This document is intended as an **exact transcript** of the lecture, with extra summary and vocabulary sections for your convenience. By design, the material covered in lecture is exactly the same as what is in these notes. 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.

- A *line of best fit* through data points $(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)$ is an equation $y = \beta_0 + \beta_1 x$ where $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \in \mathbb{R}^2$ is a least-squares solution to $Ax = b$ where $A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_n \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$.
- A matrix A is *symmetric* if $A^\top = A$. This can only hold if A is square. For example:

  $$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}.$$

  If A is symmetric then so is A^2, A^3, A^4, etc.

  If A is symmetric and invertible then so is A^{-1}, A^{-2}, A^{-3}, etc.

  If A is symmetric and u and v are eigenvectors for A with **different** eigenvalues, then $u \bullet v = 0$.
- A list of vectors $u_1, u_2, \dots, u_p$ is *orthonormal* if $u_i \bullet u_i = 1$ and $u_i \bullet u_j = 0$ for all $i \neq j$.

  A square matrix P is invertible with $P^{-1} = P^\top$ if and only if its columns are orthonormal.

  An $n \times n$ matrix A is *orthogonally diagonalizable* if there is a diagonal matrix D and an invertible matrix P with $P^{-1} = P^\top$ such that $A = PDP^{-1}$.
- When $A = PDP^{-1}$ where D is diagonal and $P^{-1} = P^\top$, the diagonal entries of D are the eigenvalues of A, and the columns of P are an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors for A.

  Conversely, an $n \times n$ matrix A is orthogonally diagonalizable if and only if there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors for A.
- Surprising fact: all (complex) eigenvalues of a symmetric matrix $A = A^\top$ belong to $\mathbb{R}$.

  Surprising fact: an $n \times n$ matrix A is orthogonally diagonalizable if and only if $A = A^\top$.

  Much of this lecture is spent proving these facts.
- To orthogonally diagonalize a given $n \times n$ symmetric matrix A, you need to find an orthogonal basis of $\mathbb{R}^n$ consisting of eigenvectors $v_1, v_2, \dots, v_n$ for A.

  Once you find this, let $u_i = \frac{1}{\|v_i\|} v_i$ and $U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$.

  Then $A = UDU^\top$ where D is the diagonal matrix whose ith diagonal entry is the eigenvalue of v_i.
- To find the orthogonal basis of eigenvectors $v_1, v_2, \dots, v_n$ for A:
  1. Factor the characteristic polynomial of A to compute its eigenvalues.
  2. For each eigenvalue λ, do the usual row reduce procedure to find a basis for $\mathrm{Nul}(A - \lambda I)$.
  3. Apply the Gram-Schmidt process to convert your basis of $\mathrm{Nul}(A - \lambda I)$ to an orthogonal basis.
  4. Finally combine these orthogonal bases — the combined list of vectors will still be orthogonal.

1 Last time: least-squares problems

Definition. Suppose A is an $m \times n$ matrix and $b \in \mathbb{R}^m$.

The linear system $A^\top Ax = A^\top b$ is always consistent, so has at least one solution.

A solution to $A^\top Ax = A^\top b$ is called a *least-squares solution* to the equation $Ax = b$.

Let $\|v\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \geq 0$ for $v \in \mathbb{R}^n$. Recall that $\|v\| = 0$ if and only if $v = 0$.

Fact. A vector $s \in \mathbb{R}^n$ is a least-squares solution to $Ax = b$ if and only if $\|b - As\| \leq \|b - Ax\|$ for all x.

The linear system $Ax = b$ is consistent if and only if $\|b - Ax\| = 0$ for some $x \in \mathbb{R}^n$.

This means that if $Ax = b$ is consistent then all least-squares solutions s satisfy $\|b - As\| = 0$ so $As = b$.

If $Ax = b$ is inconsistent, there is still at least one least-squares solution s (but in this case $\|b - As\| > 0$).

Theorem. Let A be an $m \times n$ matrix. The following properties are equivalent:

(a) $Ax = b$ has a unique least-squares solution for each $b \in \mathbb{R}^m$.

(b) The columns of A are linearly independent.

(c) $A^\top A$ is invertible.

Example (Lines of best fit)**.** Suppose we have n data points $(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$.

We want to find parameters $\beta_0, \beta_1 \in \mathbb{R}$ such that $y = \beta_0 + \beta_1 x$ describes the *line of best fit* for this data. If our points are all on the same line, then for some $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \in \mathbb{R}^2$ we would have

$$b_i = \beta_0 + \beta_1 a_i \quad \text{for } i = 1, 2, \ldots, n,$$

meaning that $x = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ is an exact solution to the linear system $Ax = b$ where

$$A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_n \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

If the given points are not on the same line, then no exact solution to $Ax = b$ exists, and we should instead try to find a least-squares solution to this linear system.

To be concrete, suppose we have four points $(2, 1), (5, 2), (7, 3)$, and $(8, 3)$ so that

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

The least-squares solutions to $Ax = b$ are the exact solutions to $A^\top Ax = A^\top b$. We have

$$A^\top A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

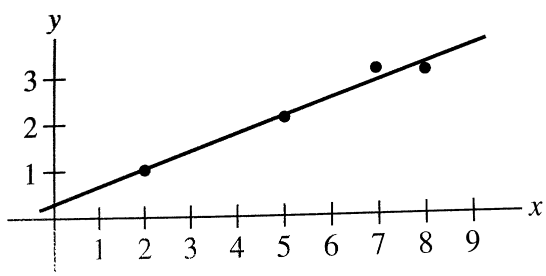
and

$$A^\top b = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

The matrix $A^\top A$ is invertible. (Why?) It follows that a least-squares solution is provided by

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (A^\top A)^{-1} A^\top b = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}.$$

Thus our line of best fit for the data is $y = \frac{2}{7} + \frac{5}{14}x$:



2 Symmetric matrices

A matrix A is *symmetric* if $A^\top = A$. This happens if A is square and $A_{ij} = A_{ji}$ for all i, j.

Example. $\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$ and $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}$ and $\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$ are symmetric matrices.

$\begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 6 & -6 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \end{bmatrix}$ are not symmetric.

Proposition. If A is a symmetric matrix and k is a positive integer then A^k is also symmetric.

Proof. If $A = A^\top$ then $(A^k)^\top = (AA\cdots A)^\top = A^\top \cdots A^\top A^\top = (A^\top)^k = A^k$. □

Proposition. If A is an invertible symmetric matrix then A^{-1} is also symmetric.

Proof. This is because $(A^{-1})^\top = (A^\top)^{-1}$. □

Recall how we can diagonalize a matrix.

Example. Let $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$.

Then $\det(A - xI) = (8-x)(6-x)(3-x)$ so the eigenvalues of A are 8, 6, and 3. By constructing bases for the null spaces of $A - 8I$, $A - 5I$, and $A - 3I$, we find that the following are eigenvectors of A:

$$v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \text{ with eigenvalue 8.}$$

$$v_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \text{ with eigenvalue 6.}$$

$$v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ with eigenvalue 3.}$$

These eigenvectors are actually an orthogonal basis for $\mathbb{R}^3$.

Converting these vectors to unit vectors gives an orthonormal basis of eigenvectors:

$$u_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \qquad u_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \qquad u_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}.$$

We then have $A = PDP^{-1}$ where

$$P = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \qquad \text{and} \qquad D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

(Why does this hold? It is enough to check that $PDP^{-1}v = Av$ for $v \in \{u_1, u_2, u_3\}$.)

Since the columns of P are orthonormal, we actually have $P^\top = P^{-1}$ so $A = PDP^\top$.

The special properties in this example will turn out to hold for all symmetric matrices.

Theorem. Suppose A is a symmetric matrix. Then any two eigenvectors from different eigenspaces of A are orthogonal. In other words, if $A = A^\top$ is $n \times n$ and $u, v \in \mathbb{R}^n$ are such that $Au = au$ and $Av = bv$ for numbers $a, b \in \mathbb{R}$ with $a \neq b$, then $u \bullet v = 0$.

Proof. Let u and v be eigenvectors of A with eigenvalues a and b, where $a \neq b$.

Then $au \bullet v = Au \bullet v = (Au)^\top v = u^\top A^\top v = u^\top Av = u \bullet Av = u \bullet bv$.

But $au \bullet v = a(u \bullet v)$ and $u \bullet bv = b(u \bullet v)$, so this means $a(u \bullet v) = b(u \bullet v)$ and therefore $(a-b)(u \bullet v) = 0$.

Since $a - b \neq 0$, it follows that $u \bullet v = 0$. □

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

Definition. A matrix A is *orthogonally diagonalizable* if there is an orthogonal matrix P and a diagonal matrix D such that $A = PDP^{-1} = PDP^\top$.

When A is orthogonally diagonalizable and $A = PDP^{-1} = PDP^\top$, the diagonal entries of D are the eigenvalues of A, and the columns of P are the corresponding eigenvectors; moreover, these eigenvectors form an orthonormal basis of $\mathbb{R}^n$.

In fact, it follows by the arguments in our earlier lectures about diagonalizable matrices that an $n \times n$ matrix A is orthogonally diagonalizable if and only if there is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for A.

Surprisingly, there is a much more direct characterization of orthogonally diagonalizable matrices:

Theorem. A square matrix is orthogonally diagonalizable if and only if it is symmetric.

We prove this after a sequence of lemmas.

Lemma. If A is orthogonally diagonalizable then A is symmetric.

Proof. If X, Y, Z are $n \times n$ matrices then $(XYZ)^\top = Z^\top (XY)^\top = Z^\top Y^\top X^\top$.

Suppose $A = PDP^\top$ where D is diagonal. Then $D = D^\top$ and $(P^\top)^\top = P$, so

$$A^\top = (PDP^\top)^\top = (P^\top)^\top D^\top P^\top = PDP^\top = A.$$

□

Lemma. All (complex) eigenvalues of an $n \times n$ symmetric matrix A with real entries belong to $\mathbb{R}$.

Proof. Suppose A is a symmetric $n \times n$ matrix with real entries, so that $A = A^\top = \overline{A}$.

Let $v \in \mathbb{C}^n$. Then $\overline{v}^\top Av$ is some complex number.

For example, if $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ and $v = \begin{bmatrix} 1+i \\ 1-i \end{bmatrix}$ then

$$\overline{v}^\top Av = \begin{bmatrix} 1-i & 1+i \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} = \begin{bmatrix} 3+i & 3-i \end{bmatrix} \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} = (3+i)(1+i) + (3-i)(1-i) = 4.$$

In fact, the number $\overline{v}^\top Av$ belongs to $\mathbb{R}$ since $\overline{\overline{v}^\top Av} = v^\top A\overline{v} = \left(\overline{v}^\top Av\right)^\top = \overline{v}^\top Av$.

(The last equality holds since both sides are 1×1 matrices, i.e., scalars.)

Now suppose $v \in \mathbb{C}^n$ is an eigenvector for A with eigenvalue $\lambda \in \mathbb{C}$. Then $\overline{v}^\top Av = \overline{v}^\top(\lambda v) = \lambda(\overline{v}^\top v) \in \mathbb{R}$. The complex number $\overline{v}^\top v$ always belongs to $\mathbb{R}$ (why?) so it must also hold that $\lambda \in \mathbb{R}$. □

Lemma. An $n \times n$ matrix A with all real eigenvalues can be written as $A = URU^\top$ where U is an $n \times n$ orthogonal matrix (i.e., has orthonormal columns) and R is an $n \times n$ upper-triangular matrix.

One calls $A = URU^\top$ with U and R of this form a *Schur factorization* of A.

Proof. Suppose A is an $n \times n$ matrix with all real eigenvalues.

Let $u_1 \in \mathbb{R}^n$ be a unit eigenvector for A with eigenvalue $\lambda \in \mathbb{R}$.

Let $u_2, \dots, u_n \in \mathbb{R}^n$ be any vectors such that $u_1, u_2, \dots, u_n$ is an orthonormal basis for $\mathbb{R}^n$.

(One way to construct these vectors: let $u_1 = x_1, x_2, \dots, x_n$ be any basis, apply the Gram-Schmidt process to get $u_1 = v_1, v_2, \dots, v_n$, and then convert each v_i to a unit vector.)

Define $U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ so that $U^\top = U^{-1}$.

By considering the product $U^\top AUe_i$ for $i = 1, 2, \dots, n$, one finds that $U^\top AU$ has the form

$$U^\top AU = \begin{bmatrix} \lambda & * \\ 0 & B \end{bmatrix}$$

for some $(n-1) \times (n-1)$ matrix B. Here, $*$ stands for $n-1$ arbitrary entries.

The matrix $U^\top AU = U^{-1}AU$ has the same characteristic polynomial as A.

This polynomial is just $(\lambda - x)\det(B - xI)$, which is $\lambda - x$ times the characteristic polynomial of B.

Since the characteristic polynomial of A has all real roots, the same must be true of the characteristic polynomial of B. Thus B must also have all real eigenvalues.

By repeating the argument above, we deduce that there is an eigenvalue $\mu \in \mathbb{R}$ for B, an $(n-1) \times (n-1)$ orthogonal matrix V, and an $(n-2) \times (n-2)$ matrix C with all real eigenvalues such that

$$V^\top BV = \begin{bmatrix} \mu & * \\ 0 & C \end{bmatrix}.$$

The matrix $\begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}$ is also orthogonal, and the product of orthogonal matrices is orthogonal. (Why?)

It follows for the orthogonal matrix $W = U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}$ that $W^\top A W = \begin{bmatrix} \lambda & * & * \\ 0 & \mu & * \\ 0 & 0 & C \end{bmatrix}$.

By continuing in this way, we will eventually construct an orthogonal matrix X and an upper-triangular matrix R such that $X^\top A X = R$, in which case $A = XX^\top AXX^\top = XRX^\top$. □

Now we can prove the theorem.

Proof of theorem. The first lemma shows that if A is orthogonally diagonalizable then A is symmetric.

Suppose conversely that A is symmetric. Then A has all real eigenvalues, so there exists a Schur factorization $A = URU^\top$. We then have $A^\top = (URU^\top)^\top = UR^\top U^\top$ but also $A^\top = A = URU^\top$.

Since $U^\top = U^{-1}$, it follows that $R = R^\top$. Since R is upper-triangular, this can only hold if R is diagonal.

But if R is diagonal then $A = URU^\top$ is orthogonally diagonalizable. □

To orthogonally diagonalize an $n \times n$ symmetric matrix A, we just need to find an orthogonal basis of eigenvectors $v_1, v_2, \dots, v_n$ for $\mathbb{R}^n$. Then $A = UDU^\top$ with $U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ where $u_i = \frac{1}{\|v_i\|} v_i$ and D is the diagonal matrix of the corresponding eigenvalues.

If all eigenspaces of A are 1-dimensional, then any basis of eigenvectors will be orthogonal. If A has an eigenspace of dimension greater than one, then after finding a basis for this eigenspace, it is necessary to apply the Gram-Schmidt process to convert this basis to one that is orthogonal.

Corollary. If $A = UDU^\top$ where $U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ has orthonormal columns and

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

is diagonal, then $A = \lambda_1 u_1 u_1^\top + \lambda_2 u_2 u_2^\top + \dots + \lambda_n u_n u_n^\top$.

Each product $u_i u_i^\top$ is an $n \times n$ matrix of rank 1. One calls this expression a *spectral decomposition* of A.

Example. Let $A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$. A spectral decomposition of A is given by

$$\begin{aligned} A &= \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \\ &= 8 \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} + 3 \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \\ &= \begin{bmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{bmatrix} + \begin{bmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{bmatrix}. \end{aligned}$$

3 Vocabulary

Keywords from today's lecture:

1. **Symmetric matrix**.

   A matrix A that is equal to its transpose, so that $A = A^\top$. Such a matrix is square.

   Symmetric matrices are precisely the square matrices A that are **orthogonally diagonalizable**, in other words, the matrices that can be expressed as

   $$A = PDP^\top$$

   where D is a diagonal matrix and P is an invertible matrix with $P^{-1} = P^\top$.

   Example: $\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ or any diagonal matrix.

2. **Schur factorization** of an $n \times n$ matrix A.

   A decomposition $A = URU^\top$ where R is an $n \times n$ upper triangular matrix and U is an orthogonal matrix (i.e., U is invertible with $U^{-1} = U^\top$).