Chapter 7

The Singular Value Decomposition (SVD)

7.1 Image Processing by Linear Algebra

1 An image is a large matrix of grayscale values, one for each pixel and color.

2 When nearby pixels are correlated (not random) the image can be compressed.

3 The SVD separates any matrix A into rank one pieces $\boldsymbol{uv}^{\mathrm{T}} =$ (**column**)(**row**).

4 The columns and rows are eigenvectors of symmetric matrices AA^{T} and $A^{\mathrm{T}}A$.

The singular value theorem for A is the eigenvalue theorem for $A^{\mathrm{T}}A$ and AA^{T}.

That is a quick preview of what you will see in this chapter. A has *two* sets of singular vectors (the eigenvectors of $A^{\mathrm{T}}A$ and AA^{T}). There is *one* set of positive singular values (because $A^{\mathrm{T}}A$ has the same positive eigenvalues as AA^{T}). A is often rectangular, but $A^{\mathrm{T}}A$ and AA^{T} are square, symmetric, and positive semidefinite.

The Singular Value Decomposition (SVD) separates any matrix into simple pieces.

Each piece is a column vector times a row vector. An m by n matrix has m times n entries (a big number when the matrix represents an image). But a column and a row only have **$\boldsymbol{m+n}$ components, far less than $\boldsymbol{m}$ times $\boldsymbol{n}$**. Those (column)(row) pieces are full size matrices that can be processed with extreme speed—they need only m *plus* n numbers.

Unusually, this image processing application of the SVD is coming before the matrix algebra it depends on. I will start with simple images that only involve one or two pieces. Right now I am thinking of an image as a large rectangular matrix. The entries a_{ij} tell the grayscales of all the pixels in the image. Think of a pixel as a small square, i steps across and j steps up from the lower left corner. Its grayscale is a number (often a whole number in the range $0 \leq a_{ij} < 256 = 2^8$). An all-white pixel has $a_{ij} = 255 = 11111111$. That number has eight 1's when the computer writes 255 in binary notation.

You see how an image that has m times n pixels, with each pixel using 8 bits (0 or 1) for its grayscale, becomes an m by n matrix with 256 possible values for each entry a_{ij}.

In short, an image is a large matrix. To copy it perfectly, we need $8\,(m)(n)$ bits of information. High definition television typically has $m = 1080$ and $n = 1920$. Often there are 24 frames each second and you probably like to watch in color (3 color scales). This requires transmitting $(3)(8)(48,470,400)$ bits per second. That is too expensive and it is not done. The transmitter can't keep up with the show.

When compression is well done, you can't see the difference from the original. *Edges in the image* (sudden changes in the grayscale) are the hard parts to compress.

Major success in compression will be impossible if every a_{ij} is an independent random number. We totally depend on the fact that *nearby pixels generally have similar grayscales*. An edge produces a sudden jump when you cross over it. Cartoons are more compressible than real-world images, with edges everywhere.

For a video, the numbers a_{ij} don't change much between frames. **We only transmit the small changes**. This is *difference coding* in the H.264 video compression standard (on this book's website). We compress each change matrix by linear algebra (and by nonlinear "quantization" for an efficient step to integers in the computer).

The natural images that we see every day are absolutely ready and open for compression—but that doesn't make it easy to do.

Low Rank Images (Examples)

The easiest images to compress are all black or all white or all a constant grayscale g. The matrix A has the same number g in every entry : $a_{ij} = g$. When $g = 1$ and $m = n = 6$, here is an extreme example of the central SVD dogma of image processing :

Example 1 **Don't send** $A = \begin{bmatrix} 1&1&1&1&1&1\\ 1&1&1&1&1&1\\ 1&1&1&1&1&1\\ 1&1&1&1&1&1\\ 1&1&1&1&1&1\\ 1&1&1&1&1&1 \end{bmatrix}$ **Send this** $A = \begin{bmatrix} 1\\1\\1\\1\\1\\1 \end{bmatrix} \begin{bmatrix} 1&1&1&1&1&1 \end{bmatrix}$

36 numbers become 12 numbers. With 300 by 300 pixels, $90,000$ numbers become 600. And if we define the all-ones vector $\boldsymbol{x}$ in advance, we only have to send **one number**. That number would be the constant grayscale g that multiplies $\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$ to produce the matrix.

Of course this first example is extreme. But it makes an important point. If there are special vectors like $\boldsymbol{x}$ = **ones** that can usefully be defined in advance, then image processing can be extremely fast. The battle is between **preselected bases** (the Fourier basis allows speed-up from the FFT) and **adaptive bases** determined by the image. The SVD produces bases from the image itself—this is adaptive and it can be expensive.

I am not saying that the SVD always or usually gives the most effective algorithm in practice. The purpose of these next examples is instruction and not production.

Example 2
"ace flag"
French flag $\boldsymbol{A}$
Italian flag $\boldsymbol{A}$
German flag $\boldsymbol{A}^{\mathrm{T}}$

Don't send $A = \begin{bmatrix} a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \end{bmatrix}$ **Send** $A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} a & a & c & c & e & e \end{bmatrix}$

This flag has 3 colors but it still has rank 1. We still have one column times one row. The 36 entries could even be all different, provided they keep that rank 1 pattern $A = \boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$. But when the rank moves up to $r = 2$, we need $\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$. Here is one choice:

Example 3
Embedded square

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

The 1's and the 0 in A could be blocks of 1's and a block of 0's. *We would still have rank* 2. We would still only need two terms $\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$ and $\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$. A 6 by 6 image would be compressed into 24 numbers. An N by N image (N^2 numbers) would be compressed into $4N$ numbers from the four vectors $\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2$.

Have I made the best choice for the $\boldsymbol{u}$'s and $\boldsymbol{v}$'s ? This is *not* the choice from the SVD! I notice that $\boldsymbol{u}_1 = (1, 1)$ is not orthogonal to $\boldsymbol{u}_2 = (1, 0)$. And $\boldsymbol{v}_1 = (1, 1)$ is not orthogonal to $\boldsymbol{v}_2 = (0, 1)$. The theory says that orthogonality will produce a smaller second piece $c_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$. (**The SVD chooses rank one pieces in order of importance.**)

If the rank of A is much higher than 2, as we expect for real images, then A will add up many rank one pieces. We want the small ones to be really small—they can be discarded with no loss to visual quality. Image compression becomes lossy, but good image compression is virtually undetectable by the human visual system.

The question becomes: **What are the orthogonal choices from the SVD**?

Eigenvectors for the SVD

I want to introduce the use of eigenvectors. But the eigenvectors of most images are not orthogonal. Furthermore the eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2$ give only one set of vectors, and we want two sets ($\boldsymbol{u}$'s and $\boldsymbol{v}$'s). The answer to both of those difficulties is the SVD idea:

Use the eigenvectors $\boldsymbol{u}$ of AA^{T} and the eigenvectors $\boldsymbol{v}$ of $A^{\mathrm{T}}A$.

Since AA^{T} and $A^{\mathrm{T}}A$ are automatically symmetric (but not usually equal!) the $\boldsymbol{u}$'s will be one orthogonal set and the eigenvectors $\boldsymbol{v}$ will be another orthogonal set. We can and will make them all unit vectors: $||\boldsymbol{u}_i|| = 1$ and $||\boldsymbol{v}_i|| = 1$. Then our rank 2 matrix will be $A = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$. The size of those numbers σ_1 and σ_2 will decide whether they can be ignored in compression. *We keep larger σ's, we discard small σ's.*

The $\boldsymbol{u}$'s from the SVD are called **left singular vectors** (unit eigenvectors of AA^{T}). The $\boldsymbol{v}$'s are **right singular vectors** (unit eigenvectors of $A^{\mathrm{T}}A$). The σ's are **singular values**, square roots of the equal eigenvalues of AA^{T} and $A^{\mathrm{T}}A$:

Choices from the SVD $\quad AA^{\mathrm{T}}\boldsymbol{u}_i = \sigma_i^2\boldsymbol{u}_i \quad A^{\mathrm{T}}A\boldsymbol{v}_i = \sigma_i^2\boldsymbol{v}_i \quad A\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i \quad$ (1)

In Example 3 (the embedded square), here are the symmetric matrices AA^{T} and $A^{\mathrm{T}}A$:

$$AA^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{2} \end{bmatrix} \qquad A^{\mathrm{T}}A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{2} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} \end{bmatrix}.$$

Their determinants are 1, so $\lambda_1\lambda_2 = 1$. Their traces (diagonal sums) are 3:

$$\det\begin{bmatrix} 1-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} = \lambda^2 - 3\lambda + 1 = 0 \quad \text{gives} \quad \lambda_1 = \frac{3+\sqrt{5}}{2} \quad \text{and} \quad \lambda_2 = \frac{3-\sqrt{5}}{2}.$$

The square roots of λ_1 and λ_2 are $\sigma_1 = \dfrac{\sqrt{5}+1}{2}$ and $\sigma_2 = \dfrac{\sqrt{5}-1}{2}$ with $\sigma_1\sigma_2 = 1$.

The nearest rank 1 matrix to A will be $\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$. The error is only $\sigma_2 \approx 0.6 =$ best possible.

The orthonormal eigenvectors of AA^{T} and $A^{\mathrm{T}}A$ are

$$\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ \sigma_1 \end{bmatrix} \quad \boldsymbol{u}_2 = \begin{bmatrix} \sigma_1 \\ -1 \end{bmatrix} \quad \boldsymbol{v}_1 = \begin{bmatrix} \sigma_1 \\ 1 \end{bmatrix} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -\sigma_1 \end{bmatrix} \quad \text{all divided by } \sqrt{1+\sigma_1^2}. \quad (2)$$

Every reader understands that in real life those calculations are done by computers! (Certainly not by unreliable professors. I corrected myself using svd(A) in MATLAB.) And we can check that the matrix A is correctly recovered from $\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$:

$$A = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{bmatrix}\begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix}\begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \end{bmatrix} \quad \text{or more simply} \quad A\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} \sigma_1\boldsymbol{u}_1 & \sigma_2\boldsymbol{u}_2 \end{bmatrix} \quad (3)$$

Important The key point is not that images tend to have low rank. **No**: Images mostly have full rank. But they do have **low effective rank**. This means: Many singular values are small and can be set to zero. *We transmit a low rank approximation.*

Example 4 Suppose the flag has two triangles of different colors. The lower left triangle has 1's and the upper right triangle has 0's. The main diagonal is included with the 1's. Here is the image matrix when $n = 4$. It has full rank $r = 4$ so it is invertible:

Triangular flag matrix $$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad A^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

With full rank, A has a full set of n singular values σ (all positive). The SVD will produce n pieces $\sigma_i\, \boldsymbol{u}_i\, \boldsymbol{v}_i^{\mathrm{T}}$ of rank one. Perfect reproduction needs all n pieces.

In compression *small* σ's can be discarded with no serious loss in image quality. We want to understand and plot the σ's for $n = 4$ and also for large n. Notice that Example 3 was the special case $n = 2$ of this triangular Example 4.

Working by hand, we begin with AA^{T} (a computer would proceed differently):

$$AA^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \quad \text{and} \quad (AA^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}} A^{-1} = \begin{bmatrix} \mathbf{2} & -1 & 0 & 0 \\ -1 & \mathbf{2} & -1 & 0 \\ 0 & -1 & \mathbf{2} & -1 \\ 0 & 0 & -1 & \mathbf{1} \end{bmatrix}. \tag{4}$$

That $-1, 2, -1$ inverse matrix is included because its eigenvalues all have the form $2 - 2\cos\theta$. So we know the λ's for AA^{T} and the σ's for A:

$$\lambda = \frac{1}{2 - 2\cos\theta} = \frac{1}{4\sin^2(\theta/2)} \quad \text{gives} \quad \sigma = \sqrt{\lambda} = \frac{1}{2\sin(\theta/2)}. \tag{5}$$

The n different angles θ are equally spaced, which makes this example so exceptional:

$$\theta = \frac{\pi}{2n+1}, \frac{3\pi}{2n+1}, \ldots, \frac{(2n-1)\,\pi}{2n+1} \qquad \left(n = 4 \text{ includes } \theta = \frac{3\pi}{9} \text{ with } 2\sin\frac{\theta}{2} = 1 \right).$$

That special case gives $\lambda = 1$ as an eigenvalue of AA^{T} when $n = 4$. So $\sigma = \sqrt{\lambda} = 1$ is a singular value of A. You can check that the vector $\boldsymbol{u} = (1, 1, 0, -1)$ has $AA^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{u}$ (a truly special case).

The important point is to graph the n singular values of A. Those numbers drop off (unlike the eigenvalues of A, which are all 1). But the dropoff is not steep. So the SVD gives only moderate compression of this triangular flag. *Great compression for Hilbert.*

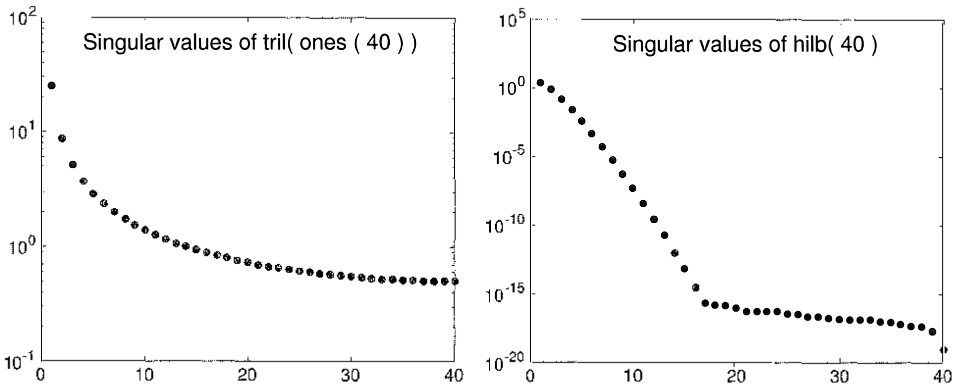


Figure 7.1: Singular values of the triangle of 1's in Examples 3-4 (not compressible) and the evil Hilbert matrix $H(i,j) = (i+j-1)^{-1}$ in Section 8.3: compress it to work with it.

Your faithful author has continued research on the ranks of flags. Quite a few are based on horizontal or vertical stripes. Those have *rank one*—all rows or all columns are multiples of the *ones* vector $(1, 1, \ldots, 1)$. Armenia, Austria, Belgium, Bulgaria, Chad, Colombia, Ireland, Madagascar, Mali, Netherlands, Nigeria, Romania, Russia (and more) have three stripes. Indonesia and Poland have two ! Libya was the extreme case in the Gadaffi years 1977 to 2011 (*the whole flag was green*).

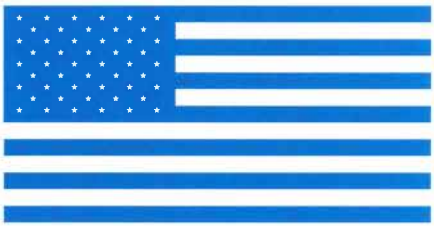
At the other extreme, many flags include diagonal lines. Those could be long diagonals as in the British flag. Or they could be short diagonals coming from the edges of a star—as in the US flag. The text example of a triangle of ones shows how those flag matrices will have large rank. The rank increases to infinity as the pixel sizes get small.

Other flags have circles or crescents or various curved shapes. Their ranks are large and also increasing to infinity. These are still compressible! The compressed image won't be perfect but our eyes won't see the difference (with enough terms $\sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}}$ from the SVD). Those examples actually bring out the main purpose of image compression:

Visual quality can be preserved even with a big reduction in the rank.

For fun I looked back at the flags with finite rank. They can have stripes and they can also have crosses—provided the edges of the cross are horizontal or vertical. Some flags have a thin outline around the cross. This artistic touch will increase the rank. Right now my champion is the flag of Greece shown below, with a cross and also stripes. Its rank is **three** by my counting (three different columns). I see no US State Flags of finite rank !

The reader could google "national flags" to see the variety of designs and colors. I would be glad to know any finite rank examples with rank > 3. Good examples of all kinds will go on the book's website **math.mit.edu/linearalgebra** (and flags in full color).



Problem Set 7.1

1 What are the ranks r for these matrices with entries i times j and i plus j? Write A and B as the sum of r pieces $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ of rank one. Not requiring $\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{u}_2 = \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_2 = \boldsymbol{0}$.

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 9 & 12 \\ 4 & 8 & 12 & 16 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

2 We usually think that the identity matrix I is as simple as possible. But why is I completely incompressible? *Draw a rank* 5 *flag with a cross.*

3 These flags have rank 2. Write A and B in any way as $\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$.

$$\boldsymbol{A}_{\textbf{Sweden}} = \boldsymbol{A}_{\textbf{Finland}} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 1 & 2 & 1 & 1 \end{bmatrix} \qquad \boldsymbol{B}_{\textbf{Benin}} = \begin{bmatrix} 1 & 2 & 2 \\ 1 & 3 & 3 \end{bmatrix}$$

4 Now find the trace and determinant of BB^{T} and $B^{\mathrm{T}}B$ in Problem 3. The singular values of B are close to $\sigma_1^2 = 28 - \frac{1}{14}$ and $\sigma_2^2 = \frac{1}{14}$. Is B compressible or not?

5 Use $[U, S, V] = \mathrm{svd}\,(A)$ to find two orthogonal pieces $\sigma\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ of $\boldsymbol{A}_{\textbf{Sweden}}$.

6 Find the eigenvalues and the singular values of this 2 by 2 matrix A.

$$A = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix} \quad \text{with} \quad A^{\mathrm{T}}A = \begin{bmatrix} 20 & 10 \\ 10 & 5 \end{bmatrix} \quad \text{and} \quad AA^{\mathrm{T}} = \begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix}.$$

The eigenvectors $(1, 2)$ and $(1, -2)$ of A are not orthogonal. How do you know the eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ of $A^{\mathrm{T}}A$ are orthogonal? Notice that $A^{\mathrm{T}}A$ and AA^{T} have the same eigenvalues (25 and 0).

7 How does the second form $\boldsymbol{AV} = \boldsymbol{U\Sigma}$ in equation (3) follow from the first form $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$? That is the most famous form of the SVD.

8 The two columns of $AV = U\Sigma$ are $A\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1$ and $A\boldsymbol{v}_2 = \sigma_2\boldsymbol{u}_2$. So we hope that

$$A\boldsymbol{v}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \sigma_1 \\ 1 \end{bmatrix} = \sigma_1\begin{bmatrix} 1 \\ \sigma_1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -\sigma_1 \end{bmatrix} = \sigma_2\begin{bmatrix} \sigma_1 \\ -1 \end{bmatrix}$$

The first needs $\sigma_1 + 1 = \sigma_1^2$ and the second needs $1 - \sigma_1 = -\sigma_2$. Are those true?

9 The MATLAB commands $A = \mathrm{rand}\,(20, 40)$ and $B = \mathrm{randn}\,(20, 40)$ produce 20 by 40 random matrices. The entries of A are between 0 and 1 with uniform probability. The entries of B have a normal "bell-shaped" probability distribution. Using an svd command, find and graph their singular values σ_1 to σ_{20}. Why do they have 20 σ's?

7.2 Bases and Matrices in the SVD

1 The SVD produces **orthonormal basis** 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 $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$. $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$ is the largest!

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 basis vectors **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 :

$$\text{“}\boldsymbol{A}\textbf{ is diagonalized”} \qquad A\boldsymbol{v}_1 = \sigma_1 \boldsymbol{u}_1 \quad A\boldsymbol{v}_2 = \sigma_2 \boldsymbol{u}_2 \quad \ldots \quad A\boldsymbol{v}_r = \sigma_r \boldsymbol{u}_r \tag{1}$$

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

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

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*** $\boldsymbol{AV=U\Sigma}$.

$$\begin{array}{c}(m \text{ by } n)(n \text{ by } n)\\ \boldsymbol{AV} \text{ equals } \boldsymbol{U\Sigma}\\ (m \text{ by } m)(m \text{ by } n)\end{array} \quad A\begin{bmatrix} v_1 \cdots v_r \cdots v_n \end{bmatrix} = \begin{bmatrix} u_1 \cdots u_r \cdots u_m \end{bmatrix}\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ & & & \end{bmatrix} \qquad (3)$$

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 matrices and $V^{-1}=V^{\mathrm{T}}$. So $AV=U\Sigma$ becomes $\boldsymbol{A=U\Sigma V^{\mathrm{T}}}$. This is the ***Singular Value Decomposition***. I can multiply columns $\boldsymbol{u}_i\sigma_i$ from $U\Sigma$ by rows of V^{T}:

$$\textbf{SVD} \qquad \boldsymbol{A=U\Sigma V^{\mathrm{T}}=u_1\sigma_1v_1^{\mathrm{T}}+\cdots+u_r\sigma_rv_r^{\mathrm{T}}}. \qquad (4)$$

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 is the fourth way to multiply matrices.

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 $A=U\Sigma V^{\mathrm{T}}$ (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 $A=\boldsymbol{xy}^{\mathrm{T}}$ (rank 1) with unit vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, what is the SVD of A?

Solution The reduced SVD in (2) is exactly $\boldsymbol{xy}^{\mathrm{T}}$, 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** $\boldsymbol{A^{\mathrm{T}}A}$. This must be true because we are aiming for

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

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 $Av_i = \sigma_i u_i$ tells us the unit vectors u_1 to u_r. This is the key equation (1). The essential point—the whole reason that the SVD succeeds—is that those unit vectors u_1 to u_r are automatically orthogonal to each other (*because the v's are orthogonal*):

Key step $i \neq j$
$$u_i^{\mathrm{T}} u_j = \left(\frac{Av_i}{\sigma_i}\right)^{\mathrm{T}} \left(\frac{Av_j}{\sigma_j}\right) = \frac{v_i^{\mathrm{T}} A^{\mathrm{T}} A v_j}{\sigma_i \sigma_j} = \frac{\sigma_j^2}{\sigma_i \sigma_j} v_i^{\mathrm{T}} v_j = \textbf{zero}. \quad (6)$$

chosen

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

Finally we complete the v's and u's to n v's and m 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 all 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}:

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

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):

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

Then v_1 and v_2 are those orthogonal eigenvectors rescaled to length 1. Divide by $\sqrt{2}$.

Right singular vectors $v_1 = \dfrac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad v_2 = \dfrac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ **Left singular vectors** $u_i = \dfrac{Av_i}{\sigma_i}$

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

$$\begin{aligned} Av_1 &= \frac{3}{\sqrt{2}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} &= \sqrt{45} \frac{\mathbf{1}}{\sqrt{\mathbf{10}}} \begin{bmatrix} \mathbf{1} \\ \mathbf{3} \end{bmatrix} &= \sigma_1 u_1 \\ Av_2 &= \frac{1}{\sqrt{2}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} &= \sqrt{5} \frac{\mathbf{1}}{\sqrt{\mathbf{10}}} \begin{bmatrix} \mathbf{-3} \\ \mathbf{1} \end{bmatrix} &= \sigma_2 u_2 \end{aligned}$$

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 of A is U times Σ times V^{T}.

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

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

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

An Extreme Matrix

Here is a larger example, when the $\boldsymbol{u}$'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.

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

$A^{\mathrm{T}}A$ and AA^{T} are diagonal (with easy eigenvectors, but in different orders):

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

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. Those eigenvalues are $\sigma^2 = 9, 4, 1$.

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

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 (for this extreme example) exactly from the numbers 3, 2, 1 in A.

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

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 and also the last row and column of U. Then $(3 \times 4) = U\Sigma V^{\mathrm{T}} = (3 \times 3)(3 \times 4)(4 \times 4)$. 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 $\boldsymbol{v}_1$.

Section 7.3 will apply the SVD to finance and genetics and search engines.

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

$$\lambda = 0,0,0,0 \text{ to } \boldsymbol{\lambda} = \frac{\mathbf{1}}{\mathbf{10}}, \frac{\boldsymbol{i}}{\mathbf{10}}, \frac{\mathbf{-1}}{\mathbf{10}}, \frac{\boldsymbol{-i}}{\mathbf{10}}$$

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 of A. The expansions in eigenvectors and singular vectors are perfectly parallel.

Symmetric S	$S = Q\Lambda Q^{\mathrm{T}} = \lambda_1 \boldsymbol{q}_1 \boldsymbol{q}_1^{\mathrm{T}} + \lambda_2 \boldsymbol{q}_2 \boldsymbol{q}_2^{\mathrm{T}} + \cdots + \lambda_r \boldsymbol{q}_r \boldsymbol{q}_r^{\mathrm{T}}$
Any matrix 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 the 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:

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

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

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 $\boldsymbol{x}$:

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

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 same as 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 optional for readers who aim to see the SVD in action (Section 7.3).

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 $S\boldsymbol{q}_1 = \lambda_1 \boldsymbol{q}_1$:

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

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 7.4 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 Singular Values 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 a *symmetric eigenvalue problem* (good). But in the end we don't want to multiply A^{T} times A (squaring is time-consuming: not good).

The first idea is ***to produce zeros in A and S without changing any σ's and λ's***. Singular vectors and eigenvectors will change—no problem. The similar matrix $Q^{-1} S Q$ has the same λ's as S. If Q is orthogonal, this matrix is $Q^{\mathrm{T}} S Q$ and still symmetric.

Section 11.3 will show how to build Q from 2 by 2 rotations so that $Q^{\mathrm{T}} S Q$ is **symmetric and tridiagonal** (many zeros). But rotations can't get all the way to a diagonal matrix. To show all the eigenvalues of S needs a new idea and more work.

For the SVD, what is the parallel to $Q^{\mathrm{T}} S Q$? 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 $Q_1^{\mathrm{T}} A Q_2$. The σ's don't change:

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

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

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. We certainly don't use those polynomials!

The favorite way to find λ's and σ's in **LAPACK** 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 (zeros first) 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 ■

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

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

7.2 B Show that $\sigma_1 \geq |\lambda|_{\max}$. 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*** $||A\boldsymbol{x}||/||\boldsymbol{x}||$ ***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.

Problem Set 7.2

1 Find the eigenvalues of these matrices. Then find singular values from $A^{\mathrm{T}}A$:

$$A = \begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$$

For each A, construct V from the eigenvectors of $A^{\mathrm{T}}A$ and U from the eigenvectors of AA^{T}. Check that $A = U\Sigma V^{\mathrm{T}}$.

2 Find $A^{\mathrm{T}}A$ and V and Σ and $\boldsymbol{u}_i = A\boldsymbol{v}_i/\sigma_i$ and the full SVD:

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = U\Sigma V^{\mathrm{T}}.$$

3 In Problem 2, show that AA^{T} is diagonal. Its eigenvectors $\boldsymbol{u}_1, \boldsymbol{u}_2$ are _____. Its eigenvalues σ_1^2, σ_2^2 are _____. The rows of A are orthogonal but they are not _____. So the columns of A are not orthogonal.

4 Compute $A^{\mathrm{T}}A$ and AA^{T} and their eigenvalues and unit eigenvectors for V and U.

$$\textbf{Rectangular matrix} \qquad A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Check $AV = U\Sigma$ (this decides $\pm$ signs in U). Σ has the same shape as A: 2×3.

5 (a) The row space of $A = \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}$ is 1-dimensional. Find $\boldsymbol{v}_1$ in the row space and $\boldsymbol{u}_1$ in the column space. What is σ_1? Why is there no σ_2?