value Decomposition***. I can multiply columns $\boldsymbol{u}_i\sigma_i$ from $U\Sigma$ by rows of V^{T}:

SVD $\qquad \boxed{\boldsymbol{A = U\Sigma V^{\mathrm{T}} = u_1\sigma_1 v_1^{\mathrm{T}}} + \cdots + \boldsymbol{u_r\sigma_r v_r^{\mathrm{T}}}.} \qquad (4)$

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 $\boldsymbol{u}_i\sigma_i\boldsymbol{v}_i^{\mathrm{T}}$ of rank one: column times row.

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

Example 1 When is $\Lambda = U\Sigma V^{\mathrm{T}}$ (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 $Q\Lambda Q^{\mathrm{T}}$ coincide with $A = U\Sigma V^{\mathrm{T}}$.

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

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

Proof of the SVD

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

$$\boldsymbol{A}^{\mathbf{T}}\boldsymbol{A} = (U\Sigma V^{\mathrm{T}})^{\mathrm{T}}(U\Sigma V^{\mathrm{T}}) = V\Sigma^{\mathrm{T}}U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{\Sigma}^{\mathbf{T}}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathbf{T}}. \tag{5}$$

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

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

$$\textbf{Key step} \qquad \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j = \left(\frac{A\boldsymbol{v}_i}{\sigma_i}\right)^{\mathrm{T}}\left(\frac{A\boldsymbol{v}_j}{\sigma_j}\right) = \frac{\boldsymbol{v}_i^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{v}_j}{\sigma_i\sigma_j} = \frac{\sigma_j^2}{\sigma_i\sigma_j}\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j = \textbf{zero}. \tag{6}$$

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

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

An Example of the SVD

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

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

With rank 2, this A has positive singular values σ_1 and σ_2. We will see that σ_1 is larger than $\lambda_{\max} = 5$, and σ_2 is smaller than $\lambda_{\min} = 3$. Begin with $A^{\mathrm{T}}A$ and AA^{T}:

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

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

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

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

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

Right singular vectors $\boldsymbol{v}_1 = \dfrac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix} \quad \boldsymbol{v}_2 = \dfrac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\end{bmatrix}$. $\boldsymbol{u}_i =$ **left singular vectors**.

Now compute $A\boldsymbol{v}_1$ and $A\boldsymbol{v}_2$ which will be $\sigma_1\boldsymbol{u}_1 = \sqrt{45}\,\boldsymbol{u}_1$ and $\sigma_2\boldsymbol{u}_2 = \sqrt{5}\,\boldsymbol{u}_2$:

$$\begin{aligned}A\boldsymbol{v}_1 &= \frac{3}{\sqrt{2}}\begin{bmatrix}1\\3\end{bmatrix} &= \sqrt{45}\frac{\mathbf{1}}{\sqrt{\mathbf{10}}}\begin{bmatrix}\mathbf{1}\\\mathbf{3}\end{bmatrix} &= \sigma_1\,\boldsymbol{u}_1\\ A\boldsymbol{v}_2 &= \frac{1}{\sqrt{2}}\begin{bmatrix}-3\\1\end{bmatrix} &= \sqrt{5}\frac{\mathbf{1}}{\sqrt{\mathbf{10}}}\begin{bmatrix}\mathbf{-3}\\\mathbf{1}\end{bmatrix} &= \sigma_2\,\boldsymbol{u}_2\end{aligned}$$

The division by $\sqrt{10}$ makes $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ orthonormal. Then $\sigma_1 = \sqrt{45}$ and $\sigma_2 = \sqrt{5}$ as expected. The Singular Value Decomposition is $A = U\Sigma V^{\mathrm{T}}$:

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

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 $\boldsymbol{U^{\mathrm{T}}AV = \Sigma}$ is diagonal**.

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

$$\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} = \frac{\sqrt{\mathbf{45}}}{\sqrt{\mathbf{20}}}\begin{bmatrix}\mathbf{1} & \mathbf{1}\\\mathbf{3} & \mathbf{3}\end{bmatrix} + \frac{\sqrt{\mathbf{5}}}{\sqrt{\mathbf{20}}}\begin{bmatrix}\mathbf{3} & \mathbf{-3}\\\mathbf{-1} & \mathbf{1}\end{bmatrix} = \begin{bmatrix}3 & 0\\4 & 5\end{bmatrix} = A.$$

An Extreme Matrix

Here is a larger example, when the $\boldsymbol{u}$'s and the $\boldsymbol{v}$'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 $A^{\mathrm{T}}A$. 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 $A^{\mathrm{T}}A$ and AA^{T}. They are diagonal (with easy $\boldsymbol{v}$'s and $\boldsymbol{u}$'s):

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

Their eigenvectors ($\boldsymbol{u}$'s for AA^{T} and $\boldsymbol{v}$'s for $A^{\mathrm{T}}A$) 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!

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

Those first columns $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ have 1's in positions 3 and 4. Then $\boldsymbol{u}_1\sigma_1\boldsymbol{v}_1^{\mathrm{T}}$ 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.

$$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathbf{T}} = \mathbf{3}\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \mathbf{2}\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + \mathbf{1}\boldsymbol{u}_3\boldsymbol{v}_3^{\mathrm{T}}.$$

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 $A^{\mathrm{T}}A$ and AA^{T}, producing the same singular values $\sigma = 3, 2, 1$ in Σ.

Removing the zero row of A (now 3×4) just removes the last row of Σ 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 $\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$ could have $\boldsymbol{u}_1$ = **combination course** and $\boldsymbol{v}_1$ = **combination student**. And σ_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 $\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$. Then σ_1 is the largest frequency for a hyperword (the word combination $\boldsymbol{u}_1$) in the hyperarticle $\boldsymbol{v}_1$.

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.

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

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

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 $A^{\mathrm{T}}A$. At the other extreme, if $A^{\mathrm{T}}A = AA^{\mathrm{T}}$ (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 $\mathbf{3}, \mathbf{2}, \mathbf{1}$, and $\mathbf{1/60,000}$. The matrices U and V stay the same. The new fourth piece of A is $\sigma_4 \boldsymbol{u}_4 \boldsymbol{v}_4^{\mathrm{T}}$, with fifteen zeros and that small entry $\sigma_4 = 1/60,000$.

Singular Vectors of A and Eigenvectors of $S = A^{\mathrm{T}}A$

Equations (5–6) "proved" the SVD *all at once*. The singular vectors $\boldsymbol{v}_i$ are the eigenvectors $\boldsymbol{q}_i$ of $S = A^{\mathrm{T}}A$. The eigenvalues λ_i of S are the same as σ_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	$S = Q\Lambda Q^{\mathrm{T}}$	$= \lambda_1 \boldsymbol{q}_1 \boldsymbol{q}_1^{\mathrm{T}} + \lambda_2 \boldsymbol{q}_2 \boldsymbol{q}_2^{\mathrm{T}} + \cdots + \lambda_r \boldsymbol{q}_r \boldsymbol{q}_r^{\mathrm{T}}$
SVD of A	$A = U\Sigma V^{\mathrm{T}}$	$= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$

The $\boldsymbol{q}$'s are orthonormal, the $\boldsymbol{u}$'s are orthonormal, the $\boldsymbol{v}$'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 λ 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 $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$ before $\sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}}$. We want to understand the eigenvalues λ (of S) and singular values σ (of A) **one at a time instead of all at once**.

Start with the largest eigenvalue λ_1 of S. It solves this problem:

$$\boldsymbol{\lambda_1} = \textbf{maximum ratio } \frac{\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}}. \quad \text{The winning vector is } \boldsymbol{x} = \boldsymbol{q}_1 \text{ with } S\boldsymbol{q}_1 = \lambda_1 \boldsymbol{q}_1. \tag{8}$$

Compare with the largest singular value σ_1 of A. It solves this problem:

$$\boldsymbol{\sigma_1} = \textbf{maximum ratio } \frac{||A\boldsymbol{x}||}{||\boldsymbol{x}||}. \quad \text{The winning vector is } \boldsymbol{x} = \boldsymbol{v}_1 \text{ with } A\boldsymbol{v}_1 = \sigma_1 \boldsymbol{u}_1. \tag{9}$$

This "one at a time approach" applies also to λ_2 and σ_2. But not all $\boldsymbol{x}$'s are allowed:

$$\boldsymbol{\lambda_2} = \textbf{maximum ratio } \frac{\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}} \text{ among all } \boldsymbol{x}\text{'s with } \boldsymbol{q}_1^{\mathrm{T}} \boldsymbol{x} = 0. \text{ The winning } \boldsymbol{x} \text{ is } \boldsymbol{q}_2. \tag{10}$$

$$\boldsymbol{\sigma_2} = \textbf{maximum ratio } \frac{||A\boldsymbol{x}||}{||\boldsymbol{x}||} \text{ among all } \boldsymbol{x}\text{'s with } \boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{x} = 0. \text{ The winning } \boldsymbol{x} \text{ is } \boldsymbol{v}_2. \tag{11}$$

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

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

$$\textbf{The derivatives of } \boldsymbol{r(x)} = \frac{\boldsymbol{x}^{\mathbf{T}} \boldsymbol{S} \boldsymbol{x}}{\boldsymbol{x}^{\mathbf{T}} \boldsymbol{x}} \textbf{ are zero when} \qquad S\boldsymbol{x} = r(\boldsymbol{x})\boldsymbol{x}. \tag{12}$$

So the winning $\boldsymbol{x}$ is an eigenvector of S. The maximum ratio $r(\boldsymbol{x})$ is the largest eigenvalue λ_1 of S. All good. Now turn to A—and notice the connection to $S = A^{\mathrm{T}} A$!

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

So the winning $\boldsymbol{x} = \boldsymbol{v}_1$ in (9) is the top eigenvector $\boldsymbol{q}_1$ of $S = A^{\mathrm{T}} A$ in (8).

Now I have to explain why $\boldsymbol{q}_2$ and $\boldsymbol{v}_2$ are the winning vectors in (10) and (11). We know they are orthogonal to $\boldsymbol{q}_1$ and $\boldsymbol{v}_1$, 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 $\boldsymbol{q}_1$ in its first column. The other $n-1$ orthonormal columns just have to be orthogonal to $\boldsymbol{q}_1$. Then use $S\boldsymbol{q}_1 = \lambda_1 \boldsymbol{q}_1$:

$$SQ_1 = S \left[\boldsymbol{q}_1 \;\; \boldsymbol{q}_2 \ldots \boldsymbol{q}_n \right] = \left[\boldsymbol{q}_1 \;\; \boldsymbol{q}_2 \ldots \boldsymbol{q}_n \right] \begin{bmatrix} \lambda_1 & \boldsymbol{w}^{\mathrm{T}} \\ \mathbf{0} & S_{n-1} \end{bmatrix} = Q_1 \begin{bmatrix} \lambda_1 & \boldsymbol{w}^{\mathrm{T}} \\ \mathbf{0} & S_{n-1} \end{bmatrix}. \tag{13}$$

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

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

The requirement $\boldsymbol{q}_1^{\mathrm{T}} \boldsymbol{x} = 0$ 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 $\boldsymbol{\lambda_2}$.** The winning vector in (10) will be the eigenvector $\boldsymbol{q}_2$ with $S\boldsymbol{q}_2 = \lambda_2 \boldsymbol{q}_2$.

We just keep going—or use the magic word *induction*—to produce all the eigenvectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ and their eigenvalues $\lambda_1, \ldots, \lambda_n$. The Spectral Theorem $S = Q\Lambda Q^{\mathrm{T}}$ 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 $\boldsymbol{\lambda}$'s and $\boldsymbol{\sigma}$'s actually computed**?

Computing the Eigenvalues of S and the SVD of A

The singular values σ_i of A are the square roots of the eigenvalues λ_i of $S = A^{\mathrm{T}}A$. 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 λ's for S and singular values σ for A?

The first idea is *to produce zeros in A and S without changing the σ's and the λ's.* Singular vectors and eigenvectors will change—no problem. The similar matrix $Q^{-1}SQ$ has **the same $\boldsymbol{\lambda}$'s as $\boldsymbol{S}$**. If Q is orthogonal, this matrix is $Q^{\mathrm{T}}SQ$ and still symmetric. Section 11.3 will show how to build Q from 2 by 2 rotations so that $Q^{\mathrm{T}}SQ$ is **symmetric and tridiagonal** (many zeros). We can't get all the way to a diagonal matrix Λ—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 $Q_1^{\mathrm{T}}AQ_2$. The σ's and λ's don't change:

$$(Q_1^{\mathrm{T}}AQ_2)^{\mathrm{T}}(Q_1^{\mathrm{T}}AQ_2) = Q_2^{\mathrm{T}}A^{\mathrm{T}}AQ_2 = Q_2^{\mathrm{T}}SQ_2 \text{ gives the same } \sigma(A) \text{ from the same } \lambda(S).$$

The freedom of two Q's allows us to reach $Q_1^{\mathrm{T}}AQ_2 =$ **bidiagonal matrix** (2 diagonals). This compares perfectly to $Q^{\mathrm{T}}SQ = 3$ diagonals. It is nice to notice the connection between them: (*bidiagonal*)$^{\mathrm{T}}$ (*bidiagonal*) $=$ *tridiagonal*.

The final steps to a *diagonal* Λ and a *diagonal* Σ 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 λ's and σ's uses simple orthogonal matrices to approach $Q^{\mathrm{T}}SQ = \Lambda$ and $U^{\mathrm{T}}AV = \Sigma$. **We stop when very close to Λ and Σ.**

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 $U\Sigma V^{\mathrm{T}}$, with r singular values $\sigma_1 \geq \ldots \geq \sigma_r > 0$.

2. The numbers $\sigma_1^2, \ldots, \sigma_r^2$ are the nonzero eigenvalues of AA^{T} and $A^{\mathrm{T}}A$.

3. The orthonormal columns of U and V are eigenvectors of AA^{T} and $A^{\mathrm{T}}A$.

4. Those columns hold orthonormal bases for the four fundamental subspaces of A.

5. Those bases diagonalize the matrix: $A\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i$ for $i \leq r$. This is $\boldsymbol{AV = U\Sigma}$.

6. $A = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^{\mathrm{T}}$ and σ_1 is the maximum of the ratio $||A\boldsymbol{x}|| \,/\, ||\boldsymbol{x}||$.

WORKED EXAMPLES

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

1. *Orthogonal* columns	$\boldsymbol{u}_1\sigma_1,\ldots,\boldsymbol{u}_r\sigma_r$	times	*orthonormal* rows	$\boldsymbol{v}_1^{\mathrm{T}},\ldots,\boldsymbol{v}_r^{\mathrm{T}}$.
2. *Orthonormal* columns	$\boldsymbol{q}_1,\ldots,\boldsymbol{q}_r$	times	*triangular* rows	$\boldsymbol{r}_1^{\mathrm{T}},\ldots,\boldsymbol{r}_r^{\mathrm{T}}$.
3. *Triangular* columns	$\boldsymbol{l}_1,\ldots,\boldsymbol{l}_r$	times	*triangular* rows	$\boldsymbol{u}_1^{\mathrm{T}},\ldots,\boldsymbol{u}_r^{\mathrm{T}}$.

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 $\boldsymbol{A = U\Sigma V^{\mathrm{T}}}$**
2. **Gram-Schmidt Orthogonalization $\boldsymbol{A = QR}$**
3. **Gaussian Elimination $\boldsymbol{A = LU}$**

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

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

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 $A = U\Sigma V^{\mathrm{T}} = LDU = QHR$:

$$|\,\mathbf{det}\,\boldsymbol{A}\,| = |\,\textit{\textbf{product of}}\ \boldsymbol{\sigma}\text{'s}\,| = |\,\textit{\textbf{product of}}\ \boldsymbol{d}\text{'s}\,| = |\,\textit{\textbf{product of}}\ \boldsymbol{h}\text{'s}\,|.$$

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

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

$$\|A\boldsymbol{x}\| = \|U\Sigma V^{\mathrm{T}}\boldsymbol{x}\| = \|\Sigma V^{\mathrm{T}}\boldsymbol{x}\| \leq \sigma_1\|V^{\mathrm{T}}\boldsymbol{x}\| = \sigma_1\|\boldsymbol{x}\|. \tag{14}$$

An eigenvector has $\|A\boldsymbol{x}\| = |\lambda|\|\boldsymbol{x}\|$. So (14) says that $|\lambda|\|\boldsymbol{x}\| \leq \sigma_1\|\boldsymbol{x}\|$. Then $\boldsymbol{|\lambda| \leq \sigma_1}$.

Apply also to the unit vector $\boldsymbol{x} = (1, 0, \ldots, 0)$. Now $A\boldsymbol{x}$ 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 $\boldsymbol{||A\boldsymbol{x}||/||\boldsymbol{x}||}$ equals $\boldsymbol{\sigma_1}$.***

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