

This document is an **exact transcript** of the lecture, with extra summary and vocabulary sections. 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.

- The *inner product* or *dot product* of two vectors $u, v \in \mathbb{R}^n$ is the scalar $u \bullet v = u^\top v \in \mathbb{R}$.
- We always have $v \bullet v \geq 0$. The *length* of $v \in \mathbb{R}^n$ is $\|v\| = \sqrt{v \bullet v}$.

If $c \in \mathbb{R}$ and $v \in \mathbb{R}^n$ then $\|cv\| = |c|\|v\|$.

The *distance* between $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ is defined to be the length $\|u - v\|$.

- A *unit vector* is a vector $u \in \mathbb{R}^n$ with $\|u\| = 1$.

If $v \in \mathbb{R}^n$ is any nonzero vector, then the *unit vector in the direction of v* is $u = \frac{1}{\|v\|}v \in \mathbb{R}^n$.

- Two vectors $u, v \in \mathbb{R}^n$ are *orthogonal* if $u \bullet v = 0$.

If $V \subseteq \mathbb{R}^n$ is a subspace then its *orthogonal complement* is the subspace

$$V^\perp = \{w \in \mathbb{R}^n : v \bullet w = 0 \text{ for all } v \in V\}.$$

We always have $V \cap V^\perp = \{0\} \subseteq \mathbb{R}^n$. Next time, we'll see that $\dim V + \dim V^\perp = n$.

If A is an $m \times n$ matrix then $(\text{Col } A)^\perp = \text{Nul}(A^\top)$.

- An *orthogonal basis* is a basis in which any two vectors are orthogonal.

Suppose $v_1, v_2, \dots, v_p \in \mathbb{R}^n$ are nonzero vectors with $v_i \bullet v_j = 0$ for all $i \neq j$.

Then these vectors are linearly independent, and therefore an orthogonal basis for their span.

1 Last time: properties of eigenvalues

The *trace* of a square matrix A is the sum of its diagonal entries.

We denote this by the symbol $\text{tr}(A)$. For 2×2 matrices we have $\text{tr}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = a + d$.

Suppose A and B are $n \times n$ matrices. Although in general $\text{tr}(AB) \neq \text{tr}(A)\text{tr}(B)$, we have both

$$\text{tr}(AB) = \text{tr}(BA) \quad \text{and} \quad \det(AB) = \det(A)\det(B) = \det(B)\det(A) = \det(BA).$$

Theorem. Let A be an $n \times n$ matrix and write I for the $n \times n$ identity matrix. The fundamental theorem of algebra tells us that there are complex numbers $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$ such that

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

For these numbers it holds that $\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$ and $\text{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

In words: the product of the eigenvalues of A , repeated with multiplicity, is the determinant of A , while the sum of the eigenvalues of A , repeated with multiplicity, is the trace of A .

The theorem is easy to observe when A is a triangular matrix: for example if $A = \begin{bmatrix} \lambda_1 & a & b \\ 0 & \lambda_2 & c \\ 0 & 0 & \lambda_3 \end{bmatrix}$ then

$$\det(A - xI) = (\lambda_1 - x)(\lambda_2 - x)(\lambda_3 - x) \quad \text{and} \quad \text{tr} A = \lambda_1 + \lambda_2 + \lambda_3 \quad \text{and} \quad \det A = \lambda_1 \lambda_2 \lambda_3.$$

A few other properties of eigenvalues and eigenvectors worth noting:

Proposition. If A is a square matrix then A and $A^\top$ have the same eigenvalues.

Proof. This follows since $\det(A - xI) = \det((A - xI)^\top) = \det(A^\top - xI^\top) = \det(A^\top - xI)$. □

Proposition. Let A be a square matrix. Then A is invertible if and only if 0 is not one of its eigenvalues.

Proof. 0 is an eigenvalue of A if and only if $\det A = 0$ which occurs precisely when A is not invertible. □

Proposition. Assume A is invertible. Then A and A^{-1} have the same eigenvectors, but v is an eigenvector of A with eigenvalue λ if and only if v is an eigenvector of A^{-1} with eigenvalue λ^{-1} .

Proof. If A is invertible and $Av = \lambda v$ then $v = A^{-1}Av = A^{-1}\lambda v = \lambda A^{-1}v$ so $A^{-1}v = \lambda^{-1}v$. □

Corollary. If A is invertible and diagonalizable then A^{-1} is diagonalizable.

Proof. If A is invertible and diagonalizable, then $\mathbb{R}^n$ has a basis consisting of eigenvectors of A , but this basis is then also made up of eigenvectors of A^{-1} , so A^{-1} is diagonalizable. □

Corollary. If A is diagonalizable then $A^\top$ is diagonalizable.

Proof. Suppose $A = PDP^{-1}$ where D is diagonal. Let $Q = (P^{-1})^\top = (P^\top)^{-1}$.

Then $D^\top = D$ so $A^\top = (PDP^{-1})^\top = (P^{-1})^\top D^\top P^\top = QDQ^{-1}$. □

2 Inner products and orthogonality

In this lecture, we will only work with vectors in $\mathbb{R}^n$ and with matrices that have all real entries.

Definition. The *inner product* or *dot product* of two vectors

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

in $\mathbb{R}^n$ is the scalar $u \bullet v = u_1v_1 + u_2v_2 + \cdots + u_nv_n = u^\top v = v^\top u = v \bullet u$.

For example, $\begin{bmatrix} a \\ b \end{bmatrix} \bullet \begin{bmatrix} -b \\ a \end{bmatrix} = -ab + ab = 0$ for any $a, b \in \mathbb{R}$.

Definition. The *length* of a vector $v \in \mathbb{R}^n$ is $\|v\| = \sqrt{v \bullet v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.

Essential properties of length and inner product.

Let $u, v, w \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

- (a) $u \bullet v = v \bullet u$ and $(u + v) \bullet w = u \bullet w + v \bullet w$ and $(cv) \bullet w = c(v \bullet w)$, while $\|cv\| = |c|\|v\|$.
- (b) $v \bullet v = v_1^2 + v_2^2 + \cdots + v_n^2 \geq 0$ and $\|v\| \geq 0$.
- (c) $v \bullet v = 0$ if and only if $\|v\| = 0$ if and only if $v = 0 \in \mathbb{R}^n$.
- (d) There is a general trigonometric identity relating $u \bullet v$ to the angle θ between u and v :

$$u \bullet v = \|u\|\|v\|\cos\theta.$$

This holds even when $u = 0$ or $v = 0$ (as both sides are zero), although then θ is not defined.

We won't need to use this identity directly very often.

However, it is useful for gaining intuition about the sign of $u \bullet v$: this value is negative if and only if $\cos\theta \in [-1, 0)$, which happens precisely when u and v form an obtuse angle θ .

The *distance* between two vectors $u, v \in \mathbb{R}^n$ is the length of the difference $\|u - v\|$.

A *unit vector* is a vector $u \in \mathbb{R}^n$ with $\|u\| = 1$.

If $v \in \mathbb{R}^n$ is any nonzero vector, then the *unit vector in the direction of v* is $u = \frac{1}{\|v\|}v \in \mathbb{R}^n$.

Note that for this u we have $\|u\| = \|\frac{1}{\|v\|}v\| = \frac{1}{\|v\|}\|v\| = \frac{1}{\|v\|}\|v\| = 1$.

Example. The unit vector in the direction of $v = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ is $u = \frac{1}{\sqrt{1^2+1^2+1^2+1^2}}v = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$.

Definition. Two vectors $u, v \in \mathbb{R}^n$ are *orthogonal* if $u \bullet v = 0$.

When u and v are orthogonal we also say that “ u is orthogonal to v .”

Proposition. Suppose $u, v \in \mathbb{R}^2$ are nonzero vectors that are orthogonal to each other, so that $u \bullet v = 0$. Then u and v , drawn as arrows in the xy -plane, belong to perpendicular lines through the origin. In other words, these vectors are perpendicular in the usual sense of planar geometry.

Concretely, if $u, v \in \mathbb{R}^2$ are orthogonal and $0 \neq u = \begin{bmatrix} a \\ b \end{bmatrix}$, then v is a scalar multiple $\begin{bmatrix} -b \\ a \end{bmatrix}$, which is the vector obtained by rotating u counterclockwise by 90 degrees.

Proof. This follows directly from the identity $u \bullet v = \|u\|\|v\|\cos\theta$, which implies that $u \bullet v = 0$ if and only if the angle θ between u and v is $\pm\frac{\pi}{2}$. Below is a more self-contained proof.

Write $u = \begin{bmatrix} a \\ b \end{bmatrix}$ and $v = \begin{bmatrix} x \\ y \end{bmatrix}$. Then $u \bullet v = ax + by = 0$.

If $a = 0$ then $b \neq 0$ since $u \neq 0$, so $y = -\frac{a}{b}x = 0$ and $v = \begin{bmatrix} x \\ 0 \end{bmatrix} = -\frac{x}{b} \begin{bmatrix} -b \\ 0 \end{bmatrix}$.

If $a \neq 0$ then $x = -\frac{b}{a}y$ so $v = \begin{bmatrix} -\frac{b}{a}y \\ y \end{bmatrix} = \frac{y}{a} \begin{bmatrix} -b \\ a \end{bmatrix}$. Thus v is a scalar multiple of $\begin{bmatrix} -b \\ a \end{bmatrix}$.

To see that $\begin{bmatrix} a \\ b \end{bmatrix}$ and $\begin{bmatrix} -b \\ a \end{bmatrix}$ are perpendicular, note that $\begin{bmatrix} -b \\ a \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$.

The 2-by-2 matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ acts by rotating a vector 90 degrees counterclockwise. □

3 Orthogonal complements

Let $V \subseteq \mathbb{R}^n$ be a subspace. The *orthogonal complement* of V is $V^\perp = \{w \in \mathbb{R}^n : v \bullet w = 0 \text{ for all } v \in V\}$.

We pronounce “ $V^\perp$ ” as “vee perp.”

Proposition. If $V \subseteq \mathbb{R}^n$ is a subspace then its orthogonal complement $V^\perp \subseteq \mathbb{R}^n$ is also a subspace.

Proof. Since $v \bullet 0 = 0$ for all $v \in \mathbb{R}^n$ it holds that $0 \in V^\perp$. This confirms that $V^\perp$ is nonempty.

If $x, y \in V^\perp$ and $c \in \mathbb{R}$ then $v \bullet cx = c(v \bullet x) = 0$ and $v \bullet (x + y) = v \bullet x + v \bullet y = 0 + 0 = 0$ for all $v \in V$ so cx and $x + y$ both belong to $V^\perp$. Hence $V^\perp$ is a subspace. □

The operation $(\cdot)^\perp$ relates the column space, null space, and transpose of a matrix in the following way:

Theorem. Suppose A is an $m \times n$ matrix. Then $(\text{Col } A)^\perp = \text{Nul}(A^\top) \subseteq \mathbb{R}^m$.

Proof. Write $A = [a_1 \ a_2 \ \dots \ a_n]$ where $a_i \in \mathbb{R}^m$. Let $v \in \mathbb{R}^m$.

If $v \in (\text{Col } A)^\perp$ then we must have $v \bullet a_i = a_i^\top v = 0$ for all i .

Conversely, if $v \bullet a_i = a_i^\top v = 0$ for all i then

$$(c_1 a_1 + c_2 a_2 + \dots + c_n a_n) \bullet v = c_1 \underbrace{(a_1 \bullet v)}_{=0} + c_2 \underbrace{(a_2 \bullet v)}_{=0} + \dots + c_n \underbrace{(a_n \bullet v)}_{=0} = 0$$

for any scalars $c_1, c_2, \dots, c_n \in \mathbb{R}$ so $v \in (\text{Col } A)^\perp$.

Thus $v \in (\text{Col } A)^\perp$ if and only if $v \bullet a_i = a_i^\top v = 0$ for all i . This holds if and only if

$$A^\top v = \begin{bmatrix} a_1^\top \\ a_2^\top \\ \vdots \\ a_n^\top \end{bmatrix} v = \begin{bmatrix} a_1 \bullet v \\ a_2 \bullet v \\ \vdots \\ a_n \bullet v \end{bmatrix} = 0 \in \mathbb{R}^m, \quad \text{which means that } v \in \text{Nul}(A^\top).$$

□

Lemma. Let $V \subseteq \mathbb{R}^n$ be a subspace. If $w \in V \cap V^\perp$ then $w = 0$.

Proof. If $w \in V$ and $w \in V^\perp$ then $w \bullet w = 0$ so $w = 0$.

□

Proposition. Let $V \subseteq \mathbb{R}^n$ be a subspace. If $S \subseteq V$ and $T \subseteq V^\perp$ are two sets of linearly independent vectors, then $S \cup T$ is also linearly independent.

Proof. Suppose there was a nontrivial linear dependence among the elements of $S \cup T$ equal to zero. Rewrite this linear dependence so that the terms from S are on the left side of the equals sign and the terms from T are on the other side. Then we would have an equation of the form

$$\underbrace{a_1 v_1 + \cdots + a_k v_k}_{\in V} = \underbrace{b_1 w_1 + \cdots + b_l w_l}_{\in V^\perp}$$

where $v_1, \dots, v_k \in S$ and $w_1, \dots, w_l \in T$, for some coefficients $a_1, a_2, \dots, a_k, b_1, b_2, \dots, b_l \in \mathbb{R}$ which are not all zero. But such an equation would imply that a nonzero element of V is equal to a nonzero element of $V^\perp$, which is impossible by the lemma. □

Corollary. If $V \subseteq \mathbb{R}^n$ is a subspace then $\dim V^\perp \leq n - \dim V$.

Proof. If S is a basis for V and T is a basis for $V^\perp$ then $\dim V + \dim V^\perp = |S| + |T| = |S \cup T|$. Since $S \cup T$ is a set of linearly independent vectors in $\mathbb{R}^n$, its size must be at most n . □

We will see next lecture that the inequality $\leq$ in this corollary is actually always equality $=$.

4 Orthogonal bases and orthogonal projections

The following proposition is called the *Generalized Pythagorean theorem*.

Proposition. Two vectors $u, v \in \mathbb{R}^n$ are orthogonal if and only if $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.

Proof. The proof is just a little algebra:

$$\|u + v\|^2 = (u + v) \bullet (u + v) = u \bullet (u + v) + v \bullet (u + v) = u \bullet u + u \bullet v + v \bullet u + v \bullet v = \|u\|^2 + \|v\|^2 + 2(u \bullet v).$$

Then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$ if and only if $u \bullet v = 0$.

The equivalence of this proposition to the classical Pythagorean theorem boils down to our observation earlier that orthogonal vectors in $\mathbb{R}^2$ form the sides of a right triangle. □

A collection of vectors $u_1, u_2, \dots, u_p \in \mathbb{R}^n$ is **orthogonal** if $u_i \bullet u_j = 0$ whenever $1 \leq i < j \leq p$.

In particular, an **orthogonal basis** of $\mathbb{R}^n$ is a basis in which any two vectors are orthogonal.

For example, the standard basis $e_1, e_2, \dots, e_n$ is an orthogonal basis for $\mathbb{R}^n$.

Theorem. Suppose the vectors $u_1, u_2, \dots, u_p \in \mathbb{R}^n$ are orthogonal and all nonzero.

Then $u_1, u_2, \dots, u_p$ are linearly independent.

Proof. Suppose $c_1 u_1 + c_2 u_2 + \dots + c_p u_p = 0$ for some coefficients $c_1, c_2, \dots, c_p \in \mathbb{R}$.

For each $i = 1, 2, \dots, p$, we then have

$$0 = (c_1 u_1 + c_2 u_2 + \dots + c_p u_p) \bullet u_i = c_1 (u_1 \bullet u_i) + c_2 (u_2 \bullet u_i) + \dots + c_p (u_p \bullet u_i) = c_i \|u_i\|^2$$

since $u_j \bullet u_i = 0$ if $i \neq j$. But since u_i is nonzero, $\|u_i\|^2 \neq 0$, so it must hold that $c_i = 0$. As this argument applies to each index i , we deduce that $c_1 = c_2 = \dots = c_p = 0$.

In other words, the only way we can have $c_1 u_1 + c_2 u_2 + \dots + c_p u_p = 0$ is if all of the coefficients are zero, which is the definition of linear independence. $\square$

Corollary. Any set of nonzero, orthogonal vectors is an orthogonal basis for the subspace they span.

Any set of n nonzero, orthogonal vectors in $\mathbb{R}^n$ is an orthogonal basis for $\mathbb{R}^n$.

Proposition. Suppose $u_1, u_2, \dots, u_p$ is an orthogonal basis for a subspace $V \subseteq \mathbb{R}^n$.

Let $y \in V$. Then we can write $y = c_1 u_1 + c_2 u_2 + \dots + c_p u_p$ where

$$c_i = \frac{y \bullet u_i}{u_i \bullet u_i} = \frac{y \bullet u_i}{\|u_i\|^2}.$$

Proof. A basis must span V , so $y = c_1 u_1 + c_2 u_2 + \dots + c_p u_p$ for some coefficients $c_1, c_2, \dots, c_p \in \mathbb{R}$.

Since $y \bullet u_i = c_i (u_i \bullet u_i)$ for each $i = 1, 2, \dots, p$, the result follows. $\square$

Example. Suppose $u_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$ and $u_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$ and $u_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$.

You can check that these three vectors are orthogonal.

For example, $u_1 \bullet u_3 = -3/2 - 2 + 7/2 = 0$.

The vectors are therefore linearly independent, so are an orthogonal basis for $\mathbb{R}^3$.

For $y = \begin{bmatrix} 6 \\ 1 \\ 8 \end{bmatrix}$ we have $y \bullet u_1 = 11$ and $y \bullet u_2 = -12$ and $y \bullet u_3 = -33$.

We also have $u_1 \bullet u_1 = 11$ and $u_2 \bullet u_2 = 6$ and $u_3 \bullet u_3 = 33/2$. Therefore $y = u_1 - 2u_2 - 2u_3$.

5 Vocabulary

Keywords from today's lecture:

1. **Inner product** of vectors $u, v \in \mathbb{R}^n$.

The scalar $u \bullet v = u^\top v \in \mathbb{R}$.

Example: $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \bullet \begin{bmatrix} -1 \\ -10 \\ -100 \end{bmatrix} = -1 - 20 - 300 = -321.$

2. **Length** of a vector $v \in \mathbb{R}^n$ and **distance** between $u, v \in \mathbb{R}^n$.

The *length* of $v \in \mathbb{R}^n$ is $\|v\| = \sqrt{v \bullet v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$ where $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$

The *distance* from $u \in \mathbb{R}^n$ to $v \in \mathbb{R}^n$ is $\|u - v\|.$

3. **Unit vector.**

A *unit vector* is a vector in $\mathbb{R}^n$ with length 1.

The unit vector in the same direction as a nonzero vector $v \in \mathbb{R}^n$ is $u = \frac{1}{\|v\|}v.$

4. **Orthogonal vectors.**

Two vectors $u, v \in \mathbb{R}^n$ are *orthogonal* if $u \bullet v = 0.$

A collection of vectors in $\mathbb{R}^n$ is orthogonal if any two of the vectors are orthogonal.

A basis of a subspace is *orthogonal* if any two vectors in the basis are orthogonal.

Example: In $\mathbb{R}^2$, the vectors $\begin{bmatrix} a \\ b \end{bmatrix}$ and $\begin{bmatrix} -b \\ a \end{bmatrix}$ are always orthogonal.

5. **Orthogonal complement** of a subspace $V \subseteq \mathbb{R}^n$.

The subspace $V^\perp = \{w \in \mathbb{R}^n : v \bullet w = 0 \text{ for all } v \in V\}.$

Example: If $V = \mathbb{R}\text{-span}\{e_1, e_2, \dots, e_i\} \subseteq \mathbb{R}^n$ then $V^\perp = \mathbb{R}\text{-span}\{e_{i+1}, e_{i+2}, \dots, e_n\}.$

If $V = \mathbb{R}^n$ then $V^\perp = \{0\}.$ If $V = \{0\} \subseteq \mathbb{R}^n$ then $V^\perp = \mathbb{R}^n.$