This document is an **exact transcript** of the lecture, with extra summary and vocabulary sections for your convenience. Due to time constraints, the lectures 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.

- Let A be an $m \times n$ matrix. Then $A^\top A$ is a symmetric $n \times n$ matrix.

  The eigenvalues of $A^\top A$ are nonnegative real numbers. This means that there are real numbers $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ such that $\det(A^\top A - xI) = (\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x)$.

  Define $\sigma_i = \sqrt{\lambda_i}$. Then the numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ are the *singular values* of A.

  The rank of A is the same as its number of nonzero singular values.

- Recall that an *orthogonal matrix* is an invertible square matrix U with $U^{-1} = U^\top$.

  Suppose A is any $m \times n$ matrix with $\operatorname{rank} A = r$.

  Suppose $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the nonzero singular values of A.

  Then 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 $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ where $D = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$.

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

- To compute an SVD for A, first find the eigenvalues of $A^\top A$.

  Then construct an orthonormal basis $v_1, v_2, \ldots, v_n$ of $\mathbb{R}^n$ consisting of eigenvectors for $A^\top A$.

  Let λ_i be the eigenvalue such that $A^\top A v_i = \lambda_i v_i$ and define $\sigma_i = \sqrt{\lambda_i}$.

  Order the basis vectors such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

  Then set $V = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}$ and define Σ in terms of the σ_i's as above.

  Let $r = \operatorname{rank} A$. This is the largest index with $\sigma_r > 0$.

  For $i = 1, 2, \ldots, r$ define $u_i = \frac{1}{\sigma_i} A v_i$.

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

  Finally set $U = \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix}$.

  The matrices U and V will then be orthogonal and $A = U\Sigma V^\top$ is a singular value decomposition.

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

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

  Every matrix has a pseudo-inverse, which can be computed from a singular value decomposition.

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

1 Last time: symmetric matrices

A matrix A is *symmetric* if $A^\top = A$.

This happens if and only if A is square and $A_{ij} = A_{ji}$ for all i, j.

Example. $\left[\begin{array}{cc} 1 & 2 \\ 2 & 3 \end{array}\right]$ is symmetric but $\left[\begin{array}{cc} 1 & 2 \\ 3 & 2 \end{array}\right]$ is not.

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

This happens precisely when U is square with orthonormal columns.

An $n \times n$ matrix A is *orthogonally diagonalizable* if there is an orthogonal matrix U and a diagonal matrix D such that $A = UDU^{-1} = UDU^\top$. In this case, the columns of U are an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for A, and the eigenvalues of these eigenvectors are the diagonal entries of D.

The following summarizes the main results from last time:

Theorem.

(1) A square matrix is orthogonally diagonalizable if and only if it is symmetric.

(2) Eigenvectors with different eigenvalues for a symmetric matrix are orthogonal.

(3) All (complex) eigenvalues of a symmetric matrix A are real. The characteristic polynomial of A has all real roots and can be expressed as $\det(A - xI) = (\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x)$ for some (not necessarily distinct) real numbers $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$.

Example. Suppose $A = \left[\begin{array}{cc} a & b \\ b & a \end{array}\right]$ for some $a, b \in \mathbb{R}$.

How does the preceding theorem apply to this generic 2-by-2 matrix? Since

$$\det(A - xI) = \det \left[\begin{array}{cc} a - x & b \\ b & a - x \end{array}\right] = (a - x)^2 - b^2 = (a - b - x)(a + b - x),$$

the eigenvalues of A are $a - b$ and $a + b$.

It's not too hard to guess the eigenvectors corresponding to these eigenvectors, though the usual method of row reducing $A - \lambda I$ to find a basis for $\mathrm{Nul}(A - \lambda I)$ will also produce the answer:

The vector $\left[\begin{array}{c} 1 \\ -1 \end{array}\right]$ is an eigenvector for A with eigenvalue $a - b$.

The vector $\left[\begin{array}{c} 1 \\ 1 \end{array}\right]$ is an eigenvector for A with eigenvalue $a + b$.

These eigenvectors are orthogonal, as predicted by the theorem. We can convert them to unit vectors by multiplying each vector by the reciprocal of its length. This gives the eigenvectors

$$\tfrac{1}{\sqrt{2}}\left[\begin{array}{c} 1 \\ -1 \end{array}\right] \qquad \text{and} \qquad \tfrac{1}{\sqrt{2}}\left[\begin{array}{c} 1 \\ 1 \end{array}\right]$$

which form an orthonormal basis for $\mathbb{R}^2$.

It follows that $A = UDU^{-1} = UDU^\top$ where $U = \tfrac{1}{\sqrt{2}}\left[\begin{array}{cc} 1 & 1 \\ -1 & 1 \end{array}\right]$ and $D = \left[\begin{array}{cc} a - b & 0 \\ 0 & a + b \end{array}\right]$.

2 Singular value decomposition

Today, we'll apply the results from last time to prove the existence of *singular value decompositions*, which give a sort of approximate orthogonal diagonalization for any matrix, not just symmetric ones.

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

Then $A^\top A$ is a symmetric $n \times n$ matrix, since $(A^\top A)^\top = A^\top (A^\top)^\top = A^\top A$.

It follows from our results last time that $A^\top A$ has all real eigenvalues. A stronger statement holds:

Lemma. All eigenvalues of $A^\top A$ are 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 $\lambda = \|Av\|^2$.

Proof. If $v \in \mathbb{R}^n$ has $\|v\| = 1$ and $A^\top A v = \lambda v$ then

$$0 \leq \|Av\|^2 = (Av) \bullet (Av) = (Av)^\top (Av) = v^\top A^\top A v = v^\top (\lambda v) = \lambda (v^\top v) = \lambda \|v\|^2 = \lambda.$$

□

The preceding lemma allows us to make the following definition.

Definition. 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 $i = 1, 2, \ldots, n$. We call the numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ the *singular values* of A.

In other words, the singular values of a matrix A are the squares roots of the eigenvalues of $A^\top A$, which are guaranteed to be nonnegative real numbers (and therefore always have well-defined square roots).

Example. Suppose $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$. Then $A^\top A = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$.

This matrix $A^\top A$ has characteristic polynomial

$$\det(A^\top A - xI) = (360 - x)(90 - x)x$$

so the eigenvalues of $A^\top A$ are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$.

The singular values of A are therefore $\sigma_1 = \sqrt{360} = 6\sqrt{10}$, $\sigma_2 = \sqrt{90} = 3\sqrt{10}$, and $\sigma_3 = 0$.

As a sequel to the lemma above, we have this nontrivial statement about the eigenvectors of $A^\top A$.

Theorem. Suppose $v_1, v_2, \ldots, v_n$ is an orthonormal basis of $\mathbb{R}^n$ composed of eigenvectors of $A^\top A$, arranged so that if $\lambda_i \in \mathbb{R}$ is the eigenvalue of v_i then $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

Assume A has r nonzero singular values.

Then $Av_1, Av_2, \ldots, Av_r$ is an orthogonal basis for the column space of A and consequently $\operatorname{rank} A = r$.

Proof. Choose indices $i \neq j$. Then $v_i \bullet v_j = 0$ so also $v_i \bullet \lambda_j v_j = 0$. Then

$$(Av_i)^\top A v_j = v_i^\top A^\top A v_j = v_i^\top (\lambda_j v_j) = v_i \bullet \lambda_j v_j = 0.$$

This shows that $Av_1, Av_2, \ldots, Av_r$ are orthogonal vectors in $\operatorname{Col} A$.

Since $\|Av_i\| = \sqrt{\lambda_i} > 0$, these vectors are all nonzero and therefore are linearly independent.

To see that these vectors span the column space of A, suppose $y \in \operatorname{Col} A$.

Then $y = Ax$ for some vector $x \in \mathbb{R}^n$, which we can write as

$$x = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

for some coefficients $c_1, c_2, \ldots, c_n \in \mathbb{R}$. If $i > r$ then $Av_i = 0$ since $\|Av_i\| = \sqrt{\lambda_i} = 0$. Therefore

$$y = Ax = c_1 Av_1 + c_2 Av_2 + \cdots + c_r Av_r + \underbrace{c_{r+1} Av_{r+1} + \cdots + c_n Av_n}_{=0} = c_1 Av_1 + c_2 Av_2 + \cdots + c_r Av_r.$$

We conclude that $Av_1, Av_2, \ldots, Av_r$ is a basis for $\mathrm{Col}\, A$. □

Corollary. The rank of a matrix is the same as its number of nonzero singular values.

We arrive at today's main result.

Theorem (Existence of SVDs)**.** Let A be an $m \times n$ matrix with rank r.

Suppose $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ are the nonzero singular values of A.

Then 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 $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ where $D = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$.

Comments. The three zeros in the matrix defining Σ represent blocks of zeros: the upper right 0 stands for an $r \times (n-r)$ zero submatrix, the lower right 0 stands for an $(m-r) \times (n-r)$ zero submatrix, and the lower left 0 stands for an $(m-r) \times r$ zero submatrix.

Another way to think of Σ: place the diagonal matrix D in the upper left corner of an $m \times n$ matrix, and then fill all of the remaining entries with zeros.

Definition. A factorization $A = U\Sigma V^\top$ with U, V, Σ as above is a *singular value decomposition* of A.

We sometimes abbreviate by writing *SVD* instead of *singular value decomposition*.

The matrices U and V in an SVD $A = U\Sigma V^\top$ are not uniquely determined by A, but Σ is.

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

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

Proof that an SVD of A exists. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the decreasing list of eigenvalues of $A^\top A$.

The singular values of A are $\sigma_i = \sqrt{\lambda_i}$ for each $i = 1, 2, \ldots, n$.

Let $v_1, v_2, \ldots, v_n$ be a list of corresponding orthonormal eigenvectors for $A^\top A$.

Then we have $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$ and $Av_1, Av_2, \ldots, Av_r$ is an orthogonal basis for $\mathrm{Col}\, A$.

For each $i = 1, 2, \ldots, r$, define $u_i = \frac{1}{\|Av_i\|} Av_i = \frac{1}{\sqrt{\lambda_i}} Av = \frac{1}{\sigma_i} Av_i$.

Then $u_1, u_2, \ldots, u_r$ is an orthonormal basis for $\mathrm{Col}\, A$.

We can choose vectors $u_{r+1}, u_{r+2}, \ldots, u_m \in \mathbb{R}^m$ such that the extended list of vectors $u_1, u_2, \ldots, u_m$ is an orthonormal basis for $\mathbb{R}^m$. Make any such choice, and define

$$U = \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix} \qquad \text{and} \qquad V = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}.$$

These matrices are orthogonal by construction, and

$$\begin{aligned} AV &= \begin{bmatrix} Av_1 & Av_2 & \dots & Av_n \end{bmatrix} \\ &= \begin{bmatrix} Av_1 & Av_2 & \dots & Av_r & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_r u_r & 0 & \dots & 0 \end{bmatrix}. \end{aligned}$$

If Σ is the matrix given in the theorem, then we also have

$$U\Sigma = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_r u_r & 0 & \dots & 0 \end{bmatrix} = AV$$

so $U\Sigma V^\top = AVV^\top = AI = A$, which confirms the theorem statement. □

We conclude this lecture with a small example, continuing from before.

Example. Again suppose $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$.

To find a singular value decomposition for A, there are three steps.

1. Find an orthogonal diagonalization of $A^\top A$.

   In this case $A^\top A$ is a 3×3 matrix, and by the usual methods (of row reducing $A - \lambda I$ to find a basis for $\mathrm{Nul}(A - \lambda I)$ for each eigenvalue λ), you can find that

   $$v_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \qquad \text{and} \qquad v_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

   is an orthonormal basis of $\mathbb{R}^3$ consisting of eigenvectors of $A^\top A$.

   The corresponding eigenvalues are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$.

2. Set up V and Σ.

   Following the proof of the theorem, we have

   $$V = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix} \qquad \text{and} \qquad D = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

   for $\sigma_1 = \sqrt{\lambda_1} = \sqrt{360}$ and $\sigma_2 = \sqrt{\lambda_2} = \sqrt{90}$.

   Since Σ has the same size as A, we get $\Sigma = \begin{bmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{bmatrix}$.

3. Construct U.

   We have $U = \begin{bmatrix} u_1 & u_2 \end{bmatrix}$ where $u_i = \frac{1}{\sigma_i} A v_i$.

   In this case you can compute that

   $$u_1 = \frac{1}{\sqrt{360}}\begin{bmatrix} 18 \\ 6 \end{bmatrix} \qquad \text{and} \qquad u_2 = \frac{1}{\sqrt{90}}\begin{bmatrix} 3 \\ -9 \end{bmatrix}$$

   which means that we can write $U = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix}$.

Putting everything together produces the singular value decomposition

$$A = U\Sigma V^\top = \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix} \begin{bmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{bmatrix} \begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}. \qquad (*)$$

Be careful to note that the third matrix factor is the transpose $V^\top$ rather than V.

One application of singular value decompositions is to show the existence of pseudo-inverses:

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

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

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

Theorem. Every matrix A has a pseudo-inverse, which can be computed as follows. 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.

Example. If A is as in (*) then a pseudo-inverse is provided by

$$A^+ = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{360} & 0 \\ 0 & 1/\sqrt{90} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}.$$

One can show that the pseudo-inverse is unique (but we won't prove this in these notes).

Proof. We have

$$AA^+A = (U\Sigma V^\top)(V\Sigma^+U^\top)(U\Sigma V^\top) = U\Sigma\Sigma^+\Sigma V^\top$$

and

$$A^+AA^+ = (V\Sigma^+U^\top)(U\Sigma V^\top)(V\Sigma^+U^\top) = V\Sigma^+\Sigma\Sigma^+U^\top$$

so it suffices to check that $\Sigma\Sigma^+\Sigma = \Sigma$ and $\Sigma^+\Sigma\Sigma^+ = \Sigma^+$. This is easy to check because of the simple form Σ and Σ^+ (they only have nonzero entries in diagonal positions). Rather than write down a formal argument, here is an example which captures the main idea: when $a \neq 0$ and $b \neq 0$ we have

$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \end{bmatrix} \begin{bmatrix} 1/a & 0 \\ 0 & 1/b \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 1/a & 0 \\ 0 & 1/b \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \end{bmatrix} \begin{bmatrix} 1/a & 0 \\ 0 & 1/b \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/a & 0 \\ 0 & 1/b \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1/a & 0 \\ 0 & 1/b \\ 0 & 0 \end{bmatrix}.$$

$\square$

3 Vocabulary

Keywords from today's lecture:

1. *Singular values* of an $m \times n$ matrix A.

   The square roots of the eigenvalues of $A^\top A$, which are all nonnegative real numbers.

   Example: if A is diagonal then its singular values are the absolute values of its diagonal entries.

2. *Singular value decomposition* of an $m \times n$ matrix A.

   A decomposition $A = U\Sigma V^\top$ where U is an $m \times m$ matrix with $U^{-1} = U^\top$, V is an $n \times n$ matrix with $V^{-1} = V^\top$, and Σ is the $m \times n$ matrix whose first r diagonal entries are the singular values of A in decreasing order, and whose other entries are all zero.

   There may be more than one singular value decomposition for A.

   Example:

$$\underbrace{\begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}}_{=A} = \underbrace{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}}_{=U} \underbrace{\begin{bmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{bmatrix}}_{=\Sigma} \underbrace{\begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}^\top}_{=V^\top}.$$

3. *Pseudo-inverse* of an $m \times n$ matrix A.

   An $n \times m$ matrix A^+ with $AA^+A = A$ and $A^+AA^+ = A^+$.

   Example: a pseudo-inverse for $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.