Linear Algebra Notes — Lecture 11/3

Orthogonal Projections, the Gram–Schmidt Process and Orthogonal Matrices

Lecture pages 273–288 (draft) · Lay §6.2–6.4

Contents

Overview 1

1 Orthogonal projection: definition and first examples (pp. 273–274) 2

2 The matrix of a projection: UU^T (p. 275) 3

3 Image and kernel of P (pp. 276–277) 4

4 $P^2 = P$, and $x - P(x)$ is in $W^\perp$ (p. 277) 5

5 $P(x)$ is the closest point of W to x (p. 278) 6

6 Splitting x into W and $W^\perp$ parts (p. 279) 7

7 The Gram–Schmidt process (pp. 280–283) 8

8 Orthogonal matrices (pp. 284–285) 10

9 Proof of the equivalences (pp. 285–287) 11

10 Products, transposes and inverses (p. 288) 12

Summary 13

Corrections to the handwritten notes 14

Overview

This lecture uses orthonormal vectors in three ways. The plan on p. 273 is:

1. **Orthogonal projections:** the map P that sends x to the closest point of a subspace W. Its matrix is UU^T, where the columns of U are an orthonormal basis of W.
2. **The Gram–Schmidt process:** it turns an independent set of vectors into an orthonormal one.

So every subspace has an orthonormal basis.

3. **Linear maps that preserve the dot product:** these are the orthogonal matrices, $A^{-1} = A^T$.

Every blank and question in the notes is answered in a blue box. Typos are listed at the end.

Notation used throughout.

- $u \cdot v$, $\|u\|$, $W^\perp$ and "orthonormal" are as in the 10/29 notes (pp. 257–272).
- P is the orthogonal projection onto W; the notes also write $\text{proj}_W(x)$ for $P(x)$.
- $\text{Im}\, P$ and $\text{Ker}\, P$ are the image and the kernel of P.
- x_W and $x_{W^\perp}$ are the parts of x in W and in $W^\perp$ (p. 279).

1 Orthogonal projection: definition and first examples (pp. 273–274)

Projecting onto W means adding up, over an orthonormal basis of W, the part of x along each basis vector.

Definition (orthogonal projection, p. 273)

Let W be a subspace of $\mathbb{R}^n$ and let $u_1, \ldots, u_k$ be an orthonormal basis of W. The **orthogonal projection onto** W is the map $P : \mathbb{R}^n \to \mathbb{R}^n$,

$$P(x) = u_1(u_1 \cdot x) + \cdots + u_k(u_k \cdot x).$$

Each $u_i \cdot x$ is a number, so $P(x)$ is a linear combination of $u_1, \ldots, u_k$, and therefore lies in W. For x in W, this is the coordinate formula of p. 271 (10/29 notes).

Note (p. 273). We shall see later that P depends only on W, and not on the choice of $u_1, \ldots, u_k$. This is proved on p. 278.

Ex (p. 273): $n = 2$, $k = 1$, $u_1 = e_1$

Fill-in: W is the x_1-axis, $P(x) = (x_1, 0)$, and the standard matrix is $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

$W = \text{span}\{e_1\}$ is the x_1-axis. Since $x \cdot e_1 = x_1$,

$$P(x) = e_1(x \cdot e_1) = x_1 e_1 = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}.$$

So P drops x straight down onto the x_1-axis. The columns of the standard matrix are $P(e_1) = e_1$ and $P(e_2) = 0$, which gives the matrix above.

Check: $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$. ✓

Ex (p. 274): $\{e_1, \ldots, e_k\}$ is a basis of $W \subseteq \mathbb{R}^n$

Fill-in: $P(x) = (x_1, \ldots, x_k, 0, \ldots, 0)$, and the standard matrix is $\begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$

Since $x \cdot e_i = x_i$,

$$P(x) = e_1(x \cdot e_1) + \cdots + e_k(x \cdot e_k) = x_1 e_1 + \cdots + x_k e_k = (x_1, \ldots, x_k, 0, \ldots, 0).$$

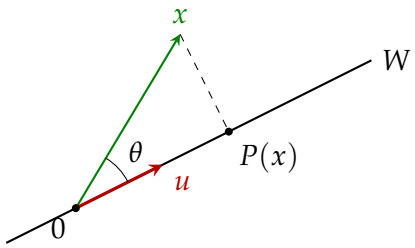
So P keeps the first k entries and sets the others to 0. The columns of the standard matrix are $P(e_i) = e_i$ for $i \leq k$ and $P(e_i) = 0$ for $i > k$. This is the diagonal matrix with k ones followed by $n - k$ zeros (I_k is the $k \times k$ identity).
Check: $k = 1$, $n = 2$ gives the matrix of the previous example. ✓

Ex (p. 274): u a unit vector in $\mathbb{R}^n$, $W = \operatorname{span}\{u\}$

The notes use the angle formula $x \cdot u = \|x\| \, \|u\| \cos\theta$ (definition on p. 263), with $\|u\| = 1$:

$$P(x) = u(x \cdot u) = u \, \|x\| \, \|u\| \cos\theta = u \, \|x\| \cos\theta.$$

Meaning: $\|x\| \cos\theta$ is the adjacent side of the right triangle in the picture, so $P(x)$ is the shadow of x on the line W. If $\theta > \pi/2$, then $\cos\theta < 0$ and the shadow points opposite to u. In the picture below, the dashed segment from x to $P(x)$ is perpendicular to W.



2 The matrix of a projection: UU^T (p. 275)

Put the orthonormal basis in the columns of U; then the projection is multiplication by UU^T.

Theorem (p. 275)

If $u_1, \ldots, u_k$ is an orthonormal basis of $W \subseteq \mathbb{R}^n$ and $U = [u_1 \ \cdots \ u_k]$ ($n \times k$), then UU^T is the standard matrix of P, the orthogonal projection onto W.

Proof (p. 275). The product $U^T x$ has entries $u_i^T x = u_i \cdot x$. Then U times this column is the

combination of the columns of U with these weights (Lay §1.4):

$$UU^Tx = U\begin{bmatrix} u_1^T \\ \vdots \\ u_k^T \end{bmatrix} x = U\begin{bmatrix} u_1 \cdot x \\ \vdots \\ u_k \cdot x \end{bmatrix} = [u_1 \ \cdots \ u_k]\begin{bmatrix} u_1 \cdot x \\ \vdots \\ u_k \cdot x \end{bmatrix} = u_1(u_1 \cdot x) + \cdots + u_k(u_k \cdot x) = P(x). \ \blacksquare$$

In particular, P is linear, because it is multiplication by a matrix.

Ex (p. 275): $W = \operatorname{span}\{u\}$, u a unit vector

Here $U = [u]$ is a single column, so $P(x) = uu^Tx$.

Fill-in: $u = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ **gives** $uu^T = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$**, and** $u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ **gives** $uu^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

A column times a row is a matrix: entry (i, j) is u_iu_j. So

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}[0 \ \ 1] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 \\ 0 \end{bmatrix}[1 \ \ 0] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

The first is the projection onto the x_2-axis, the second the projection onto the x_1-axis (as on p. 273). (For $u = (0, 1)$, the notes write the row as $[1 \ \ 0]$; it should be $[0 \ \ 1]$, see the corrections.)

Check: $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \end{bmatrix} = u(x \cdot u)$, since $x \cdot u = 5$. ✓

Don't confuse UU^T with U^TU. Entry (i, j) of U^TU is $u_i \cdot u_j$, so $U^TU = I_k$ always. But UU^T is $n \times n$ and equals I_n only when $k = n$. Otherwise it is the projection onto W, and it is not invertible.

3 Image and kernel of P (pp. 276–277)

P fixes W and kills $W^\perp$.

Theorem (p. 276)

If P is the orthogonal projection onto W, then

① $\operatorname{Im} P = W$;

② if x is in W, then $P(x) = x$;

③ $\operatorname{Ker} P = W^\perp$.

Proof of ②, blanks filled in: $P(u_i) = u_i$**, so** $P(x) = c_1u_1 + \cdots + c_ku_k = x$

Write $x = c_1u_1 + \cdots + c_ku_k$. First apply P to one basis vector. By orthonormality, $u_j \cdot u_i = 0$ for $j \neq i$ and $u_i \cdot u_i = 1$, so only the term $j = i$ survives:

$$P(u_i) = \sum_{j=1}^{k} u_j(u_j \cdot u_i) = u_i.$$

Since P is linear,

$$P(x) = c_1P(u_1) + \cdots + c_kP(u_k) = c_1u_1 + \cdots + c_ku_k = x. \blacksquare$$

Proof of ① (p. 276). $P(x)$ is in W for any x, so $\operatorname{Im} P \subseteq W$. Conversely, let x be in W. Then $P(x) = x$ by ②, so x is in $\operatorname{Im} P$. So $W \subseteq \operatorname{Im} P$. $\blacksquare$

Proof of ③, blanks filled in: $\dim W^\perp = n - k$**,** $\dim \operatorname{Ker} P = n - k$**, and** $W^\perp \subseteq \operatorname{Ker} P$

Let $k = \dim W$.

- $\dim W^\perp = n - k$ (theorem on p. 266).
- $\dim \operatorname{Ker} P = n - \dim \operatorname{Im} P = n - k$ (Rank Theorem, Lay §4.6; and $\operatorname{Im} P = W$ by ①).

The two subspaces have the same dimension, so it is enough to show $W^\perp \subseteq \operatorname{Ker} P$ (p. 277). A subspace contained in another of the same dimension equals it, as on p. 270.
Take x in $W^\perp$. Each u_i is in W, so $x \cdot u_i = 0$ for $1 \leq i \leq k$. So

$$P(x) = u_1(u_1 \cdot x) + \cdots + u_k(u_k \cdot x) = u_1 \cdot 0 + \cdots + u_k \cdot 0 = 0. \blacksquare$$

4 $P^2 = P$, and $x - P(x)$ is in $W^\perp$ (p. 277)

Projecting twice changes nothing, and what is left over after projecting is orthogonal to W.

Theorem (p. 277)

① $P^2 = P$;

② $x - P(x)$ is in $W^\perp$, for every x in $\mathbb{R}^n$.

Proof of ①, blank filled in: $P^2(x) = u_1(x \cdot u_1) + \cdots + u_k(x \cdot u_k) = P(x)$

By linearity of P (the $x \cdot u_i$ are numbers), and then $P(u_i) = u_i$ (p. 276):

$$\begin{aligned} P^2(x) = P\big(u_1(x \cdot u_1) + \cdots + u_k(x \cdot u_k)\big) &= P(u_1)(x \cdot u_1) + \cdots + P(u_k)(x \cdot u_k) \\ &= u_1(x \cdot u_1) + \cdots + u_k(x \cdot u_k) = P(x). \blacksquare \end{aligned}$$

(Shorter: $P(x)$ is in W, and P fixes W.)

Proof of ②, blanks filled in: $P(x - P(x)) = 0$, so $x - P(x)$ is in $\operatorname{Ker} P = W^{\perp}$

By linearity and ①, $P(x - P(x)) = P(x) - P^2(x) = P(x) - P(x) = 0$. So $x - P(x)$ is in $\operatorname{Ker} P$, which is $W^{\perp}$ by the theorem on p. 276. ■

Meaning: in the picture below, x splits into $P(x)$, which lies in W, plus a piece perpendicular to W.

5 $P(x)$ is the closest point of W to x (p. 278)

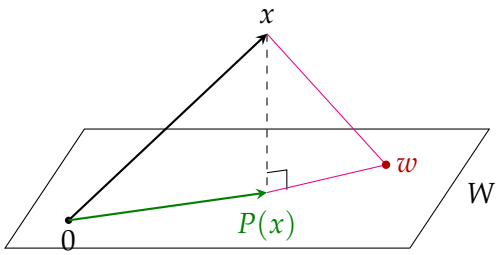
Among all points of W, the projection $P(x)$ is strictly the closest to x. This shows P depends only on W.

Theorem (p. 278)

$\|P(x) - x\| < \|w - x\|$ for every w in W and x in $\mathbb{R}^n$ with $w \neq P(x)$.

In other words, $P(x)$ is the unique point of W that is closer to x than any other w in W. **By uniqueness, P is independent of the choice of $u_1, \ldots, u_k$:** the closest point is described using only W and x. This settles the note on p. 273.

The picture shows W as a plane, with x above it. The right angle is at $P(x)$, and w is another point of W.



Proof (p. 278). Add and subtract $P(x)$:

$$\|w - x\|^2 = \| \underbrace{w - P(x)}_{\text{in } W} + \underbrace{P(x) - x}_{\text{in } W^{\perp}} \|^2 = \|w - P(x)\|^2 + \|P(x) - x\|^2 > \|P(x) - x\|^2.$$

Here $w - P(x)$ is in W because both terms are, and $P(x) - x$ is in $W^{\perp}$ by the theorem on p. 277. So the two pieces are orthogonal, and the second equality is the Pythagorean theorem (p. 262). The strict inequality holds because $\|w - P(x)\|^2 \neq 0$, as $w \neq P(x)$. ■

Worked example (addition)

Let $W = \text{span}\{(1,1,0),(1,0,1)\}$ in $\mathbb{R}^3$. An orthonormal basis, computed by Gram–Schmidt in Section 7, is $u_1 = \frac{1}{\sqrt{2}}(1,1,0)$, $u_2 = \frac{1}{\sqrt{6}}(1,-1,2)$. By the theorem on p. 275,

$$P = UU^T = \frac{1}{2}\begin{bmatrix}1&1&0\\1&1&0\\0&0&0\end{bmatrix} + \frac{1}{6}\begin{bmatrix}1&-1&2\\-1&1&-2\\2&-2&4\end{bmatrix} = \frac{1}{3}\begin{bmatrix}2&1&1\\1&2&-1\\1&-1&2\end{bmatrix}.$$

For $x = (3,0,0)$, the first column gives $P(x) = (2,1,1)$. So $x - P(x) = (1,-1,-1)$.

Check: $(1,-1,-1)$ has dot product 0 with $(1,1,0)$ and with $(1,0,1)$, so it is in $W^\perp$. The other basis $\frac{1}{\sqrt{6}}(2,1,1)$, $\frac{1}{\sqrt{2}}(0,1,-1)$ of W gives the same P: its $(1,1)$ entry is $\frac{4}{6} + 0 = \frac{2}{3}$. And the distance $\|x - P(x)\| = \sqrt{3}$ is less than the distance $\sqrt{5}$ from x to the point $w = (1,1,0)$ of W. ✓

6 Splitting x into W and $W^\perp$ parts (p. 279)

A vector splits in exactly one way as a part in W plus a part in $W^\perp$, and the parts are $P(x)$ and $x - P(x)$.

Theorem (p. 279)

If $x = x_W + x_{W^\perp}$, where x_W is in W and $x_{W^\perp}$ is in $W^\perp$, then $x_W = P(x)$ and $x_{W^\perp} = x - P(x)$.

Proof, blank filled in: both sides are in $W \cap W^\perp = \{0\}$, so $x_W = P(x)$ and $x_{W^\perp} = x - P(x)$

Write x in two ways, using the theorem on p. 277 for the second:

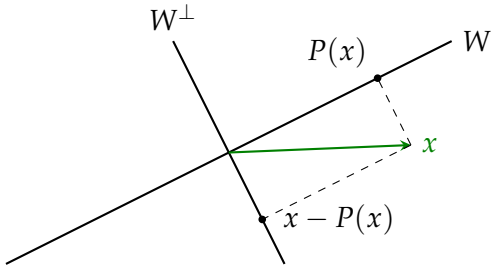
$$x_W + x_{W^\perp} = x = \underbrace{x - P(x)}_{\text{in } W^\perp} + \underbrace{P(x)}_{\text{in } W}.$$

Rearranging,

$$\underbrace{x_W - P(x)}_{\text{in } W} = \underbrace{x - P(x) - x_{W^\perp}}_{\text{in } W^\perp}.$$

This vector is in both W and $W^\perp$, so it is 0 (note on p. 267). Hence $x_W = P(x)$ and $x_{W^\perp} = x - P(x)$. ■

The picture shows the two perpendicular subspaces, and x as the sum of its parts along them.



7 The Gram–Schmidt process (pp. 280–283)

Take the vectors one at a time. From each, subtract its projection onto what has been built so far, then normalize.

Gram–Schmidt process (p. 280)

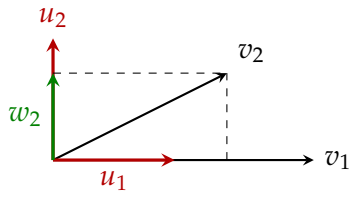
Input: an independent set of vectors $\{v_1, \dots, v_k\}$.
Output: an orthonormal basis $\{u_1, \dots, u_k\}$ of $\operatorname{span}\{v_1, \dots, v_k\}$ such that

$$\operatorname{span}\{u_1, \dots, u_\ell\} = \operatorname{span}\{v_1, \dots, v_\ell\} \quad \text{for all } \ell = 1, \dots, k.$$

Idea in $\mathbb{R}^2$ (p. 280). Normalize v_1 to get u_1. Then remove from v_2 its part along u_1, and normalize what remains:

$$u_1 = \frac{v_1}{\|v_1\|}, \qquad w_2 = v_2 - u_1(v_2 \cdot u_1) = v_2 - \operatorname{proj}_{\operatorname{span}\{u_1\}}(v_2), \qquad u_2 = \frac{w_2}{\|w_2\|}.$$

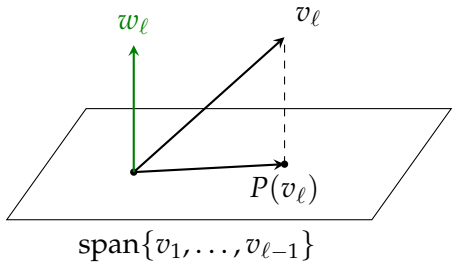
In the picture, w_2 is the part of v_2 perpendicular to v_1. The vectors u_1 and u_2 have length 1, while w_2 may be shorter or longer.



Inductive idea in $\mathbb{R}^n$ (p. 281). Suppose $\{u_1, \dots, u_{\ell-1}\}$ is orthonormal in $\operatorname{span}\{v_1, \dots, v_{\ell-1}\}$. Then let

$$w_\ell = v_\ell - \operatorname{proj}_{\operatorname{span}\{v_1, \dots, v_{\ell-1}\}}(v_\ell), \qquad u_\ell = \frac{w_\ell}{\|w_\ell\|}.$$

(The notes' indices are one too high here; see the corrections.) The picture shows the plane $\operatorname{span}\{v_1, \dots, v_{\ell-1}\}$; w_ℓ is the part of v_ℓ perpendicular to it.



Algorithm (pp. 281–282)

Input $\{v_1, \dots, v_k\}$.
Step 1: $u_1 = \dfrac{v_1}{\|v_1\|}$.

Step 2: $w_2 = v_2 - u_1(v_2 \cdot u_1), \quad u_2 = \dfrac{w_2}{\|w_2\|}.$

Step ℓ: $w_\ell = v_\ell - u_1(v_\ell \cdot u_1) - \cdots - u_{\ell-1}(v_\ell \cdot u_{\ell-1}), \quad u_\ell = \dfrac{w_\ell}{\|w_\ell\|}.$

The formula in step ℓ is the definition of the projection on p. 273, for the orthonormal basis $u_1, \ldots, u_{\ell-1}$.

Worked example (addition): $v_1 = (1,1,0)$, $v_2 = (1,0,1)$

Step 1: $u_1 = \frac{1}{\sqrt{2}}(1,1,0)$. **Step 2:** $v_2 \cdot u_1 = \frac{1}{\sqrt{2}}$, so

$$w_2 = (1,0,1) - \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{\sqrt{2}}(1,1,0) = \left(\tfrac{1}{2}, -\tfrac{1}{2}, 1\right).$$

Then $\|w_2\|^2 = \frac{1}{4} + \frac{1}{4} + 1 = \frac{3}{2}$, so $u_2 = w_2/\|w_2\| = \frac{1}{\sqrt{6}}(1,-1,2)$. *Check:* $u_1 \cdot u_2 = \frac{1}{\sqrt{12}}(1-1+0) = 0$ and $\|u_2\|^2 = \frac{1+1+4}{6} = 1$. ✓

Need to justify for each ℓ (p. 282):

① $w_\ell \neq 0$, so that we can divide by $\|w_\ell\|$;

② $\operatorname{span}\{u_1, \ldots, u_\ell\} = \operatorname{span}\{v_1, \ldots, v_\ell\}$.

The argument is by induction on ℓ: assume the claims for smaller values. Also, u_ℓ is orthogonal to the earlier u_j: by the theorem on p. 277, w_ℓ is in $\operatorname{span}\{u_1, \ldots, u_{\ell-1}\}^\perp$, and u_ℓ is a multiple of w_ℓ (the notes leave this to that theorem).

① Why can't w_ℓ be 0? Blank filled in: it would contradict the independence of $v_1, \ldots, v_\ell$

If $0 = w_\ell = v_\ell - u_1(v_\ell \cdot u_1) - \cdots - u_{\ell-1}(v_\ell \cdot u_{\ell-1})$, then v_ℓ is in $\operatorname{span}\{u_1, \ldots, u_{\ell-1}\}$, which by the induction assumption is $\operatorname{span}\{v_1, \ldots, v_{\ell-1}\}$. But the set $\{v_1, \ldots, v_k\}$ is independent, so v_ℓ is not a combination of $v_1, \ldots, v_{\ell-1}$. This is a contradiction.
(For $\ell = 1$: $v_1 \neq 0$, since an independent set does not contain 0.)

② Why $\operatorname{span}\{v_1, \ldots, v_\ell\} = \operatorname{span}\{u_1, \ldots, u_\ell\}$ (p. 283).

"$\subseteq$": By induction, assume $\operatorname{span}\{v_1, \ldots, v_{\ell-1}\} = \operatorname{span}\{u_1, \ldots, u_{\ell-1}\}$. By the formula for w_ℓ, v_ℓ is in $\operatorname{span}\{u_1, \ldots, u_{\ell-1}, w_\ell\}$. Since $w_\ell = \|w_\ell\| u_\ell$, that span is $\operatorname{span}\{u_1, \ldots, u_\ell\}$.

"$\supseteq$": $u_\ell = w_\ell/\|w_\ell\|$ is in $\operatorname{span}\{u_1, \ldots, u_{\ell-1}, v_\ell\}$, by the formula for w_ℓ. By the induction assumption, this is contained in $\operatorname{span}\{v_1, \ldots, v_\ell\}$. ■

Theorem (p. 283)

Every subspace W of $\mathbb{R}^n$ has an orthonormal basis.

Proof. Start with any basis of W (Lay §4.5) and hit it with Gram–Schmidt. The output is orthonormal and has the same span, W. Orthonormal sets are independent (p. 271), so it is a basis. (For $W = \{0\}$, the basis is empty.) ■

8 Orthogonal matrices (pp. 284–285)

The real square matrices that preserve the dot product are exactly those with orthonormal columns.

Q (p. 284). What can we say about a real $n \times n$ matrix A with $Ax \cdot Ay = x \cdot y$ for all x, y in $\mathbb{R}^n$?

Ex (p. 284): take $x = e_i$, $y = e_j$

Fill-in: $Ae_i \cdot Ae_j = e_i \cdot e_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$, **so the columns of A are orthonormal**

By the hypothesis, $Ae_i \cdot Ae_j = e_i \cdot e_j$, and this is 1 if $i = j$ and 0 otherwise (p. 258). Recall that Ae_i is column i of A. So the columns of A are orthonormal: this is the definition on p. 259.

Ex (p. 284): rotation and reflection

$A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ is a rotation by θ. (It turns e_1 to $(\cos\theta, -\sin\theta)$, so the rotation is clockwise.)

$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ is the reflection in the x_1-axis. Rotations and reflections preserve lengths and angles, so both should have orthonormal columns.

Check: rotation: $\cos\theta \sin\theta - \sin\theta \cos\theta = 0$, and each column has length $\sqrt{\cos^2\theta + \sin^2\theta} = 1$. Reflection: the columns are e_1 and $-e_2$. ✓

Definition (orthogonal matrix, p. 285)

An $n \times n$ matrix A is **orthogonal** if its columns are orthonormal.

(The name is standard, though "orthonormal matrix" would be more accurate. It applies only to square matrices.)

Theorem (p. 285)

For an $n \times n$ real matrix A, the following are equivalent:

(a) $A^{-1} = A^T$;

(b) $\|Ax\| = \|x\|$ for all x in $\mathbb{R}^n$;

(c) $Ax \cdot Ay = x \cdot y$ for all x, y in $\mathbb{R}^n$;

(d) the columns of A are orthonormal.

9 Proof of the equivalences (pp. 285–287)

We show (a) $\iff$ (d), then (c) $\Rightarrow$ (b), (c) $\Rightarrow$ (d), (a) $\Rightarrow$ (c), and finally (b) $\Rightarrow$ (c).

(a) $\iff$ (d) (pp. 285–286). (a) says $A^TA = I$. (For a square matrix, one-sided inverses are enough: Invertible Matrix Theorem, Lay §2.3.) If $A = [a_1 \ \cdots \ a_n]$, then

$$A^TA = \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} [a_1 \ \cdots \ a_n] = I \iff \text{the } (i,j) \text{ entry of } A^TA \text{ is } \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

That entry is $a_i^Ta_j = a_i \cdot a_j$. So $A^TA = I$ says exactly that the columns are orthonormal. Thus (a) $\iff$ (d).

(c) $\Rightarrow$ (b) because: take $y = x$

With $y = x$, (c) gives $\|Ax\|^2 = Ax \cdot Ax = x \cdot x = \|x\|^2$. Lengths are ≥ 0, so $\|Ax\| = \|x\|$.

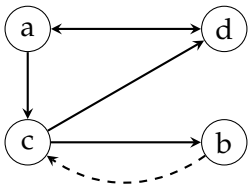
(c) $\Rightarrow$ (d) because: take $x = e_i$, $y = e_j$

This is the example on p. 284: $a_i \cdot a_j = Ae_i \cdot Ae_j = e_i \cdot e_j$, which is 1 if $i = j$ and 0 otherwise.

(a) $\Rightarrow$ (c) (p. 286). Write the dot product as a matrix product, and use (a) in the form $A^TA = I$:

$$Ax \cdot Ay = (Ax)^TAy = x^TA^TAy \overset{(a)}{=} x^Ty = x \cdot y.$$

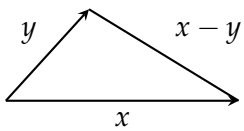
In summary (p. 287). So far we have the solid arrows below. It remains to show (b) $\Rightarrow$ (c), the dashed arrow.



(b) $\Rightarrow$ (c) (p. 287). That is, we must show that

$$\|Ax\| = \|x\| \text{ for all } x \quad \Longrightarrow \quad Ax \cdot Ay = x \cdot y \text{ for all } x, y.$$

The left side is about lengths; the right side is about lengths and angles. **Idea: lengths determine angles.** A triangle is determined by its three side lengths, so its angles are too. In the picture, the sides are x, y and $x - y$.



Proof, finishing the notes' computation: apply (b) to $x - y$ and cancel

By (b) applied to $x - y$, and because $A(x - y) = Ax - Ay$,

$$A(x - y) \cdot A(x - y) = (x - y) \cdot (x - y).$$

Expand both sides as on p. 260:

$$Ax \cdot Ax - 2\,Ax \cdot Ay + Ay \cdot Ay = x \cdot x - 2\,x \cdot y + y \cdot y.$$

By (b), $Ax \cdot Ax = \|Ax\|^2 = \|x\|^2 = x \cdot x$, and likewise $Ay \cdot Ay = y \cdot y$. Cancel these and divide by -2: $Ax \cdot Ay = x \cdot y$. ■

10 Products, transposes and inverses (p. 288)

Orthogonal matrices are closed under products, transposes and inverses.

Theorem (p. 288)

If A, B are orthogonal $n \times n$ matrices, then so are AB, A^T and A^{-1}.

Proof (p. 288). We use condition (b) of the theorem on p. 285. Since $A^{-1} = A^T$ by (a), it is enough to treat AB and A^{-1}.

AB: $\|ABx\| = \|Bx\| = \|x\|$, using (b) for A and then for B.

A^{-1}: given x, write $x = Ay$ (possible because A is invertible). Then

$$\|A^{-1}x\| = \|A^{-1}Ay\| = \|y\| = \|Ay\| = \|x\|. \blacksquare$$

Check (addition): Take the rotation with $\theta = \pi/2$ times the reflection on p. 284:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix},$$

and the columns $(0, -1)$, $(-1, 0)$ are orthonormal. ✓

Draft. These notes were made from the draft of the 11/3 notes; most blanks are empty there. They will be updated when the final version arrives.

Summary

Fact	Formula
Orthogonal projection	$P(x) = u_1(u_1 \cdot x) + \cdots + u_k(u_k \cdot x) = UU^T x$
Image and kernel	$\operatorname{Im} P = W$, $\operatorname{Ker} P = W^\perp$; $P(x) = x$ for x in W
Properties	$P^2 = P$; $x - P(x)$ is in $W^\perp$
Closest point	$\|P(x) - x\| < \|w - x\|$ for w in W, $w \neq P(x)$
Decomposition	$x = P(x) + (x - P(x))$ is the only split into W and $W^\perp$ parts
Gram–Schmidt	$w_\ell = v_\ell - \sum_{j<\ell} u_j(v_\ell \cdot u_j)$, $u_\ell = w_\ell / \|w_\ell\|$
Orthogonal matrix	$A^{-1} = A^T \iff \|Ax\| = \|x\| \iff Ax \cdot Ay = x \cdot y \iff$ orthonormal columns
Closure	AB, A^T, A^{-1} are orthogonal if A, B are

Corrections to the handwritten notes

Page	As written	Should be	Why
275	"$u_1, \dots, u_k$ is orthonormal in W"	"is an orthonormal basis of W"	otherwise UU^T projects onto a smaller subspace
275	$u = (0,1)$: $uu^T = \begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}0 & 1\end{bmatrix}$	u^T has the entries of u
276	$P(u_i) = \sum_j u_i(u_i \cdot u_j)$	$\sum_j u_j(u_j \cdot u_i)$	the sum runs over the basis vectors u_j (same value, u_i)
278	"closer to x than any w in W"	"than any other w in W"	$P(x)$ is itself in W
280	Output: basis in $\text{span}\{u_1, \dots, u_k\}$	in $\text{span}\{v_1, \dots, v_k\}$	the u_i are what we are building
281	$\{u_1, \dots, u_\ell\}$ in $\text{span}\{v_1, \dots, v_\ell\}$; $w_\ell = v_\ell - \text{proj}_{\text{span}\{v_1,\dots,v_\ell\}}(v_\ell)$	index $\ell - 1$ in all three places	as written, v_ℓ is in the span, so $w_\ell = 0$
281	Step 2: $w_2 = v_2 - u_1(v_1 \cdot u_1)$	$v_2 - u_1(v_2 \cdot u_1)$	see p. 280 and step ℓ
283	$u_\ell = \|w_\ell\|\, w_\ell$	$u_\ell = w_\ell / \|w_\ell\|$	definition on p. 282
284	"rotation by θ"	clockwise rotation by θ	$Ae_1 = (\cos\theta, -\sin\theta)$ (fine if θ is measured clockwise)