This document is a **transcript** of the lecture, with extra summary and vocabulary sections for your convenience. Due to time constraints, the notes may sometimes only contain limited illustrations, proofs, and examples; for a more thorough discussion of the course content, **consult the textbook**.

Summary

Quick summary of today's notes. Lecture starts on next page.

- The singular values of a symmetric matrix are the absolute values of its eigenvalues.
- Order the eigenvalues of $A = A^\top$ such that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| \geq 0$. Let

  $$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

  We know from previous lectures that there is an orthogonal $n \times n$ matrix U such that $A = UDU^\top$.

  Define $\epsilon_i = 1$ if $\lambda_i \geq 0$ and let $\epsilon_i = -1$ of $\lambda_i < 0$. Then let

  $$E = \begin{bmatrix} \epsilon_1 & & & \\ & \epsilon_2 & & \\ & & \ddots & \\ & & & \epsilon_n \end{bmatrix}.$$

  A singular value decomposition for $A = A^\top = UDU^\top$ is $A = U\Sigma V^\top$ where $\Sigma = DE$ and $V = UE$.
- Every 2×2 orthogonal matrix is a rotation matrix times a permutation matrix.
- The image of the unit disc

  $$\mathcal{D} = \{v \in \mathbb{R}^2 : v \bullet v \leq 1\}$$

  under any linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ is an ellipse.
- Suppose $A = U\Sigma V^\top$ is an SVD for a 2×2 matrix.

  Suppose the ellipse $\mathcal{E} = \{Av : v \in \mathcal{D}\}$ has radii of lengths $\sigma_1 \geq \sigma_2 \geq 0$. Then $\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$.

  The columns of $V = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ are two orthogonal radii of the unit disc $\mathcal{D}$.

  These vectors have the property that Av_i is a radius of $\mathcal{E}$ with length r_i for each $i = 1, 2$.

  The matrix U is always an orthogonal matrix whose inverse tranforms the ellipse $\mathcal{E}$ back to a standard ellipse (whose radii belong to the x- and y-axes).

1 Last time: definition of singular value decomposition

Let A be an $m \times n$ matrix.

Then $A^\top A$ is a symmetric $n \times n$ matrix, whose eigenvalues are all nonnegative real numbers.

If λ is an eigenvalue of $A^\top A$ and $v \in \mathbb{R}^n$ is a unit vector with $A^\top A v = \lambda v$, then $\lambda = \|Av\|^2$.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of $A^\top A$ arranged in decreasing order.

Define $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, 2, \dots, n$.

The nonnegative real numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ are the *singular values* of A.

Remember that a matrix U is *orthogonal* if U is invertible and $U^{-1} = U^\top$.

Theorem. Suppose $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n$.

Then $\operatorname{rank} A = r$ and we can write $A = U \Sigma V^\top$ where

U is some $m \times m$ orthogonal matrix.

V is some $n \times n$ orthogonal matrix.

Σ is the $m \times n$ matrix with σ_i in position (i, i) for $i = 1, 2, \dots, r$ and zeros in all other positions.

The factorization $A = U \Sigma V^\top$ is called a *singular value decomposition* or *SVD* of A.

The columns of U are called *left singular vectors* of A.

The columns of V are called *right singular vectors* of A.

A matrix A may have more than one SVD, but the middle matrix Σ will be the same in all of these.

A *pseudo-inverse* of an $m \times n$ matrix A is an $n \times m$ matrix A^+ such that

$$AA^+A = A \qquad \text{and} \qquad A^+AA^+ = A^+.$$

If A is a square, invertible matrix, then $A^+ = A^{-1}$ is the pseudo-inverse of A.

If $A = U \Sigma V^\top$ is a singular value decomposition, and Σ^+ is the matrix formed by transposing Σ and then replacing all of its nonzero entries by their reciprocals, then $A^+ = V \Sigma^+ U^\top$ is a pseudo-inverse for A.

To find a singular value decomposition for an $m \times n$ matrix A, do the following steps:

1. Find the nonnegative eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ of $A^\top A$.

   Find a basis of eigenvectors for $A^\top A$ for each eigenspace.

   Convert each basis to an orthonormal basis using the Gram-Schmidt process.

   Combine these orthonormal bases to get an orthonormal list of eigenvectors $v_1, v_2, \dots, v_n$.

2. Let $V = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$.

   Form Σ as the $m \times n$ matrix with $\sigma_i = \sqrt{\lambda_i}$ in position (i, i) for $i = 1, 2, \dots, m$ and zeros elsewhere.

3. Let $u_i = \frac{1}{\sigma_i} A v_i$ for $i = 1, 2, \dots, r$ where $r = \operatorname{rank} A$ is maximal such that $\sigma_r \neq 0$.

   Find vectors $u_{r+1}, u_{r+2}, \dots, u_m \in \mathbb{R}^m$ such that $u_1, u_2, \dots, u_m$ are orthonormal.

   This can be done by finding the pivot columns of $\begin{bmatrix} u_1 & u_2 & \dots & u_r & e_1 & e_2 & \dots & e_m \end{bmatrix}$ and then applying the Gram-Schmidt process. Finally let $U = \begin{bmatrix} u_1 & u_2 & \dots & u_m \end{bmatrix}$.

2 SVDs for symmetric matrices

When we first introduced singular value decompositions we said that they generalized the notion of "orthogonal diagonalization" for symmetric matrices. Let's briefly explain how SVDs can be seen as a generalization of the decomposition $A = UDU^\top = UDU^{-1}$ that exists for a symmetric matrix.

Suppose $A = A^\top$ is an $n \times n$ symmetric matrix.

We know there are real numbers $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$ such that

$$\det(A - xI) = (\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x).$$

These are the eigenvalues of A. Some of these numbers could be negative.

Suppose the eigenvalues are ordered such that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| \geq 0$. Let

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

We know from previous lectures that there is an orthogonal $n \times n$ matrix U such that $A = UDU^\top$.

Define $\epsilon_i = 1$ if $\lambda_i \geq 0$ and let $\epsilon_i = -1$ of $\lambda_i < 0$. Then let

$$E = \begin{bmatrix} \epsilon_1 & & & \\ & \epsilon_2 & & \\ & & \ddots & \\ & & & \epsilon_n \end{bmatrix}.$$

Proposition. A singular value decomposition for the symmetric matrix $A = A^\top = UDU^\top$ is

$$A = U\Sigma V^\top$$

where $\Sigma = DE$ and $V = UE$. The singular values of A are $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$.

In general, the singular values of any symmetric matrix are just the absolute values of its eigenvalues.

Proof. We have $E = E^\top = E^{-1}$ so $U\Sigma V^\top = UDE(UE)^\top = UDEE^\top U^\top = UDU^\top = A$.

Thus E is orthogonal, so $V = UE$ is orthogonal since product of orthogonal matrices are orthogonal.

To show that $A = U\Sigma V^\top$ is a singular value decomposition, we just need to check that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$$

are the singular values of A.

For this, it is enough to show that $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_n^2$ are the eigenvalues of $A^\top A = A^2$.

The follows since $A^2 = (UDU^\top)^2 = (UDU^{-1})^2 = UD^2U^{-1}$ is similar to D^2.

Now recall that similar matrices have the same eigenvalues; the eigenvalues of a diagonal matrix are its diagonal entries; and the numbers $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_n^2$ are the diagonal entries of D^2. $\square$

3 SVDs for 2×2 matrices

To get some physical intuition for what an SVD means, let's consider SVDs for 2×2 matrices.

It is possible describe all 2×2 orthogonal matrices in a simple way:

Proposition. Every orthogonal 2×2 matrix is a rotation matrix times a permutation matrix.

Specifically, every 2×2 orthogonal matrix has the form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} = \begin{bmatrix} \cos(\theta - \frac{\pi}{2}) & -\sin(\theta - \frac{\pi}{2}) \\ \sin(\theta - \frac{\pi}{2}) & \cos(\theta - \frac{\pi}{2}) \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

for an angle $0 \leq \theta < 2\pi$.

Proof. Since $(\cos\theta)^2 + (\sin\theta)^2 = 1$ the given matrices are orthogonal.

Now suppose $U = \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix}$ is orthogonal, so that $u_1^2 + u_2^2 = 1$.

Every point (u_1, u_2) on the unit circle has the form $(\cos\theta, \sin\theta)$ for some angle θ.

For the columns of U to be orthogonal, $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ must be a scalar multiple of $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$.

Since both columns are unit vectors we must have $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ or $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \sin\theta \\ -\cos\theta \end{bmatrix}$. $\square$

Suppose U is a 2×2 orthogonal matrix. The columns of U are two perpendicular radii of the unit circle.

If the second column is 90 degrees counterclockwise from the first column, then

$$\det U = 1 \qquad \text{and} \qquad U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

for some angle θ. Otherwise, the second column must be 90 degrees clockwise from the first column, so

$$\det U = -1 \qquad \text{and} \qquad U = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

for some angle θ.

We can also describe the effect of the mapping $v \mapsto Uv$ for $v \in \mathbb{R}^2$ as follows:

- If $\det U = 1$ then v is rotated counter-clockwise by some angle.
- If $\det U = -1$ then v is reflected across $y = x$ and then rotated counter-clockwise by some angle.

In both cases the angle of rotation depends on U but not on v.

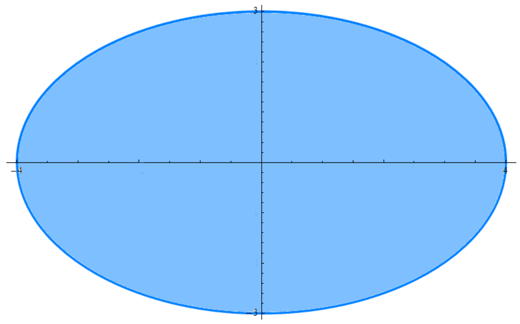
The *unit disc* $\mathcal{D}$ is the set of vectors $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$ with $v_1^2 + v_2^2 \leq 1$.

Fix real numbers $r_1, r_2 \geq 0$.

Consider the set $\mathcal{E}$ of vectors $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$ with $(v_1/r_1)^2 + (v_2/r_2)^2 \leq 1$.

When $r_i = 0$ we consider $(v_i/r_i)^2$ to be zero if $v_i = 0$ and $+\infty$ if $v_i \neq 0$.

We call $\mathcal{E}$ a *(solid) standard ellipse*.



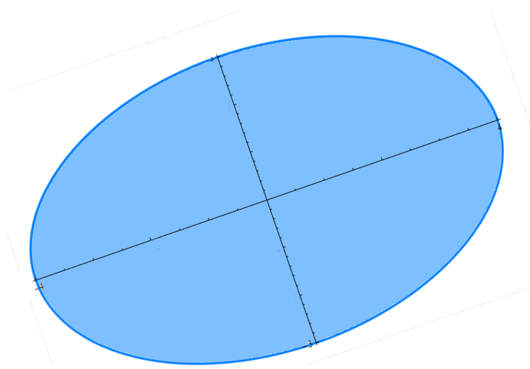
Proposition. It holds that $\mathcal{E} = \left\{ \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} v : v \in \mathcal{D} \right\}$.

Proof. We have $\begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} v = \begin{bmatrix} r_1 v_1 \\ r_2 v_2 \end{bmatrix} \in \mathcal{E}$ if and only if $(r_1 v_1/r_1)^2 + (r_2 v_2/r_2)^2 = v_1^2 + v_2^2 \leq 1$.
This is equivalent to having $v \in \mathcal{D}$. $\square$

The radii of $\mathcal{E}$ are the vectors $\pm \begin{bmatrix} r_1 \\ 0 \end{bmatrix}$ and $\pm \begin{bmatrix} 0 \\ r_2 \end{bmatrix}$. These vectors are allowed to be zero.

For each radius there is a choice of direction, but any two orthogonal radii uniquely determine $\mathcal{E}$.

More generally, we refer to any rotation of the region $\mathcal{E}$ as a *(solid) ellipse*.



The radii of an ellipse formed by rotating $\mathcal{E}$ by some angle θ counterclockwise are the vectors

$$\pm \begin{bmatrix} r_1 \cos\theta \\ r_1 \sin\theta \end{bmatrix} \qquad \text{and} \qquad \pm \begin{bmatrix} -r_2 \sin\theta \\ r_2 \cos\theta \end{bmatrix}$$

formed by rotating the radii of $\mathcal{E}$ counterclockwise by the same angle.

Any two orthogonal radii once again completely determine the ellipse.

Proposition. Suppose U is some orthogonal 2×2 matrix and $\Sigma = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}$. Then the set of vectors

$$\left\{ U\Sigma v \in \mathbb{R}^2 : v \in \mathcal{D} \right\}$$

is an ellipse whose radii have lengths r_1 and r_2, and every such ellipse arises as a set of this form.

Proof. Reflecting a standard ellipse across the line $y = x$ gives another standard ellipse. The result follows since $\{\Sigma v : v \in \mathcal{D}\}$ is a standard ellipse and U is a rotation matrix times a permutation matrix. $\square$

Proposition. Let A be a 2×2 matrix. Then the region $\{Av : v \in \mathcal{D}\}$ is an ellipse.

The lengths of the radii of this ellipse are the singular values of A.

Proof. Let $A = U\Sigma V^\top$ be a singular value decomposition.

Then $\{V^\top v : v \in \mathcal{D}\} = \mathcal{D}$ since multiplication by orthogonal matrices preserves lengths.

Therefore $\{Av : v \in \mathcal{D}\} = \{U\Sigma v : v \in \mathcal{D}\}$ so the result follows by the previous proposition. $\square$

Let's now try to say what a singular value decomposition $A = U\Sigma V^\top$ means physically for a 2×2 matrix.

Suppose the ellipse $\mathcal{E} = \{Av : v \in \mathcal{D}\}$ has radii of lengths $\sigma_1 \geq \sigma_2 \geq 0$.

As noted in the proposition, we then have $\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$.

The columns of $V = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ are two orthogonal radii of the unit disc $\mathcal{D}$.

These vectors have the property that Av_i is a radius of $\mathcal{E}$ with length r_i for each $i = 1, 2$.

This holds since $Av_1 = U\Sigma V^\top v_1 = U\Sigma V^{-1} v_1 = U\Sigma e_1 = U \begin{bmatrix} \sigma_1 \\ 0 \end{bmatrix}$ and likewise $Av_2 = U \begin{bmatrix} 0 \\ \sigma_2 \end{bmatrix}$.

The matrix U is always an orthogonal matrix whose inverse transforms the ellipse $\mathcal{E}$ back to a standard ellipse (whose radii belong to the x- and y-axes). If $\det A$ and $\det V$ have the same sign then U is a rotation matrix. Otherwise U is a rotation matrix with its columns interchanged.

A 2×2 matrix A parametrizes a linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ by telling us the images of the standard basis elements $e_1, e_2 \in \mathbb{R}^2$ (these images are the columns of A).

The SVD of A parametrizes a linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ in a different way, by telling us which orthogonal radii of the unit disc (the columns of V) are mapped to which orthogonal radii of the image ellipse (the columns of $U\Sigma$).

We can extend this interpretation of the SVD to higher dimensions, after setting

$$\mathcal{D}^n = \{v \in \mathbb{R}^n : v \bullet v \leq 1\}$$

defining an m-dimensional ellipse to be a set of the form $\{U\Sigma v : v \in \mathcal{D}^n\}$ where U is an orthogonal $m \times m$ matrix and Σ is an $m \times n$ matrix with nonzero entries only on the main diagonal.

If A is $m \times n$, then the first $r = \operatorname{rank} A$ columns of V in an SVD $A = U\Sigma V^\top$ are still orthogonal vectors of the unit disc that are transformed to orthogonal radii of some m-dimensional ellipse (in which $m - r$ radii have length zero), while the last $n - r$ columns are an orthogonal basis for $\operatorname{Nul} A$.