

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 *inner product* or *dot product* of two vectors  $u, v \in \mathbb{R}^n$  is the scalar

$$u \bullet v = u_1v_1 + u_2v_2 + \cdots + u_nv_n = u^\top v = v^\top u \in \mathbb{R}^1 = \mathbb{R}.$$

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

- 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

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

- A set of nonzero vectors  $v_1, v_2, \dots, v_p \in \mathbb{R}^n$  is *orthogonal* if  $v_i \bullet v_j = 0$  for all  $i \neq j$ .

Any such set is automatically linearly independent and therefore a basis for a subspace.

- An orthogonal basis is *orthonormal* if it consists entirely of unit vectors.

If  $u_1, u_2, \dots, u_n \in \mathbb{R}^n$  are orthonormal and  $U = [u_1 \ u_2 \ \dots \ u_n]$  then  $U^\top U = I_n$ .

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

This occurs if and only if the columns of  $U$  are orthonormal.

- Any subspace  $V \subseteq \mathbb{R}^n$  has an orthogonal basis.

Any subspace  $V \subseteq \mathbb{R}^n$  therefore also has an orthonormal basis.

If  $u_1, u_2, \dots, u_p$  is an orthogonal basis for  $V$  then the *projection* of  $y \in \mathbb{R}^n$  onto  $V$  is the vector

$$\text{proj}_V(y) = \frac{y \bullet u_1}{u_1 \bullet u_1} u_1 + \frac{y \bullet u_2}{u_2 \bullet u_2} u_2 + \cdots + \frac{y \bullet u_p}{u_p \bullet u_p} u_p \in V.$$

This formula does not depend on the choice of orthogonal basis for  $V$ .

The projection of  $y$  onto  $V$  is the unique vector in  $V$  such that  $y - \text{proj}_V(y) \in V^\perp$ .

The projection of  $y$  onto  $V$  is also characterized as the vector in  $V$  that is the shortest distance away from  $y$ . If  $v \in V$  and  $v \neq \text{proj}_V(y)$  then  $\|y - \text{proj}_V(y)\| < \|y - v\|$ .

- The *Gram-Schmidt process* is an algorithm that takes a basis  $x_1, x_2, \dots, x_p$  for a subspace of  $\mathbb{R}^n$  as input, and produces an orthogonal basis  $v_1, v_2, \dots, v_p$  of the same subspace as output.

The orthogonal basis  $v_1, v_2, \dots, v_p$  is defined from the input basis  $x_1, x_2, \dots, x_p$  by these formulas:

$$\begin{aligned} v_1 &= x_1. \\ v_2 &= x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1. \\ v_3 &= x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2. \\ v_4 &= x_4 - \frac{x_4 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_4 \bullet v_2}{v_2 \bullet v_2} v_2 - \frac{x_4 \bullet v_3}{v_3 \bullet v_3} v_3. \\ &\vdots \\ v_p &= x_p - \frac{x_p \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_p \bullet v_2}{v_2 \bullet v_2} v_2 - \cdots - \frac{x_p \bullet v_{p-1}}{v_{p-1} \bullet v_{p-1}} v_{p-1}. \end{aligned}$$

# 1 Last time: inner products and orthogonality

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  $u \bullet v = u_1v_1 + u_2v_2 + \cdots + u_nv_n = u^\top v = v^\top u = v \bullet u$ .

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

A vector with length 1 is a *unit vector*. Note that  $\|v\|^2 = v \bullet v$ .

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

In  $\mathbb{R}^2$ , two vectors are orthogonal if and only if they belong to perpendicular lines through the origin.

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

The *orthogonal complement* of a subspace  $V \subseteq \mathbb{R}^n$  is the subspace  $V^\perp$  whose elements are the vectors  $w \in \mathbb{R}^n$  such that  $w \bullet v = 0$  for all  $v \in V$ .

The only vector that is in both  $V$  and  $V^\perp$  is the zero vector.

We have  $\{0\}^\perp = \mathbb{R}^n$  and  $(\mathbb{R}^n)^\perp = \{0\}$ . If  $A$  is an  $m \times n$  matrix then  $(\text{Col } A)^\perp = \text{Nul}(A^\top)$ .

We also showed last time that  $\dim V + \dim V^\perp \leq n$ .

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

**Theorem.** Any list of orthogonal nonzero vectors is linearly independent and so is an orthogonal basis of the subspace it spans.

*Second proof.* Suppose  $u_1, u_2, \dots, u_p \in \mathbb{R}^n$  are orthogonal and nonzero.

$$\text{Let } A = [u_1 \quad u_2 \quad \cdots \quad u_p] \text{ and } d_i = u_i \bullet u_i > 0 \text{ and } D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_p \end{bmatrix}.$$

Check that  $A^\top A = D$ . Our vectors are linearly dependent if and only if  $Ax = 0$  has a nonzero solution. This is impossible since if  $Ax = 0$  then  $A^\top Ax = 0$  which implies  $x = 0$  since  $A^\top A = D$  is invertible.  $\square$

If  $u_1, u_2, \dots, u_p$  is an orthogonal basis for a subspace  $V \subseteq \mathbb{R}^n$  and  $y \in V$ , then

$$y = c_1u_1 + c_2u_2 + \cdots + c_pu_p \quad \text{where } c_i = \frac{y \bullet u_i}{u_i \bullet u_i} \in \mathbb{R}.$$

This is a very useful property of orthogonal bases.

In general, to determine the coefficients that express a vector in a given basis, we have to solve an entire linear system. For orthogonal bases, we can just compute inner products.

**Example.** Let's work through this statement for the standard orthogonal basis  $e_1, e_2, \dots, e_n$  for  $\mathbb{R}^n$ . If

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y_1 e_1 + y_2 e_2 + \dots + y_n e_n$$

then  $y = c_1 e_1 + c_2 e_2 + \dots + c_n e_n$  where  $c_i = \frac{y \bullet e_i}{e_i \bullet e_i}$ . But  $e_i \bullet e_i = 1$  and  $y \bullet e_i = y_i$ , so we just have  $c_i = y_i$ .

## 2 Orthogonal projection onto a line

Let  $L \subseteq \mathbb{R}^n$  be a one-dimensional subspace.

Then  $L = \mathbb{R}\text{-span}\{u\}$  for any nonzero vector  $u \in L$ .

Let  $y \in \mathbb{R}^n$ . The *orthogonal projection* of  $y$  onto  $L$  is the vector

$$\text{proj}_L(y) = \frac{y \bullet u}{u \bullet u} u \quad \text{for any } 0 \neq u \in L.$$

The value of  $\text{proj}_L(y)$  does not depend on the choice of the nonzero vector  $u$ .

The *component of  $y$  orthogonal to  $L$*  is the vector  $z = y - \text{proj}_L(y)$ .

**Proposition.** The only vector  $\hat{y} \in L$  with  $y - \hat{y} \in L^\perp$  is the orthogonal projection  $\hat{y} = \text{proj}_L(y)$ .

*Proof.* Let  $u \in L$  be nonzero. Then  $y - \text{proj}_L(y) = y - \frac{y \bullet u}{u \bullet u} u$  and it holds that

$$\left( y - \frac{y \bullet u}{u \bullet u} u \right) \bullet u = y \bullet u - \frac{y \bullet u}{u \bullet u} u \bullet u = y \bullet u - y \bullet u = 0.$$

This shows that  $y - \text{proj}_L(y) \in L^\perp$ , and clearly  $\text{proj}_L(y) \in L$ .

To see that  $\text{proj}_L(y)$  is the only vector in  $L$  with this property, suppose  $\hat{y} \in L$  is such that  $y - \hat{y} \in L^\perp$ .

Then  $(y - \hat{y}) \bullet \hat{y} = y \bullet \hat{y} - \hat{y} \bullet \hat{y} = 0$  so  $y \bullet \hat{y} = \hat{y} \bullet \hat{y}$ .

But  $\hat{y} = cu$  for some nonzero  $c \in \mathbb{R}$ .

So we have  $c(y \bullet u) = y \bullet cu = (cu) \bullet (cu) = c^2(u \bullet u)$ .

Thus  $c = \frac{y \bullet u}{u \bullet u}$  so  $\hat{y} = \text{proj}_L(y)$ . □

**Example.** If  $y = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$  and  $L = \mathbb{R}\text{-span}\left\{ \begin{bmatrix} 4 \\ 2 \end{bmatrix} \right\}$  then

$$\text{proj}_L(y) = \frac{\begin{bmatrix} 7 \\ 6 \end{bmatrix} \bullet \begin{bmatrix} 4 \\ 2 \end{bmatrix}}{\begin{bmatrix} 4 \\ 2 \end{bmatrix} \bullet \begin{bmatrix} 4 \\ 2 \end{bmatrix}} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \frac{28 + 12}{16 + 4} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}.$$

In  $\mathbb{R}^2$ , the distance from a point  $(x, y)$  to a line  $L = \mathbb{R}\text{-span}\{u\}$  is the length  $\left\| \begin{bmatrix} x \\ y \end{bmatrix} - \text{proj}_L \left( \begin{bmatrix} x \\ y \end{bmatrix} \right) \right\|$ .

**Example.** To find the distance from the point  $(x, y) = (7, 6)$  to the line  $L$  defined by  $y = \frac{1}{2}x$ , note that  $L$  contains the vector  $u = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ . Let  $w = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ . Then  $\text{proj}_L \left( \begin{bmatrix} 7 \\ 6 \end{bmatrix} \right) = \frac{w \bullet u}{u \bullet u} u = \frac{28 + 12}{16 + 4} u = \frac{40}{20} u = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

so the distance is  $\left\| \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} \right\| = \left\| \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\| = \sqrt{1 + 4} = \sqrt{5}$ .

### 3 Orthonormal vectors

A set of vectors  $u_1, u_2, \dots, u_p$  is *orthonormal* if the vectors are orthogonal and each vector is a unit vector. In other words, if  $u_i \bullet u_j = 0$  when  $i \neq j$  and  $u_i \bullet u_i = 1$  for all  $i$ .

An *orthonormal basis* of a subspace is a basis that is orthonormal.

**Confusing convention:** a square matrix with orthonormal columns is called an *orthogonal matrix*.

It would make more sense to call such a matrix an “orthonormal matrix” but the term “orthogonal matrix” is standard and widely used.

**Example.** The standard basis  $e_1, e_2, \dots, e_n$  is an orthonormal basis for  $\mathbb{R}^n$ .

**Example.** The vectors  $\frac{1}{\sqrt{11}} \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$ ,  $\frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$ , and  $\frac{1}{\sqrt{66}} \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix}$  are an orthonormal basis for  $\mathbb{R}^3$ .

**Theorem.** Let  $U$  be an  $m \times n$  matrix.

The columns of  $U$  are orthonormal vectors if and only if  $U^\top U = I_n$ .

If  $U$  is square then its columns are orthonormal if and only if  $U^\top = U^{-1}$ .

(In other words, a matrix  $U$  is *orthogonal* if and only if  $U$  is square and  $U^\top = U^{-1}$ .)

*Proof.* Suppose  $U = [u_1 \ u_2 \ \dots \ u_n]$  where each  $u_i \in \mathbb{R}^m$ .

The entry in position  $(i, j)$  of  $U^\top U$  is then  $u_i^\top u_j = u_i \bullet u_j$ .

Therefore  $u_i \bullet u_i = 1$  and  $u_i \bullet u_j = 0$  for all  $i \neq j$  if and only if  $U^\top U$  is the  $n \times n$  identity matrix. □

**Corollary.** If  $U$  is an orthogonal matrix then  $\det(U) \in \{-1, 1\}$ .

*Proof.* We have  $\det(U)^2 = \det(U^\top) \det(U) = \det(U^\top U) = \det(I) = 1$ . □

**Theorem.** Let  $U$  be an  $m \times n$  matrix with orthonormal columns. Suppose  $x, y \in \mathbb{R}^n$ . Then:

1.  $\|Ux\| = \|x\|$ .
2.  $(Ux) \bullet (Uy) = x \bullet y$ .
3.  $(Ux) \bullet (Uy) = 0$  if and only if  $x \bullet y = 0$ .

*Proof.* The first and third statements are special cases of the second since  $\|Ux\| = \|x\|$  if and only if  $(Ux) \bullet (Ux) = x \bullet x$ . The second statement holds since  $(Ux) \bullet (Uy) = x^\top U^\top U y = x^\top I y = x^\top y = x \bullet y$ . □

### 4 Orthogonal projections onto subspaces

We have already seen that if  $y \in \mathbb{R}^n$  and  $L \subseteq \mathbb{R}^n$  is a 1-dimensional subspace then  $y$  can be written uniquely as  $y = \hat{y} + z$  where  $\hat{y} \in L$  and  $z \in L^\perp$ . This generalizes to arbitrary subspaces as follows:

**Theorem.** Let  $W \subseteq \mathbb{R}^n$  be any subspace. Let  $y \in \mathbb{R}^n$ .

Then there are unique vectors  $\hat{y} \in W$  and  $z \in W^\perp$  such that  $y = \hat{y} + z$ .

If  $u_1, u_2, \dots, u_p$  is an orthogonal basis for  $W$  then

$$\hat{y} = \frac{y \bullet u_1}{u_1 \bullet u_1} u_1 + \frac{y \bullet u_2}{u_2 \bullet u_2} u_2 + \dots + \frac{y \bullet u_p}{u_p \bullet u_p} u_p \quad \text{and} \quad z = y - \hat{y}. \quad (*)$$

It doesn't matter which orthogonal basis is chosen for  $W$ ; this formula gives the same value for  $\hat{y}$  and  $z$ .

*Proof.* To prove the theorem, we need to assume that  $W$  has an orthogonal basis. This nontrivial fact will be proved later in this lecture. Choose one such basis  $u_1, u_2, \dots, u_p \in W$ .

Define  $\hat{y}$  by the given formula. Then  $\hat{y} \in W$  and  $y - \hat{y} \in W^\perp$  since for each  $i = 1, 2, \dots, p$  we have

$$(y - \hat{y}) \bullet u_i = y \bullet u_i - \frac{y \bullet u_i}{u_i \bullet u_i} u_i \bullet u_i = 0.$$

To show uniqueness, suppose  $y = \hat{u} + v$  where  $\hat{u} \in W$  and  $v \in W^\perp$ .

Since we already have  $y = \hat{y} + z$ , we must have  $\hat{u} - \hat{y} = z - v$ . But  $\hat{u} - \hat{y}$  is in  $W$  while  $z - v$  is in  $W^\perp$ , so both expressions must be zero as  $W \cap W^\perp = \{0\}$ . This means we must have  $\hat{u} = \hat{y}$  and  $v = z$ .  $\square$

**Definition.** The vector  $\hat{y}$ , defined relative to  $y$  and  $W$  by the formula (\*) in the preceding theorem, is the *orthogonal projection* of  $y$  onto  $W$ . From now on we will write  $\boxed{\text{proj}_W(y) = \hat{y}}$  to refer to this vector.

**Corollary.** If  $W \subseteq \mathbb{R}^n$  is any subspace then  $\dim W^\perp = n - \dim W$ .

*Proof.* The preceding theorem shows that  $W$  and  $W^\perp$  together span  $\mathbb{R}^n$ . Therefore the union of any basis for  $W$  with a basis for  $W^\perp$  also spans  $\mathbb{R}^n$ .

The size of such a union is at most  $\dim W + \dim W^\perp$  and at least  $n$ , so  $n \leq \dim W + \dim W^\perp$ . This means that  $\dim W^\perp \geq n - \dim W$ . We showed last time that  $\dim W^\perp \leq n - \dim W$ , so  $\dim W^\perp = n - \dim W$ .  $\square$

Properties of orthogonal projections onto a subspace  $W \subseteq \mathbb{R}^n$ .

**Fact.** If  $y \in W$  then  $\text{proj}_W(y) = y$ . If  $y \in W^\perp$  then  $\text{proj}_W(y) = 0$ .

**Proposition.** If  $v \in W$  and  $y \in \mathbb{R}^n$  and  $v \neq \text{proj}_W(y)$  then  $\|y - \text{proj}_W(y)\| < \|y - v\|$ .

In words:  $\boxed{\text{the projection } \text{proj}_W(y) \text{ is the vector in } W \text{ that is closest to } y.}$

*Proof.* Let  $\hat{y} = \text{proj}_W(y)$ . Then  $y - v = (y - \hat{y}) + (\hat{y} - v)$ .

The first term in parentheses is in  $W^\perp$  while the second term is in  $W$ .

Therefore by the Pythagorean theorem  $\|y - v\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - v\|^2 > \|y - \hat{y}\|^2$  since  $\|\hat{y} - v\| > 0$ .  $\square$

**Fact.** Suppose  $u_1, u_2, \dots, u_p$  is an orthonormal basis of  $W$ . Then

$$\text{proj}_W(y) = (y \bullet u_1)u_1 + (y \bullet u_2)u_2 + \dots + (y \bullet u_p)u_p.$$

Define the matrix  $U = [u_1 \quad u_2 \quad \dots \quad u_p]$ . Then  $\text{proj}_W(y) = UU^\top y$ .

## 5 The Gram-Schmidt process

The *Gram-Schmidt process* is an algorithm that takes an arbitrary basis for some subspace of  $\mathbb{R}^n$  as input, and produces an orthogonal basis of the same subspace as output.

**Theorem.** Let  $W \subseteq \mathbb{R}^n$  be a nonzero subspace. Then  $W$  has an orthogonal basis.

(The zero subspace  $\{0\}$  has an orthogonal basis given by the empty set, but we exclude this trivial case.)

**Gram-Schmidt process.** Suppose  $x_1, x_2, \dots, x_p$  is any basis for  $W$ .

Then an orthogonal basis is given by the vectors  $v_1, v_2, \dots, v_p$  defined by the following formulas:

$$\begin{aligned} v_1 &= x_1. \\ v_2 &= x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1. \\ v_3 &= x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2. \\ v_4 &= x_4 - \frac{x_4 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_4 \bullet v_2}{v_2 \bullet v_2} v_2 - \frac{x_4 \bullet v_3}{v_3 \bullet v_3} v_3. \\ &\vdots \\ v_p &= x_p - \frac{x_p \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_p \bullet v_2}{v_2 \bullet v_2} v_2 - \dots - \frac{x_p \bullet v_{p-1}}{v_{p-1} \bullet v_{p-1}} v_{p-1}. \end{aligned}$$

These formulas are inductive: to compute any  $v_i$  you need to have already computed  $v_1, v_2, \dots, v_{i-1}$ .

More strongly, we can say the following. Let  $W_i = \mathbb{R}\text{-span}\{v_1, v_2, \dots, v_i\}$  for each  $i = 1, 2, \dots, p$ .

Then  $v_1, v_2, \dots, v_i$  is an orthogonal basis for  $W_i$  and  $v_{i+1} = x_{i+1} - \text{proj}_{W_i}(x_{i+1})$ .

(Our proof of the existence of orthogonal projections relies on this theorem.)

*Proof.* For  $i = 1, 2, \dots, p$  and  $y \in \mathbb{R}^n$  define  $\text{proj}_{W_i}(y) = \frac{y \bullet v_1}{v_1 \bullet v_1} v_1 + \frac{y \bullet v_2}{v_2 \bullet v_2} v_2 + \dots + \frac{y \bullet v_i}{v_i \bullet v_i} v_i$ .

We want to show that  $v_1, v_2, \dots, v_i$  is an orthogonal basis for  $W_i$  for each  $i$ .

If we assume that this is true for any particular value of  $i$ , then the formula  $v_{i+1} = x_{i+1} - \text{proj}_{W_i}(x_{i+1})$  automatically holds, which means that  $v_{i+1} \in W_i^\perp$  so  $v_1, v_2, \dots, v_i, v_{i+1}$  is also an orthogonal set, and therefore an orthogonal basis for  $W_{i+1}$ .

The single vector  $v_1 = x_1$  is necessarily an orthogonal basis for  $W_1 = \mathbb{R}\text{-span}\{v_1\}$ .

Therefore  $v_1, v_2$  is an orthogonal basis for  $W_2$ , which means that  $v_1, v_2, v_3$  is an orthogonal basis for  $W_3$ ; continuing in this way, we deduce that  $v_1, v_2, \dots, v_i$  is an orthogonal basis for  $W_i$  for each  $i = 1, 2, \dots, p$ . In particular  $v_1, v_2, \dots, v_p$  is an orthogonal basis for  $W_p = W$ .  $\square$

**Remark.** To find an orthonormal basis for a subspace  $W$ , first find an orthogonal basis  $v_1, v_2, \dots, v_p$ . Then replace each vector  $v_i$  by  $u_i = \frac{1}{\|v_i\|} v_i$ . The vectors  $u_1, u_2, \dots, u_p$  will then be an orthonormal basis.

**Example.** Suppose  $x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$  and  $x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$  and  $x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ .

These vectors are linearly independent and so are a basis for the subspace  $W = \mathbb{R}\text{-span}\{x_1, x_2, x_3\}$ .

To compute an orthogonal basis for  $W$ , we carry out the Gram-Schmit process as follows:

- We set  $v_1 = x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ . Then  $v_2 = x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$ .

- Finally let  $v_3 = x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$ .

The vectors  $v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ ,  $v_2 = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$ ,  $v_3 = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$  are then an orthogonal basis for  $W$ .

## 6 Vocabulary

Keywords from today's lecture:

### 1. Orthonormal vectors.

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

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

A set of vectors in  $\mathbb{R}^n$  is *orthonormal* if the vectors are orthogonal and each vector is a unit vector.

Example: the standard basis  $e_1, e_2, \dots, e_n$  of  $\mathbb{R}^n$  is orthonormal.

### 2. Orthogonal projection of a vector $y \in \mathbb{R}^n$ onto a subspace $W \subseteq \mathbb{R}^n$ .

The unique vector  $\text{proj}_W(y) \in W$  such that  $y - \text{proj}_W(y)$  is orthogonal to every element of  $W$ .

If  $u_1, u_2, \dots, u_p$  is an orthonormal basis for  $W$  then

$$\text{proj}_W(y) = (y \bullet u_1)u_1 + (y \bullet u_2)u_2 + \dots + (y \bullet u_p)u_p.$$

### 3. Orthogonal matrix.

A square matrix  $U$  whose columns are orthonormal. A better name for an orthogonal matrix would be “orthonormal matrix,” but this term is not commonly used.

Equivalently, a matrix  $U$  is orthogonal if and only if  $U$  is invertible and  $U^{-1} = U^\top$ .

Example: every rotation matrix  $\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$  is orthogonal.

### 4. Gram-Schmidt process.

A specific algorithm whose input is an arbitrary basis  $x_1, x_2, \dots, x_p$  for a subspace of  $\mathbb{R}^n$  and whose output is an orthogonal basis  $v_1, v_2, \dots, v_p$  for the same subspace. Explicitly:

$$\begin{aligned} v_1 &= x_1. \\ v_2 &= x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1. \\ v_3 &= x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2. \\ v_4 &= x_4 - \frac{x_4 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_4 \bullet v_2}{v_2 \bullet v_2} v_2 - \frac{x_4 \bullet v_3}{v_3 \bullet v_3} v_3. \\ &\vdots \\ v_p &= x_p - \frac{x_p \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_p \bullet v_2}{v_2 \bullet v_2} v_2 - \dots - \frac{x_p \bullet v_{p-1}}{v_{p-1} \bullet v_{p-1}} v_{p-1}. \end{aligned}$$