1 Last time: orthogonal vectors and projections

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

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

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

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 $u \bullet v = 0$.

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

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

The *orthogonal complement* of a subspace $V \subset \mathbb{R}^n$ 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 $(\mathrm{Col}\, A)^\perp = \mathrm{Nul}(A^T)$. We also showed last time that that $\dim V^\perp \leq n - \dim V$.

A list of vectors $u_1, u_2, \ldots, u_p \in \mathbb{R}^n$ 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 they span.

If $u_1, u_2, \ldots, u_p$ is an orthogonal basis for a subspace $V \subset \mathbb{R}^n$ and $y \in V$, then $y = c_1u_2 + c_2u_2 + \cdots + c_pu_p$ where the coefficients $c_1, c_2, \ldots, c_p \in \mathbb{R}$ are defined by

$$c_i = \frac{y \bullet u_i}{u_i \bullet u_i}.$$

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

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y_1e_1 + y_2e_2 + \cdots + y_ne_n$$

then $y = c_1e_1 + c_2e_2 + \cdots + c_ne_n$ 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 $c_i = y_i$.

Let $L \subset \mathbb{R}^n$ 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

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

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

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

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

Proof. If $u \in L$ is nonzero then $y - \text{proj}_L(y) = y - \frac{y \bullet u}{u \bullet u} u$ and it holds that

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

To see that $\text{proj}_L(y)$ is the only vector in L with this property, suppose $\widehat{y} \in L$ is such that $y - \widehat{y} \in L^\perp$. Then $(y - \widehat{y}) \bullet \widehat{y} = y \bullet \widehat{y} - \widehat{y} \bullet \widehat{y} = 0$ so $y \bullet \widehat{y} = \widehat{y} \bullet \widehat{y}$. But then $\widehat{y} = \frac{y \bullet u}{u \bullet u} u = \text{proj}_L(y)$ for $u = \widehat{y} \in L$. $\square$

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

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

Check that

$$\left(\begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix}\right) \bullet \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \bullet \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 0.$$

2 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 which is orthonormal.

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

Example. The vectors $v_1 = \frac{1}{\sqrt{11}} \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$, $v_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$, and $v_3 = \frac{1}{\sqrt{66}} \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix}$ form another 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 $U^T U = I_n$. If U is square then its columns are orthonormal if and only if $U^T = U^{-1}$.

Proof. Suppose $U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ where each $u_i \in \mathbb{R}^n$. The entry in position (i, j) of $U^T U$ is then $u_i^T u_j = u_i \bullet u_j$. Therefore $u_i \bullet u_i = 1$ and $u_i \bullet u_j = 0$ for all $i \neq j$ if and only if $U^T U$ is the $n \times n$ identity matrix. $\square$

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

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 $(Ux) \bullet (Uy) = x^T U^T U y = x^T I_n y = x^T y = x \bullet y$. $\square$

Somewhat confusingly, a square matrix U with orthonormal columns is called an *orthogonal matrix*.

3 Orthogonal projections onto subspaces

We have already seen that if $y \in \mathbb{R}^n$ and $L \subset \mathbb{R}^n$ is a 1-dimensional subspace then y can be written uniquely as $y = \widehat{y} + z$ where $\widehat{y} \in L$ and $z \in L^\perp$.

This generalises to arbitrary subspaces as follows:

Theorem. Let $W \subset \mathbb{R}^n$ be any subspace. Let $y \in \mathbb{R}^n$. Then there are unique vectors $\widehat{y} \in W$ and $z \in W^\perp$ such that $y = \widehat{y} + z$.

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

$$\widehat{y} = \frac{y \bullet u_1}{u_1 \bullet u_1} u_1 + \frac{y \bullet u_2}{u_2 \bullet u_2} u_2 + \cdots + \frac{y \bullet u_p}{u_p \bullet u_p} u_p \qquad \text{and} \qquad z = y - \widehat{y}. \tag{*}$$

It doesn't matter which orthogonal basis is chosen for W; this formula gives the same value for $\widehat{y}$ 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. Fix one such basis $u_1, u_2, \dots, u_p \in W$.

Define $\widehat{y}$ by the given formula. Then $\widehat{y} \in W$ and

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

for each $i = 1, 2, \dots, p$, so $y - \widehat{y} \in W^\perp$.

To show uniqueness, suppose $y = \widehat{u} + v$ where $\widehat{u} \in W$ and $v \in W^\perp$. Then $\widehat{u} - \widehat{y} = v - z$. But $\widehat{u} - \widehat{y}$ is in W while $v - z$ is in $W^\perp$, so both expressions must be zero as $W \cap W^\perp = \{0\}$. This means we must have $\widehat{u} = \widehat{y}$ and $v = z$. $\square$

Definition. The vector $\widehat{y}$, 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 usually write

$$\mathrm{proj}_W(y) = \widehat{y}$$

to refer to this vector.

Corollary. If $W \subset \mathbb{R}^n$ 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 $\mathbb{R}^n$. This size of such a union is at most $\dim W + \dim W^\perp$ (since $\dim W$ and $\dim W^\perp$ are the sizes of the two bases that we are combining) and at least n (since fewer than n vectors cannot span $\mathbb{R}^n$), 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 $\dim W^\perp = n - \dim W$. $\square$

<u>Properties of orthogonal projections onto a subspace $W \subset \mathbb{R}^n$.</u>

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

Proposition. If $v \in W$ and $y \in \mathbb{R}^n$ and $v \neq \mathrm{proj}_W(y)$ then $\|y - \mathrm{proj}_W(y)\| < \|y - v\|$. In words: the projection $\mathrm{proj}_W(y)$ is the vector in W which is closest to y.

Proof. Let $\widehat{y} = \mathrm{proj}_W(y)$. Then $y - v = (y - \widehat{y}) + (\widehat{y} - v)$. The first term in parentheses is in $W^\perp$ while the second term is in W. Therefore by the Pythagorean theorem we have

$$\|y - v\|^2 = \|y - \widehat{y}\|^2 + \|\widehat{y} - v\|^2 > \|y - \widehat{y}\|^2$$

since $\|\widehat{y} - v\| > 0$. $\square$

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

$$\mathrm{proj}_W(y) = (y \bullet u_1)u_1 + (y \bullet u_2)u_2 + \cdots + (y \bullet u_p)y_p.$$

If $U = \begin{bmatrix} u_1 & u_2 & \ldots & u_p \end{bmatrix}$ then $\mathrm{proj}_W(y) = UU^T y$.

4 The Gram-Schmidt process

The *Gram-Schmidt process* is an important algorithm which 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 $W \subset \mathbb{R}^n$ 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 $x_1, x_2, \ldots, x_p$ is any basis for W. Then an orthogonal basis is given by the vectors $v_1, v_2, \ldots, v_p$ defined by the following formulas:

$$v_1 = x_1.$$

$$v_2 = x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1.$$

$$v_3 = x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2.$$

$$v_4 = x_4 - \frac{x_4 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_4 \bullet v_2}{v_2 \bullet v_2} v_2 - \frac{x_4 \bullet v_3}{v_3 \bullet v_3} v_3.$$

$$\vdots$$

$$v_p = x_p - \frac{x_p \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_p \bullet v_2}{v_2 \bullet v_2} v_2 - \cdots - \frac{x_p \bullet v_{p-1}}{v_{p-1} \bullet v_{p-1}} v_{p-1}.$$

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

More strongly, we can say the following. Let $W_i = \mathbb{R}\text{-span}\{v_1, v_2, \ldots, v_i\}$ for each $i = 1, 2, \ldots, p$. Then $v_1, v_2, \ldots, v_i$ is an orthogonal basis for W_i, and $v_{i+1} = x_{i+1} - \mathrm{proj}_{W_i}(x_{i+1})$.

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

To avoid circular arguments, define

$$\mathrm{proj}_{W_i}(y) = \frac{y \bullet v_1}{v_1 \bullet v_1} v_1 + \frac{y \bullet v_2}{v_2 \bullet v_2} v_2 + \cdots + \frac{y \bullet v_i}{v_i \bullet v_i} v_i$$

for $i = 1, 2, \ldots, p$ and $y \in \mathbb{R}^n$.

We want to show that $v_1, v_2, \ldots, v_i$ 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 $v_{i+1} = x_{i+1} - \mathrm{proj}_{W_i}(x_{i+1})$ automatically holds, which means that $v_{i+1} \in W_i^\perp$ so $v_1, v_2, \ldots, v_i, v_{i+1}$ 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:

1. First let $v_1 = x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.
2. Next let $v_2 = x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$.
3. 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

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

then form an orthogonal basis for W.

We note one final result related to the Gram-Schmidt process.

Theorem (QR factorization)**.** Let A be an $m \times n$ matrix with linearly independent columns. Then $A = QR$ where Q is an $m \times n$ matrix whose columns are an orthonormal basis for $\mathrm{Col}\, A$ and R is an $n \times n$ upper-triangular matrix with positive entries on the diagonal.

One calls the decomposition $A = QR$ a *QR factorization* of A.

Proof. Let $A = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$ where each $x_i \in \mathbb{R}^m$.

Perform the Gram-Schmidt process on $x_1, x_2, \dots, x_n$ to get an orthogonal basis $v_1, v_2, \dots, v_n$ for $\mathrm{Col}\, A$.

Then define $Q = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ where $u_i = \frac{1}{\|v_i\|} v_i$ for $i = 1, 2, \dots, n$.

These vectors have the property that $\mathbb{R}\text{-span}\{u_1, u_2, \dots, u_k\} = \mathbb{R}\text{-span}\{x_1, x_2, \dots, x_k\}$ for each $k = 1, 2, \dots, n$, and $x_i \in \|v_i\| u_i + \mathbb{R}\text{-span}\{u_1, u_2, \dots, u_{i-1}\}$. It follows that $A = QR$ for an upper-triangular matrix R of the desired form. $\square$

5 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 $W \subset \mathbb{R}^n$.

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

   If $u_1, u_2, \dots, u_p$ is an orthonormal basis for W then $\mathrm{proj}_W(y) = (y \bullet u_1)u_1 + (y \bullet u_2)u_2 + \dots + (y \bullet u_p)y_p$.

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 $U^{-1} = U^T$.

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

4. **Gram-Schmidt process**.

   An 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:

$$v_1 = x_1.$$

$$v_2 = x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1.$$

$$v_3 = x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2.$$

$$v_4 = x_4 - \frac{x_4 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_4 \bullet v_2}{v_2 \bullet v_2} v_2 - \frac{x_4 \bullet v_3}{v_3 \bullet v_3} v_3.$$

$$\vdots$$

$$v_p = x_p - \frac{x_p \bullet v_1}{v_1 \bullet v_1} - \frac{x_p \bullet v_2}{v_2 \bullet v_2} - \dots - \frac{x_p \bullet v_{p-1}}{v_{p-1} \bullet v_{p-1}} v_{p-1}.$$