I.8 Singular Values and Singular Vectors in the SVD

The best matrices (real symmetric matrices S) have real eigenvalues and orthogonal eigenvectors. But for other matrices, the eigenvalues are complex or the eigenvectors are not orthogonal. If A is not square then $A\boldsymbol{x} = \lambda\boldsymbol{x}$ is impossible and eigenvectors fail (left side in $\mathbf{R}^m$, right side in $\mathbf{R}^n$). We need an idea that succeeds for every matrix.

The Singular Value Decomposition fills this gap in a perfect way. In our applications, A is often a matrix of data. The rows could tell us the age and height of 1000 children. Then A is 2 by 1000: definitely rectangular. Unless height is exactly proportional to age, the rank is $r = 2$ and that matrix A has two positive singular values σ_1 and σ_2.

The key point is that we need **two sets of singular vectors**, the $\boldsymbol{u}$'s and the $\boldsymbol{v}$'s. For a real m by n matrix, the n right singular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthogonal in $\mathbf{R}^n$. The m left singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ are perpendicular to each other in $\mathbf{R}^m$. The connection between n $\boldsymbol{v}$'s and m $\boldsymbol{u}$'s is not $A\boldsymbol{x} = \lambda\boldsymbol{x}$. That is for eigenvectors. **For singular vectors, each $A\boldsymbol{v}$ equals $\sigma\boldsymbol{u}$:**

$$\boxed{A\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1 \quad \cdot\cdot \quad A\boldsymbol{v}_r = \sigma_r\boldsymbol{u}_r} \qquad \boxed{A\boldsymbol{v}_{r+1} = \mathbf{0} \quad \cdot\cdot \quad A\boldsymbol{v}_n = \mathbf{0}} \qquad (1)$$

I have separated the first r $\boldsymbol{v}$'s and $\boldsymbol{u}$'s from the rest. That number r is the *rank of* A, the number of independent columns (and rows). Then r is the dimension of the column space and the row space. **We will have r positive singular values in descending order $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$.** The last $n - r$ $\boldsymbol{v}$'s are in the nullspace of A, and the last $m - r$ $\boldsymbol{u}$'s are in the nullspace of A^{T}.

Our first step is to write equation (1) in matrix form. All of the right singular vectors $\boldsymbol{v}_1$ to $\boldsymbol{v}_n$ go in the columns of V. The left singular vectors $\boldsymbol{u}_1$ to $\boldsymbol{u}_m$ go in the columns of U. Those are **square orthogonal matrices** ($V^{\mathrm{T}} = V^{-1}$ and $U^{\mathrm{T}} = U^{-1}$) because their columns are orthogonal unit vectors. Then equation (1) becomes the full SVD, with square matrices V and U:

$$\boxed{\boldsymbol{AV = U\Sigma} \qquad A\begin{bmatrix} \boldsymbol{v}_1 \,..\, \boldsymbol{v}_r \,..\, \boldsymbol{v}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 \,..\, \boldsymbol{u}_r \,..\, \boldsymbol{u}_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \mathbf{0} \\ & & \sigma_r & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right]} \qquad (2)$$

You see $A\boldsymbol{v}_k = \sigma_k\boldsymbol{u}_k$ in the first r columns above. That is the important part of the SVD. It shows the basis of $\boldsymbol{v}$'s for the row space of A and then $\boldsymbol{u}$'s for the column space. After the positive numbers $\sigma_1, \ldots, \sigma_r$ on the main diagonal of Σ, the rest of that matrix is all zero from the nullspaces of A and A^{T}.

The eigenvectors give $AX = X\Lambda$. But $AV = U\Sigma$ needs **two sets of singular vectors.**

Example 1 $\quad AV = U\Sigma$

$$\begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 3\sqrt{5} & \\ & \sqrt{5} \end{bmatrix}$$

The matrix A is not symmetric, so V is different from U. The rank is 2, so there are two singular values $\sigma_1 = 3\sqrt{5}$ and $\sigma_2 = \sqrt{5}$. Their product $3 \cdot 5 = 15$ is the determinant of A (in this respect singular values are like eigenvalues). The columns of V are orthogonal and the columns of U are orthogonal. Those columns are unit vectors after the divisions by $\sqrt{2}$ and $\sqrt{10}$, so **V and U are orthogonal matrices**: $V^{\mathrm{T}} = V^{-1}$ and $U^{\mathrm{T}} = U^{-1}$.

That orthogonality allows us to go from $AV = U\Sigma$ to the usual and famous expression of the SVD: Multiply both sides of $AV = U\Sigma$ by $V^{-1} = V^{\mathrm{T}}$.

$$\boxed{\textbf{The Singular Value Decomposition of } \boldsymbol{A} \textbf{ is} \quad \boldsymbol{A = U\Sigma V^{\mathrm{T}}}.} \tag{3}$$

Then column-row multiplication of $U\Sigma$ times V^{T} separates A into r pieces of rank 1:

Pieces of the SVD
$$\boxed{\boldsymbol{A = U\Sigma V^{\mathrm{T}} = \sigma_1 u_1 v_1^{\mathrm{T}} + \cdots + \sigma_r u_r v_r^{\mathrm{T}}}.} \tag{4}$$

In the 2 by 2 example, the first piece is more important than the second piece because $\sigma_1 = 3\sqrt{5}$ is greater than $\sigma_2 = \sqrt{5}$. To recover A, add the pieces $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}}$:

$$\frac{3\sqrt{5}}{\sqrt{10}\sqrt{2}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} + \frac{\sqrt{5}}{\sqrt{10}\sqrt{2}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix} = \frac{3}{2} \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{3} & \mathbf{3} \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \mathbf{3} & \mathbf{-3} \\ \mathbf{-1} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{3} & \mathbf{0} \\ \mathbf{4} & \mathbf{5} \end{bmatrix}$$

This simplified because $\sqrt{5}/\sqrt{10}\sqrt{2}$ equals $1/2$. Notice that the right singular vectors $(1, 1)$ and $(-1, 1)$ in V are transposed to rows $\boldsymbol{v}_1^{\mathrm{T}}, \boldsymbol{v}_2^{\mathrm{T}}$ of V^{T}. We have not yet explained how V and U and Σ were computed!

The Reduced Form of the SVD

The full form $AV = U\Sigma$ in equation (2) can have a lot of zeros in Σ when the rank of A is small and its nullspace is large. Those zeros contribute nothing to matrix multiplication. The heart of the SVD is in the first r $\boldsymbol{v}$'s and $\boldsymbol{u}$'s and σ's. We can reduce $AV = U\Sigma$ to $AV_r = U_r\Sigma_r$ by removing the parts that are sure to produce zeros. This leaves the **reduced SVD where Σ_r is now square:**

$$\boxed{\boldsymbol{AV_r = U_r\Sigma_r} \qquad A \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r \\ \multicolumn{3}{c}{\text{row space}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r \\ \multicolumn{3}{c}{\text{column space}} \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}} \tag{5}$$

We still have $V_r^{\mathrm{T}} V_r = I_r$ and $U_r^{\mathrm{T}} U_r = I_r$ from those orthogonal unit vectors $\boldsymbol{v}$'s and $\boldsymbol{u}$'s. But when V_r and U_r are not square, we can no longer have two-sided inverses: $V_r V_r^{\mathrm{T}} \neq I$ and $U_r U_r^{\mathrm{T}} \neq I$.

Example $V_r = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}$ and $V_r^{\mathrm{T}} V_r = \begin{bmatrix} 1 \end{bmatrix}$ but $V_r V_r^{\mathrm{T}} = \dfrac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} =$ rank 1.

Problem 21 shows that **we still have** $\boldsymbol{A = U_r \Sigma_r V_r^{\mathrm{T}}}$. The rest of $U\Sigma V^{\mathrm{T}}$ contributes nothing to A, because of those blocks of zeros in Σ. The key formula is still $A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$. The SVD sees only the r nonzeros in the diagonal matrix Σ.

The Important Fact for Data Science

Why is the SVD so important for this subject and this book? Like the other factorizations $A = LU$ and $A = QR$ and $S = Q\Lambda Q^{\mathrm{T}}$, it separates the matrix into rank one pieces. A special property of the SVD is that **those pieces come in order of importance**. The first piece $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$ is the closest rank one matrix to A. More than that is true: *The sum of the first k pieces is best possible for rank k.*

$\boldsymbol{A_k = \sigma_1 u_1 v_1^{\mathrm{T}} + \cdots + \sigma_k u_k v_k^{\mathrm{T}}}$ **is the best rank $\boldsymbol{k}$ approximation to $\boldsymbol{A}$:**

Eckart-Young **If $\boldsymbol{B}$ has rank $\boldsymbol{k}$ then $\boldsymbol{||A - A_k|| \leq ||A - B||}$.** (6)

To interpret that statement you need to know the meaning of the symbol $||A - B||$. This is the "**norm**" of the matrix $A - B$, a measure of its size (like the absolute value of a number). The Eckart-Young theorem is proved in Section I.9.

Our first job is to find the $\boldsymbol{v}$'s and $\boldsymbol{u}$'s for equation (1), to reach the SVD.

First Proof of the SVD

Our goal is $A = U\Sigma V^{\mathrm{T}}$. We want to identify the two sets of singular vectors, the $\boldsymbol{u}$'s and the $\boldsymbol{v}$'s. One way to find those vectors is to form the symmetric matrices $A^{\mathrm{T}}A$ and AA^{T}:

$$\boldsymbol{A^{\mathrm{T}}A} = (V\Sigma^{\mathrm{T}}U^{\mathrm{T}})\ (U\Sigma V^{\mathrm{T}}) = \boldsymbol{V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}}} \tag{7}$$

$$\boldsymbol{AA^{\mathrm{T}}} = (U\Sigma V^{\mathrm{T}})\ (V\Sigma^{\mathrm{T}}U^{\mathrm{T}}) = \boldsymbol{U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}}} \tag{8}$$

Both (7) and (8) produced symmetric matrices. Usually $A^{\mathrm{T}}A$ and AA^{T} are different. Both right hand sides have the special form $Q\Lambda Q^{\mathrm{T}}$. Eigenvalues are in $\Lambda = \Sigma^{\mathrm{T}}\Sigma$ or $\Sigma\Sigma^{\mathrm{T}}$. **Eigenvectors are in $\boldsymbol{Q = V}$ or $\boldsymbol{Q = U}$.** So we know from (7) and (8) how V and U and Σ connect to the symmetric matrices $A^{\mathrm{T}}A$ and AA^{T}.

V contains orthonormal eigenvectors of $A^{\mathrm{T}}A$
U contains orthonormal eigenvectors of AA^{T}
σ_1^2 to σ_r^2 are the nonzero eigenvalues of both $A^{\mathrm{T}}A$ and AA^{T}

We are not quite finished, for this reason. **The SVD requires that $A\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k$.** It connects each right singular vector $\boldsymbol{v}_k$ to a left singular vector $\boldsymbol{u}_k$, for $k = 1, \ldots, r$. When I choose the $\boldsymbol{v}$'s, that choice will decide the signs of the $\boldsymbol{u}$'s. If $S\boldsymbol{u} = \lambda\boldsymbol{u}$ then also $S(-\boldsymbol{u}) = \lambda(-\boldsymbol{u})$ and I have to know the correct sign. More than that, there is a whole plane of eigenvectors when λ is a double eigenvalue. When I choose two $\boldsymbol{v}$'s in that plane, then $A\boldsymbol{v} = \sigma\boldsymbol{u}$ will tell me both $\boldsymbol{u}$'s. This is in equation (9).

The plan is to start with the $\boldsymbol{v}$'s. **Choose orthonormal eigenvectors** $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ of $A^{\mathrm{T}}A$. **Then choose $\sigma_k = \sqrt{\lambda_k}$. To determine the $\boldsymbol{u}$'s we require $A\boldsymbol{v} = \sigma\boldsymbol{u}$:**

$$\boxed{\boldsymbol{v}\text{'s then }\boldsymbol{u}\text{'s} \qquad A^{\mathrm{T}}A\boldsymbol{v}_k = \sigma_k^2\boldsymbol{v}_k \quad \text{and then} \quad \boldsymbol{u}_k = \frac{A\boldsymbol{v}_k}{\sigma_k} \quad \text{for} \quad k = 1, \ldots, r} \tag{9}$$

This is the proof of the SVD ! Let me check that those $\boldsymbol{u}$'s are eigenvectors of AA^{T} :

$$AA^{\mathrm{T}}\boldsymbol{u}_k = AA^{\mathrm{T}}\left(\frac{A\boldsymbol{v}_k}{\sigma_k}\right) = A\left(\frac{A^{\mathrm{T}}A\,\boldsymbol{v}_k}{\sigma_k}\right) = A\,\frac{\sigma_k^2\boldsymbol{v}_k}{\sigma_k} = \sigma_k^2\boldsymbol{u}_k \tag{10}$$

The $\boldsymbol{v}$'s were chosen to be orthonormal. I must check that the $\boldsymbol{u}$'s are also orthonormal:

$$\boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{u}_k = \left(\frac{A\boldsymbol{v}_j}{\sigma_j}\right)^{\mathrm{T}}\left(\frac{A\boldsymbol{v}_k}{\sigma_k}\right) = \frac{\boldsymbol{v}_j^{\mathrm{T}}(A^{\mathrm{T}}A\boldsymbol{v}_k)}{\sigma_j\,\sigma_k} = \frac{\sigma_k}{\sigma_j}\,\boldsymbol{v}_j^{\mathrm{T}}\,\boldsymbol{v}_k = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \tag{11}$$

Notice that $(AA^{\mathrm{T}})A = A(A^{\mathrm{T}}A)$ was the key to equation (10). The law $(AB)C = A(BC)$ is the key to a great many proofs in linear algebra. Moving the parentheses is a powerful idea. This is the *associative law*.

Finally we have to choose the last $n - r$ vectors $\boldsymbol{v}_{r+1}$ to $\boldsymbol{v}_n$ and the last $m - r$ vectors $\boldsymbol{u}_{r+1}$ to $\boldsymbol{u}_m$. This is easy. **These $\boldsymbol{v}$'s and $\boldsymbol{u}$'s are in the nullspaces of A and A^{T}.** We can choose any orthonormal bases for those nullspaces. They will automatically be orthogonal to the first $\boldsymbol{v}$'s in the row space of A and the first $\boldsymbol{u}$'s in the column space. This is because the whole spaces are orthogonal: $\mathbf{N}(A) \perp \mathbf{C}(A^{\mathrm{T}})$ and $\mathbf{N}(A^{\mathrm{T}}) \perp \mathbf{C}(A)$. *The proof of the SVD is complete.*

Now we have U and V and Σ in the full size SVD of equation (1). You may have noticed that the eigenvalues of $A^{\mathrm{T}}A$ are in $\Sigma^{\mathrm{T}}\Sigma$, and *the same numbers σ_1^2 to σ_r^2 are also eigenvalues of AA^{T} in $\Sigma\Sigma^{\mathrm{T}}$*. An amazing fact: **$BA$ always has the same nonzero eigenvalues as AB : 5 pages ahead.**

Example 1 (completed) Find the matrices U, Σ, V for $A = \begin{bmatrix} \mathbf{3} & \mathbf{0} \\ \mathbf{4} & \mathbf{5} \end{bmatrix}$.

With rank 2, this A has two positive singular values σ_1 and σ_2. We will see that σ_1 is larger than $\lambda_{\text{max}} = 5$, and σ_2 is smaller than $\lambda_{\text{min}} = 3$. 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):

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

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

Right singular vectors $\boldsymbol{v}_1 = \dfrac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\boldsymbol{v}_2 = \dfrac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ **Left singular vectors** $\boldsymbol{u}_i = \dfrac{A\boldsymbol{v}_i}{\sigma_i}$

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 $\sigma_2\boldsymbol{u}_2 = \sqrt{5}\,\boldsymbol{u}_2$:

$$A\boldsymbol{v}_1 = \frac{3}{\sqrt{2}}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sqrt{45}\frac{\mathbf{1}}{\sqrt{\mathbf{10}}}\begin{bmatrix} \mathbf{1} \\ \mathbf{3} \end{bmatrix} = \sigma_1\,\boldsymbol{u}_1$$

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

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 of A is U times Σ times V^{T}.

$$\boldsymbol{U} = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{45} & \\ & \sqrt{5} \end{bmatrix} \qquad \boldsymbol{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \tag{12}$$

U and V contain orthonormal bases for the column space and the row space of A (both spaces are just $\mathbf{R}^2$). The real achievement is that those two bases diagonalize A: AV equals $U\Sigma$. The matrix $A = U\Sigma V^{\mathrm{T}}$ splits into two rank-one matrices, columns times rows, with $\sqrt{2}\,\sqrt{10} = \sqrt{20}$.

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

Every matrix is a sum of rank one matrices with orthogonal $\boldsymbol{u}$'s and orthogonal $\boldsymbol{v}$'s.

Question: If $S = Q\Lambda Q^T$ is symmetric positive definite, what is its SVD?

Answer: The SVD is exactly $U\Sigma V^T = Q\Lambda Q^T$. The matrix $U = V = Q$ is orthogonal. And the eigenvalue matrix Λ becomes the singular value matrix Σ.

Question: If $S = Q\Lambda Q^T$ has a negative eigenvalue ($S\boldsymbol{x} = -\alpha\,\boldsymbol{x}$), what is the singular value and what are the vectors $\boldsymbol{v}$ and $\boldsymbol{u}$?

Answer: The singular value will be $\boldsymbol{\sigma} = +\boldsymbol{\alpha}$ (positive). One singular vector (either $\boldsymbol{u}$ or $\boldsymbol{v}$) must be $-\boldsymbol{x}$ (reverse the sign). Then $S\boldsymbol{x} = -\alpha\,\boldsymbol{x}$ is the same as $S\boldsymbol{v} = \sigma\boldsymbol{u}$. The two sign changes cancel.

Question: If $A = Q$ is an orthogonal matrix, why does every singular value equal 1?

Answer: All singular values are $\sigma = 1$ because $A^TA = Q^TQ = I$. Then $\Sigma = I$. But $U = Q$ and $V = I$ is only one choice for the singular vectors $\boldsymbol{u}$ and $\boldsymbol{v}$:

$$Q = U\Sigma V^T \text{ can be } Q = QII^T \text{ or any } Q = (QQ_1)IQ_1^T.$$

Question: **Why are all eigenvalues of a square matrix A less than or equal to σ_1?**

Answer: Multiplying by orthogonal matrices U and V^T does not change vector lengths:

$$||\boldsymbol{A}\boldsymbol{x}|| = ||U\Sigma V^T\boldsymbol{x}|| = ||\Sigma V^T\boldsymbol{x}|| \le \sigma_1||V^T\boldsymbol{x}|| = \boldsymbol{\sigma_1}||\boldsymbol{x}|| \text{ for all } \boldsymbol{x}. \tag{13}$$

An eigenvector has $||A\boldsymbol{x}|| = |\lambda|\,||\boldsymbol{x}||$. Then (13) gives $|\lambda|\,||\boldsymbol{x}|| \le \sigma_1\,||\boldsymbol{x}||$ and $|\boldsymbol{\lambda}| \le \boldsymbol{\sigma_1}$.

Question: If $A = \boldsymbol{x}\boldsymbol{y}^T$ has rank 1, what are $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ and σ_1? **Check that $|\boldsymbol{\lambda_1}| \le \boldsymbol{\sigma_1}$.**

Answer: The singular vectors $\boldsymbol{u}_1 = \boldsymbol{x}/||\boldsymbol{x}||$ and $\boldsymbol{v}_1 = \boldsymbol{y}/||\boldsymbol{y}||$ have length 1. Then $\boldsymbol{\sigma}_1 = ||\boldsymbol{x}||\,||\boldsymbol{y}||$ is the only nonzero number in the singular value matrix Σ. Here is the SVD:

$$\textbf{Rank 1 matrix} \qquad \boldsymbol{x}\boldsymbol{y}^T = \frac{\boldsymbol{x}}{||\boldsymbol{x}||}\left(||\boldsymbol{x}||\,||\boldsymbol{y}||\right)\frac{\boldsymbol{y}^T}{||\boldsymbol{y}||} = \boldsymbol{u}_1\sigma_1\boldsymbol{v}_1^T.$$

Observation The only nonzero eigenvalue of $A = \boldsymbol{x}\boldsymbol{y}^T$ is $\lambda = \boldsymbol{y}^T\boldsymbol{x}$. The eigenvector is $\boldsymbol{x}$ because $(\boldsymbol{x}\boldsymbol{y}^T)\,\boldsymbol{x} = \boldsymbol{x}\,(\boldsymbol{y}^T\boldsymbol{x}) = \lambda\boldsymbol{x}$. Then $|\lambda_1| = |\boldsymbol{y}^T\boldsymbol{x}| \le \sigma_1 = ||\boldsymbol{y}||\,||\boldsymbol{x}||$.

The key inequality $|\lambda_1| \le \sigma_1$ becomes exactly the Schwarz inequality.

Question: **What is the Karhunen-Loève transform and its connection to the SVD?**

Answer: KL begins with a covariance matrix V of a zero-mean random process. V is symmetric and positive definite or semidefinite. In general V could be an infinite matrix or a covariance function. Then the KL expansion will be an infinite series.

The eigenvectors of V, in order of decreasing eigenvalues $\sigma_1^2 \ge \sigma_2^2 \ge \ldots \ge 0$, are the basis functions $\boldsymbol{u}_i$ for the KL transform. The expansion of any vector $\boldsymbol{v}$ in an orthonormal basis $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots$ is $\boldsymbol{v} = \sum(\boldsymbol{u}_i^T\boldsymbol{v})\boldsymbol{u}_i$.

In this stochastic case, that transform decorrelates the random process: the $\boldsymbol{u}_i$ are independent. More than that, the ordering of the eigenvalues means that the first k terms, stopping at $(\boldsymbol{u}_k^T\boldsymbol{v})\boldsymbol{u}_k$, minimize the expected square error. This fact corresponds to the Eckart-Young Theorem in the next section I.9.

The KL transform is a stochastic (random) form of Principal Component Analysis.

The Geometry of the SVD

The SVD separates a matrix into $A = U\Sigma V^{\mathrm{T}}$: **(orthogonal)** $\times$ **(diagonal)** $\times$ **(orthogonal)**. In two dimensions we can draw those steps. The orthogonal matrices U and V rotate the plane. The diagonal matrix Σ stretches it along the axes. Figure I.11 shows **rotation** times **stretching** times **rotation**. **Vectors $\boldsymbol{x}$ on the unit circle go to $A\boldsymbol{x}$ on an ellipse.**

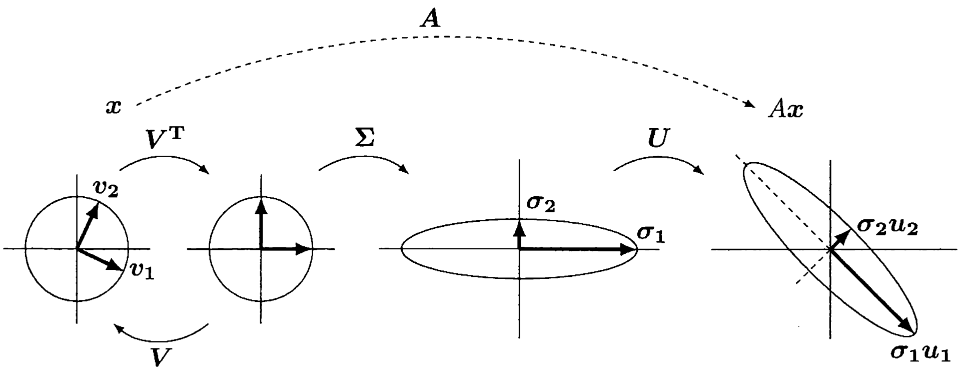

Figure I.10: U and V are rotations and possible reflections. Σ stretches circle to ellipse.

This picture applies to a 2 by 2 invertible matrix (because $\sigma_1 > 0$ and $\sigma_2 > 0$). First is a rotation of any $\boldsymbol{x}$ to $V^{\mathrm{T}}\boldsymbol{x}$. Then Σ stretches that vector to $\Sigma V^{\mathrm{T}}\boldsymbol{x}$. Then U rotates to $A\boldsymbol{x} = U\Sigma V^{\mathrm{T}}\boldsymbol{x}$. We kept all determinants positive to avoid reflections. The four numbers a, b, c, d in the matrix connect to *two angles* θ and ϕ and *two numbers* σ_1 and σ_2.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}. \qquad (14)$$

Question. If the matrix is symmetric then $b = c$ and A has only 3 (not 4) parameters. *How do the 4 numbers θ, ϕ, σ_1, σ_2 reduce to 3 numbers for a symmetric matrix S?*

The First Singular Vector $\boldsymbol{v}_1$

The next page will establish a new way to look at $\boldsymbol{v}_1$. The previous pages chose the $\boldsymbol{v}$'s as eigenvectors of $A^{\mathrm{T}}A$. Certainly that remains true. But there is a valuable way to understand these singular vectors **one at a time instead of all at once**. We start with $\boldsymbol{v}_1$ and the singular value σ_1.

$$\boxed{\textbf{Maximize the ratio } \frac{||A\boldsymbol{x}||}{||\boldsymbol{x}||}. \textbf{ The maximum is } \sigma_1 \textbf{ at the vector } \boldsymbol{x} = \boldsymbol{v}_1.} \qquad (15)$$

The ellipse in Figure I.10 showed why the maximizing $\boldsymbol{x}$ is $\boldsymbol{v}_1$. When you follow $\boldsymbol{v}_1$ across the page, it ends at $A\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1$ (the longest axis of the ellipse). Its length started at $||\boldsymbol{v}_1|| = 1$ and ended at $||A\boldsymbol{v}_1|| = \sigma_1$.

But we aim for an independent approach to the SVD ! We are not assuming that we already know U or Σ or V. How do we recognize that the ratio $||A\boldsymbol{x}||/||\boldsymbol{x}||$ is a maximum when $\boldsymbol{x} = \boldsymbol{v}_1$? Calculus tells us that the first derivatives must be zero. The derivatives will be easier if we square our function:

$$\textbf{Problem: Find the maximum value } \boldsymbol{\lambda} \textbf{ of } \frac{||A\boldsymbol{x}||^2}{||\boldsymbol{x}||^2} = \frac{\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \frac{\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}. \tag{16}$$

This "Rayleigh quotient" depends on $x_1, \ldots, x_n$. Calculus uses the quotient rule, so we need

$$\frac{\partial}{\partial x_i}\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right) = \frac{\partial}{\partial x_i}\left(x_1^2 + \cdots + x_i^2 + \cdots + x_n^2\right) = 2(\boldsymbol{x})_i \tag{17}$$

$$\frac{\partial}{\partial x_i}\left(\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}\right) = \frac{\partial}{\partial x_i}\left(\sum_i\sum_j S_{ij}x_ix_j\right) = 2\sum_j S_{ij}x_j = 2\left(S\boldsymbol{x}\right)_i \tag{18}$$

The quotient rule finds $\partial/\partial x_i\,(\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}/\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})$. Set those n partial derivatives of (16) to zero:

$$\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)2\left(S\boldsymbol{x}\right)_i - \left(\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}\right)2\left(\boldsymbol{x}\right)_i = \mathbf{0} \text{ for } i = 1, \ldots, n \tag{19}$$

Equation (19) says that the best x is an eigenvector of $S = A^{\mathrm{T}}A$!

$$2S\boldsymbol{x} = 2\lambda\boldsymbol{x} \textbf{ and the maximum value of } \frac{\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \frac{||A\boldsymbol{x}||^2}{||\boldsymbol{x}||^2} \textbf{ is an eigenvalue } \lambda \textbf{ of } S.$$

The search is narrowed to eigenvectors of $S = A^{\mathrm{T}}A$. The eigenvector that maximizes is $\boldsymbol{x} = \boldsymbol{v}_1$. The eigenvalue is $\lambda_1 = \sigma_1^2$. Calculus has confirmed the solution (15) of the maximum problem—the first piece of the SVD.

For the full SVD, we need *all* the singular vectors and singular values. To find $\boldsymbol{v}_2$ and σ_2, we adjust the maximum problem so it looks only at vectors $\boldsymbol{x}$ orthogonal to $\boldsymbol{v}_1$.

$$\textbf{Maximize } \frac{||A\boldsymbol{x}||}{||\boldsymbol{x}||} \textbf{ under the condition } \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{x} = \mathbf{0}\textbf{. The maximum is } \boldsymbol{\sigma_2} \textbf{ at } \boldsymbol{x} = \boldsymbol{v}_2.$$

"Lagrange multipliers" were invented to deal with constraints on $\boldsymbol{x}$ like $\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{x} = 0$. And Problem 3 gives a simple direct way to work with this condition $\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{x} = 0$.

In the same way, every singular vector $\boldsymbol{v}_{k+1}$ gives the maximum ratio over all vectors $\boldsymbol{x}$ that are perpendicular to the first $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$. The left singular vectors would come from maximizing $||A^{\mathrm{T}}\boldsymbol{y}||/||\boldsymbol{y}||$. We are always finding the axes of an ellipsoid and the eigenvectors of symmetric matrices $A^{\mathrm{T}}A$ or AA^{T}.

The Singular Vectors of A^{T}

The SVD connects $\boldsymbol{v}$'s in the row space to $\boldsymbol{u}$'s in the column space. When we transpose $A = U\Sigma V^{\mathrm{T}}$, we see that $\boldsymbol{A^{\mathrm{T}} = V\Sigma^{\mathrm{T}}U^{\mathrm{T}}}$ goes the opposite way, from $\boldsymbol{u}$'s to $\boldsymbol{v}$'s :

$$\boldsymbol{A^{\mathrm{T}}u_k = \sigma_k v_k} \textbf{ for } \boldsymbol{k = 1, \ldots, r} \qquad \boldsymbol{A^{\mathrm{T}}u_k = 0} \textbf{ for } \boldsymbol{k = r+1, \ldots, m} \tag{20}$$

Multiply $A\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k$ by A^{T}. Remember $A^{\mathrm{T}}A\boldsymbol{v}_k = \sigma_k^2 \boldsymbol{v}_k$ in equation (9). Divide by σ_k.

A Different Symmetric Matrix Also Produces the SVD

We created the SVD from two symmetric matrices $A^{\mathrm{T}}A$ and AA^{T}. Another good way uses one symmetric block matrix S. *This matrix has r pairs of plus and minus eigenvalues.* The nonzero eigenvalues of this matrix S are σ_k and $-\sigma_k$, and its size is $m+n$:

$$\boldsymbol{S} = \begin{bmatrix} \mathbf{0} & \boldsymbol{A} \\ \boldsymbol{A^{\mathrm{T}}} & \mathbf{0} \end{bmatrix} \quad \text{has eigenvectors} \quad \begin{bmatrix} \boldsymbol{u}_k \\ \boldsymbol{v}_k \end{bmatrix} \text{ and } \begin{bmatrix} -\boldsymbol{u}_k \\ \boldsymbol{v}_k \end{bmatrix}.$$

We can check those eigenvectors directly, remembering $A\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k$ and $A^{\mathrm{T}}\boldsymbol{u}_k = \sigma_k \boldsymbol{v}_k$:

$$\begin{bmatrix} 0 & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \pm\boldsymbol{u}_k \\ \boldsymbol{v}_k \end{bmatrix} = \begin{bmatrix} A\boldsymbol{v}_k \\ \pm A^{\mathrm{T}}\boldsymbol{u}_k \end{bmatrix} = \sigma_k \begin{bmatrix} \boldsymbol{u}_k \\ \boldsymbol{v}_k \end{bmatrix} \text{ and } -\sigma_k \begin{bmatrix} -\boldsymbol{u}_k \\ \boldsymbol{v}_k \end{bmatrix}. \tag{21}$$

That gives $2r$ eigenvalues. The eigenvectors are orthogonal: $-\boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{u}_k + \boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{v}_k = -1+1$. Can you see the other $(m-r)+(n-r)$ eigenvectors with $\lambda = 0$ for that block matrix? They must involve the remaining $\boldsymbol{u}$'s and $\boldsymbol{v}$'s in the nullspaces of A^{T} and A.

AB and BA : Equal Nonzero Eigenvalues

If A is m by n and B is n by m, then AB and BA have the same nonzero eigenvalues.

Start with $AB\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\lambda \neq 0$. Multiply both sides by B, to get $BAB\boldsymbol{x} = \lambda B\boldsymbol{x}$. This says that $B\boldsymbol{x}$ is an eigenvector of BA with the same eigenvalue λ—exactly what we wanted. We needed $\lambda \neq 0$ to be sure that this eigenvector $B\boldsymbol{x}$ is not zero.

Notice that if B is square and invertible, then $B^{-1}(BA)B = AB$. This says that BA is *similar* to AB : *same eigenvalues.* But our first proof allows A and B to be m by n and n by m. This covers the important example of the SVD when $B = A^{\mathrm{T}}$. In that case $A^{\mathrm{T}}A$ and AA^{T} both lead to the singular values of A.

If $m > n$, then AB has $m-n$ extra zero eigenvalues compared to BA.